[SQUEAKING]

[RUSTLING]

[CLICKING]

PROFESSOR: So next, we will-- so we're switching away from metric spaces, and we talk about differentiation. So if we have a function on R, say, and then we say that x that f is differentiable at x0, if the limit here f of dx minus f of x0, divided by x minus x0, if this limit here exists as x goes to x0.

This quotient here is usually called the difference quotient. And it's always, of course, when you form this difference quotient, it's always assumed, obviously, that x is not equal to x0 so that the denominator makes sense. So if this limit here, as x goes to x0, if this exists, then we say that the function f is differentiable at x0. And the limit is called the derivative of f at x0. And this is how it's denoted.

So now the simplest possible example of this would be that suppose that you're looking at a function that is constant. So f is equal to f for some constant C. Then if you're looking at this difference quotient, so f of dx minus f of x0 x minus x0, well, the value at any point is C-- so this is C minus C divided by x minus y, x0-- and so the denominator is 0, so this whole thing is 0.

And so of course, the limit-- this is 0 for each x and also each x0. So in particular, of course, the limit here exists. So the limit here of the difference quotient exists, and it's equal to 0 because it's 0 for all x for a given x0. So that would be the simplest possible example.

The second simplest example. Second most simple example would be that you're looking at a linear function. So let's say that f of x is equal to x. In this case, the difference quotient here-- well, this here is x. This here is x0. So this here is for each x and x0. This is 1. So the limit of course, in this case, also exists. The limit here exists, and it's equal to 1. And so in this case, f is differentiable at all points. And the derivative at all points is equal to 0. In this case, f here is also differentiable. And the derivative at all points is equal to 1.

So now, in general, how do we compute the-- so how do we compute the derivative of a function? How do we compute the derivative? And for that, there are some basic laws-- how to compute it. So there's these laws-- the sum rule, what's usually called the Leibniz rule, the quotient rule, and then also the chain rule.

Now let me look at the first three rules first. So I have two functions, f and g. They are defined, say, on R, but it could have been a subset of R. And I'm assuming that they are both and both-- actually, we'll need one. So both f and g are differentiable at x0.

Then we have the following. So we have the following lemma. One is we have that-- so 1, and this is really the sum rule. We have that the sum here f plus g-- so the sum of these two functions, this is the function that is just in x. It's just the sum of f of x and g of x We have that this function here, f plus g, is differentiable at x0. And the derivative, f plus g, the derivative of this at x0 is just the sum of the derivative.

Let me prove this one here, by the way. So let's prove this. So we have the sum-- so we are going to look at-- so we formed the difference quotient, so f plus g at x minus f plus g at x0 x minus x0. So now what is this? So this is just f of x plus g of x minus and then f of x0 plus g of x0, and then divided by x minus x0. And so then you can split this up into two. You can write this difference quotient as if you split it up if you collect the f's in the numerator. So I just collect the f's. And then, afterwards, I collect and I put it on two different fractions, separate them out, and then I collect the g's.

And now you see that because f is differentiable at x0, this thing here, as x goes to x0, this here goes to f prime. So this here goes to f prime of x0. This is what this difference quotient here. And this difference quotient goes to g prime of 0. So you have that indeed, it does converge. And this is what it converges to. And so this proves that the sum of these two functions is also differentiable, and their derivative is the sum of the derivative.

The next thing we will prove is Leibniz rule. But before that, I need to step back and I want to-- so maybe I'll just state it and then we'll go back and we need something. The next is what's called Leibniz rule. So you could also call it a product rule, but it's really quite generally called the Leibniz rule.

And so this is when you're looking at the products of these functions. So the product function is a function in x that is just the product of f of x and g of x.

And Leibniz rule says the following, that if f and g are differentiable at x0, then the product of these two function here is also differentiable at x0, and the derivative of the product is the derivative of the first function at x0 times the value of the second function at x0 plus the symmetric thing in f and g, which is the value of f at x0 times the derivative of g at x0.

So the product is differentiable at x0 also. And this is what the derivative is. And in the notes, I just copied a little bit of thing just for fun from Wikipedia about Leibniz that was kind of an interesting figure and played an important role a long time ago.

So before I want to prove this, let me step back and let me make an observation about what it means to be differentiable at a point. So this is the next lemma. So you have a function f on R into R. And you assume that f-- so if f here is differentiable at x0, then I claim that then f here is also continuous at x0. So in other words, being differentiable is stronger than being continuous. In a minute, we will see an example of something that's continuous but not differentiable.

And so now let's try to prove that. So I'm assuming that it's differentiable. So this means that this difference quotient here has a limit. And that limit is the derivative as x goes to x0. So this means that there exists a delta greater than 0 such that if x here minus x0 is less than delta y-- of course, I'm assuming that it's not equal to 0. If it is less than, the difference between x and x0 is less than delta, then this difference quotient here, then f of dx minus f of x here x minus x minus x0 minus f prime of x0-- this thing here have to be less than 1. I mean, this is just using that this limit exists, and I'm using-- in place of epsilon I'm using 1.

Now so this means just multiplying over and multiplying over. You get that f of x minus f of x minus 0 minus f prime of x0, this thing here-- and I'm multiplying both sides, so multiply both sides by x minus x0. This is a positive thing. So multiplying both things by that, so then I get this thing here, x minus x0. And then I get less or less strictly less than x minus x0 because this is strictly positive because x is not equal to x0. So I have that.

And so I get that-- so now I just split it up so I have that f of dx minus f of x0 is now less than x minus x0 plus f prime of x minus 0 times x minus x0. I mean, this is just the triangle inequality that we've used many times, just depend on R.

And so now you see that if you now-- and this here, of course, I can also write like this. I can just factor this out, and I can write it as 1 plus f prime of x0 norm, like this. That's fine.

And so now if I want this thing here-- so now you want to prove that f is continuous at x0. So you want this here to be less than some given epsilon. So given epsilon greater than 0, you let delta 1 be the minimum of-- if you set delta 1 to be the minimum of the delta that I already chose, so then I have this inequality, and then epsilon over 1 plus f prime of x0. So if I choose delta like this, then I have this inequality. I fill in that.

So now I'm assuming that this thing here is less than delta 1. So again, I have this inequality because it's less than delta. So then when I fill in here delta 1 for this, then I see that this thing here times this is less than epsilon. You see this holds when x minus x0 is less than delta 1. Again, because if it's less than delta 1, it's less than delta, so I have this inequality. And then I can use that delta 1 since this is less than this. And then I get this inequality.

So this is really what we want to prove. We want to prove that if it's differentiable, then it's also continuous. And the continuity here, I proved it with epsilon and delta that really should prove it.

Now let's look at some examples. So this means that if a function is differentiable, then it's continuous. Let me give you an example of a function that is -- so let me give you three examples. The first example is maybe the simplest one. So this is the function f of x that is equal to the absolute value of x. So f here is defined on R into R. And it's just the absolute value. And so the absolute value here we've seen before. So f here is continuous. It's continuous at all points, in particular, of course, at 0.

But if you now form the difference quotient. Suppose that x here is bigger than 0. And now look at the difference quotient. So the difference quotient-- quotient at x0 equal to 0 if you're looking at that difference quotient. And you first look at where x is positive. If x is positive, then f of x minus f of 0-- x minus 0, this difference quotient here. If x is positive, then this thing here is just x. So this is x minus 0, x minus 0. So this here is 1. This difference quotient is 1. This is when x is positive.

When x is negative, then f of x-- if x is negative, then the absolute value is minus x. So the difference quotient here f of x minus f of 0, which is 0, x minus 0-- this here is just minus x divided by x. But this thing here is minus 1. So if you're looking at the difference quotient that comes from looking at the derivative is 0, then when you approach 0 from the right, then the difference quotient is 1.

When you approach it from the-- and when you approach it from the left, then the difference quotient is minus 1. So it doesn't have a limit. It has a limit from the right from the left, but it doesn't have a limit if you don't restrict where it's coming from. So this function here is not differentiable. Yeah?

AUDIENCE: Can you define what it means when they have a function of x [INAUDIBLE]?

PROFESSOR: So we did this earlier in the class. So that's the difference quotient here that this thing here is converging to some number, say a. This here means that for all epsilon greater than 0, there exists a delta greater than 0 such that if x minus x0 is less than delta and, again, x is assumed to be different from x0. If you have that, then this difference quotient here minus this a is supposed to be less than epsilon. And that's this.

And so you see, again, in this example that we just looked at, where in the example that we just looked at-- so in the example where f of x was equal to the absolute value of x, then if you were graphing it, then this looks like this. And we saw that the difference quotient when you approach it over here, the difference quotient was 1. When you approach it over here, the difference quotient is minus 1.

So if you take a tiny little interval around 0, then the difference quotient could be either minus 1 or 1. But however small you make this little interval, it's not like that it converges to some number because it has-- it is both. There are both cases where it's 1 and cases where it's minus 1. So that's like the first example. And again, you know of a function that is continuous, but it's not differentiable at 0. In this case, it was a 0.

And maybe a little bit more interesting example, along the same line, is the following. So this is, again, a function on R into R, and I set f of x equal to x times sine of 1 over x. This here is not really defined at 0. So I need to define it also at 0. So when x is equal to 0, then I just set it equal to 0. And otherwise, I set it equal to x times sine of 1 over x.

And so the thing is that this function here, so f here, is continuous. In particular, it's continuous at 0. It's actually continuous everywhere, but I'm mostly interested in it at 0. The graph of this-- so remember that sine function looks like this, That at 0, at pi, at 2 pi, at 3 pi, and so on. And all of these-- or at all integer multiples, also, for the negative ones, at all integer multiple of pi, sine function vanishes. And in between here, it is at pi over 2, it achieves its maximum, which is 1. And so the function looks like this. So this here is 1. That's how a sine function looks like.

And so now if you're looking at sine of 1 over x, that's going to be a function as you're going down. So if you're looking at this-- here's a sine of x. But if you're looking at sine of 1 over x, as x goes to 0, this becomes very large. So this means that as you're going to 0, it goes very rapidly through these periods, where it oscillates between negative 1 and 1. So the value here is minus 1.

And so it will look like this that it will again fluctuate. It will fluctuate, but it will fluctuate much more rapidly. So if this here, say, was 1 over pi. And 1 over pi is the same as pi, there, it would be 0. And it would like fluctuate. So it will look like something like this, but it will fluctuate. And then, as you're going away, it will fluctuate less often. So as you're going down towards 0, it will fluctuate a lot.

Because sine always is between minus 1 and 1, the value here will always be between minus 1 and 1. But as you're going towards 0 from the right or from the left, it's going to fluctuate more and more. As you're going away from 0, it's fluctuating less and less.

So this is the function I'm interested in. But when you multiply through by x, then you see, when you multiply through by x, it's fluctuating, but it's always fluctuating between minus 1 and 1. If you multiply it by x, then as you're going to 0, this here will squeeze down, and the fluctuation will be 0 at 0. So it's easy to see that this function here-- so is indeed continuous.

But let's look at the difference quotient. So the difference quotient-- so if we look at the difference quotient, so we're looking at-- and we just look at the difference quotient at 0. So this is the difference quotient. And so it's x sine 1 over x. The value at 0 by definition was 0. And so you see that now this here, this here becomes just x sine 1 over x divided by x. And so you see that this here cancel with that. So you just get sine 1 over x.

And so you see that now, as x goes to-- so when x here goes to 0, this here fluctuates. This here fluctuates between minus 1 and 1. But it doesn't have a limit because it's fluctuating wider and wider. So this is an example again-- this is another example of a function that is continuous at 0 but not differentiable at 0. It's actually much worse than the previous. The previous example that was-- this example is much more tame. It is not differentiable at 0. But if you were to restrict to just positive things, then it would be differentiable.

You can also define derivative-- so, in general, when we define derivative, we let it be defined in a neighborhood of that point, so on either side. But if you were considering where you were, allowing the function just to be defined at this point and then to the right, say, or to the left, then this function would be differentiable. But in this example here, it doesn't matter if you restrict the interval to just to the right or to the left. It's still not differentiable.

So this says something about this example, something about that it fluctuates wildly and that is causing it not to be differentiable. Again, if you defined it-- if you didn't have this kind of-- the x here is serving as some of-- it's dampening how much it's fluctuating. So if there wasn't an x here, then this function wouldn't-- it would just fluctuate, and it wouldn't even be continuous at x0. But once I put this x on, then it becomes continuous.

If I put another x, so if I'm looking at f of x equal to x squared times sine 1 over x. Suppose I'm looking at this function. So now I have dampened the fluctuation even more. And so if I came looking at the difference quotient-- so I'm looking at this function. I should be more careful. So again, at 0, I need to define it separately. So I define a function that is 0 at x equal to 0. And then it is x squared times sine 1 over x otherwise.

If I'm looking at this function, then I claim that this function is, in fact, differentiable at x at 0. And so let's look at that. So again, I'm looking at the difference quotient. And the value at 0 is 0. So I have this. And so this here is now just x squared sine 1 over x. This thing here is then x. And so you see that this difference quotient is just this. But this here does converge to 0 as x goes to 0. So this function is actually differentiable. It's fluctuates, but it's more mild the way it fluctuates.

Now let me get back to the rules. So we proved one rule that was this sum rule. Then the second rule we're interested in is when we're looking at the Leibniz rule. So we're looking at the product of these two functions. And each function is assumed to be differentiable at x0. And then we claim that the product function is also differentiable at x0. And this is the formula for the derivative. To prove this one here, I need these facts that we have now proved that if it is differentiable, then it's also continuous at that point. So I'm going to use that fact.

So this is proof of Leibniz rule. So we have these two functions. We're looking at f of x. We form the difference quotient. So this is the difference quotient, x minus x0. We do this trick that we've done a bunch of times before, when we also looked at products if we had two sequences, and we looked at the sequence that was the product of these two.

So just a reminder that-- I mean, it's not like-- so remember that if you had a sequence an and a sequence bn, and you knew that an converges to a and bn converges to b, then the product here converges to a times b. And when we prove that, we use this little trick that we're splitting the difference between this and this. And so I'm going to do exactly the same trick here.

So I'm looking at this thing here. And so I write this thing here as f of x times g of x minus-- and it doesn't really matter where you're inserting it. So this here is splitting the difference. And then I have to add it again. So this is certainly the same, just subtracting it and adding it right back.

And now I split this into two. So I write this thing here. So this fraction, I split into two. And the first one, I take this one here together. And you can factor out a g of x. You can write it like this, f of x minus f of x0, x minus x0. So I'm just here. I'm factoring out the g of x, and then I'm left with this. And this is keeping the denominator, like that. This is this part here.

And then I have the other part here. But there, you see that the f of x0 is factoring out. And then I have g of x minus g of x0, x minus x0, like that.

But now you see that as x goes to x0, well, g was continuous at x0. So this thing here, as x goes to x0, this thing here goes to g of x0-- this one here. That was because g was continuous. Because it was differentiable, so it was also continuous. This here goes to the derivative. And the second thing here this is just constant. And this goes to the derivative, like that.

So what we have proven this here is as x goes to x0. So you see that we have proven that this difference quotient here for the product has a limit. And this is what the limit is. And so this means that the product is differentiable. And this here is the derivative. And that's exactly the Leibniz formula, which was this over here, this formula here.

The next thing is that we want to prove the next rule is like this quotient rule. So let me prove the quotient rule. So Leibniz and Newton was the one that was responsible for calculus.

And so then the next thing I want to prove is this quotient rule. And so, again, you have these two functions, f and g. And they are differentiable at x0. And I'm also assuming that g here at x0 is not equal to x. Let's just assume that g is not equal to 0 everywhere. It's different from 0 everywhere.

And then the quotient rule says that if you're looking at the quotient here and you take the derivative of the-- so it say that if you're looking at the quotient here-- so the quotient will say that if you're looking at this thing here, this now makes sense because g does not vanish. This here is differentiable at x0.

And the derivative of this function at x0 is given by the following formula, that it is the derivative of f at x0 times g of x0 minus f at x0 times g prime at x0 divided by g squared. So g of x minus 0 times itself-- I'll write that as a square. Write this here again-- this here, this is just g of x0 times g of x0.

So now let's try to prove that. So we're looking at f of x like this. At x minus-- so this is the difference version-- at x0 divided by x minus x0. That's the difference equation. And so written out, this is f of x g of x minus f of x0 divided by g of x0 x minus x0.

So let me put this on a fraction with a common denominator. So if I do that, then this thing here is-- so then I'm multiplying by g of x0 both in the numerator and the denominator here. So I'm getting f of x times g of x0. And then here, I'm multiplying by g of x in both numerator and denominator, so getting this thing here, g of x. And then it becomes x minus x0 and then g of x, g of x0, like that.

And so now you can write this thing here as-- and so now you do the same trick as before. I mean, I could have gotten a way slightly easier with looking at a special case of this quotient rule, where f was a constant function 1. So I could do that first. And then the general case, I could then from-- get from this special case together with Leibniz rule. It's most often I want to do it together anyway, but if you have to do it together, you have to use the same trick that we just did.

So we're using that over here, so f of g of x0 minus and then f of x0, g of x0, Plus f of x0 g of x0 minus f of x0 g of x-- so this is, again, where I'm subtracting in the denominator or something, and then I'm adding it back up immediately. So I have this, like that.

And so now if you do that, then you see that-- and now you split it up, again, into two fractions. So we split it up. And so this thing here, this thing here, but there, you see that there's a g of x here that factor out. So this is equal to g of x0. And then you have f of x minus f of x0 and then x minus x0, g of x, g of x0, like that. And for the second one, you see that there, you also have something that factor out. And this is f of x0, like that. And you're getting g of x0 minus g of x over x minus x0, g of x, g of x0, like that.

And so now I can make this looks nicer. I can, this one here, cancel with that. And so I can write this thing here as 1 over g of x f of x minus f of x0 divided by x minus x0. So let's take care of this here. Here, I will write it as I will take this one here out. So I'll write it as f of x0 of x times g of x0.

And then I'm left with this thing here. And what I don't love about this thing here is it's not quite the difference quotient because the order of these two is wrong. So I switched the order. But if I switch the order, then I have to pick up a minus sign. So I get minus here. And now it is g of x, g of x0, x minus x0, like that.

And now you see that as x goes to x0, so now as x goes to x0, well, this one here, the function was differentiable. So it's also continuous. So this goes to 1 over g of x0. This one here just goes to the derivative at x0. This one here goes to-- again, because the function is continuous, this goes to minus f of x0. And then it's what I wrote before as g squared. It's g of x0 times itself. And then this thing here goes to the derivative of x0 at 0.

And now if I want to, I can rewrite this thing here. If I want to put it back on the common denominator, then I have to multiply in the numerator and denominator by g of x0. If I do this, I get f prime of x0, g of x0, and then I get g squared of x0 in the denominator. And then here it's just those two minus f of x0 g prime of x0.

And so you see that this here is exactly this formula here, this quotient rule. So this proof, that if you take two functions, the one you have in the denominator, if that is nonvanishing, So You can form the quotient. And if each of these two functions is differentiable, then the quotient is also differentiable. And this is the formula for the derivative.

So now why is this already really, really useful? So we calculated-- so we started with the example of where f is a constant. And we saw that it's differentiable, and the derivative is 0. Then the second example we had was where the function was x, and we saw that it was differentiable and the derivative is 1. If you now take-- if you now take ax times-- so let's just take ax as the next example.

This is like a constant function times this function. So Leibniz rule-- so if f of x is equal to a times x, where A is a constant, Leibniz will say that the product of two differentiable function is differentiable, and the derivative of this is the derivative of that. But the derivative of that, that's a constant function. So that's 0. So I don't even have to think about that. So it's just the a times the derivative of the second function. But that's one. So this here is the derivative of that.

And the next example would be that you're looking at ax plus, say, b. f of x is equal to this. Then you can use the sum plus-- we know what each of these is. So you get that this one here is also differentiable. And the derivative of this function here is also just a. And you can continue like this, that if you take-- and if you now have a polynomial, suppose you take a to the power n. So if f of x is x to the power n, then you can think about this thing here as x times x times x times x. And you can use the Leibniz rule.

And so if you use the Leibniz rule, so you just do it first in the case where f of x is equal to x squared, you can think about this as x times x. These are two differentiable functions. So the derivative here is now the derivative of this. So that's 1 times the other function. Plus you keep this one here and take the derivative of that. So this is this. So this is 2 times x. And so if you take this function here using Leibniz rule. In this case, the derivative is differentiable, and the derivative is n times x to the power n minus 1.

And in fact, even more generally, just using the sum and the Leibniz rule, if you take any polynomial an xn plus an minus 1 xn minus 1 down to the last thing is being a constant-- if you take a polynomial like this just using the Leibniz rule and the sum rule, using those repeatedly, we know that this is differentiable and the derivative is just like the sum of the derivative of these.

So this is going to be an times n times x to the power n minus 1 plus n minus 1 times n minus 1 xn minus 2. And then this one here, the derivative is 0. So the last one you're getting something in nontrivial form is a1. So this is just using sum and Leibniz rule. And so this is just using the sum plus Leibniz rule repeatedly.

Now likewise, if you take a rational function, so meaning that if you take something-- if f of x is equal to some polynomial-- so let's go p of x divided by q of x, so where this here is a polynomial and so is that. So now you can use the quotient rule.

So the quotient rule will tell you that as long as you are looking at point where this here does not vanish, so then the quotient here is also differentiable, and it gives you a formula for the derivative in terms of the derivative of each of these two. And each of those two, they can be calculated like that. So you see, already now, we have quite the derivative of a lot of functions. We have of all function and of all rational functions. Rational functions, again, are quotients of a polynomial by other polynomial.

So now the next rule that we want to talk about is the chain rule. So what is the chain rule? So the chain rule is a little bit different from those three rules here. And it is where you take the composition of functions. So the next is the chain rule. And again, this is about composition of functions.

So I have a function f that goes, say, from-- to say that it goes from some interval a to b into another interval from c to d. And then I have a function g that goes from this interval here into R because g is defined whatever the image of f always, if it's in the set, this is where g is defined.

Then you have that you can form if you take some x in here-- so if x here is in a to b, then this here makes sense. And this is now an element here. So this is where g is defined. So this here makes sense. And this is now what we call the composition. And we write it like this. It's important to write it in the correct order. And so this is the composition function.

So now the chain rule is the following. So the chain rule is a foreign theorem. Then if f is differentiable at x0 and g is differentiable at y0, where y0 is equal to the value of f at x0, then the composition-- then the composition of these two functions is differentiable at x0, and the derivative is equal to the derivative of g at y0 times the derivative of f at x0. So this is the chain rule. The chain rule says that the composition is also differentiable at x0, and this is the formula for the derivative. Let's try to prove that.

And so, in this case, it's useful to divide up to two cases. And the two cases to divide up to-- so let me just look at a special case first, and then once we understand that, the general case will be clear.

So let's make one assumption. So let's try to prove it. But we assume first that the derivative-- I'm going to assume that the derivative at x0 is not equal to 0. So I'm assuming that, first, what does this guarantee me? Well, remember that the derivative is the limit of this difference quotient. So this is the limit. If this here is not equal to 0, this means that when x is very close to x0, this thing here is very close to this thing that is not equal to 0.

But then it means that if you are very close to x0, f of x cannot be equal to. So if you have this here, this implies that f of x is not equal to x0 for-- sorry, it's not equal to f of x0 for x sufficiently close to x0. I'm always assuming, of course, that x is not but x is not equal to 0. So I'm going to first make this assumption. And if I make this assumption, what I get out of it is that this thing here is not equal to that. Why is this useful? Because now I can actually look at this difference and I can divide by it.

So I'm going to just to make things a little bit clearer. So I have my function. So this here is f. And here I have g. Here, I have x0. But under f, it's taken to this, which is f of x0. But this I also will denote by y0. So I'm denoting just to make things clearer, y0 is just the value here.

And then I will use for-- f of x I will denote by y. So now I'm looking at the difference quotient. So I'm looking at f of-- sorry, I'm looking at g of f of x minus g of f of x0 x minus x0. That's what I'm supposed to look at. And I can write this thing here as g of y because this here I also call y. So you can write it like this g of y0. And then I will write it as y minus y0, but I didn't have this in the denominator up here, so I have to multiply it back on. And then I still have to divide by that. So I can write it like this.

This here I'm happy with. Get to this in just a second, but I'm happy with this expression. The other one I really want to put it back in, in terms of f. So I'm writing this in here. And this thing here, I really want to put it back in terms of f. So if I do that, I write it as f of dx minus f of x0 x minus x0.

But now you see that as x goes to x0, well, as x goes to x0, because the function f was continuous at x0, it means that if x goes to x0, then the value at x, which is y, will go to the value at x0. So when x goes to x0, we have that y goes to y0. So this means that this thing here, this here converges. That's just because the function f is differentiable. So this here will converge to the derivative. So the second factor here is converging to f prime at x0.

The first factor is-- but now you see that we just argued that y goes to y0. So this is also a difference quotient. But it's a difference quotient for g. And since y goes to y0, this difference quotient here goes to the derivative of g at y0. But that's exactly the formula that we wanted to prove, except that we made one assumption that the derivative of f at x0 was not equal to 0. We made that assumption only so that this number here was not equal to 0. And so we could divide by it.

But actually, otherwise, if you, for some reason, if y was not equal to y0, but then this formula here would still hold. So it's only the possibility that y could be equal to y0. But now you see that if you were just looking at the difference quotient g of f-- g of y minus-- sorry, g of f of x minus-- so this is the difference quotient. That is the one we really have to look at for the composition. And this is what we also call-- this thing here is what we called y0. This thing here is what we call y.

And the only case we haven't dealt with is where f prime of 0 is equal to 0, and y is equal to y0. This here, again, was f of x. And this here was equal to f of x0. So this is only this case here we haven't dealt with. But you see that if it so happened that f of x was equal to f of 0, then this difference quotient here, the denominator would trivially be 0.

And so you want to prove we have that if this here was not equal to 0, then we have that this is converging to the derivative. And the only case we have to look at where this derivative is 0-- so we are claiming that-- so the only case we are left with that we haven't dealt with is, first of all, where the derivative-- so let's see.

So let me just write down here clearly. The remaining case for the chain rule-- so the many cases that the derivative at x0 is 0. And we have already observed that if the derivative is 0 but y is not equal to y0-- if this is the case, then actually g of f of x minus g of f of x0-- this thing here still converged. In this case here, where this is the case, this has still converged to 0.

So we still have the chain rule in that case if this here was not the case. The only thing we have to worry about is what if y was equal to y0? But if y was equal to y0, when we're looking at the difference quotient. This is the difference quotient, after all. And so in that case, this thing here is y0. This is y0. But in that case, the difference quotient is actually zero already. So of course, it's going to be very close to 0. It actually is 0.

So you see that the remaining case, that's trivial. But we wanted to first exclude it because we wanted to be able to divide by this difference. And then once we have seen that under that assumption, we got the formula we wanted, then we go back and then we say that not only the remaining issue, first of all, was the case where the derivative was 0, but in fact, the only case that was potentially a problem was when y was equal to y0. But for that, the difference quotient was actually trivially 0. So that's the chain rule.

And so we only have a few minutes left. And I won't quite be able to get through all of the stuff. But I can get through one more lemma. And then I'll pick up from here next time. The notes are already posted. But so the last bit of the notes we'll do next time.

So the following lemma here-- so if you have a function, say, from R to R, and you have that x0 is a local max or min-- so f is a function from R to R that is differentiable. And x0 is a local max or local min for f, then the derivative at x0 is equal to 0.

Let me remind you what a local max and local min is. So if you take a function-- so local max, similarly for local min, if you have a function here, you have here x0, a local max means that in a neighborhood of this point, then the value at x0 is the maximum.

So local max at x0 means that there exists delta greater than 0 such that if you're looking at the interval to either side of x0, this interval, and you're looking at f here restricted to this subset, then x0 is a local max-- is a max. So this means that f of x0 is equal to the max of f, where x here is in this little interval here, like that.

So the idea is that the function may look like this. And this here is x0. So you see that in a neighborhood like this. It looks like that you have that this x0 is a maximum. But that might even be somewhere else, where the value is much higher. So that's a local maximum, and similarly for a local minimum.

And so I won't be able to actually prove this lemma. We will prove that next time. And then once we have that lemma, we'll prove what's called Rolle's theorem. And then we will use it to prove the mean value theorem. But I'll do that next time. But it's actually in the notes that I posted already.