

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high-quality educational resources for free. To make a donation or view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at [ocw.mit.edu](http://ocw.mit.edu).

**HERBERT  
GROSS:**

Hi. Today we do a somewhat computational bit. And actually, the lecture for today is not nearly as difficult, once you get through the maze of symbolism, as it is to apply the material. In other words, we're going to devote the next two units of our course to this particular topic, which is known as the chain rule.

But we'll give one lecture to cover both units. And again, the idea is that it's not so much that the concept becomes more difficult as much as it is that you must develop a certain amount of dexterity keeping track of the various partial derivatives and the like. At any rate, maybe I think the best way is to just barge into a hypothetical situation and see what the situation really is.

The idea is essentially the following. We're given some function, say,  $w$ .  $w$  is a function of, say, the three independent variables  $x$ ,  $y$ , and  $z$ . Now, for some reason or other, which we won't worry about right now, it turns out that  $x$ ,  $y$ , and  $z$  are, in turn, conveniently expressible in terms of the two variables  $r$  and  $s$ . In fact, if you want a physical interpretation of this, you can think of if  $x$ ,  $y$ , and  $z$  are functions of the two independent variables  $r$  and  $s$ , that means that we have two degrees of freedom. So we may think of this as parametrically representing the equation of a surface.

And what we're talking about here is  $w$  being a function of something in space and asking, what does  $w$  look like when you restrict your space to a particular surface? I mean, that's just a geometrical interpretation that one could talk about. But the idea is the following. After all, if  $w$  depends on  $x$ ,  $y$ , and  $z$ , and  $x$ ,  $y$ , and  $z$  each depend on  $r$  and  $s$ , in particular then, it's clear that  $w$  itself is some function of  $r$  and  $s$  where, again, I use the usual notation of using a  $g$  here rather than the  $f$  up here to indicate that the relationship between  $r$  and  $s$ , which specifies  $w$ , may very well be a different

algebraic relationship than that which relates  $x$ ,  $y$ , and  $z$ , to give  $w$ .

But the point that we have in mind is the following. Given that  $w$  is a function of  $x$ ,  $y$ , and  $z$ , given that  $x$ ,  $y$ , and  $z$  are functions of  $r$  and  $s$ , hence  $w$  is a function of  $r$  and  $s$ . The question that we ask in calculus of several variables is, first of all, if we can be sure that these were all continuously differentiable functions, can we be sure that  $w$  will be a continuously differentiable function of  $r$  and  $s$ ? That's the first question.

And the second question is, OK, assuming that the answer to the first question is in the affirmative, that  $w$  is a continuously differentiable function of  $r$  and  $s$ , how could we compute, for example, the partial of  $w$  with respect to  $r$ , knowing all of the so-called obvious other partial derivatives? What do I mean by that? Well, what I mean is if you were to look just at this equation, just looking at this equation, what are the obvious partial derivatives to take?

You say, well, we'll take the partial of  $w$  with respect to  $x$ , the partial of  $w$  with respect to  $y$ , and the partial of  $w$  with respect to  $z$ . And if you were to look, say, at this equation, the natural thing to ask is, what is the partial of  $x$  with respect to  $r$ ? What is the partial of  $x$  with respect to  $s$ ? et cetera. In other words, what we're saying is, in this particular problem, we would like to figure out how, for example, to compute the partial of  $w$  with respect to  $r$ , knowing that we have at our disposal the partials of  $w$  with respect to  $x$ ,  $y$ , and  $z$ ; the partials of  $x$ ,  $y$ , and  $z$  with respect to  $r$ ; et cetera; meaning we also have the partials of  $x$ ,  $y$ , and  $z$  with respect to  $\theta$ .

Before I go any further, notice, by the way, that if I left out that phrase that we were talking about in our last lecture, "continuously differentiable," notice that all of this would make sense, provided that the derivatives existed. There was no place here do I make any statement that the partials have to not only exist but be continuous. I never say that at all. This is what the problem is. I would like to use the chain rule.

Do you see why it's called the chain rule?  $w$  is a function of  $x$ ,  $y$ , and  $z$ .  $x$ ,  $y$ , and  $z$  are each functions of  $r$  and  $s$ . Now, what I claim is that not only is it possible to do this but the recipe for doing this is a very, very suggestive thing, one which is very, very easy to remember, once you see how it's put together. If you don't see how it's

put together, the thing is just a mess-- namely, the claim is that the partial of  $w$  with respect to  $r$  is the partial-- I'll just read it to you-- the partial of  $w$  with respect to  $x$  times the partial of  $x$  with respect to  $r$ , plus the partial of  $w$  with respect to  $y$  times the partial of  $y$  with respect to  $r$ , plus the partial of  $w$  with respect to  $z$  times the partial of  $z$  with respect to  $r$ . And as I say, if you try to memorize that, it's a very, very nasty business.

But let's look at this in three separate pieces. In a way, can you sense that this is nothing more than the change in  $w$  with respect to  $r$  due to the change in  $x$  alone? In other words, you're taking here what? The change in  $w$  due to  $x$  and multiplying that by the change in  $x$  with respect to  $r$ . So this is the contribution of the change in  $w$  with respect to  $r$  due to  $x$  alone.

On the other hand, this is the partial of  $w$  with respect to  $r$  due to the change in  $y$  alone. And this is the partial of  $w$  with respect to  $r$  due to the change in  $z$  alone. And since  $x$ ,  $y$ , and  $z$  are independent, the change in  $x$ , the change in  $y$ , and the change in  $z$  are also independent variables. Consequently, it seems reasonable to assume that to find the total change of  $w$  with respect to  $r$ , we just add up all of the partial contributions. Namely, we take the partial of  $w$  with respect to  $r$  due to  $x$  alone, add on to that the partial of  $w$  with respect to  $r$  due to  $y$  alone, add on to that the partial of  $w$  with respect to  $r$  due to  $z$  alone, and that that sum should be the total change in  $w$  with respect  $r$ , treating  $s$  as a constant.

And by the way, let me point out a pitfall with this notation. We're so used to using fractional notation here. Have you noticed that if you're not careful here, you're almost tempted to cancel-- I don't want to write this, because you'll think that it's the right way of doing it. But see if we say, let's cancel the partials with respect to  $x$  here, let's cancel the partials with respect to  $y$  here, and let's cancel the partials with respect to  $z$  here. By the way, if you did that, notice what you would get is the contradiction that the partial of  $w$  with respect to  $r$  is equal to the partial of  $w$  with respect to  $r$ , plus the partial of  $w$  with respect to  $r$ , plus the partial of  $w$  with respect to  $r$ . In other words, it seems that you would get that the partial of  $w$  with respect to  $r$  is always three times itself, which is, I hope, a glaring enough contradiction so I

don't have to go into any more detail about the contradiction part.

Notice again, though, why I have made such a fetish over labeling the variables. Notice that when you're taking the partial of  $w$  with respect to  $x$ , you're assuming that  $y$  and  $z$  are the variables that are being held constant. And when you're taking the partial of  $x$  with respect to  $r$ , it's  $s$  that you're assuming is being held constant. And as soon as you look at these subscripts here, somehow or other that should put you on your guard to be careful about crossing out because, after all, the changes are being made with respect to different sets of variables. At any rate, this is the statement.

And my other claim is that the proof follows immediately from the main, key theorem that we stressed last time, even though we didn't prove it. But we've had ample exercises using this. Namely, notice that we have already seen that if  $w$  does happen to be a continuously differentiable function of  $x$ ,  $y$ , and  $z$ , then  $\Delta w$  is the partial of  $w$  with respect to  $x$  times  $\Delta x$ , plus the partial of  $w$  with respect to  $y$  times  $\Delta y$ , plus the partial of  $w$  with respect to  $z$ , times  $\Delta z$ , plus an error term. And what is that error term? It's  $k_1 \Delta x$ , plus  $k_2 \Delta y$ , plus  $k_3 \Delta z$ , where  $k_1$ ,  $k_2$ , and  $k_3$  all approach 0 as  $\Delta x$ ,  $\Delta y$ , and  $\Delta z$  approach 0.

Now again, the key step in all of this is that this amount here I could always call  $\Delta w_{\text{tan}}$ , or as Professor Thomas calls it for more than two variables,  $\Delta w_{\text{sub lin}}$ , meaning that this is a linear equation. Remember-- I've made an abbreviation here-- these partials are assumed to be evaluated at a particular point that we're interested in. But the idea is, granted that I can always call this  $\Delta w_{\text{tan}}$ , to say that the error has this small a magnitude depends on the fact that  $w$  is a continuously differentiable function of  $x$ ,  $y$ , and  $z$ .

That's why the theory is so important. What happens in real life is that most examples you encounter in real-life engineering, the functions that you're dealing with are continuously differentiable. So it seems like we're making a big issue over nothing. I should point out that on the frontiers of knowledge, enough situations occur where the functions that we're dealing with are not continuously differentiable,

that some horrible mistakes can be made by assuming that you can replace  $\Delta w$  by this, without any significant error. But as long as this is the case, we can do this.

And now notice, what does the partial of  $w$  with respect to  $r$  mean? It means you take  $\Delta w$  divided by  $\Delta r$ . And let me just do that here. I'll just divide every term by  $\Delta r$ . And now, what do I have to do next to get the partial? I have to take the limit as  $\Delta r$  approaches 0.

Now the interesting point is as  $\Delta r$  approaches 0, holding  $s$  fixed, this term obviously becomes the partial of  $x$  with respect to  $r$ , by definition. This term becomes the partial of  $y$  with respect to  $r$ , by definition. And this term becomes a partial of  $z$  with respect to  $r$ , by definition.

By the way, notice that even though  $\Delta x$ ,  $\Delta y$ , and  $\Delta z$  are all going to 0 as  $\Delta r$  goes to 0, you can not immediately conclude that these terms drop out. Because after all,  $\Delta r$  is also approaching 0. So  $\Delta x$  over  $\Delta r$  is that 0 over 0 form. In fact, that's precisely the partial of  $x$  with respect to  $r$  term that we're talking about. The beauty is what? That as  $\Delta x$ ,  $\Delta y$ , and  $\Delta z$  approach 0, each of the  $k$ 's approach 0.

You see, the reason that the error term becomes negligible, becomes 0 in the limit, isn't because  $\Delta x$ ,  $\Delta y$ , and  $\Delta z$  are becoming small. Because these small numbers are being divided by another small number. It's because the  $k_1$ ,  $k_2$ , and  $k_3$  are getting small.

At any rate, putting this all together, notice that now, in a manner completely analogous to our part-one treatment of the chain rule, except that we're now dealing with several variables, these three terms drop out. And these three terms become the claim that we made before. In other words, this is how the partial of  $w$  with respect to  $r$  is computed.

And again, the theory is the easiest part of this. That's the easy part. The hard part is getting familiarity with how to work with this. And I think the best way to get some familiarity for working with this is to pick particularly simple problems for the lecture,

problems where it's so easy to do the problem both ways that no hangup can possibly occur.

Let's take a very simple example. Let's suppose that  $w$  equals  $x$  squared plus  $y$  squared plus  $z$  squared. Suppose we also know that  $x$  is  $r$  plus  $s$ ,  $y$  is  $r$  minus  $s$ , and  $z$  happens to be  $2r$ .

In this particular case, notice that we would find the partial of  $w$  with respect to  $r$  very conveniently by direct substitution. Namely, we simply replace  $x$  by  $r$  plus  $s$ , we replace  $y$  by  $r$  minus  $s$ , we replace  $z$  by  $2r$ . And then  $w$  simply becomes this expression here, which when we collect terms, becomes  $6r$  squared plus  $2s$  squared. And again, the arithmetic there is simple enough so I'm not even going to bother worrying about how we justify these steps.

At which stage, to take the partial of  $w$  with respect to  $r$ , holding  $s$  constant, this is simply what?  $12r$ . Because  $s$  is being treated as a constant, its derivative with respect to  $r$  is  $0$ .

You see, in an example like this, one would not really be tempted to use the chain rule. The chain rule is used in many cases not just for convenience, but in cases of great theory where you're only given that  $w$  is some function of  $x$ ,  $y$ , and  $z$ , and you're not told explicitly what the function is. You're just given  $f(x,y,z)$ . In the case where the function is given explicitly, it's sometimes very easy to substitute directly.

At any rate, what the chain rule says is roughly this. They say, look. From this equation, you could immediately say the partial of  $w$  with respect to  $x$  is  $2x$ , the partial of  $w$  with respect to  $y$  is  $2y$ , the partial of  $w$  with respect to  $z$  is  $2z$ . From this equation, you could immediately say that the partial of  $x$  with respect to  $r$  is  $1$ , the partial of  $x$  with respect to  $s$  is  $1$ , the partial of  $y$  with respect to  $r$  is  $1$ , the partial of  $y$  with respect to  $s$  is minus  $1$ , the partial of  $z$  with respect to  $r$  is  $2$ , and the partial of  $z$  with respect to  $s$  is  $0$ . In particular, summarizing our results, we have these here.

Now, what the chain rule says is what? To find the partial of  $w$  with respect to  $r$ , you just take the partial of  $w$  with respect to  $x$  times the partial of  $x$  with respect to  $r$ ,

plus the partial of  $w$  with respect to  $y$  times the partial of  $y$  with respect to  $r$ , plus the partial of  $w$  with respect to  $z$  times the partial of  $z$  with respect to  $r$ . And if we do that in this case, we simply get what?  $2x$  plus  $2y$  plus  $4z$ .

Now again, I picked, deliberately, a very simple problem here. Remember, by definition,  $x$  is  $r$  plus  $s$ ,  $y$  is  $r$  minus  $s$ , and  $z$  happens to be  $2r$ . And now you can see very quickly here that when I substitute in, I get what?  $2r$  plus  $2r$  is  $4r$ , plus  $8r$  is  $12r$ , and  $2s$  minus  $2s$  is  $0$ . The partial of  $w$  with respect to  $r$  is also  $12r$ , also meaning what? We found that same answer before. At least that's how the chain rule works.

And again, we have to remember that the chain rule does not depend on the number of variables, even though this may start to look a little bit sticky. Let's word it as follows. Suppose  $w$  happens to be a continuously differentiable function of the  $n$  independent variables  $x_1$  up to  $x_n$ . See, that's what this parenthetical remark means. I'm saying that not only do the partials of  $f$  with respect to  $x_1$  up to  $x_n$  exist at a given point, but they are continuous there.

And why do I want that in there? So I can say that my error term is never any greater than that  $k_1 \Delta x_1$ , plus  $k_2 \Delta x_2$ , plus, et cetera,  $k_n \Delta x_n$ , where the  $k$ 's go to  $0$  as the  $\Delta x$ 's go to  $0$ . I'm going to spare you the details of proofs. But I just want you to keep seeing why these things are necessary.

At any rate, let's suppose now that each of the  $n$  variables  $x_1$  up to  $x_n$  turn out to be functions of the  $m$  variables.  $n$  and  $m$  could conceivably be equal. But  $m$  could even be more than  $n$ . It can be less than  $n$ . There's no reason why they have to be equal.

All we're saying is, speaking in the most general terms, suppose each of the  $n$   $x$ 's is a continuously differentiable function of the  $m$  independent variables  $y_1$  up to  $y_m$ . In fact, that's what this "et cetera" means here. The et cetera refers to the parenthetical remark here. I mean that not only are the  $x$ 's functions of  $y_1$  up to  $y_m$ , but they're continuously differentiable functions.

Now obviously, what we're saying is that if  $w$  can be expressed in terms of the  $x$ 's, the  $x$ 's can be expressed in terms of the  $y$ 's, obviously then,  $w$  can be expressed in

terms of the  $y$ 's. In other words,  $w$  is some function of  $y_1$  up to  $y_m$ . Now the question that comes up is that just looking at this, I can talk about the partial of  $w$  with respect to  $y_1$ , the partial of  $w$  with respect to  $y_2$ , the partial of  $w$  with respect to  $y_3$ , et cetera, all the way up to the partial of  $w$  with respect to  $y$  sub  $m$ .

And the question is, look. From the original form of  $w$ , it was easy to talk about the partials of  $w$  with respect to the  $x$ 's. From how the  $x$ 's are given in terms of the  $y$ 's, it's easy to talk about the derivatives of the  $x$ 's with respect to the  $y$ 's. And so the question is, how do you find the partial of  $w$  with respect to, say,  $y$  sub  $1$ , given all of these other partial derivatives? And the answer, again, is something that you just have to get used to. The proof goes through for  $n$  and  $m$  the same way as it did for the lower-dimensional case.

And the intuitive interpretation is the same. Namely, to find the partial of  $w$  with respect to  $y_1$ , we simply see how much  $w$  changed with respect to  $y_1$  due to the change in  $x_1$  alone, add on to that the change in  $w$  with respect to  $y_1$  due to the change in  $x_2$  alone, et cetera, add on to that, finally, the change in  $w$  with respect to  $y$  sub  $1$  due to the change in  $x$  sub  $n$  alone. In other words, again, if you think of this in terms of cancellation, if you cross these things out, don't think of adding them, but think of them as what? Giving you the individual components that tell you how the partial of  $w$  with respect to  $y_1$  is made up.

By the way, there is one parenthetical remark that I haven't written on the board that I would like to make at this time. In Professor Thomas's text, he has elected to introduce matrix algebra prior to this particular chapter. It again turns out that one does not need matrices to talk about the chain rule but that if one had matrix notation, the matrix notation is particularly convenient for summarizing the chain rule. I have elected to hold off on matrix algebra till the near future because it comes up in a much better motivated way, I think, in terms of these linear approximations.

But the point is if, as you're reading the text, you see the matrix notation, and you are not familiar with the matrices, forget it. All the matrix is, is a shortcut notation for

saying this. And if I want a shortcut notation here, I don't need matrices for saying this. I can say this in terms of our sigma notation.

Notice that one other way of writing this thing very compactly that may be more suggestive is the following. Notice that I'm adding up  $n$  terms here. Each term consists of two factors, each of which looks like a fraction. The numerator of the first fraction is always a partial of  $w$ . The denominator of the second fraction is always the partial  $y_1$ .

And it appears that the denominator of the first and the numerator of the second always have the same subscript, but they seem to vary consecutively from 1 to  $n$ . And that's precisely where the sigma notation comes in handy. Why don't we just write, therefore, that this is the sum, partial of  $w$  with respect to  $x_{sub\ k}$ , plus the partial of  $x_{sub\ k}$  with respect to  $y_1$ , as the subscript  $k$  ranges through all integral values from 1 to  $n$ ?

In other words, notice that in this particular form, we have simply rewritten this thing compactly. But if you look at this and look at this, I think it's very suggestive to see how the chain rule works. You see, here's your partial of  $w$  with respect to  $y_1$ .

And these are what? The contributions due to each of the changes of the  $n$  variable. This is the change due to the  $x_{sub\ k}$  variable.

And you add these all up because they're independent variables as  $k$  goes from 1 to  $n$ . And I write "et cetera" here simply to point out that I could have computed the partial of  $w$  with respect to  $y_2$  instead of  $y_{sub\ 1}$ . By the way, the recipe would've looked exactly the same, except that if there was a 2 here, there would have been a 2 here. If there were a 3 here, they would've been a 3 here. If there were an  $m$  here, there would have been an  $m$  here.

OK, now look. At this particular stage of our lecture today, this could end with the idea that for the unit that's now assigned, this is as far as you have to go. In other words, for the exercises that I've given you in this particular unit, we do nothing higher than using the chain rule for first-order derivatives.

The point is that in many applications in real life, we must take higher-order derivatives. In other words, there are many differential equations, partial differential equations, where we must work with higher-order derivatives. And for that reason, it becomes very important, sometimes, to be able to take a second derivative or a third derivative or a fourth derivative by means of the chain rule.

Now the interesting point is that the theory that we've used so far doesn't change at all. What does happen is that the average student, in learning this material for the first time, get swamped by the notation. Consequently, what I want to do is to give you the lecture on this material at the same time that I'm lecturing on first-order derivatives, simply because the continuity follows smoother this way, so that you see what the whole overall picture is, then to make sure that you cement these things down.

The next unit after this will give you drill on taking higher-order derivatives. What this may mean is that many of you may prefer to watch this half of the film a second time, after you've already tried working some of the problems with higher-order derivatives, if you're still confused by this. But at any rate, let's take a look at a hypothetical situation.

Since we're so used to polar coordinates, let's talk in terms of polar coordinates. Suppose  $w$  happens to be a continuously differentiable function of  $x$  and  $y$ .  $x$  and  $y$ , in turn, are continuously differentiable functions of the polar coordinates  $r$  and  $\theta$ . In fact, they're  $x$  equals  $r \cos \theta$ ,  $y$  equals  $r \sin \theta$ .

Now look. If all I want to do is find the partial of  $w$  with respect to  $r$ , I can do that by the ordinary chain rule. Namely, it's the partial of  $w$  with respect to  $x$  times the partial of  $x$  with respect to  $r$ , plus the partial of  $w$  with respect to  $y$  times the partial of  $y$  with respect to  $r$ .

Now, knowing what  $x$  looks like explicitly in terms of  $r$  and  $\theta$  and what  $y$  looks like explicitly in terms of  $r$  and  $\theta$ , I can certainly compute the partials of  $x$  and  $y$  with respect to  $r$ , holding  $\theta$  constant. In particular, the partial of  $x$  with respect to  $r$  is simply  $\cos \theta$ . And the partial of  $y$  with respect to  $r$  is simply  $\sin \theta$ . So the

partial of  $w$  with respect to  $r$  is partial of  $w$  with respect to  $x$  times cosine theta, plus partial of  $w$  with respect to  $y$  times sine theta.

And by the way, notice that I cannot simplify these terms. I cannot simplify these terms in general, because all I'm given is that  $w$  is some function of  $x$  and  $y$ . I don't know what  $w$  looks like explicitly in terms of  $x$  and  $y$ . So all I can do is talk about the partials of  $w$  with respect to  $x$ , partial of  $w$  respect to  $y$ , without worrying any more about this, with the understanding that if I knew what  $w$  looked like explicitly in terms of  $x$  and  $y$ , I could work out what this thing was.

Now, here's the key point. That's why I've accentuated it. In the same way that  $w$  is a function of both  $x$  and  $y$ , so also are the partials of  $w$  with respect to  $x$  and the partials of  $w$  with respect to  $y$ . In other words, even though this looks like this emphasizes the  $x$ , notice that when you take the derivative of a function of both  $x$  and  $y$  with respect to  $x$ , in general, the resulting function will again be a function on both  $x$  and  $y$ . And so what we're saying is that if the partials of  $w$  with respect to  $x$  and the partials of  $w$  with respect to  $y$  also happen to be continuously differentiable functions of  $x$  and  $y$ , we could, if we wished, use the chain rule again.

In other words, suppose in the particular problem that I was dealing with, it wasn't enough to know the partial of  $w$  with respect to  $r$ . Suppose, for example, I wanted the second partial of  $w$  with respect to  $r$ . Well obviously, that simply means what? Take the partial of this with respect to  $r$ .

In other words, the second partial of  $w$  with respect to  $r$  is just the partial of the partial of  $w$  with respect to  $r$ , with respect to  $r$ . I'm just going to differentiate this thing with respect to  $r$ . In other words, writing this out more succinctly for you, the second partial derivative of  $w$  with respect to  $r$  is the partial with respect to  $r$  of cosine theta partial of  $w$  with respect to  $x$ , plus sine theta partial of  $w$  with respect to  $y$ .

Now, here's the key point. When we differentiate here, we're assuming that theta is being held constant. Isn't that right? So consequently, when I'm differentiating with respect to  $r$ , cosine theta is a constant. I can skip over that, see, and differentiate what's left with respect to  $r$ .

In other words, that's what? It's the partial of  $w$  with respect to  $x$  differentiated with respect to  $r$ . See, I'm using the ordinary rule for the derivative of a sum. Now, the derivative of  $\sin \theta$ -- see,  $\sin \theta$  is a constant with respect to  $r$ . So the derivative of this term is just  $\sin \theta$  times the derivative of the partial of  $w$  with respect to  $y$ , with respect to  $r$ , written this way.

Now, the key point is that both of these functions here, both of these are functions of  $x$  and  $y$ .  $x$  and  $y$ , in turn, are functions of  $r$  and  $\theta$ . So in other words, to differentiate this thing with respect to  $r$ , I must use the chain rule again. Now, because this may seem difficult for you, all I'm really saying is, look. If this term here looks messy, since we know that the partial of  $w$  with respect to  $x$  is some function of  $x$  and  $y$ , let's call that  $h(x,y)$ . Then all we're saying is that the partial of the partial of  $w$  with respect to  $x$ , with respect to  $r$ , is just the partial of  $h$  with respect to  $r$ .

But to find the partial of  $h$  with respect to  $r$ , we know how to use the chain rule there. It's just what? It's the partial of  $h$  with respect to  $x$  times the partial of  $x$  with respect to  $r$ , plus the partial of  $h$  with respect to  $y$  times the partial of  $y$  with respect to  $r$ . Of course, if we now remember what  $h$  is-- see,  $h$  is the partial of  $w$  with respect to  $x$ . So if I differentiate again with respect to  $x$ , I get the second partial of  $w$  with respect to  $x$ .

We've already seen that the partial of  $x$  with respect to  $r$  is  $\cos \theta$ , so I have this term. The partial of  $h$  with respect to  $y$  really says what? Differentiate the partial of  $w$  with respect to  $x$ , with respect to  $y$ . And the usual way of abbreviating that is like this, which, again, is explained in the reading material. And we now multiply that by the partial of  $y$  with respect to  $r$ , which happens to be  $\sin \theta$ .

Now, look at this a few times in your spare time, if it's bothering you. It is not really that difficult. It is messy notation in the sense that you're not used to notation that's quite that messy. That's why it's messy. Once you get used to it, it is not any tougher than the chain rule for one independent variable.

In fact, to take the partial of the partial of  $w$  with respect to  $y$ , with respect to  $r$ , I'll do

that in one step without using a substitution. All I'm saying is that this function depends on both  $x$  and  $y$ . So to see what its derivative is with respect to  $r$ , I'll see what the contribution of its derivative with respect to  $r$  is due to just  $x$  alone. Then I'll see what contribution of its derivative with respect to  $r$  is due to just  $y$  alone.

And by the way, when I say it that way, notice how quick it is to write this thing down. I differentiate this with respect to  $x$  multiplied by the partial of  $x$  with respect to  $r$ . Add on to that the partial of this with respect to  $y$ . Multiply that by the partial of  $y$  with respect to  $r$ .

If I do this, notice now I have what? I have the partial with respect to  $y$ . And I differentiate that with respect to  $x$ . That's written this way.

And by the way, notice that this is the reverse order of what we did over here. Namely, in one case, we first differentiated with respect to  $x$  and then with respect to  $y$ . In the other case, we differentiated first with respect to  $y$  and then with respect to  $x$ . So that actually, conceptually there is a difference. That's why we write these things differently.

It does, again, turn out that in most cases, the answer that you get-- thank goodness-- doesn't depend on the order in which you perform the derivatives. But this is not at all self evident, even though you'd like to believe that it is. But we'll talk about that more in the exercises and the like.

But all I'm saying now is that if we put everything together of what we've had before, we can obtain, in this particular case, that the second partial of  $w$  with respect to  $r$  is this somewhat messy but nonetheless straightforward expression. And see, I've circled these things to sort of tell you that if it is permissible to interchange the order of differentiation, we could combine these two terms. On the other hand, if you couldn't interchange the order, this would be a rather dangerous thing to do over here because these might be different answers.

As I say again, if you have enough continuity, it turns out that these two factors are the same. But that's not the important issue here. The important issue here is that I

can keep using the chain rule to take higher-order derivatives.

And even though the notation is messier, this happened when we dealt with functions of a single variable. Remember when we used the chain rule to find  $dy/dx$  when  $y$  and  $x$  were given, say, as functions of  $t$ ? We could also use the chain rule to find the second derivative of  $y$  with respect to  $x$ . But we had to be a little bit more careful of the computation because certain factors crept in that we had to keep track of.

At any rate, again, to illustrate this idea rather than to keep droning on about it, let me take a particularly simple computational problem to check this thing on. In other words, what I'm going to do is take this messy formula over here and apply it to a case where the arithmetic happens to be very, very simple. I'm going to rig this very, very nicely.

I'm going to let  $f(x,y)$  just be  $x$  squared plus  $y$  squared, in this case. Let  $w$  be  $x$  squared plus  $y$  squared. In polar coordinates, notice that  $x$  squared plus  $y$  squared is just  $r$  squared. So  $w$  is just  $r$  squared.

What is the partial of  $w$  with respect to  $r$ , then? The partial of  $w$  with respect to  $r$  is  $2r$ . And if I now differentiate that with respect to  $r$ , the second partial of  $w$  with respect to  $r$  is  $2$ . Obviously, one would not use the chain rule in real life to find the answer to this particular problem. We've chosen this problem simply to emphasize how the chain rule would work here.

At any rate, going back here, notice that it's very simple to see from this equation that the partial of  $w$  with respect to  $x$  is  $2x$ . Therefore, the second partial of  $w$  with respect to  $x$  is  $2$ . The partial of  $w$  with respect to  $y$  is  $2y$ . Therefore, the second partial of  $w$  with respect to  $y$  is also  $2$ .

The partial of  $w$  with respect to  $x$  is a function of  $x$  alone, in this case. Consequently, the derivative with respect to  $y$  will be  $0$ . Similarly, the partial of  $w$  with respect to  $y$  is a function of  $y$  alone. Consequently, when I differentiate that with respect to  $x$ , meaning I'm holding  $y$  constant, that derivative will also be  $0$ .

And the interesting point now is if I take these values and substitute those into this equation, what happens? Look. The second partial of  $w$  with respect to  $x$  is just 2. The second partial of  $w$  with respect to  $y$  is just 2.

The mixed partials are both 0, regardless of which order you did them in. That's what we saw over here. So consequently, according to this recipe, the second partial of  $w$  with respect to  $r$  is  $2 \cos^2 \theta + 0$ , plus  $2 \sin^2 \theta + 0$ , where the reason I've written these 0's in is simply so that when you're looking at your notes later, that traces the analog of these terms over here.

At any rate, notice now that if I add these up,  $2 \cos^2 \theta + 2 \sin^2 \theta$ , since  $\sin^2 \theta + \cos^2 \theta = 1$ , this sum is just-- I'll write that in white chalk just so we don't accentuate it. Let it just be part of the answer. This is  $2 + 0$ , which is 2. And this certainly does check with the result that we got the so-called easier way.

And again, I don't want to leave you with the idea that the second way was just the hard way of doing the same problem that we did easily the first way. I picked a simple example so you can see how this works. I'm going to have a multitude of exercises for you to do in the next unit, simply so that you'll pick up the kind of know-how that will allow you to change variables using the chain rule, with a minimum degree of difficulty. In fact, hopefully, I would like to feel by the time we're through with the next two units, you will be doing this almost as second nature.

Well, we have other topics to consider in terms of our linear approximations and the like. We'll talk about that more as the course unfolds. For the time being, I would like you to concentrate simply on mastering the chain rule. And so until we meet next time, good bye.

Funding for the publication of this video was provided by the Gabriella and Paul Rosenbaum Foundation. Help OCW continue to provide free and open access to MIT courses by making a donation at [ocw.mit.edu/donate](https://ocw.mit.edu/donate).