

ANNOUNCER: 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 to view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at ocw.mit.edu.

**HERBERT
GROSS:**

Hi, our lecture today is entitled differentiation of inverse functions. And it pulls together two previous topics that we've discussed. Namely, inverse functions themselves, and secondly, the chain rule that we've discussed just a short time ago. And perhaps the best way to introduce the power of differentiation of inverse functions is to start out with such a problem. Let's actually try to differentiate a particular function, which at least up until now, we have not been able to differentiate.

The function happens to be y equals the cube root of x . In other words, y equals x to the $1/3$. Let's find dy/dx if y equals x to the $1/3$.

Now, the whole idea of inverse functions is what? That it gives us a chance to paraphrase. That we can interchange the role of the dependent and the independent variables. And like any other form of paraphrase, even though two things may be synonymous, psychologically one of the two may be easier for us to visualize than the other. In particular, in this particular case, if y equals x to the $1/3$, another way of writing the same thing is to say what? x equals y cubed.

And now, given that x equals y cubed, in other words, with treating x as a function of y , we certainly know how to differentiate y cubed with respect to y . Namely, we know that for a positive exponent to differentiate, all we have to do is bring the exponent down and replace it by one less. In other words, right away what we can say is that dx/dy is $3y$ squared.

Now if we use the result of last time that we talked about when we were discussing the chain rule, and I'll review that result in just a few moments. But for the time being, let's assume that we have the result that dx/dy is the reciprocal of dy/dx . In other words, if dy/dx is 1 over dx/dy , that tells us that dy/dx is 1 over $3y$ squared.

And I guess to write that in a more convenient form, that's $1/3 y$ to the minus 2 if you're used to exponential notation. If we now recall from above that y is equal to x to the $1/3$, this can now be written as $1/3 x$ to the $1/3$ to the minus 2 . And we now arrive at an answer dy/dx is

' $\frac{1}{3}x$ to the minus $\frac{2}{3}$ '.

By the way, let me point out there certainly was nothing wrong with leaving our answer in this particular form. It's just conventional that wherever possible, if the original problem was given as a function of 'x', we would like our answer to also be a function of 'x'. I mean, here for example, we could have said that the answer is ' $\frac{1}{3}y$ to the minus 2', where 'y' equals ' x to the $\frac{1}{3}$ '. All we've done here is to fill this thing in explicitly.

By the way, there may have been a tendency-- and here's an example of circular reasoning. There may have been a tendency to say, didn't we already learn that to differentiate a power, we just bring the power down and replace it by one less? In other words, if we did that bringing down the $\frac{1}{3}$ would give us a factor of $\frac{1}{3}$ in front. Replacing $\frac{1}{3}$ by one less, $\frac{1}{3}$ minus 1 is minus $\frac{2}{3}$. We would then see that ' $dy dx$ ' should be ' $\frac{1}{3}x$ to the minus $\frac{2}{3}$ '. Which is exactly what we got this way.

The point that we should mention at this particular stage is that the derivative of ' x to the 'n' being ' nx to the 'n - 1"' was proven only for the case that 'n' is an integer, either positive or negative. One uses the binomial theorem to prove the result for a positive integer. One uses the quotient rule to prove the result for a negative integer. And now, even though I didn't do this thing in general, I think you can see how this will generalize. For fractional exponents, one uses the inverse function idea. Namely, we did have to use the fact that if 'y' is equal to ' x to the $\frac{1}{3}$ ', ' x equals 'y cubed' was an equivalent equation. And we could differentiate that.

Now, what is the hang up here, in so far as how certain are we that ' $dy dx$ ' and ' $dx dy$ ' are reciprocals? You may recall that when we used the chain rule, we showed that if 'y' is a differentiable function of 'x', and if 'x' is a differentiable function of 'u', then 'y' is a differentiable function of 'u'. And in particular, ' $dy du$ ' is " $dy dx$ ' times ' $dx du$ ".

And if we now take the particular case where the first variable equals the third, way 'u' equals 'y', we get what? ' $dy dy$ ', which is 1, equals " $dy dx$ ' times ' $dx dy$ ".

And at first glance, it might seem that we've proven rigorously now the result that ' $dx dy$ ' and ' $dy dx$ ' are reciprocals of each other. Product is 1. The one logical hang up that we have right now is simply this.

In the statement of the chain rule, it did not say that 'y' had to be a function of 'x' and 'x' had to be a function of 'u'. It said 'y' had to be a differentiable function of 'x'. That was the first

variable had to be a differentiable function of the second. And the second had to be a differentiable function of the third. In other words, coming down to here, if we know that 'y' is a differentiable function of 'x' and 'y' has an inverse function, and if we also knew that the inverse function was differentiable. See, in other words, this must be a differentiable function of this and this must be a differentiable function of this. In other words, the one point that was missing was is that if we knew that if a function is differentiable, then its inverse if it exists, is also differentiable. The chain rule would have given us a rigorous proof.

The point that we're missing though is we do not as yet know that the inverse of a differentiable function is also a differentiable function.

Now if you recall on our earlier lecture on inverse functions, we pointed out that there was a rather interesting graphical interpretation between 'y' equals 'f of x' and 'y' equals 'f inverse of x'. By the way whenever I say, you may recall, that's just my polite way of saying perhaps you don't, but you'd better look it up because we had it. So recall that the two curves are symmetric with respect to the line 'y' equals 'x'.

By the way again, just a brief aside here. Notice that either one of these functions could have been called 'f of x' and the other one could have been called 'f inverse'. In other words, just another piece of brief knowledge here that the inverse of the inverse is the original function again. In other words, thinking in terms of our function machine, if you interchanged the input and the output, and then of the resulting machine again, interchange the input and the output, you're back to the original machine again.

So in this particular diagram, I certainly could have labeled this curve 'y' equals 'f of x' or 'g of x', then this curve here would have been 'y' equals 'g inverse of x'. But the important point is what? That 'y' equals 'f of x' and 'y' equals 'f inverse of x' are symmetric with respect to the line 'y' equals 'x'. Now you see we have a particularly simple geometric argument as to why an inverse function should be differentiable if the original function is differentiable.

Namely, pictorially, what does it mean to say that a function is differentiable? It means that when you plot its graph, the graph is smooth. In other words, if 'f' is a differentiable function, the curve 'y' equals 'f of x' will be a smooth curve.

Now simply ask yourself the following question. If you take a smooth curve and take its mirror image with respect to the line 'y' equals 'x', or for that matter with respect to any line, do you expect the curve to become un-smooth? You see, in other words, the mirror image of a

smooth curve will again, be smooth. And that's perhaps the most intuitive way of picking off in your mind why if a function is differentiable, its inverse function will also be differentiable. Of course, as we've seen many times already in this course, we must distinguish between geometric intuition and mathematical analysis. That on more occasions than one, what seemed to be happening geometrically was complicated by something unforeseen when we tried to get the results in terms of analytic methods. Let me give you an illustration of this.

You see, what we're really saying is, granted that a picture can be a good visual aid, let's suppose we're given the curve 'y' equals 'f inverse of x'. Well, first of all, what is that the same as saying? It's the same as saying that 'x' equals 'f of y'. At any rate, the question is this. We're assuming that 'f' is a differentiable function. What we would like to do is to prove that 'f inverse' is also differentiable.

Now you see the whole thing again is that whenever you want to prove anything, what do you mean if you take 'f inverse' and differentiate it? And by the way, that may look like a funny notation. Think of 'f inverse' as being one symbol. Call it 'g'. Call it whatever you want. All we're saying is, how would we, by definition, find the derivative of the inverse function say, at 'x' equals 'x1'? And the answer is that by definition, it's just what? It's the limit as 'delta x' approaches 0. Same definition as before. "f of 'x1 plus 'delta x' minus 'f of x1" over 'delta x'.

And in fact, if you want to write that a little bit more explicitly, what is another way of writing 'delta x'? 'Delta x', of course, is "x1 plus 'delta x' minus x1'.

Now the idea is-- this is 'f inverse'. See we want to find the derivative of 'f inverse'. Who cares what the name of the function is? Whatever the name of the function is, what do you do? You compute the function at 'x1 plus 'delta x', subtract off its value at 'x1', and divide by 'delta x'. So by the same definition that we had the first time we defined derivative, this is the basic definition for finding the derivative of 'f inverse'.

Now, how do we use the fact that we already know what 'f' is like? Remember, we mentioned when we talked about inverse functions before is at the time you use-- the way you really effectively handle inverse functions is when you know properties of the original function. We're not just working blindly with 'f inverse' here, we're working with the case that 'f inverse' is the inverse of the function 'f', and that we know that 'f' is differentiable. And now we want to see if knowing that 'f' is differentiable, can we prove that 'f inverse' is differentiable? And you see, the idea is not really that difficult. We can work this thing out step by step from our little diagram

over here.

You see, notice that another name for 'f inverse of 'x1 plus 'delta x' is 'y1 plus delta y'. Let me just work with what's in the bracketed expression over here.

See, the numerator is 'y1 plus 'delta y'. That's this term. Now, what's 'f inverse of x1'? As we come up over here, remember the curve is 'y equals 'f inverse'. So 'f inverse of x1' is just 'y1'.

Now what's our denominator? 'x1 plus 'delta x' maps into 'y1 plus 'delta y'. Well, the idea is this. In terms of inverse functions, 'x1 plus 'delta x' is just the back map of 'y1 plus 'delta y'. In other words, since 'f inverse'-- let's write that down. Since 'f inverse of 'x1 plus 'delta x' is equal to 'y1 plus 'delta y', that's another way of saying that 'x1 plus 'delta x'-- we might as well write this because this is what we're emphasizing. In other words, this becomes what? 'x1 plus 'delta x' is just 'f of 'y1 plus 'delta y'. And 'x1' is just 'f of y1'.

See again, I used the picture as a visual aid. But notice that everything I've written down here follows analytically by my basic definitions. I don't want to overwhelm you with formal proofs here. These are all done in the text. And I think that again, for those of you who are proof-oriented, the proofs are done excellently enough so that you'll get them from that. And for those of you who are not proof-oriented, an extra few minutes here will not make that much of a difference. But what I want you to see over here is how this thing starts to set up now. In other words, notice that this starts to look like what? Let me just come over here where we have some more space. This is what? The limit as 'delta x' approaches 0. 'f 'delta y' over "f of 'y1 plus delta y' minus 'f of y1'".

And see, if you look at this thing, remember 'f' is a differentiable function of 'y'. If this had been a 'delta y' approaching 0, this would have just been what? If we could assume that as 'delta x' approaches 0, 'delta y' approaches 0, this just would have been what? This is the reciprocal of the derivative of 'f of y' with respect to 'y'. In other words, this would be what one would call 'dx dy' evaluated at 'y' equals 'y1'.

This is what? "f of 'y1 plus delta y' minus 'f of y1'" over 'delta y' as 'delta y' approaches 0 is that derivative. And we have what? The reciprocal of this thing. And in other words, what we will have proven is that 'dy dx' evaluated at 'x' equals 'x1' is the same as the reciprocal of 'dx dy' evaluated at 'y' equals 'y1'. The only thing we have to be sure of in terms of the formal proof is to make sure that as 'delta x' approaches 0, 'delta y' approaches 0. And that is not too difficult a thing to do. As I say, the proof is done in the book. We could do it here, but I think that that

would take away from the flavor of what we're trying to show.

The idea is it's fine to think in terms of intuitive ideas. In fact, to level with you as much as I can, of all of my mathematician friends who are outstanding in various aspects of mathematics, to my knowledge not one of them works without some sort of mental picture as to what's going on. In other words, you can take something that's very, very abstract and somehow or other, you associate in your mind some kind of a visual aid that gives you a hint as to what to do next. But once you know what to do next, you always formulate the thing in terms of mathematical precision. In other words, another way of looking at this-- let's give this a broad title. Let's call this 'Proof versus Intuition'.

And this is a topic that comes up very, very early in mathematics. Perhaps the first place that it's extremely noticeable is in the subject called plane geometry. Let me give you a for instance.

Let's take a typical traditional plane geometry problem. We'll take an isosceles triangle ABC , with AB equal to AC . And what we would like to prove is that the base angles of this triangle are equal. We'd like to prove that angle B equals angle C .

Now you remember how you tackled this problem in high school geometry. You said something like this. Well, let me draw the angle bisector here. That meets BC at D . AD equals itself by identity. These two angles are equal by definition of angle bisector. Therefore, triangle ABD is congruent to triangle ACD . And corresponding parts of congruent triangles are equal. And you then proved that the base angles of an isosceles triangle were equal.

Now, at this stage of the game if you were anything like me, what you would have done has said, this is the end of the problem. Let's go onto the next one. But if you had passed this in this particular way, you would have got a 0. And why'd you get a 0?

Well, if you remember from plane geometry, there was a particular format that had to be followed. It was called the statement reason format. For every statement that you wrote down, you had to give a reason. And the reason couldn't be things like because, or why not, or obvious. The reasons had to be what? Either definitions, or rules, or previously proven theorems. In other words, notice that even though we never emphasized it, back in plane geometry when you were drawing this diagram and getting the result, that was the geometric intuition part. In other words, this was where you showed the result was plausible.

The logic part-- and this is why geometry is being stressed particularly in the modern curriculum. In terms of logic, notice that once you had your intuitive picture, the statement reason part followed independently of the picture. You used the picture to set yourself up, but the final proof hinged on what? Having the result follow purely from the axioms themselves. From the assumptions.

And by the way, this is the basic difference between traditional and modern geometry. In modern geometry, let's go back to the same proof. And it's a rather interesting point and I think you'll see a connection between what's happening in geometry and what's happening in calculus.

You see, remember how we proved this. We said draw the angle bisector and call this point D. And then we went through this and got this result.

In modern geometry they say look it. Without looking at the picture, how do you know that D falls between B and C? It's obvious in the picture that it does. But if everything has to follow inescapably from your rules when you go through the statement reason part, then unless you have some rule or definition that tells you that D must fall between B and C, you can't use this result. In other words, your result will be plausible from a picture, but not provable analytically.

So in modern geometry, we add a few axioms, a few rules of the game. They're called the axioms of betweenness, the axioms of separation. How one point separates other points and what this thing means analytically, so that we can continue on this way.

Now you see what I'm driving at is simply this. The Utopian way I think of learning is to first have an intuitive picture of what's going on. Then you proceed gradually to learn what rigor means. As I may have said to you before, in the language of functions, rigor is a function of the 'rigoree'. In other words, if a person is perfectly willing to accept a result, and he's not going to get into any trouble using it, let him use the intuitive result. On the other hand, if you wanted to teach him later that there are pitfalls using his intuition as a background, then we can come ahead and start to do things in a little bit more of a sophisticated manner.

Now, you see what I'm driving at I guess is this. If a youngster survives this procedure of going from the intuitive approach to the rigorous approach to the logical difference between analysis and geometry, he'll be in great shape when he gets to calculus. You see, the idea is exactly the same. What we're saying in the calculus is simply this.

Given the derivative of an inverse function, we do it first in a way that makes good geometric sense to us. Then to make sure that the results do not depend on our picture, and that our results can be generalized to more variables or to tougher analytic situations where we can't draw the picture, then we tried to pick up the sophistication that allows us to remove the picture proceed purely by analysis. What I'm telling you as you read the text is if you can do both, fine. Learn the proofs and the intuition.

If you have trouble with the proofs, at least satisfy yourself that they're there. That they do seem to follow from the basic axioms and other assumptions. But meanwhile, rely heavily on the intuitive results. And the important thing being that you have a picture as to what's going on.

The second aside that I'd like to make-- and in fact, I would like to conclude our lesson for today with this very important aside-- is the following. You remember when we were first learning limits and we talked about-- well, let me just make an aside right at the beginning here. Hate to do that, but it just occurred to me.

Remember when we said, let's compute the limit of 'f of x' as 'x' approaches 'a'. And our first approach was to say, OK, let's just replace 'x' by 'a'. And you said, OK, this is fine, but what happens if you get a 0/0 form? And the counterargument to that was, well, if 'f' and 'a' are chosen at random, how likely is it that we're going to get a 0/0 form? Answer's well, it's not too likely at all.

And then the answer to that was, well, but look it. Every time you take a derivative, you're going to get a 0/0 form. And so the question was, in calculus 0/0 was very important. Now we're going to ask the same kind of a question about 1:1 functions. I guess the best way to state it is bluntly.

How likely is it that a given function f is 1:1? For example, if I draw a curve like this, is this graph 1:1? The answer, of course, is no.

For example, if I pick the point 'y1' over here and come across here, I find at least in this picture, three different candidates, three different "x's". 'x1', 'x2', and 'x3' for which what? 'f of x1' equals 'f of x2', equals 'f of x3' equals 'y1'. And at first glance, you might be tempted to say, oops, we can't apply any of our theory to this particular function.

But here's the very important point. Very often in calculus we do not start with 'y1' and look to

see whether we have ' x_1 ', ' x_2 ', or ' x_3 '. Very often in calculus we're starting at something like, oh, for the sake of argument, ' x_3 '. And we say, hey, I wonder what's going on in a neighborhood of ' x_3 '. If you want a fancy word to take care of that, it's what the mathematician calls the difference between local and global properties. And those words are exactly what they sound like. Local means in a neighborhood of a point and global means let's look at the curve in the large. And the point is this. That very, very often in calculus, we are not interested in what's happening globally.

For example, when you're driving in a car and you're driving along say, the New York Thruway, and you're near Albany and somebody says, what's our gas situation? Somehow or other, what your gas situation was when you were near Buffalo has no bearing on the problem here. How full the gas tank is and the problems involved with a full gas tank are local properties. And whether this gets more abstract or not is irrelevant. All we're saying is that in calculus, very often you're dealing with a neighborhood of a point. And notice this, and we'll do it intuitively. The book again, supplies the rigorous proof.

Notice that if ' f' of x_3 ' here is not 0. Well, for the sake of argument, in this case, we notice that the curve is always rising. Notice that with a smooth curve, if it's rising at a particular point, obviously it's going to be rising in a neighborhood of that point. Just look at the picture here. And what we're saying is this. What does it mean for a function to be 1:1?

For a function to be 1:1 on an interval, it's sufficient that either ' f' prime' never be negative, or ' f' prime' never be positive. In other words, what it means is this. That as long as the derivative is not 0, we can find a neighborhood, a local neighborhood, that will make that function 1:1 on that particular neighborhood.

In other words, once I'm working with this particular neighborhood, and I can't stress this point enough because it's going to come up over and over again. It's going to come up in more sophisticated forms when we deal with functions of several variables. But the idea is what? That in a neighborhood of a point where the derivative is not 0, the function may be viewed as being 1:1. You see, the tough part is that if you start with ' y_1 ', you have no way of knowing just from that whether you want a neighborhood near ' x_1 ', or ' x_2 ', or ' x_3 '. But if you know what neighborhood you want, in a sufficiently small neighborhood, the function always behaves like it's 1:1.

In fact, the only problem that one runs into is the case where at the point in question, the

derivative is 0. And I think you can see pictorially what happens in that case right away. As soon as the derivative is 0, notice that no matter how small an interval-- well, I shouldn't say this is a possibility. Let me show you what I mean by that.

Suppose the curve has a low point here. What I'm saying is if the curve does this, then no matter how small an interval we choose surrounding x_1 , the function will not be 1:1 in that interval. No matter what interval you pick here, if you look at the image for every point in the image, there are going to be what? Two back mappings, two points that come from here.

The reason I say you have to be careful is this. You see, you can have a case where the curve does this. It comes in, gets tangent to the x-axis, and then goes down like this. You see, at this particular point, the derivative at 0 is 0. Yet, the curve is still never falling in this area. This particular curve is 1:1. I guess what we're saying here is that when a derivative is 0, be careful because something like this can happen.

If the derivative is not 0, then we know that in the neighborhood of the point in question, as long as the curve is smooth, it represents a 1:1 function. And this is what we'll be doing very, very often in calculus, is using neighborhoods of points at which the derivative is not 0.

Now, what this leads to is this. When you start talking about things like a derivative being 0, and intervals and one-to-oneness, I think you can see that this suggests a rather powerful means or need for doing the geometry of curve plotting. This will be the topic of our next investigation. And so until next time, goodbye.

ANNOUNCER: 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.