

GUEST SPEAKER: 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.

HERBERT

GROSS:

Hi. Our lecture today involves the backbone of Calculus, namely the concept of limit, and perhaps no place in the history of mathematics is there a more subtle topic than this particular one. I'm reminded of the anecdote of the professor who was staring quite intently, philosophically, in his home, and his wife asked, what are you contemplating, and he said, that it's amazing how electricity works. And the wife said, well what's so amazing about that? All you have to do is flick the switch. And this somehow or other is exactly the predicament that the limit concept falls into.

From the point of view of computation, it seems that what comes naturally 99% of the time gives us the right answer, but unfortunately, the 1% of the time turns up almost all the time in differential Calculus. For this reason, what we will do is introduce the concept of limit through the eyes of differential Calculus, and we shall call our lecture today 'Derivatives and Limits'.

And what we shall do is go back to our old friend who makes an appearance in almost all of our lectures so far, the case of the freely falling object. A ball is dropped. It falls freely in a vacuum, and the distance, 's', that it falls in feet, at the end of 't' seconds, is given by 's' equals '16t squared'. The question that we would like to raise is, how fast is the ball falling when 't' equals one?

In other words, notice that the ball is covering a different distance here as it falls. What we want to know is at the instant that the time is one. And by the way, don't confuse the speed with the displacement. We know that when the time is one, the object has fallen a distance of 16 feet. When the time is 1, the object has fallen 16 feet. What we want to know is how fast is it falling at that particular point.

Now keep in mind that mathematics being a logical subject proceeds from the idea of trying to study the unfamiliar in terms of the familiar. What we are familiar with is the concept known as average rate of speed. So what we will do is duck the main question temporarily and proceed to a different question. What we shall ask is, what is the average speed of the ball during the time interval from t equals one to t equals two?

And we all know how to solve this problem from our pre-calculus courses, namely, we compute where the ball is when t is two. We compute where the ball was when t is one. The difference between these two distances is the distance that the ball has fallen during this time interval, and we then divide by the time interval. And the distance divided by the time is precisely the definition of average rate of speed.

In other words, during the time interval from t equals one to t equals two, the ball falls at an average speed of 48 feet per second. Again, in terms of our diagram, you see at t equals one, the object is over here. At t equals two, the object is over here, it's fallen 64 feet, and so we compute the average speed by taking the total distance and dividing by the time that it took.

Now again, what we have done is found the right answer, but to the wrong question. The question was not what was the average rate of speed from t equals one to t equals two, it was, what is the speed at the instant t equals one? And we sense that between t equals one and t equals two, an awful lot can happen, and that therefore, our average speed need not be a good approximation for the instantaneous speed.

What we do next, you see, is we say OK, let's do something a little bit differently. Let's now compute the average speed, but not from t equals one to t equals two, but rather from t equals one to t equals 1.1. And computationally, we mimic the same procedure as before. We compute s when t is 1.1, we compute s when t is one. By the way, the only difference is that it's a little bit tougher to square 1.1 and multiply that by 16 than it was to square two and multiply that by 16. The arithmetic gets slightly messier if you want to call it that, but the concept stays the same. We find the total distance traveled during the time interval, we divide by the length of the time interval, and that quotient, by definition, is the average rate of speed.

Again, in terms of our diagram, at t equals one, the object has fallen 16 feet. I'll distort this so that we can see what's happening here. At t equals 1.1, the object has fallen 19.36 feet, and now we compute the average speed from here to here in the usual high school way. Now again, the same question arises as before, namely, this is still an average speed. We wanted an instantaneous speed.

And our come back is to say look at, we at least suspect intuitively that between one and 1.1, we will hopefully get a better idea as to what's happening at exactly one than when we did on the bigger interval from one to two. In other words, whereas we don't believe that 33.6 is the

answer to our original question, we do believe that 33.6 is probably a better approximation to the answer to our original question than the answer 48 feet per second that was obtained over the larger interval.

What the mathematician does next is he says well, let's stop playing games. Let's not go between one and 1.1, or one and 1.01. Why don't we go between one and $1+h$, where h is any non-zero amount? In other words, let's find the average speed from t equals one to t equals $1+h$, and then what we'll do is we will investigate to see what happens when h gets very, very small.

Well again, we mimic exactly our procedure before. We feed in t equals $1+h$ and compute s , which of course is 16 times $1+h$ squared. Then we subtract off the distance that the particle had fallen when t is one, that's 16 feet, we divide by the length of the time interval. Well between one and $1+h$, the time interval is h , we simplify algebraically, and we wind up with this particular expression, $32h$ plus $16h$ squared over h .

And now we come to the crux of what limits, and derivatives, and instantaneous rate of speeds are all about. If you recall, it's quite tempting to look at this thing and say, ah-ha, I'll cancel an h from both the numerator and the denominator, and that will leave me 32 plus $16h$. And notice though, the very important thing, that as we mentioned in previous lectures, since you cannot divide by 0 , it is crucial at this stage that we emphasize that h is not equal to 0 .

Now of course, this is not a difficult rationalization to make. Namely, in terms of our physical situation, we wouldn't want to compute an average rate of speed over a time interval during which no time transpired. And if we let h be 0 , obviously, from one to $1+h$, no time transpires. But the key statement is, and we'll come back to exploit this not only for this lecture, but for the next one too, is to exploit the idea that in our entire discussion, it was crucial that h not be allowed to equal 0 . In other words, the average speed between one and $1+h$ is 32 plus $16h$.

Now again, without trying to be rigorous, observe that our senses tell us that as h gets us close to 0 as we want without ever being allowed to get there, $16h$ gets us close to 0 as we want also, and hence, 32 plus $16h$ apparently gets us close to 32 as we wish. In other words, it appears that the speed is 32 feet per second when t equals one. And to reuse the vernacular, what we're saying is we compute the average rate of speed-- see s of $1+h$ minus s of $1/h$ -- and see what happens to that as h approaches 0 . And in our particular case, notice that this simply means what? We came down to 32 plus $16h$ over here, and then as h

got arbitrarily small, we became suspicious, and believed that as 'h' became arbitrarily small, '32 plus 16h' became 32.

And I'm saying that in great detail because this is not quite nearly as harmless as it may seem. There is a great deal of difficulty involved in just saying as 'h' gets close to 0, '32 plus 16h' gets close to 32. Well there's no difficulty in saying it. There's difficulty that comes up in trying to show what this means precisely, as we shall see in just a little while.

By the way, it's rather easy to generalize this result. Namely, you may have noticed in our presentation there was nothing sacred about asking what was the speed when 't' equals one. One could just as easily have chosen any other time 't', say 't sub one', and asked what was the speed between 't1' and 't1 plus h'? We could have done exactly the same computation. In fact, notice down here, when we do this computation, the answer '32t one' plus '16h' becomes precisely '32 plus 16h' when 't' one happens to be one. In other words, this result here is the generalization of our previous result that had us work with the limit of '32 plus 16h'.

At any rate, by this approach, it becomes clear then, if the equation of motion, the distance versus the time, is given by 's' equals '16t squared', then that time 't' equals 't one', the instantaneous speed appears to be 32 times 't' one. And in fact, since 't' one happens to be any arbitrary time that we choose, it is customary to drop the subscript, and simply say what? At any time, 't', our freely falling body in this case, will have a speed given by '32t'. And many of us will remember, as this is a revisited course, that this is precisely the recipe that turns out when one talks about bringing down the exponent and replacing it by one less. 16 times two is 32, and 't' to a power one less than two is just t to the first power or 't'.

But notice now, no recipes involved here. Just a question of intuitively defining instantaneous speed in terms of being a limit of average speeds. Now in order that we don't become hypocrites in what we're doing, recall also that we have spent an entire lecture talking about the value of analytic Geometry in the study of Calculus. The idea that a picture is worth 1,000 words. So what we would like to do now is to revisit our previous discussion in terms of a graph.

You see, suppose we have a curve 'y' equals 'f of x', and I've drawn the curve over on this side over here. We'll make more reference to it later as we go along, but what does it mean if I now mechanically compute 'f of x1 plus delta x' minus 'f of x1' over 'delta x'? And by the way, again, notice just a question of symbolism. 'Delta x' happens to be the conventional symbol

that one uses in the analytical Geometry approach where we were using h before, but this is simply a question of notation.

Well the idea is this: observe that if we go to the graph, ' f of x_1 ' plus ' Δx ' is this particular height. On the other hand, this height is ' f of x_1 '. Consequently, our numerator is just this distance here, and our denominator, which is ' Δx ', is just this distance here. And therefore, what we have done in our quotient, is we have found the slope of the straight line that joins ' p ' to ' q .' And here then is why the question of straight line becomes so important in the study of differential Calculus. That what is average speed, in terms of analysis, becomes the slope of a straight line in terms of Geometry. And the idea, you see, is that this gives me the average speed of the straight line that joins the average rise of the straight line that joins ' p ' to ' q '.

And by the way, this is quite deceptive as any average is. It's like the man who sat with his feet in the oven and an ice pack on his head, and when somebody said how do you feel, he said, on the average, pretty good. You see, the same thing is happening over here. Notice if I had taken a completely different curve, a curve which looked nothing at all like my original curve, only that also passed through the points ' p ' and ' q ', notice that even though these two curves are quite different, the average rise from ' p ' to ' q ' is the same for both curves. In other words, when one deals with the average, one is not too concerned with how one got from the first point to the second.

You see, this is the problem with average rate of space. In fact, the bigger a time interval we work over, the more danger there is that the average rate of speed will not coincide intuitively with what we would like to believe that the instantaneous rate of speed is. And you see, that's exactly what this idea here means. What we do to define the derivative, the instantaneous rise at ' x ' equals ' x_1 ', is we take this quotient and investigate what happens to it as ' Δx ' is allowed to get arbitrarily close to 0. But let's emphasize this again. ' Δx ' is not allowed to equal 0.

You see, it's over a smaller and smaller interval, and you see what's happening here geometrically? To let ' Δx ' get smaller and smaller means, that for a given ' x_1 ', you are holding ' p ' fixed, and allowing ' q ' to move in closer and closer to ' p '. And what we are saying is, is that when ' q ' assumes this position, let's call it ' $q_{sub 1}$ ', the slope of the line that joints ' p ' to ' q ' one looks a lot more like the slope of the line that will be tangent to the curve at the point ' p '. And for this reason, what is instantaneous rate of speed, when we're talking about functions, become slopes of curves at a given point when we are talking about curves. You

see that the tangent line to a curve is viewed as being the limiting position of a cord drawn from a point 'p' to a point 'q'.

Now at any rate, this will be emphasized in the text. It will be emphasized in our exercises. The major issue now is what did we really do when it came time to compute this little gadget. In other words, when we talked about the limit of 'f of x' as 'x' approached 'a', how did we compute this? The danger is, and this is why I have underlined a question mark here, the intuitive approach is to say something like, let's just replace x by a in here. In other words, let the limit of 'f of x', as 'x' approaches 'a', be 'f of a'. In fact, this is what we did with our '32 plus 16h'.

In a way, what we did was, with '32 plus 16h', do you remember what we did? We said, let's let 'h' get arbitrarily close to 0. And in our minds, we let 'h' equal 0. '16h' was then 0, and then '32 plus 16h' was then 32. But what we had done is somehow or other, if our intuition is correct, we had arrived at the correct answer but for wrong reasons. Because you see, in this expression, it is mandatory, as we've outlined this, is that 'x' never be allowed to equal 'a'.

Well you say, who cares whether you allow it or you don't? Aren't we going to get the same answer? And just to give you a quickie, one that we talked about in our introductory lecture, consider the function 'f of x' to be 'x squared minus nine' over 'x minus three'. And let's evaluate what this limit is as 'x' approaches three. If we were to blindly assume that all we have to do is replace 'x' by 'a', that would mean what here? Replace 'x' by three. Notice that what we got is what? Three squared, which is nine, minus nine, (which is 0) over three minus three, which is also 0.

In other words, if in the expression 'x squared minus nine' over 'x minus three' you replace 'x' by three, you do get 0 over 0. Unfortunately, or perhaps fortunately, somehow or other, this definition is supposed to incorporate the fact that you never allow 'x' to equal three. See, the way we got 0 over 0 was we violated what the intuitive meaning of limit was. We wound up with this 0 over 0 mess by allowing 'x' to equal the one value it is not allowed to equal here.

To see what happened more anatomically in this particular problem, let's go back to our old friend of cancellation again. We observe that 'x squared minus nine' can be written as 'x plus three' times 'x minus three', and then we divide by 'x minus three', and here's where the main problem seems to come up. One automatically, somehow or other, gets into the habit of canceling these things out.

But the point is, since we have already agreed that you cannot divide by 0, the idea then becomes that the only time this is permissible is when 'x' is unequal to three. In fact, to have been more precise, what we should have said was not that 'x squared minus nine' over 'x minus three' equals 'x plus three', but rather, 'x squared minus nine' over 'x minus three' equals 'x plus three', unless 'x' equals three. And by the way, I should add here what? That if 'x' equals three, then 'x squared minus nine' over 'x minus three' is undefined. You see, that's what we call 0 over 0. Undefined, indeterminate.

But here's what bails us out. As soon as we say the limit as 'x' approaches three, what are we assuming? We're assuming that 'x' can be any value at all except three, and the interesting point now becomes, that as long as 'x' is anything but three, the only time you could tell these two expressions apart was when 'x' happened to equal three. So if you now say we won't allow 'x' to equal three, then it is true, whereas the bracketed expression by itself is not equal to the parenthetical expression by itself. The limit as 'x' approaches three of these two expressions are equal, and in fact, one becomes tempted to say "what?" at this case. Oh, here it's quite easy to see, that as 'x' approaches three, 'x plus three' is equal to six.

And let me put a question mark here also, because it certainly is true that if you replace 'x' by three, this will be six. But this says don't replace 'x' by three. Now before, when we replaced 'x' by three we got into trouble, so we said we can't do that. Now we replace 'x' by three, we get an answer that we like, and that is the danger that we'll say OK, we won't allow this to be done, unless we like the result.

And of course this gives you a highly subjective subject matter here. You see, the point is that somehow or other we must come up with a more objective criteria for distinguishing what limits are. And by the way, for the Polyanna who has the intuitive feeling for saying things like well look at, this may give you trouble if you wind up with 0 over 0. But what is the likelihood that given 'f of x' at random, and 'a' at random, that the limit of 'f of x' as 'x' approaches 'a' is going to be 0 over 0 if you're careless? And the answer is that the probability is quite small, except in Calculus, in which case, every time, you take a derivative. Every time you are going to wind up with 0 over 0. This is why we said earlier that differential Calculus has been referred to as 'the study of 0 over 0'.

You see, the whole point is, let's go back now to our basic definition of average rate of change of 'f', 'f of x1 plus 'delta x' minus 'f of x1' over 'delta x'. I guess when you're writing in black, you have to use white for emphasis, so we'll use this. Notice that every time you let 'delta x'

equal zero, you are going to wind up with what? 'f of x1' minus 'f of x1' , which is 0, over 'delta x', which is 0. In other words, you are going to get 0 over 0.

If you believe that all this means is to replace 'delta x' by 0, you see, and the point is, this is not the definition of this, and this is why you don't wind up with this particular thing. Well I've spent enough time, I think, trying to give a picture as to what limits mean intuitively, now what I'd like to do is to devote the remainder of this lecture, plus our next one, to investigating, more stringently, what limit means in a way that is unambiguous.

And you know, it's like everything else in this world. When you want to get something that is ironclad, you don't get something for nothing. To get an ironclad expression that doesn't lend itself to misinterpretations usually involves a rigorous enough language that scares off the uninitiated. Sort of like a legal document. The chances are if you're not a trained lawyer, and the document is simple enough for you to understand, it must have thousands of loopholes in it.

But that same rigorous language that makes it almost impossible for the layman to understand what's happening, is precisely the language that the lawyer needs to make the document well defined to him, hopefully. Well with that as a background, let's now try to give a more rigorous definition of limit. And let's try to do it in such a way that not only will the definition be rigorous, but it will agree, to the best of our ability, with what we intuitively believe a limit should mean.

Now what I want to do here is scare you a little bit, and the reason I want to scare you a little bit is, I want to show you, once and for all, how in many cases, what appears to be ominous mathematical notation, is simply a rigorous way of saying something which was quite intuitive pictorially to do. The mathematical definition of limit is the following. The limit of 'f of x' as 'x' approaches 'a' equals 'l' means, and get this, given epsilon greater than 0, we can find delta greater than 0, such that when the absolute value of 'x minus a' is less than delta, but greater than 0, then the absolute value of 'f of x' minus 'l' is less than epsilon. Now doesn't that kind of turn you on?

See, this is the whole problem. I will now cut you in sort of on the trade secret that makes this thing very, very readable. First of all, to say that you want to get arbitrarily close in value, you want 'f of x' to be arbitrarily close to 'l', another way of saying that is what? You specify any tolerance limit whatsoever, which we call epsilon, and I can make 'f of x' get to within epsilon of 'l' just by choosing 'x' sufficiently close to 'a'. And by the way again, in terms of geometry, that's

all the absolute value of a difference means. The absolute value of 'x minus a' simply means what? The distance between 'x' and 'a'.

All this says is what? I can make 'f of x' fall within epsilon of 'l' by finding a delta such that 'x' is within delta of 'a', and by the way, this little tidbit over here is simply a fancy way of saying that you're not letting 'x' equal 'a'. You see, the only way the absolute value of a number can be 0 is if the number itself is 0. The only way this could be 0 is if 'x' equals 'a', so by saying this little tidbit over here, that this is greater than 0, that's our fancy way of saying that we are not going to allow 'x' to equal to 'a'.

Now again, whereas this may sound a little bit simpler than the original definition, let's see how much easier the picture makes it. Let's go to a graph of 'y' equals 'f of x'. And in my diagram here, here is 'a', and here is 'l', and what I'm saying is, I intuitively suspect. Let me write down what I suspect. I suspect that the limit of 'f of x', as 'x' approaches 'a', is 'l' in this case. Namely I suspect that as 'x' gets closer and closer to 'a', no matter which side you come in on, that's why we use absolute value. It makes no difference whether it's positive or negative. As 'x' gets in tighter and tighter on 'a', I suspect that I can make 'f of x' come arbitrarily equal to 'l'. In other words, if 'x' is close enough to 'a', 'f of x', it seems, will be close to 'l'.

And by the way, to complete this problem geometrically, pick your epsilon. Let's say that epsilon is this 'y', and since epsilon is positive, this will be called 'l plus epsilon'. This will be called 'l minus epsilon'. Now what we do is we come over to the curve here. All right, hopefully these will look like horizontal lines to you. When I get down to here, I then project down vertical lines to meet the x-axis. And now all I'm saying, and I think this is fairly intuitive clear, is what? That if 'x' is this close to 'a', 'f of x' will be within these tolerance limits of 'l'. In other words, for any value of 'x' in here, see, for this neighborhood of 'a', every input has its output come within epsilon of 'l'.

By the way, just one brief note here, something that we'll have to come back to in more detail-- in fact, I'll do that in our next example-- notice here how important it was that our curve was not single value, but one to one. In other words, suppose our curve had doubled back, say something like this. Notice then, when you try to project over to the curve from 'l plus epsilon', you don't know whether to stop at this point or at this point. You see, in other words, notice that there is another point down here which has the same 'l plus epsilon' value.

In other words, whenever our function is not one to one, there is going to be a problem, when

we work analytically, of being able to solve algebraic equations, and trying to distinguish this point of intersection from this point. You see, if we made a mistake and bypass this point and came all the way over to here, and thought that this was our tolerance limit, we would be in a great deal of difficulty. Because you see, for example, for this value of 'x', its 'f of x' value projects up here, which is outside of the tolerance limits that are given.

You see, many things which are self evident from the picture are not nearly so self evident analytically, to which we then raise the question, why use analysis at all? Let's just use pictures. And again the answer is, in this case, we can get away with it, but in more complicated situations where either we can't draw the diagram because it's too complicated, or because we have several variables, and we get into a dimension problem, the point is that sometimes we have no recourse other than the analysis. Let me give you an exercise that we'll do in terms of putting the geometry and the analysis side by side.

Let's look at the expression the limit of 'x squared minus 2x' as 'x' approaches three. Now of course, we don't want 'x' to equal three here, but let's cheat again, and see what this thing means. Our intuition tells us that when 'x' is three, 'x squared' is nine, '2x' is six. So nine minus six is three, and we would suspect that the limit here should be three. Now to see what this means, let's draw a little diagram. The graph 'y' equals 'x squared minus 2x' crosses the x-axis at 0 and two. It has its low point at the 0.1 comma minus one. We can sketch these curves, it's a parabola.

And now what we're saying is, is it true that when 'x' gets very close to three, that 'f of x' gets very close to three? And the answer is, from the diagram, it appears very obvious that this is the case. Fact, here's that example of non-single value again. If you call this thing 'l', and you now pick tolerance limits which we'll call 'l plus epsilon' and 'l minus epsilon', and now you want to see what interval you need on the x-axis, notice that in this particular diagram, you will get two intervals, one of which surrounds the value 'x' equals minus one, and the other of which surrounds the 'x' value 'x' equals three. In other words, both when 'x' is minus one and 'x' is three, 'x squared minus 2x' will equal three. At any rate, leaving the diagram as an aid, let's see what our epsilon delta definition says, and how we handle the stuff algebraically, and how we correlate the algebra with the geometry.

Given epsilon greater than 0, what must I do? I must find a delta greater than 0 such that, and I'm going to read this colloquially, I'll write it formally, but read it colloquially, such that whenever 'x' is within delta of three, but not equal to three, then 'x squared minus 2x' will be

within epsilon of three. Now here's the whole point. We know, algebraically, how to handle absolute values. Namely, the absolute value of 'x squared minus 2x minus three' being less than epsilon, is the same as saying that 'x squared minus 2x minus three' itself must be between epsilon and minus epsilon.

Now again, we rewrite things in fancy ways if we wish. There other ways of doing this. I call this completing the square. Well, people for 2,000 years have called this completing the square. Namely, I rewrite minus three as one minus four, so that 'x squared minus 2x plus one' can be factored as 'x minus one squared'. I'm now down to this form here. Then I add four to all sides of my inequality, and have that 'x minus one squared' is between four plus epsilon and four minus epsilon. Some fairly elementary Algebra of inequalities here.

Now here's the key point. Remembering my diagram, I said to you, how do we know whether we're near three or near minus one? How do we distinguish, when we draw the lines to the curve, what neighborhood we want? And notice that all we're saying is that if 'x' is near three, that means what? 'x' is close to three. If 'x' is reasonably close to three, then 'x minus one' is going to be reasonably close to two, and hence be positive.

Point is, that as long as you're dealing with positive numbers over here, see if epsilon is sufficiently small, these will then all be positive numbers. For positive numbers, if the squares obey a certain inequality, the square roots will obey the same inequality, as we have emphasized in one of our exercises. In other words, from here, we can now say this, and from this, we can now conclude, that if you want 'x squared minus 2x' to be within epsilon of three, 'x' itself is going to have to be between 'one plus the square root of 'four plus epsilon"', and 'one plus the square root of 'four minus epsilon"'.

And by the way, this is quite rigorous, but I don't think it turns you on. I don't think there's any picture that you associate with this, and so what I thought I'd like to do for our next little thing, is to come back to our diagram here, and show you what this really means. Namely, notice that if epsilon is a small positive number, let's take a look at this again. If epsilon is a small positive number, this is slightly less than four. Therefore, the square root is slightly less than two, therefore, this number will be slightly less than three. On the other hand, 'four plus epsilon' is slightly more than four, so its square root will be slightly more than two, and therefore, one plus that square root will be slightly more than three.

And what these two numbers, as abstract as they look like, correspond to, is nothing more

than this. Coming back to our diagram and choosing an epsilon, and coming over to the curve like this, and projecting down like this, all we're saying is, see, what are we saying pictorially? That for 'f of x' to be within epsilon of 'l', 'x' itself has to be in this range here. And all this says is, this very simple point to compute, just by projecting down like this, its rigorous name would be 'one plus the square root of 'four minus epsilon''. That's this number over here.

And the furthest point, namely this point here, which again, in terms of the picture is very easy to find, that point corresponds to one plus the square root of 'four plus epsilon'. And again notice, in terms of the Algebra, I need no recourse to a picture. But in terms of the picture, I get a heck of a good feeling as to what these abstract symbols are telling me.

Well let's continue on with the solution of this exercise. You see, we didn't want to find where 'x' was, we wanted to see where 'x minus three' had to be. In other words, we're trying to find the delta. We know that 'x minus three' is between these two extremes here. Consequently, and here's a place we have to be a little bit careful here, this of course, is a positive number, because this is more than two. This, on the other hand, is a negative number, because you see 'four minus epsilon' is less than four, so its square root is less than two. So if I subtract two from-- it's a negative number-- and since delta has to be positive, if this thing is negative, two minus the square root of 'four minus epsilon' will be positive.

So what I do is, to solve this problem, is I simply choose delta to be the minimum of these two numbers. And then by definition, what do I have? That when the absolute value of 'x minus three' is less than delta, and at the same time, this is crucial, greater than 0. In other words, we never let 'x' equal three. For this choice of delta, by how we worked backwards, this will come out to be less than epsilon. And that is exactly what you mean by our formal definition of saying that the limit of 'x squared minus 2x' as 'x' approaches three, in this case, equals three.

For someone who wants more concrete evidence, I think all one has to do is, for example, take a number like 'one plus the square root of 'four plus epsilon'', that is the widest point on our interval, and actually plug that in for 'x', and go through this computation of squaring 'one plus the square root of 'four plus epsilon''. Subtract twice, 'one plus the square root of 'four plus epsilon''. Carry out that computation and you get what? 'Three plus epsilon'. That using the upper extreme, you wind up with what? In excess of three by epsilon, which by the way is exactly what's supposed to happen.

In fact, I hope this doesn't give you an eye sore, I will pull down the top board again to show

you also what these two distances mean, and what this means. You see, all these two numbers mean is the following, and I'll pull this down just far enough to see it. That these two numbers were the widths from this end point to three on the one extreme, and from this end point to three on the other extreme. In other words, those two numbers that we took the minimum of were the half-width intervals that surrounded three.

You see the problem, as we mentioned before is, is that even though epsilon and minus epsilon are symmetric with respect to 'l', when you translate over to the curve, because this curve is not a straight line when you project down, these two widths will not be equal. By picking the minimum of these two widths, you guarantee that you have locked yourself inside the required area. You see, that's all this particular thing means.

Now the trouble is that one might now get the idea that there are no shorter ways of doing the same problem. You see, I showed you a rigorous way. There was no law that said we had to find the biggest delta. In other words, if delta equals, a half will work, in other words, if you're within a half of three, something's going to happen, then certainly any smaller number will work just as well. In other words, I can get estimates. Let me show you what I mean by that.

Another way of tackling how to make ' $x^2 - 2x - 3$ ' smaller than epsilon is to use our properties of absolute values, and to factor this. In other words, we know that ' $x^2 - 2x - 3$ ' is ' $x - 3$ ' times ' $x + 1$ '. We know that the absolute value of a product is a product of the absolute values, so these two things here are synonyms. Then we know that near ' x ' equals three, which we're interested in here, ' $x + 1$ ' is near four. Well what we mean more rigorously is this, choose an epsilon, and if the absolute value of ' $x - 3$ ' is less than epsilon, in other words, if ' $x - 3$ ' is less than epsilon but greater than minus epsilon, then by adding on four to all three sides of the inequality here, we see that ' $x + 1$ ' is less than 'four plus epsilon', but greater than 'four minus epsilon'.

If the size of epsilon is less than one, then certainly this is less than five, and this is greater than three. Now of course, somebody may say to us, but what if epsilon is more than one? Well obviously, if a guy says make this thing within 10, and I can make it within one, certainly within 10 is within one. I can always pick the smaller number. It's like that nonsense of the fellow asking for an ice cream cone, and the waitress said what flavor, and he said anything except chocolate. And she said, I'm all out of chocolate, will you take anything except vanilla?

The idea here is that if somebody says make this less than 15, if you've made it less than 10,

in particular, you've made it less than 15. And so all we do over here, you see, is something like this. We say look it, if we want to make this number here very small, we know that this number here is less than five, we know that ' ϵ over five' is a positive number if ϵ is small, so why don't we just pick the absolute value of ' x minus three' to be less than ' ϵ over five?' In fact, that's what this δ will be in that case.

In other words, if the absolute value of x minus three is within ' ϵ over five', if ' x ' is within ' ϵ over five' of three, notice what this product becomes. This is less than ' ϵ over five', this we know from before is less than five, and therefore, this product is less than ϵ . In other words, we have exhibited the δ such that when this happens, the absolute value of what we want to make less than ϵ indeed becomes less than ϵ .

Now because I recognize that this is a hard topic, you'll notice in the reading assignments that these problems are covered in great detail. Everything that I've said in the lesson so far is repeated in great computational depth in our learning exercises. The point is that even if I can make the ϵ s and δ s seem a little bit more meaningful for you than by the formal definition, notice that it's simple only in comparison to what we had before. But it's still very, very difficult. The beauty of Calculus is that in many cases, we do not have to know what δ looks like for a given ϵ .

What we shall do there therefore, in our next lecture, is to develop recipes that will allow us to get the answers to these limit problems without having to go through this genuinely difficult problem of finding a given δ to match a given ϵ . Though we must admit, in real life, in many cases where you're making approximations, we will want to know what the tolerance limits are. I am not belittling the ϵ δ approach. All I'm saying is that in certain problems, you do not need the ϵ δ approach to get nice results, and that will be the topic of our next lecture. So until next time, goodbye.

GUEST SPEAKER: 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.