PROFESSOR: Hi. Today's lecture starts off with what probably is the single most important topic in the entire course, not just to date, but in the entire course. And it's a rather sneaky thing in the sense that takes a while to grow into it. Things have been going so smoothly so far, that perhaps this point is going to be a little bit subtle to make, and we'll try to lead into it gradually.

There are probably ten different ways to introduce this topic. And whichever one you pick, I think, any one of the other nine would've been better. But without any further ado, let me talk about today's lesson in terms of something called the directional derivative and indicate, in a manner of speaking, that when we deal with functions of several variables, especially in the case that n = 2, there is a very obvious geometrical interpretation, one that we made some use of in the previous lecture, but which I hope to make more use of in this particular lecture.

Let's take a look and see what the situation is. Let's suppose I'm given that w is a function of the two independent variables, x and y, and I have the w as therefore some function of x and y. And I can now talk about the graph of w, which we assume is some surface. And I've drawn it in this particular way.

I take some point (a, b) in the xy plane on which f is defined. And at that particular point, I go to the corresponding point on the surface, which I call P0, which has coordinates what? (a, b, c), where c is simply f(a,b). That, of course is how you graph a function of two independent variables. The w coordinate, so to speak, is simply the functional value of the x- and y-coordinates.

Now, what we did last time was we essentially said, look, if we were to slice this surface by a plane which was either parallel to the wy plane or to the wx plane, we get special curves of intersection, and we use this to indicate the idea of partial derivatives. Now, without too much imagination, I think it should be very easy for you to visualize an arbitrary surface over your heads. You're standing in a particular point on the floor. At that point on the floor, you visualize a coordinate axis, which we'll call the x- and y-axis.

Now, obviously, you can move out from the point at which you're standing along the x-axis. You can move out in the direction along the y-axis. You can visualize that how rapidly the height above your head is changing as you move out certainly will depend on whether you're moving along the x-axis or the y-axis.

But the key point-- at least, the key point from the point of view of today's lecture-- is the fact that why were you restricted to either of these two directions? Why couldn't you have moved in any one of infinitely many different directions, namely from the point at which originating? You can move in any direction at all, because we're assuming, at least, that the floor on which you're standing is continuous. It's unbroken. You can move in these directions.

And the idea that you get can be shown pictorially by saying, look. Let's assume that we're at the point (a, b). We have this surface over our head just as we did before. But now, instead of picking a direction either parallel to the wx plane or to the wy plane, let's pick some arbitrary direction s in which to move.

And now, what I'm saying is if I now take the plane that passes through this direction s perpendicular to the xy plane, that plane will also intersect this surface, passing through the point P0. I get a curve of intersection, and I can talk about the slope of that particular curve. See, ultimately, this is what we're going to be talking about.

Now, what is the slope of that particular curve? Well it's a derivative. It's a derivative of w with respect to s, where all of a sudden we've now chosen the s direction over here to be what we're going to take our derivative with respect to. In other words, it seems to make sense to talk about the derivative of f in the direction of s, evaluated at the point a comma b. In other words, just like we talked about f sub x (a, b) and f sub y (a, b), why can't we talk about the derivative of f evaluated at (a, b) in the direction of x?

And if we want to correlate this with the notation in the text, observe that we call that dw/ds. By the way, notice-- not the partial of w with respect to s, but the derivative of w with respect to s. Remember, the partials were essentially defined in terms of holding either x or y constant.

Notice that in our general directional derivative idea, neither x nor y is being held constant. Notice that x and y are varying as we move along s. Notice, also by the way, that once x and y are restricted to move along s, they are no longer independent variables.

It's a rather interesting point. We talk about w being a function of the two independent variables x and y, but as soon as we pick that direction s in the xy plane, x and y have to be very specially related, namely, according to the equation of a straight line that determines s for that point to be on the line. And that's why we talk about dw/ds. w is a function of a single variable once you restrict yourself to the particular direction s. At any rate, the question that we want to come to grips with is how do you find dw/ds.

Well, here again is the beauty of our logical structure. By definition-- this is the same definition we knew early in part one of our course. dw/ds, by definition, is the limit as delta s approaches 0, delta w divided by delta s.

Now, the thing is what's very difficult to compute in real life is delta w. After all, this w, being f(x,y), f can be a very, very complicated surface. And to actually find the true change in w-- well, heck. We already saw this in part one when we compared this delta y with delta y tan. To actually find the change in y was a much more difficult job then to find the change in y to the tangent line.

What we're saying here is, look. We already know how to find the change in w, not to the surface, but to the plane which is tangent to the surface at our point P sub 0. In other words, we have already discussed that delta w tan is the partial of f with respect to x evaluated at (a, b) times delta x plus the partial of f with respect to y at the point (a, b) times delta y.

And now, what we say is-- and I've written this to accentuate, because I have to talk very strongly about this. It's a point that, if I don't make, most of you, at least, will allow me to slip over this and not even notice that I've missed something very, very crucial. But let me assume that I can approximate delta w by delta w tan. In other words, let's suppose delta w tan is a reasonable approximation for delta w.

You say, well, what's such a big assumption about that? And I'm going to save that for the very last part of the lecture, because my belief is that the subtlety is so great that I would like to leave that for the very end and go through as if the subtlety didn't exist so that you get the computational feeling as to what's going on here. But here's the interesting point. Notice that delta w is a change in w as you move from the point (a, b) to some other point in the plane. It's a change in height.

Now, obviously, that change in w is going to depend very strongly on what direction you're moving in. On the other hand, how was delta w tan computed? Delta w tan was computed just by knowing two special directional derivatives known as the partial derivatives.

You see, notice that to get delta w tan, I have made the assumption that all I have to know is what's happening in the x direction and what's happening in the y direction, everything else being determined from the function evaluated at (a, b). So this is really a very strong assumption, that delta w is determined pretty closely by delta w tan.

And it turns out to be almost universally true and we're going to save that part, as I say, for the end. But for now, let's suppose that we are allowed to replace delta w by delta w tan. If we do that, and we go back to our definition for dw/ds-- namely, the limit as delta s approaches 0, delta w divided by delta s-- we now replace delta w by delta w tan and divide through by delta s to find delta w over delta s.

We have, assuming that this is a legal substitution, that delta w over delta s is the partial of f with respect to x, evaluated at (a, b) times delta x divided by delta s plus the partial of f with respect to y evaluated at (a, b) times the change in y divided by the change in s-- delta y divided by delta s.

Now, remember, in the s direction, x and y are not independent variables. In fact, how are they related? Let's isolate our little diagram here so that we see in what direction we're moving here. We're starting at the point (a, b).

And by the way, to get the proper orientation here, as I've drawn the xy plane here, imagine the surface coming out from the blackboard. In other words, the height is really being measured away from the blackboard here. That's where my surface is. I move in the direction of s.

Here is a delta x. Here is a delta y. s has a constant direction, a constant slope. It's a straight line. Call the angle that it makes with the positive x-axis phi. Notice that no matter how big delta x and delta y are, they are related by similar triangles to what? The delta x divided by delta s will always be cosine phi, and delta y divided by delta s will always be sine phi.

And therefore, if I replace delta x divided by delta s by cosine phi, delta y divided by delta s by sine phi, I obtain that-- what? Delta w divided by delta s is equal to the partial of f with respect to x evaluated at point (a, b) times the cosine of phi plus the partial of f with respect to y evaluated at a comma b times the sine of phi, where phi is the what? The angle that the direction s makes with the positive x-axis.

At any rate, notice by the way, that these things are all constants once s is chosen, once the point (a, b) is fixed. Therefore, when I pass to the limit, nothing really changes. In other words, this thing that I'm calling the directional derivative of f in the direction of s evaluated at (a, b) turns out to be what? The partial of f with respect to x at the point (a, b) times cosine phi plus the partial of f with respect to y evaluated at (a, b) times the sine of phi.

And what I'd like you to notice is that in these two terms, one factor of each term is determined solely by the point a comma b. In other words, notice that these two factors here are just partial derivatives and have nothing to do with direction. They determined solely by the choice of the function f and the point (a, b).

On the other hand, notice that these two factors, cosine phi and sine phi, have nothing to do with f and have to do only with the direction s itself, which, again, should make intuitive sense to you. But if you're asking in terms of the surface being up over your ahead for a directional derivative, obviously, once the surface is given, the directional derivative should depend on two things-- one, what point you start at, and the other, what direction you move in once you've started at that particular point. And what direction you move in has nothing to do with what the surface looks like over your head.

At any rate, what we now do is invoke our dot product notation-- in other words, this little trick that we talked about when we learned dot products. This particular sum can be written very suggestively as the dot product of two vectors. Namely, I will write this as what? The vector whose components are f sub x and f sub y dotted with the vector whose components are cosine phi and sine phi, because remember, when you dot two vectors in Cartesian coordinates, you multiply them coefficient by coefficient.

To make a long story short, dw/ds is what? It's this vector, f sub, whose i component is f sub x evaluated at (a, b), whose j complement is f sub y evaluated at (a, b). I call that vector g(a, b). I'm going to give that a special name a little bit later.

But for now, it's very important to notice that this is not a number. It's an ordered pair of numbers. In other words, it's a vector, and if you want to think of this as a vector, what we're saying is what? Think of the vector whose i component is the partial of f with respect to x evaluated at (a, b), whose j complement is f sub y evaluated at (a, b).

Notice that these are numbers, because a and b are fixed constants here. Therefore, this vector, g(a, b) is a constant vector, and it's in the xy plane. It's a 2-tuple. On the other hand, the other vector (cosine phi, sine phi)-- hopefully, you recognize by now is nothing more than the unit vector in the direction of s. You see, this is the unit vector in the direction of s. So if I now use my abbreviation, dw/ds evaluated at (a, b)-- in other words, the directional derivative of w in the direction of s evaluated at (a, b) is just my vector g(a, b) dotted with the unit vector in the direction of s.

Now, observe have two things. First of all, when you dot, this is a constant. I can't change this once a and b are given. This is fixed. So all I can vary is u sub s. But u sub s is a unit vector, so the only way I can vary u sub s is to change its direction.

Notice that for two vectors of constant magnitude, their dot product is maximum when the two vectors are parallel. In other words, this will be as big as possible when u sub s is chosen to be in the direction of my vector g. In other words, dw/ds at (a, b) is maximum in the direction of g(a, b). That's the first thing to observe.

The second thing is that when they act parallel, the cosine of the angle between them is 1. So the magnitude of this vector will just be the product of these two magnitudes. But the magnitude of u sub s, u sub s being a unit vector, is 1. Therefore, the maximum magnitude not only occurs in the direction of g, but it is also numerically equal to the magnitude of g.

In other words, the maximum value of dw/ds evaluated at (a, b) not only occurs in the direction of the vector g, but that maximum magnitude is the magnitude of g evaluated at (a, b). For that reason, g of (a, b) is given a very important name. And I decided to hold off on the name until as late as possible so that the name wouldn't frighten you.

But the name is the gradient vector. In other words, the vector g of (a, b) is called the gradient of f at (a, b). And it's usually written in this notation-- in an upside down delta. It's called del, usually, with an arrow over it, or in boldface print in the text. And it's written this way and it's read what? The gradient of f evaluated at (a, b).

What is the gradient of f evaluated at (a, b)? It's the vector, which gives you the hint as to how to compute the directional derivative in any direction that you wish. Namely, in terms of the gradient vector, the directional derivative of f in the direction of s evaluated at (a, b) is the gradient of f evaluated at (a, b) dotted with the unit vector in the direction of s. By the way, you may recall that when you dot a vector with a unit vector, you get the projection of that vector in the direction of the unit vector. In other words, the directional derivative-- another way of looking at this physically is nothing more than the projection of the gradient vector onto the given direction in which you're moving.

And the important point is that this particular definition does not depend on our coordinate system. What is interesting is that, in Cartesian coordinates, there is a very simple way of computing the gradient vector. Namely, the i component of the gradient vector is just the partial of f with respect to x and the j component is just the partial of f with respect to y.

But that was a very special case, because, you see, i and x happen to have the same direction, as do y and j. For arbitrary coordinate systems, this need not be true. And I'm going to drill you on that in the exercises. But in other words, what I'm saying is remember the gradient vector in terms of a maximum directional derivative. Don't memorize it as a formula, because if you do, you're going to get in trouble.

For example, if I were to give my surface in polar coordinates, say w of some function of r and theta, then it turns out-- and there's an exercise on this in the notes-- that the gradient of f is the partial of w with respect to r times u sub r plus-- and here's the big difference-- 1 over r times the partial of w with respect to theta times u sub theta. In other words, the gradient vector is not the partial of w with respect to r times u sub r plus the partial of w with respect to theta times u sub theta. In other words, you don't just mechanically differentiate with respect to these variables.

And the key reason that you can't do this-- well, look. Let's just look at this little diagram. And I think the whole idea will become very clear. Remember that in polar coordinates r is denoted this way. u sub theta is at right angles to r. Notice that u sub theta is not in the direction of theta. Notice that the direction of theta is sort of the tangent to this circle of radius r at this point. If I call this increment d(theta), notice that this arc length is r d(theta), so the vector in the direction of u sub theta is r d(theta), not d(theta).

I don't know if you noticed that, but coming back up here for a moment, notice that this was OK here, because r was in the same direction as u sub r. Notice, however, that it's r d(theta) which is in the u sub theta direction. Again, I leave most of these details for the notes. But I feel that if I don't say these things to you, it becomes very easy to miss these points when we talk about them or write about them, but somehow I hope that by you hearing me say this, you will be keyed in when you come to these concepts in the unit that we're studying.

But I think the best way to augment what we're doing is by means of a specific example. Let us suppose that we're given the surface w equals f(x,y) where f(x,y) is x to the fifth plus x cubed y plus y to the fifth. And we want to compute the directional derivative of f at the point (1, 1) in the direction-- let's call it s sub 1, where s sub 1 is the direction that goes from the point (1, 1) to the point (4, 5).

Now, what we're saying is-- and I guess, maybe, if we look at these two diagrams concurrently, maybe this'll be easier to see. Here we are at the point (1, 1) We want to see how fast the slope over our head-- the w value-- is changing in the direction of s1, where s1 is chosen to be what? We're moving from the point (1, 1) on the xy plane to the point (4, 5).

See, we're moving in this direction and we want to see how fast w is changing over our heads, which geometrically means you draw this plane, intersect it with the particular surface here. And this point, P0-- what we really want geometrically is what? The slope of the line tangent to this curve in the ws1 plane tangent to this curve at the point P0. And my claim is that this can be done very, very easily from a mechanical point of view now that we have our gradient vector behind us.

Namely, what we do is, given what w looks like as a function of x and y, we take the partial of w with respect to both x and y, which, hopefully, you can all do quite mechanically now based on our last unit's work. We differentiate, first holding y constant, then holding x constant. At any rate, we obtain what? That the partial of w with respect to x is 5x to the fourth plus 3x squared y. And so if we compute that at the point (1, 1) when x and y are both 1, this simply turns out to be 8.

In a similar way, the partial of w with respect to y is x cubed plus 6y to the fifth. So if we compute that at the point (1, 1), that turns out to be 7. In other words, then, by definition of our gradient, which is the partial of f with respect to x evaluated at (1, 1) times i plus the partial of f with respect to y evaluated at (1, 1) times j, the gradient of f at (1, 1) is just 8i plus 7j. Very easy to write down mechanically when you're using Cartesian coordinates.

Now, let me make a brief aside, an interruption here. The idea is to emphasize what the gradient vector means. What this tells me is that if I were to leave the point (1, 1) in the direction of the vector 8i plus 4j-- if I were to leave in that direction, that would be the direction in which the directional derivative would be maximum. And moreover, that maximum directional derivative would just be the magnitude of this gradient vector.

The magnitude of that gradient vector is just the square root of 8 squared plus 7 squared, which is the square root of 113. In other words, what this tells us is that the maximum directional derivative leaving the point 1 comma 1 is the square root of 113, and it occurs in a direction 8i plus 7j as you leave the point 1 comma 1.

At any rate, getting back to the main stream of the problem, what we want is a directional derivative in the direction of s1. That means what? We take our gradient vector, which is (8, 7), and dot that with the unit vector in the direction of s1.

You may recall from this diagram here that the vector in the direction of s1 has its i component equal to 3, its j complement equal to 4. This makes this a 3, 4, 5 right triangle. So the unit vector in this direction has as its components 3/5 and 4/5.

In other words, the directional derivative of f at the point (1, 1) in the given direction s1 is just the gradient dotted with the unit vector 3/5 i plus 4/5 j. Just mechanically carrying out this operation leads to 52/5. And by the way, this had better turn out to be less than this, because this is what? The maximum value that the directional derivative can have. In other words, if we haven't made a mistake here, one of the checkpoints is what? That the vector can't project to be any longer than what it really is in this. It can't be more than the gradient vector.

But at any rate, let's now conclude the lecture by coming to the part which is probably the hardest thing that we're going to encounter in the whole course. In a way, I feel a little bit like a man who fell off the Empire State Building. And when he went past the 40th floor, somebody said, "How are you doing?" And he said, "So far so good."

And that is, we've taken some tremendous liberties here. And the biggest liberty that we've taken-- and it's not just a liberty. It's the kind of a liberty that to solve involves the foundations of our entire course. We are now at the grassroots of what at least the calculus of functions of several variables is all about. And that's this trouble spot.

First of all, does delta w tan exist? That's the first question. Namely, how do you know that there is a tangent plane? Just because the surface happens to be smooth when you cut it by a plane parallel to the wy plane and smooth when you cut it by a plane parallel to wx plane, how do you that it's going to be smooth for any given direction?

See, that's the first intellectual question that comes up in the reading assignment that has to be solved effectively. First of all, does delta w tan exist meaningfully? And secondly, if it does exist, how is it related to delta w?

And now, we come to that key theorem, the proof of which is quite hairy. It's done in the text. It's also done as an optional exercise to help you generalize what's done in the text. And it's the counterpart of what happens with differentials in functions of a single real variable. But the key theorem, which I'll state here without proof, is simply this.

Suppose that w is a function of x and y and that f sub x and f sub y both happen to exist in some neighborhood of the point (a, b). All right? So far so good. Now, here is they key additional hypothesis. Suppose also that f sub x and f sub y happen to be continuous at (a, b).

My claim is that if this additional hypothesis is obeyed, the tangent plane will exist. In other words, it's not enough for the directional derivative to exist in the x and y directions in order to guarantee that the directional derivative will exist in every direction. But it is enough provided that these directional derivatives happen to be continuous. And by the way, if these conditions are met, we say that f is a continuously differentiable function of x and y.

But I'll talk about that more next time or in the notes or in the exercises. We're going to make a big issue over the sooner or later. But for now what I do want to do is just end with what the key theorem is.

The key theorem says, look. Just like with one variable, if these conditions are met, then there is a very reasonable approximation to delta w by delta w tan. Namely, what the theorem says is, in this case, delta w will be the partial of f with respect to x evaluated at (a, b) times delta x plus the partial of f with respect to y evaluated at (a, b) times delta y-- and notice, of course, that this is a thing that we've been calling delta y tan-- plus an arrow. And the arrow has the form k1 delta x plus k2 delta y, where k1 and k2 both approach 0 as delta x and delta y approach 0.

And this is very, very crucial. It's not enough, as we're going to see in the very next lecture, that delta x and delta y approach 0. It's that these things go to 0 as delta x and delta y go to 0. Consequently, these terms go to 0 faster. They go to 0 as a second order, infinitesimal. And what this really says is, look, for very small values of delta x and delta y, even when you're dealing with 0/0 forms, if you pick a sufficiently small neighborhood of the point (a, b)-- and that's the key point, a sufficiently small neighborhood of the point (a, b)-- then delta w is approximately equal to delta w tan

And by the way, this holds also in several variables. In other words, I picked the case n equals 2 here simply so that we can utilize the geometry. Namely, what we're saying is, in terms of the geometry, if f happens to be a continuously differentiable function of x and y, and we look at the surface w equals f(x, y) above the point (a, b), what we're saying is that in a neighborhood of that point, there is--

well, first of all we're saying what?

A tangent plane exists to the surface above that point and that in a neighborhood of that point of tangency, the tangent plane is an excellent approximation s the true change in w. Now, what happens is if n is greater than 2, we can no longer use the geometric interpretation. But what is important is that the analytic proof never makes use of the picture.

And the key point is-- and I'm going to exploit this in future lectures. The really key point is that the delta w tan never gets messy, that the variables-- delta x, delta y, et cetera-- all occur as linear terms. And this is why the so-called linear algebra subject becomes so important in the study of functions of several variables.

At any rate, I think this is enough for one lesson. And in our next lesson, what we will do is show how using this key theorem has its analog in something called the chain rule, just as it did in the case of part one when we studied functions of a single independent variable.

At any rate, then, until next time. Good bye.

Funding for the publication of this video was provided by the Gabrielle 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.