

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. so -- OK, so remember last time, on Tuesday we learned about the chain rule, and so for example we saw that if we have a function that depends, sorry, on three variables, x, y, z , that x, y, z themselves depend on some variable, t , then you can find a formula for df/dt by writing down $w_x dx/dt + w_y dy/dt + w_z dz/dt$. And, the meaning of that formula is that while the change in w is caused by changes in x, y , and z , x, y , and z change at rates $dx/dt, dy/dt, dz/dt$. And, this causes a function to change accordingly using, well, the partial derivatives tell you how sensitive w is to changes in each variable. OK, so, we are going to just rewrite this in a new notation. So, I'm going to rewrite this in a more concise form as gradient of w dot product with velocity vector dr/dt . So, the gradient of w is a vector formed by putting together all of the partial derivatives. OK, so it's the vector whose components are the partials. And, of course, it's a vector that depends on x, y , and z , right? These guys depend on x, y, z . So, it's actually one vector for each point, x, y, z . You can talk about the gradient of w at some point, x, y, z . So, at each point, it gives you a vector. That actually is what we will call later a vector field. We'll get back to that later. And, dr/dt is just the velocity vector $dx/dt, dy/dt, dz/dt$. OK, so the new definition for today is the definition of the gradient vector. And, our goal will be to understand a bit better, what does this vector mean? What does it measure? And, what can we do with it? But, you see that in terms of information content, it's really the same information that's already in the partial derivatives, or in the differential. So, yes, and I should say, of course you can also use the gradient and other things like approximation formulas and so on. And so far, it's just notation. It's a way to rewrite things. But, so here's the first cool property of the gradient. So, I claim that the gradient vector is perpendicular to the level surface corresponding to setting the function, w , equal to a constant. OK, so if I draw a contour plot of my function, so, actually forget about z because I want to draw a two variable contour plot. So, say I have a function of two variables, x and y , then maybe it has some contour plot. And, I'm saying if I take the gradient of a function at this point, (x, y) . So, I will have a vector. Well, if I draw that vector on top of a contour plot, it's going to end up being perpendicular to the level curve. Same thing if I have a function of three variables. Then, I can try to draw its contour plot. Of course, I can't really do it because the contour plot would be living in space with x, y , and z . But, it would be a bunch of level faces, and the gradient vector would be a vector in space. That vector is perpendicular to the level faces. So, let's try to see that on a couple of examples. So, let's do a first example. What's the easiest case? Let's take a linear function of x, y , and z . So, I will take w equals a_1 times x plus a_2 times y plus a_3 times z . Well, so, what's the gradient of this function? Well, the first component will be a_1 . That's partial w partial x . Then, a_2 , that's partial w partial y , and a_3 , partial w partial z . Now, what is the levels of this? Well, if I set w equal to some constant, c , that means I look at the points where $a_1x + a_2y + a_3z$ equals c . What kind of surface is that? It's a plane. And, we know how to find a normal vector to this plane just by looking at the coefficients. So, it's a plane with a normal vector exactly this gradient. And, in fact, in a way, this is the only case you need to check because of linear approximations. If you replace a function by its linear approximation, that means you will replace the level surfaces by their tangent planes. And then, you'll actually end up in this situation. But maybe that's not very convincing. So, let's do another example. So, let's do a second example. Let's say we look at the function $x^2 + y^2$. OK, so now it's a function of just two variables because that way we'll be able to actually draw a picture for you. OK, so what are the level sets of this function? Well, they're going to be circles, right? w equals c is a circle, $x^2 + y^2 = c$. So, I should say, maybe, sorry, the level curve is a circle. So, the contour plot looks something like that. Now, what's the gradient vector? Well, the gradient of this function, so, partial w partial x is $2x$. And partial w partial y is $2y$. So, let's say I take a point, x comma y , and I try to draw my gradient vector. So, here at x, y , so, I have to draw the vector, $\langle 2x, 2y \rangle$. What does it look like? Well, it's going in that direction. It's parallel to the position vector for this point. It's actually twice the position vector. So, I guess it goes more or less like this. What's interesting, too, is it is perpendicular to this circle. OK, so it's a general feature. Actually, let me show you more examples, oops, not the one I want. So, I don't know if you can see it so well. Well, hopefully you can. So, here I have a contour plot of a function, and I have a blue vector. That's the gradient vector at the pink point on the plot. So, you can see, I can move the pink point, and the gradient vector, of course, changes because the gradient depends on x and y . But, what doesn't change is that it's always perpendicular to the level curves. Anywhere I am, my gradient stays perpendicular to the level curve. OK, is that convincing? Is that visible for people who can't see blue? OK, so, OK, so we have a lot of evidence, but let's try to prove the theorem because it will be interesting. So, first of all, sorry, any questions about the statement, the example, anything, yes? Ah, very good question. Does the gradient vector, why is the gradient vector perpendicular in one direction rather than the other? So, we'll see the answer to that in a few minutes. But let me just tell you immediately, to the side, which side it's pointing to, it's always pointing towards higher values of a function. OK, and we'll see in that maybe about half an hour. So, well, let me say actually points towards higher values of w . OK, any other questions? I don't see any questions. OK, so let's try to prove this theorem, at least this part of the theorem. We're not going to prove that just yet. That will come in a while. So, well, maybe we want to understand first what happens if we move inside the level curve, OK? So, let's imagine that we are taking a moving point that stays on the level curve or on the level surface. And then, we know, well, what happens is that the function stays constant. But, we can also know how quickly the function changes using the chain rule up there. So, maybe the chain rule will actually be the key to understanding how the gradient vector and the motion on the level surface relate. So, let's take a curve, r equals r of t , that stays inside, well, maybe I should say on the level surface, w equals c . So, let's think about what that means. So, just to get you used to this idea, I'm going to draw a level surface of a function of three variables. OK, so it's a surface given by the equation w of x, y, z equals some constant, c . And, so now I'm going to have a point on that, and it's going to move on that surface. So, I will have some parametric curve that lives on this surface. So, the question is, what's

going to happen at any given time? Well, the first observation is that the velocity vector, what can I say about the velocity vector of this motion? It's going to be tangent to the level surface, right? If I move on a surface, then at any point, my velocity is tangent to the curve. But, if it's tangent to the curve, then it's also tangent to the surface because the curve is inside the surface. So, OK, it's getting a bit cluttered. Maybe I should draw a bigger picture. Let me do that right away here. So, I have my level surface, w equals c . I have a curve on that, and at some point, I'm going to have a certain velocity. So, the claim is that the velocity, v , equals dr/dt is tangent -- -- to the level, w equals c because it's tangent to the curve, and the curve is inside the level, OK? Now, what else can we say? Well, we have, the chain rule will tell us how the value of w changes. So, by the chain rule, we have dw/dt . So, the rate of change of the value of w as I move along this curve is given by the dot product between the gradient and the velocity vector. And, so, well, maybe I can rewrite it as $w \cdot v$, and that should be, well, what should it be? What happens to the value of w as t changes? Well, it stays constant because we are moving on a curve. That curve might be complicated, but it stays always on the level, w equals c . So, it's zero because w of t equals c , which is a constant. OK, is that convincing? OK, so now if we have a dot product that's zero, that tells us that these two guys are perpendicular. So -- So if the gradient vector is perpendicular to v , OK, that's a good start. We know that the gradient is perpendicular to this vector tangent that's tangent to the level surface. What about other vectors tangent to the level surface? Well, in fact, I could use any curve drawn on the level of w equals c . So, I could move, really, any way I wanted on that surface. In particular, I claim that I could have chosen my velocity vector to be any vector tangent to the surface. OK, so let's write this. So this is true for any curve, or, I'll say for any motion on the level surface, w equals c . So that means v can be any vector tangent to the surface tangent to the level. See, for example, OK, let me draw one more picture. OK, so I have my level surface. So, I'm drawing more and more levels, and they never quite look the same. But I have a point. And, at this point, I have the tangent plane to the level surface. OK, so this is tangent plane to the level. Then, if I choose any vector in that tangent plane. Let's say I choose the one that goes in that direction. Then, I can actually find a curve that goes in that direction, and stays on the level. So, here, that would be a curve that somehow goes from the right to the left, and of course it has to end up going up or something like that. OK, so given any vector tangent -- -- let's call that vector v tangent to the level, we get that the gradient is perpendicular to v . So, if the gradient is perpendicular to this vector tangent to this curve, but also to any vector, I can draw that tangent to my surface. So, what does that mean? Well, that means the gradient is actually perpendicular to the tangent plane or to the surface at this point. So, the gradient is perpendicular. And, well, here, I've illustrated things with a three-dimensional example, but really it works the same if you have only two variables. Then you have a level curve that has a tangent line, and the gradient is perpendicular to that line. OK, any questions? No? OK, so, let's see. That's actually pretty neat because there is a nice application of this, which is to try to figure out, now we know, actually, how to find the tangent plane to anything, pretty much. OK, so let's see. So, let's say that, for example, I want to find -- -- the tangent plane -- -- to the surface with equation, let's say, $x^2 y^2 - z^2 = 4$ at the point $(2, 1, 1)$. Let me write that. So, how do we do that? Well, one way that we already know, if we solve this for z , so we can write z equals a function of x and y , then we know tangent plane approximation for the graph of a function, z equals some function of x and y . But, that doesn't look like it's the best way to do it. OK, the best way to it, now that we have the gradient vector, is actually to directly say, oh, we know the normal vector to this plane. The normal vector will just be the gradient. Oh, I think I have a cool picture to show. OK, so that's what it looks like. OK, so here you have the surface $x^2 y^2 - z^2$ equals four. That's called a hyperboloid because it looks like when you get when you spin a hyperbola around an axis. And, here's a tangent plane at the given point. So, it doesn't look very tangent because it crosses the surface. But, it's really, if you think about it, you will see it's really the plane that's approximating the surface in the best way that you can at this given point. It is really the tangent plane. So, how do we find this plane? Well, you can plot it on a computer. That's not exactly how you would look for it in the first place. So, the way to do it is that we compute the gradient. So, a gradient of what? Well, a gradient of this function. OK, so I should say, this is the level set, w equals four, where w equals $x^2 y^2 - z^2$. And so, we know that the gradient of this, well, what is it? $2x$, then $2y$, and then negative $2z$. So, at this given point, I guess we are at x equals two. So, that's four. And then, y and z are one. So, two, negative two. OK, and that's going to be the normal vector to the surface or to the tangent plane. That's one way to define the tangent plane. All right, it has the same normal vector as the surface. That's one way to define the normal vector to the surface, if you prefer. Being perpendicular to the surface means that you are perpendicular to its tangent plane. OK, so the equation is, well, $4x - 2y - 2z$ equals something, where something is, well, we should just plug in that point. We'll get eight plus two minus two looks like we'll get eight. And, of course, we could simplify dividing everything by two, but it's not very important here. OK, so now if you have a surface given by an evil equation, and a point on the surface, well, you know how to find the tangent plane to the surface at that point. OK, any questions? No. OK, let me give just another reason why, another way that we could have seen this. So, I claim, in fact, we could have done this without the gradient, or using the gradient in a somehow disguised way. So, here's another way. So, the other way to do it would be to start with a differential, OK? dw , while it's pretty much the same content, but let me write it as a differential, dw is $2x dx + 2y dy - 2z dz$. So, at a given point, at $(2, 1, 1)$, this is $4 dx + 2 dy - 2 dz$. Now, if we want to change this into an approximation formula, we can. We know that the change in w is approximately equal to $4 \Delta x + 2 \Delta y - 2 \Delta z$. OK, so when do we stay on the level surface? Well, we stay on the level surface when w doesn't change, so, when this becomes zero, OK? Now, what does this approximation sign mean? Well, it means for small changes in x , y , z , this guy will be close to that guy. It also means something else. Remember, these approximation formulas, they are linear approximations. They mean that we replace the function, actually, by some closest linear formula that will be nearby. And so, in particular, if we set this equal to zero instead of approximately zero, it means we'll actually be moving on the tangent plane to the level set. If you want strict equalities in approximations means that we replace the function by its tangent approximation. So -- [APPLAUSE] OK, so the level corresponds to Δw equals zero. and its tangent plane corresponds to $4 \Delta x + 2 \Delta y - 2 \Delta z$ equals zero.

$\Delta y - 2\Delta z = 0$. That's what I'm trying to say, basically. And, what's Δx ? Well, that means it's the change in x . So, what's the change in x here? That means, well, we started with $x = 2$, and we moved to some other value, x . So, that's actually $x - 2$, right? That's how much x has changed compared to 2. And, two times $(y - 1)$ minus two times $z - 1 = 0$. That's the equation of a tangent plane. It's the same equation as the one over there. These are just two different methods to get it. OK, so this one explains to you what's going on in terms of approximation formulas. This one goes right away, by using the gradient factor. So, in a way, with this one, you don't have to think nearly as much. But, you can use either one. OK, questions? No? OK, so let's move on to new topic, which is another application of a gradient vector, and that is directional derivatives. OK, so let's say that we have a function of two variables, x and y . Well, we know how to compute $\frac{\partial w}{\partial x}$ or $\frac{\partial w}{\partial y}$, which measure how w changes if I move in the direction of the x axis or in the direction of the y axis. So, what about moving in other directions? Well, of course, we've seen other approximation formulas and so on. But, we can still ask, is there a derivative in every direction? And that's basically, yes, that's the directional derivative. OK, so these are derivatives in the direction of \hat{i} or \hat{j} , the vectors that go along the x or the y axis. So, what if we move in another direction, let's say, the direction of some unit vector, let's call it \hat{u} . OK, so if I give you a unit vector, you can ask yourself, if I move in the direction, how quickly will my function change? So -- So, let's look at the straight trajectory. What this should mean is I start at some value, x, y , and there I have my vector \hat{u} . And, I'm going to move in a straight line in the direction of \hat{u} . And, I have the graph of my function -- -- and I'm asking myself how quickly does the value change when I move on the graph in that direction? OK, so let's look at a straight line trajectory. So, we have a position vector, \mathbf{r} , that will depend on some parameter which I will call s . You'll see why very soon, in such a way that the derivative is this given unit vector \hat{u} . So, why do I use s for my parameter rather than t . Well, it's a convention. I'm moving at unit speed along this line. So that means that actually, I'm parameterizing things by the distance that I've traveled along a curve, sorry, along this line. So, here it's called s in the sense of arc length. Actually, it's not really an arc because it's a straight line, so it's the distance along the line. OK, so because we are parameterizing by distance, we are just using s as a convention just to distinguish it from other situations. And, so, now, the question will be, what is $\frac{dw}{ds}$? What's the rate of change of w when I move like that? Well, of course we know the answer because that's a special case of the chain rule. So, that's how we will actually compute it. But, in terms of what it means, it really means we are asking ourselves, we start at a point and we change the variables in a certain direction, which is not necessarily the x or the y direction, but really any direction. And then, what's the derivative in that direction? OK, does that make sense as a concept? Kind of? I see some faces that are not completely convinced. So, maybe you should show more pictures. Well, let me first write down a bit more and show you something. So I just want to give you the actual definition. Sorry, first of all in case you wonder what this is all about, so let's say the components of our unit vector are two numbers, a and b . Then, it means we'll move along the line x of s equals some initial value, the point where we are actually at the directional derivative plus s times a , or I meant to say plus a times s . And, y of s equals $y_0 + bs$. And then, we plug that into w . And then we take the derivative. So, we have a notation for that which is going to be $\frac{dw}{ds}$ with a subscript in the direction of \hat{u} to indicate in which direction we are actually going to move. And, that's called the directional derivative -- -- in the direction of \hat{u} . OK, so, let's see what it means geometrically. So, remember, we've seen things about partial derivatives, and we see that the partial derivatives are the slopes of slices of the graph by vertical planes that are parallel to the x or the y directions. OK, so, if I have a point, at any point, I can slice the graph of my function by two planes, one that's going along the x , one along the y direction. And then, I can look at the slices of the graph. Let me see if I can use that thing. So, we can look at the slices of the graph that are drawn here. In fact, we look at the tangent lines to the slices, and we look at the slope and that gives us the partial derivatives in case you are on that side and want to see also the pointer that was here. So, now, similarly, the directional derivative means, actually, we'll be slicing our graph by the vertical plane. It's not really colorful, something more colorful. We'll be slicing things by a plane that is now in the direction of this vector, \hat{u} , and we'll be looking at the slope of the slice of the graph. So, what that looks like here, so that's the same applet the way that you've used on your problem set in case you are wondering. So, now, I'm picking a point on the contour plot. And, at that point, I slice the graph. So, here I'm starting by slicing in the direction of the x axis. So, in fact, what I'm measuring here by the slope of the slice is the partial in the x direction. It's really partial f partial x , which is also the directional derivative in the direction of \hat{i} . And now, if I rotate the slice, then I have all of these planes. So, you see at the bottom left, I have the direction in which I'm going. There's this, like, rotating line that tells you in which direction I'm going to be moving. And for each direction, I have a plane. And, when I slice by that plane, I will get, so I have this direction here going maybe to the southwest. So, that gives me a slice of my graph by a vertical plane, and the slice has a certain slope. And, the slope is going to be the directional derivative in that direction. OK, I think that's as graphic as I can get. OK, any questions about that? No? OK, so let's see how we compute that guy. So, let me just write again just in case you want to, in case you didn't hear me it's the slope of the slice of the graph by a vertical plane -- -- that contains the given direction, that's parallel to the direction, \hat{u} . So, how do we compute it? Well, we can use the chain rule. The chain rule implies that $\frac{dw}{ds}$ is actually the gradient of w dot product with the velocity vector $\frac{d\mathbf{r}}{ds}$. But, remember we say that we are going to be moving at unit speed in the direction of \hat{u} . So, in fact, that's just gradient w dot product with the unit vector \hat{u} . OK, so the formula that we remember is really $\frac{dw}{ds}$ in the direction of \hat{u} is gradient w dot product of \hat{u} . And, maybe I should also say in words, this is the component of the gradient in the direction of \hat{u} . And, maybe that makes more sense. So, for example, the directional derivative in the direction of \hat{i} is the component along the x axes. That's the same as, indeed, the partial derivatives in the x direction. Things make sense. $\frac{dw}{ds}$ in the direction of \hat{i} is, sorry, gradient w dot \hat{i} , which is w_x , maybe I should write, partial w of partial x . OK, now, so that's basically what we need to know to compute these guys. So now, let's go back to the gradient and see what this tells us about the gradient. [APPLAUSE] I see you guys are having fun. OK, OK, let's do a little bit of geometry here. That should calm you down. So, we said $\frac{dw}{ds}$ in the direction of \hat{u} is gradient w dot \hat{u} . That's the same as the length of

gradient w times the length of u . Well, that happens to be one because we are taking the unit vector times the cosine of the angle between the gradient and the given unit vector, u , so, have this angle, θ . OK, that's another way of saying we are taking the component of a gradient in the direction of u . But now, what does that tell us? Well, let's try to figure out in which directions w changes the fastest, in which direction it increases the most or decreases the most, or doesn't actually change. So, when is this going to be the largest? If I fix a point, if I set a point, then the gradient vector at that point is given to me. But, the question is, in which direction does it change the most quickly? Well, what I can change is the direction, and this will be the largest when the cosine is one. So, this is largest when the cosine of the angle is one. That means the angle is zero. That means u is actually in the direction of the gradient. OK, so that's a new way to think about the direction of a gradient. The gradient is the direction in which the function increases the most quickly at that point. So, the direction of gradient w is the direction of fastest increase of w at the given point. And, what is the magnitude of w ? Well, it's actually the directional derivative in that direction. OK, so if I go in that direction, which gives me the fastest increase, then the corresponding slope will be the length of the gradient. And, with the direction of the fastest decrease? It's going in the opposite direction, right? I mean, if you are on a mountain, and you know that you are facing the mountain, that's the direction of fastest increase. The direction of fastest decrease is behind you straight down. OK, so, the minimal value of dw/ds is achieved when cosine of θ is minus one. That means θ equals 180° . That means u is in the direction of minus the gradient. It points opposite to the gradient. And, finally, when do we have dw/ds equals zero? So, in which direction does the function not change? Well, we have two answers to that. One is to just use the formula. So, that's one cosine θ equals zero. That means θ equals 90° . That means that u is perpendicular to the gradient. The other way to think about it, the direction in which the value doesn't change is a direction that's tangent to the level surface. If we are not changing a , it means we are moving along the level. And, that's the same thing -- -- as being tangent to the level. So, let me just show that on the picture here. So, if actually show you the gradient, you can't really see it here. I need to move it a bit. So, the gradient here is pointing straight up at the point that I have chosen. Now, if I choose a slice that's perpendicular, and a direction that's perpendicular to the gradient, so that's actually tangent to the level curve, then you see that my slice is flat. I don't actually have any slope. The directional derivative in a direction that's perpendicular to the gradient is basically zero. Now, if I rotate, then the slope sort of increases, increases, increases, and it becomes the largest when I'm going in the direction of a gradient. So, here, I have, actually, a pretty big slope. And now, if I keep rotating, then the slope will decrease again. Then it becomes zero when I perpendicular, and then it becomes negative. It's the most negative when I pointing away from the gradient and then becomes zero again when I'm back perpendicular. OK, so for example, if I give you a contour plot, and I ask you to draw the direction of the gradient vector, well, at this point, for example, you would look at the picture. The gradient vector would be going perpendicular to the level. And, it would be going towards higher values of a function. I don't know if you can see the labels, but the thing in the middle is a minimum. So, it will actually be pointing in this kind of direction. OK, so that's it for today.