

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](http://ocw.mit.edu). Let me start by basically listing the main things we have learned over the past three weeks or so. And I will add a few complements of information about that because there are a few small details that I didn't quite clarify and that I should probably make a bit clearer, especially what happened at the very end of yesterday's class. Here is a list of things that should be on your review sheet for the exam. The first thing we learned about, the main topic of this unit is about functions of several variables. We have learned how to think of functions of two or three variables in terms of plotting them. In particular, well, not only the graph but also the contour plot and how to read a contour plot. And we have learned how to study variations of these functions using partial derivatives. Remember, we have defined the partial of  $f$  with respect to some variable, say,  $x$  to be the rate of change with respect to  $x$  when we hold all the other variables constant. If you have a function of  $x$  and  $y$ , this symbol means you differentiate with respect to  $x$  treating  $y$  as a constant. And we have learned how to package partial derivatives into a vector, the gradient vector. For example, if we have a function of three variables, the vector whose components are the partial derivatives. And we have seen how to use the gradient vector or the partial derivatives to derive various things such as approximation formulas. The change in  $f$ , when we change  $x$ ,  $y$ ,  $z$  slightly, is approximately equal to, well, there are several terms. And I can rewrite this in vector form as the gradient dot product the amount by which the position vector has changed. Basically, what causes  $f$  to change is that I am changing  $x$ ,  $y$  and  $z$  by small amounts and how sensitive  $f$  is to each variable is precisely what the partial derivatives measure. And, in particular, this approximation is called the tangent plane approximation because it tells us, in fact, it amounts to identifying the graph of the function with its tangent plane. It means that we assume that the function depends more or less linearly on  $x$ ,  $y$  and  $z$ . And, if we set these things equal, what we get is actually, we are replacing the function by its linear approximation. We are replacing the graph by its tangent plane. Except, of course, we haven't see the graph of a function of three variables because that would live in 4-dimensional space. So, when we think of a graph, really, it is a function of two variables. That also tells us how to find tangent planes to level surfaces. Recall that the tangent plane to a surface, given by the equation  $f$  of  $x$ ,  $y$ ,  $z$  equals  $c$ , at a given point can be found by looking first for its normal vector. And we know that the normal vector is actually, well, one normal vector is given by the gradient of a function because we know that the gradient is actually pointing perpendicularly to the level sets towards higher values of a function. And it gives us the direction of fastest increase of a function. OK. Any questions about these topics? No. OK. Let me add, actually, a cultural note to what we have seen so far about partial derivatives and how to use them, which is maybe something I should have mentioned a couple of weeks ago. Why do we like partial derivatives? Well, one obvious reason is we can do all these things. But another reason is that, really, you need partial derivatives to do physics and to understand much of the world that is around you because a lot of things actually are governed by what is called partial differentiation equations. So if you want a cultural remark about what this is good for. A partial differential equation is an equation that involves the partial derivatives of a function. So you have some function that is unknown that depends on a bunch of variables. And a partial differential equation is some relation between its partial derivatives. Let me see. These are equations involving the partial derivatives -- -- of an unknown function. Let me give you an example to see how that works. For example, the heat equation is one example of a partial differential equation. It is the equation -- Well, let me write for you the space version of it. It is the equation  $\text{partial } f \text{ over partial } t$  equals some constant times the sum of the second partials with respect to  $x$ ,  $y$  and  $z$ . So this is an equation where we are trying to solve for a function  $f$  that depends, actually, on four variables,  $x$ ,  $y$ ,  $z$ ,  $t$ . And what should you have in mind? Well, this equation governs temperature. If you think that  $f$  of  $x$ ,  $y$ ,  $z$ ,  $t$  will be the temperature at a point in space at position  $x$ ,  $y$ ,  $z$  and at time  $t$ , then this tells you how temperature changes over time. It tells you that at any given point, the rate of change of temperature over time is given by this complicated expression in the partial derivatives in terms of the space coordinates  $x$ ,  $y$ ,  $z$ . If you know, for example, the initial distribution of temperature in this room, and if you assume that nothing is generating heat or taking heat away, so if you don't have any air conditioning or heating going on, then it will tell you how the temperature will change over time and eventually stabilize to some final value. Yes? Why do we take the partial derivative twice? Well, that is a question, I would say, for a physics person. But in a few weeks we will actually see a derivation of where this equation comes from and try to justify it. But, really, that is something you will see in a physics class. The reason for that is basically physics of how heat is transported between particles in fluid, or actually any medium. This constant  $k$  actually is called the heat conductivity. It tells you how well the heat flows through the material that you are looking at. Anyway, I am giving it to you just to show you an example of a real life problem where, in fact, you have to solve one of these things. Now, how to solve partial differential equations is not a topic for this class. It is not even a topic for 18.03 which is called Differential Equations, without partial, which means there actually you will learn tools to study and solve these equations but when there is only one variable involved. And you will see it is already quite hard. And, if you want more on that one, we have many fine classes about partial differential equations. But one thing at a time. I wanted to point out to you that very often functions that you see in real life satisfy many nice relations between the partial derivatives. That was in case you were wondering why on the syllabus for today it said partial differential equations. Now we have officially covered the topic. That is basically all we need to know about it. But we will come back to that a bit later. You will see. OK. If there are no further questions, let me continue and go back to my list of topics. Oh, sorry. I should have written down that this equation is solved by temperature for point  $x$ ,  $y$ ,  $z$  at time  $t$ . OK. And there are, actually, many other interesting partial differential equations you will maybe sometimes learn about the wave equation that governs how waves propagate in space, about the diffusion equation, when you have maybe a mixture of two fluids. how they somehow mix over time and so on. Basically, to every problem

you might want to consider there is a partial differential equation to solve. OK. Anyway. Sorry. Back to my list of topics. One important application we have seen of partial derivatives is to try to optimize things, try to solve minimum/maximum problems. Remember that we have introduced the notion of critical points of a function. A critical point is when all the partial derivatives are zero. And then there are various kinds of critical points. There is maxima and there is minimum, but there is also saddle points. And we have seen a method using second derivatives -- -- to decide which kind of critical point we have. I should say that is for a function of two variables to try to decide whether a given critical point is a minimum, a maximum or a saddle point. And we have also seen that actually that is not enough to find the minimum of a maximum of a function because the minimum of a maximum could occur on the boundary. Just to give you a small reminder, when you have a function of one variables, if you are trying to find the minimum and the maximum of a function whose graph looks like this, well, you are going to tell me, quite obviously, that the maximum is this point up here. And that is a point where the first derivative is zero. That is a critical point. And we used the second derivative to see that this critical point is a local maximum. But then, when we are looking for the minimum of a function, well, it is not at a critical point. It is actually here at the boundary of the domain, you know, the range of values that we are going to consider. Here the minimum is at the boundary. And the maximum is at a critical point. Similarly, when you have a function of several variables, say of two variables, for example, then the minimum and the maximum will be achieved either at a critical point. And then we can use these methods to find where they are. Or, somewhere on the boundary of a set of values that are allowed. It could be that we actually achieve a minimum by making  $x$  and  $y$  as small as possible. Maybe letting them go to zero if they had to be positive or maybe by making them go to infinity. So, we have to keep our minds open and look at various possibilities. We are going to do a problem like that. We are going to go over a practice problem from the practice test to clarify this. Another important cultural application of minimum/maximum problems in two variables that we have seen in class is the least squared method to find the best fit line, or the best fit anything, really, to find when you have a set of data points what is the best linear approximately for these data points. And here I have some good news for you. While you should definitely know what this is about, it will not be on the test. [APPLAUSE] That doesn't mean that you should forget everything we have seen about it, OK? Now what is next on my list of topics? We have seen differentials. Remember the differential of  $f$ , by definition, would be this kind of quantity. At first it looks just like a new way to package partial derivatives together into some new kind of object. Now, what is this good for? Well, it is a good way to remember approximation formulas. It is a good way to also study how variations in  $x$ ,  $y$ ,  $z$  relate to variations in  $f$ . In particular, we can divide this by variations, actually, by  $dx$  or by  $dy$  or by  $dz$  in any situation that we want, or by  $d$  of some other variable to get chain rules. The chain rule says, for example, there are many situations. But, for example, if  $x$ ,  $y$  and  $z$  depend on some other variable, say of variables maybe even  $u$  and  $v$ , then that means that  $f$  becomes a function of  $u$  and  $v$ . And then we can ask ourselves, how sensitive is  $f$  to a value of  $u$ ? Well, we can answer that. The chain rule is something like this. And let me explain to you again where this comes from. Basically, what this quantity means is if we change  $u$  and keep  $v$  constant, what happens to the value of  $f$ ? Well, why would the value of  $f$  change in the first place when  $f$  is just a function of  $x$ ,  $y$ ,  $z$  and not directly of you? Well, it changes because  $x$ ,  $y$  and  $z$  depend on  $u$ . First we have to figure out how quickly  $x$ ,  $y$  and  $z$  change when we change  $u$ . Well, how quickly they do that is precisely  $\partial x / \partial u$ ,  $\partial y / \partial u$ ,  $\partial z / \partial u$ . These are the rates of change of  $x$ ,  $y$ ,  $z$  when we change  $u$ . And now, when we change  $x$ ,  $y$  and  $z$ , that causes  $f$  to change. How much does  $f$  change? Well,  $\partial f / \partial x$  tells us how quickly  $f$  changes if I just change  $x$ . I get this. That is the change in  $f$  caused just by the fact that  $x$  changes when  $u$  changes. But then  $y$  also changes.  $y$  changes at this rate. And that causes  $f$  to change at that rate. And  $z$  changes as well, and that causes  $f$  to change at that rate. And the effects add up together. Does that make sense? OK. And so, in particular, we can use the chain rule to do changes of variables. If we have, say, a function in terms of polar coordinates on  $\theta$  and we like to switch it to rectangular coordinates  $x$  and  $y$  then we can use chain rules to relate the partial derivatives. And finally, last but not least, we have seen how to deal with non-independent variables. When our variables say  $x$ ,  $y$ ,  $z$  related by some equation. One way we can deal with this is to solve for one of the variables and go back to two independent variables, but we cannot always do that. Of course, on the exam, you can be sure that I will make sure that you cannot solve for a variable you want to remove because that would be too easy. Then when we have to look at all of them, we will have to take into account this relation, we have seen two useful methods. One of them is to find the minimum of a maximum of a function when the variables are not independent, and that is the method of Lagrange multipliers. Remember, to find the minimum or the maximum of the function  $f$ , subject to the constraint  $g$  equals constant, well, we write down equations that say that the gradient of  $f$  is actually proportional to the gradient of  $g$ . There is a new variable here,  $\lambda$ , the multiplier. And so, for example, well, I guess here I had functions of three variables, so this becomes three equations.  $f_x = \lambda g_x$ ,  $f_y = \lambda g_y$ , and  $f_z = \lambda g_z$ . And, when we plug in the formulas for  $f$  and  $g$ , well, we are left with three equations involving the four variables,  $x$ ,  $y$ ,  $z$  and  $\lambda$ . What is wrong? Well, we don't have actually four independent variables. We also have this relation, whatever the constraint was relating  $x$ ,  $y$  and  $z$  together. Then we can try to solve this. And, depending on the situation, it is sometimes easy. And it sometimes it is very hard or even impossible. But on the test, I haven't decided yet, but it could well be that the problem about Lagrange multipliers just asks you to write the equations and not to solve them. [APPLAUSE] Well, I don't know yet. I am not promising anything. But, before you start solving, check whether the problem asks you to solve them or not. If it doesn't then probably you shouldn't. Another topic that we solved just yesterday is constrained partial derivatives. And I guess I have to re-explain a little bit because my guess is that things were not extremely clear at the end of class yesterday. Now we are in the same situation. We have a function, let's say,  $f$  of  $x$ ,  $y$ ,  $z$  where variables  $x$ ,  $y$  and  $z$  are not independent but are constrained by some relation of this form. Some quantity involving  $x$ ,  $y$  and  $z$  is equal to maybe zero or some other constant. And then, what we want to know, is what is the rate of change of  $f$  with respect to one of the variables. say  $x$ .  $v$  or  $z$  when I keep the others constant? Well. I cannot

keep all the other constant because that would not be compatible with this condition. I mean that would be the usual or so-called formal partial derivative of  $f$  ignoring the constraint. To take this into account means that if we vary one variable while keeping another one fixed then the third one, since it depends on them, must also change somehow. And we must take that into account. Let's say, for example, we want to find -- I am going to do a different example from yesterday. So, if you really didn't like that one, you don't have to see it again. Let's say that we want to find the partial derivative of  $f$  with respect to  $z$  keeping  $y$  constant. What does that mean? That means  $y$  is constant,  $z$  varies and  $x$  somehow is mysteriously a function of  $y$  and  $z$  for this equation. And then, of course because it depends on  $y$ , that means  $x$  will vary. Sorry, depends on  $y$  and  $z$  and  $z$  varies. Now we are asking ourselves what is the rate of change of  $f$  with respect to  $z$  in this situation? And so we have two methods to do that. Let me start with the one with differentials that hopefully you kind of understood yesterday, but if not here is a second chance. Using differentials means that we will try to express  $df$  in terms of  $dz$  in this particular situation. What do we know about  $df$  in general? Well, we know that  $df$  is  $f_{\text{sub } x} dx + f_{\text{sub } y} dy + f_{\text{sub } z} dz$ . That is the general statement. But, of course, we are in a special case. We are in a special case where first  $y$  is constant.  $y$  is constant means that we can set  $dy$  to be zero. This goes away and becomes zero. The second thing is actually we don't care about  $x$ . We would like to get rid of  $x$  because it is this dependent variable. What we really want to do is express  $df$  only in terms of  $dz$ . What we need is to relate  $dx$  with  $dz$ . Well, to do that, we need to look at how the variables are related so we need to look at the constraint  $g$ . Well, how do we do that? We look at the differential  $g$ . So  $dg$  is  $g_{\text{sub } x} dx + g_{\text{sub } y} dy + g_{\text{sub } z} dz$ . And that is zero because we are setting  $g$  to always stay constant. So,  $g$  doesn't change. If  $g$  doesn't change then we have a relation between  $dx$ ,  $dy$  and  $dz$ . Well, in fact, we say we are going to look only at the case where  $y$  is constant.  $y$  doesn't change and this becomes zero. Well, now we have a relation between  $dx$  and  $dz$ . We know how  $x$  depends on  $z$ . And when we know how  $x$  depends on  $z$ , we can plug that into here and get how  $f$  depends on  $z$ . Let's do that. Again, saying that  $g$  cannot change and keeping  $y$  constant tells us  $g_{\text{sub } x} dx + g_{\text{sub } z} dz$  is zero and we would like to solve for  $dx$  in terms of  $dz$ . That tells us  $dx$  should be  $-g_{\text{sub } z} dz / g_{\text{sub } x}$ . If you want, this is the rate of change of  $x$  with respect to  $z$  when we keep  $y$  constant. In our new terminology this is  $\partial x / \partial z$  with  $y$  held constant. This is the rate of change of  $x$  with respect to  $z$ . Now, when we know that, we are going to plug that into this equation. And that will tell us that  $df$  is  $f_{\text{sub } x} dx + f_{\text{sub } z} dz$ . Well, what is  $dx$ ?  $dx$  is now  $-g_{\text{sub } z} / g_{\text{sub } x} dz + f_{\text{sub } z} dz$ . So that will be  $-f_{\text{sub } x} g_{\text{sub } z} / g_{\text{sub } x} + f_{\text{sub } z}$  times  $dz$ . And so this coefficient here is the rate of change of  $f$  with respect to  $z$  in the situation we are considering. This quantity is what we call partial  $f$  over partial  $z$  with  $y$  held constant. That is what we wanted to find. Now, let's see another way to do the same calculation and then you can choose which one you prefer. The other method is using the chain rule. We use the chain rule to understand how  $f$  depends on  $z$  when  $y$  is held constant. Let me first try the chain rule brutally and then we will try to analyze what is going on. You can just use the version that I have up there as a template to see what is going on, but I am going to explain it more carefully again. That is the most mechanical and mindless way of writing down the chain rule. I am just saying here that I am varying  $z$ , keeping  $y$  constant, and I want to know how  $f$  changes. Well,  $f$  might change because  $x$  might change,  $y$  might change and  $z$  might change. Now, how quickly does  $x$  change? Well, the rate of change of  $x$  in this situation is  $\partial x / \partial z$  with  $y$  held constant. If I change  $x$  at this rate then  $f$  will change at that rate. Now,  $y$  might change, so the rate of change of  $y$  would be the rate of change of  $y$  with respect to  $z$  holding  $y$  constant. Wait a second. If  $y$  is held constant then  $y$  doesn't change. So, actually, this guy is zero and you didn't really have to write that term. But I wrote it just to be systematic. If  $y$  had been somehow able to change at a certain rate then that would have caused  $f$  to change at that rate. And, of course, if  $y$  is held constant then nothing happens here. Finally, while  $z$  is changing at a certain rate, this rate is this one and that causes  $f$  to change at that rate. And then we add the effects together. See, it is nothing but the good-old chain rule. Just I have put these extra subscripts to tell us what is held constant and what isn't. Now, of course we can simplify it a little bit more. Because, here, how quickly does  $z$  change if I am changing  $z$ ? Well, the rate of change of  $z$ , with respect to itself, is just one. In fact, the really mysterious part of this is the one here, which is the rate of change of  $x$  with respect to  $z$ . And, to find that, we have to understand the constraint. How can we find the rate of change of  $x$  with respect to  $z$ ? Well, we could use differentials, like we did here, but we can also keep using the chain rule. How can I do that? Well, I can just look at how  $g$  would change with respect to  $z$  when  $y$  is held constant. I just do the same calculation with  $g$  instead of  $f$ . But, before I do it, let's ask ourselves first what is this equal to. Well, if  $g$  is held constant then, when we vary  $z$  keeping  $y$  constant and changing  $x$ , well,  $g$  still doesn't change. It is held constant. In fact, that should be zero. But, if we just say that, we are not going to get to that. Let's see how we can compute that using the chain rule. Well, the chain rule tells us  $g$  changes because  $x$ ,  $y$  and  $z$  change. How does it change because of  $x$ ? Well,  $\partial g / \partial x$  times the rate of change of  $x$ . How does it change because of  $y$ ? Well,  $\partial g / \partial y$  times the rate of change of  $y$ . But, of course, if you are smarter than me then you don't need to actually write this one because  $y$  is held constant. And then there is the rate of change because  $z$  changes. And how quickly  $z$  changes here, of course, is one. Out of this you get, well, I am tired of writing  $\partial g / \partial x$ . We can just write  $g_{\text{sub } x} \partial x / \partial z$  with  $y$  constant plus  $g_{\text{sub } z}$ . And now we found how  $x$  depends on  $z$ .  $\partial x / \partial z$  with  $y$  held constant is  $-g_{\text{sub } z} / g_{\text{sub } x}$ . Now we plug that into that and we get our answer. It goes all the way up here. And then we get the answer. I am not going to, well, I guess I can write it again. There was  $\partial f / \partial x$  times this guy, minus  $g_{\text{sub } z} / g_{\text{sub } x}$ , plus  $\partial f / \partial z$ . And you can observe that this is exactly the same formula that we had over here. In fact, let's compare this to make it side by side. I claim we did exactly the same thing, just with different notations. If you take the differential of  $f$  and you divide it by  $dz$  in this situation where  $y$  is held constant and so on, you get exactly this chain rule up there. That chain rule up there is this guy,  $df / dz$  with  $y$  held constant. And the term involving  $dy$  was replaced by zero on both sides because we knew, actually, that  $y$  is held constant. Now, the real difficulty in both cases comes from  $dx$ . And what we do about  $dx$  is we use the constant. Here we use it by writing  $da = 0$ . Here we write the chain rule for  $a$ . which is the same thing. iust

divided by  $dz$  with  $y$  held constant. This formula or that formula are the same, just divided by  $dz$  with  $y$  held constant. And then, in both cases, we used that to solve for  $dx$ . And then we plugged into the formula of  $df$  to express  $df$  over  $dz$ , or partial  $f$ , partial  $z$  with  $y$  held constant. So, the two methods are pretty much the same. Quick poll. Who prefers this one? Who prefers that one? OK. Majority vote seems to be for differentials, but it doesn't mean that it is better. Both are fine. You can use whichever one you want. But you should give both a try. OK. Any questions? Yes? Yes. Thank you. I forgot to mention it. Where did that go? I think I erased that part. We need to know -- -- directional derivatives. Pretty much the only thing to remember about them is that  $df$  over  $ds$ , in the direction of some unit vector  $u$ , is just the gradient  $f$  dot product with  $u$ . That is pretty much all we know about them. Any other topics that I forgot to list? No. Yes? Can I erase three boards at a time? No, I would need three hands to do that. I think what we should do now is look quickly at the practice test. I mean, given the time, you will mostly have to think about it yourselves. Hopefully you have a copy of the practice exam. The first problem is a simple problem. Find the gradient. Find an approximation formula. Hopefully you know how to do that. The second problem is one about writing a contour plot. And so, before I let you go for the weekend, I want to make sure that you actually know how to read a contour plot. One thing I should mention is this problem asks you to estimate partial derivatives by writing a contour plot. We have not done that, so that will not actually be on the test. We will be doing qualitative questions like what is the sign of a partial derivative. Is it zero, less than zero or more than zero? You don't need to bring a ruler to estimate partial derivatives the way that this problem asks you to. [APPLAUSE] Let's look at problem 2B. Problem 2B is asking you to find the point at which  $h$  equals 2200, partial  $h$  over partial  $x$  equals zero and partial  $h$  over partial  $y$  is less than zero. Let's try and see what is going on here. A point where  $f$  equals 2200, well, that should be probably on the level curve that says 2200. We can actually zoom in. Here is the level 2200. Now I want partial  $h$  over partial  $x$  to be zero. That means if I change  $x$ , keeping  $y$  constant, the value of  $h$  doesn't change. Which points on the level curve satisfy that property? It is the top and the bottom. If you are here, for example, and you move in the  $x$  direction, well, you see, as you get to there from the left, the height first increases and then decreases. It goes for a maximum at that point. So, at that point, the partial derivative is zero with respect to  $x$ . And the same here. Now, let's find partial  $h$  over partial  $y$  less than zero. That means if we go north we should go down. Well, which one is it, top or bottom? Top. Yes. Here, if you go north, then you go from 2200 down to 2100. This is where the point is. Now, the problem here was also asking you to estimate partial  $h$  over partial  $y$ . And if you were curious how you would do that, well, you would try to figure out how long it takes before you reach the next level curve. To go from here to here, to go from  $Q$  to this new point, say  $Q'$ , the change in  $y$ , well, you would have to read the scale, which was down here, would be about something like 300. What is the change in height when you go from  $Q$  to  $Q'$ ? Well, you go down from 2200 to 2100. That is actually minus 100 exactly. OK? And so  $\Delta h$  over  $\Delta y$  is about minus one-third, well, minus 100 over 300 which is minus one-third. And that is an approximation for partial derivative. So, that is how you would do it. Now, let me go back to other things. If you look at this practice exam, basically there is a bit of everything and it is kind of fairly representative of what might happen on Tuesday. There will be a mix of easy problems and of harder problems. Expect something about computing gradients, approximations, rate of change. Expect a problem about reading a contour plot. Expect one about a min/max problem, something about Lagrange multipliers, something about the chain rule and something about constrained partial derivatives. I mean pretty much all the topics are going to be there.