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. Today we try to wrap up our present discussion of partial derivatives by means of a rather practical example, which not only has wide application but which ties in many of the individual principles that we've talked about in the last two blocks of material. The particular topic that I have in mind today is the topic known as the theory of maxima minima of functions in several variables.

You see, in part one of our course we studied this special case where we had a function from the real numbers into the real numbers. And what we were looking for were values of the independent variable for which f was either maximum or minimum. And so a natural extension of this is simply the following, given a real-valued function of several real variables-- in other words, assume that f is a mapping from n dimensional space into the real numbers, f is a function from E n into E.

Then the n-tuple a bar in E sub n is called a local maximum. And I suppose we might as kill two birds with one stone here and put in the definition for a local minimum at the same time. It's called a local maximum of f if, and only if, there exists a neighborhood n of a bar such that f of a bar is greater than or equal to f of x bar for every x bar in the neighborhood.

I suppose, by the way, while we're at this, for a local minimum the condition would be what? For a local minimum, instead of f of a bar being greater than or equal to f of x bar, it would be less than or equal to. In other words, what you're saying is that what you mean by a local high point or low point is what? That in a neighborhood of the point in question, for example, if that output exceeds every other possible output in a sufficiently small neighborhood, then we say that that particular input is the local maximum. And a similar definition for local minimum.

In other words, again, keep in mind that this is precisely and almost a word for word translation of what the terms relative maximum and relative minimum or local max and local min meant for functions of a single real variable. Consequently, except for the fact that in several variables it's much more difficult to describe domains because of all the degrees of freedom that you have, one would expect that these same particular tests for membership would occur for functions of several variables when we're looking for high-low points or max-min points as in the case of one independent variable.

So let's just very quickly summarize this. How do we test for candidates for max-min points? Now, the idea is simply this, that if we're going to have a relative high or a relative low point, think of the thing graphically. If you take a cross section with respect to any one of the independent variables.

In other words, if you look at f as a function just of x 1, say, and hold x 2 up to x n constant, then when w is viewed as a function of x 1 alone, then one would expect that you have what? That just in the wx1 plane, you must have a candidate, which means that one would expect that if you took the partial of f with respect to x 1, that must be 0.

Similarly, the partial of f with respect to x 2 must be 0, et cetera. The partial of f with respect to x sub n must be 0. In other words, to find a candidate for a max-min point, notice that right off the bat we're back to a practical application of the study of systems of several equations and several unknowns.

Namely, we must solve the systems of equations, the partial of f with respect to x 1 equals 0, et cetera, the partial of f with respect to x n equals zero. Find simultaneous solutions, whenever such simultaneous solutions exist.

Now of course, there may be points in our domain where the partials don't exist, just like there was in the case of calculus of a single variable. In other words, the second thing we do in testing for candidates is we find points in the domain where f is not differentiable. For example, f might not be continuous. F might not be defined.

Whatever the reason is, we look to see where f is not differentiable. And all points in the domain at which f is not differentiable, they also become candidates for max-min points.

And thirdly, we check the boundary of the domain of f. In other words, if the domain of f happens to be a two dimensional region, we check the boundary of the region. Again, the reason being the same as in the calculus of a single real variable. If a function is differentiable, it must take on its maximum and minimum values someplace, if the domain happens to be a closed set, in other words, a connected set with a boundary. And we won't go into that in any more detail at this particular time.

But essentially, what do we do? We look at the function wherever it's differentiable, take all of its partials, set them equal to 0, solve that system simultaneously to see what values of x 1 up to x n give us permissible candidates for max-min points, we check to see where the derivative does not exist, and that gives us another batch of points, and then we check the boundary values to see if anything peculiar happens there. Same three tests as we had for calculus of a single variable.

And why is this so closely allied to calculus of a single variable? Well, if we take the case n equals 2, we again get a nice geometric interpretation. Namely, if w is a function of x and y, notice again our notation for what happens when we have two independent variables. They're called x and y, and usually we let the dependent variable be w in that case.

Let's take a look and see what happens over here. Suppose for the sake of argument that we know that we have a relative low point corresponding to the input a comma b. In other words, suppose we know that f of a comma b is the lowest value of f, the lowest height on the surface in a sufficiently small neighborhood of the point a comma b.

All I'm saying is this, take any slice whatsoever, any plane through the point a comma b, pick any direction s. Take that plane perpendicular to the xy plane in the s direction, and slice the surface w equals f of x, y with that plane, and we get a slice

something like this. Not something like this, this is the slice that we get.

Now, let's take a look at where that low point is. Since that is to be a low point in the entire region, obviously it must be a low point in particular with respect to the particular slice that we took. How could it be the lowest point every place if it's not the lowest point with respect to any particular slice?

And all we're saying then is that with respect to this slice, notice that w is a function of s alone. In other words, we can talk about the directional derivative, df ds. And again, all we're saying is that directional derivative, df ds, evaluated at the point a comma b must be 0 for all directions s. And as we have seen throughout our course, if s happens to be a continuously differentiable function of the variables x 1 up to x n, then the directional derivative is determined completely by our knowledge of the partials with respect to x 1 up to x n.

And you see, as far as the theory is concerned, that's all there is to it. As the cliche goes, the rest is commentary. Now what kind of commentary do I mean? Well, among other things, once we have located all the particular max-min candidates-- See, notice why I call these things candidates.

All we're saying is that wherever the system off possible derivatives, setting them equal to 0, yields a value, that point is simply a candidate for a max-min point. Just, again, like in the calculus of a single variable. If f prime of a is 0, we cannot conclude that a is a local max or a local min. it might be a saddle point, a stationary point, where the thing just levels off.

It's just that once we have the candidate, how do we test whether it really is a max or a min? Well, what we have to do is looking at f of a comma b, we have to see whether that's the lowest possible value or the highest possible value in a sufficiently small neighborhood of that point.

Well, how you characterize a neighborhood of the point? You look at some nearby point, which we can denote as what? a plus h comma b plus k. As if this h and k seem strange to you, notice that h is often what we call delta x, and k is what we

often call delta y. In other words, we look at some nearby point, a plus delta x comma b plus delta y.

And we're saying is that if this is to be, for example, a high point, it means that when you compute this difference, this difference had better be negative all the time. Because when you look at this particular thing over here, if this is to be the greatest possible value in a neighborhood when you subtract it from something else in the neighborhood, you should get a negative value.

In other words, to put it and still other terms, what we're saying is that to-- Let's just read this again. Once a max-min candidate a comma b is found, we must investigate the sine of f of a plus h comma b plus k minus f of a, b for all sufficiently small values of h and k. OK?

That's for all sufficiently small values of h of k, which can be messy. I don't mean that. I need for all sufficiently small values of h and k, in other words, in a neighborhood of the point a comma b. And what I'm saying is that this particular computation can itself be very messy.

You see? This is, again, going back to this idea of how we invert equations and the like. It's difficult to compute f of a plus h comma b plus k, in general, if f is a messy function, if f is a computationally complicated thing. And not only that, but we may have several unknowns, more than two unknowns.

And you see, this was true in one variable. We saw in the case of one variable that, technically speaking, to test whether f of a was a high point of a low point, we had to look at f of a plus h minus f of a for all sufficiently small values of h. And that could have been a messy computation too.

Of course, the thing that happened in the single variable that was very helpful to us was that in the case of a single variable, we were often able to use f double prime of a as a hint. In other words, whenever f double prime of a wasn't equal to 0, we can conclude, or could conclude, whether that f of a was a max or min men, once we know that f prime of a was 0.

Remember, that was that holding water versus spilling water routine. If f double prime of a was positive, that meant that the curve was holding water. Holding water meant that you had a minimum value. If f double prime of a was negative, that meant that the curve was spilling water. And spilling water yielded a maximum value.

And the only problem was is when the second derivative was 0, in which case the test failed. What did it mean that the test failed? When f double prime of a was 0, the only way we could test to see whether a was a max or a min on neither, meaning a saddle point, what was to actually look at f of a plus h minus f of a and see what happened in that particular case.

Now, one would like to believe that an analogous result held for the case of several real variables, in particular for the case of two independent variables. The point is that in a manner of speaking, it does. But in another manner of speaking, things are much more complicated than what happened in the case of one independent variable. In particular, what goes wrong is the following, and that is that the second derivative, in the case of two independent variables, involves three separate partials.

See what do you mean by a second derivative? You mean you must differentiate something twice. Well, you could've differentiated the function twice with respect to x at the point a comma b. That's what we mean by f sub xx, recall. Or you might have differentiated first with respect to x and then with respect to y, in which case it would have been f sub xy of a comma b.

In fact, I should be careful here. There should be a fourth one here too, and that is f sub yx. Namely, first differentiate with respect to y and then with respect to x. The reason I left that out over here was simply because if we have a nice function, meaning one that's continuous, and the derivatives are continuous, and the mixed derivatives exist and are continuous, we showed that the order in which we take the partials made no difference, that f sub xy was equal to f sub yx. But getting back to this idea, the third possibility is that you can have differentiated twice with respect to

y and formed and f sub yy of a comma b.

And so the question is, with all of these different second order partial derivatives floating around, what do you mean by the second partial derivative? And the key expression, and I think this is far from intuitive, but the key expression turns out to be that the determining factor is the second partial with respect to x, in other words f sub xx, multiplied by the second partial with respect to y, f sub yy, minus the square of the mixed partial. And that particular factor, the sine of that factor, determines whether you have a maximum point, a minimum point, a saddle point, or else the test might fail.

By the way, this is a rather difficult proof to come by. The proof is done in chapter 18 of the Thomas text, and is assigned for you. I tried to give you learning exercises that take you through the proof step by step. And I have also included an optional supplementary lecture for those of you who may still have difficulty following both the text and the supplementary notes and the learning exercises, because it's all written out and might like to hear the thing spoken. What I will do is derive this for you in an optional lecture for those of you want it.

But for the time being, to give us our overview, let me simply state what the properties of this particular quantity are. That the main result, and as I just write here to remind you, that the details are derived later, both in the text, in the learning exercises, and in optional lecture, that suppose we have solved our simultaneous system and have found values-- points a comma b, where the partial of f with respect to x and the partial of f with respect to y equals 0.

So we now have a candidate, meaning a comma b now is eligible to be tested to see whether it yields a maximum or a minimum value. The test turns out to be this, you compute f sub xx times f sub yy minus f sub xy squared at the point a comma b. And if that particular number turns out to be greater than 0, then a comma b yields a local minimum of f if f sub xx happens to be positive, and a local maximum if f sub xx happens to be negative.

Now, the easiest way to remember that is to think in terms of a partial derivative,

again. Imagine that we've sliced the surface so that we're looking at a cut in the wx plane. In the wx plane, notice that if the second derivative of w with respect to x is positive, that means that the curve is holding water. And holding water seems to indicate a minimum, you see. And similarly, if it's negative it's spilling water, and that would indicate a maximum.

You might say, what happens if f sub xx happens to be 0. And the answer is, look, if f sub xx happens to be 0, this case couldn't have occurred in the first place, because if f sub xx happens to be 0, this term drops out, in which case this could not be a positive expression. With this term missing, the smallest the square can be is 0, and a positive or non-zero number can't be negative. In other words, you could not obey this inequality if f sub xx happened to be 0.

If you want to argue why couldn't you look at f sub yy instead of f sub xx, the answer is f sub xx and f sub yy must both have the same sign. Because if we just look at this thing, notice that this term has to be positive. You're subtracting off something positive, therefore the term you're subtracting from must be positive. And the only way the product of two numbers can be positive is if each of the factors has the same sign.

But again, I don't want to belabor that. I just want to go through this thing fairly rapidly with you.

It turns out, by the way, if this key factor, this key term, f sub xx f sub yy minus the square of the mixed partial, happens to be negative, then you can be sure that you have a saddle point. In other words, what that means is for any neighborhood of the point a comma b, for some values of f, for some values of x comma y in that neighborhood, the function will be greater than f of a comma b, and for others it will be less than f of a comma b. So it's neither a max nor a min. And it turns out that the situation in which to test fails is if this particular expression happens to equal 0.

So again, you see that from a purely mechanical point of view, this test is rather easy to memorize. The hard part is the proof. And that's why, as I say, there's extra drill on that part, if you happen to be interested in it.

And by the way, if you're not interested in it, skip the proof. In fact, for the sake of this course, I am not concerned with how well you handle max-min problems. I'm interested more in showing you overall what max-min problems mean and how all of the principles of partial differentiation seem to come up in that particular application of max-min problems.

So from a theoretical point of view, that would complete the study of how one handles max-min problems, except that an even more difficult and subtle form of computational difficulty comes up, in terms of some of the practical applications that we have, which motivate such mechanical and computational devices known as, for example, the Lagrange multipliers and things of this sort that, again, are discussed in the text and in the learning exercises and which I will not discuss in the lecture other than to motivate for you why they occur.

And let me just finish up today's lesson in terms of one more topic. And this is a bad name for this topic, because we've already used this word in a different context. But this very often happens in mathematics that the same word is used in more than one context. But one often talks about constraints when one deals with functions of several variables. In other words, in many cases when one wants to maximize or minimize a function of several variables, it turns out that certain external conditions happen to be imposed.

Now, if that sounds like a difficult mouthful to comprehend, I would like to start off with an example that's already occurred in a max-min problem in part one of our course, but in such a subtle way that we never noticed that it was really involving a function of several variables. In fact, if it hadn't have been that subtle, we'd have been in trouble, because in part one of our course we did not talk about functions of more than one real variable.

But for example, let's revisit a type of problem that says let's find the minimum distance, say, from the origin, 0 comma 0, the minimum distance from the origin to the curve xy equals 1. Well, to find that minimum distance, notice that what we have to do is minimize a distance function, namely the square of the distance-- I use the

square simply to eliminate the square root sign here-- the square of the distance from the origin to any point x comma y is x squared plus y squared.

Well, notice that in this form, this is a function of two independent variables. The trouble is, you don't want the distance from the origin to any old point. The point that you're investigating has to be on the curve of xy equals 1. And that means that x and y, for your investigation in this problem, are not independent. Namely, x and y are related.

Now by the way, sometimes this equation can be so messy that we cannot solve for y explicitly in terms of x. In this particular case, as long as x is not 0, we can solve specifically for y in terms of x, in which case we get y equals 1/x. And that is, in this particular case then, the function that we want to minimize, even though it looks like a function of two independent variables, is really a function of one independent variable, because since y is equal to 1/x in our investigation, f of x comma y is really f of x comma 1/x. In other words, going back to what f is explicitly in this case, f of x comma y is just x squared plus 1/x quantity squared. And that in turn is also just a function of x.

Now you see, what makes this thing more difficult is, first of all, we may not be dealing with a function of just two variables to minimize. And secondly, we may have a very difficult constraint imposed. Or what makes things even more difficult is that if we have to minimize a function, say, a five variables, there may be two or three are even four constraints imposed.

And the question is, how do you maximize or minimize a function, taking into consideration the fact that there are constraints imposed? And this is what brings up all of the type of material that I gave you as exercises in the last unit, where we talked about the Jacobian matrix, and handled the inverting systems, and how we could solve explicitly or implicitly for functions, implicit function theorems, things that we talked about in those exercises. All of those things come up in solving max-min problems in several unknowns.

See, more generally, what we're saying is we have more variables than two, and we

have more implicit constraints, meaning more constraints than just a single constraint, and also using more variables, and also more involving modifying implicit that you just can't solve for one of the variables explicitly in terms of the other, even though there is an implicit relationship involved.

Now, I hate to work abstractly in general, but in this particular lecture I'm going to do that. I'm going to talk with you abstractly here, but all of the exercises will deal with concrete situations so that you'll see all of the theory come alive in the problems. But because I want to give you the material as compactly as possible, let me just state what the situation is.

For example, suppose we want to maximize or minimize a function of what appears to be three independent variables, say, f of x, y, z. And all of a sudden, somebody tells us, hey, in the domain that we're interested in, x, y, and z are not independent. There's a certain constraint.

And let's write that symbolically as some function of x, y, and z equals 0. Remember, that's simply our abstract way of saying that there is some functional relationship that relates x, y, and z. And we'll simply call that g of x, y, z equals 0. Maybe I could solve for z explicitly in terms of x and y from this particular relationship. Maybe I can.

But at any rate, what we do know from the learning exercises of last time, that g of x, y, z equals 0 will implicitly define z as some function of x and y, say k of x, y, except in that case where the partial of g with respect to z happens to be 0. And I put this thing in parentheses for you simply to give you motivation to review the exercises of last time if this seems a bit vague to you.

At any rate, what this thing means in plain English is that subject to the constraint, that z is some function of x and y. We take this constraint, we put that back into our original function that we're trying to maximize. Notice that f of x, y, z now becomes f of x, y, k of x, y. See? z is k of x, y.

If we now look at this expression, notice only x and y appear so that subject to the

constraint g of x, y, z equals 0, the function, f, is a function of only two independent variables, not three. And to indicate that, let's simply say that f of x, y, z, f of x, y, k of x, y in this case, is some function h of x and y. And what our problem is now saying is, minimize or maximize the function h, which is a function of two independent variables.

Now notice here-- it's been quite a while since we've dealt with the chain rule-- but notice here that the chain rule now comes up in a very important practical application, namely, this looks like an eyesore f of x, y comma k of x, y. How can we handle that?

Notice there's another way of saying this, utilizing the chain rule, is to say h of x, y is f of x, y, z where z is some function k of x and y. See, f is a function of x, y and z, and z is a function of x and y. In fact, if you wanted to say it another way, you could say what? f is some function of x, y, z when x equals x, y equals y, and z equals k of x, y. You see? That's your particular chain rule.

Now, look at it. See, this is, again, one of the problems of mathematics, which I hope is crystal clear by this time. Granted, that we discussed the chain rule in block three. That's no reason why in block four we can beg off and say, we had a long time ago, I don't remember that. No. Hopefully, we've made the chain rule so clear that any time I tell you that we have to invoke it, you can just write it down very, very quickly.

How does the chain rule work now? To differentiate this with respect to x, say, we take the partial of this with respect to x, times the partial of x with respect to x, plus the partial of f with respect to y times the partial of y with respect to x, plus the partial of f with respect to z times the partial of z with respect to x.

In other words, the general theory hasn't changed at all. To maximize or minimize h, what we're going to do is we're going to take the partial of h with respect to x and set that equal to 0, we're going to take the partial of h with respect to y and set that equal to 0, and solve that system simultaneously. But the hard point computationally is, sure, you can say let's set the partials is equal to 0. But before you can do that,

you had better be able to take the partials. And that's where the computationally skill comes in.

So how do we take the partial of h with respect to x here? Well, we just said that. The possible of h with respect to x using the chain rule is f sub x times the partial of x with respect to x, plus f sub y times the partial of y with respect to x, plus f sub z times the partial of z with respect to x.

Now notice that the partial of x with respect to x is 1. So that gives me f sub x over there. Keep in mind that whereas z is a function of x and y-- let's go back here and take a quick look at that-- noticed that in our function h, we're assuming what? That x and y are the independent variables, but that z is a function of x and y.

The point, therefore, is that since x and y are independent variables, by definition that means that the partial of y with respect to x is 0. In other words, the fact that y and x are independent means that the change in y with respect to a change in x is 0, because we can change x without changing y.

And finally, given that z is k of xy, we can compute the partial of z with respect to x. And so, what we wind up with is that the partial of h with respect to x is the partial of f with respect to x, plus the partial of f with respect to z times the partial of z with respect x. And we must set that equal to 0.

Again, leaving the details to you as a review of the chain rule. In a similar way, we show that the partial of h with respect to y is the partial of f with respect to y, plus the partial of f with respect to z, times the partial of z with respect to x, and we set that equal to 0.

Now, the thing to keep in mind is to observe that f of x, y, z was a given function. We know what that looks like. Consequently, these four quantities are known. The trouble is that g of x, y, z equals 0 defines z implicitly as a function of x and y. And consequently, if it turns out that we could not have solved our system for z explicitly in terms of x and y, the question mark would be, for example--

I'm sorry. This is a misprint over here. This should be, of course, the partials would

differentiate with respect to y. This is the partial of f with respect to y, plus the partial of f with respect to z, times the partial of z with respect to y.

The key point over here is simply that we must know what? If we can't solve for z explicitly in terms of x and y, how do we know what these two quantities are? We only know them implicitly. And this gives us our review, again, of setting differentials equal to 0 and the like. Namely, if g of x, y, z is identically 0, we can equate the derivative on both sides to 0.

And to differentiate this thing implicitly, it's, again, using the chain rule, we get what? The partial of g with respect to x, plus the partial of g with respect to z, times the partial of z with respect to x is 0.

Notice, by the way, that g is given explicitly. We're told what the function g of x, y, z looks like. Consequently, these things here are known. And we can now solve for the partial of z with respect to x.

In fact, what is the partial of z with respect to x? That's nothing more than what? It's minus the partial of g with respect to x divided by the partial of g with respect to z. And that's why the partial of g with respect to z had better not be 0, otherwise we wind up in trouble here.

Similarly, we can find what the partial of z with respect to y is. That's going to turn out to be what? Minus the partial of g with respect to y divided by the partial of g with respect to z. Knowing what these two values are from these two equations, we come back into the here. And now we know what these are, and now we simply have all of the known functions on the right hand-side, and we solve this system.

Now hopefully, this is one of the times I hope things don't sound too clear to you, meaning you have an overview, but you begin to suspect that this is a very messy situation. Because, you see, what I wanted to see is that solving such a system like this can be extremely cumbersome. And that's why, in the exercises, we do two things. We have to solve systems like this to get the experience of seeing that there's a big difference between knowing theoretically how to solve a system of

equations and knowing pragmatically how to carry it out.

And secondly, we hope that some of these computations become cumbersome enough so that you practically beg to find some shortcuts. Because if you're not begging to find a shortcut, then such things as Lagrange multipliers and the like, which are techniques for solving max-min problems subject to constraints, that those shortcuts don't appeal to you and you fail to see their significance, and you say, why do I have to learn these things. Don't I have enough problems without it?

You see, again, notice the difference between what's happening here abstractly and what's happening computationally. Abstractly the theory of max-min is not that difficult. But computationally, to handle it, you have to be extremely adept at handling systems of n equations and n unknowns, not necessarily linear equations, being able to throw in constraints and seeing what's happening here. And at any rate, that's what the learning exercises will be all about and the material in the text.

And I think that's enough of a mouthful for this time. So until next time, good bye.