IIT Madras

ONLINE DEGREE

Mathematics for Data Science - 2

Professor Sarang Sane
Department of Mathematics
Indian Institute of Technology, Madras

Higher order partial derivatives and the Hessian matrix

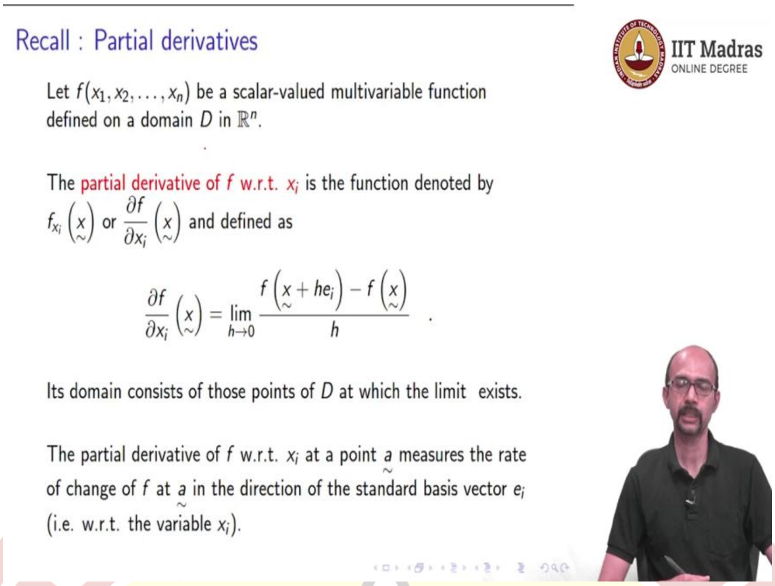
Hello, and welcome to the Maths 2 component of the online BSc program on Data Science and Programming. This video is about higher order partial derivatives and the Hessian matrix. So, in the earlier video, we studied the notion of critical points for multivariable functions. So, what that meant was that you take the function f is a scalar valued multivariable function, and then you find its gradient and then you set it to 0.

And you compute all those points for which those equations are satisfied, which means that the gradient at that point is 0 or the other thing that can happen is that one or more partial derivatives are undefined. So, you collect together all those points and those are called the critical points of f.

And the reason we were interested in studying these was that the local extrema, meaning local maxima or local minima are all critical points. So, they satisfy that the gradient is 0, or maybe some partial is not defined. So, they must satisfy one of these. And the intuition there was that the tangent plane is parallel to the x-y plane or in general $x_1, x_2, .., x_n$ hyperplane.

In one variable calculus, we have studied that there is a second order test, meaning a test using the second derivative, and that allows us to classify which of these critical points are local maxima, local minima or saddle points, and sometimes it can fail also. So, we would like an analogue of such a test for the multivariable case. And in order to do that, we will first study the notion of higher order partial derivatives and the Hessian matrix.

(Refer Slide Time: 2:10)



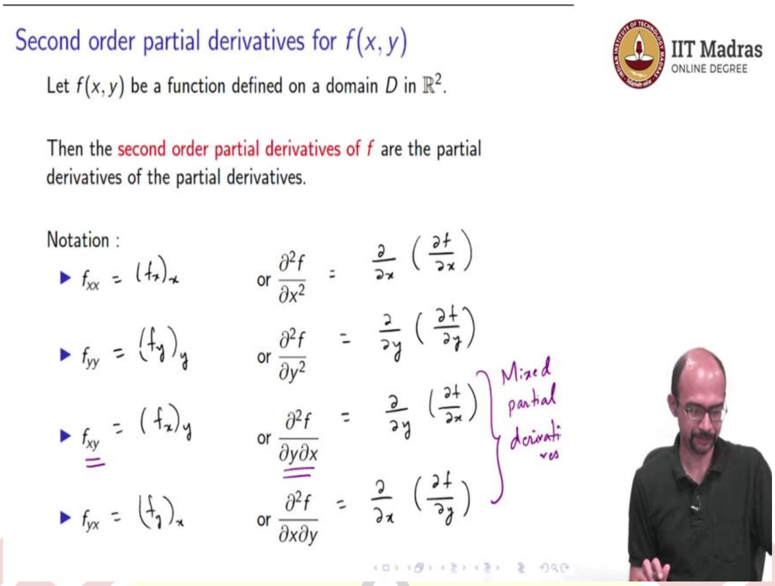
So, let us recall what are partial derivatives first, suppose $f(x_1, x_2, .., x_n)$ is a scalar valued multivariable function defined on a domain D in R^n. The partial derivative of f with respect to the variable x_i is the function denoted by $f_{x_i}(\underset{\smile}{x})$, or $\frac{\partial f}{\partial x_i}(\underset{\smile}{x})$. And it is defined by taking the limit $\lim_{h\to 0}\frac{f\left(\underset{\smile}{x}+he_i\right)-f(\underset{\smile}{x})}{h}$.

So, this is the definition of the partial derivative function of f with respect to the i^{th} variable x_i. And this domain, it need not always, this limit may not always be, it need not always exist. And so the domain of this function, the partial derivative is all those points within D where this limit does exist.

And what does the partial derivative do, the partial derivative at a point measures the rate of change of the function in the direction of the standard basis vector e_i at the point $\underset{\smile}{a}$. So, or equivalently with respect to the variable x_i, meaning it treats the function as a function of only x_i.

So, you restrict to that line parallel to the xi axis passing through a tilde and restrict your function to that line and then ask what is the, how is the function changing? What is the rate of change of that function? That is exactly what the ith partial derivative does.

(Refer Slide Time: 3:53)



So, now the partial derivative is a function. And so, you can say what is the partial derivative of that function, the partial derivative is a scalar valued multivariable function, and we can very well ask what is the partial derivative of that function and that is exactly what are second order partial derivatives for $f(x, y)$.

So, if $f(x, y)$ is a function defined in a domain D in R^n, then the second order partial derivatives of f are the partial derivatives of the partial derivatives. So, you take partial derivatives and then you again take partial derivatives. So, just for the sake of simplicity, the notations are either f_{xx} or $\frac{\partial^2 f}{\partial x^2}$. So, note this strange thing that the notation here is $\frac{\partial^2 f}{\partial x^2}$ and not $\frac{\partial^2 f}{\partial^2 x}$. So, this is for something that will come ahead.

The other possibility is that you have f_{yy} which is also $\frac{\partial^2 f}{\partial y^2}$ and the third possibility is that you have f_{xy}, which is the same as $\frac{\partial^2 f}{\partial y \partial x}$ or and the fourth possibility is f_{yx}, which is $\frac{\partial^2 f}{\partial x \partial y}$. So, what do these mean? So, f_{xx} means, you consider the partial derivative f with respect to x, of f with respect to x and then take the partial derivative of that function again with respect to x.

So, f subscript x again subscript x or the other way of writing this del x of the function del f del x. So, del f del x is the function the partial with respect to x and then you take the partial again with respect to x. Similarly, if you have f_{yy}, what that means is you take partial with respect to y and then again take partial with respect to y equivalently you could write it as del del y of del f del y. What is f_{xy}?

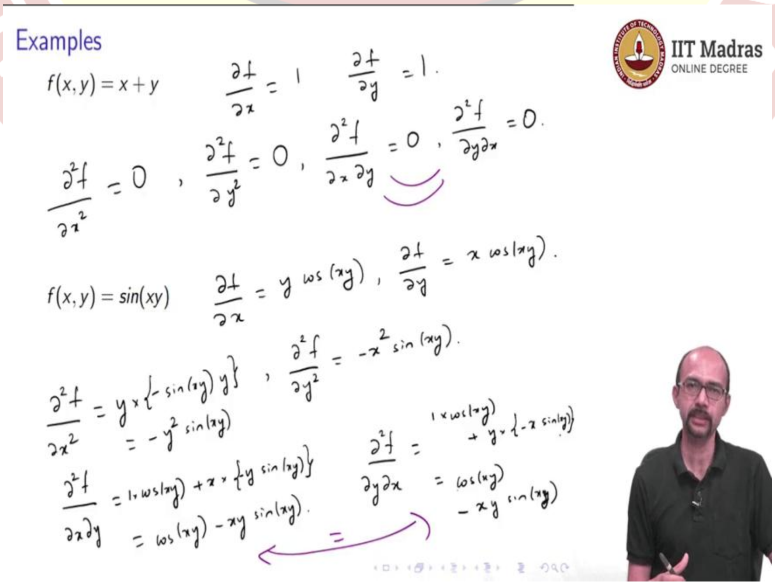
So, this is where you first take the partial with respect to x and then take the partial with respect to y of that function, which is the same as saying you take del del y of del f del x and then here this is f_{yx} and this is del f del x of del f del y. So, now, just notice one small thing, this is a fairly obvious but sometimes students make mistakes with this.

In the notation when you do f_{xy} so, here the order is first differentiate with respect to x, then take partial with respect to y, but when you write it like this, it gets kind of flipped over, so you have to remember that this means first take partial with respect to x, and then take partial with respect to y. So, because of the notation on the right is sort of saying you take del del of something and that is why it is written like this.

Whereas, on the left, it is x is coming on the right, the variables are coming on the right, that is why you write it like this. So, you have to just remember that when you go between these 2, whatever is written here, gets it in the opposite order towards the left, so the subscript notation gets written towards the right, the del del notation gets written towards the left, this is what you should remember, because this is a common source of error.

Now, it would be of course, convenient if we could flip these as we want. Like, if we have f_{xy} and f_{yx}, then we can interchange them. That would be very convenient. And indeed, we will soon see a theorem that says that if your function is nice, meaning under some hypothesis, indeed, that does happen.

(Refer Slide Time: 8:04)

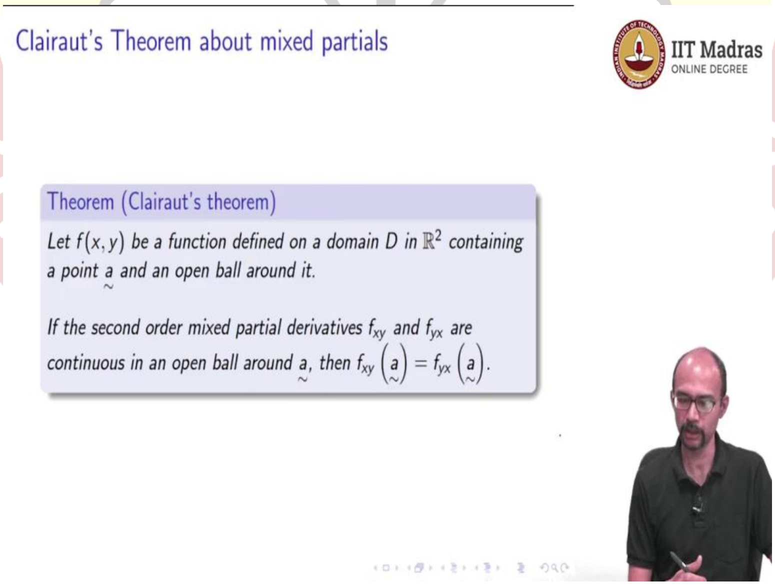


Let us start with some examples. So, let us compute all the second order partial derivatives of these functions. So first, what are the first order partial derivatives, $\frac{\partial f}{\partial x}$. So, we have actually computed these before it is 1, $\frac{\partial f}{\partial y}$ is 1. And so $\frac{\partial^2 f}{\partial^2 x}$ is 0, because $\frac{\partial f}{\partial x}$ is a constant, $\frac{\partial^2 f}{\partial^2 y}$ is similarly 0, in fact all these functions are 0; $\frac{\partial^2 f}{\partial x \partial y}$ is also 0 and so is $\frac{\partial^2 f}{\partial y \partial x}$..

And indeed do notice that here, these two are equal. So, at least for easy functions, it may happen that these two are often equal. So, it does not matter in which order you do your partial differentiation, the answer will be the same. Let us see if that holds for this as well. So, you have $f(x, y) = \sin(xy)$, we have again computed the gradient here, so we know what the partial derivatives look like.

So, the first order partial derivatives are $\frac{\partial f}{\partial x} = y\cos(xy), \frac{\partial f}{\partial y} = x\cos(xy)$. Now, let us look at the higher order partial derivatives. So, $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \cos(xy) - xy\sin(xy)$, $\frac{\partial^2 f}{\partial y \partial x} = \cos(xy) - \ xy\cos(x)$, $\frac{\partial^2 f}{\partial x^2} = -y^2\sin(xy), \frac{\partial^2 f}{\partial y^2} = -x^2\sin(xy)$.

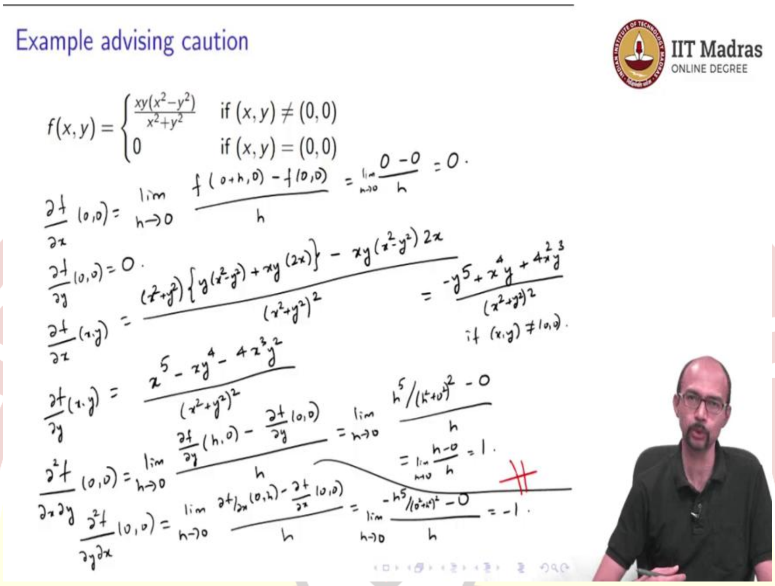
So, it seems that the mixed partials are indeed often equal. Maybe I should have mentioned in the previous slide that these are called the mixed partials, the mixed partial derivatives. So, as I was saying, the mixed partial derivatives, at least in these examples seem to be equal. (Refer Slide Time: 12:27)



And indeed, we have a very nice theorem called Clairaut's theorem about the mixed partials. So, it says the following, so suppose $f(x, y)$ is a function defined on a domain D in R^2 containing a point A and an open ball around it. If the second order mixed partial derivatives, f_{xy}, f_{yx} are continuous in an open ball around A, then they are equal at A.

So, what do we need to check that these mixed partial derivatives are continuous in an open ball around A. Now, in our previous examples, they were all very nice functions. So, continuity was not at all a problem. And from there, we could as a result, apply Clairaut's theorem. And we saw that they are equal although we actually computed them, but we could have directly concluded it from Clairaut's theorem.

(Refer Slide Time: 13:18)



So, let us see an example where we will try to compute all these and see that the hypothesis is indeed necessary. So, let us compute what are the partial derivatives. So, $\frac{\partial f}{\partial x}(0,0) = \lim_{h\to 0} \frac{f(0+h,0)-f(0,0)}{h} = 0, \frac{\partial f}{\partial y}(0,0) = 0$ by the same argument.

So, we know that these two are 0. Of course, we have to compute now the partial derivatives at the other points as well. So, now let us do that. So, what is $\frac{\partial f}{\partial x}(x,y)$ where (x,y) is not equal to (0, 0). Then this is a rational function and the denominator is non-zero so I can apply my u bv v rule. So, if we do that, let us see what we get.

So, $\frac{\partial f}{\partial x}(x,y) = \frac{(x^2+y^2)\{y(x^2-y^2)+xy(2x)\}-xy(x^2-y^2)2x}{(x^2+y^2)^2} = \frac{-y^5+x^4y+4x^2y^3}{(x^2+y^2)^2}$.

And you can do the same thing for $\frac{\partial f}{\partial y}$. And if we do that, what you will get is, again, you use a u by v rule and go through the expressions, you will get exactly the same thing. Except that now the x term the highest x term comes with a positive sign and the lower ones come with a negative sign. So, the numerator is $x^5 - xy^4 - 4x^3y^2$.

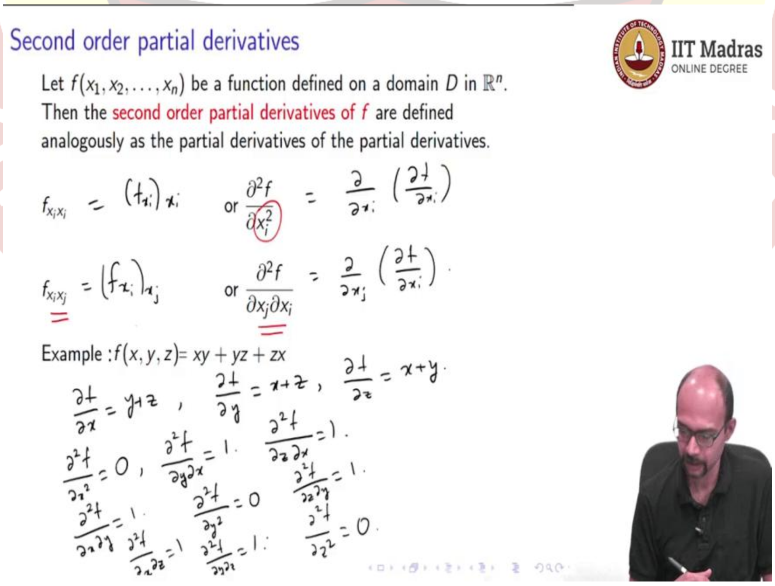
So, it is kind of symmetric, except that there is this gap between the plus and the minus and that gap is exactly what is going to cause the second order partials to not work out properly. So, so this is what is. And now let us compute what is the second order partial, mixed partials for at 0, 0. So, at 0, 0, I want to compute what is $\frac{\partial^2 f}{\partial x \partial y}(0,0) = \lim_{h \to 0} \frac{\frac{\partial f}{\partial y}(h,0) - \frac{\partial f}{\partial y}(0,0)}{h}$.

So, this limit simplifies to the value 1. Now, let us compute $\frac{\partial^2 f}{\partial y \partial x}(0,0)$. So, this is the limit $\lim_{h \to 0} \frac{\frac{\partial f}{\partial x}(0,h) - \frac{\partial f}{\partial x}(0,0)}{h}$.

So, you can see now this is -1 and so these do not match. So, the values of the mixed partials do not match.

So, what was the problem? The problem was that these mixed partials are actually not continuous. So, for that, you will have to evaluate what they are and check what happens. So, the hypothesis in Clairaut's theorem is important, and without that, it may not happen that the mixed partials are indeed equal. So, for whatever is next we will assume in that the hypothesis of Clairaut's theorem holds, because we want that to hold in order to get whatever the second order partials are useful for fine.

(Refer Slide Time: 20:21)



So, we have done second order partial derivatives for $f(x, y)$. Now, let us talk about second partials for $f(x_1, x_2, .., x_n)$. So, if $f(x_1, x_2, .., x_n)$ is a function defined on the domain D in R^n, then the second order partial derivatives are defined analogously as a partial derivatives of the partial derivatives. So, here there are for the n=2 case there were 4 second order partial derivatives. So, for the general case there are going to be n^2 second order partial derivatives.

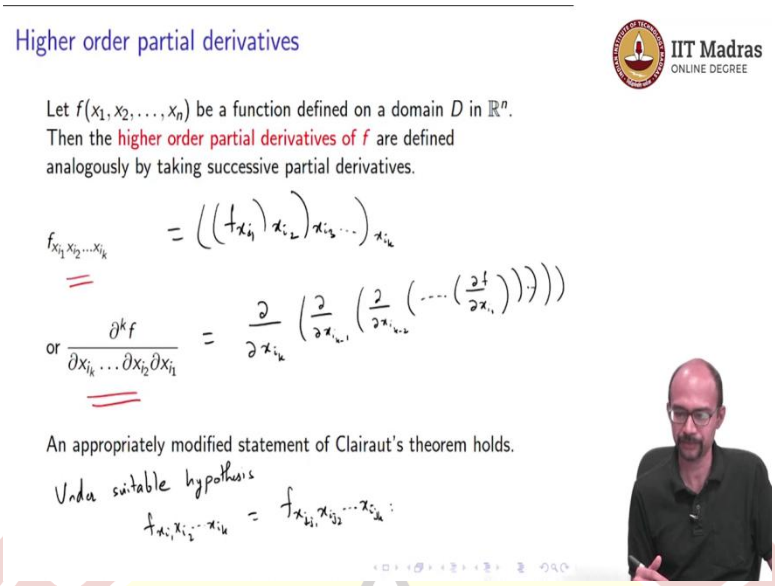
So, they are given by $f_{x_i x_i}, \frac{\partial^2 f}{\partial {x_i}^3}$. So, again, important point is to notice here how here it is written on the right for the mixed partials, and so it goes, the order is from left to right, here it is written towards the left and so the order is from right to left. That is the only thing you have to recall, remember, as far as notation is concerned. And the other difference meaning why did I write it differently? See, here the notations are same x_i, x_j. So, it does not matter if i and j are different or i and j are the same, but when you write it in terms of this del notation, the $\partial x_i\, \partial x_i$ becomes $\partial {x_i}^2$. This is just a convention, there is no reason for this notation other than convention.

Fine, let us do this example. If $f(x, y, z) = xy + yz + zx$, let us first compute the partial so $\frac{\partial f}{\partial x} = y + z, \frac{\partial f}{\partial z} = y + x, \frac{\partial^2 f}{\partial x^2} = 0, \frac{\partial^3 f}{\partial y \partial x} = 1, \frac{\partial^3 f}{\partial x \partial z} = 1, \frac{\partial^2 f}{\partial y^2} = 0, \frac{\partial^2 f}{\partial z^2} = 0$. Again we have done this example before meaning the partials first order partials.

And I think you can see the general pattern emerging in these partial derivatives.

So, from Clairaut's theorem or an equivalent version of the theorem in n dimensions, you can check that $\frac{\partial^2 f}{\partial x \partial y} = 1, \frac{\partial^2 f}{\partial z \partial x} = 1$. So, you can check this directly from the expressions for the first order partials or you can use the fact that Clairaut's theorem applies.

(Refer Slide Time: 24:35)



So, I will make that as a comment in this slide. So, before which let us define the higher order partial derivatives. So, you can keep repeating this process. So, you have the second order partials you can take the third order partials, fourth order partials. So, for one variable functions, you can differentiate as many times as you want provided the derivative exists. So, the same thing can be done for partial derivative, you can take partial derivatives as many times as you want. And that is exactly what higher order partial derivatives do.

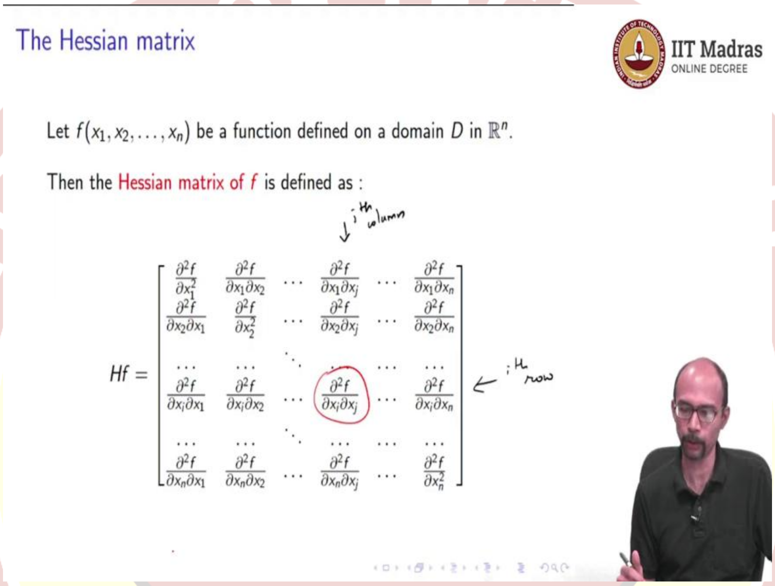
So, if $f(x_1, x_2, .., x_n)$ is a function defined in a domain D in R^n, then the higher order partial derivatives of f are defined analogously by taking successive partial derivatives. So, the general notation so, if you take the kth order partial derivative, the notations are $f_{x_{i1}x_{i2}\ldots x_{ik}}$ which means you take the partial with respect to x_{i1}, then you take the partial with respect to x_{i2}, then you take the partial with respect to x_{i3} and all the way up to x_{ik}, this is what it means.

And you can write the same thing except remember our usual thing about the shift in the order. So, this is exactly the same as taking $\frac{\partial}{\partial x_{ik}}(\frac{\partial}{\partial x_{ik-1}}\left(..\left(\frac{\partial f}{\partial x_{i1}}\right)\right))$. Of course, once again one should point out that there is no guarantee that this higher order partial derivative exists the same way as there is no guarantee that the first order partial derivative exists.

So, there will be some points where it exists some points it may not exist and so, the domain of this function is whichever points it exists and an appropriately modified statement of Clairaut's theorem holds. What that means is, if suppose you take the kth order partial derivatives, then if all the, if this guy is continuous and so is in the neighborhood and so is any, you take this denominator in some other order, if that k^{th} order partial is also continuous in some neighborhood, then they are going to match at the point, at any point in that neighborhood.

So, in general what you can say is under suitable hypothesis, you can shift the orders without any problem. So, if I do x_{i1}, x_{i2},., x_{ik} that is the same as doing it in some other order x_{ij1}, x_{ij2}, ..,x_{ijk}.

(Refer Slide Time: 28:12)

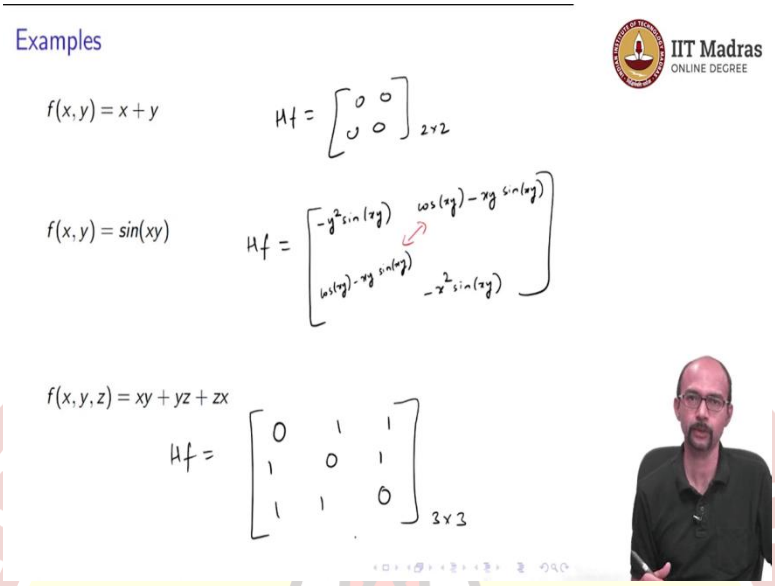


Finally, let us talk about the Hessian matrix which is really what all this is going to be used for. So, the Hessian matrix is the following if you have a function of n variables defined on domain D in R^n,, then the Hessian matrix of f is defined as this complicated looking $n \times n$ matrix. So, remember that you had n squared second order partial derivatives, so, you place them in a matrix.

So, the first row is you take all your, you take your gradient vector and then you take the partial with respect to x_1 of each component of that gradient vector. So, the first row of the matrix contains the derivatives $\frac{\partial^2 f}{\partial x_1{}^2}, \frac{\partial^2 f}{\partial x_1 \partial x_2}, \ldots, \frac{\partial^2 f}{\partial x_1 \partial x_j}, \ldots, \frac{\partial^2 f}{\partial x_1 \partial x_n}$.

So, the important part is you have to remember what happens for the $(ij)^{th}$ entry. That is how you describe any matrix by saying what happens to the $(ij)^{th}$ entry. So, the $(ij)^{th}$ entry is given by taking $\frac{\partial^2 f}{\partial x_i \partial x_j}$. This is what if you remember. So, that is what is happening. So, this is the Hessian matrix and this is somehow going to come in, this is somehow going to be useful when we talk about the second derivative test whatever, so the analogy of the second derivative test when we classify critical points.

(Refer Slide Time: 30:04)



Fine, let us do a couple of examples. So, for this example, $f(x, y) = x + y$, we saw that the second order partials are actually all 0. So, the Hessian here is a null matrix as shown. So, this was a relatively complicated Hessian, $f(x, y) = \sin(xy)$. So, it is of, it is a 2×2 because you have two variables. And the terms here are. So, they are $-y^2\sin(xy), \cos(xy) - xy\sin(xy), \cos(xy) - xy\sin(xy), -x^2\sin(xy)$. So, these off-diagonal terms typically match. So, often for the Hessian matrix, these terms will match.

And then let us do this $f(x, y) = xy + yz + zx$. This one was again rather easy matrix. Now, this is a 3×3 matrix, why 3, because there are 3 variables. So, what was this matrix? The matrix is as shown above. The kind of matrices we found in linear algebra were somewhat useful. And indeed, we will have some use of linear algebra when we study this Hessian matrix in the context of the classification of critical points. Thank you.