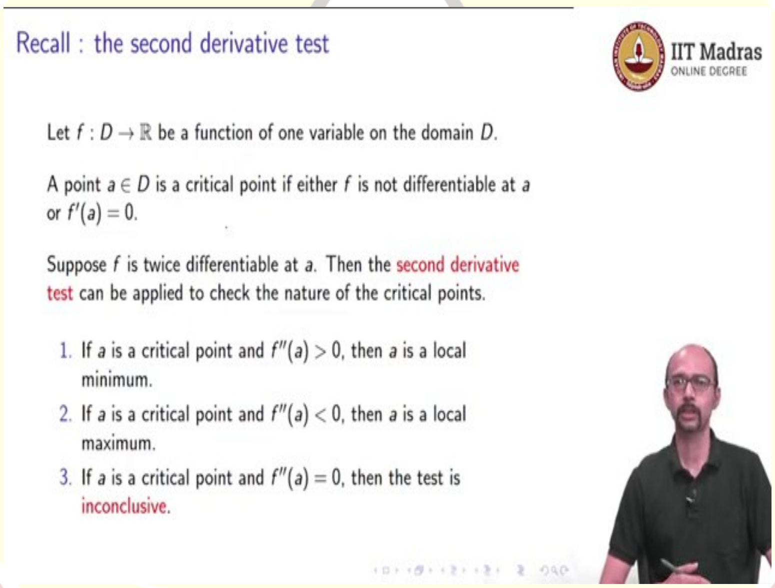
IIT Madras

ONLINE DEGREE

Mathematics for Data Sciences - 2
Professor. Sarang Sane
Department of Mathematics
Indian Institute of Technology, Madras
The Hessian matrix and local extrema for f(x,y)

Hello and welcome to the match 2 component of the online Bsc program on Data Science and Programming. In this video we are going to talk about the Hessian matrix and local extrema. So, we have studied the notion of critical points for f(x, y) and in previous video we also studied the Hessian matrix which consisted of the second order partial derivatives placed in a square matrix. So, in the context of f(x, y) this will be a 2 by 2 matrix.

(Refer Slide Time: 0:43)

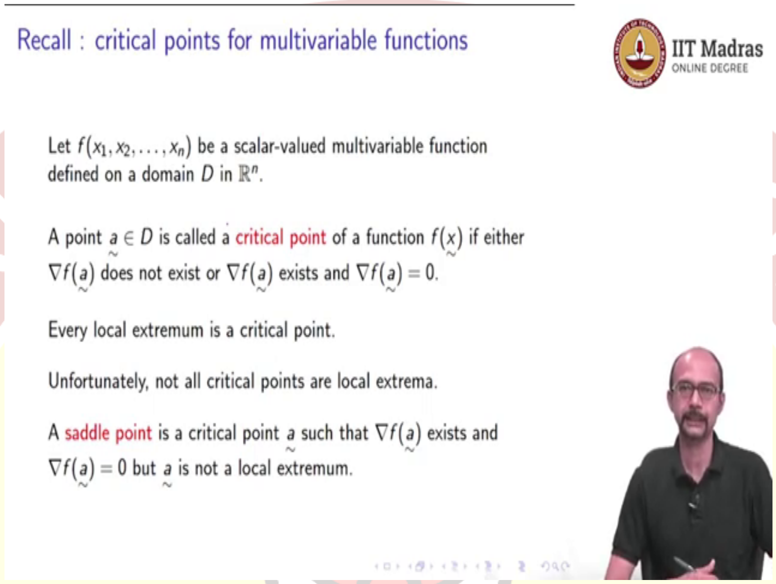


Let us recall first what was the second derivative test for a one variable function. So, suppose you have a function of one variable defined in the domain D, a point $a \in D$ is a critical point if either the function f is not differentiable at that point or $f'(a)=0$ And now if the function is twice differentiable, we applied something called the second derivative test, so the second derivative test told us the nature of the critical points.

So, the test said the following; if a is a critical point and if the double derivative is positive, then this point is a local minimum, if it is a critical point and the double derivative is negative, then it is a local maximum and if it is a critical point and the double prime is 0, then the test is inconclusive, so we do not know what happens in that case.

Fine, so this was the second derivative test for functions of one variable. Now, we would like to have something analogous for higher order meaning functions, multivariable functions and indeed the Hessian matrix is what is going to allow us to do that.

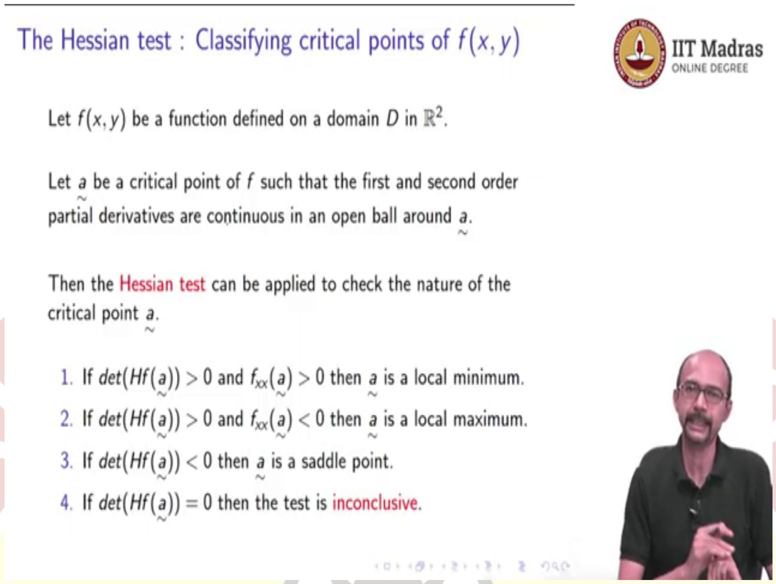
(Refer Slide Time: 1:56)



So, let us first recall critical points for multi-variable functions. So, if we have a scalar valued multivariable function defined in our domain D in $\mathbb{R}^n$ then a point $a \in D$ is called a critical point if either the gradient is 0 or the gradient does not exist. So, the main point here was that every local extremum, meaning a local maximum or a local minimum was a critical point, so every local extremum is a critical point.

And what we want to do is once we get the collection of critical points we want to identify which of these are local maxima, which of these are local minima and if possible if which of these are saddle points. So, we know that not all critical points are local extrema because there are things called saddle points and the saddle point is exactly a critical point such that the gradient exists and the gradient is 0 but a is not a local extrema. So, let us now see how we can use the hessian in order to classify these critical points.

(Refer Slide Time: 2:59)

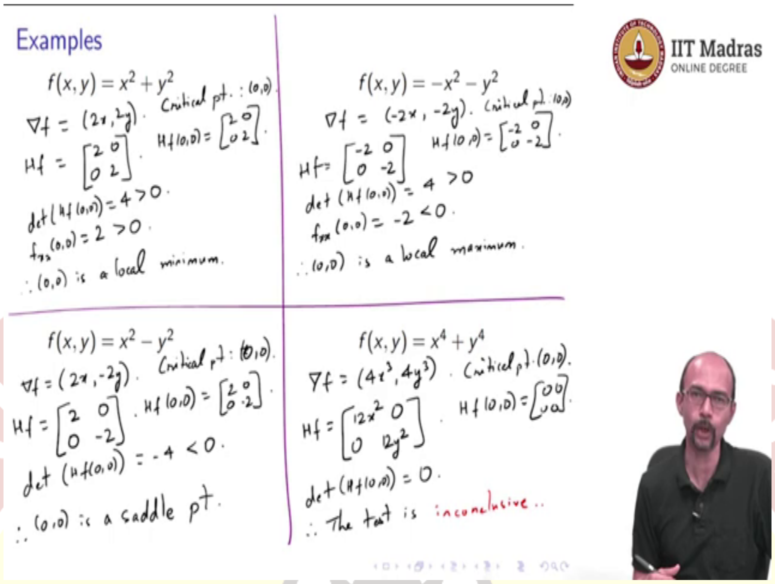


So, again if f (x, y) is a function defined in the domain D and a is a critical point such that the first and second order partial derivatives are continuous in an open ball around a. So, this is a very important hypothesis, in particular this hypothesis allows us to apply Clairaut’s theorem and say that the mixed partials are going to be equal.

So, then the hessian test can be applied to check the nature of the critical point a and what is the test? If the determinant of the Hessian matrix at that point is positive and the second order partial with respect to x at that point is positive, then it is a local minimum, if the determinant is positive and the second order partial with respect to x is negative, then it is a local maximum and if the determinant is negative, then it is a saddle point, and if the determinant is 0 then we do not know, the test is inconclusive.

So, this is a slightly more involved test as you can see then the second derivative test for one variable functions because here you have to write down the matrix, compute its determinant and also check for an entry. So, let us try to understand how to keep these in mind. Suppose, I forget the Hessian test, how do I remember what happens?

(Refer Slide Time: 4:27)



Well, let us look at these functions, so these functions we already know what the situation is, so you have x^2+y^2 we know how this function looks, it is like a bowl and (0,0) is a global minimum, this is something you just know from the fact that squares are positive, so (0, 0) is a local minimum, is a global minimum in fact and so the test hopefully should tell us that let us see if that happens.

Similarly, $-x^2-y^2$ is the same thing inverted down. So, now (0, 0) is a local maximum in fact a global maximum but in particular global, local maximum. Similarly, x^2-y^2 (0, 0) was a saddle point, this was the prototype of how a saddle point look like.

And then x^4+y^4 here the, again (0, 0) is a local in fact a global minimum, but since it is the fourth power something strange happens, so we. So, these are the four prototype examples to keep in mind and these, so each of these will correspond to one of the four conditions that we had before.

So, first let us compute the gradient. So, ∇f here is (2x, 2y), and from here we can compute the Hessian, so the Hessian is $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ very nice Hessian. So, from here we get that the only critical point is, so if I set my gradient to 0, then the only critical point is (0, 0) and the Hessian is actually independent of the point, so the Hf(0, 0) is in particular $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ And now let us apply the test.

So, let us compute determinant of this matrix. So, the determinant of this matrix is 4, so it is positive, so good, so we know already that we are not in the inconclusive case and we know that we do not have a saddle point because for saddle point the determinant will be negative. So, now it will depend on the value of the (1, 1) term. So, here the (1, 1) term which is f_{xx} is 2, this is positive so therefore, it is a meaning (0, 0) is a local minimum and this is exactly what we knew already by inspection because x^2+y^2 is always a positive function with its only 0 at (0, 0).

We can do the same thing for $-x^2-y^2$ so here the gradient is (- 2x, - 2y). If I set it to 0 that gives me that the only critical point is (0, 0) and the Hessian here that means you take the partial derivatives of you take - 2x and take its partial derivatives so that is - 2 and 0 and then you take -2y and take its partial derivatives which is 0 and - 2.

So, notice what I did here by definition actually I should have taken partial of this gradient with respect to x so that is partial of - 2x and then partial of - 2y, but I just took partials of - 2x in the first row and partials of - 2y in the second and the reason that worked is because I know it is a symmetric matrix from Clairaut's theorem.

So, again this Hessian is independent of whatever choice, meaning of whatever point, so in particular at (0, 0) you get the same Hessian and now let us see what happens so again the determinant is non-zero, it is 4, so it is greater than 0. So, what that tells us is one it is non-zero that means it is not in the conclusive situation and two that it is positive which means it is either a local maximum or a local minimum and that will be determined by whatever happens to f_{xx} and f_{xx} means the $(1,1)^{th}$ entry. So, at (0, 0) this is - 2 which is negative so therefore (0, 0) is a local maximum and this indeed conforms to what we already know that this is a global maximum in fact.

Let us do x^2-y^2 again if you compute the gradient, this is (2x, -2y), if you set it to 0 you get that the critical point x is 0 and y is 0, so that is (0, 0). Let us find the Hessian, so for the Hessian you differentiate 2x with respect to x and with respect to y, so with respect to x it is 2, with respect to y it is 0, you take - 2y differentiate with respect to x and y with respect to x it is 0, with respect to y it is - 2, and now you can see it is a bit different than earlier.

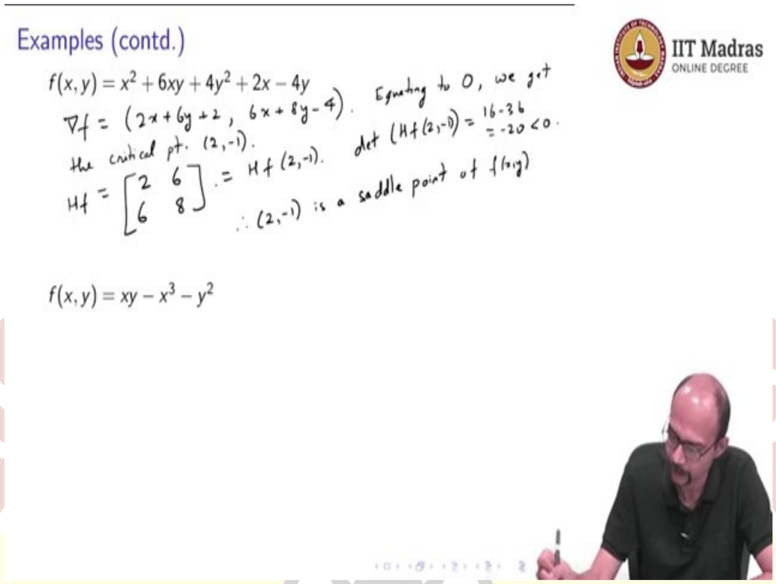
Of course, again it is independent of the point so at (0, 0) also you have the same Hessian. And now what is the determinant, so this determinant is $2\times-2-0$ so that is - 4 which is negative and you stop right here because once the determinant of the Hessian is negative in the 2 by 2 situation, then we know that this is a saddle point and indeed this is something that we have seen already, so this is a saddle point.

And finally let us do the example of x^4+y^4 find the gradient $(4x^3, 4y^3)$ set it to 0 that means $x^3=0, y^3=0$ that means x is 0, y is 0 so the only critical point is (0, 0). Find the Hessian. So, now the Hessian actually involves functions, so it is $\begin{bmatrix} 12x^2 & 0 \\ 0 & 12y^2 \end{bmatrix}$ but what happens at (0, 0)? So, at (0, 0) the Hessian matrix actually is $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ and in particular that means that the determinant of the Hessian is 0 and therefore, this we cannot conclude anything from the test.

Therefore, the test is inconclusive and this is a warning that even for simple functions like x^4+y^4 where we actually know that (0, 0) is a global minimum not just a local minimum it is actually a global minimum because it is a bowl which is x^2+y^2 is more like this, x^4+y^4 is more like this, so it goes up faster, (0, 0) is a global minimum. But the test is inconclusive.

So, these are the 4 prototypes of corresponding to each of the cases that you can remember and if you remember these you will know instead of by hearting the cases from here you can recall which case tells you what.

(Refer Slide Time: 12:03)

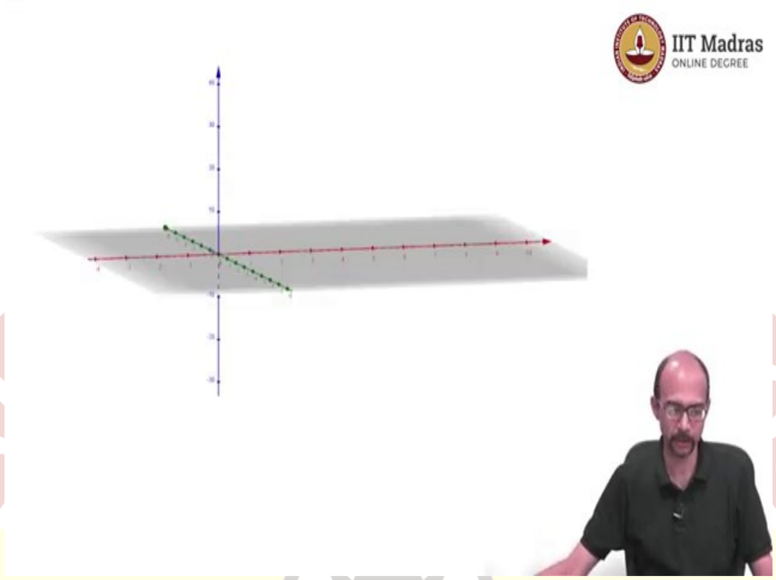


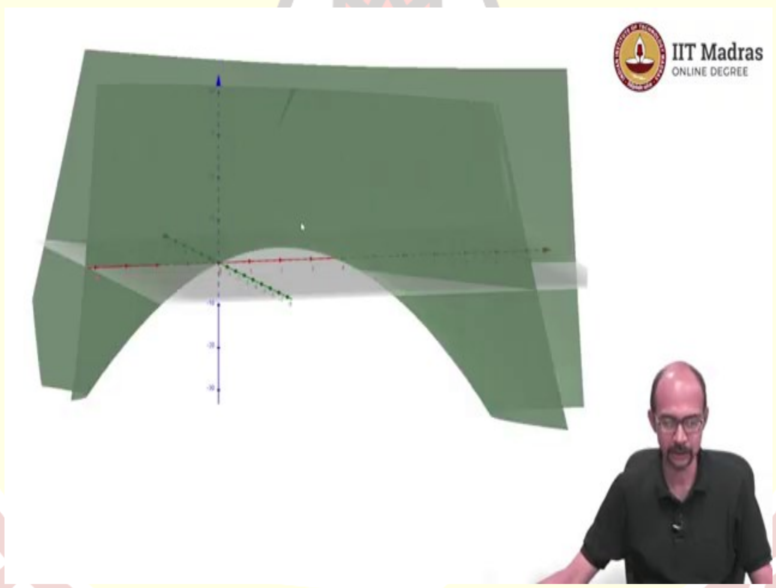
Let us do a couple of other examples. So, this is $x^2+6xy+4y^2+2x-4y$ So, let us find the gradient here, these are slightly more involved. So, (2x + 6y + 2, 6x + 8y - 4). So, if you set the gradient to 0, so equating to 0, we get the critical point, we have actually done this before so I will not spend time here on this, so we get the critical point (2, -1), so this is something that we did in the video on critical points. So, I will not repeat this computation, you get a system of 2 linear equations in x and y and which you can solve, so we used Gaussian elimination to solve that, fine.

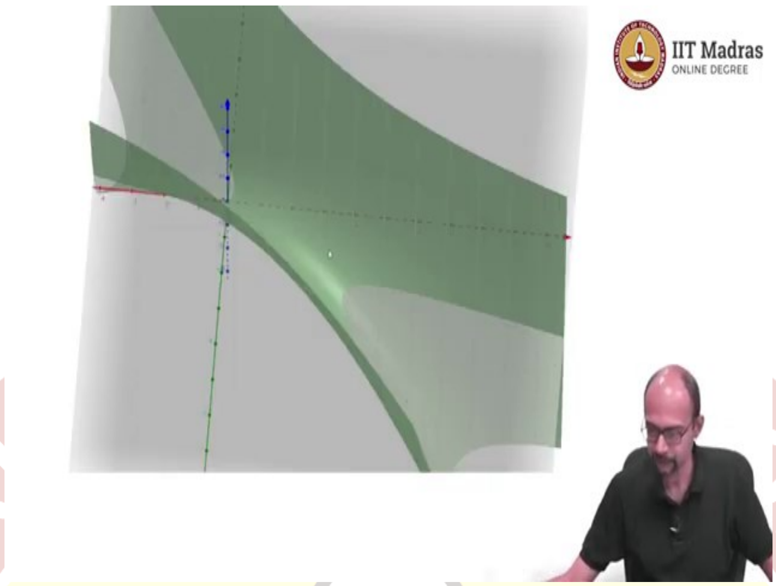
So, now let us ask what is the nature of this critical point, is it a saddle point, is it a local minimum, is it a local maximum, can we say that from the test? So, for that we have to find the Hessian matrix first. So, what is the Hessian matrix? So, you have 2x + 6y + 2 let us take the partial with respect to x and y, so with respect to x you get 2, with respect to y you get 6, so $6x+8y-4$ the partial with respect to x is 6, with respect to y is 8.

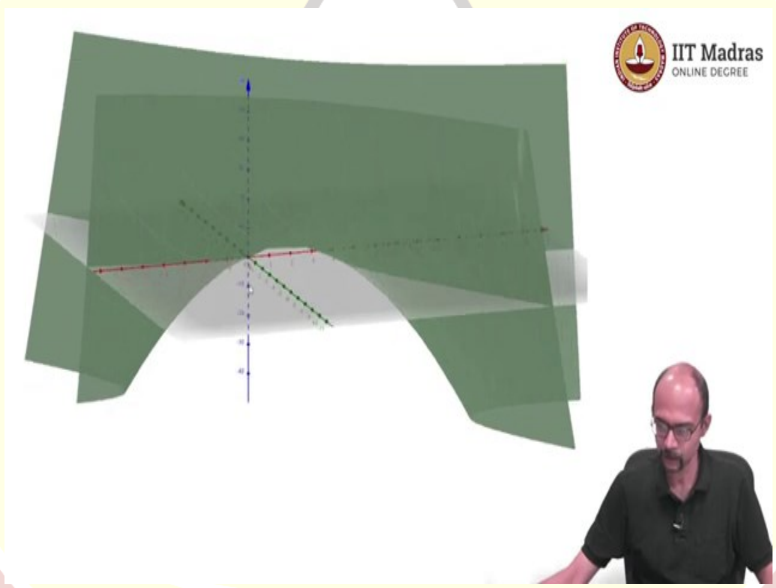
And again this hessian does not depend on any point, so at (2, -1) also it is the same matrix. Let us compute its determinant, so $det(Hf(2,-1))=8\times2-36=-20$ which is negative. So, this tells us that (2, 1) is a saddle point of f (x, y).

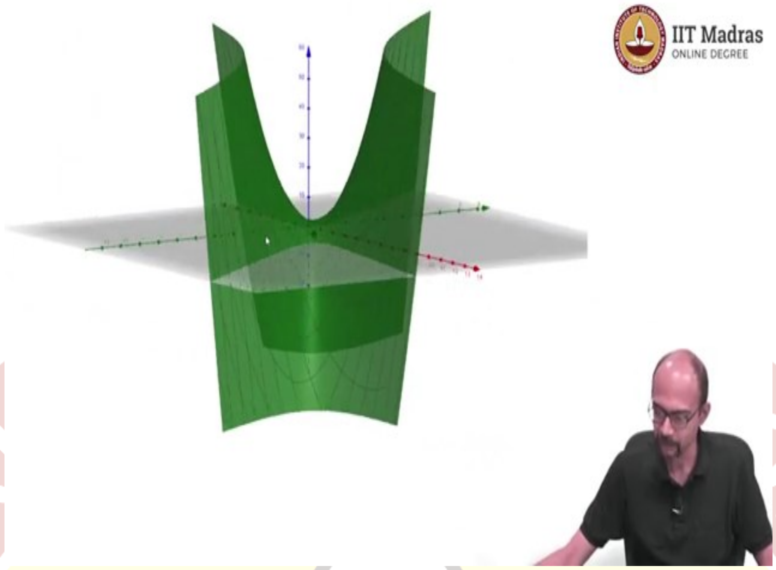
(Refer Slide Time: 14:22)






IIT Madras
ONLINE DEGREE

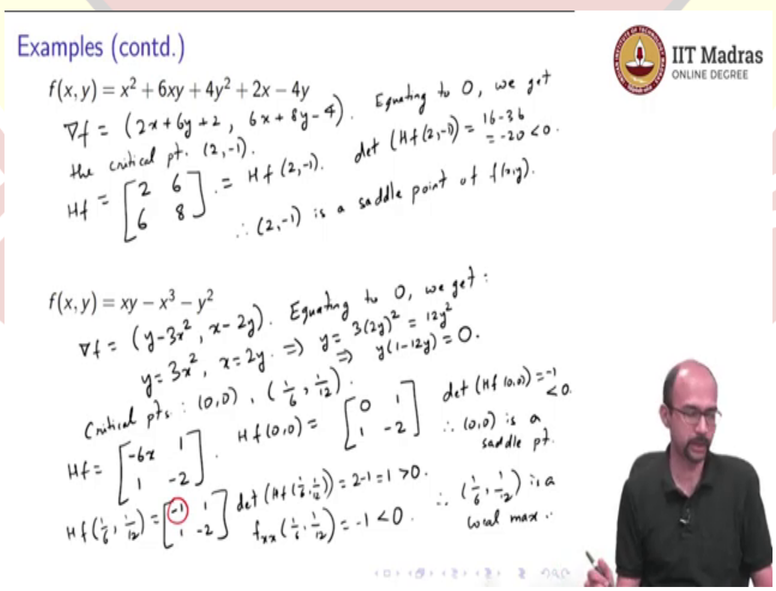

IIT Madras
ONLINE DEGREE



So, let us view this function in Geogebra and check exactly what is happening at (2, -1). So, here is the function $f(x, y) = x^2 + 6xy + 4y^2 + 2x - 4y$ So, you can see that there is indeed something interesting going on here, it looks like a, function looks quite nice actually.

And if we view it like this let us see what is happening at the point (2, -1). So, here is the point (2, -1) and you can see right above the point (2, -1), there is that saddle, so that saddle is coming at the point (2, -1) and indeed we are being able to detect that using the Hessian matrix.

(Refer Slide Time: 15:15)



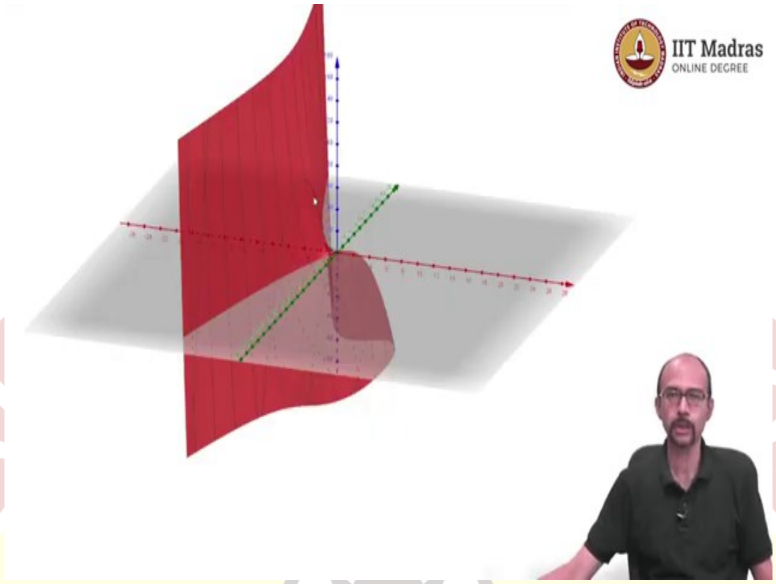
Let us do this second example which is f (x, y) is $xy-x^3-y^2$ So, let us find the gradient so the gradient here is $(y-3x^2, x-2y)$ So, if we equate this to 0. So, now we have 2 equations $x=2y$ and $y=3x^2$ So, let us substitute x $x=2y$ in the first equation, so if we do that. So, one point we already know at (0, 0) so the list of critical points one point we already know, one is (0, 0) let us see if there is any other point.

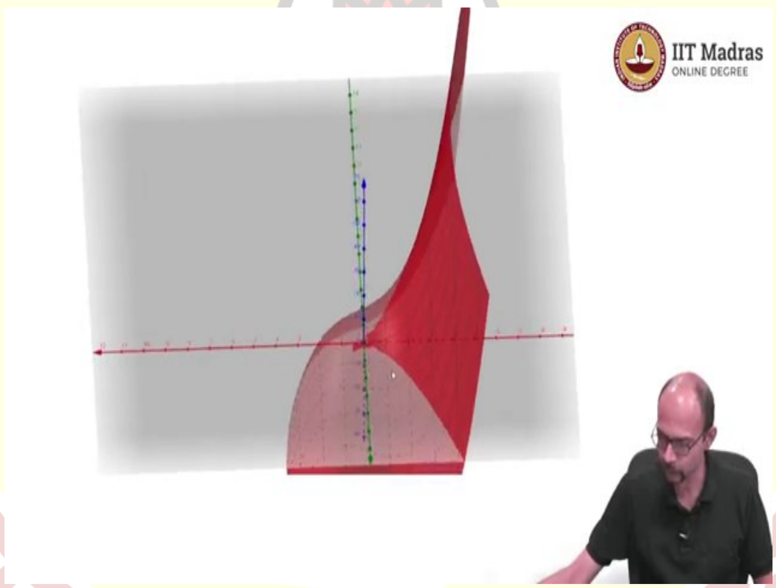
So, for that we substitute x = 2y in the first equation, so we get $y=3(2y^2)=12y^2$ which means $y(1-12y)=0$ So, one option is y is 0 that is exactly the point (0, 0) that corresponds to the point (0, 0). The other option is that y is not 0 in which case $1-12y$ must be 0, so which means y must be $\frac{1}{12}$ And if y is $\frac{1}{12}$ then we can get x is $\frac{2}{12}$ so which is $\frac{1}{6}$ so this $(\frac{1}{6}, \frac{1}{12})$

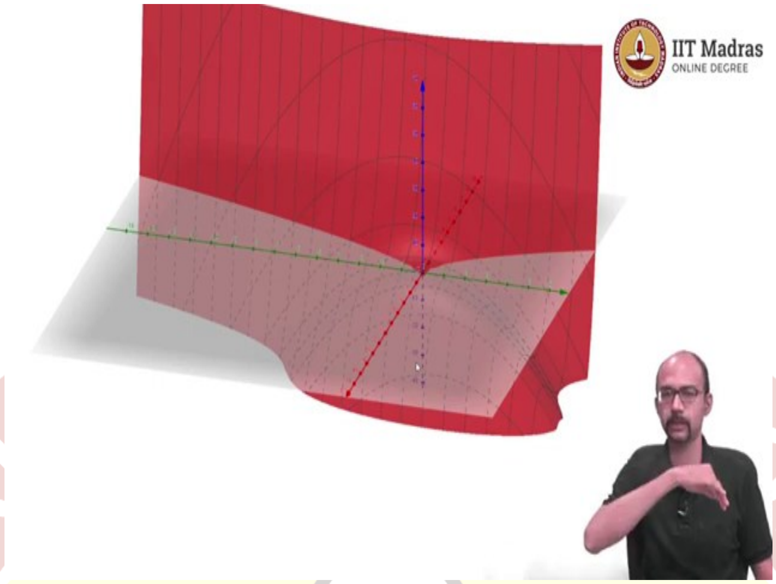
So, there are two critical points (0, 0) and $(\frac{1}{6}, \frac{1}{12})$ Let us check the nature of these critical points, so for that we have to find the Hessian matrix. So, I will take partials again, so $y-3x^2$ the partial with respect to x is $-6x$ with respect to y is 1, $x-2y$ partial with respect to x is 1 and partial with respect to y is -2. So, at the point (0, 0) this becomes $\begin{bmatrix} 0 & 1 \\ 1 & -2 \end{bmatrix}$ so the determinant is negative, so this is -1 which is less than 0 and that tells us it is a saddle point. So, therefore (0, 0) is a saddle point.

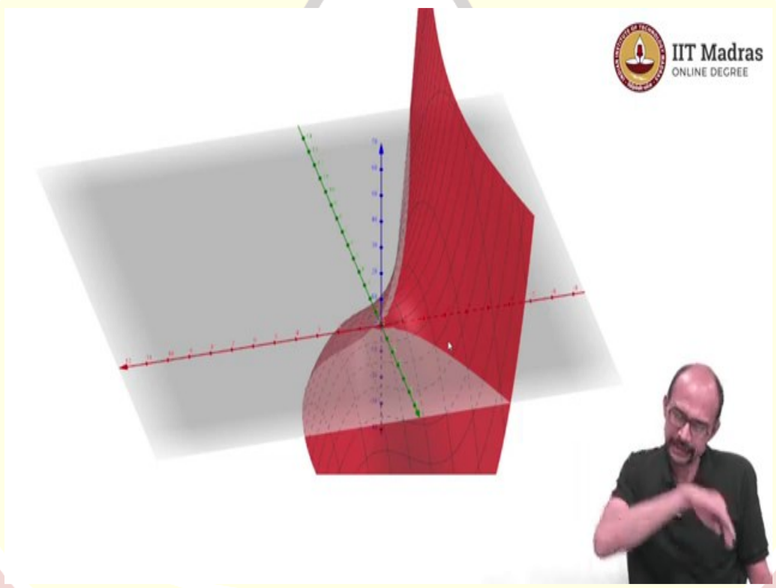
The other critical point is $(\frac{1}{6}, \frac{1}{12})$ let us find out what the Hessian is. So, if I substitute 1 /6 in the $(1,1)^{th}$ place, I get $Hf(\frac{1}{6}, \frac{1}{12})=\begin{bmatrix} -1 & 1 \\ 1 & -2 \end{bmatrix}$ so the determinant of the Hessian matrix is 2-1 which is 1, so this is positive, so we know this is either a local minimum or a local maximum. And for that it is determined by the $(1,1)^{th}$ term, so $f_{xx}(\frac{1}{6}, \frac{1}{12})$ so that is a $(1,1)^{th}$ term which is this term here and that is -1 which is less than 0, so therefore $(\frac{1}{6}, \frac{1}{12})$ is a local maximum. Let us see in Geogebra if this is indeed what is happening.

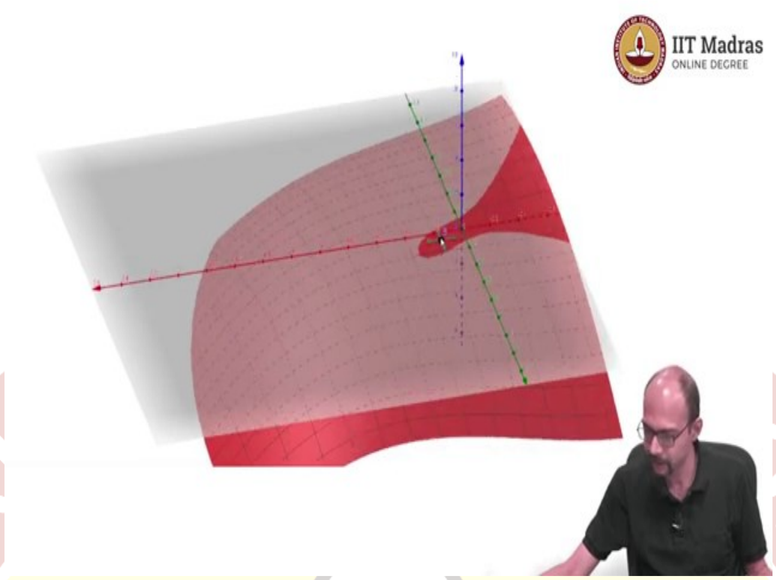
(Refer Slide Time: 19:33)

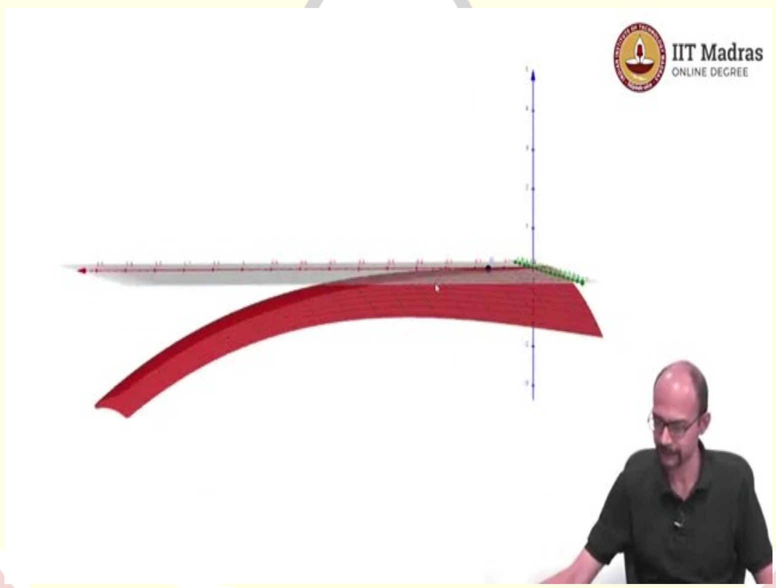





IIT Madras
ONLINE DEGREE


IIT Madras
ONLINE DEGREE



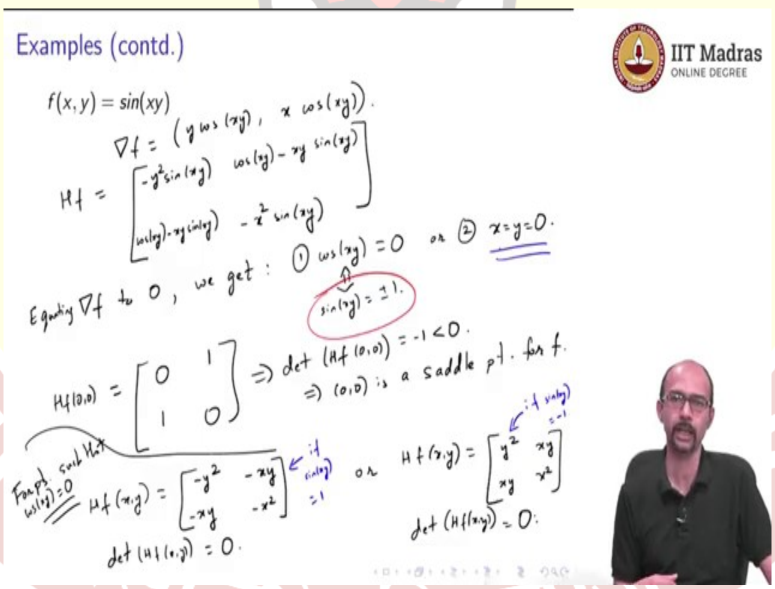


So, this is how the graph of the function looks like in Geogebra and you can see at (0, 0) it really has very interesting behavior and that it is clearly a saddle point. So, if we zoom in, let us see if I can come close to (0, 0). So at (0, 0) you can see at if you take the, take it along the green line, then the function values it is like this, so it is a local maximum, whereas if you take it along the other axis it is decreasing like this, increasing a little and then it again goes down, so it is a local minimum along that, so indeed it is like a saddle point, although it looks a bit different than the standard saddle which is a little more clear to see. But it is clear that at (0, 0) something strange happens.

The more interesting point is $(\frac{1}{6},\frac{1}{12})$ that is a bit hard to pick up from the graph. So, here is that point, it is a bit difficult to see but let us zoom in further and maybe we can see what is there, so now we can see this point here, here is the point b which is $(\frac{1}{6},\frac{1}{12})$ and you can see that close to this point the function is indeed sort of positive, so it is coming down, remember at before (0, 0) from the on the other side it was coming down but then just for a brief period it kind of goes up again and then it goes down and that is what is being picked up by our Hessian matrix over there.

So, this indeed conforms to our understanding of what is happening. So, maybe here one can see, yeah. So, you can see from the graph it may not be very easy to pick up these points but the algebra tells you what is happening and even from the function directly the equation of the function it is not clear why this is a local maximum but the second order test is going to, the Hessian test is going to tell you that indeed it is a local maximum.

(Refer Slide Time: 21:50)



Let us do this last example of $f(x,y)=\sin(xy)$ So, we have computed the gradient for $\sin(xy)$ which now let us write that down again, this was $(y\cos(xy), x\cos(xy))$ In fact we also wrote down the Hessian, the Hessian was $-y^2\sin(xy)$ then the other diagonal term is $-x^2\sin(xy)$ and then the off diagonal terms were slightly more complicated.

So, $\cos(xy)-xy\sin(xy)$ and then the same thing over here by Clairaut' theorem or by computation. So, this is how our Hessian matrix looks like, so first what are the critical points. So, we actually know for this function what is happening, we have seen a graph of this function and several times and we know what is happening for this function because the local maxima or minima are actually global maxima or minima and they are given by when the function this is a sin function, so it takes the value 1 and -1 respectively.

So, when does it take the value 1 when xy is of the form $\frac{\pi}{2}+k\times 2\pi$ where k is an integer. So, for all such points the function takes the value 1, so the value 1 will be taken whenever xy attains one of those values so it is curves and we have actually seen this picture, we will see it again. And similarly, when does it take the value -1 when it is of the form $3\frac{\pi}{2}+k\times 2\pi$ where k is an integer again and again that is a curve and on those points it takes value -1.

And the only other point which ends up being a critical point is the following. So, let us equate, so equating gradient to 0, we get, so two things can happen, one is $\cos(xy)=0$ the other is if $\cos(xy)\neq 0$ then both x and y should be 0. So, $\cos(xy)=0$ is 0 is exactly the set of points that I just said, so $\cos(xy)=0$ means that $\sin(xy)\pm 1$ because we know that $\sin^2(x)+\cos^2(x)=1$

So, this is the same as saying that sin of $\sin(xy)\pm 1$ so +1 is exactly the set of local, global in fact maxima and -1 is exactly the set of global minimum. So, we know actually what happens at all these points, so these are points that we are already understand. What happens at x = y =0, so this is the point that we do not know what happens. Well, it is fairly clear that this is probably not a point of local maximum or local minimum, so most likely this is a saddle point and let us see if our Hessian is telling us that.

So, if you, if we compute our Hessian what do we get? So, Hessian at (0, 0), so I should say Hessian at (0, 0), well, substitute x = y = 0, so the diagonal entries become 0 what happens to the off diagonal entries? Well, $xy\sin(xy)$ becomes 0, so the only contributing term is cos(xy) that is cos(0) when x = y = 0 and cos(0) we know is 1, so this is the matrix we get, this implies determinant of this matrix, so sometimes this is called the Hessian determinant is -1 and that tells us that (0, 0) is a saddle point for this function.

So, the point that was sort of we were unsure of what happened, the test told us it is a saddle point. Let us see what happens for those points for which we are sure of what happens, let us see what happens to points where of this form where $\cos(xy)=0$ or $\sin(xy)=\pm 1$ So, for points such that $\cos(xy)=0$ let us compute the Hessian.

So, as I said there are two options either $\sin(xy)=1$ or $\sin(xy)=-1$ let us say sin(xy) is 1. So, there are two options so I will say this or this and we will see what happens in both cases. So, if sin(xy) is 1 then this gives me $Hf(x,y)=\begin{bmatrix} -y^2 & -xy \\ -xy & -x^2 \end{bmatrix}$ And if sin(xy) is -1, then this gives me $Hf(x,y)=\begin{bmatrix} y^2 & xy \\ xy & x^2 \end{bmatrix}$ So, I will just reiterate that this is if sin(xy) is 1 and this is if sin(xy) is -1. And what happens to the determinant?

So, suppose we compute the determinant so the Hessian determinant. Well, you can see what is happening, this is $x^2y^2-x^2y^2$ which is actually 0. So, this test is inconclusive even though we actually know what is happening at these points and the same thing happens here, this is also 0, $x^2y^2-x^2y^2$

So, the point of this example is to say that the Hessian test has several features to it and what can happen is that points where it is obvious that there are local maxima or local minima it may end up being inconclusive, whereas points that we actually do not know what is happening from the function it gives us a result, so such things can happen.

So, the Hessian test it is not a full proof test but it is our first step towards understanding critical points. So, we have studied the Hessian test in this video for f (x, y) we will study this in the forthcoming video for more number of variables. Thank you.