INDIAN INSTITUTE OF TECHNOLOGY MADRAS
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

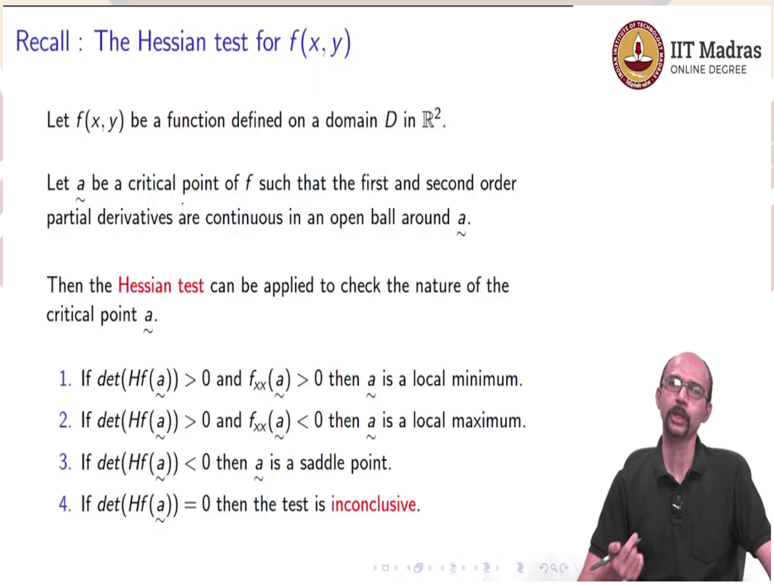
Mathematics for Data Science -2
Professor Sarang Sane
Department of Mathematics
Indian Institute of Technology, Madras
The Hessian matrix and local extrema for f(x,y,z)

Hello, and welcome to the Maths 2 component of the online BSc program on data science and programming. This video is about the Hessian matrix and local extrema for f(x,y,z). In the previous video we have seen the Hessian test to classify critical points for the function f(x,y). In this video, we will see the same or an analogous test for f(x, y, z).

So, as you can see that as your number of variables increases, the test sort of becomes a bit harder, which is why we have these separate videos. Now, one can make sense of all this in terms of linear algebra for the Hessian matrix for n variables. And in particular, if we allow ourselves the notions of eigenvalues, which you will study later in your machine learning course, and also use these ideas of diagonalization and orthogonal, using orthogonal matrices. So, we did this when we did similarity and equivalence of matrices.

Using all those ideas, we can formulate a general test for n variables. But in the interest of time, and functionality, as far as this course is concerned, I am not doing this for general variables, although we have developed some of the linear algebra already required for this, but we need a little bit more. So, I am restricting myself to 3 variables. And this will be you enough probably for most of the problems that you will come across.

(Refer Slide Time: 1:51)



So, let us recall what is the Hessian test for a function of 2 variables. So, if you have a function defined on a domain D in R^2, and ã is a critical point such that the first and second order partial derivatives are continuous in an open ball around ã, then there is something called the Hessian test, which means that you compute the Hessian matrix of the function f and evaluated at the point ã, that gives you a 2 x 2 matrix.

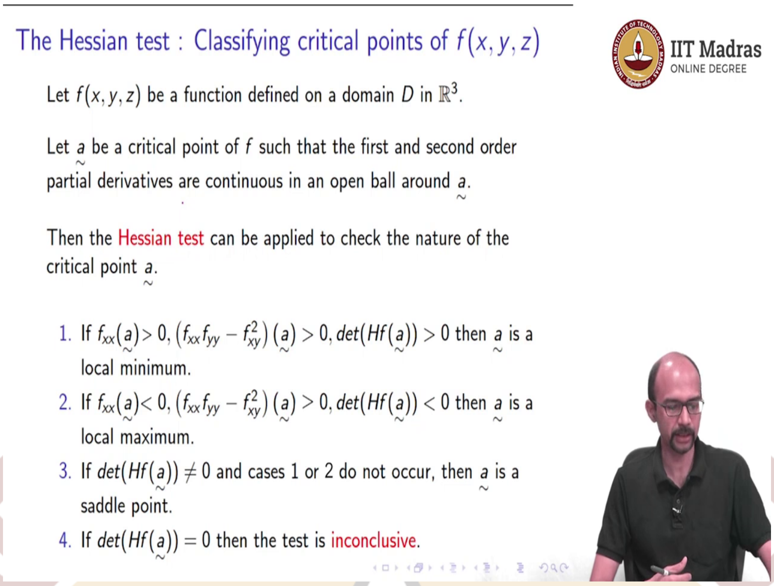
You compute its determinant and the value of the 1 1 term. If that determinant is positive, and the 1 1 term is also positive, then the point is a local minimum. If the determinant is positive, and the 1 1^{th} term, meaning the $f_{xx}(\tilde{a})$ is negative, then it is a local maximum. If the determinant is negative, it is a saddle point. And if the determinant is 0, which is called the degenerate case.

So, the previous 3 are called the non-degenerate cases, which means that the determinant is non-zero. But in the degenerate case, which is to say the Hessian determinant is 0, this test is inconclusive. So, that is really the bad case. That is the case we want to avoid, because that means that the test does not tell us anything. So, we saw some examples of this test, we also saw that it told us the answer in some surprising cases.

Whereas, in some cases where it was obvious what the answer should be, it turned out to be inconclusive. So, the test has a, it comes with, it is a double-edged test, sometimes it will be it will work in situations where you really do not know what is happening. And sometimes it will not work, even though you absolutely know what happened. So, we will see something similar for the Hessian test in 3 variables.

I am not going to do as many examples here. Because we already know the phenomenon that are taking place, namely that it can work sometimes in spite of the fact that we have no clue of what is happening. And it may not work at all, despite us knowing what is happening.

(Refer Slide Time: 4:01)



So, let us see what the test is. So, again, this is a function f is a function defined on a domain D in 3 variables. Suppose ã is a critical point and the first and second partial derivatives are continuous in an open ball around ã. I will remind you that now there are 9 second order partial derivatives f_{xx}, f_{xy}, f_{xz}, f_{yx}, f_{yy}, f_{yz} and then f_{zx}, f_{zy}, f_{zz}. Of course, thanks to Clairaut's theorem, we know that we can interchange the order.

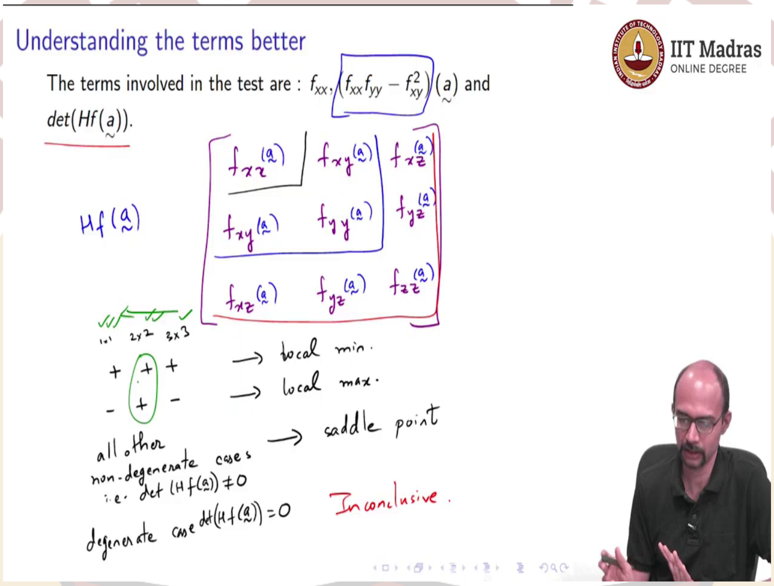
So, f_{xy} and f_{yx} are the same, f_{xz} and f_{zx} are the same, and f_{yz} and f_{zy} are the same. So, effectively, there are 6 partial derivatives. So, this is a fairly long computation if we want to compute this, and then the Hessian matrix is where we put these second order partial derivatives into a 3 x 3 matrix, remember, it is a 3 x 3 matrix now, because you have 3 variables and then the test tells us that if the following things happen, then we can classify the critical points.

So, if the term f_{xx} is positive, the term $f_{xx}f_{yy} - f_{xy}^2$, which was the determinant that we used in the 2 x 2 case if that at ã is positive, and if the determinant of the entire Hessian matrix is positive, then ã is a local minimum. If on the other hand, the f_{xx} is negative, the second term $(f_{xx}f_{yy} - f_{xy}^2)$(ã) is positive and the determinant is negative. So, this is an alternating sign, then ã is a local maximum.

If the determinant is non-zero, and neither of cases 1 or 2 occur, then it is a saddle point. So, that means in these two, the first two cases we have local minimum or local maximum, and in all the other non-degenerate cases, meaning where the determinant is non-zero, it is a saddle point. And if the determinant is 0, which means it is a degenerate situation, then the test is inconclusive. So, this is the no Hessian test in the three variables situation.

This is getting slightly complicated as you can see, and which is why we, I am not describing this in the general n x n setup. It actually has an easier description, but for that, we need to know what are Eigen values.

(Refer Slide Time: 7:01)



So, let us understand these terms better. So, the terms involved in the test are f_{xx},

$(f_{xx}f_{yy}-f_{xy}^2)$(ã) and det(Hf(ã)). So, what are these terms? Let us write down the matrix. So, here is the matrix f_{xx}, f_{xy}, f_{xz}, so I am not really caring now about the order whether it is f_{xy} or f_{yx} because of Clairaut's theorem, f_{yy}, f_{yz} and then f_{xz}, f_{yz} and then f_{zz} and let me draw this like this. So, this is our hessian matrix. So, the Hessian matrix and maybe let us do this at ã. So in that case, I would evaluate all these at ã.

So, once I evaluate, these are numbers, these are all numbers. And what do these terms correspond. So, f_{xx}(ã) corresponds to this term there, so it is the one-oneth term or the 1 x 1 matrix in the top left corner. This second term. So, the second term here, this term so,this term corresponds to the determinant of the 2 x 2 matrix in the top left corner. And the third term which is the determinant of Hf(ã) corresponds to the determinant of the entire matrix.

So, what we are doing is, we are taking the determinants of the top left corner successively larger. So, the 1 x 1, 2 x 2, 3 x 3 and in fact, we can generalize this to n x n, only thing is those signs, we have to be careful of how they work. And what we are saying is that if that top left corner, if the sign is like this, if it is + ,+ and +, so which means that this is 1 x 1, 2 x

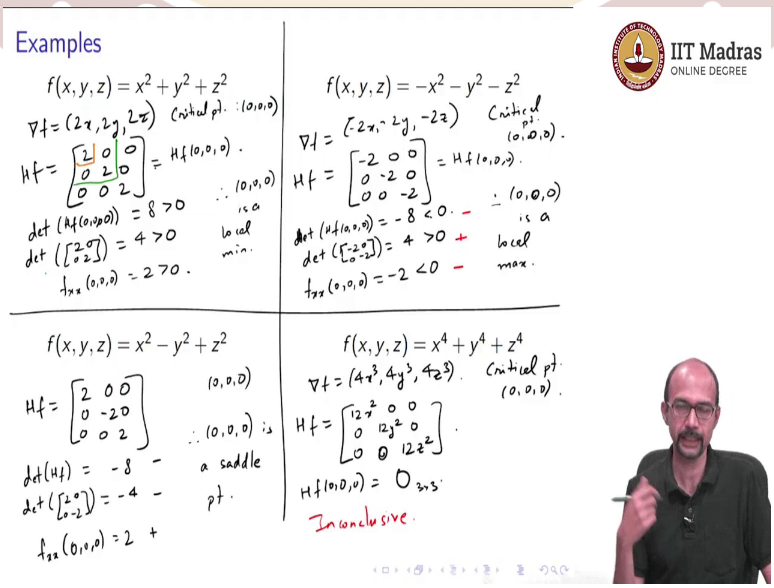
2, 3 x 3. If it is + , + and +, that means they are greater than 0, greater than 0 and greater than 0, then this is a local minimum.

If it is a -, + and - . Note, the alternating thing, this is very, very important. This is where students often make mistakes. If it is -, + and -, it is a local maximum and in all the other cases, where this is not 0, where the determinant is not 0, all other non-degenerate cases by which we mean that the determinant is not zero, this is a saddle point. And if it is the degenerate case, which means the determinant is 0. This, we cannot really say this is the really bad, bad part. So, this is what it seems.

So, what you have to keep track of is you have to compute 3 successive determinants, the first one is a 1 x 1 determinant. So, that really there is nothing to say. The second one is a 2 x 2 determinant. And the third one is a 3 x 3 determinant. So, in practice, you do it the other way. Because you want to check that the determinant of the entire matrix first is non-zero, if the determinant of the entire matrix is 0, then you stop the test right there.

So, what you do is you your testing should start in the opposite direction. So, you do your testing in this direction. First, you test this, then you test this, and then you test this right. So, first, you test the entire matrix, what is the determinant, whether it is positive or negative, or 0. If it is 0, end of story. If it is positive or negative, then you test for the 2 x 2, if the 2 x 2 by chance happens to be negative, stop right there. Because as you can see, here, you want it to be positive, so then it is a saddle point. If it is positive, then you check for the 1 x 1 that is the way to go.

(Refer Slide Time: 12.00)



So, let us do a couple of examples. So, these are again, sort of our test examples to remember how the rules work. So, the functions are also very familiar here. So, let us first do

$f(x,y,z) = x^2 + y^2 + z^2$. Clearly this is a non-negative function, and the only place it 0 is the point (0,0,0). So, this is, so (0,0,0) is a local minimum. And in particular, in fact, it is a global minimum, and in particular, it is a local minimum.

So hopefully, the Hessian test will tell us that so $\nabla f = (2x, 2y, 2z)$. Hessian is, well, you can compute this, now I am doing this rather fast. $Hf = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$. Let us compute the various things that we have to compute. So the critical point is (0,0,0). And so, the Hessian at (0,0,0) is the same thing. And so, we should compute the determinant of all the top left matrices of decreasing sides.

So, $\det(Hf(0,0,0)) = 8 > 0$. Actually, you can see all of them are going to be positive.

$\det(\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}) = 4 > 0$. And then determinant of the first, meaning the entry in the 1-1, one-oneth place, that is this entry here, so that entry is positive. So, $f_{xx}(0,0,0) = 2 > 0$. So therefore, this (0,0,0) is a local minimum. And that is exactly what we had in our case 1. So, this is how you can remember case 1.

Similarly, if you take $f(x,y,z) = -x^2-y^2-z^2$ by the same argument, it is always a non-positive function, the only place it 0 is the point (0,0,0). It is a global maximum. And so hopefully, the Hessian test picks that out. So, I am going to sort of do this really fast, $\nabla f = (-2x, -2y, -2z)$, the Hessian matrix now you can write down is

$Hf = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}$, critical point is clearly (0,0,0). So, this is the Hessian also at that critical point.

And now let us see what the signs are. So, the determinant of the entire matrix is-8, it is a diagonal matrix, this is negative. So, we have a - sign, the determinant of the 2 x 2, it is 4, it is positive. And then the one-oneth entry is -2, which is negative. So, the signs here are, this is the important part, what are the signs, the signs are -, +, -, and that tells us this is a local maximum. And indeed, this is what we already know and so this is how you can keep this in mind.

Similarly, if you take $f(x,y,z) = x^2 - y^2 + z^2$, you will get it as a saddle point. So, I will just write down the Hessian and you can do the checking. So, the Hessian is going to be

$$Hf = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

. So, the signs you are going to get are determinant of the entire so the critical point is (0,0,0) and the signs you are going to get are so -8, 2 x 2 is = -4, and already you can stop right here, because Hessian is non, the Hessian determinant is non-zero and the signs are not corresponding to either alternating or + ,+, +.

So, you can stop right here, but just to finish it at $f_{xx}(0,0,0) = 2 > 0$. So, this is a saddle point. Indeed, we already know this, because of, well, you can check that along certain axis it is a local maximum, along certain axes it is a local minimum. And finally, if you take $f(x,y,z) = x^4+y^4+z^4$, what happens $\nabla f = (4x^3, 4y^3, 4z^3)$, critical point is clearly (0,0,0).

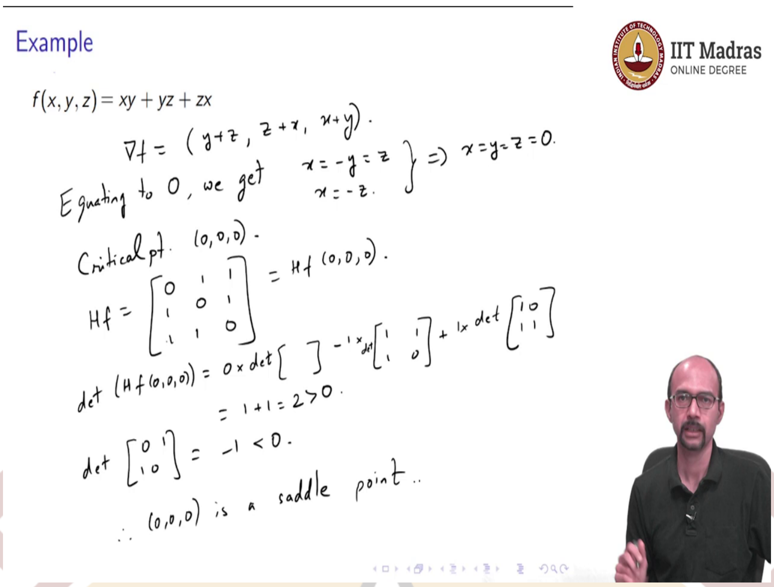
By equating this to 0, and the Hessian you can compute from here is

$$Hf = \begin{bmatrix} 12x^2 & 0 & 0 \\ 0 & 12y^2 & 0 \\ 0 & 0 & 12z^2 \end{bmatrix}$$

. And what happens at the point, so at the point (0,0,0), this is the 0 matrix. And as a result, it is the degenerate case, we cannot say anything. So, this is inconclusive, as far as the Hessian test is concerned.

On the other hand, we already know that $x^4+y^4+z^4$ is always non-negative, it is a fourth power, they will always be non-negative. And the only point where it takes the value 0 is the point (0,0,0). So, this is in fact a global minimum. So, in particular, a local minimum. So, in spite of the fact that we easily know that this is a global minimum, the test is inconclusive. So, these are the 4 prototype examples you should keep in mind to, in case you forget the tests. So, this will tell you.

(Refer Slide Time: 18:25)



Let us do a couple of examples of, of computing this. So, here f(x, y, z)= xy + yz + zx. So, gradient f, we have actually computed the gradient and the Hessian, so $\nabla f = (y+z,\ x+z,\ x+y)$. So, you equate this to 0. So, then we get these 3 equations, y+z = 0, z + x = 0, x + y = 0. So, that tells us that x = -y, and also tells us that x = z, that is what we get from equation 1 and 3, but equation 2 tells us that x = -z.

So, the only way this can happen is if x = y = z = 0. So, the only solutions are x = y = z = 0. So, the critical point is (0,0,0). And what happens to the Hessian matrix? So, what is the Hessian, so if we write down the Hessian, which we have actually done before, we get

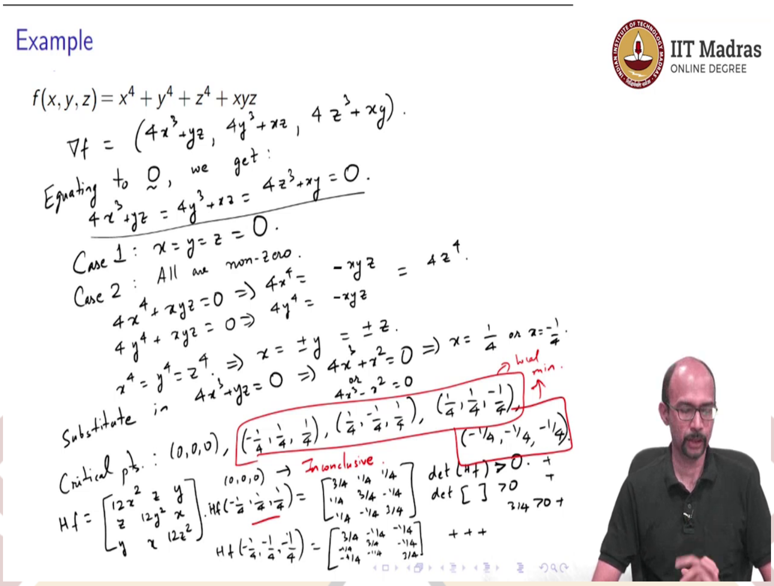
$Hf = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$. So, this is in particular, the Hessian at (0,0,0). And what happens? So, let us do our test.

So first, what is the determinant of the hessian matrix? So, this is, so if we do our determinants, det(Hf) = 0 (0 -1) -1(0-1) + 1 (1-0) = 1 + 1 = 2 > 0 So, at least it is non-zero. So, we are in business, this test is going to be conclusive.

So, now we take the 2 x 2 left, top left matrix calculated det($\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$) = 0 - 1 = -1 < 0. And, well as soon as we know it is negative, this does not correspond to case 1 or case 2. So, this means it is a saddle point. So therefore, (0,0,0) is a saddle point. And I will encourage you to check why it is a saddle point. So, find a sequence of points which go to (0,0,0) for which

(0,0,0) on that curve is a minimum value, and find some other curve on which it is a maximum (0,0,0) is a point of maximum. So, I hope it is clear how we are using this test.

(Refer Slide Time: 21:48)



So, let us do this final example, $f(x,y,z) = x^4+y^4+z^4+xyz$. I want to do this example just to show you that sometimes these calculations do become rather difficult, doable, but difficult. So, what is the gradient, so $\nabla f = \left(4x^3 + yz,\ 4y^3 + xz,\ 4z^3 + xy\right)$. Let us set this to 0. So, equating to 0 and we have to really work this out, we get, so we have 3 equations. $4x^3+yz = 4y^3+xz= 4z^3+xy = 0$.

So, now we have to solve these 3 equations. And note that these are not linear equations. So, our linear algebra would not really help us over here. So, we have to solve sort of see this and solve this by brute force. And also using the fact that, they have a very nice form. So here, here is where your comfort with, with equations or algebra in general will kick in.

So, case 1 is, there is one solution staring us in the face, namely, where $x = y = z = 0$. And, it is very important to note that, once one of them is 0, all the others are 0. For example, if $x=0$, then you can put that into the equation $4y^3+xz = 0$ and that will give you that $4y^3 = 0$ and that will say that $y = 0$. And you can put it into the equation $4z^3+ xy = 0$ and that will say that $z=0$. So, by symmetry, if one of them is 0, all of them are 0. So that is case 1.

The other cases where none of them are 0. So, all are non-zero, all are not zero. So, none of them are 0. So, in that case, what we can do is we can multiply each of our equations suitably. So, the first equation we can multiply by x, and we can say that $4x^4 + xyz = 0$, which means $4x^4 = -xyz$ but I can do the same thing for the second equation by multiplying it by y and I can say that $4y^4 + xyz = 0$, which is saying that $4y^4 = -xyz$ and by symmetry we will also get that this is $4z^4 = -xyz$.

So, what that means is$4x^4=4y^4=4z^4$, so $x^4 = y^4 = z^4$ and they are not all 0. Fine and if we take the fourth root, what that will tell us is that $x = \pm y = \pm z$. And now we can, so, now we can substitute this into our equations. So, substitute in $4x^3 + yz = 0$.

So, if you substitute this, what you are going to get is $4x^3+x^2 = 0$. And since x is non-zero, we can cancel x^2 on both sides and what we get is $x = -1/4$ either you get this or $4x^3-x^2=0$. And then what we get is $x = 1/4$ or $x = -1/4$. And then so again, for y and z also, you get these three. Of course, all possible solutions are may not be correct, because they have to satisfy these equations over here.

So, then you have to back check for what values these are satisfied. And if you do that, you will get your critical points as (0,0,0), which was case 1, and then from here, we will get (-¼, ¼, ¼) , (¼,-¼,¼), and (¼,¼,-¼), and finally (-¼,-¼,-¼). So, these are the 5 critical points that you will get. And now we have to check the nature of these critical points.

So, let us see what the Hessian is. So, if you compute the Hessian, what you are going to get is $Hf = \begin{bmatrix} 12x^2 & z & y \\ z & 12y^2 & x \\ y & x & 12z^2 \end{bmatrix}$. Fine, so now we have to evaluate this Hessian at all these various points, and check what happens. So, already, you can see that (0,0,0) this is inconclusive. So, we cannot really say anything about (0,0,0). So, (0,0,0) it is inconclusive.

And again, I will encourage you to check what happens at (0,0,0). My claim is that this should probably be a saddle point. What happens at the other points, and by symmetry, for these three, at least, we should get the same thing and then we have to check(-¼,-¼,-¼). So, let us check(-¼, ¼, ¼). So, in this case, the Hessian is, so the Hessian is, let us write that down, $Hf = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} & \frac{-1}{4} \\ \frac{1}{4} & \frac{-1}{4} & \frac{3}{4} \end{bmatrix}$.

Now, you can use, use our good old determinants, and check what we are getting. So, notice here that there are constants in this, so you can pull out the constant. So, this matrix is 1/4 multiplied by the matrix, which consists of integers, if that makes your computation easier.

And remember that when you take determinant, since it is a 3 x 3 matrix, that constant will be to the power 3.

So, if you do that, the determinant of the entire thing is, I maybe, I not do the actual computation and I will just say that det(Hf) > 0, you can see that is the case because on the diagonal the terms are rather large. So, this is positive. And then for the 2 x 2, again, this is positive. And the 1 x 1 is again positive, positive, so this is a +,+,+ situation that means this is a local minima. So, all these 3 are local minimum. And finally, we have the last point. And, again, actually the same thing happens.

So, in this case, you get $$\mathrm{Hf} = \begin{bmatrix} \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{bmatrix}$$. And if you compute the determinant, again you are going to get +, + , + , which means this is also a local minima. So, once again, what you will find here is that for the points, where it kind of not so clear how the function behaves, the Hessian test is telling us something, whereas for (0,0,0), it is comparatively easier to see what is happening, because the function value there is 0.

And you can try to sort of understand from which side do I come and along with curve do I come, in order to get local maxima, local minima, etcetera. So, that behaves like a certain point. So, the conclusion here is that is that again, the Hessian test is kind of the behavior is kind of surprising, sometimes for things that you may understand relatively better, it gives you inconclusive results, but for things that you may not understand relating, you understand relatively less clearly, it gives you results.

Now, of course, we have not gone into proofs of the Hessian test. And maybe that is, that is a separate topic in itself. So, we will not go there. There is a, we can extend this, as I said, for the n x n case, but we will not do that in this course. So, the main theme of this video was that we now have a Hessian test analogue to the second derivative test for 3 variables. Thank you.