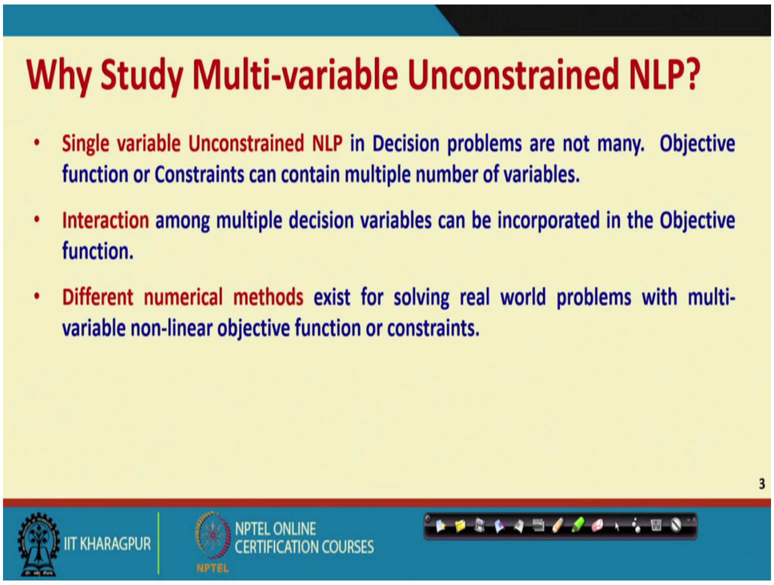
Selected Topics in Decision Modeling
Prof. Biswajit Mahanty
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

Lecture - 23
Multi-variable Unconstrained NLP

So, in our course Selected Topics in Decision Modeling we are in our 23rd lecture. And this will be on Multi-variable Unconstrained Non Linear Programming.

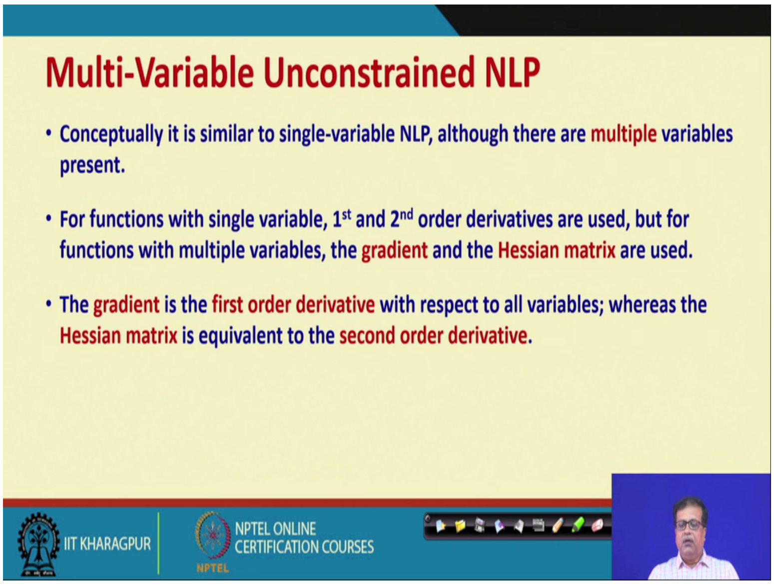
(Refer Slide Time: 00:34)



Now, last lecture if you recall we had studied the single variable unconstrained non-linear programming. But you know they are most of the situations they are not going to be single variable, there could be multiple number of variables that it may not be a single variable that could be number of variables that may be present.

And you know they could be interacting and that can be represented in the objective function and the constraints would also be similar for multiple resources. There are different numerical methods that exist for solving multivariable non-linear objective functions or constraints. So, let us see what are they.

(Refer Slide Time: 01:26)

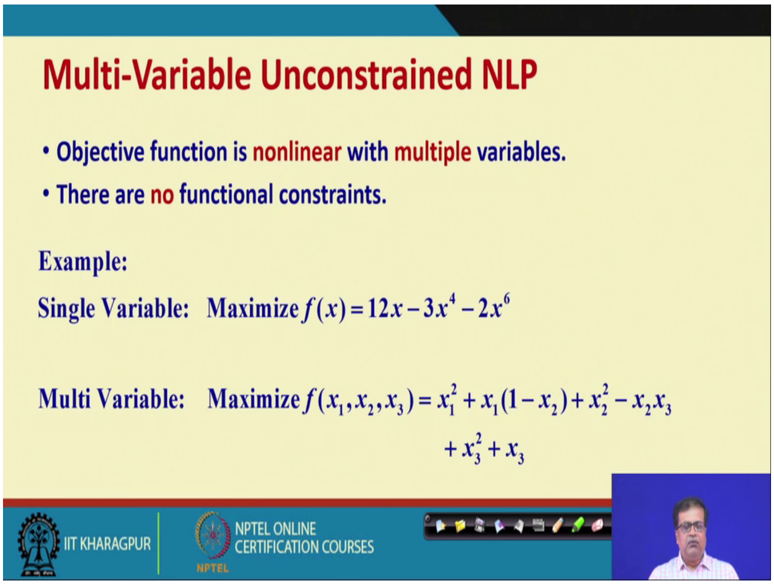


Now, although conceptually they are similar, but you know there are multiple variables, they are present.

So, earlier if you recall we had taken the first order derivative and equated it to 0 to really indicator slope change and therefore, the position of an extreme up. And thereafter taken the second derivative as a sufficient condition to find out whether it is a maxima or minima. But when you have multiple variables instead of first and second order derivative, we really have to take the gradient that is the del operato and the Hessian matrix, which will be an equivalent of the second order derivative.

So, essentially we need the two things the first of all the gradient and second of all the Hessian matrix. And again the equivalent of putting the gradient equal to 0 and the Hessian matrix whether it is negative or positive should be established also. So, those are the challenges that will be there for the multivariable unconstrained problems.

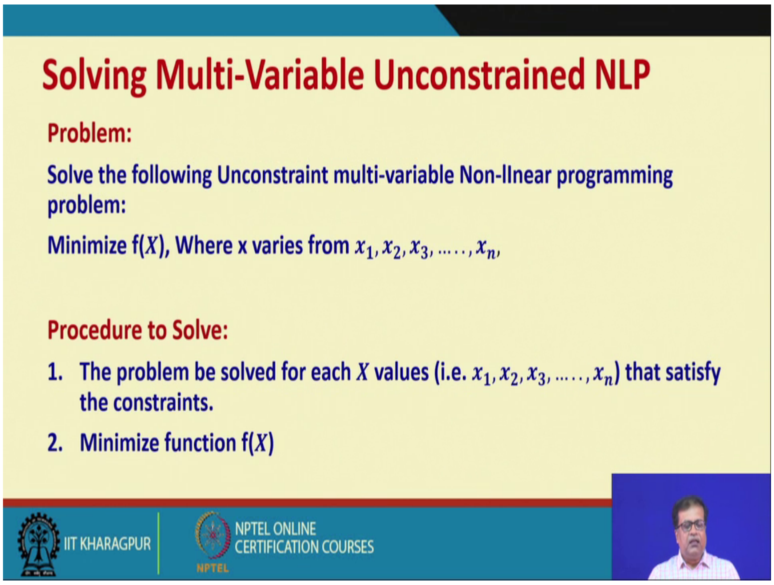
(Refer Slide Time: 02:52)



Now, let us see how exactly they are; so, here is some here are some examples that is a single variable problem f x equal to 12 x minus 3 x to the power 4 minus 2 x to the power 6. And the other one, if you can see that x 1 square plus x 1 into 1 minus x 2 plus x 2 square minus x 2 x 3 plus x 3 square plus x 3 you know that is a multivariable problem.

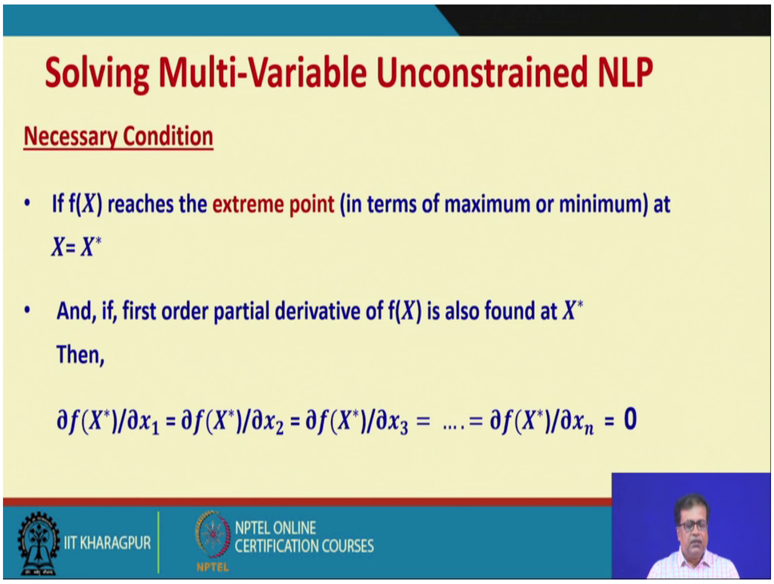
So, as they are all unconstrained problem we have not taken any constraints. So, basically what we need to find out; the maxima for these two problems that example you can see and in other situations you can have minima as well right.

(Refer Slide Time: 03:48)



So, if there are the multi variable unconstraint problem; now we have come to the next stage that is how do we solve this multivariate unconstrained problem. So, essentially the problem is that we have to minimize or maximize f X, where x varies from x 1, x 2, x 3 to x n. And the problem is to be solved for each X values that satisfies the constraints and we have to have the minimize the function.

(Refer Slide Time: 04:24)



So, there are two conditions; one is called the necessary condition in terms of the first order derivatives, but in this case partial derivatives; is it alright? So, that is our del operator, if f X reaches the extreme point whether maxima or minima at X equal to X star, then the first order partial derivative of f X is also found at X star; that means, we must be able to have the first order partial derivative defined at that point; is it alright.

So, it should not be that the function is not continuous and therefore, we do not have the partial derivative existing at that point. So, that is a first condition that the partial derivative should actually exist. After that once we know that yes said that particular point we have the partial derivative existing, then we put individual partial derivatives equal to 0 right.

So, the solution of these you know the individual equations that you shall find by putting the partial derivatives equal to 0, will be our you know the value of the extrema.

(Refer Slide Time: 05:48)

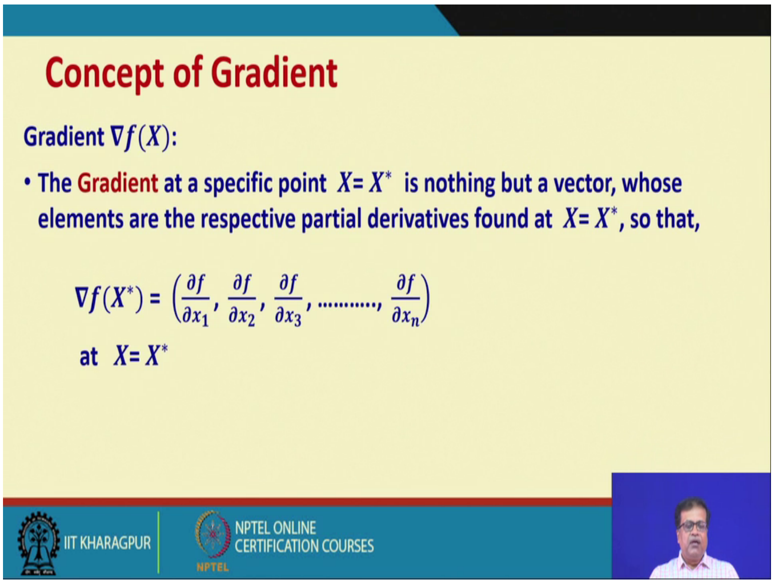
But we have to also have what is known as the sufficient condition; the sufficient condition that should be in terms of the second order partial derivatives. And these matrix that the second order partial derivatives will form is called a Hessian matrix right. The Hessian matrix has to be either positive definite or negative definite or there could be situations where they are neither positive definite nor negative definite.

Now, how to know whether the Hessian matrix is positive definite or negative definite? We shall see in due course of time, but just know this that if you know that the Hessian matrix is a minimum then you know sorry positive definite, then we are going to have a minimum.

But is it a global minima or a local minima that is also to be ascertained; now exactly what are we saying here is that when it comes to the global minima the global minima has to be you know the; suppose there is a card which is actually you know having multiple peaks. For multiple peaks we shall have what is known as a; you know and if you reach a particular one that could be called as a local minima or maxima. So, Hessian matrix you know can indicate yes it is a minima or maxima.

But you know it does not really tell that whether it is an global one right. So, that require further knowledge about the concavity or convexity of the graph which has to be established separately right.

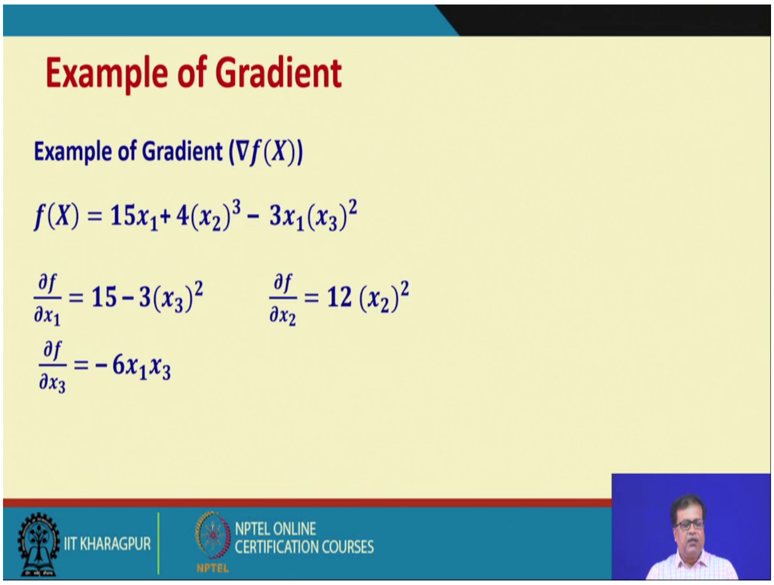
(Refer Slide Time: 07:40)



So, let us look at the gradient; now the gradient at a specific point X equal to X star is nothing, but a vector whose elements are the respective partial derivatives found at X equal to X star right. So, that we call it let us say the del operate del f X star equal to del f del x 1, del f del x 2, del f del x 3 and up to del f del x n at X equal to X star; is it alright, so that is our the concept of gradient.

So, it is all the partial derivative with respect to all the variables that we have at a given us a point.

(Refer Slide Time: 08:29)

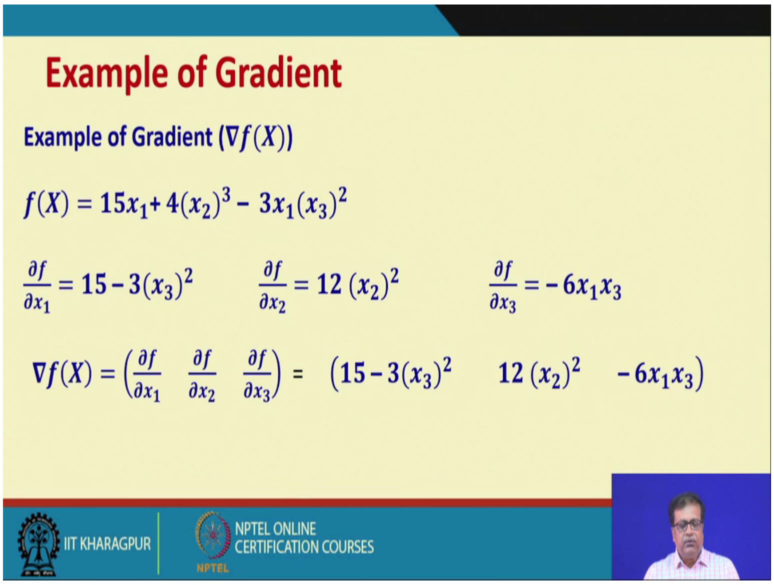


So, that is the definition of gradient. So, here is an example supposing we have a function f X equal to 15 x 1 plus 4 x 2 cube minus 3 x 1 and x 3 to the power whole square. So, what should be the value of the gradient? Now we need to find out the first order partial derivative with respect to x 1, x 2 and x 3.

So, when we take the first order you know the partial derivative of the function with respect to x 1; tell me which are the terms which will be useful for us right? So, as you know in case of partial derivative, the first term which is 15 x 1 you know these term is going to be used, the second term that is 4 x 2 to the power 3; I know this will not be taken simply because there is no x 1 term in that, the third one that is 3 x 1 x 3 whole square the reason x 1 term.

So, when you differentiate the first 15 x 1 will give 15, second term will return 0 and the third term will return 3 x 3 square right; so, that will be our del f del x 1. So, similarly del f del x 2 will be only the second term is having x 2. So, 12 x 2 whole square and third term will be obtained from the third term again because first 2 terms do not have x 3 and that would be minus 6 x 1 x 3.

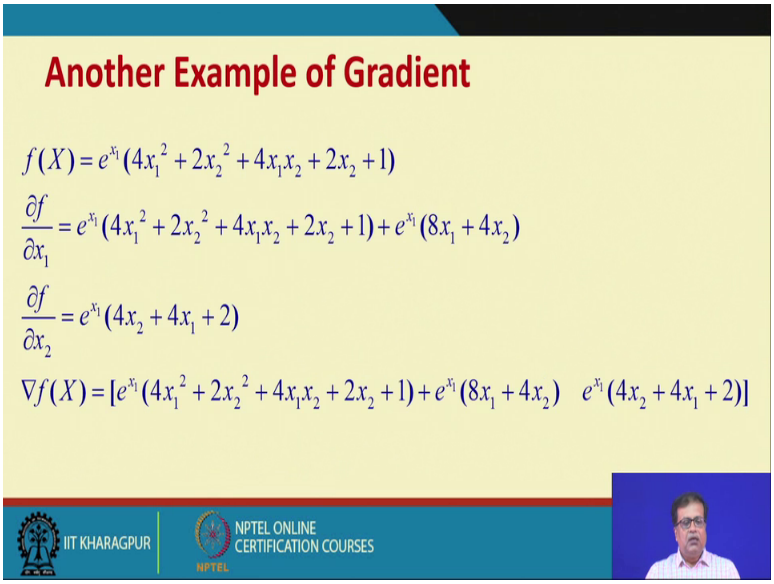
(Refer Slide Time: 10:11)



So, now what do we do? We combine them in the form of a matrix that you know you can see that matrix here.

So, this is our matrix. So, you know this is 15 minus 3 x 3; 12 x 2 whole square minus 6 x 1 x 3. So, these entire thing will be our del f X that is the gradient; is it alright. So, we have the gradient and for this particular function that is obtained.

(Refer Slide Time: 10:46)



So, another example this example is slightly little more complex because we have an e term that is you know f X is e to the power x 1 and within bracket ah; 4 x 1 whole square plus 2 x 2 whole square plus 4; x 1, x 2 plus 2 x 2 plus 1.

Now, this has to be; if you if you differentiate it with respect to you know the x 1, then you know these differentiation has to be done in parts right. The in the in the first one we differentiate e to the power x 1 and e to the power x 1 differentiation is e to the power x 1 itself. So, you can see these term that is you know e to the power x 1 and the whole of the second term will come and in the next stage we keep e to the power x 1 as it is and we differentiate the term.

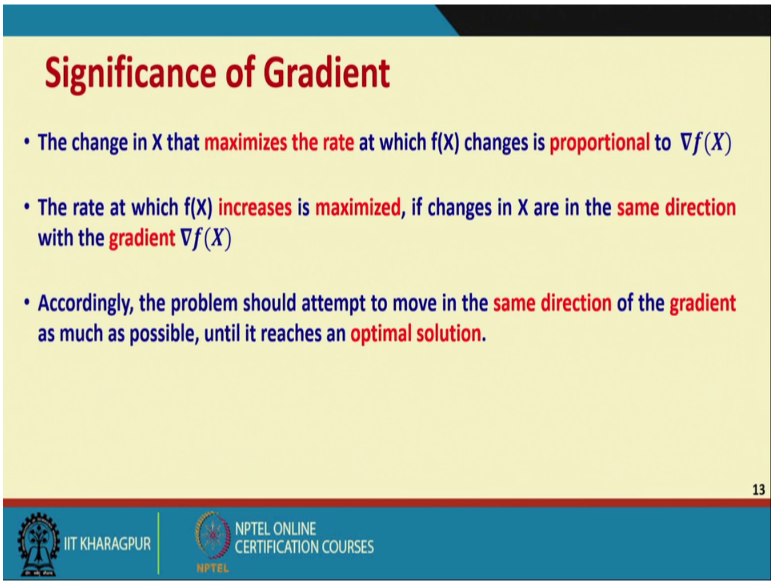
So, we take the x 1 things; so, we have 8 x 1 and there is one more that is here; so we get 4 x 2 right. So, it will be e to the power x 1 8 x 1 plus 4 x 2, but with respect to x 2 it will be easier because e to the power x 1 will be you know not coming into the picture here, there is no need to do that differentiation. So, if you simply multiply them; then actually we have the first term is not a function of x 1, second term is a function; so, it come to 4 x 2.

Then third term gives 4 x 1 next term gives 2; so, e to the power x 1; 4 x 2 plus 4 x 1 plus 2. So, that will be our del f del x 2 and when you combine them all you know all these terms. So, this is this one is the first term that is you know this entire term is here and this is the second term that is here when you put them together that will be our gradient; is it alright.

So, this is how we actually calculate what is known as the gradient. So, first of all we have calculated for 2 examples; in the first example we had 3 variables. So, we took the partial derivative with respect to each of the 3 variables and when you put them in the form of a matrix that will be our gradient. And in the second example is a 2 variable problem, we separately again found partial derivative with respect to x 1 and x 2 and when you combine them that time we got the gradient.

So, you know how to really find out the gradient, we have to put gradient equal to 0 to find out our extreme point.

(Refer Slide Time: 14:01)

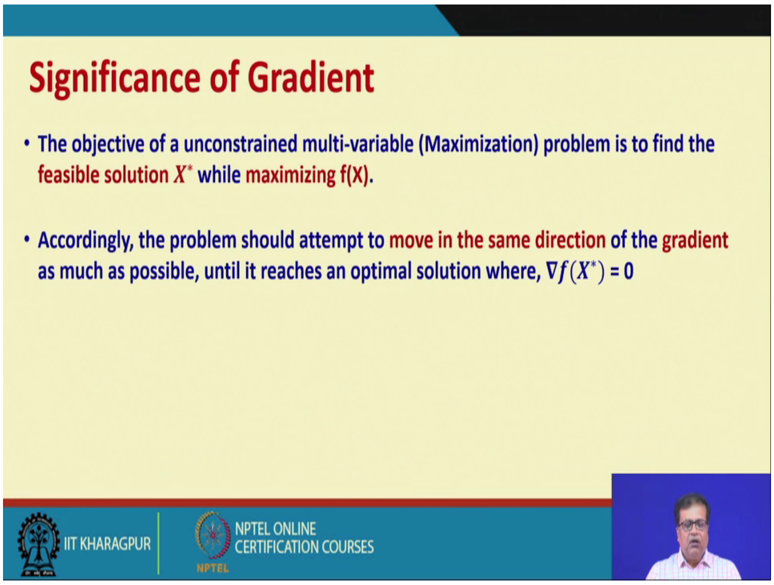


Now, the significance is this the change in X that maximizes the rate at which f X changes is proportional to the gradient. The rate at which f X increases is maximized; if changes in X are in the same direction with the gradient right; so, as long as the direction is same then you keep going, but you know at a point when the slope changes.

So, gradient really represents the slope; so at some point the slope really changes and when the slope changes then the directions will be different also. So, accordingly the problem should attempt to move in the same direction of the gradient as much as possible; until it reaches at optimal point. Because what happens at optimal point? The optimal point the slope changes right, we have seen in those convex or concave you know functions that the slope keep increasing in one case and then it decreases.

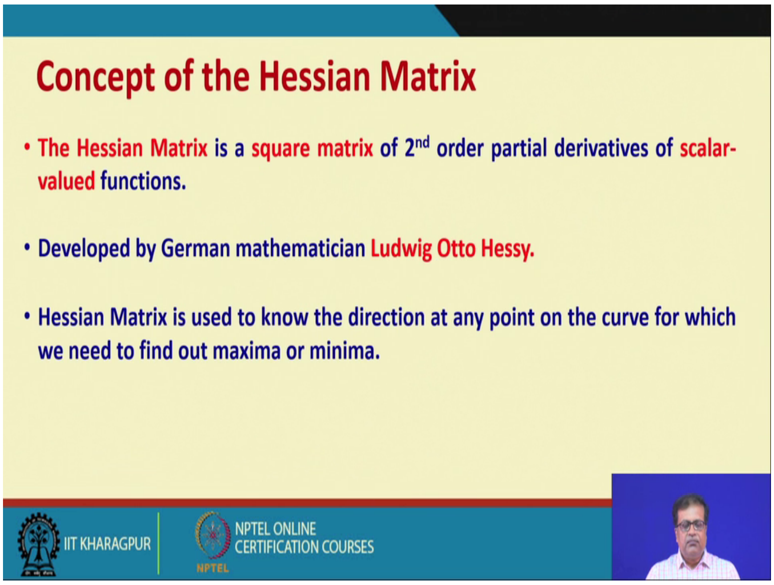
And in the another case the slope decreases in case of minima and then it increases after the extreme point right.

(Refer Slide Time: 15:16)



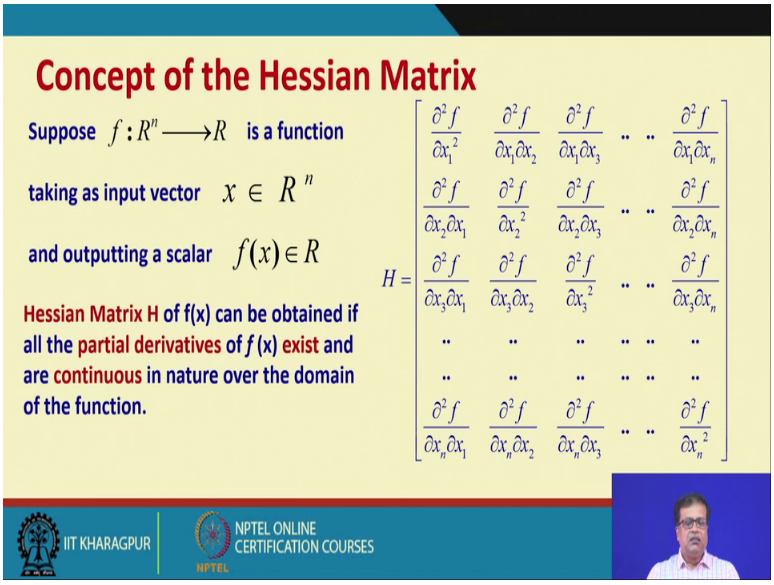
So, the objective of an unconstrained multivariable problem that is maximizing is to find the feasible solution X star, while maximizing f X. And the problem should attempt to move in the same direction of the gradient as much as possible until it reaches the optimal solution; so that is how we make use of gradient to solve problems. Now, having seen the gradient part; let us see the other one that is the second order partial derivative matrix or the Hessian matrix.

(Refer Slide Time: 15:41)



The Hessian matrix is a square matrix of second order partial derivatives of scalar valued functions and developed by German mathematician Ludwig Otto Hessy right and he is again used to know the direction at any point of the curve for which we need maxima or minima.

(Refer Slide Time: 16:12)

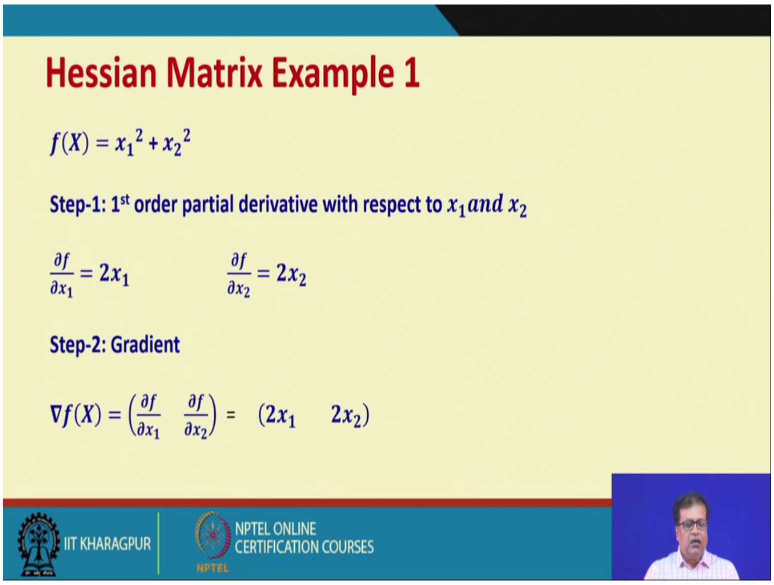


So, what is Hessian matrix? So, you look on this side of the screen we have a Hessian matrix format. So, supposing a function f which is really having you know a number of variables let us say x 1, x 2, x 3, x 4, x 5, x 6 and all the way up to x n, then Hessian matrix will be del square f, del x 1 square. Then del square f, del x 1, del x 2 basically it is nothing, but del del x of you know the del f, del x 2 right.

So, del del x 1; so all of these terms are basically taking del del x 1 of del f del x 1; del f del x 2, del f del x 3 and up to del f del x n. And then the second term is with regard to del del x 2 and; obviously, the second term there will be then del square f del x del x 2 square.

And finally, at the end del square f, del x n, del x 1; del square f, del x n, del x 2; del square f, del x n, del x 3 and finally, del square f del x n whole square; so, these entire matrix it can be called as the Hessian matrix.

(Refer Slide Time: 17:40)



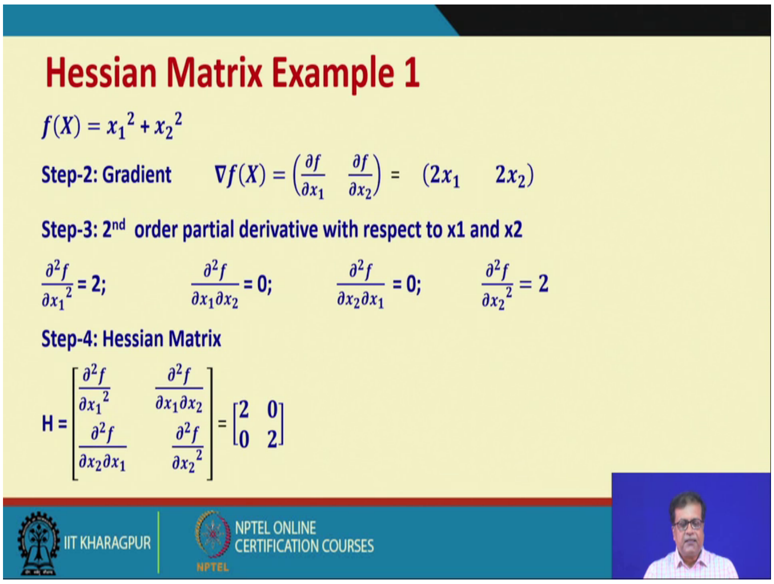
Now, look how to really find out Hessian matrix. So, you know the let us take an example that is f X equal to x 1 square plus x 2 square. So, first we need to find out the gradient the what will be the gradient? What is the partial derivative of this function with regard to x 1; del del x 1 of the function only for first term; is it not. So, we take the first term and we find the derivative and we get what is known as 2 x 1 and if we do it with respect to x 2; that is del f del x 2, then we get 2 x 2 because only the second term will come.

So, the gradient will be the del f del x 1; del f del x 2 that is 2 x 1 and 2 x 2 right. So, now how what will be your Hessian matrix? The Hessian matrix we have to find out first of all del square f, del x 1 square tell me what it is? If you take del del x 1 of del f del x 1; now the del f del x 1 is 2 x 1, if you take it again del del x del del x 1 what will be the value? You have guessed it correctly; it will be 2 right.

And what will be del del x 2 of del f del x 2? Because again the value is 2 x 2; so, it will be 2 again. So, those two terms we know now we need to find out what is del del x 1 of del f del x 2 and what is del del x 2 of del f del x 1. Look here del f del x 1 is a function of x 1 because it is 2 x 1. So, if you take del del x 2; it will be 0 similarly del f del x 2 comes out to be a function of x 2 only that is 2 x 2.

So, if you take again the partial derivative with respect to x 1 that is del del x 1 that will be 0 again. So, what will be the Hessian matrix? You know I hope you have understood now what will be the Hessian matrix.

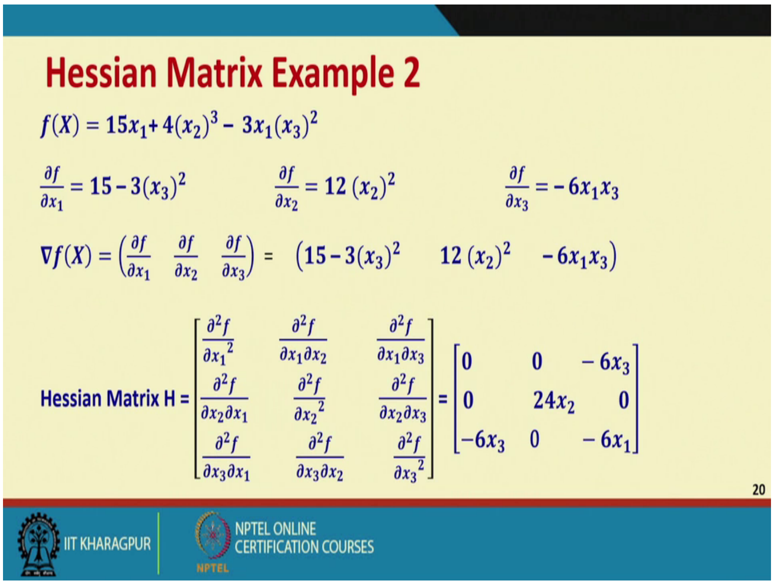
(Refer Slide Time: 20:05)



So, let us see what it is. So, you know we have the function x 1 square plus x 2 square and we have taken the gradient and the gradient has come out to be 2 x 1 and 2 x 2. Now, we compute all these 4 terms; the first term as you have seen that del square f del x 1 square is 2 because this value is 2 x 1.

So, it becomes 2 the last term del square f del x 2 whole square is the del del x 2 of these. So, it will be 2, but the other terms as 0 as you have seen just now. So, if you if you put them all in the form of a matrix that sees that is our Hessian matrix and that Hessian matrix will be 2 0 and 0 2 right. So, I hope you understood how to really obtain the Hessian matrix, I know that is the first part of the thing.

(Refer Slide Time: 21:03)



So, let us take one more example. So, you know these example we have also seen earlier that is the other example that where f X is 15 x 1 plus 4 x 2 cube minus 3 x 1; x 3 whole square. So, when you have taken del f del x 1 del f del x 2 and del f del x 3 you know we really got the gradient. Now, what will be our Hessian matrix? So, when you take the Hessian matrix the first term is a differentiation of these with respect to x 1.

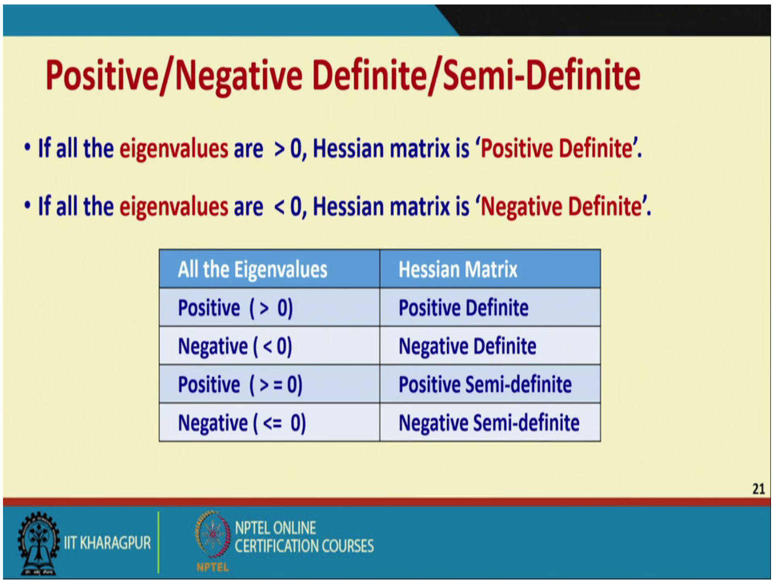
So, what will be that first term? It will be definitely 0 what will be the second term? The second term will be you know, second term is a required del f del x 2. So, del f del x 2 is 12 x 2 whole square; so if you if you differentiate is by del del x 1; what will you get? 0 again and what will be the third term? This is your del f del x 3. So, so if you if you differentiate it with respect to del del x 1 what will you get? You will get you know minus 6 x 3 is it not.

So, like that you can get the Hessian matrix. So, this is how the Hessian matrix is obtained let us see each term; supposing this term just now I said that del f del x 3 is minus 6 x 1, x 3. So, if you differentiate with regard to del del x 1 you get minus 6 x 3. So, again let us see how this 24 x 2 is coming. So, this is our you know this is our term that is del f del x 2; again if you differentiate that is del del x 2 of the term then we will get 24 x 2.

Similarly, third term minus 6 x 1 x 3 is del f del x 3, again you differentiate it with respect to that is del del x 3 it will get minus 6 x 1 right. So, rest of the terms are 0; let us see only this term. So, this term is del del x 3 of del f del x 1; del f del x 1 is this. So, if you take del del x 3 the first term we will get nothing, the second term we will get minus 6 x 3 right. So, I hope you have now understood how to obtain the Hessian matrix.

The Hessian matrix is a matrix of second order partial derivative in a particular form and the Hessian matrix has got a very important significance and what is that significance? In terms of non-linear programming problem right; so, we must understand that. So, in order to understand that let us see what is the meaning of positive and negative definiteness of a Hessian matrix right?

(Refer Slide Time: 24:18)



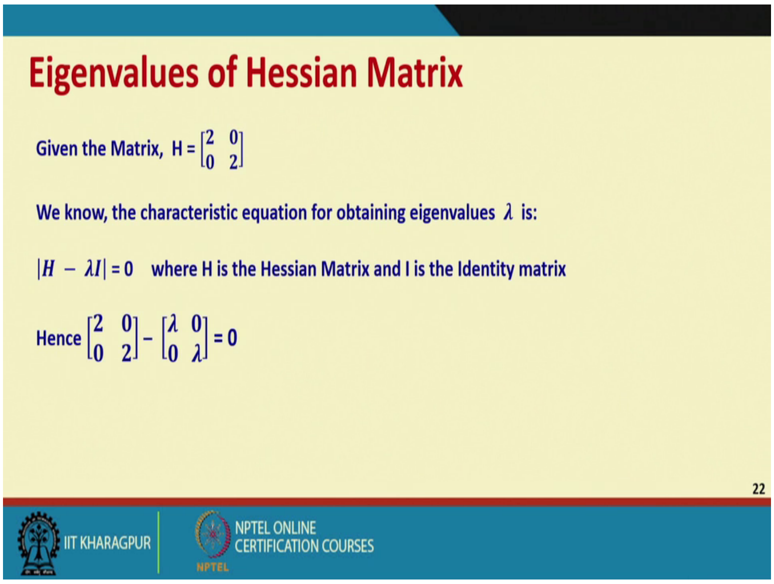
So, you know a matrix we can find out their eigen values right. So, you all have studied the concept of eigen values, I will show you how to find eigen values very quickly this is not very difficult rather easy.

So, once you find out the eigen values then you check that whether those eigen values are positive or negative, but the question is see the eigen values should be all positive or all negative or all greater than equal to 0 or all negative that is all less than equal to 0. So, only these 4 cases are considered here ah; can there be some other cases as well? Is it a complete set? Answer is no, it is not a complete set. Supposing we have 3 eigen values, two could be positive, one can be negative right or one is positive rest two are negative.

So, if we have combination of the eigen values both positive and negative; then we cannot really conclude whether the Hessian matrix is a either positive definite negative definite or positive or negative semi definite right. So, only in these 4 cases are defined; rest of the cases we are not going to take up in the sense that in that case it will be difficult to conclude right. So, now, if the all the eigen values are positive then we call it the Hessian matrix has positive definite.

If all the eigen values are negative then we call all the Hessian matrix as negative definite and if all the eigen values are greater than equal to 0; that means, positive, but greater than equal to 0; then we call the Hessian matrix as positive semi definite and finally, if all are less than equal to 0 then we call the Hessian matrix as negative semi definite.

(Refer Slide Time: 26:42)

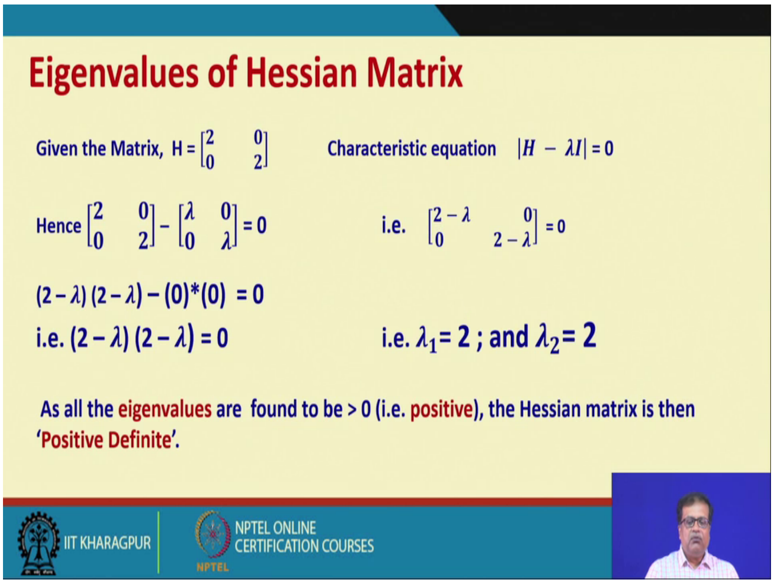


Now, let us see how to find out the eigen values? Now you see, we got in our example one of the situations where Hessian matrix is a 2 0; 0 2 that was our Hessian matrix.

So, we know that we have to have a characteristic equation; the characteristic equation for obtaining eigen value is given by the absolute value of H minus lambda I equal to 0, where H is the Hessian matrix and I is the identity matrix. So, what we do? 2 0 0 2 minus lambda 0 0 lambda you know that is what you get by multiplying and unity matrix 1 0 0 1 with lambda matrix right; so, that is lambda 0 0 lambda.

So, we put that equal to 0. So, if you put that equal to 0 then what are we going to get? See we get a matrix 2 minus lambda 0 0 and 2 minus lambda is it not that is the matrix that is now put equal to 0.

(Refer Slide Time: 27:53)

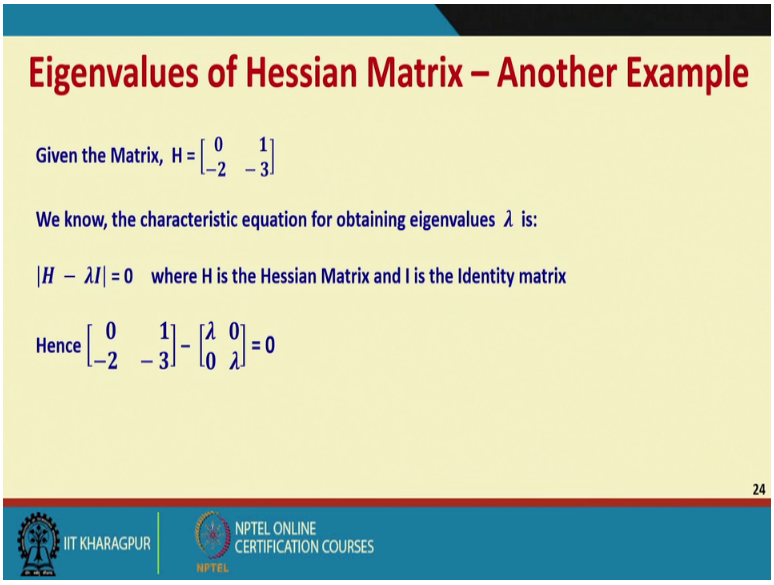


So, that is exactly what is written here that this will become 2 minus lambda 0 0 2 minus lambda equal to 0.

So, when you try to solve them then you multiply 2 minus lambda into 2 minus lambda minus 0 in to 0. So, that is exactly what is written here and that gives us 2 minus lambda into 2 minus lambda equal to 0. So, it is factorized already; so we do not have to do anything therefore, we find lambda 1 equal to 2 and lambda 2 equal to 2 also. So, all the eigen values are found to be greater than equal to 0 that is positive. So, we then say the Hessian matrix is then positive definite.

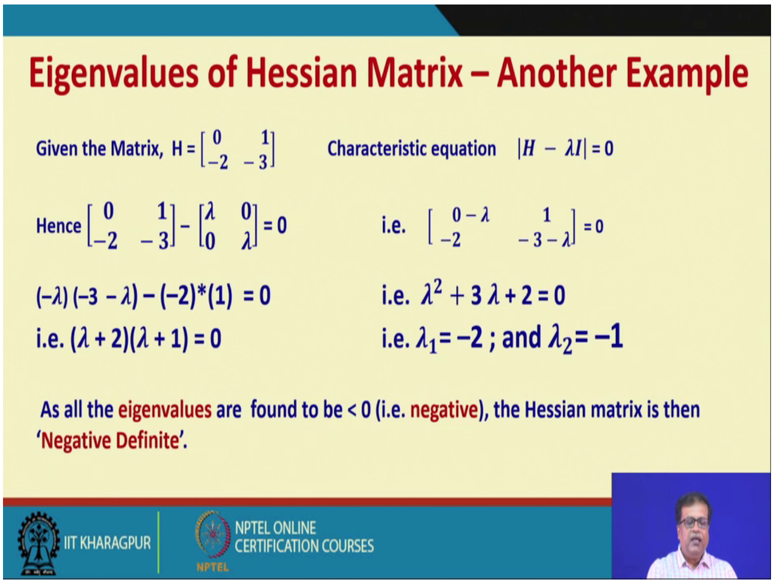
So, what does it mean? It really means that since the Hessian matrix is positive definite that function, if you put equal to 0 that is the gradient of the corresponding matrix you know we then get the minima because we have a positive definite Hessian matrix right that is the significance.

(Refer Slide Time: 29:06)



Let us take another example; suppose the Hessian matrix is 0 1 and minus 2 and minus three. So, follow the same procedure H minus lambda I absolute equal to 0.

(Refer Slide Time: 29:18)



Then we get 0 minus lambda 1 minus 2 minus 3 minus lambda. So, you get this kind of format and when you put them into an equation we get lambda square plus 3 lambda plus 2 equal to 0 by simple factorization we see lambda plus 2 lambda plus 1 equal to 0 or lambda 1 equal to minus 2 and lambda 2 equal to minus 1. So, all the eigen values are found to be negative and therefore, the Hessian matrix is negative definite this is all right.

So, this is how you know we find the gradient, we find the Hessian matrix and we find the Eigen values of the Hessian matrix and from there try to conclude whether the Hessian matrix is positive definite or negative definite right. The positive definite Hessian matrix is connected to a minima when you put the gradient equal to 0. And negative definite Hessian matrix is connected to maxima, if you put the gradient equal to 0. So, in our next lecture we shall see some problems related to this, right.

Thank you very much.