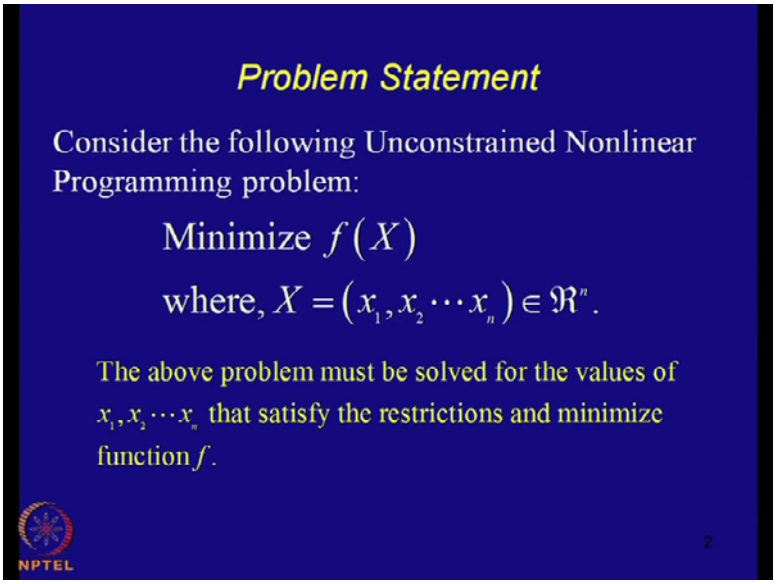
Optimization
Prof. Debjani Chakraborty
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 22
Unconstrained Multivariable Optimization

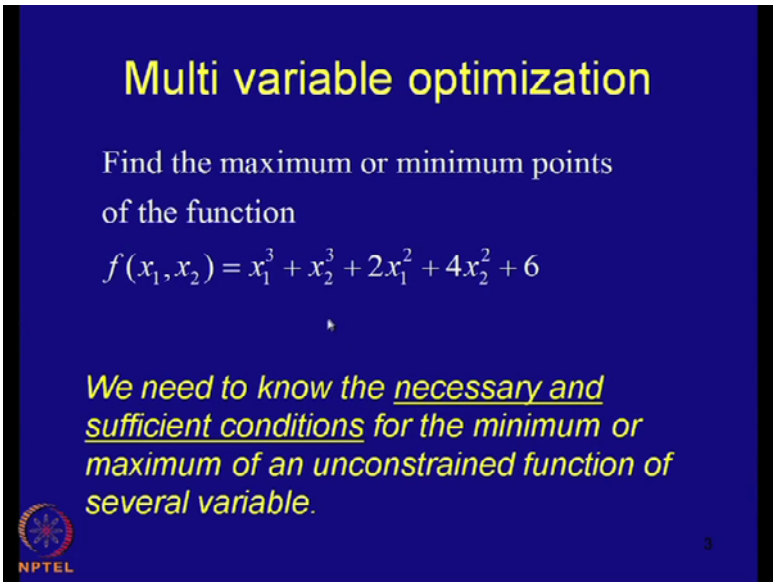
Today, the topic is classical optimization technique for unconstrained several variable optimization. In the last I talked about unconstrained optimization with a single variable. Today, I will talk on the necessary and sufficient condition for tackling unconstraint multivariable optimization problem.

(Refer Slide Time: 00:39)



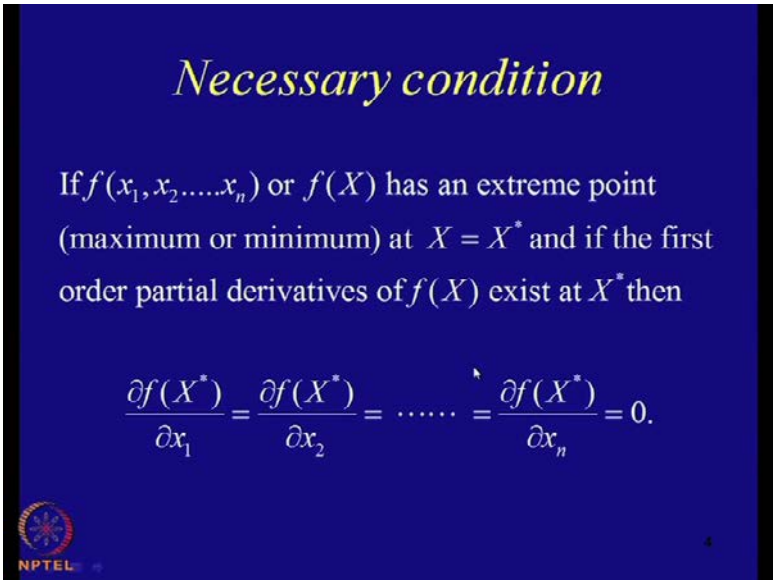
This is the model for the unconstrained general non-linear programming problem, where the objective function is f X. And the function involves n number of decision variables and we want to minimize the function f of X. Similarly, we can maximize the function as well. Now, this is the general form of unconstrained optimization problem. Now, our aim is to find out the values for x 1, x 2, x n, which satisfy the restrictions – means here only the range of the decision variables are given to us; there is no other constraint as such; and we want to minimize the function f here.

(Refer Slide Time: 01:23)



Now, this is the simple example for two variables optimization problem, where we need to find out the maximum or minimum points of the functions – function x 1 to the power 3 plus x 2 to the power 3 plus 2 x 1 square plus 4 x 2 square plus 6. If we see here that, this function involves two variables and there is no other constraints given to us. That is why the variables range is given; there is no restriction on the range even. Now, we want to find out the necessary and sufficient conditions for getting the minimum or maximum of this unconstrained function of several variables.

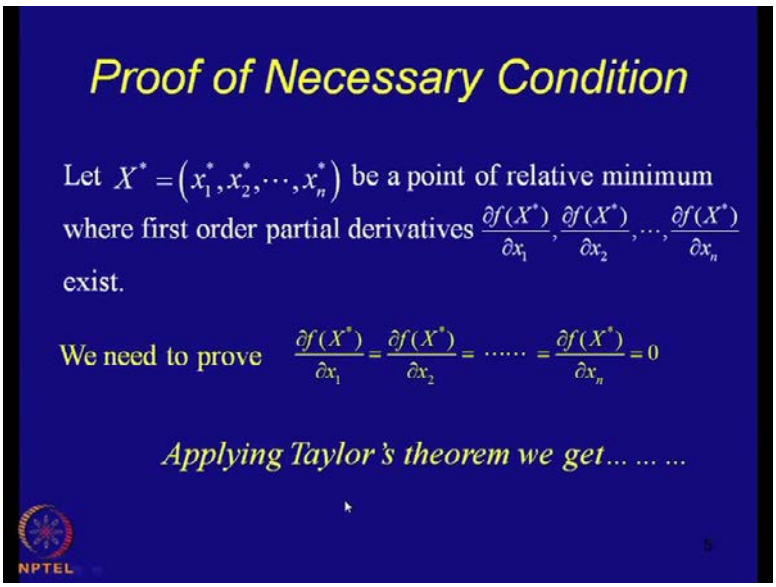
(Refer Slide Time: 02:12)



Now, we will just consider a general decision optimization problem, where the problem involves n number of decision variables. That is why the necessary condition tells us that, if f of X has an extreme point at the point X equal to X star and the first order partial derivatives exist for f; then first order partial derivative derivatives of f with respect to n number of decision variables; or, will be 0 at the extreme point. This is the necessary condition for us.

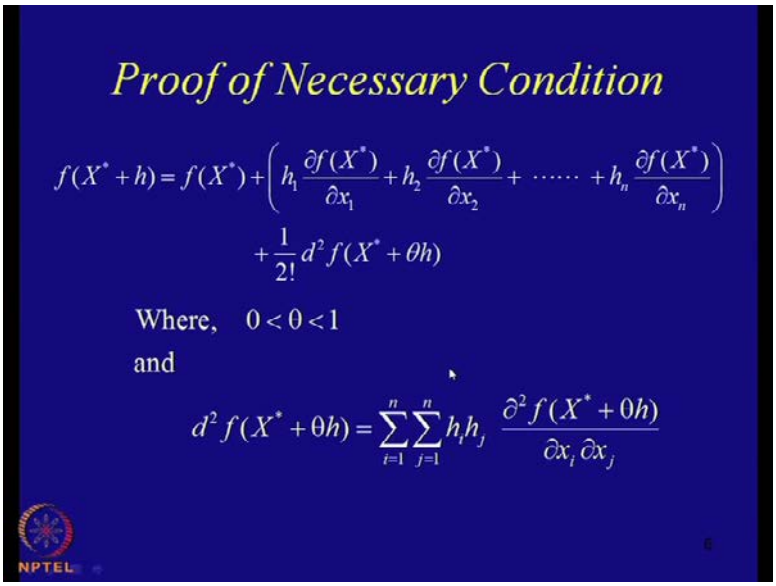
If you remember for the necessary condition involving function of single variable; we had the same condition; instead of the partial derivative, there we had the first order derivative of the function. Since here the function is the function of several variables; that is why we need to consider the first order partial derivatives of f with respect to x 1, with respect to x 2. Similarly, for all other up to n variables – x n variable and we will equate to 0 at the ((Refer Time: 03:17)); then we will get the extreme point x equal to x star.

(Refer Slide Time: 03:23)



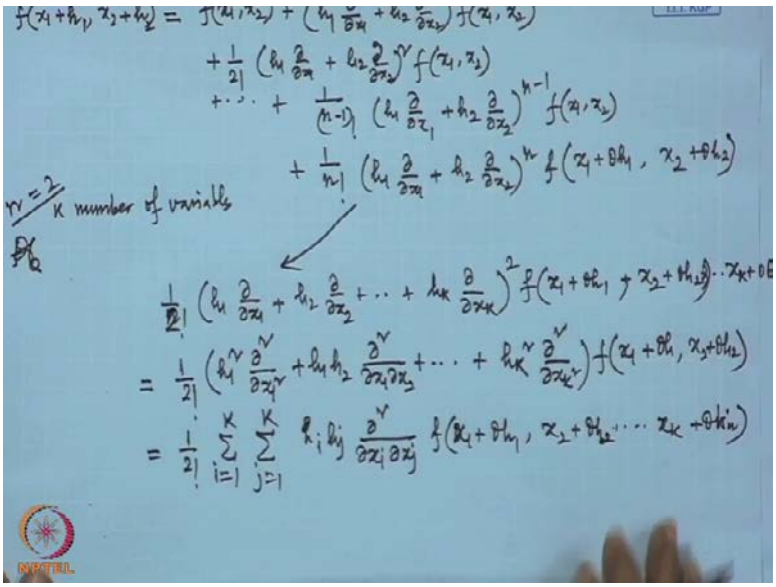
We need to prove it. We need to prove the necessary condition. That is why we need to prove that, if X star is the optimum point; let us consider this is the point; that is the relative minimum for us; then the first order partial derivatives at x star would be del f X star by del x 1, del f X star by del x 2; like that; del f X star by del x n. And we need to prove that, all are 0 individually. Now, let us assume… We will prove this one with the help of the Taylor's theorem.

(Refer Slide Time: 03:56)



Not only that; we will prove the theorem by the contradiction process. As we know for the Taylor's series, if I just want to expand the function in Taylor's series, this is the form given to… This is the form. And we have considered after the second term, because we need to consider that, the function is a twice differentiable and this is the remainder – Taylor's remainder form with us. This is the number of n variables.

(Refer Slide Time: 04:27)



For explaining this thing, let me explain the Taylor's series in general for function of two variables; where, function is a variable. There are two variables if I want to consider;

then f x 1 plus h, f x 2 plus h. In Taylor's form, this would be is equal to f x 1, x 2 plus h – h 1 and h 2. Let me consider two increments – h 1 del x 1 plus h 2 del x 2 of f of x 1, x 2. And the third term would be 1 by 2 factorial h 1 del by del x 1 plus h 2 del by del x 2 square and f of this one.

And, if I just consider up to n number of terms, then it would be minus n minus 1 factorial and h 1 del by del x plus del by del x 1 plus h 2 del by del x 2 whole to the power n minus 1 f of x 1, x 2. And the n h term would be 1 by n factorial h 1 del by del x 1 plus h 2 del by del x 2 to the power n. But, in the remainder term, there will be the involvement of theta will be there; where, theta will lie between 0 to 1. This is a very small value; and that is why this is the expression for us.

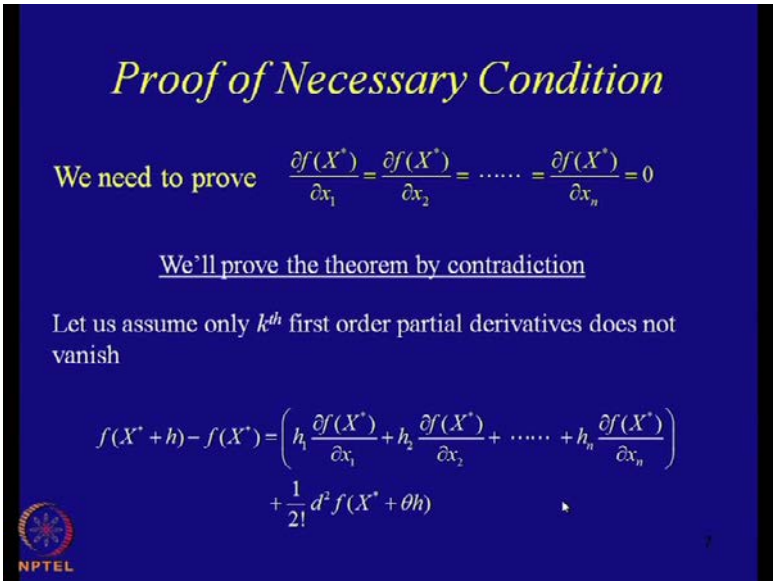
If this is so then we can generalize it for n number of variables. And if we consider that, up to this second order term; then with n number of variables, the second order term. Then considering number of variables – again k number of variables. Here we are considering n is equal to 2 only upto to the second term; not only that; k number of variables.

Then, we will have the remainder term only. Let me just write down the remainder term only. This term would be n factorial h 1 del x 1 plus h 2 del x 2 since there are k number of variables. That is why it would be h k del by del x k to the power 2. And f of x 1 theta h 1 plus x 2 theta h 2. If I expand it further, it would be two factorial rather, because n is equal to 2 here. If I expand it further, it would be is equal to 1 by 2 factorial h 1 square del 2 by del x 1 square h 1 h 2 del 2 by del x 1 del x 2. Like that it will just go further and further; and the last term would be h k square del 2 by del x k square and this would be x 1 plus theta h 1, x 2 plus theta h 2.

The same thing has been written here as well. Here we have considered f x star plus h. At the extreme point, we have expanded the… We have done the same thing. And here we are having this one; where, d 2 f is equal to summation i is equal to 1 to n, summation j is equal to 1 to n h i h j and del 2 f X star plus theta h by del x i del x j. The way we have written here; the same thing. This we have… We can consider here as 2 factorial i is equal to 1 to k; j is eqaul to 1 to k; h i h j del 2 by del x i del x j and f of x 1 plus theta h, x 2 plus theta h 2; and there are k number of variables certainly; there will be k number of variables will be there – theta h k. Similarly, here as well; it will just go up to x k plus

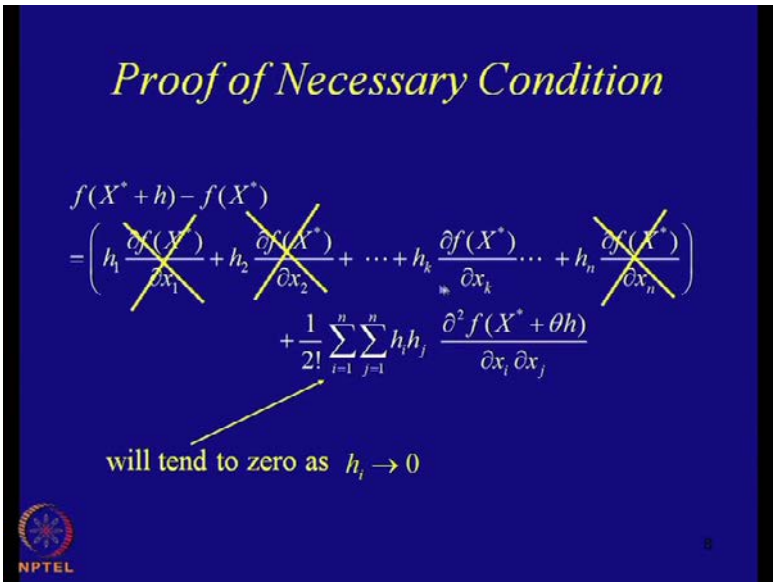
theta h k. Now, this is the way we have just… The same thing we have written with the Taylor series like this. And this is the expansion we need to prove; this is the necessary condition we are proving now.

(Refer Slide Time: 09:16)



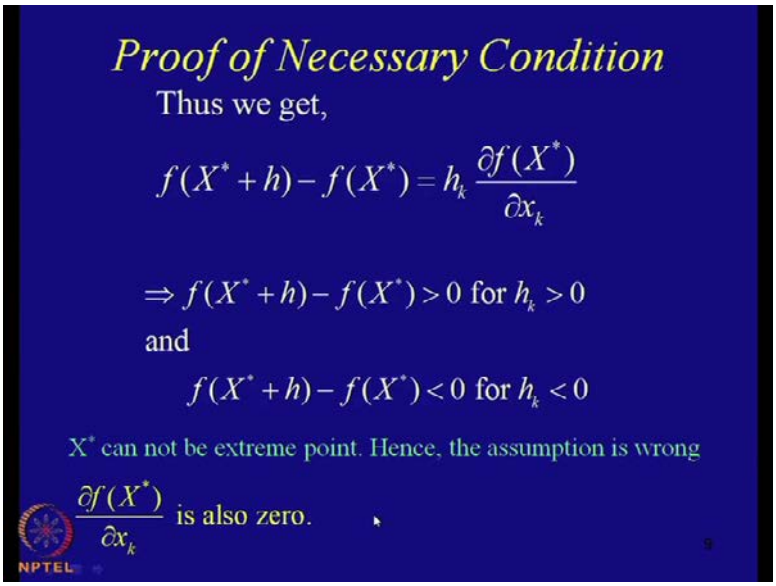
We want to prove that, these terms are all individually equal to 0. Now, we will prove the theorem by contradiction process. We will consider all of the first order partial derivatives are not equal to 0. Let us assume that, k-th first order partial derivatives does not vanish; that means the del f X star by del x k is not equal to 0 and all other 0. And let us see what is happening then. Again we are writing the same expression here. Here it is equal to f X star plus h minus f X star would be is equal to this term plus this one – this thing.

(Refer Slide Time: 09:59)



Now, we are assuming here that, the k-th first order partial derivative does not vanish. That is why if I just write down once again the same expression; we will see that, this will vanish; this equal to 0. The next one again equal to 0 up to this k minus one-th partial derivative 0, k plus one-th partial derivative 0; and the n-th partial derivative is again 0 plus k-th partial derivative is not equal to 0. That is why what is happening then; if this is so; then we are getting… And again this is tending to 0, because h i and h j – these are small increments over the decision variables. When this is tending to 0, this expression is of order h square. That is why this will vanish again. The whole expression then only depend on… The left-hand side is only depending on the right-hand side value.

(Refer Slide Time: 10:47)



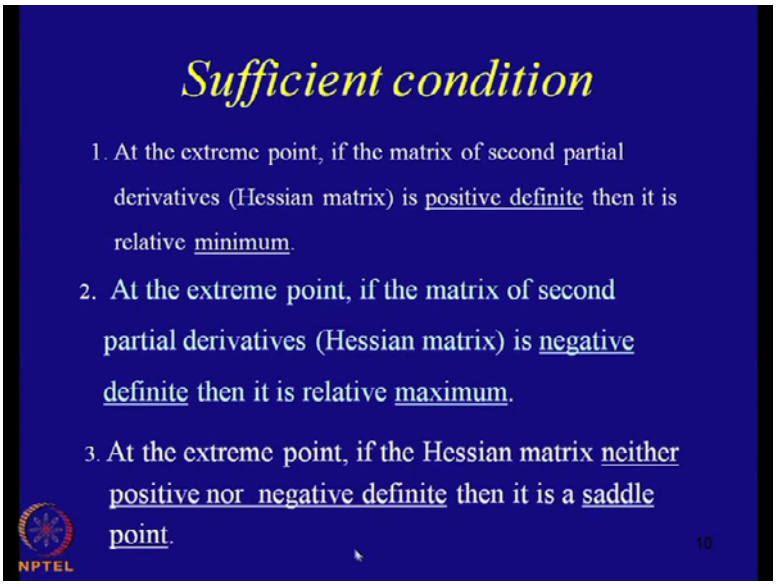
This is the expression for us; we are getting it. Now, here if we assume… We have assumed that, del f X star by del x k is not equal to 0; first, let us assume that, this is greater than 0. If this is greater than 0; that is, del f X star by del x k is greater than 0; then what is happening? For h k greater than 0, the left-hand side expression is greater than 0; for h k less than 0, the left-hand side expression is less than 0, because this we have assumed as greater than 0. If this is so then what will happen? Ultimately, we are proving that, X star cannot be the optimum point. But, this is not acceptable to us. That is why we will conclude that, whatever we have assumed that, k-th first order partial derivative is not equal to 0; that is not acceptable.

Now, the same condition we can prove. Similarly, if we consider del f X star by del x k is lesser than 0; only the condition here; the condition will just reverse. For h k less than 0, it will be greater; for h k greater than 0, it will be lesser than 0. That is why for the both the cases, we are reaching to the conclusion that, x star cannot be the optimal point, because ultimately this change of the functional value is not going to be equal to 0. That is not acceptable.

That is why whatever assumption we made, that is, the k-th partial derivative – k-th order partial derivative, first order partial derivative is not equal to 0; that is not acceptable. That is why we are concluding that, all the first order partial derivatives as well as the k-

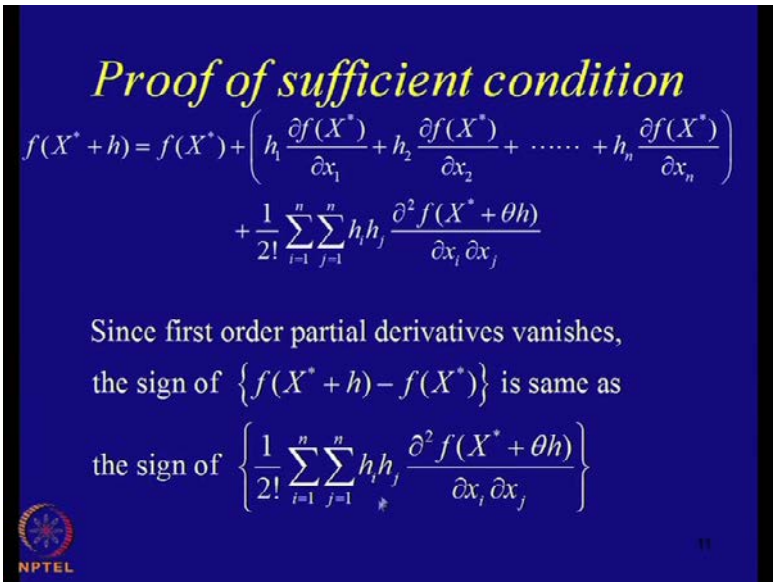
th, the first order partial derivative with respect to k-th decision variables is also equal to 0. That completes the proof of the necessary condition.

(Refer Slide Time: 12:38)



Now, let us come to the next; that is, the sufficient condition proving. In the sufficient condition, the first condition tells us that, at the extreme point, if the matrix of the second order partial derivatives, that is, the Hessian matrix is positive definite; then it is relative minimum. At the extreme point if we see; the matrix of the second order partial derivatives, that is, the Hessian matrix is negative definite; then the corresponding extreme point is the relative maximum. And if we see at the extreme point, the Hessian matrix neither positive nor negative definite; then we are concluding that, this is the saddle point. We are… Then we need to know few things to prove the sufficient condition. First of all, we need to know what is Hessian matrix; then we need to know what is the meaning of the positive definiteness, negative definiteness of the matrix in general. That is why we are now starting the proof of the sufficient condition.

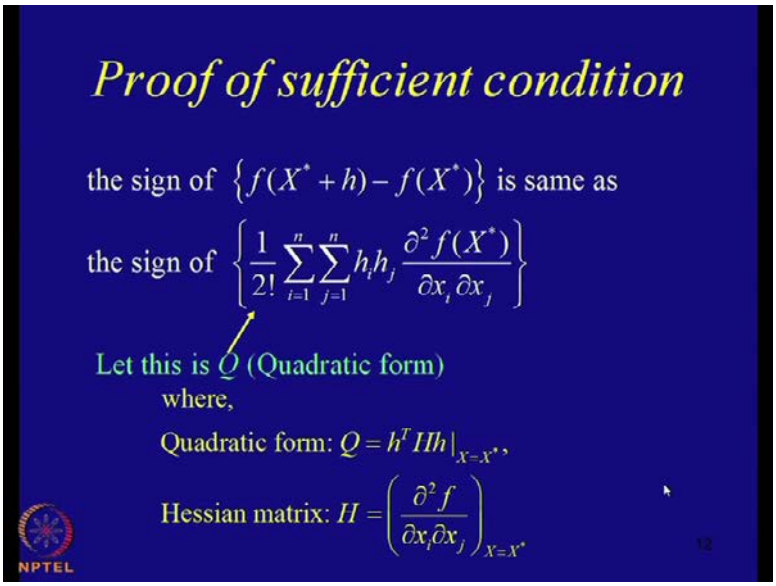
(Refer Slide Time: 13:39)



Again I am coming back to the expression as I have explained, that is, the Taylor series expansion. After the second order, after the second term, we have taken the Lagrange form of remainder. And if we just look at this expression; what this expression… This expression tells us that, as we know from the necessary condition at the extreme point; if we consider X star is the extreme point; then certainly, the first order partial derivatives – all will vanish together.

That is why only the left-hand side expression, that is, f X star plus h minus f X star is only depending on the term, that is, 1 by 2 factorial double summation over i and j; both are running from 1 to n; h i h j del 2 f X star plus theta h by del x i del x j. if you… Just I have showed it to you that, this expression is nothing but an expression, that is, second order, which involves second order partial derivatives. I will talk more on this. The only thing just I want to mention here that, the sign of the change of the functional variable is fully dependent on the the sign of this expression.

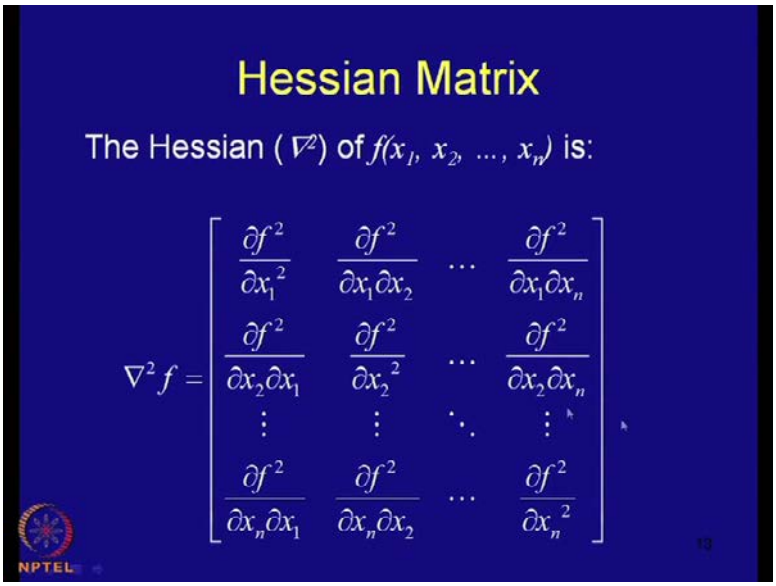
(Refer Slide Time: 15:04)



Let us analyze further what this expression means. As we see, this expression is in the form, is a quadratic form. And this is quadratic form. What is the meaning of the quadratic form we need to know? Not only that; we will show you that, this quadratic form will have the form with the Hessian matrix in this fashion. h T capital H small h at point X equal to X star. And one thing just I want to mention that, if you remember, the last order term was involving the theta. That is the remainder term for us.

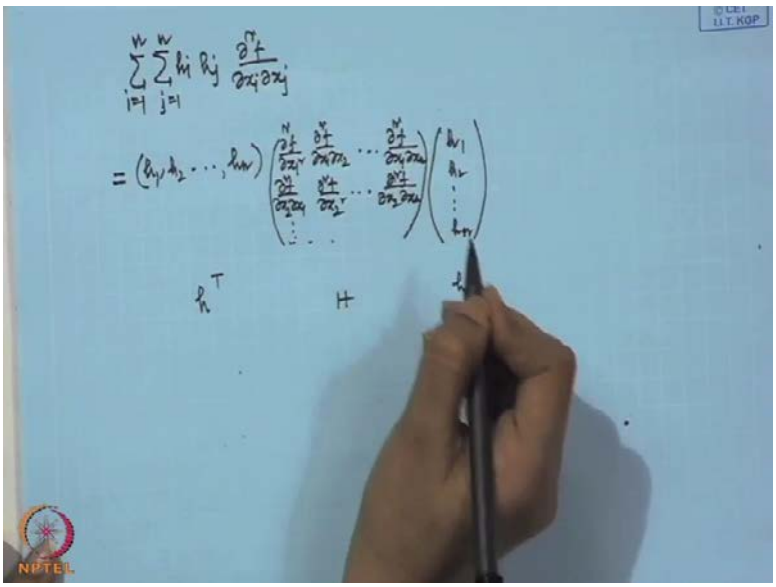
Since x star is the extreme value for us, extreme point for us, in the neighbourhood of X star, this del 2 f X star by del x i del x j will have the same sign with del 2 f X star plus theta h by del x i del x j. That is why we can say that, the sign of f X star plus h minus f X star is fully dependent on the sign of this expression. And this expression is in the quadratic form; that is, this quadratic form is very much related to the matrix notation. And this quadratic form, there are three three terms within this: one is h T; h is the vector we have considered h is h 1 to h n vector; capital is the Hessian matrix for us; and small h is the same vector. The first one was in transpose fashion and this is the in the normal; and this vector at the point X equal to X star. That is why we are concluding here that, the sign of the function is fully dependent on the quadratic form.

(Refer Slide Time: 16:44)



And, this is the Hessian matrix for us. If I just write down in detail, then we can write down the expression – this one. Just see this expression.

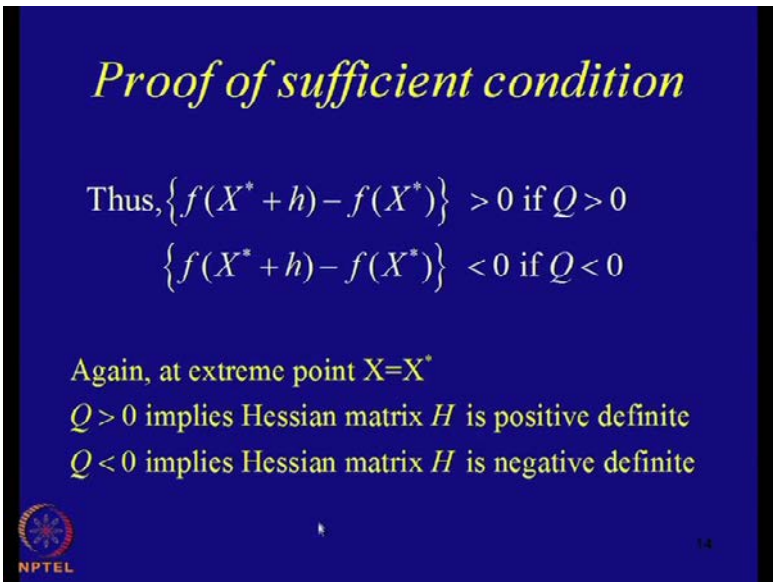
(Refer Slide Time: 16:59)



We were having h i; i was running from 1 to n; j was running from 1 to n h i h j del 2 f by del x i del x j. This thing can be written as h 1, h 2, h n. That is the transpose of the column vector. And in between we are having the Hessian matrix. How Hessian matrix looks like? Just I will just write down in the next. And this is the column vector for us – h n. That is why this is taking the form h T, H, h as I showed you. And this H is in the

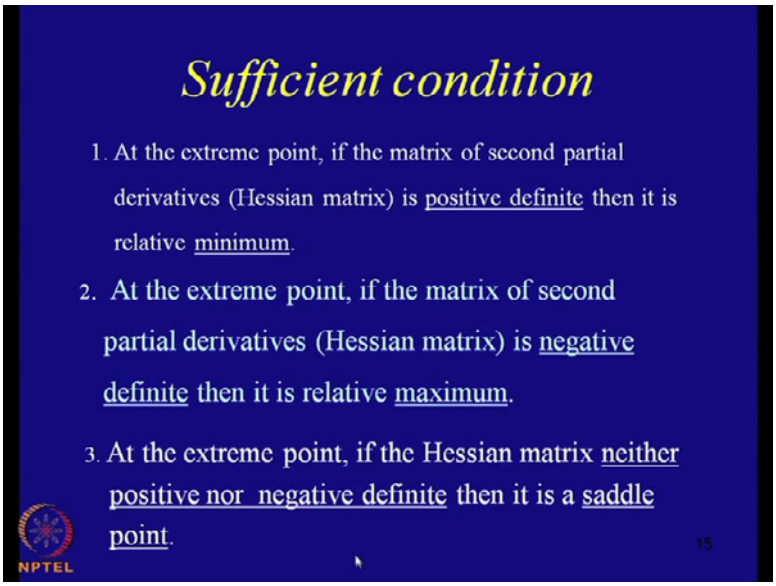
form of del 2 f by del x 1 square, del 2 f by del x 1 del x 2, dot, dot, del 2 f by del x 1 del x n. Similarly, the next row would be del 2 f by del x 2 del x 1, del 2 f by del x 2 square, del 2 f by del x 2 del x n. And similarly, we will get the n-th – the last row as well. That is why I can say this is the Hessian matrix for us. That is the Hessian matrix. As I showed you, this is the Hessian matrix for us. And this Hessian matrix is being written with this term. This is the notation we are using for the Hessian matrix.

(Refer Slide Time: 18:36)



That is why we are saying that, the sufficient condition is fully dependent on the quadratic form. And the sign and the… Rather this is fully dependent on the property of the Hessian matrix. And if we see here, f X star plus h minus f X star is greater than 0 if Q greater than 0 and f X star plus h minus f X star is lesser than 0 if Q lesser than 0. That is why we can conclude that, at the extreme point X equal to X star, if Q greater than 0, then the Hessian matrix must be positive definite. And if Q less than 0, then the Hessian matrix must be negative definite. How the quadratic form is related with the positive definiteness of the Hessian matrix? That I will just tell you in the next. But, one thing is clear that, at the extreme point, whether the extreme point is minimum or maximum, that is fully dependent on the sign of the quadratic form.

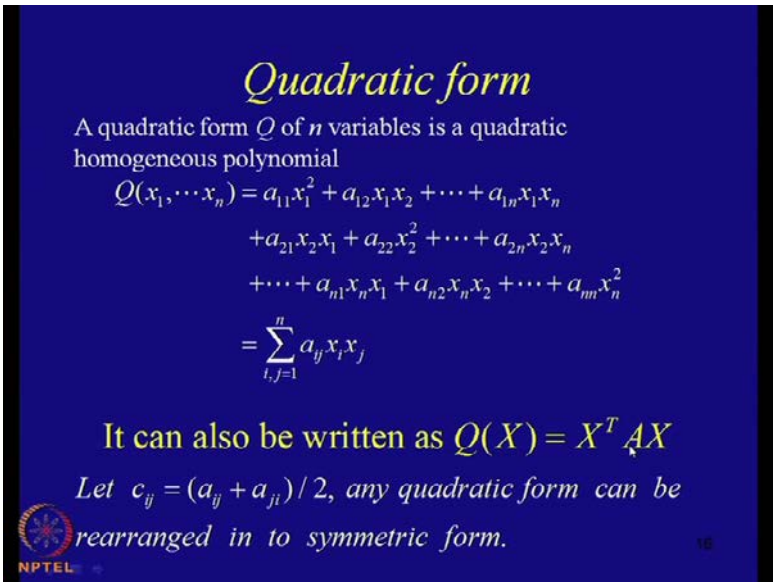
(Refer Slide Time: 12:38)



That is why we can conclude that, if the sign of the quadratic form is positive, then certainly the function is having the minimum value – relative minimum value. That is why we can see at the extreme point, at the relative minimum point, matrix of the second order partial derivative, that is, the Hessian matrix is positive definite. Rather the corresponding quadratic form is greater than 0.

Similarly, we can say that, at the relative maximum point, the Hessian matrix is negative definite; rather the quadratic form is lesser than 0. But, if we see at the extreme point, where the Hessian matrix is neither positive definite or negative definite; that means we cannot conclude whether the quadrtaic form is positive or negative. Then we will say the corresponding point is the saddle point for us.

(Refer Slide Time: 20:24)



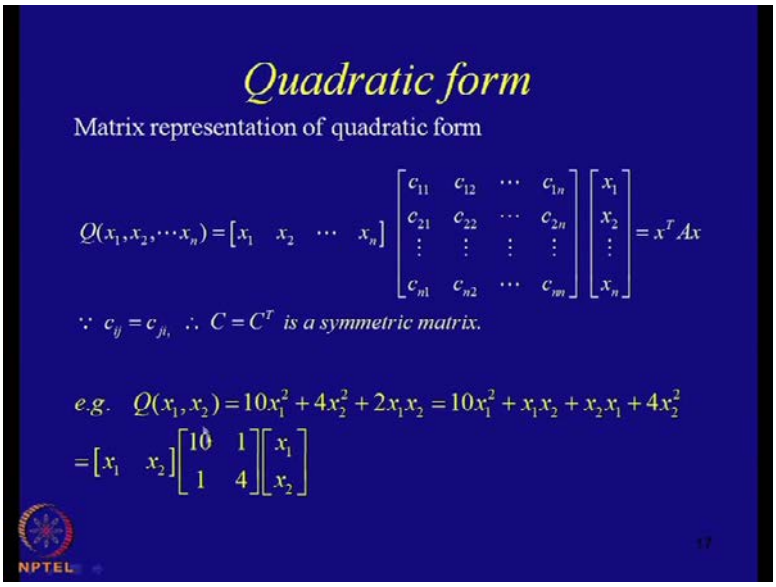
That is why before going to the detail of any example on the unconstrained optimization problem involving n number of variables, let us just go through the the properties of the quadratic form first. If we see that, if there are n number of variables, this is the polynomial of degree 2. How the polynomial look like? This is the quadratic homogenous polynomial. There are terms are there. The terms can be written in summation a i j x i x j. Here Q is equal to a 1 1 x 1 square plus a 1 2 x 1 x 2; like that a 1 n x 1 x n; like that we will get all the terms from the summation notation.

This is the quadratic polynomial for us. This quadratic polynomial is very much related to our sufficient condition for unconstrained several variable optimization, because as we see that, whether the extreme point is minimum or maximum, that is fully dependent on the sign of the quadratic form. If the sign of the quadratic form is positive, then we can see that, the corresponding – the extreme point gives us the relative minimum value. That is why the sign of quadratic form is very important form for us. That is why I am just going through that part of matrix algebra, which we need further.

Now, this can be written in the form of just see. This expression can be written as in the matrix notation X T A X. Now, one thing it is here. Instead of taking a 1 1, a 1 2 etcetera, if I just consider, if I just change the term in this way; I want to change the matrix a to a symmetric matrix. That is my target. That is why what we will do, we will just change all the coefficients with this form – c i j is equal to a i j plus a j i divided by 2. Then the

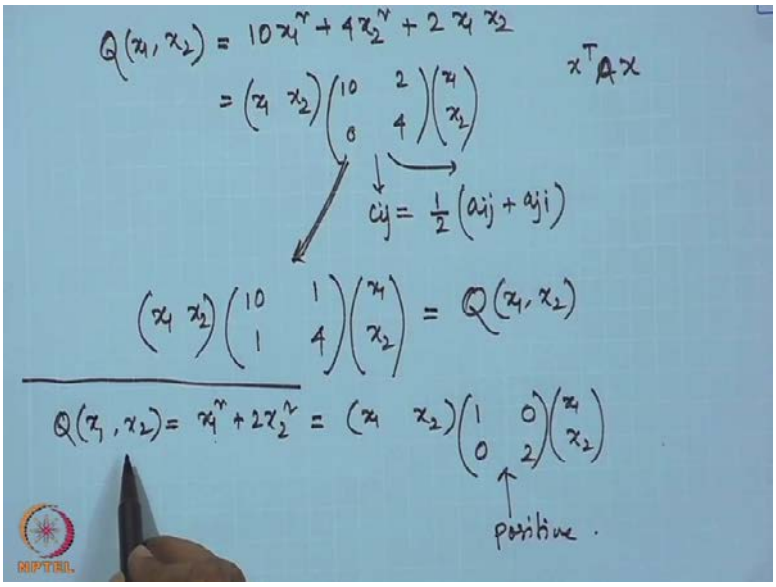
corresponding quadratic form will be a quadratic form, which involves the symmetric matrix. Symmetric matrix means if a is a symmetric matrix, then a must be is equal to a transpose. That is why we will just see how the symmetric form we are getting in the next.

(Refer Slide Time: 22:48)



That is why, we can say that, this is the symmetric matrix. This should be x T C x; and where C is a symmetric matrix.

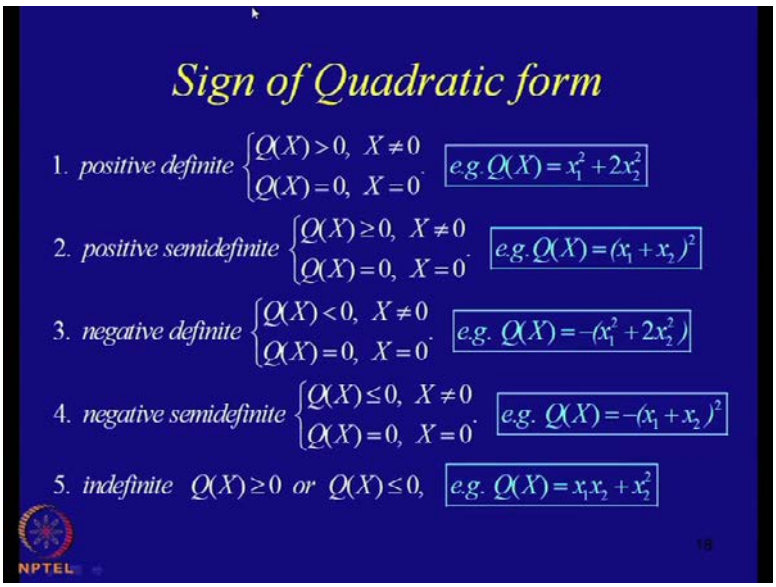
(Refer Slide Time: 23:07)



Let us see with an example here. If I look at this example; what we see here; this is the quadratic form given to us. Q of x 1 x 2 is equal to 10 x 1 square. This is a homogeneous quadratic polynomial – x 2 square plus 2 x 1 x 2. If I want to write in the form of this way; just x T A x; then I have to write this one as x 1 x 2 10, 4, 2, 0; and here x 1, x 2. Now, here you will see that, this matrix is not a symmetric matrix. If I want to make this matrix as a symmetric matrix; we will just use that equation. As I have told you, we will just take a i j plus a j i.

That is why this matrix will become 10, 1 and 1, 4. Certainly this matrix is a symmetric matrix, because transpose of this matrix is again the same matrix. And here if I just write x 1 x 2; here also x 1 x 2. And we can conclude that, this is again Q x 1 x 2. That is why the quadratic form can be written in the symmetric form. Thus, in general quadratic form is a related with a symmetric matrix in matrix algebra. That thing has been written here; that Q x 1 x 2 has been written in the quadratic form x T C x as I have just told you.

(Refer Slide Time: 24:53)

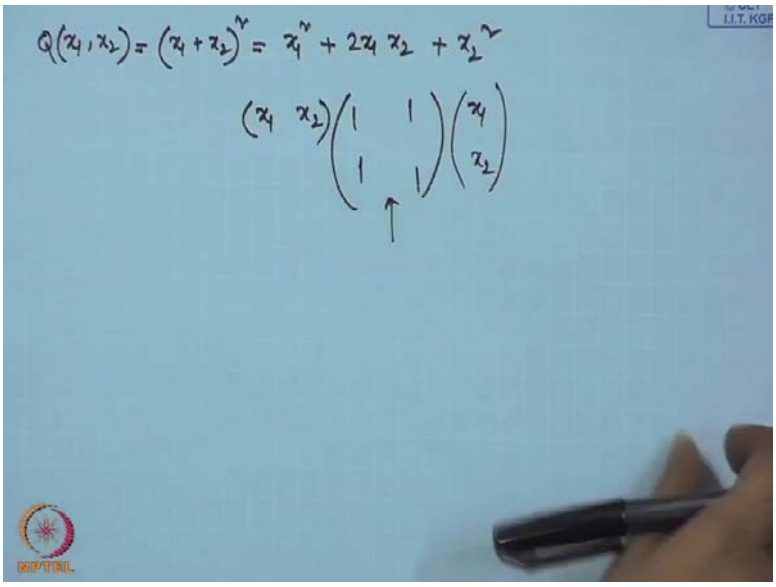


Now, we will just see further properties of the quadratic form. It tells us that, if the quadratic form is greater than 0, then the corresponding matrix rather the symmetric matrix is being said as the matrix is positive definite. For example… Just look at the example; Q X equal to x 1 square plus 2 x 2 square. If you just see the expression here; for every x 1 and x 2 in the space, always Q value will be greater than 0. That is why the corresponding matrix whatever we will get; what is the matrix for x 1 square plus 2 x 2

square? If I just write down that matrix again; we will see that… If I just write down the Q x 1, x 2 is equal to x 1 square plus 2 x 2 square; then this matrix would be x 1, x 2; the quadratic form 1, 0, 0, 2 and x 1, x 2.

Certainly this matrix is the symmetric matrix. Not only that; this matrix is positive definite. Why? Because for every x 1 and x 2, this Q value is always greater than 0. And for X equal to 0, Q value is equal to 0. That is why we say the corresponding matrix is the positive definite matrix. Similarly, for the next. If the quadratic form is greater than equal to 0 for X not equal to 0; then the corresponding matrix is the or the corresponding quadratic form is positive semi definite. Look at this expression for example. The example has been given; Q X equal to x 1 plus x 2 whole square. If we just see this expression; we see for every x 1, x 2, this value is always greater than equal to 0, when it will be equal to 0? When x 1 will be equal to x 2.

(Refer Slide Time: 26:53)



And, this quadratic form again can be written in terms of the symmetric matrix; just see. Q x 1, x 2 is given as x 1 plus x 2 whole square; that means this is equal to x 1 square plus 2 x 1 x 2 plus x 2 square. That is why this can be written in terms of x 1, x 2; 1, 1, 1, 1; and this would be is equal to x 1, x 2. And this is again a symmetric matrix, because a transpose is equal to a; all right? And not only that, for x 1 is equal to x 2, always we will get the corresponding quadratic form as equal to 0. That is why for X not equal to 0; for every x 1, x 2, Q X will be greater than equal to 0; for X is equal to 0, it will be 0. That is

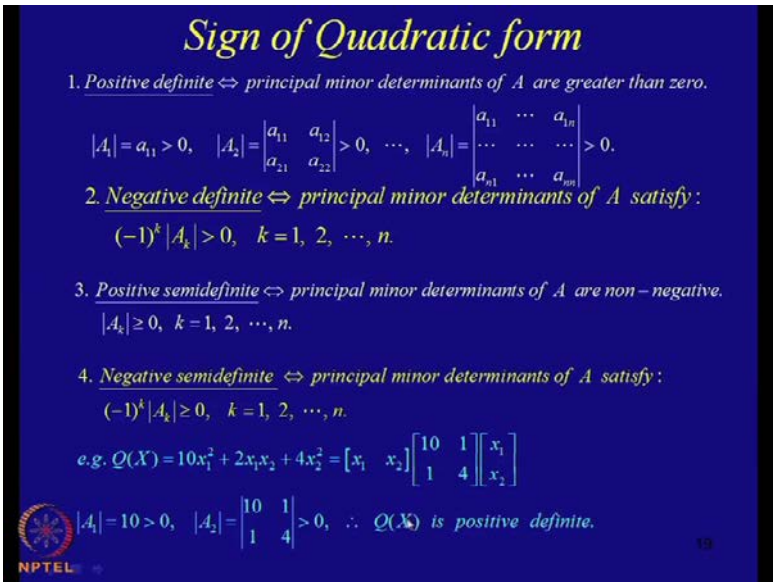
why the corresponding matrix would be positive semidefinite. And if we see, the sufficient condition involves the positive definiteness. For negative definiteness, here also another concept will be involved; that would be positive semidefiniteness and the negative semidefiniteness.

Similarly, in this line, we can say one matrix – the quadratic form would be negative definite if Q X lesser than 0 for X not equal to 0; Q X equal to 0 for X equal to 0. For example, Q X would be is equal to minus x 1 square plus 2 x 2 square. For every x 1 and x 2, this – the bracketed term would be always positive. That is why always the quadratic quadratic form would give you the negative value. That is why the corresponding form would be the negative definite. Similarly, here in the next negative semidefinite, where we are considering the relation Q X less than equal to 0. And here for x 1 equal to x 2 only, the Q X would be is equal to 0; otherwise, for every nonzero x 1 and x 2, for every nonzero capital X, that is, the vector x, the Q X value will give us the negative value. That is why this is negative definite.

But, in the next, the example, Q X equal to x 1 x 2 plus x 2 square. For different x 1, x 2, we cannot conclude anything about the sign of Q. That is why for some cases, for some combination of x 1 and x 2, Q will be positive; and for some combinations of the x 1 and x 2, the Q value would be the negative value, rather the nonpositive value. That is why in that case, we will say the corresponding quadratic form is the indefinite form. If this is so then how to check the positive definiteness or the negative definiteness of a matrix. In general, checking the positive definiteness, negative definiteness from the given expression; it is not so easy.

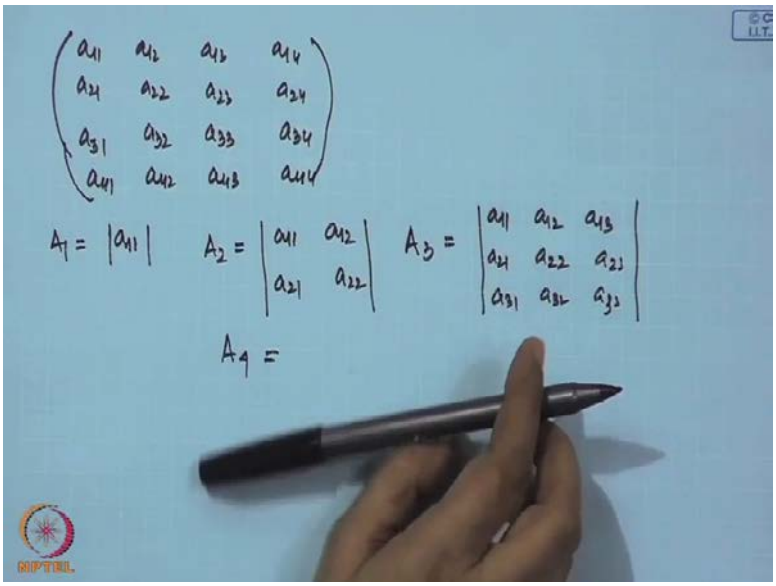
The examples I showed you. Since those are involving only the two number of variables; that is why we could say very easily whether the quadratic form is positive or negative or nothing can be said about that. But, if we have n number of variables – more than 2 number of variables – 3, 4, 5, 6; looking at the expression of the quadratic form, it is not always so easy to say whether the corresponding matrix is positive definite or positive semidefinite, etcetera. That is why they should have some process to check it.

(Refer Slide Time: 30:22)



The process is that, we will… There are two processes rather. One process is that, we will just see the principle minors of the determinant – of the corresponding determinant. From there we will conclude. This is either this way or we will just look at the eigenvalues of the corresponding matrix. From there also we can conclude about the positive definiteness or the negative definiteness or semidefiniteness or indefiniteness about the matrix. Now, if it is positive definite; if the principle minor of A are all greater than 0; principle minor determinant means that, k-th principle minor would be; we will just consider first k-th row and first k-th column together. Then that will give us k-th principle minor.

(Refer Slide Time: 31:14)

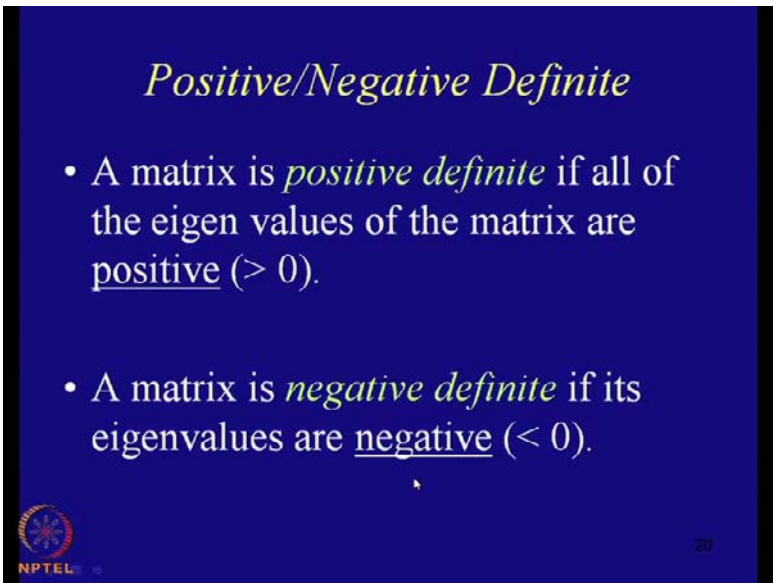


In that way, if we are having the matrix a 11, a 12, a 13, a 14, a 21, a 22, a 23, a 24, a 31, a 32, a 33, a 34, a 41, a 42, a 43, a 44. Then the first principle minor of this matrix would be A 1. The first principle minor will have give me the corresponding determinant. The second principle minor would be is equal to the first two rows and first two columns. That is why a 11, a 12, a 21, a 22. The third principle minor would be the first three rows and first three columns. That is why it would be a 11, a 12, a 13, a 21, a 22, a 23, a 31, a 32, a 33. And the fourth principle minor will coincide with the given matrix. Since the given matrix is of the order 4; that is why here if we are having a matrix of size n by n, we can say that, this is the first principle minor, second principle minor; this is the n-th principle minor.

If we see that, all the determinant values are positive; then the corresponding quadratic form is positive definite. And if we see that, the principle minor determinants are having the sign – alternate sign starting from negative; then the corresponding quadratic from would be negative definite. And that is why k will run from 1 to n. The first principle minor will give me the negative value; second principle minor will give me the positive value; third one negative; fourth – positive. In that way we will get the alternative sign. And similarly, positive semidefiniteness. We will have the principle minors, which are all non-negative values. That is why some principle minor may have the 0 value as well. Then only we can say the corresponding form is the positive semidefinite – the matrix.

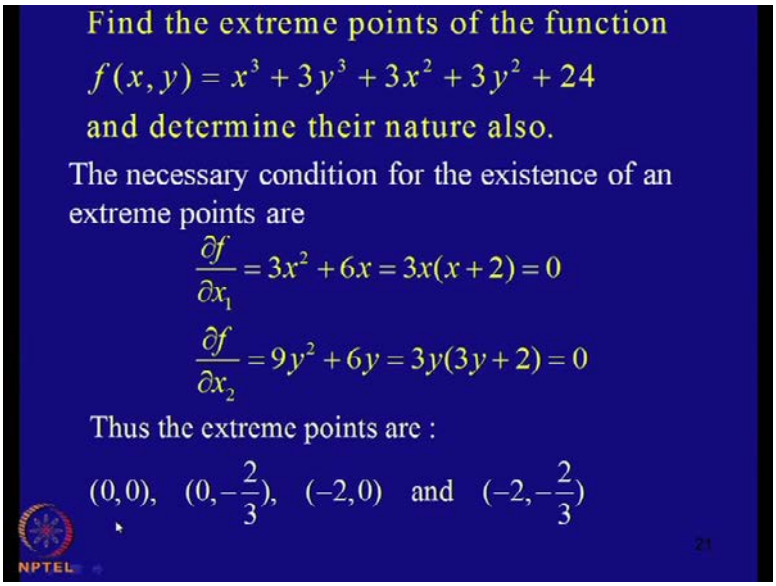
And similarly, negative semidefinite would be; if we are having the alternative sign of the determinant values starting from negative. And indefinite; if we would not get any pattern; then we will say, the corresponding matrix is indefinite in nature. This is one of that example. The same example we have taken before. Just see the example. 10 x 1 square plus 2 x 1 x 2 plus 4 x 2 square. This can be written in the quadratic form in this way – x 1, x 2; 10, 1, 1, 4; x 1, x 2. This is a symmetric matrix for us. If we look at the principle minors, the first principle minor would be greater than 0; second principle minor – the determinant value is again greater than 0. In that way, if we are having a big expression with few more number of variables; from there very easily we can conclude whether the corresponding quadratic form is positive definite or not.

(Refer Slide Time: 34:20)



And, this is… That was one of the check. And this is the alternative check; alternative check whether the matrix is positive definite or negative definite. It has been said that, a matrix is positive definte if all the eigenvalues of the matrix are positive; and it is negative definite if all the eigenvalues of them are negative, because we need this idea for when we are expressing the case whether the Hessian matrix is positive definite. We are concluding the corresponding extreme point is the minimum point – relative minimum point. If the Hessian matrix is negative definite, we are concluding that, the corresponding extreme point is the maximum point – relative maximum point. That is why checking the positive definiteness, negative definiteness – these are very important for unconstrained optimization problem.

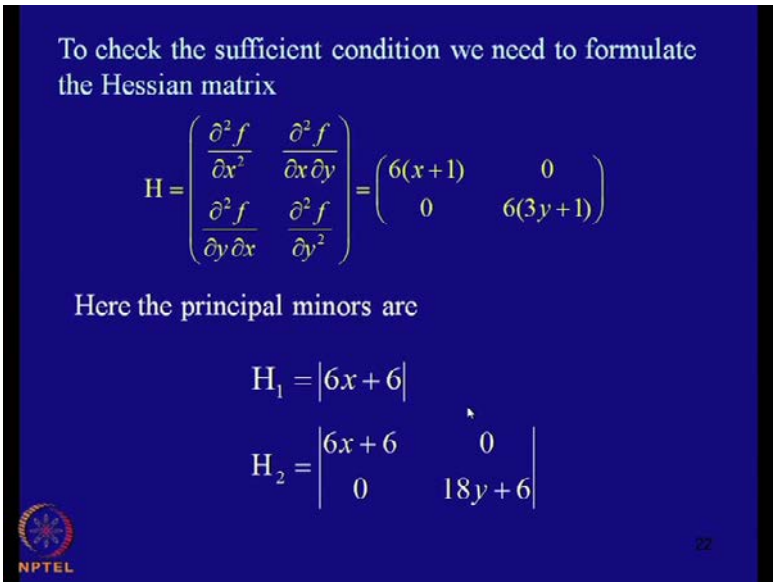
(Refer Slide Time: 35:14)



Let us take an example, where the function involves two variables: x and y. And this is the expression – x cube plus 3 y cube plus 3 x square plus 3 y square plus 24. And we want to have the extreme points first. For checking the extreme points as we know, the necessary condition offers us the possible extreme members of the extreme points. That is why we will take the first order derivative – partial derivatives with respect to individual decision variables; and we will equate to 0. del f by del x 1; it is coming 3 x square plus 6 x. That is equal to 3 x into x plus 2. And this we are equating to 0. We will get two points from here: one point x equal to 0; and another point x equal to minus 2.

Similarly, from the second – del f by del x 2, we are getting two points for y: one is y equal to 0; and another one is y equal to minus 2 by 3. That is why we can say these are all the extreme points we are getting from the necessary conditions. With every combinations of the x and y, we got just from the previous equations. We are getting four extreme points. Through these necessary conditions, we are getting the extreme points for function of several variable. Now, we do not know whether 0, 0 is the relative maxima or minima. That is why we need to have the next level check, that is, the sufficient condition check for it.

(Refer Slide Time: 36:48)



And, for checking the sufficient condition, as I have proved that, we will look at the property of the Hessian matrix, because the Hessian matrix will tell us whether the corresponding extreme point is a relative minimum or maximum or the saddle point; rather we will look for the quadratic form we are getting through the Hessian matrix. Whether the quadratic form is positive or negative, accordingly, we will conclude further. And this is the Hessian matrix for us; for this would be a 2 by 2 matrix, because we are having only two variables here. And del 2 f by del x 2 – we are getting from the previous. We are getting del f by del x 1 as 3 x square plus 6 x. That is why del 2 f by del x 1 square. Here it should not be x 1, x 2; it should be x and y.

In the next, we are having del 2 f by del x 2 is equal to 6 into x plus 1; and del 2 f by del x del y would be is equal to 0, because del f by del y was not involving any x component there; that is 0. This is our Hessian matrix. We will just see the nature of this Hessian matrix at different extreme points. That is why we will go for the principle minors of this Hessian matrix. Looking at the sign of the principle minors of this Hessian matrix, we can conclude whether the given – whether the extreme points which, we got from the necessary conditions – these are all the minimum or maximum or the saddle point.

(Refer Slide Time: 38:30)

Extreme points (X)	H_1	H_2	Nature of H	Nature of extreme points X	$f(X)$
(0,0)	+6	+36	Positive definite	Relative minimum	24
$(0,-\frac{2}{3})$	+6	-36	Indefinite	Saddle point	$\frac{76}{9}$
(-2,0)	-12	-72	Indefinite	Saddle point	28
$(-2,-\frac{2}{3})$	-6	+36	Negative definite	Relative maximum	$\frac{76}{9}$

In this table, everything has been given in a nutshell. Just see… Let us see the first extreme point 0, 0. And the Hessian matrix was 6x… First, the Hessian matrix was this; the first principle minor was 6x plus 6. At 0, 0 point, the value is 0 here. The principle minor for the second… The second principle minor at 0, 0 point – it would be 6, 6. That is why the value of this determinant would be is equal to 36 – plus 36. And we see the first principle minor is positive; second principle minor – positive. That is why we can say the corresponding Hessian matrix is positive definite. And we conclude that, this is relative minimum. 0, 0 is the relative minimum. And the corresponding functional value is 24.

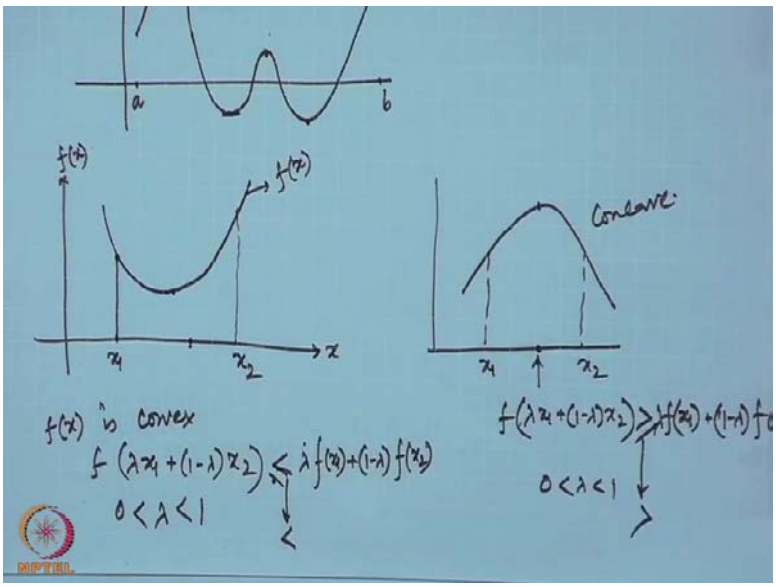
Go for the next extreme point; that is, 0 comma minus 2 by 3. Again we will go back to the Hessian matrix. 0, minus 2 by 3 – the first Hessian matrix first principle minor is 0; first principle minor is 6 for us; and the second principle minor will give us some value. And this is the value for us; that is, minus 36. That is why we cannot conclude anything from here, because the first principle minor will give us positive, second principle minor give us negative.

But, as I said, for the positive definiteness, all principle minors starting from the first – it would all be positive; and for the negative definite starting from the first principle minor – it would be alternate sign; but it will start from the negative sign. Since it is starting from the positive sign, first one is positive; next one is negative. That is why we

conclude that, the corresponding Hessian matrix is indefinite; and the corresponding extreme point is the saddle point for us; that is, saddle point means neither maximum nor minimum for the corresponding function. And this is the functional value – 76 by 9 at that point.

Similarly, the other extreme – third extreme point. Third extreme point is minus 2, 0. First hessian matrix – minus 12; second is again minus 72. Both are negative. Again we cannot conclude anything. And the corresponding Hessian matrix is indefinite. Again this point is a saddle point. And go for the last one – minus 2, minus 2 by 3. If you just see the first principle minor, it gives me the value – the negative value – minus 6; and the second principle minor gives the value, that is, positive value – plus 36. And since it is in the alternate sign starting from the negative sign only, that is why we can conclude that, corresponding Hessian matrix is negative definite. And the corresponding extreme point is the relative maximum point for us. That is why this is the corresponding value.

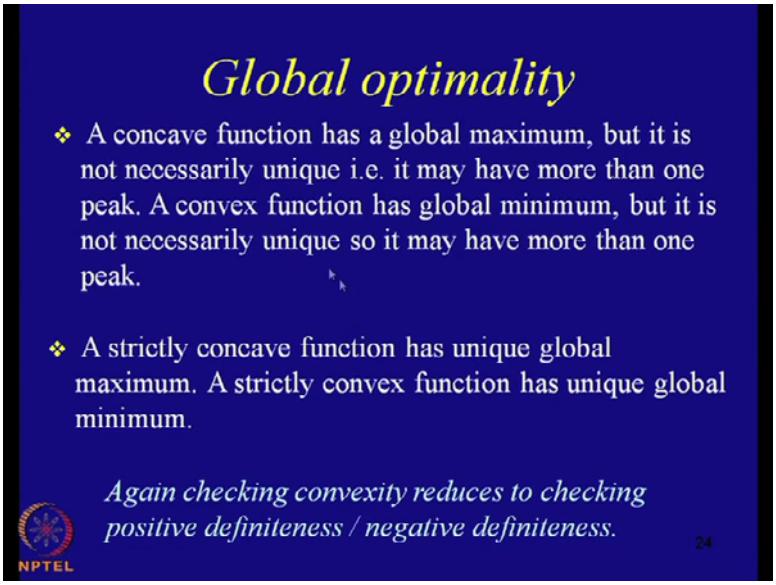
(Refer Slide Time: 41:48)



Now, till now, whatever we have learnt for function of several variables, the necessary condition gives us the possible extreme points; and sufficient condition tells us whether the extreme points we achieve. These are the relative minimum or relative maximum or the saddle point. But, nothing we have said about the global optimality. Now, that is why we need to say something about the global optimality for this function. For example, this is the function for us. Now, whatever conditions we got, we will get this point as a

relative maximum; this point as a relative minimum; this point as a relative maximum; this point as a relative minimum. But, if I just look at the graph, you see this will offer us within this range. If this is the function is defined from a to b; then within this range, this is the global maximum. That is why whatever necessary and sufficient condition we have learnt; that could not tell us anything about the global optimality. We need to know something more about this when we can say the extreme point is a global maxima or global minima. That is why we are coming to the next for this.

(Refer Slide Time: 42:39)



And it has been said that, a concave function has a global maximum; and a convex function has global minimum. Now, what is concave function? What is convex function? That we need to learn again. And we will see that, convex and concave funtion again will be related with the positive definiteness and the negative definiteness and rather the semidefiniteness as well. But, here just I want to mention that, a concave function… We are considering a function, which involves n number of variables.

And, not only that; the problem is unconstrained optimization. We will say if the objective function that which we are trying to optimize; if the corresponding function is a concave function and whatever maximum we are achieving, that is, a local maximum we are achieving; that is the global maximum for us. And if we see the function is global minimum; if we see the function is convex; then whatever minimum point we have achieved through the necessary and sufficient condition – the local, rather the relative;

this is the global minimum as well. But, one thing we should mention here. Let us see the pattern of the concave function and the convex function.

Now, this function is a convex function and this is a concave function for us. Now, how we can define the convex function and concave function? This is the convex function for us. If I consider a point here say x 1 point; this is corresponding f x 1; this is the x; this is the f x. We are considering of single variable only. This is the f function for us. We will say this f function is convex; f x is convex. If we take two points here: x 1 and x 2; and if we see that, if we just take any point in between x 1 and x 2; say this is the point for us. This point can be written as lambda x 1 plus 1 minus lambda x 2; where, lambda is lying between 0 to 1. Then we can say, if we see the functional value at this point – any point in between x 1 and x 2, is lesser than equal to lambda into f x 1 plus 1 minus lambda into f x 2; then corresponding function is the convex function.

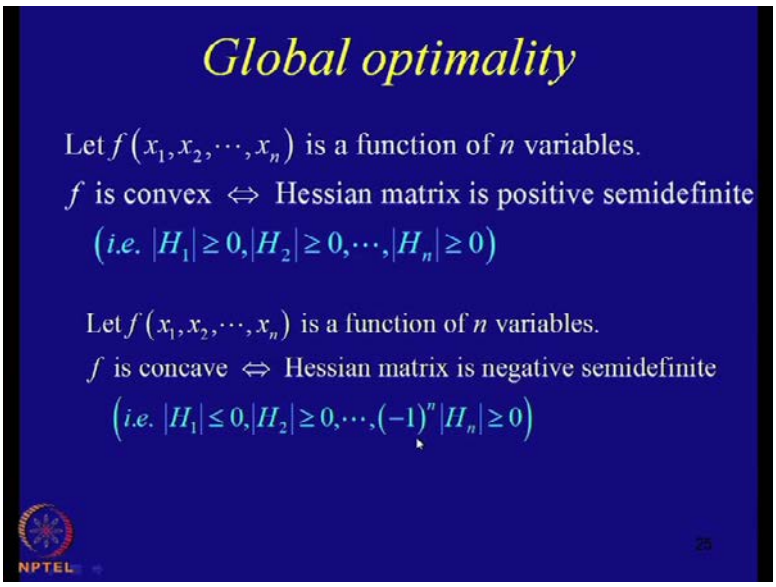
And, the reverse case; if we consider a point x 1 here and if we consider another point x 2 here; and in between if I consider any point here; then if we see at this point, that is, lambda x 1 plus 1 minus lambda x 2; for lambda in between 0 to 1, if we put lambda equal to 0; we will get the point x 2; if we put lambda is equal to 1, we will get the point x 1. And if we want to get any point in between x 1, x 2; we will just vary the value for lambda from 0 to 1.

And, if we see this is greater than equal to lambda into f x 1 plus 1 minus lambda into f x 2; then the corresponding function is the concave function. That is why if we see the objective function is concave function; then the corresponding min-maxima; we will get the global maximum. And if we see the the function is convex; then we will get the global minimum. But, here one thing we should point it out. This global maxima or minima may not be unique one.

For uniqueness, we should have the next condition. Condition is that, we shloud have a strict convex function for global minimum; and we should have a strict concave function for global maximum. Strict means what? Whatever inequality we have achieved here; that equality signs should not be there. That is why we will get only one point here; there should not be any flat area, where we will have several points, where the equality holds. That is why we want to say that, we should have only one point here. And strict convex means this inequality would be less than; strict concave means this would be greater than

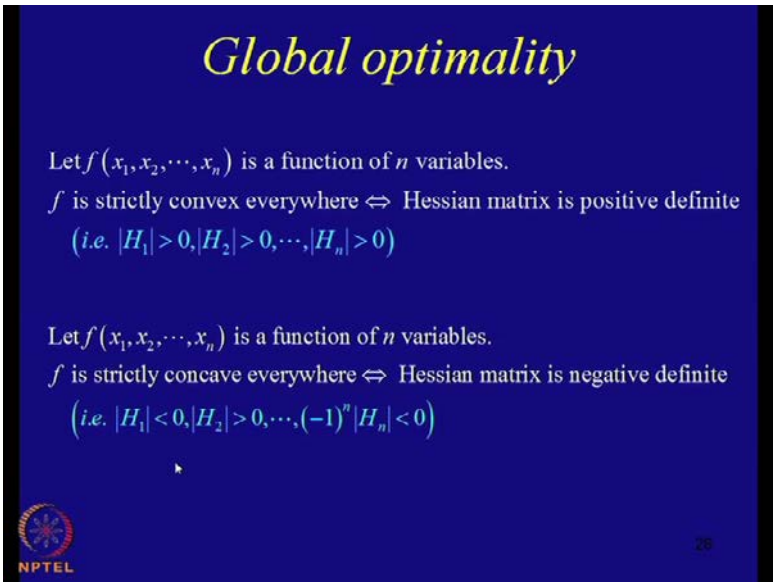
only. And for strict concave, we will have the unique global maximum; and for strict convex, we will have the unique global minimum. That is why looking at the functional pattern – whether the function is concave or convex, we can conclude above; we can say something about the global optimality of the unconstrained optimization problem. But, again the checking of the convexity reduces to the checking of positive definiteness and the negative definiteness of the matrix.

(Refer Slide Time: 47:39)



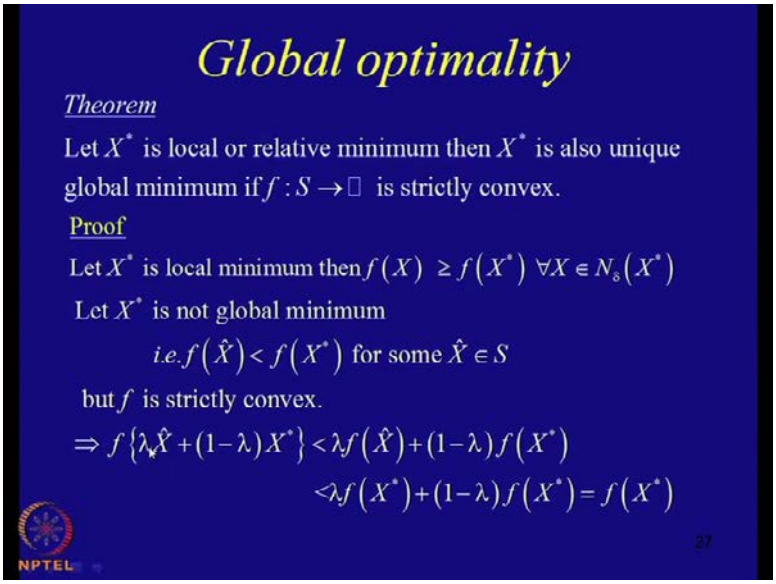
Because we know this property as well; if f is a function of n number of variables; then f is convex. If we form the Hessian matrix; if this is positive semidefinite; then the corresponding f function would be convex function. Similarly, if the Hessian matrix is negative semidefinite, then the corresponding function will be the concave function. And for checking the positive definiteness, we know the principle minors; will be all positive. The determinant values will be all positive; that semidefiniteness. That is why greater than equal to sign is there. And for negative semidefinite, we should have the alternate sign of the principle minors starting from negative value.

(Refer Slide Time: 48:26)



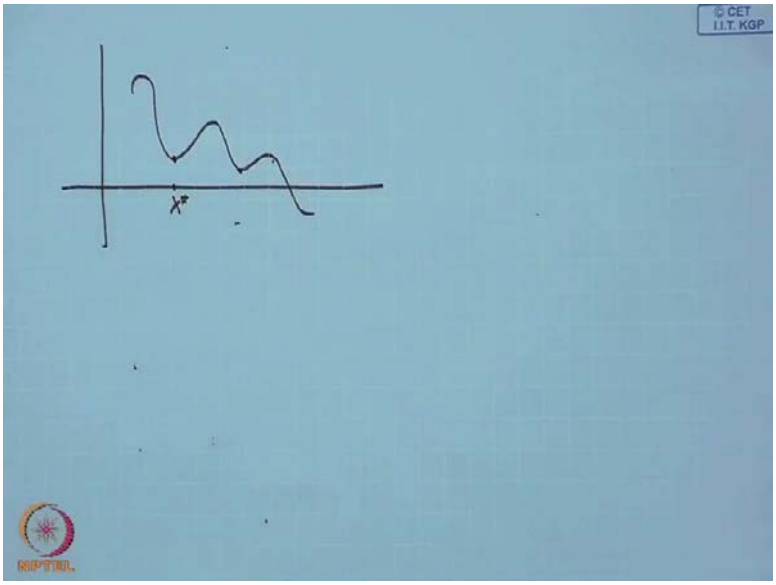
But, we know the other thing as well. If f is strictly convex everywhere within the given domain, where the function is defined; then we will see the corresponding Hessian matrix will be positive definite. And similarly, for the… If we see the Hessian matrix is negative definite, then the corresponding f must be strictly concave everywhere. That is why looking at the Hessian matrix as well, we can check whether the function is strictly concave or strictly convex. That is why the Hessian matrix not only gives us the possible extreme points; can check whether these are local maxima or local minima.

(Refer Slide Time: 49:07)



But, also it gives us the next level, that is, whether the local minima, local maxima is again the global maxima or not. Here just one of the property is given here. Let X star is a local or relative minimum; then X star is a unique global minima if there is a function from S to R; that is a real line. That is a… It is a… If function of several variable, then it should be S from to R n. If f is strictly convex, then the relative minimum would be the unique global minimum. Let us prove it. Let us consider X star is the local minimum for us. Then we know if X star is the local minimum; then in the neighbourhood…

(Refer Slide Time: 49:51)

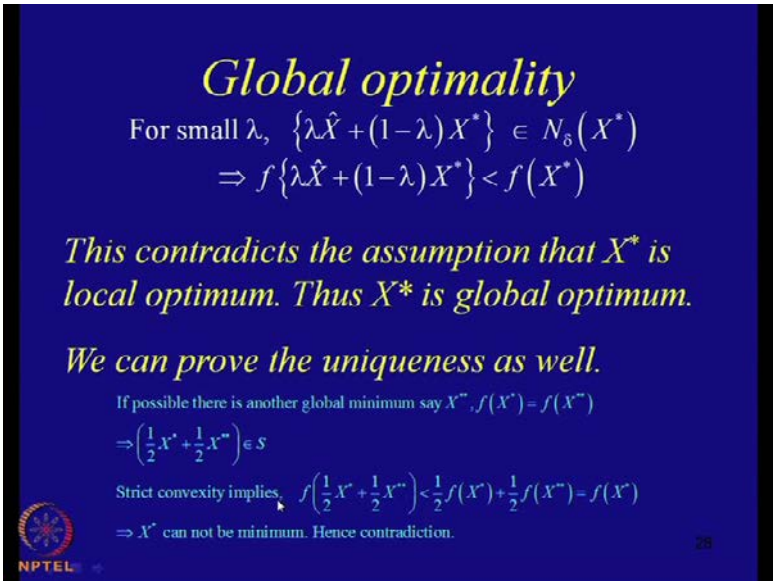


For example, this is the function for us. Now, here this is the local minima for us; this is the local minima for us. How we will check this is the local minima? We will just look at if this is X star for us; we will just see the neighbourhood of X star. If we see in the neighbourhood, this relation holds that f X greater than equal to f X star; then we will say the corresponding minima is the local minima for us; where, X belongs to the delta neighbourhood of X star.

Now, X star we are… We have to prove that, X star is also the global minima. Now, let us assume that, X star is not the global minima. We are having another global minima X hat within the domain of the function. That is S for us. Then if X hat is another global minima for us; then we can say that f X hat must be lesser than f X star. Now, within this condition, let us see further what is happening; and we have assumed that, f is strictly convex. If f is strictly convex through the condition f of lambda X hat plus 1 minus

lambda x star; this must be lesser than; it is not lesser than equal to, because this is strictly convex; lambda of f X hat plus 1 minus lambda f X star; where, this point is a point in between X hat and X star. We have considered any point. If this holds; with this condition – f X hat is lesser than f X star, we can further simplify it and we are achieving to the condition that, f lambda X hat and 1 minus lambda X star is less than f X star.

(Refer Slide Time: 51:22)



This is the condition we are getting it. But, if we consider lambda as very small amount, so that the lambda X hat plus 1 minus lambda X star would be in the delta neighbourhood of X star. If we consider lambda very small; that would be very near to X star; that will be certainly in the neighbourhood – delta neighbourhood of X star; then what we see, in the delta neighbourhood of X star, we are getting one point, where functional value is lesser than the functional value at the extreme point.

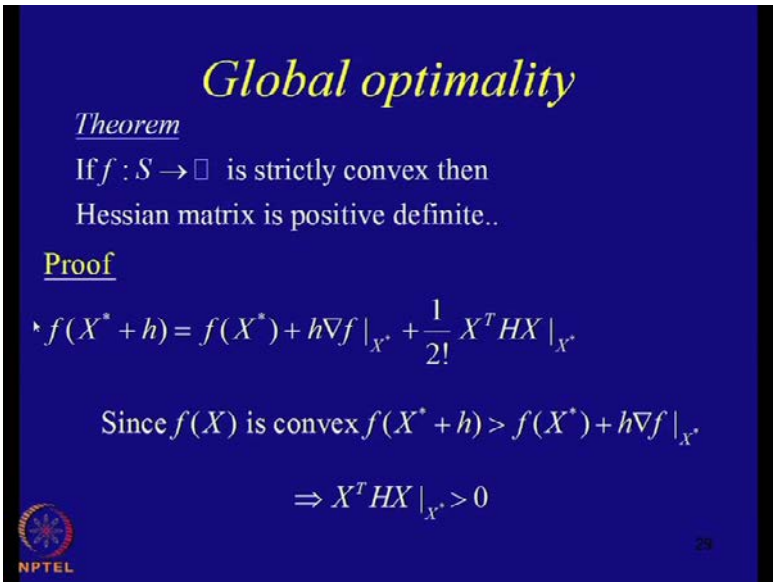
That is why whatever conclusion we made that, X star is the local optima; then it is failing. That is why whatever assumption we made before, we have considered that, X star is not global minima, but f is strictly convex. That assumption is totally wrong. That is why we can conclude that, when f is strictly convex; whatever local minima we are getting both; that is the global minimum as well. That is why we are taking the conclusion that, X star is the global minima.

Now, with the function is strictly convex; if the function is not strictly convex, if we are having the function; if we see the function is not strictly convex; function is convex only;

that is why the lesser than equal to sign is involved there; then we will see that, we will have several optima together we can have. But, if we see the function is strictly convex; then in that case, we would not get several global optima; we will get the unique global optima. That also we can prove it. We are considering one of that global optima X double star, which is different from X star.

And, if we take any point in between; that is, considering lambda is equal to half; that is again within the doamin of the function. And we see in that point, the functional value is lesser than f X star. That is not acceptable to us, because the function is strictly convex. That is why the less than sign is there. And X star we have assumed as the global minima; we should not have any other point for the functional value is lesser than that. That is why whatever we have assumed that, we are having another global minima with X double star; that is, that assumption was wrong. That is why x star cannot be… That is why we can conclude that, we cannot have more than one X star when the function is strictly convex. That we concluding here that, when f is the strictly convex function, then we can conclude that, the corresponding local minima is the global minima.

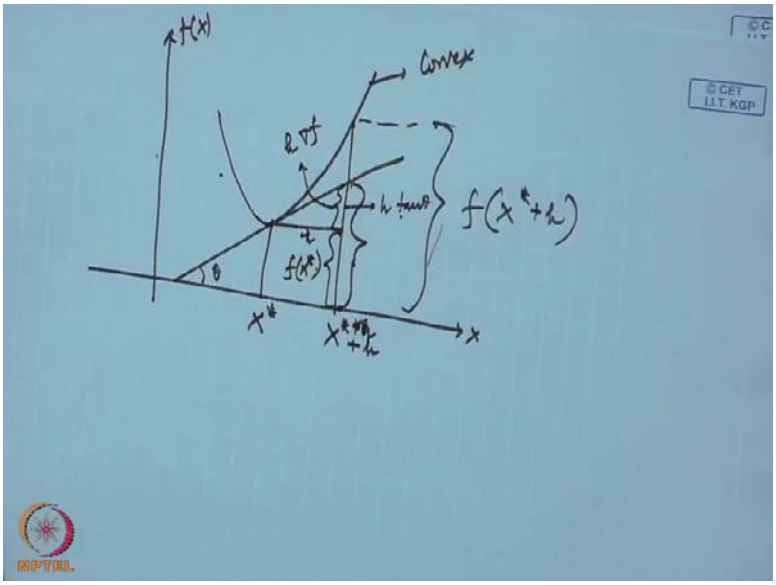
(Refer Slide Time: 54:05)



Similarly, when function f is the strictly concave function, the corresponding maxima – local maxima would be the global maxima. Now, for the next, again we can prove that, for strictly convex function, the corresponding Hessian matrix is positive definite. That is why whatever conclusion we made before that, if the function is… We should have a

relative minima if the Hessian matrix is positive definite. Here we are making further assumption that, f is strictly convex. Whatever local or relative minima we will get; that would be the global minima as well. That we can prove very easily. Just look at the function. This is the expression we are getting from the Taylor's series. And since the function is the convex function, we can say this one as well. How we can say it?

(Refer Slide Time: 55:03)



Let me just go through the graph once. Look at this graph; this is the convex function. That is why we are taking a tangent here; where, the optima exists at X star. All right? We are considering another point, that is, X star plus h; X star plus h here. This is the X; this is the f X. Then certainly function is the convex function for us. Not only that; we have considered only the function is strict convex function. Now, at X star plus h, this is the functional value. That is why the whole value upto this would be is equal to f of X star plus h.

But, upto this, if I see this part; this value would be is equal to f X star. And what about this value? This value would be is equal to h del f, because if we consider the tangent here with the angle theta; if this is h; then it would be is equal to h tan theta; tan theta would be nothing but the slope of this tangent. That would be the first order partial derivative. That is the gradient function we are considering. That is why what we see that, f X star is star plus h is greater than summation of f X star plus h of delta f at X star.

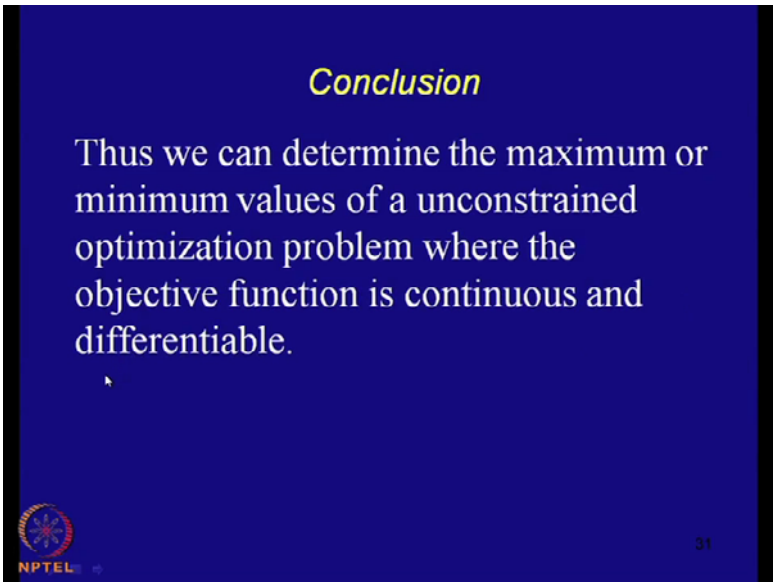
And, this is the reverse for the case of concave function. This is not the same for concave function. For the concave function, it will just reverse f X star plus h must be lesser than f X star plus h. That is the gradient of f. That is why if I just combine these two conditions together, what we get? We get f x star plus h minus f X star is greater than this value. That is why what we conclude that, if the function is strictly convex; then the corresponding Hessian matrix is greater than 0. That is the conclusion for us. That is why we can say the Hessian matrix is positive. If it is positive definite, then the function is strictly convex.

(Refer Slide Time: 57:06)

Now, whatever results we made out till now, we are summarizing all together right now. A convex function has a local minimum if at extreme point Hessian matrix is positive semidefinite. A strict convex function has a global minimum if the Hessian matrix is positive definite everywhere. And a concave function has a local maximum if at extreme point, the Hessian matrix is negative semidefinite. And a strict concave function has the global maximum if the Hessian matrix is negative definite everywhere.

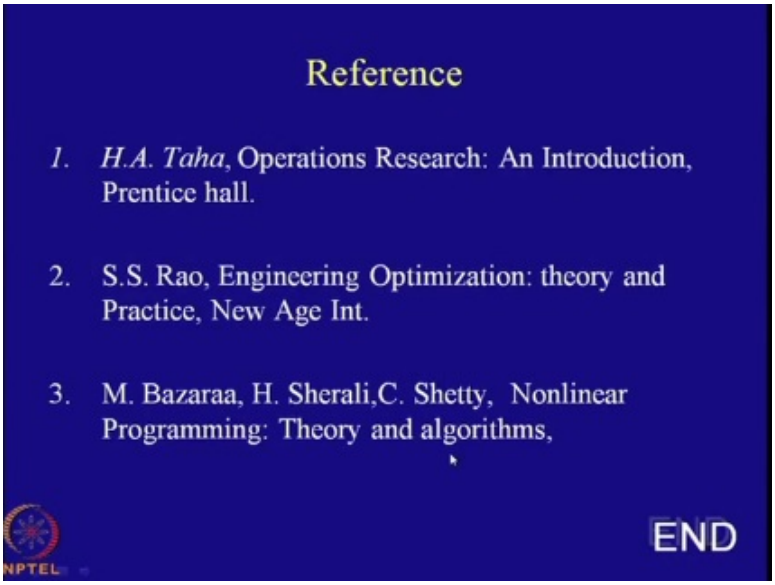
Whatever conclusion we just made up to this; this is very much related with that well-known fact, that is, a convex programming in optimization technique. We will just reconfirm all these results together in the convex programming further when we will consider some more complicated situation, that is, the constraint optimization problem with several variables in the next.

(Refer Slide Time: 58:12)



And, thus we are concluding our lecture with this. That we can determine maxima, minima for the unconstrained optimization problem, where the objective function is continuous and differentiable. And we have said even not only that; we are not getting not only the local optimum; we are also getting the global optimum. With further conditions, we are having with the function – objective function involved in the function in the problem.

(Refer Slide Time: 58:41)



And these are the references can be referred for further learning of this topic.