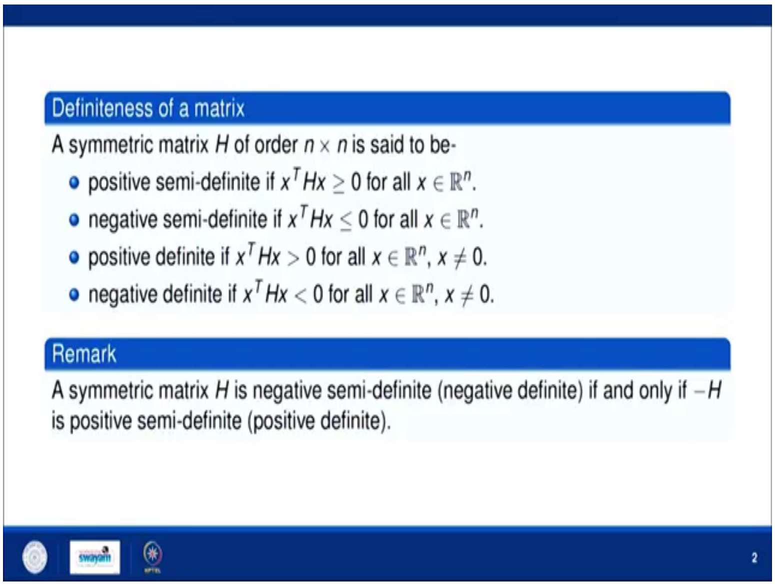
Essential Mathematics for Machine Learning
Prof. S. K. Gupta
Department of Mathematics
Indian Institute of Technology, Roorkee

Lecture – 25
Properties of Convex Functions - II

Hello friends. Welcome to lecture series on Essential Mathematics for Machine Learning. In the last lecture, we have seen that what convex sets convex functions are and some of their properties. Other Properties of Convex Function we will see in this lecture.

(Refer Slide Time: 00:44)



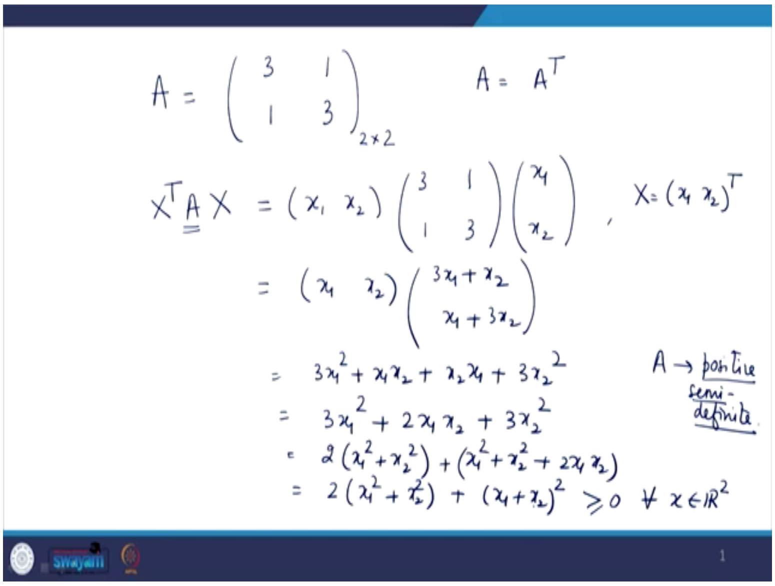
So, let us try to understand what are the other properties of convex functions. So, before stating the other property of convex functions. This is very important property of convex function. We first define that how we. What do you mean by definiteness of a matrix? When we can say that a given square matrix given square symmetric matrix in fact, is positive semi-definite, positive definite, negative semi- definite or negative definite.

So, here is the definition. A symmetric matrix H of order n cross n is said to be positive semi-definite, if x transpose H x is greater than equal to 0 for all x belongs to R n. If this is if this sign is reversed, that is less than equal to 0 for every x belongs to R n then this is called negative semi-definite.

Positive definite if x transpose H x is strictly greater than 0 for all x belongs to R n and x not equal to 0 ok. This whole does stick sense, and for every x belongs to R n and x not equal to 0. And if again this inequality reversed, that is it is less than 0 for every x belongs to R n x not equal to 0, then it is called negative definite.

So, from this definition from these four points, one thing is clear. That that if H is negative semi-definite then minus H is positive semi-definite, because there is a sign of only this inequality is reversed. And similarly, if a matrix is positive definite then minus H is negative definite. So, here is a remark that, a symmetric matrix H is negative semi-definite or negative definite if and only if minus H is positive semi-definite or positive definite.

(Refer Slide Time: 02:43)



So, let us discuss these things by an example. Suppose, we are having this example. Suppose we have matrix A which is given by suppose 3 1 1 3. So, it is a 2 cross 2 matrix and of course, it is symmetric, because A equal to A transpose ok. Now, to check whether this matrix is positive semi-definite, negative semi-definite or positive definite or negative definite, what we find? We find X transpose A X ok.

What is X transpose? See this is 2 cross 2 matrix. So, it will involve two component x 1 and x 2. A is 3 1 1 3 and this is x 1 x 2. Now, when you multiply these. So, this is x 1 x 2 this is. So, here x is what? Here x is x 1 x 2 transpose. So, this row this column gives 3 x 1 plus x 2. It is x 1 plus 3 x 2.

And when you multiply this row with this column, this is simply 3 x 1 square plus x 1 x 2 plus x 2 x 1 plus 3 x 2 square. So, this is nothing, but 3 x 1 square plus 2 x 1 x 2 plus 3 x 2 square.

And this is further can be written as twice of x 1 square plus x 2 square plus x 1 square plus x 2 square plus 2 x 1 x 2 ok.

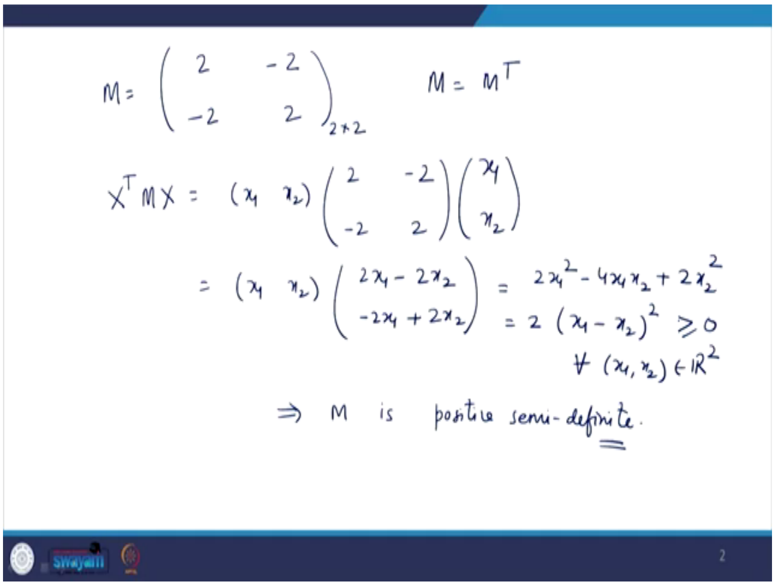
Because it is 3 x 1 square plus 3 x 2 square. So, this is further written as 2 x 1 square plus twice of x 2 square plus this is x 1 plus x 2 whole square. So, now, from this we can easily conclude that this is always greater than and equal to 0 for all x belongs to R 2. Here x is in R 2 ok.

And this is 0 only when both of 0; that means, x 1 x 2 both are 0 ok. So, for every x belongs to R 2 this quantity is always greater than equal to 0. So, we can say that, this matrix A here whatever A matrix we have chosen this matrix A is nothing, but positive semi-definite matrix ok.

So, we have concluded that this matrix A is positive semi-definite ok. Now in fact, this matrix is positive definite. Why I am saying this? Because it is 0 only when x equal to 0, x equal to 0 means x 1 x 2 both are 0. So, if you go to a definition then if you go to this definition that, x transpose H x is greater than 0 for all x belongs to R n x not equal to 0. Whenever x is not equal to 0 x is not equal to 0 means x 1 x 2 are not equal to 0 0.

So, if x 1 x 2 are not equal to 0 0, this is always strictly greater than 0. So, this matrix is in fact, positive definite ok.

(Refer Slide Time: 06:15)



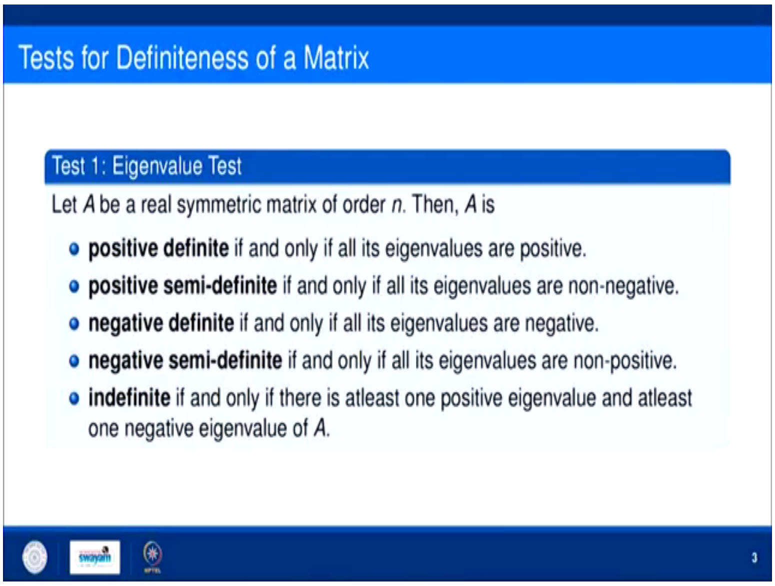
So, now consider another example. Say consider this example 2 minus 2 minus 2 2. So, suppose this is M. Again this matrix is of order 2 cross 2, it is a symmetric matrix, because M equal to M transpose. Now, again you consider X transpose M X which is x 1 x 2 m is 2 minus 2 minus 2 2 and this is x 1 x 2. So, this is nothing, but x 1 x 2 and this is 2 x 1 minus 2 x 2, this is minus 2 x 1 plus 2 x 2 by the matrix multiplication.

And it is 2 x 1 square. This x 2 multiply by this row this column that is 2 x 1 x 2 2 x 1 x 2 that is minus 4 x 1 x 2 and plus 2 x 2 square. So, this is nothing, but 2 of x 1 minus x 2 whole square and which is always greater than equal to 0 for every x 1 x 2 belongs to R 2. So, what we have concluded from here? We have concluded that this matrix M is nothing, but positive semi-definite. So, this implies M is positive semi-definite.

So, here there are some nonzero x also for this particular example there are some nonzero x also where it is 0 ok. For example, 1 1 2 2 whenever x 1 equal to x 2. So, these are not 0, but it is still this value comes out to be 0. So, it is not positive definite, but it is positive semi-definite ok. So, in this way; in this way we have seen that these are the few matrices of square matrices symmetric matrix which are positive definite or semi-definite. In the same way we can proceed for negative semi-definite.

So, now the question arises, that if a matrix is given to you these examples I have taken only 2 cross 2 or 3 cross 3 we can see, but if it is a n cross n general matrix then, how we can check whether a given matrix is positive semi-definite or negative semi-definite? How we can see? So, what are the different test? So, we have different test the test one is eigenvalue test.

(Refer Slide Time: 08:48)



So, what this test? Test is. So, let A be a real symmetric matrix of order n. Then A is positive definite, if and only if all its eigenvalues are positive. So, you find the eigen values of a given matrix A given symmetric matrix A and if all eigenvalues are greater than 0, then the given matrix is positive definite.

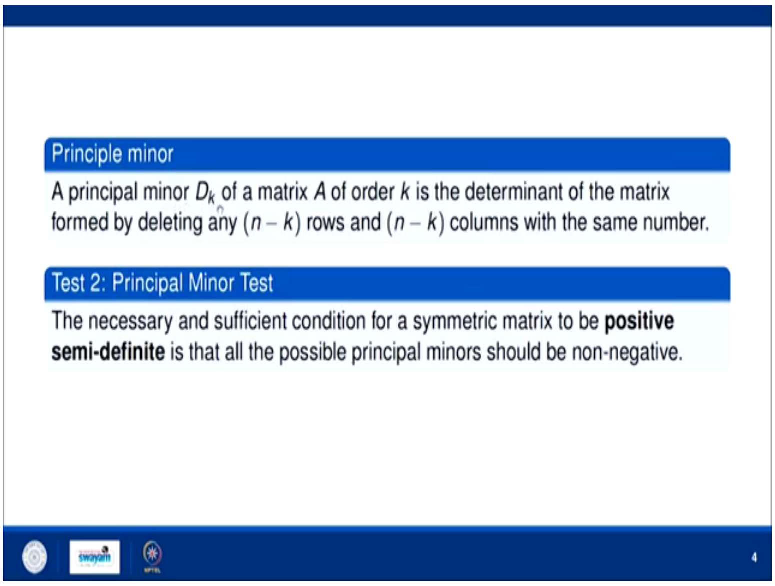
Semi-definite if and only if all its eigenvalues are greater than equal to 0. Negative definite if all eigenvalues are if and only if all its eigenvalues are negative less than 0. And negative semi-definite if and only if all its eigenvalues are less than or equal to 0. And indefinite if and only if there is at least one positive eigenvalue and at least one negative eigenvalue of A ok.

So, this is the first test ah. You find the eigenvalues of a given matrix given symmetric matrix A. And if all eigenvalues are come out to be strictly greater than 0 means positive definite, greater than equal to 0 that means positive semi-definite, less than 0 means negative definite and less than equal to 0 means negative semi-definite. This is the first test. Now, of course, it is a symmetric matrix. So, all its eigenvalues will be real, then only we can talk about positive or negative ok.

Of course, that will of course, holds, because it is a symmetric matrix. Now, again finding eigenvalues of a large matrix, say 10 cross 10 or a 5 cross 5 matrix is not an easy task. So, how can we proceed? How what is the simple test by which we can say that the given matrix is positive definite or negative semi-definite or whatever? So, we have another test that is test two which is called minor test.

So, let us see what. First of all what do you mean by minors? And then we will go for minor test.

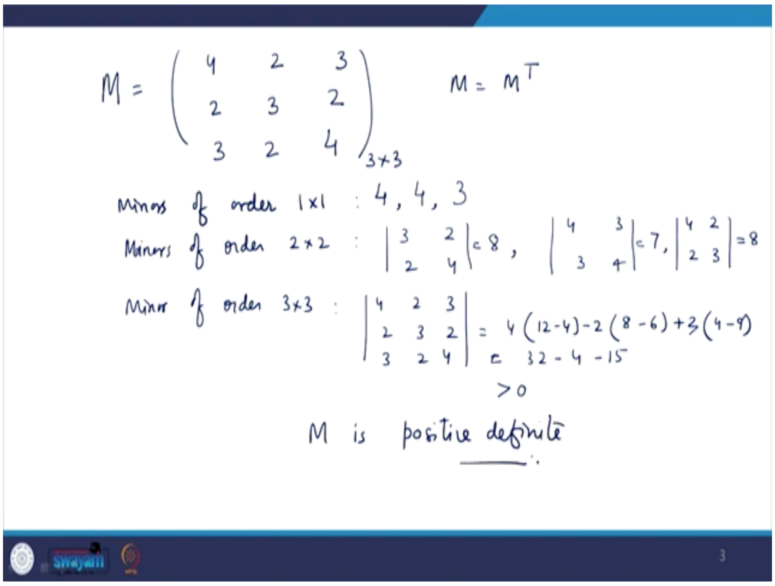
(Refer Slide Time: 10:55)



A principal minor D k of a matrix A of order k is a determinant of the matrix formed by deleting any n minus k rows and n minus k columns with the same number. So, what do you mean by this? We will discuss it by an example. So, and the first of all I have to read this principal minor test.

So, what this test is? Basically the necessary and sufficient condition for a symmetric matrix to be positive semi-definite is that all the possible principal minor should be non-negative, that it should be greater than equal to 0. And for positive definite should be strictly greater than 0. So, let us discuss it by an example. So, a things would be clear to you. What do you mean by principal minors and what are the different test?

(Refer Slide Time: 11:46)



Now, suppose you have this example. Let us take M as 4 2 3 2 3 2 3 2 4. So, first of all, is it symmetric? Yes, it is symmetric, because you can easily verify that M equal to M transpose a symmetric matrix. Now, to check whether this matrix.

What is the nature of this matrix? Is it positive definite, negative or neither? How we can check? So, the first of all either you can find out the eigenvalues of this matrix. So, by finding the eigenvalues you can say whether this matrix is positive definite or whatever or negative definite or semi-definite. The other way out is you find out the minors of this.

So, first of all we take we find out the its a 3 cross 3 matrix. So, first of all you find minors of order 1 cross 1. So, how you find minors of order 1 cross 1 ok? So, you see you delete second row and second column, you delete third row and third column. What is left with us? 4 and a determinant of 4 is 4 itself.

So, it will come out to be 4 ok. So, here k is 1 and n is 3. You see here if you see here, k is 1 and n is 3. So, 3 minus 1 means 2. So, you delete any two rows and any two columns with the same number ok. I mean the same column and the same rows ok.

If you are deleting second row second column third row third column, you are left with 4. So, 4 is the first minor. Now, if you delete first row first column first row first column sorry and second row second column, we left with 4 this 4. So, this is a second minor of 1 cross 1. Now, if you delete third row third column and first row first column. So, you left with 3.

So, this is three. So, basically minors of order 1 cross 1 are nothing, but diagonal elements. So, diagonal elements are 4 3 4 4 3 4. So, here all are greater than 0. Now, let us compute minors of order 2 cross 2. So, now, k will be 2 and n is 3 order is three. So, 3 minus 2 means 1. So, you will delete one row and the corresponding column. And whatever determinant you will obtain that is a determinant that is a minor of order 2 cross 2.

So, let us try to compute. So, first of all suppose you delete first row and first column. So, you will left with this. So, you left with 3 2 2 4 ok. Now, suppose you delete second row and second column. Now, you left with 4 3 3 4 second row and second column. So, you left with 4 3 3 4 ok.

Now, you delete third row and third column. So, you will left with 4 2 2 3 ok. Now, what are the determinant of this? The determinant of this is 12 minus 4 that is 8. What are determinant of this? The determinant of this is 16 minus 9 16 minus 9 is 7 the determinant of this is 12 minus 4 12 minus 4 is 8. So, it is 8. So, again here also the all the principal minors of 2 cross 2 are greater than 0 8 7 and 8.

Now, only one left that is minor of order 3 cross 3 which the which is the determinant of the matrix itself. Which is what 2 4 2 3 2 3 2 3 2 4. Now, what it is? It is 4 times you delete this row this column you find a determinant basically 12 minus 4 it is minus 2; that means, 8 minus

6. And that means, plus 3 times 4 minus 9 that is equal to that is 8 into 4 is 32 32 minus it is 4 it is minus 5 minus 15.

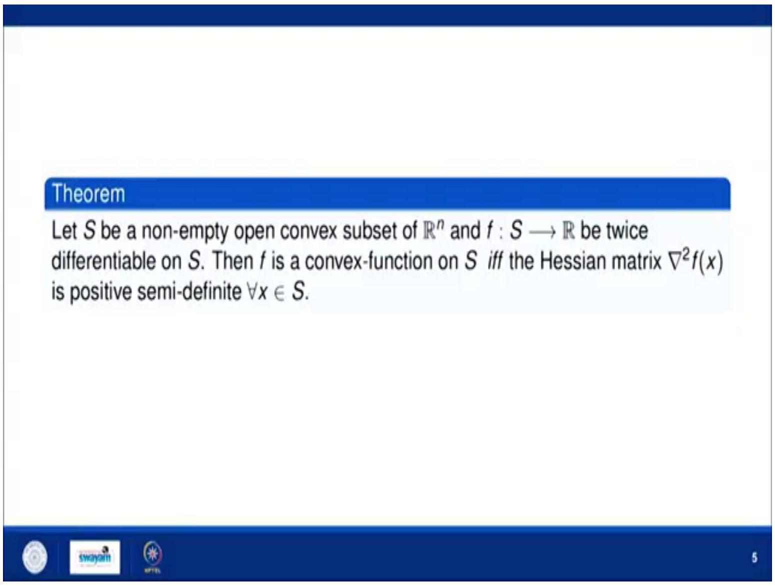
So, this value is in positive of course, and 32 minus 19 is positive. So, we can say that all the minors 1 cross 1 all the possible minors, 2 cross 2 all the possible minors, 3 cross 3 only one possible minor in this case and all are strictly greater than 0; that means, this matrix M is positive definite. So, this implies M is positive definite ok.

So, this is what given in this necessary sufficient condition, that you find out all the possible principal minors. And if all the possible principal minors are strictly greater than 0; that means, matrix is positive definite. If it is greater than equal to 0; that means, positive semi-definite. And for negative definite this should be all minor should be not all.

See, for negative definite or negative semi-definite you first convert into you first multiply that matrix with minus 1. And then test for positive definite or positive semi-definite, because otherwise sign will be alternating. If you see for say.

If you are seeing for negative semi-definite then the first order minors of order 1 cross 1 are less than 0, minors of order 2 cross 2 are greater than 0, minors of order 3 cross 3 are less than 0 and so on ok. Or the other way out is you convert you multiply the matrix by minus 1 and check for positive semi definite or positive definite. That is quite easy ok.

(Refer Slide Time: 18:13)



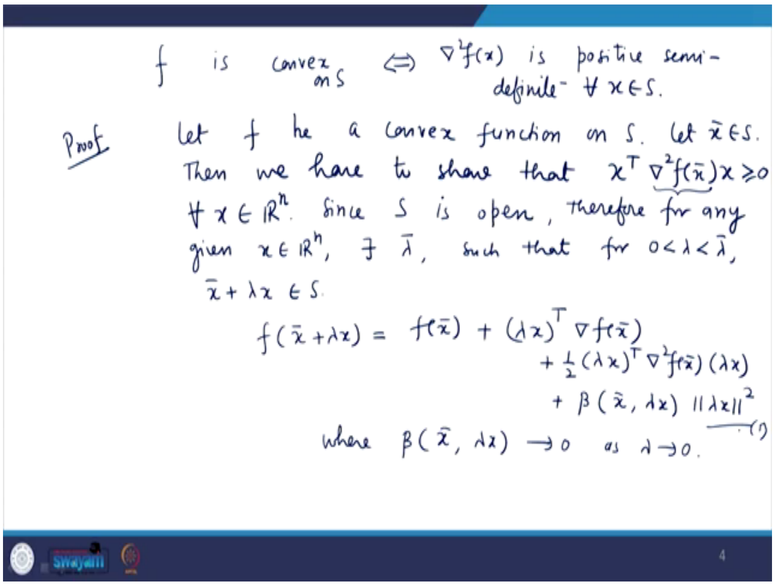
So, now let us come to the property of convex function. Now, we have already seen the basic definition of convex that when we can see the function is convex. We have also seen that if a function is once differentiable on a open convex set S, then any inequality hold that also we have discussed ok. But your function is twice differentiable suppose and its convex then, what is the important property?.

Important property is given in this theorem. What is that? Let S be a non-empty open convex subset of R n and f is a function from S to R be twice differentiable on S. Then f is a convex-function on S if and only if the Hessian matrix which is del square of f x is positive semi-definite for every x belongs to S.

So, if this matrix is positive semi-definite. The function is always convex or if function is convex then this matter is always positive semi-definite. So, this is very important theorem for checking whether a given function is convex or not ok. So, let us discuss the proof of this ok.

So, those who are interested can see the can like see the proof. Otherwise in terms of machine learning, if you know the result also then it is also sufficient ok. But here I am discussing the proof for those who are interested.

(Refer Slide Time: 19:46)



So, we have to show that f is convex if and only if del square of f x is positive semi- definite for all x belongs to S. Whereas S is an open convex subset of R n. So, let us discuss the proof.

So, first let f be a convex f is convex on S f is a f be a convex function on S. So, if you want to show that this matrix is this Hessian matrix is positive semi-definite, what we have to show? We have to show that taking any let x bar belongs to S if we take a point x bar belongs to S, then we have to show that x bar belongs to S then we have to show that x transpose gradient square of f x bar into x is greater than equal to 0 for all x belongs to R n.

See, we have already seen that when a matrix is set to be positive semi-definite. Here instead of M we have del square of f x bar. So, x transpose M x transpose M x should be greater than or equal to 0 for every x belongs to R n. If this holds then we can say that the given matrix is positive semi-definite.

So, since x bar is an arbitrary point on S. So, if we have shown this for x bar; that means, we have shown this for every x bar belongs to S ok. So, let us try to prove this inequality. So, now, since S is open it is given in the statement itself. If you see the statement it is given to us that S is an open convex subset of R n.

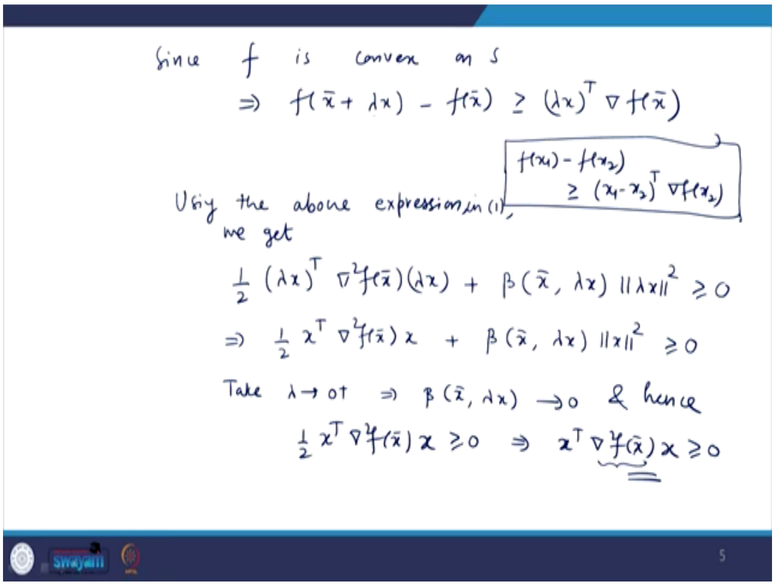
So, since S is open therefore, for any given x belongs to R n, there exist some lambda bar, such that for 0 less than lambda less than lambda bar, x bar plus lambda x will belongs to S ok. No matter how small lambda bar is it hardly matters, but there will exist some lambda bar, such that for any x belongs to R n x bar is already in s. So, this x bar plus lambda s will be lambda x will be in S.

Because S is open ok. One can one can very easily visualize the geometry of this. Now, let us use the second order differentiability of f at x bar. So, since f at x bar is second order differentiable it is given to us. So, what we have it is x bar plus lambda x will be equals to f x bar plus lambda x transpose gradient of f x bar plus 1 by 2 lambda x transpose gradient square of f x bar into lambda x plus beta it is x bar lambda x norm of lambda x whole square.

So, this we have already discussed in the last lecture that what do you mean by twice differentiability of f at x bar. It is some see it is x 1 it is x 2. If you are talking about differentiability of f at x 2; that means, f x 1 equal to f x 2 plus x 1 minus x 2 which is lambda x here gradient of f x bar plus half of x 1 minus x 2 transpose gradient square f x bar lambda x plus beta times x 1 minus x 2 into x 2 and norm of x 2 minus x 1 whole square or x 1 minus x 2 whole square. And where beta x bar lambda x will tends to 0 as lambda tending to 0.

So, this is by the definition of twice differentiability of f at x bar. Now, let us try to simplify this expression. Now, since function is still we have not used the convexity property of the function f. So, let us try to use it.

(Refer Slide Time: 24:53)



So, since f is convex on S. So, this implies f of x bar plus lambda x minus f x bar is greater than equals to lambda x transpose gradient of f x bar. This is by that property which we have discussed in the last lecture that if function is once differentiable then this property holds, then we can say that f of x 1 minus f of x 2 is greater than equals to x 1 minus x 2 whole transpose gradient of f x 2. So, this is by this property if function is convex. So, here x 1 is this x 2 is this.

So, it is x 1 minus x 2 whole transpose gradient of f x 2. Now, if we club this expression with a previous expression with this expression. So, this minus this is greater than equal to 0 ok. So, this means this expression is greater than equal to 0. So, if you club these two expression what we get finally. Now, using the above expression above expression. Suppose this expression is 1 this expression is 1 ok. Suppose this the above 1 suppose this expression is 1.

So, using this expression in 1, we get 1 by 2 lambda x transpose gradient square f x bar lambda x plus beta x bar lambda x norm of lambda x whole square is greater than or equal to 0. Now, lambda is a scalar. So, can be taken out. So, lambda into lambda square and lambda can be also taken out from the norm. So, lambda square can go can be removed.

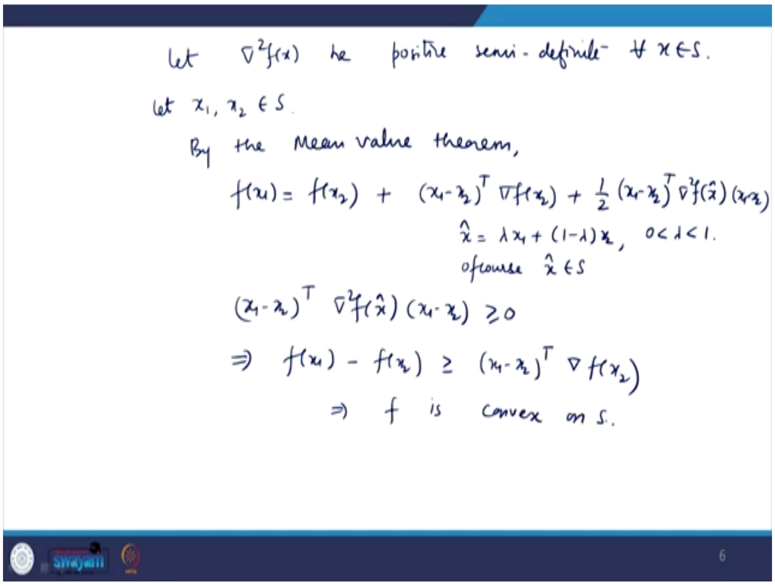
So, this implies 1 by 2 x transpose gradient square of f x bar x plus beta x bar lambda x norm of x square is greater than or equal to 0. Now, take lambda tends to 0 plus.

So, if lambda tends to 0 plus this will tends to 0, as per definition of twice differentiability of f. So, this will be greater than or equal to 0. So, take lambda tending to 0 plus, this implies beta x bar lambda x will tends to 0 and hence.

And hence, x transpose gradient square of f x bar x 1 by 2 will greater than or equal to 0. This further implies x transpose gradient square f x bar into x is greater than or equal to 0. So, we have shown that for any x belongs to R n this expression is greater than equal to 0; that means, this matrix. That means this matrix is positive semi-definite at x bar. And x bar is a variable point that we have already taken as a variable point you change x bar so; that means, it is positive semi definite for every x on S.

So, one part of the proof is over. That if f is convex then Hessian matrix of f is positive semi-definite for every x belongs to S. Now, let us try to prove the converse part. Let us take f this is positive semi-definite and we will try to show that f is convex on S. So, let us try to prove the converse part.

(Refer Slide Time: 28:55)



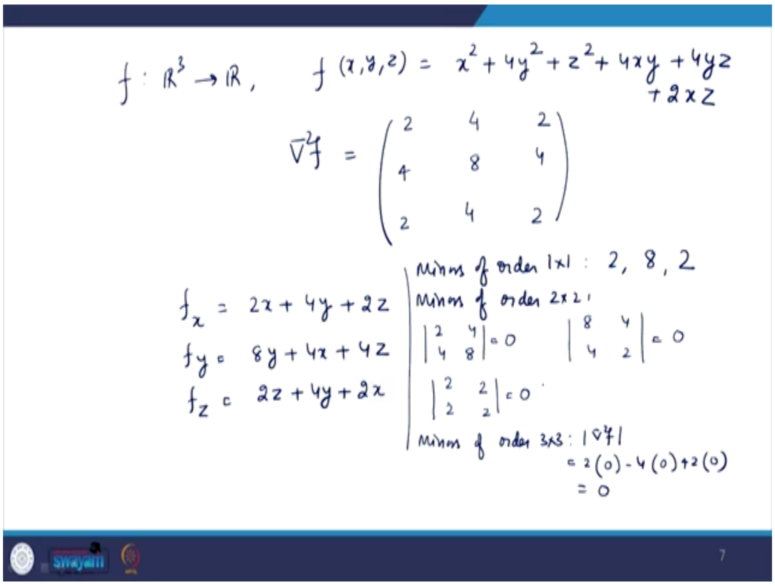
Now, let gradient square f x is be positive semi definite for all x belongs to S. So, we have to show that f is convex. So, let x 1 comma x 2 belongs to S ok. So, by the mean value theorem, what we will be having? f of x 1 minus f of x 2 f of x 1 is equal to f of x 2 plus x 1 minus x 2 whole transpose gradient of f x 2 plus 1 by 2 x 1 minus x 2 whole transpose gradient square of f x cap and x 1 minus x 2 where x cap is some convex near combination of x 1 and x 2 ok.

Now, since this is positive semi-definite and x 1 minus x 2 is any vector in R n. So, this quantity of course, this x cap belongs to S, because S is convex and x 1 x 2 is in S. So, x cap is also in S. And this is positive semi-definite and x 1 x 2 is any vector in R n. So, this part is greater than or equal to 0 ok. So, we can say that x 1 minus x 2 whole transpose gradient square of f x cap x 1 minus x 2 is greater than or equal to 0,since del square f is positive semi definite.

So, this implies if we apply this inequality in this expression in this expression finally, what we get f of x 1 minus f of x 2 is greater than equal to x 1 minus x 2 whole transpose gradient of f x 2.

And which by the property of convex function implies f is convex. So, we can say that we have proved this theorem that f is convex on S if and only if Hessian matrix of f at x is positive semi-definite for every x on S ok. Now, let us discuss one example based on this suppose you suppose you are having this example.

(Refer Slide Time: 31:37)



Suppose f is from R 3 to R and f x y z is given by this plus 2 x z. Now, we have to see whether this function is convex or not. How we will see? We will find the Hessian matrix of this f. And using the test either eigenvalue test or minor test, we will try to see whether the Hessian matrix is positive semi-definite or not.

If it is positive semi-definite, then we can easily say that this function f is convex by the twice differentiability property of f convex property of f. So, let us find Hessian matrix of f. So, Hessian matrix we already know how to find. It is del square f by del x square which is 2 x ok. It is second derivative.

So, first you find f x. What is f x? f x is 2 x plus 4 y plus 2 z. What is f y? f y is 8 y plus 4 x plus 4 z. What is f z? f z is 2 z plus 4 y plus 2 x. Now, we have to find it is f x x it is 2, then del square f by del x del y which is 4 f z is 2. So, it is also 4 it is 2 by the symmetric property, because this is always symmetric. Then it is 8 a second derivative this is 2 and x z is and y z is 4. So, it is 4 it is 4. So, this is a Hessian matrix of this f.

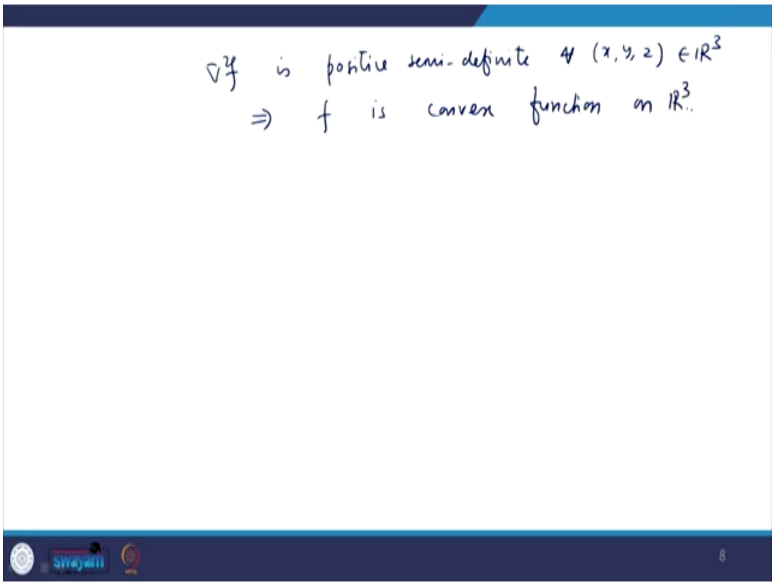
So, let us compute all the principal minors of this. First of all minors of order 1 cross 1 again. So, we have already seen minors of order 1 cross 1 are nothing, but diagonal elements. So, diagonal elements here are 2 8 and 2 ok.

Now, minors of order 2 cross 2. So, how to find minor of 2 cross 2? You delete suppose you are deleting third row and third column what left with us? 2 4 4 8 the determinant of this. Then suppose you are deleting first row and first column. So, it is eight 4 4 2 then ok. Second row second column is left here now second row second column is nothing, but 2 2 2 2 second row second column 2 2 2 2.

So, what is this determinant? 16 minus 16 is 0 here 16 minus 16 is 0 and this is also 0. All minors of order 2 cross 2 are 0. So, let us compute minors of order 3 cross 3. Minor only 1 minor is there minor of order 3 cross 3 is only one.

Which is a determinant of this matrix determinant of Hessian matrix which is equal to what? So, it is two times this row the 16 minus 16 is 0, which is minus 4 again 16 minus 16 is 0 plus 2 again same as 16 is 0 this is also comes out to be 0. So, now, here are all the principal minors are greater than equal to 0. See 1 cross 1 2 8 2 0 0 0; that means, all principal minors are greater than equal to 0. So, what we can say about the Hessian matrix of f?

(Refer Slide Time: 35:27)



So, this implies Hessian matrix of f is positive semi-definite for every x y z belongs to R 3. And by that theorem of convexity, we can say that f is convex function on R3. So, in this way in this way by simply by simply find the Hessian matrix of f and check whether this matrix is positive semi-definite or not. We can easily conclude that a given function is convex or not. So, this is the these are simple exercise that how we can find how we can see that a given function is convex or not, a function is twice differentiable.

So, in this lecture we have seen that if a function is twice differentiable then, simply finding the Hessian matrix of f and checking the positive definiteness of that Hessian matrix. We can simply conclude that a given function is convex.

Simply if similarly if you want to conclude a given function is concave, then on the same lines we can say that a function is concave on S, if and only if the Hessian matrix of f is negative semi-definite for every x in S ok. So, we will see some more properties of optimization constraint on unconstraint optimization on the next forthcoming lectures so.

Thank you.