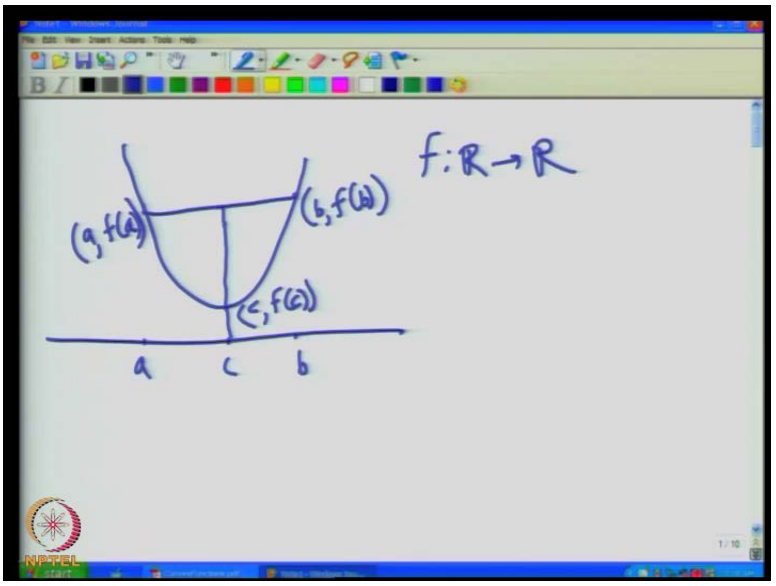
Numerical Optimization
Prof. Shirish K. Shevade
Department of Computer Science and Automation
Indian Institute of Science, Bangalore
Lecture - 9
Convex Functions (Contd)

(Refer Slide Time: 00:42)



Hello, so in the last class we started discussing about convex functions, so we defined convex functions and then started studying about the properties of the convex functions and how to characterize convex functions. So let us recall the definition of a convex function. So if we are given a function so, we are taking a function f from R to R in this case, and suppose we take any two points on the real line. So let us call them at as a and b and so the corresponding points on the functions are a f a and b f b.

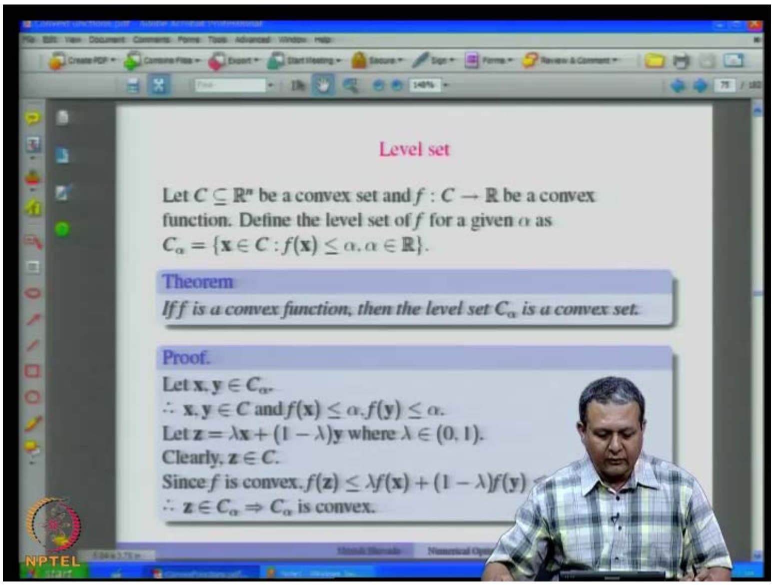
Now, if you take a chord joining these two points. So, in the case of convex functions this chord always lies on or above the function or in other words suppose, if we interpolate the points on this line segment joining a f a and b f b. So for every point will get a x coordinate and a y coordinate. Now, the y coordinate of that any point on this line segment is always has a value which is always greater than or equal to the value of the function at that point. So if you take a point c suppose here, so the value of the function at c is c comma f c right, and then if we take this point the corresponding point on this chord and if you find out its y coordinate, now that y coordinate will always be greater than or equal to f c.

So if that happens for any a and b in the domain of the function then we call it as a convex function or in other words if you take a chord joining any two points on the function the chord always lies on or above the function and then we also saw some properties and characterization of convex functions and one the important characterizations of convex functions is a epigraph. So epigraph of a convex function is a convex set and we proved that if the epigraph of a function is convex the function is convex or if the function is convex then the epigraph of the function is convex.

Now in today's class we will see more properties of convex functions especially, when the functions are continuously differentiable or wise continuously differentiable. We also will see some ways to derive more convex functions than using the existing convex functions and will also see how to extend the definition of convex functions. We have defined convex functions using only two points say x 1 and x 2 in the domain.

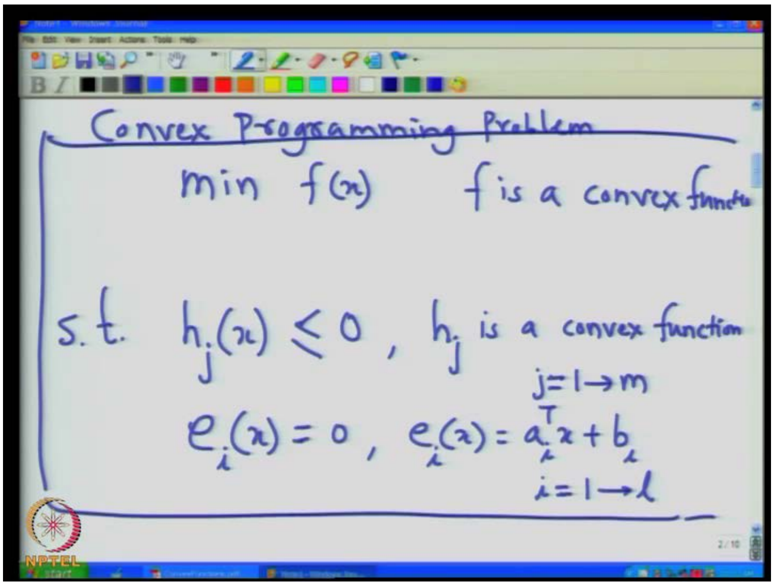
So f of lambda x 1 plus 1 minus lambda x 2 is less than or equal to lambda f of x 1 plus 1 minus lambda f of x 2, where lambda is in the range 0 to close interval 0 to 1 and x 1 and x 2 are from the convex domain set. So can we extend it to multiple number of points. We will see these properties in today's class.

(Refer Slide Time: 04:04)



Now in the last class in the last class we talked about the level set. So the level set of any function f is defined as the set of all points x in the set C such that f of x less than or equal to alpha right, where alpha is a real number and we call this as a set C alpha because there is a parameter alpha associated with the set. Now if f is a convex function, then we showed that the level set C alpha is a convex set for any alpha. So will see some of the interesting properties related to this.

(Refer Slide Time: 04:57)



Now so what we have seen is that if the function is convex then the level set of the function is convex. So suppose let us take a function h x let us take a function h x where h is a convex function and let us take the set h x less than or equal to 0 and we have h is a convex function. Now if you consider the set of all points x such that x in the domain of h, such that h x is less than or equal to 0 then by the definition of level set sorry by the properties of convex functions h x less than or equal to 0 set of all x such that h x less than or equal to 0 is a convex set, if h is a convex function.

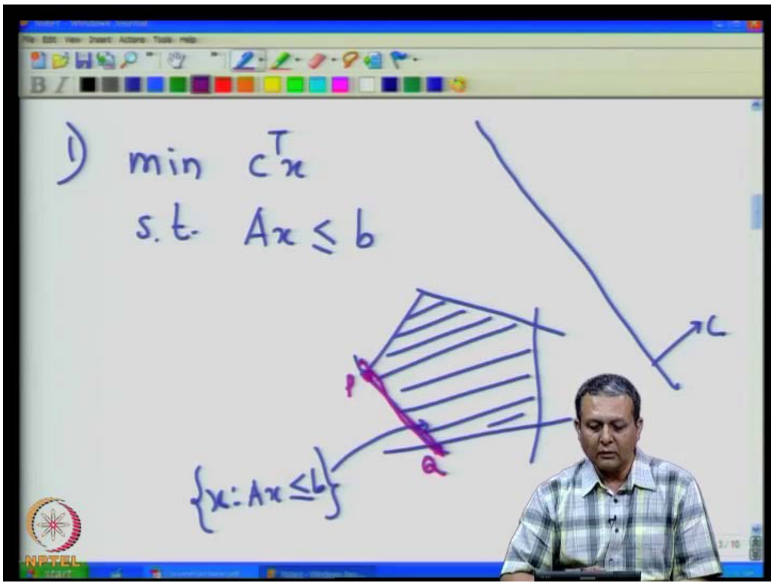
Now, we can take different convex function so let us call them as h j where each h j is a convex function. So each h j x set of all x such that h j x less than or equal to 0 is a convex set and now what we are going to do is that we are taking the intersection of all this convex sets. So as we have studied in earlier classes the intersection of convex sets is a convex sets. So the set of all x such that h j x less than or equal to 0, h j is a convex function is a convex set. So let us assume that j j are going from 1 to m. So we have m convex functions and then we are considering the set of points such that h j x less than or equal to 0 for all j going from 1 to m. Now this will form a convex set.

Now in addition to this suppose we have set of equalities so let us call them that e x equal to 0 is another type of set. Now so, if suppose let us just consider the set of that type x. So x belong to the domain such that the e x equal to 0. Now this set will be a convex set if and only if your e x is of the type a transpose x plus b. So in other words, so the sets of the sets of all points x such that e x equal to 0 is convex if and only if e x is affined. Now let us take some certain number of such affine sets. So let us assume that each e i is of the type a i transpose x plus b i where a i's are not equal to 0 and i is going from 1 to l. Now so we have h j x less than or equal to 0 where h j is a convex function and we have m such convex functions, so this set is a convex set.

So in addition we have some more sets, some more functions e i x which are affine of the type a r transpose x plus b i and this also forms a convex set. So we have a collection of convex sets and their intersection so each h j x is less than or equal to 0 is convex set, e i x equal to 0 is convex set and we have intersection of this convex sets. Now we know that any intersection of any collection of convex sets is a convex set. So let us consider a problem where we want to minimize a function f of x subject to the constraint that h j x less than or equal to 0 and e i x equal to 0 where each h j is a convex function and e i's are affine function. Now let us assume that f is also a convex function.

Then this problem that we have got, this is a convex programming problem. So this problem is convex programming and we have seen that for a convex programming problem every local minimum is a global minimum and all global minima form a convex set. So typical convex programming problem can be written in the form minimize f x subject to the constraint h j x less than or equal to 0 where, h j is a convex function, f is a convex function and e i x is equal to 0. Where, e i x is of the type a i transpose x plus b i, i going from 1 to l. So this is a typical convex programming problem. Now let us look at some of the examples of convex programming problems.

(Refer Slide Time: 10:26)



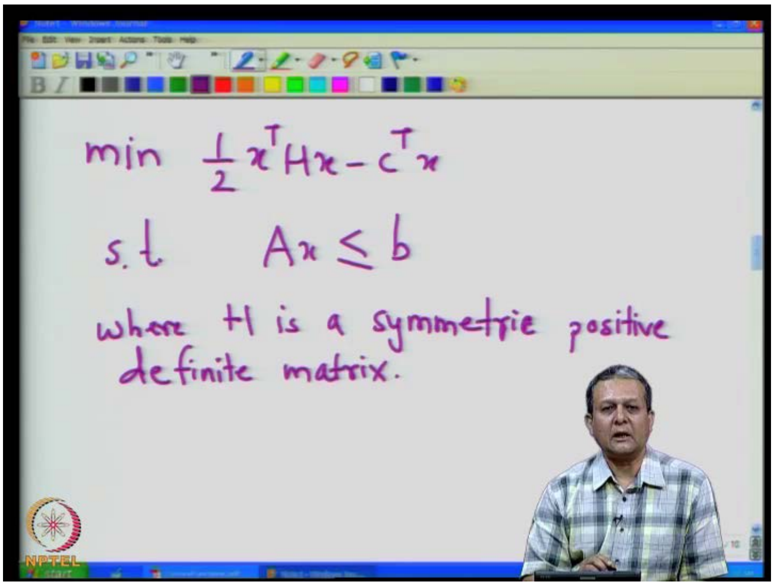
Now the first example that we can think of is a very simple example suppose we consider the problem minimize c transpose x subject to A x less than or equal to b. Now the objective function here is a linear function and we know that in linear function is a convex function. Now the constraint set we have seen that this constraint set is a convex set. So this is a convex programming problem now, if you want to see the example of one such problem so suppose so, this is our set, set of all x such that A x less than or equal to b.

So this our constraint set and suppose our function that we want to minimize is an affine function. So the normal to that hyper plane is a pointing in this direction. So this is the this is the function c transpose x. Now this affine function we want to or this linear function we want to minimize. Now you will see that if we want to minimize this function the minimum proline wood act occur at this point. It is this, this is our affine function now if we change the affine function so you will see that this is the only minimum for this if our objective function is c transpose x, where c is a vector in this direction.

Now suppose if we change the vector c so suppose if you make the vector c to be like this then if you want to minimize this function. So you will see that again this will be the minimum. So let us change the vector c again, now so if we take this as our c vector and if you try to minimize you will see that this entire line segment joining this two points forms a solution set for this problem. We will see more about such kinds of problems later in the course but, geometrically one can we convince that this entire line segment joining the points. Now let us call this points as a P and Q. So the entire line segment forms a solution set for this problem if c chosen to be like this.

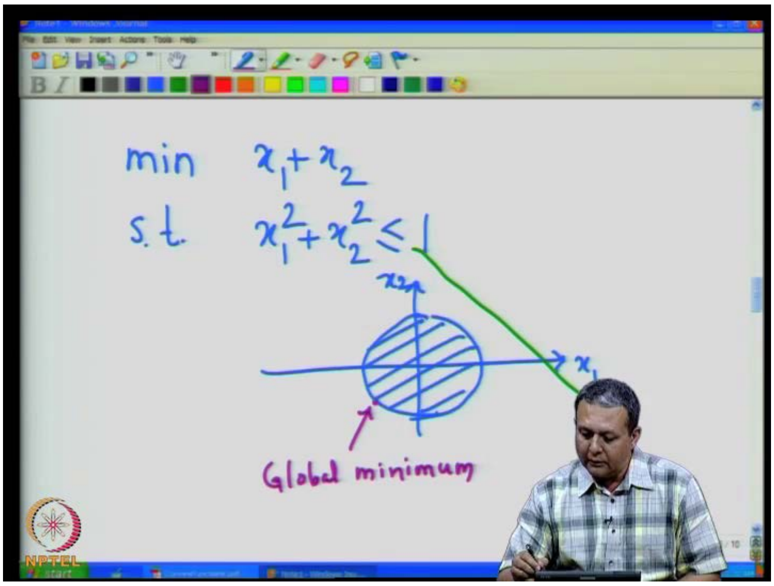
So you will see that the solutions there are multiple solutions in this case but, they form a convex sets. So as we studied in the last class, that this is the convex programming problem. So every local minimum is a global minimum so, all these are global minima at this points we get the same objective function value and in addition to that all these solutions which are on the line segments P and Q they form a convex set.

(Refer Slide Time: 14:11)



Now let us see some more examples. So let us consider the problem minimize half of x transpose H x minus c transpose x subject to the constraint that A x less than or equal to b. Where a is H is a symmetric positive definite matrix. Now if H is a symmetric positive definite matrix then one can see that, one can show that this is a convex function we will see that in today’s class. But, if under those circumstances and constraint set of the x less than or equal to b, this is going to be a convex programming problem which is again has the same properties as that of any a typical convex programming problem.

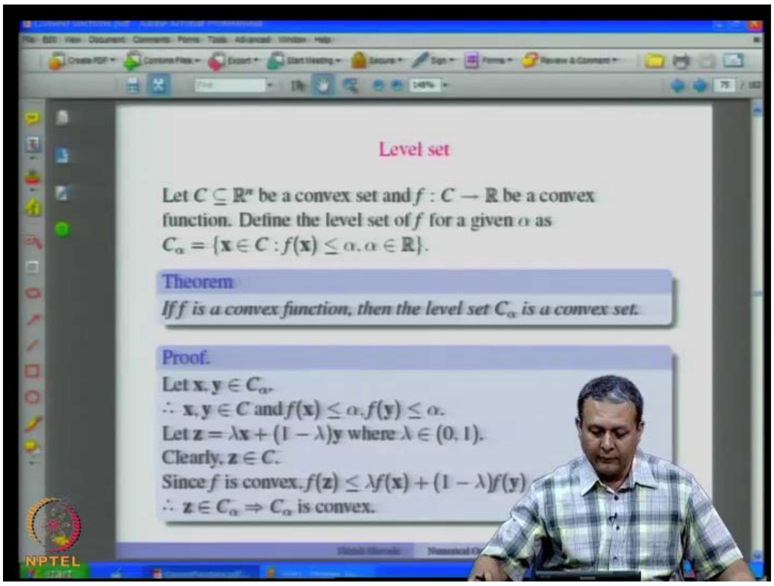
(Refer Slide Time: 15:40)



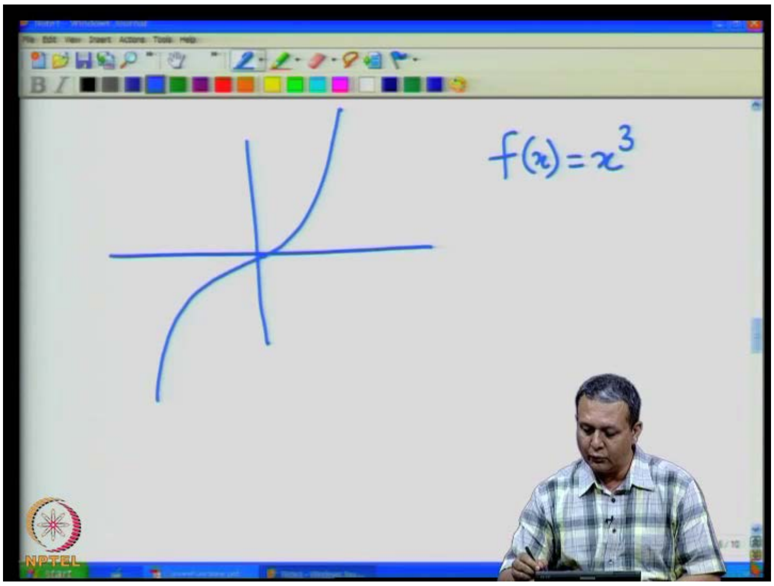
So let us see one more example. So let us consider minimize x 1 plus x 2 subject to the constraint that x 1 square plus x 2 square less than or equal to 1. Now if we take this constraint set, now it is a circle which center 0 and radius 1. So this is going to be our constraint set, where x 1 is this coordinate and x 2 is this coordinate and x 1 plus x 2 is a x 1 plus x 2 equal to constant is a line in this two dimensional space and what we are trying to do is that we are trying to minimize this the value of this objective function x 1 plus x 2 subject to the constraint that x 1 square plus x 2 square less than or equal to 1.

Now you will see that this is an affine function. So it is a convex function on the constraint set you will see that it is also a convex set the points on the boundary as well as the interior of the circle. So the constraints are convex sets. So this is a nice convex programming problem and you will see that the minimum of this function lies at some point somewhere here. So this function has a unique, so this is going to be our global minimum so, the x 1 and x 2 coordinates at this point will tell us about the global minimum and if we plug in those values of x 1 and x 2 in this objective function, then we will get the optimal objective function value. So these are some examples of convex functions and convex programming problem and you will be convinced that the global local any local minimum of convex function is a convex set is a is a global minimum and all these global minima form a convex set. Now if you look at the definition of level set we just say that the set of all x such that f x less than or equal to alpha. So this was the definition of a level set and so if we…

(Refer Slide Time: 18:23)

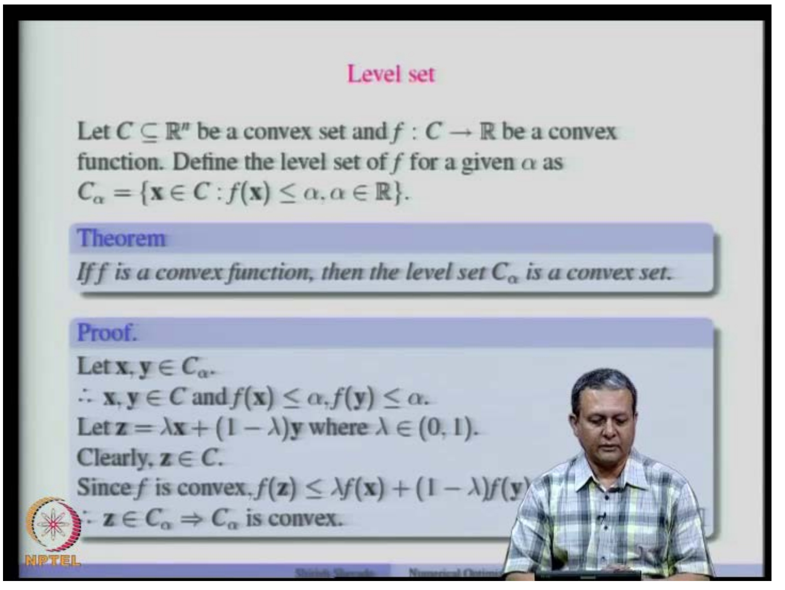


(Refer Slide Time: 18:36)

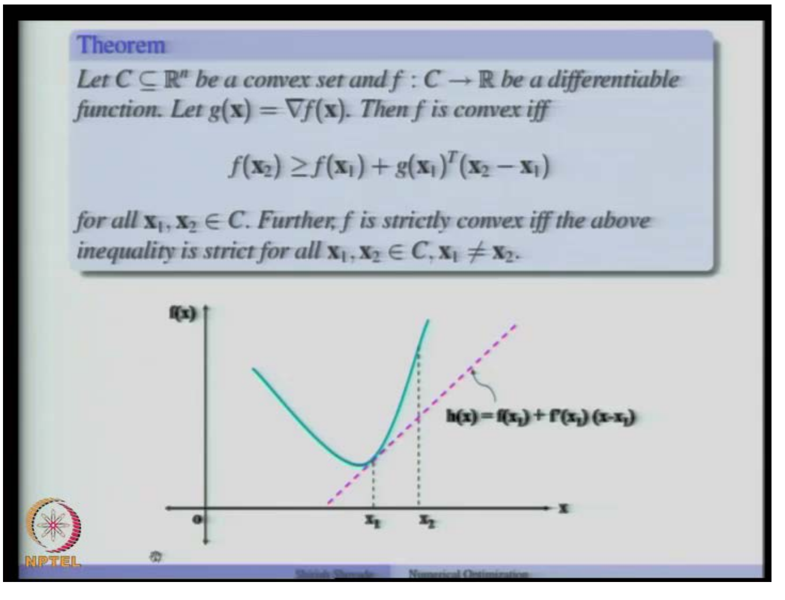


So what we have seen is that f is the convex function then the level set C alpha is a convex set. Now the converse of this statement is not true for example, if we if we consider a function f x equal to x cube and the function would look like this, now if you take the this function the level set of this function you can check that this is that is a convex set but, this function is not convex. So if f is a convex function then the level set of that function all the level sets of that function are convex but, the converse is not necessarily true.

(Refer Slide Time: 19:14)



(Refer Slide Time: 19:34)



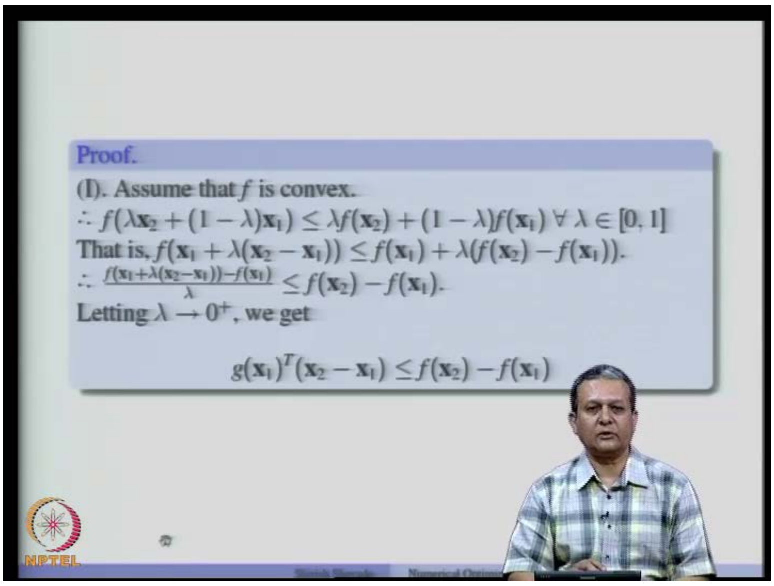
Now so far we have studied some properties of a convex functions without assuming any differentiability or twice differentiability. Now we will see some properties related to the differentiable or twice differentiable convex functions. Now the first theorem that we are going to see now is that assumes that f is a differentiable function. So let us consider C to be a convex set and f is a function differentiable function defined on the set C and the function is real valued. So let us denote the gradient of the function by g x so throughout this course will use g x to denote the gradient of the function f x and if f is a function from r to r we will the g x will be a scalar that will be the derivative.

Now, the theorem which says that if f the function f is convex if and only if, if you take any x 1, x 2 in the domain of that function and x 1 is not equal to x 2 then f of x 2 greater than or equal to f of x 1 plus g x 1 transpose x 2 minus x 1 and that happens for all x 1 and x 2 belongs to c and further the function is strictly convex is this inequality holds strictly for all x 1 not equal to x 2. Now before we study the proof of this theorem and it is the implications of this theorem. So let us try to see the geometrical interpretation of this theorem.

So let us take function f x which is shown here with the green line and so what the theorem says is that this function is convex if and only if. So if you take the right hand side of this inequality, the right hand side of this inequality says that so f of x 1 plus g x 1 transpose x 2 minus x 1 indicates the affine approximation of the function at x 1. So if you take a point x 1 and take an affine approximation of the function shown by a magenta line here the dotted dashed magenta line. So this is an affine approximation of the function f at x 1.

Now this affine approximation if we use and then find the value of the function at x 2. So this is the value of the function x 2, so f of x 2 is always greater than or equal to the corresponding value on this affine approximation at this point x 2. So in other words the affine approximation of a convex function does not over estimate the function. So this line is always on or below so this line which we the affine approximation that we have taken here, that always lies on or below the function. Now this theorem is very important to characterize convex functions. So, remember that the function is convex if and only if this holds, so if this holds then we can definitely say that the function is convex.

(Refer Slide Time: 22:57)



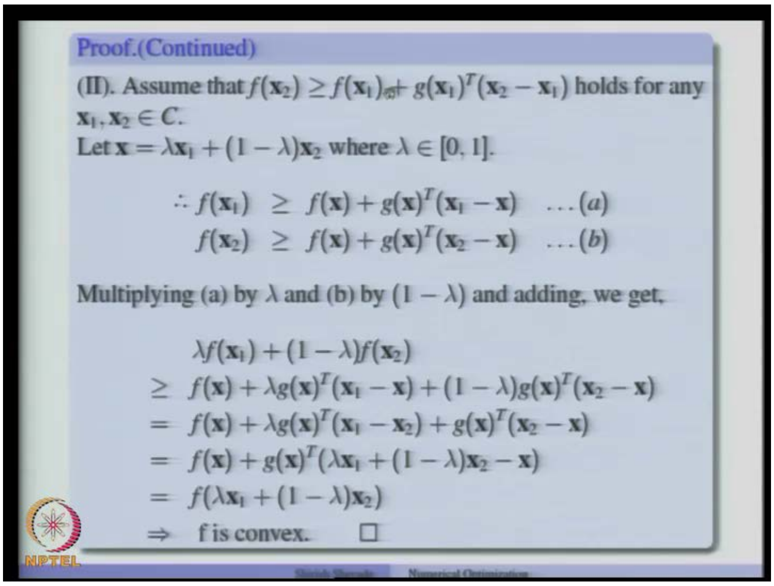
We will study some other implications of this theorem later. So let us first prove this theorem. Now the proof is in two parts. So the first part assumes that the function is convex and we prove that inequality holds and in the other part will do the will assume that the function is not convex and we will show that the inequality does not hold. So let us start proving the first part. So let us assume that the function is convex, now we can use the definition of a convex function. So we take any x 1, x 2 in the domain then f of lambda x 2 plus 1 minus lambda x 1 is less than or equal to lambda f of x 2 plus 1 minus lambda f of x 1 for all lambda in the range in the close interval 0 to 1. So this is by the definition of convex functions.

Now let us rearrange the terms here. So, let us take the x 1 term and separate it from lambda into x 2 minus x 1 and similarly, we can do the rearrangement on the right side. Now what we can do is that we can bring this f of x 1 on the left side and divide the whole expression by lambda assuming that lambda is not equal to 0. So we can do this and write f of x 1 plus lambda into x 2 minus x 1 minus f of x 1 divided by lambda less than or equal to f of x 2 minus f of x 1.

Now if you take so remember that lambda is not 0 but, lambda is in the in the open close interval 0 to 1. Now if we take the limit as lambda tends to 0, then this denotes the directional derivative of f at x 1 along the direction x 2 minus x 1. So this the letting a lambda tends to 0, where lambda is a positive number. So what we get is g of x 1 transpose x 2 minus x 1 that is the directional derivative of f at x 1 defined along the direction x 2 minus x 1.

Now that will be less than or equal to f of x 2 minus f of x 1 and this shows that f of x 2 will be greater than or equal to f of x 1 plus g x 1 transpose x 2 minus x 1. So this is the inequality that we wanted to show. So if f is convex using the convexity of f, we can show that this inequality holds for a differentiable convex function. Now let us look at the other part of the proof.

(Refer Slide Time: 25:36)



So let us assume that f of x 2, let us assume that the inequality holds and then we show that the function is convex. So let us take this inequality that we have in the theorem and that holds for any x 1, x 2 belong to c. Now since it holds for any x 1, x 2 belong to c and c is a convex set, we can take any x in the convex sets c which is the combination of x 1 and x 2 and then rewrite this equality in terms of the different points. So suppose we take any x which is a convex combination of x 1 and x 2, then we can say that the inequality holds for x 1 and x. So f of x 1 is greater than or equal to f of x plus g x transpose x 1 minus x and similarly, we can say that the inequality holds for x 2 and x because x 2 and x are again two different points in the set c.

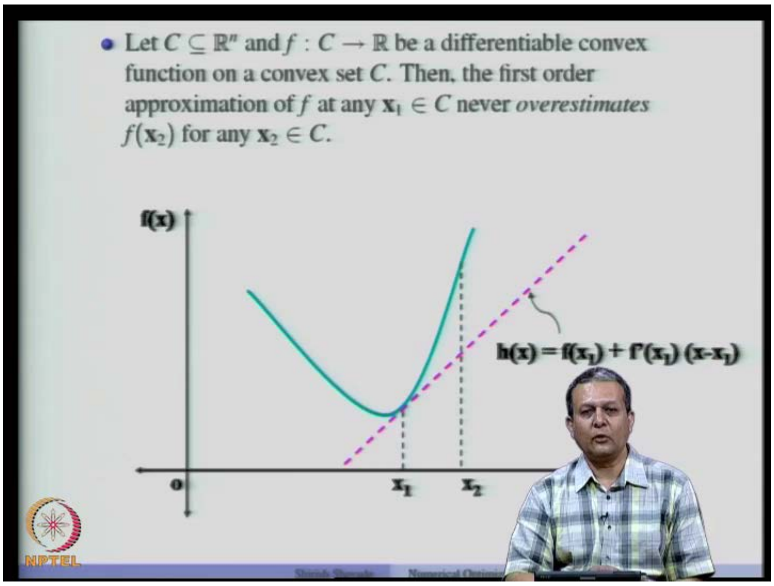
Now we have to show that the function is convex, so what we have to show is that lambda f of x 1 plus 1 minus lambda f of x 2 is greater than or equal to f of lambda x 1 plus 1 minus lambda x 2. Now if you look at x, x is nothing but, lambda x 1 plus 1 minus lambda x 2 and what we want is to show that f of x is less than or equal to lambda f of x 1 plus 1 minus lambda f of x 2. So what we do is that we multiple the first inequality by lambda and the second inequality by 1 minus lambda remember that lambda's are positive.

So the inequalities will remain as they are and what we get by multiplying the first expression by lambda and the second expression by 1 minus lambda is lambda f of x 1 plus 1 minus lambda f of x 2 and that is greater than or equal to f of so lambda f of x plus 1 minus lambda f of x will make it f x and then lambda g x transpose x 1 minus x plus 1 minus lambda g x transpose x 2 minus x.

Now we have to rearrange this terms and if you rearrange this terms we will see that what we get is f of x plus g x transpose lambda x 1 plus 1 minus lambda 2 minus x. Now if you look at this definition of x, x is nothing but, lambda x 1 plus 1 minus lambda x 2. So this x is nothing but, lambda x 1 plus 1 minus lambda x 2.

So this quantity here in this parenthesis vanishes and because of which what we get is lambda f of x 1 plus 1 minus lambda equal to f of x is greater than or equal to f of x and f of x is nothing but, f of lambda x 1 plus 1 minus lambda x 2. So you will see that this is the definition of a this is from the definition of a convex function f is convex. So we have shown that if f is convex then the inequality holds or if the inequality holds this inequality holds then the function f is convex.

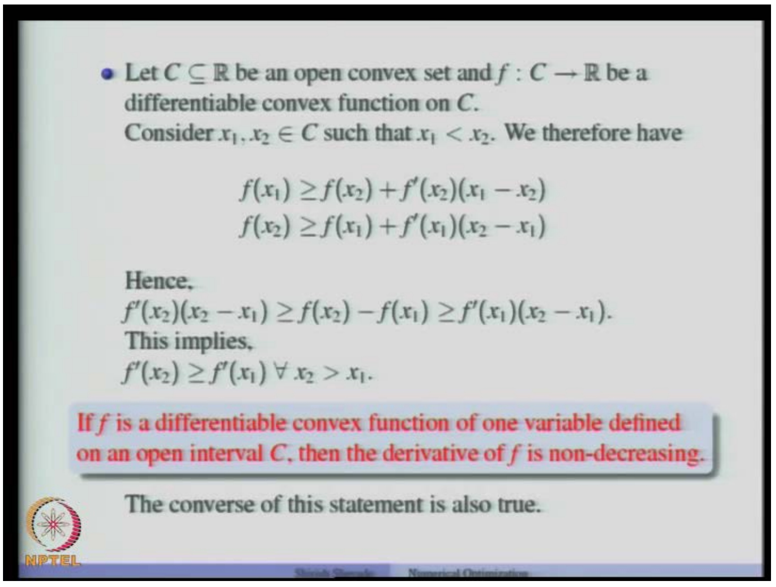
(Refer Slide Time: 29:03)



Now will see some of the results related to this. So one of the most important interpretations of this result as I mentioned earlier is that if f is the differentiable convex function on a convex set C then the first order approximation of f at any x 1 belonging c never overestimates f of x 2 for any x 2 belonging to c or in other words this affine approximation of the first order approximation of f at any x 1 it does not that line does not cut the function at any point of time, it always lies either on the function or below the function. So it does not overestimate the function and if a function is strictly convex then one can say that in those circumstances the first order.

Approximation of a strictly convex function always underestimates f of x 2 for any x 2 belonging to c. So, this result is important we will see the some more results related to this.

(Refer Slide Time: 30:13)



Now suppose we take a open convex set in the in the space of real numbers. So let that set be C which is subset of R and let us define a differentiable convex function f from C to R. Now so, let us consider any two points x 1 and x 2 in the domain of that function such that x 1 is less than x 2. Now what we are interested in finding out is that how do the remember that is now a function on one variable. So we have and in addition to that we have chosen two points where x 1 is less than x 2. So can we say something about the slopes of the functions at this two points when the function f is the differentiable convex function.

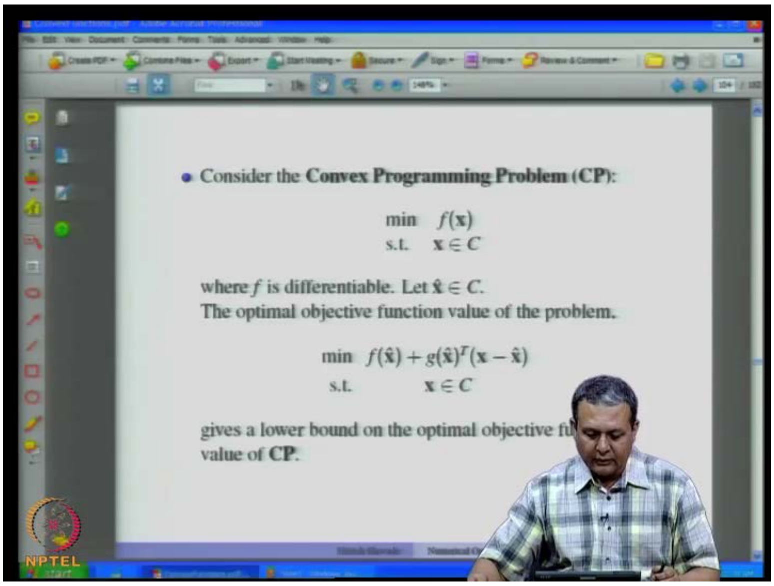
Now by using the theorem that we have studied so far, we can see that f of x 1 so, since x 1 and x 2 are this two points two points in the domain f of x 1 is greater than or equal to f of x 2 plus f dash x 2 into x 1 minus x 2 and similarly, f of x 2 is greater than or equal to f of x 1 plus f dash x 1 into x 2 minus x 1. Now we are interested in finding out what happens to the slopes of this two functions and this points especially when x 1 is less than x 2.

So what we will do is that we will write how does f of x 2 minus f of x 1 look like so, we can see that f of x 2 from this first expression we can see that f of x 2 minus f of x 1 is less than or equal to f dash x 2 into x 2 minus x 1 and from the second f of x 2 minus f of x 1 is greater than or equal to f dash x 1 into x 2 minus x 1 remember that x 1 is less than x 2. So x 2 minus x 1 is greater than 0.

So if we divide throughout by x 2 minus x 1 which is a positive number what we get is that f dash x 2 will be greater than or equal to f dash x 1 for all x 2 greater than x 1. So this means, that if we have a differentiable convex function on a open set C which is the subset of R and if we choose any two points x 1 and x 2 such that x 1 less than x 2 then the slope of the and if f is a differentiable convex function then the slope of that function at x 1 is always less than or equal to the slope of the function at x 2.

So which means that the slope of such functions we find on the set C which is a subset of R and the function is differentiable. So the slopes of such functions differentiable f is a differentiable convex functions the slope of such functions are non-decreasing.

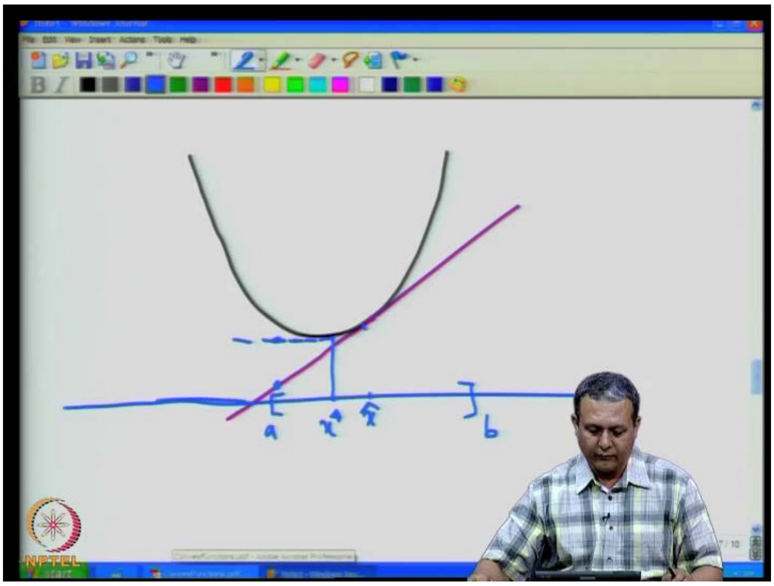
(Refer Slide Time: 33:55)



So if we have a differentiable convex function of one variable defined on an open interval C then the derivative of f is non-decreasing, in fact the converse of this statement is also true that if you have function differentiable function defined on an open interval C then the if the derivative of f is non-decreasing then the function is convex. So this is an important property of convex functions. Now let us consider a convex programming problem. So minimize f x subject to x belongs to C now since, this is a convex programming problem we have assumed that f is a convex function and c is a convex set. So only in that case we can call this as a convex programming problem.

Now let us make assumption that f is differentiable. Now if f is differentiable suppose we take a point x hat which is a domain of that function. Now by using our theorem we can say that the optimal objective function value of this convex programming problem we if you want to find out the lower bound on that optimal objective function value, what we can do is that we can take an affine approximation of f fact at x hat. So that affine approximation will look like this f of the affine approximation of f at x hat is f of x hat plus g x hat transpose x minus x hat. Now if we take this as our new objective function and then solve this. Problem now the variable is x so the objective function or this first quantity is a constant quantity. So we can ignore that so this is a affine function in x.

(Refer Slide Time: 35:48)

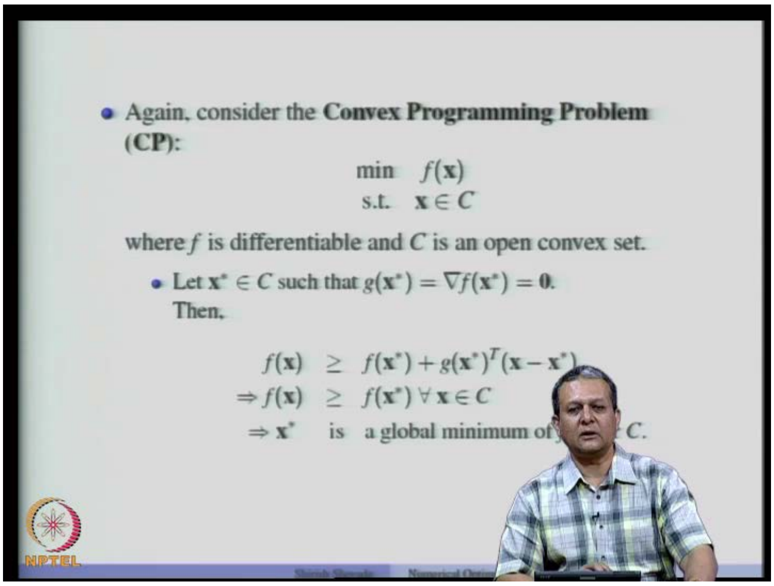


Now if we get an optimal value of this problem. Now that optimal value gives us a lower bound on the optimal objective function value of the original convex programming problem. So let us see that using an example. Now suppose we have a convex function define it like this and then suppose we take an affine approximation of f at a particular point. Now suppose we are interested in finding out the optimum of this function subject to the constraint that the interval that we are looking at is suppose a and b. Now our current result says that if we take any point so, let us call this as a x 1 and take the let us call this as x hat.

So x hat is our current point and a if we take affine approximation of the so this is a convex function and if you take a affine approximation of this. Now if you minimize this function right. So the minimum value of this affine approximation is at this point, this is an minimum value of the affine approximation over the set close interval a b.

Now the result that we have seen says that this is this gives a lower bound on the optimal objective function value. So this the optimal objective function value the optimal point of this which is our x star and the optimal objective function value at this point is this but, then we get a lower bound on the optimal objective function value by using an affine approximation of f at x hat. Now this lower bound may be a crude bound but, never the less it gives us some idea about the optimal objective function value, it gives an idea about the bound on the optimal objective function value.

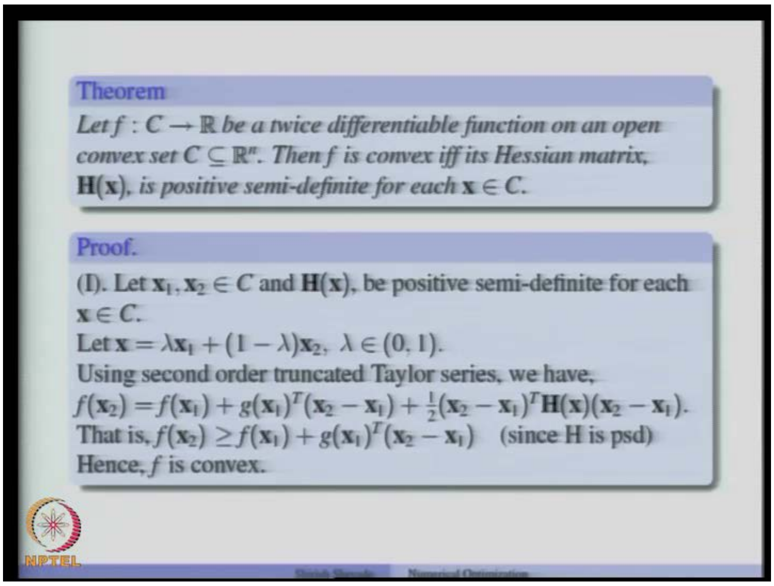
(Refer Slide Time: 38:17)



Now let us look at one more result let us again consider a convex programming problem where f is a differentiable convex function and C is an open convex set. Now let us consider some x star which belongs to the set C such that the gradient of the function vanishes at x star. So in other words g of x star is 0, then what can we say about x star. Now if we recall our earlier result on general one dimensional optimization problems you would have seen that the derivative when the derivative vanishes it does not guarantee anything whether a whether that point is a local minimum or not.

Now convex functions have this special property and will see what happens to this x star. So g of x star is 0 so suppose if we take any x in the set C then using our theorem we can say that f of x is greater than or equal to f of x star plus g x star transpose x minus x star. Now we know that g of x star is 0. So this quantity is 0 so in therefore, we can say that f of x is greater than or equal to f of x star for all x belong to C remember that is this equality holds this inequality holds for all x belong to the set C. So therefore, f of x is greater than or equal to f of x star for all x belongs to C. Now this means that x star is a global minimum of f over C. So for convex functions this condition g of x star equal to 0 is sufficient to ensure that x star is a global minimum of f over C remember that we are taking C as an open convex set.

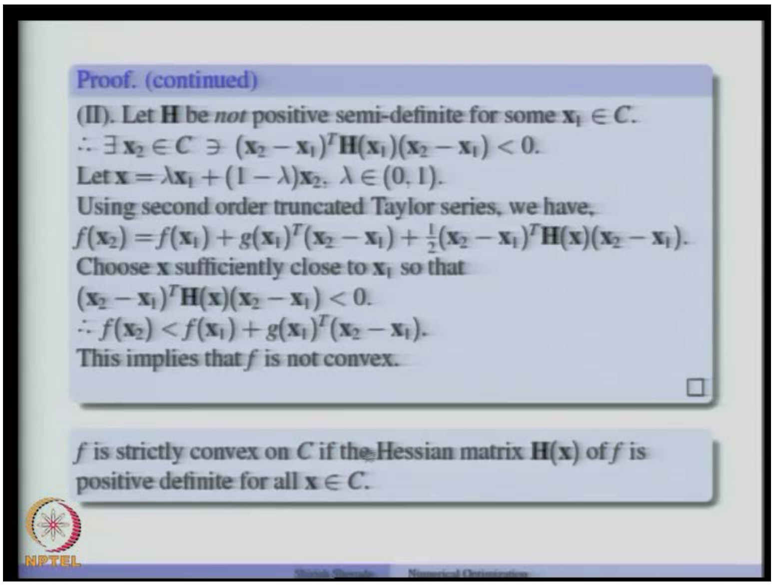
(Refer Slide Time: 40:59)



So this is the very important result and the reason why the second order information is not really important to identify the global minimum of the convex set, that we will study now. So it turns out that the Hessian matrix in the case of convex functions is always positive semi-definite. So that is why we really do not have to worry about the second order information to write the sufficient conditions for the global minimum of a convex function the first order conditions are enough to guarantee that a particular point is a global minimum of f over the open convex set C. So let us look at the second important theorem where f is a twice differentiable real valued function on an open convex set C which is the subset of R n.

Now the theorem says that if the f is convex if and only if its Hessian matrix H x is positive semi-definite or each x belong to the set C. So will prove this theorem, so again we have to prove this theorem in two parts. So will consider the first part so, let us assume that x 1 and x 2 belong to the set C and H x is positive semi-definite for each x belong to C and what we have to show is that in such a case the Hessian the function f is the Hessian matrix is positive semi-definite for each x belong to C and therefore, the function has to be convex. Now to show that the function is convex we can either show that basic property of the convexity holds or we can show that the condition which was shown in the previous theorem that also that holds.

(Refer Slide Time: 43:22)



So let us take x equal to lambda x 1 plus 1 minus lambda x 2 a point on the line segment open line segment joining x 1 and x 2 and if we use the truncated Taylor series we have we can write the truncated Taylor series at x 1 as f of x 2 is equal to f of x 1 plus g x 1 transpose x 2 minus x 1 plus half x 2 minus x 1 transpose H x x 2 minus x 1. Now remember that H x is a positive semi-definite for each x belong to C. So this quantity is always non-negative so therefore, we can write f of x 2 to be greater than or equal to f of x 1 plus g of x 1 transpose x 2 minus x 1 and by our earlier theorem this means that the function f is convex. So this is the first part of the proof.

Now let us look at the second part of the proof. So let us assume that H is not positive semi-definite for some x 1 belonging to C and then we come up with the result that if H is not positive semi-definite then f is not convex. So if H is not positive definite for some x 1 then their exist some x 2 in the set C such that x 2 minus x 1 transpose H x 1 into x 2 minus x 1 is less than 0.

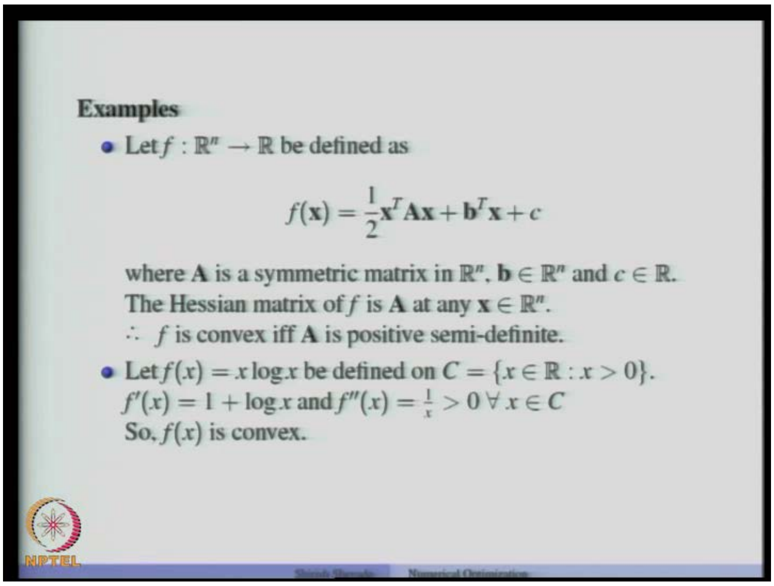
Now we will make use of this fact and to show that f is in such under this assumption f cannot be a convex function. Now so let us again take a point x which is on the open line segment joining x 1 and x 2 and then write the Taylor series truncated Taylor series of f at x 1. So f of x 2 is nothing but, f of x 1 plus the term related to the gradient of f at x 1 and then the term related to the gradient of f at any point x which is on the open line segment joining x 1 and x 2.

Now remember that H of x this quantity is less than 0, this is by our assumption. Now f is twice continuously differentiable. So H is a continuous function so, we can take x sufficiently close to x 1 such that this quantity is the third quantity in this expression is less than 0. Now if that happens so, this quantity is less than 0 then the f of x 2 will be strictly less than f of x 1 plus g x 1 transpose x 2 minus x 1 because this if x is sufficiently close to x 1 the last quantity becomes strictly less than 0 and we can write this and since this is true certainly f is not convex because of our earlier theorem.

So if we start with the assumption that H is not positive semi-definite then that implies that f is not convex. So, this means that f is convex if and only if the Hessian matrix at any point in the domain is a semi-definite positive semi-definite matrix. Now we can have one more result which says that f is strictly convex on C if the hessian matrix H x of f is positive definite for all x belonging to the set C.

Now the proof of this result goes along similar line. So will not prove this so remember that, f is convex on C in if and only if the Hessian matrix of f at any x is positive semi-definite while f is strictly if the Hessian matrix is strictly positive definite then f is convex. So remember that this is not if and only if so I leave it as an exercise to find out the case where f is strictly convex but, the Hessian matrix of the function is not positive definite but, if the Hessian matrix is positive definite then we can definitely say that f is strictly convex.

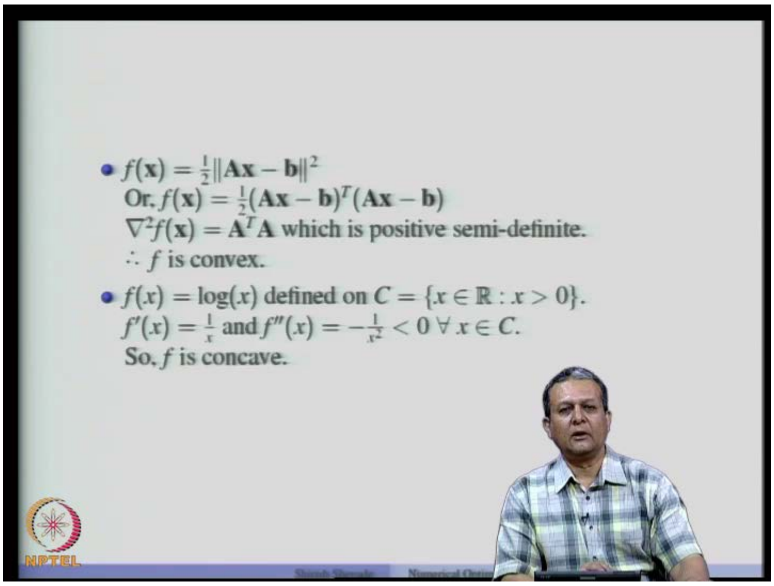
(Refer Slide Time: 46:57)



Now let us see some examples of convex functions. So let us take a function f defined from R n to R as f x is equal to half x transpose A x plus b transpose x plus c. Where A is the symmetric matrix in R n. Now if we take the Hessian of this matrix the Hessian this Hessian of this function the Hessian is the matrix A. Now by our theorem we have that f is convex if and only if the Hessian matrix is positive semi-definite. So this function is convex if and only if A is positive semi-definite. So whenever A is positive semi-definite this is our quadratic function in x. So this function is convex let us look at the another example suppose will let us take a real valued function defined on the set of positive real numbers.

Now this set of positive real numbers from a convex set and let us take the function f to be f of x equal to x log x. Now we want to test whether this function is convex or not. Now note that the function is twice differentiable so, we can take the second derivative of this function and see its behavior and based on that we can conclude whether the function is convex or not. Now we can take the derivative of this function and the derivative is 1 plus log x and the second derivative is 1 by x and that quantity is greater than 0.

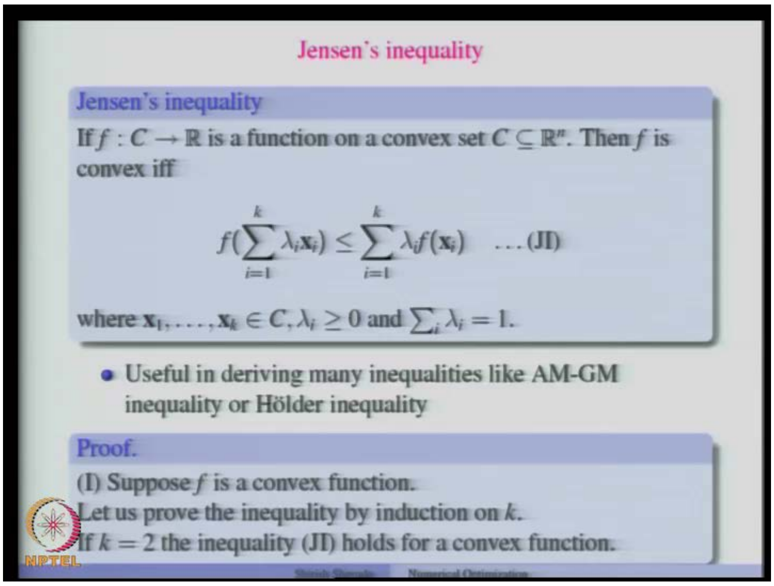
(Refer Slide Time: 49:17)



For every positive real number or for every x belonging to C so therefore, since the second derivative is always greater than 0 we can say that f is convex. In fact, one can also say that f is strictly convex function because second derivative is always greater than 0.

Now let us take a function f x equal to half norm A x minus b square remember that we use two norm suppose and we can write this rewrite this function as half of A x minus b transpose A x minus b and if you take the gradient, the gradient of this function if you take the Hessian of this function is a transpose A and A transpose A this matrix is always positive semi-definite. So this function is always a convex function, the function f is always a convex function. Let us consider a function f of log x defined again on the set of positive real numbers now if it again take the first derivative and the second derivative. So the first derivative is 1 by x and the second derivative is minus 1 by x square and that quantity is always less than 0 whenever x belongs to C. So one can say that that happens when the function is concave. So these are some examples of convex functions one can also define the new convex function using existing convex functions we will study later.

(Refer Slide Time: 50:34)

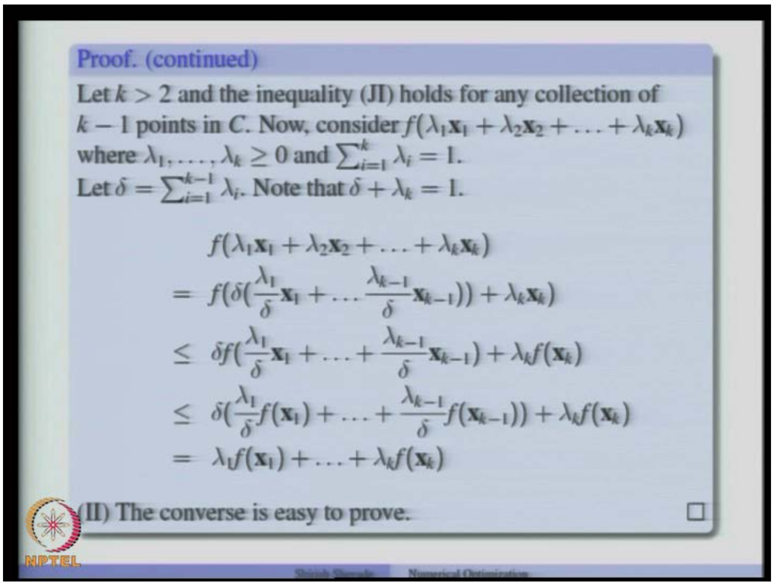


Now let us look at an important inequality called Jensen's inequality. Now let f be a real valued function defined on the convex set C which is the subset of R n. When f is convex if and only if f of sigma lambda x i is less than or equal to sigma lambda sigma lambda f of x i. So this is a Jensen's inequality where x 1 and x 1 to x k they are from the set C and lambdas are non-negative and the sum of lambda x is 1. Now this is this result is an extension of the definition of convex functions. So if you recall that we say that f is convex if and only if f of lambda 1 x 1 plus lambda 2 x 2 less than or equal to lambda 1 f of x 1 plus lambda 2 f of x 2. Where x 1 and x 2 belong to the set C and lambda 1 and lambda 2 are such that they are positive non-negative numbers and lambda 1 plus lambda 2 equal to 1.

Now instead of taking only two points if you take k any k points from the set C and take k lambdas which are non-negative such that there lambda is there sum is 1 then Jensen's inequality says that a function is convex if and only if this inequality holds. Now interestingly this inequality also can be used to derive some other inequalities like arithmetic geometric mean inequality. We will see that later but, the result is very important and it is useful in deriving other equalities other inequalities like AM GM inequality or holder inequality now let us try to prove this theorem.

Now the proof of this theorem is by method of mathematical induction and it goes in two parts. So now the first part assumes that f is convex and then this inequality holds, the other part assume that this inequality holds and then shows that f is convex. Now let us prove this inequality by the principle of mathematical induction on k. Now when k is equal to 2 the inequality holds because by the definition of convex function. So f is convex then certainly if k is equal to 2 this inequality holds. So for k equal to 2 the inequality holds or a convex function. Now what we do is that we assume that the inequality holds for a minus 1 and then show that it also holds for the case k.

(Refer Slide Time: 53:33)



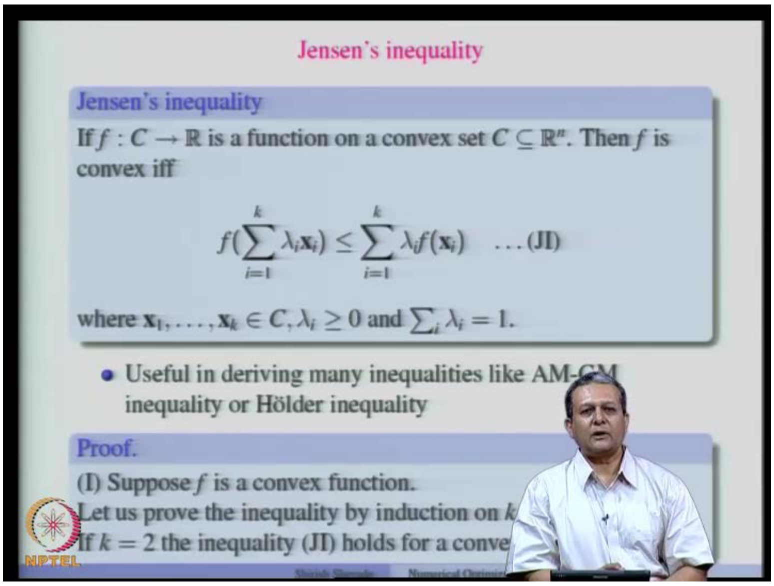
So let us assume that k is greater than 2 and the Jensen's inequality holds for any collection of k 1 k minus 1 points in the set C. Now let us add the k'th point. So suppose we have x 1 to x k minus 1 where the initial points and we have added the point x k and then we take a convex combination of those k points. So the sigma lambda x i where lambda is a non-negative and sigma lambda i i going from 1 to k is 1 that is the convex combination of these points. So we are interested in finding out what happens at this now we know that the inequality holds for two as well as k k minus 1 points. So let us make use of those facts.

So let us try to separate the first k minus 1 points from the kth point. So let us take lambdas related to the first k minus 1 points sum them up let it let us call it as delta and when delta is added to lambda k we get 1. So we have now the first k minus 1 point the combination of first k minus 1 point has 1 point and the kth point has the another point and we have a convex combination of this two. So f of lambda 1 x 1 plus lambda 2 x 2 up to lambda k x k it can be written as now what we have done is that we have separated the first k minus 1 point from the k'th point. So these the quantity in the inner parenthesis gives us the combination linear combination of the first k minus 1 points and then delta into that quantity plus lambda k x k note that delta plus lambda k is equal to 1.

So now we have two points in the domain and the convex combination of these two points. Now you will see that by the definition of convex functions we can write f of this is less than or equal to delta into f of the first point plus lambda k into f of x k. So we were able to get lambda k f of x k out of this argument of f. Now we also know that the inequality holds for k minus 1 points now we have this k minus 1 points. So since f is convex and the Jensen's inequality holds for k minus 1 points we can expand this and write it further as so f of this quantity is less than or equal to lambda 1 by delta f of x 1 plus so on up to lambda k minus 1 by delta into f of x k minus 1 and now when we simplify what we get is lambda 1 f of x 1 plus lambda 2 f of x 2 up to lambda k f of x k.

So this is this implies that the Jensen's inequality holds for k if f is a convex function. Now the converse of now the other part of this theorem, that if this inequality holds then we have to show that the f is convex that is very straight forward. We just take the first two points the remaining lambda k is equal to remaining lambdas to 0 and then we can show that it is easy to prove the from the definition of a convex function that the function is convex.

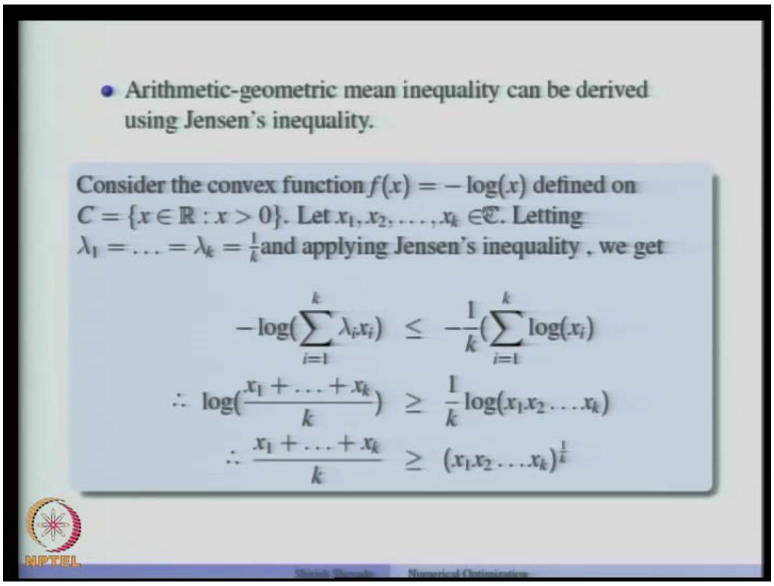
(Refer Slide Time: 57:36)



So the converse or the other part of the theorem is easy to prove. So in the next class we will study how to prove arithmetic geometric mean inequality using Jensen's inequality also will study some of the properties of convex functions and how they can be used to derive more convex functions. So if you recall if we take a function f from a convex set to set of real numbers then according to Jensen's inequality f is convex if and only if this inequality holds where x 1 and x x 1 to x k belong to the set C lambda i’s are non-negative and sigma lambda equal to 1.

So under these conditions that lambda not non-negative and sigma lambda equal to 1 sigma lambda x i ha is called a convex combinations of x i’s. So this inequality as you can see it is a generalization of the definition of convex functions to more than two points in the convex sets and in the last class we proved this inequality and I also mentioned that this inequality can also be used to derive many other inequalities like AM GM inequality or holder inequality. So today we will first see how to derive the AM GM the arithmetic mean geometric mean inequality using Jensen’s inequality.

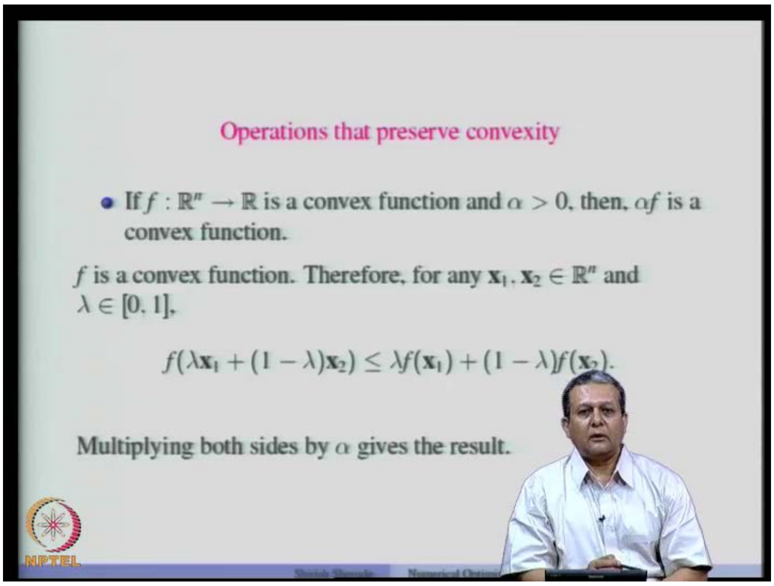
(Refer Slide Time: 58:54)



So now for that purpose let us consider a convex function from set of positive numbers to the set of real numbers to be f x equal to minus log x. So remember that the this function is different on the set C which contains the set of positive real numbers. Now let us apply Jensen’s inequality to this function. So let us consider x 1 to x k to be the k points in the set C and let us assign lambda to be 1 by k to all the points and then if you apply Jensen’s inequality what we get is that minus log of sigma lambda x i is less than or equal to minus 1 by k sigma i going from 1 to k log of x i.

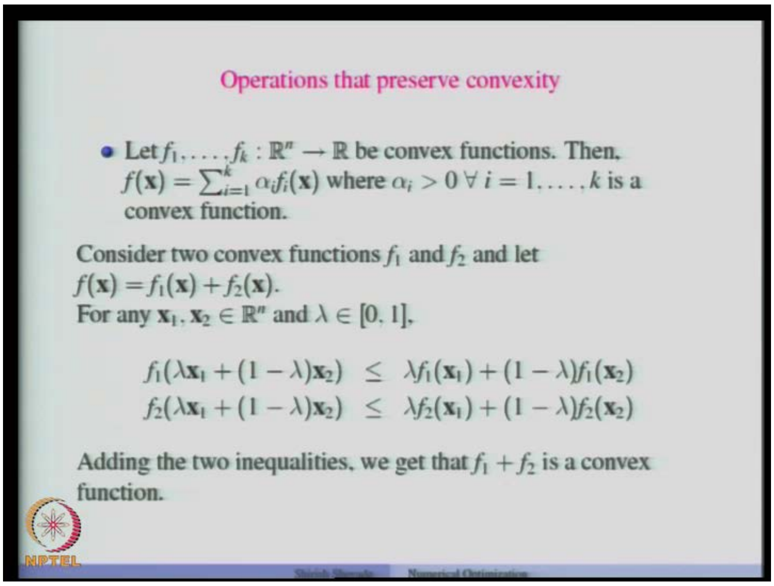
Now if we simplify, so what we get is log of x 1 plus x 2 plus up to x k by k is greater than or equal to 1 by k log of x 1 x 2 x 1 into x 2 into x 3 and so on up to x k. If we further simplify this what we get is the arithmetic mean on the left side and then the geometric mean of those k numbers x 1 to x k on the right side and this clearly says that arithmetic mean of those k numbers is greater than or equal to geometric mean of those k numbers. So this is an interesting way of proving arithmetic geometric mean inequality using Jensen's inequality and using the fact that with the function minus log x is a convex function.

(Refer Slide Time: 60:54)



Now will look at some of the operations that preserve convexity of a function. So first function the first property is that if you have function from R n to R which is a convex function and if we take any alpha which is a positive number, then alpha f is a convex function. So most of this properties are easy to easy to derive from the first principles. So we have f is a convex function so if we take any x 1 x 2 in R n and take a the lambda in the close interval 0 to 1 then by the definition of convexity f of lambda x 1 plus one minus lambda x 2 is less than or equal to lambda f of x 1 plus 1 minus lambda f of x 2. Now alpha is a positive number we can multiple by alpha on both the sides and then inequality does not change the direction and therefore, by multiplying alpha we get that alpha f is also is a convex function.

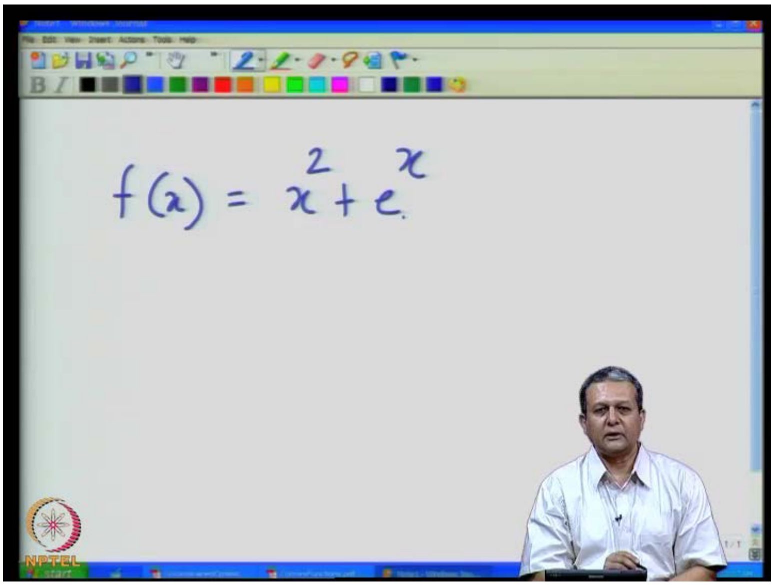
(Refer Slide Time: 61:55)



Now here is another result which considers k convex functions from R n to R, f 1 to f k are those convex function and let us define a function f of x to be non-negative or strictly positive combinations of those functions. So f of x is nothing but, sigma i going from 1 to k alpha i f i x where each f i x is a convex function and alpha is a strictly positive then the claim is that f is the convex function. Now to prove these result what we do is that we will first take only two functions f 1 and f 2 and prove the result and then the extension to this general cases will be very obvious.

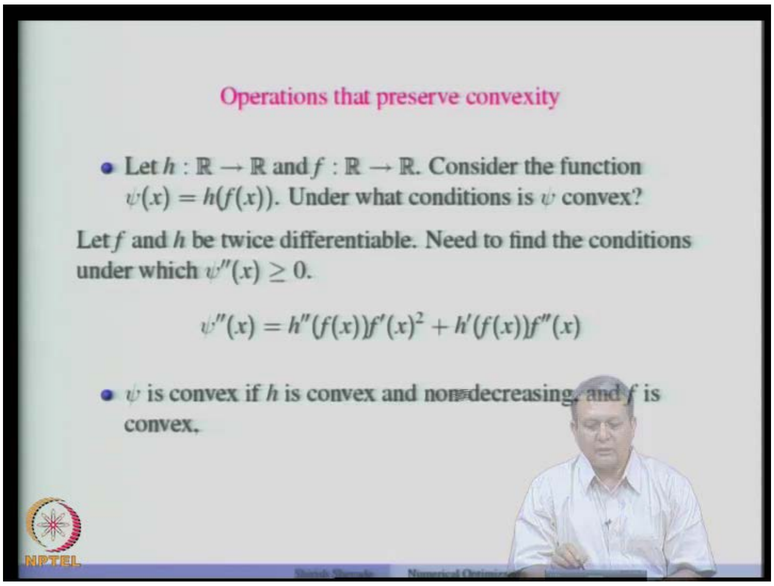
So let us consider two functions f 1 and f 2 and let f of x be a f 1 x plus f 2 x. Where in these cases we have assumed that alpha 1 is 1 and alpha 2 is 1 and then by the convexity of each of those functions we can say that for any x 1 and x 2 in R n and lambda in the close interval 0 to 1 these two inequality holds because of the convexity of the two functions. Now if we add the left sides and at right sides what we get is that f 1 plus f 2 of lambda x 1 plus 1 minus lambda x 2 is less than or equal to the addition of these two and that clearly shows that f 1 plus f 2 is a convex functions. Now this result can be extended to prove that f of x is a convex function because if you multiply this by alpha 1 and alpha 2 the result holds and similarly, one can extended to the general cases of k different convex functions.

(Refer Slide Time: 64:12)



So these are some ways to get new convex functions from the existing convex functions. So suppose we have a function f of x to be x square plus e to the power x. Now you will see that this is x square is a convex function e to the power x is the convex function. So we are taking the sum of these two convex functions so that will be a convex function.

(Refer Slide Time: 64:47)



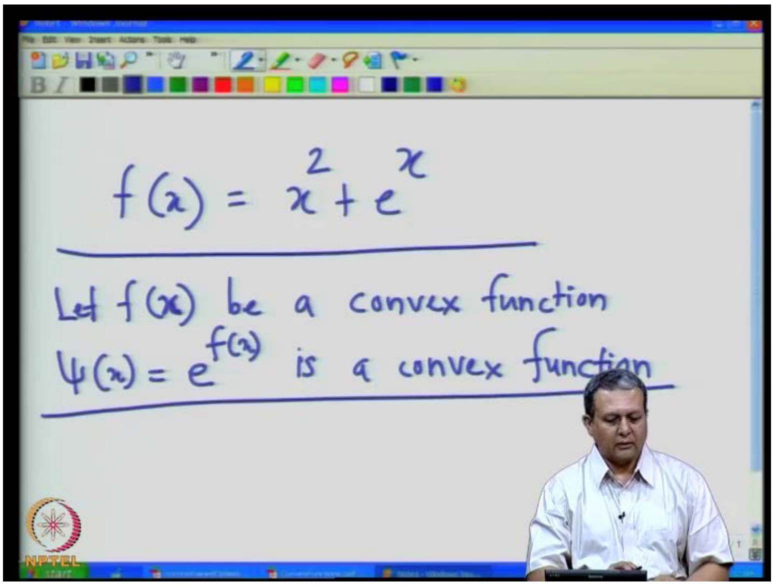
Now let us look at a another result related to convexity of functions. So let us consider a function h from R to R and f some R to R. So both are real valued functions on the domain of real variables and let us consider the composition of these two functions let us call it as psi x which is nothing but, h of x of x. Now the question is at under what conditions is psi x convex. Now we know that for such real valued functions remember that psi x is also a real valued function on the space of on the domain of real numbers. So we have already seen the result that the function is convex if its second a real valued function defined on the domain of real numbers is convex if its second derivative is non-negative. So we can use that fact now for that purpose what we need is that we need that price difference ability of h and f.

So let us assume that f and h are twice differentiable and we need to find the conditions under which the second divertive of psi x is non-negative. Now if you write the second divertive of psi x so you will get some expression something like this which is shown here. Now under what conditions psi to dash x is non-negative now just look at the first term. So the f dash x square is a non-negative quantity. So we want h 2 dash f x to be also non-negative. Now let us look at the second quantity now f f 2 dash x and h 2 dash f x this product has to be non-negative. So if that happens and h 2 dash f x is greater than or equal to 0 then we have psi 2 dash x greater than equal to 0.

Now when will this will be greater than 0 so one in one case suppose that let us assume that f is convex. So; that means, f 2 dash x is always greater than or equal to 0. Now we have to make sure that h dash f x is greater than equal to 0 and h 2 dash f x is greater than or equal to 0.

Now let us assume that h is also convex. So if h is convex then h 2 dash of f x is always greater than equal to 0. So the first quantity becomes non-negative and suppose that h dash f x is also greater than or equal to 0. So which means that the function is non-decreasing. So if f is convex and the function h is convex and non-decreasing then h of f x which is nothing but, psi x becomes a convex function. So psi is convex if h is convex and non-decreasing and f is convex.

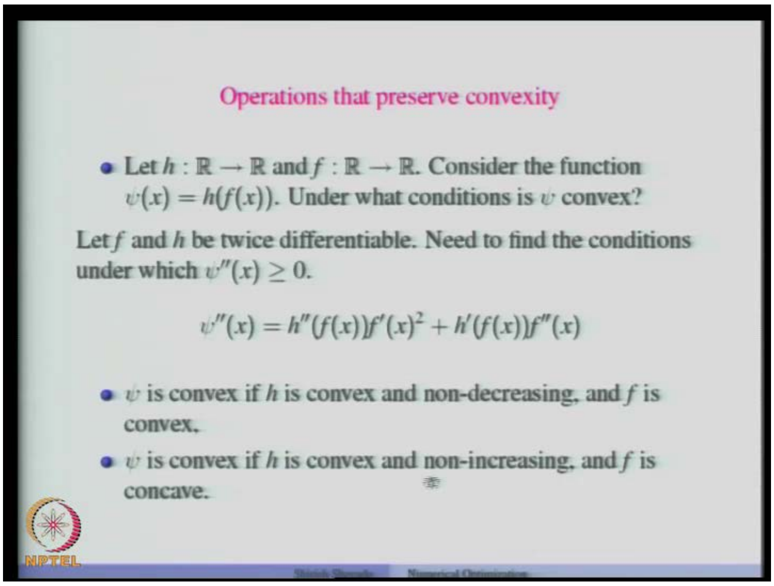
(Refer Slide Time: 68:05)



Now if u look at so let us consider the function f x is equal to f x is a so let f x be a convex function. So let us consider any convex function and then psi x to be e to the power f x. Now you will see that e to power f of x so e of y is some function which is a convex function and non-decreasing function. So this is a this e function is nothing but, the h function in our definition and then we have f which is a convex function. So psi x equal to e to the power f x is a convex function because e is a convex function and non-decreasing convex function and that compose with f of x which is a convex function will give us e to the power f x to be a convex function.

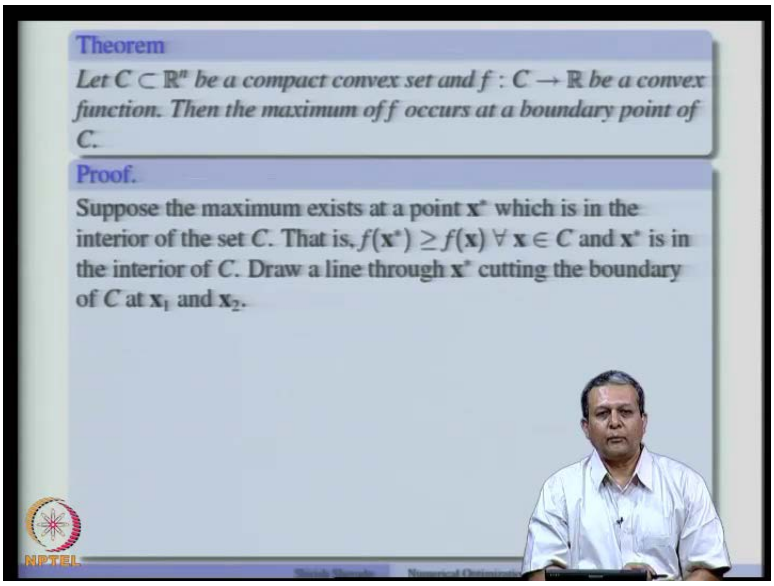
Now similarly, one can derive the condition that if h is convex and non-increasing. So h is convex means it is greater than or equal to 0 this quantity of course, is greater than or equal to 0 and non-increasing means h dash f x is a less than or equal to 0 then we want f 2 dash x to be less than 0 right and that will happen when f is concave. So when f is and h is convex and non-increasing then psi 2 dash x is greater than equal to 0 and psi is a convex function

(Refer Slide Time: 69:31)



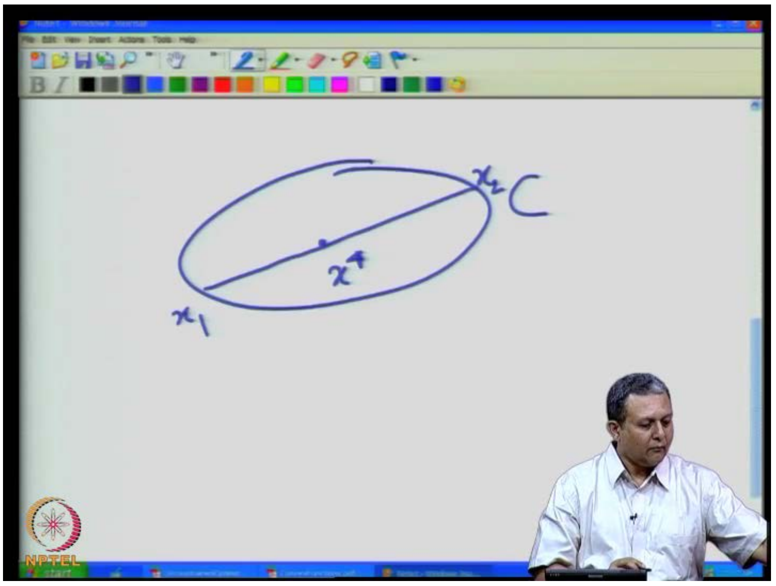
Now similar results can be derived for concave functions only thing in that cases that we have to find out the conditions under which psi 2 dash x is less than or equal to 0. So one can use the same logic to derive those conditions. Now let us see one theorem related to the convex functions and concave set remember that this theorem is not related to convex programming problem but, in this case we are trying to maximize the convex functions over a compaction.

(Refer Slide Time: 70:36)



So let us see which is subset of R n be a compact convex set and let a be a convex function we find from C to R then the claim is that maximum of f occurs at a boundary point of C. So in other words the maximum of f if you are trying to maximize the convex function over a compact convex set then the maximum would occur at a boundary point. Now let us prove by contradiction so suppose that the maximum exists at a point x star which is in the interior of the set that is f of x star greater than or equal to f of x for all x in C remember that x star is also in the set C and x star is in the interior of the set C and in such a case let us see what happens.

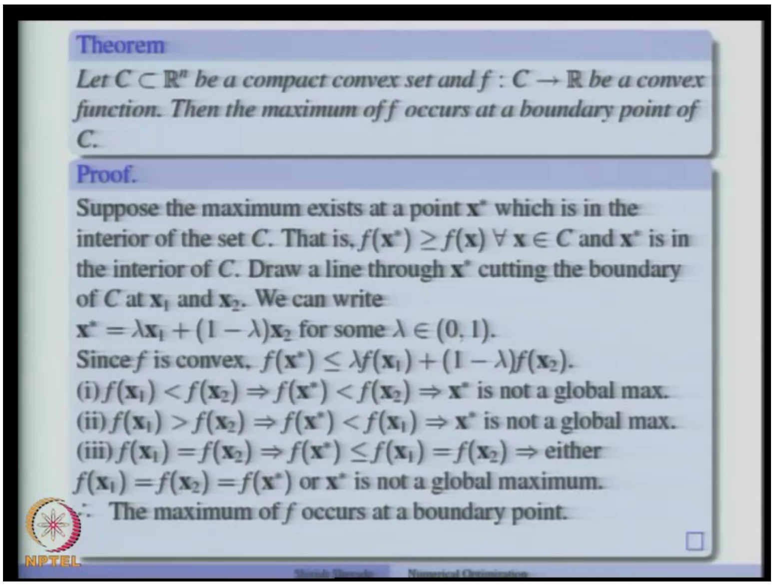
(Refer Slide Time: 72:23)



Now the compactness of the set C is very important in this case and let us try to utilize that. So let us draw a line cutting a through x star and cutting the boundary of C at x 1 and x 2. Now remember that remember that we have a compact convex set so this is our compact convex set C and then suppose that x star is a point which in the interior. Now what we are doing is that we are trying to find the line passing through x star and which cuts the set C at some boundary points say x 1 and x 2.

Now this boundary points do exists because we have assumed that C is a close bounded convex set. So we can always find this points x 1 and x 2 where that line passing through x star cuts the boundary of this and what we will prove is that if the maximum is the is at x star which is in the interior of the set C then we can show that there will be points which are on the boundary which will have a functional u higher than f of x star.

(Refer Slide Time: 73:43)



So; that means, that our claim that f of x star is the maximum or x star is the maximum point and f of x star is the maximum function value over the compact set C is not correct and this claim will be derived based on the fact that f is a convex function. So let us look at the proof of this. So let us take a line segment through x star which is cutting the boundary of C at x 1 x 2.

Now we can write x star as a convex combinations of x 1 and x 2. So we can write x star as lambda x 1 plus 1 minus lambda x 2 or some lambda in the open interval 0 to 1. So remember that x star is not equal to x 1 and x star is also not equal to x 2 because it is the point on the in the interior of the set C. Now we will use the convexity of the function f.

So since f is convex we can write f of x star is nothing but, that is less than equal to lambda f of x 1 plus 1 minus lambda f of x 2. Now let us consider the case where f of x 1 is less than f of x 2. Now f of x 1 less than f of x 2; that means, that f of x star is less than f of x 2 so which means that x star is not a global maximum because we have found a point x 2 on the boundary of the set C which has a function value higher than f of x star so which means that x star is not a global maximum.

Now similarly, we can consider the case where f of x 1 is greater than f of x 2 so then obviously, f of x star will be less than f of x 1 and which again means that x star is not a global maximum and if we consider the third case where f of x 1 equal to f of x 2. So which means that f of x star is less than or equal to f of x 1 and f of x 1 is nothing but, f of x 2.

So there are two possibilities that this inequality is strict inequality. So in that case f of x star will be less than f of x 1 so which again is a contradiction that x star is a global maximum. So the only thing which is possible is that f of x star is equal to f of x 1 equal to f of x 2. So; that means, that the function f is shows that the value of the function at x star is same as the value of the function at x 1 and same as the value of the function at x 2.

So in that case certainly we have found a point on the boundary which is a maximum but, if the inequality holds the strict inequality holds then x star is not a global maximum. So therefore, the maximum of a convex function over a compact convex set occurs at a boundary point. So with this we complete our discussion on convex functions.