Convex Optimization

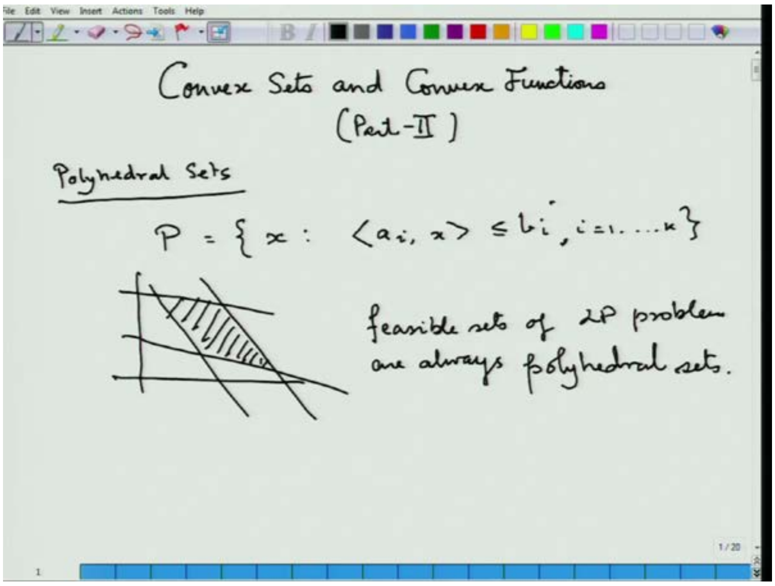
Prof Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

Module No # 01

Lecture No # 04

Yesterday, when I closed the lecture talking about Convex Sets we had talked spoken about lot of Convex Sets, various types of Convex Sets properties, and we said that we are going to talk about convex functions today. But before we do so, let me start by telling you about a very important class of Convex Sets which is fundamental to optimization basically because it is a part of linear programming which is a very important part of optimization.

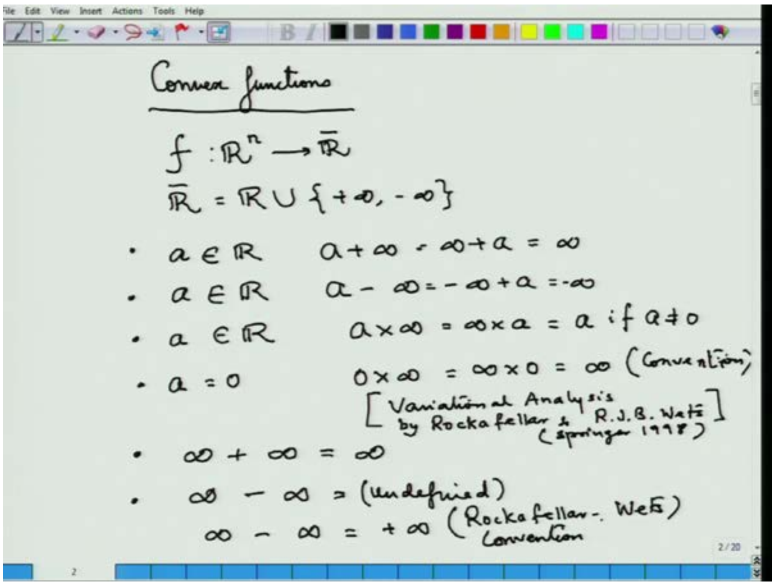
(Refer Slide Time: 00:47)



So, these types of sets are called Polyhedral Sets. So, Polyhedral set say p can be expressed as an intersection of finite number of half spaces. So, there is a i where i belongs, i runs from one to k or m or whatever some index basically. So, if you look at a, So, for example, if you take something like this or this, so this for example would be a Polyhedral Set, right, so the feasible sets of linear programming problems which we have already studied are always Polyhedral . So, l p problem feasible sets of l p problem are always Polyhedral.

So, these are very very important thing, and hence knowing the properties of Polyhedral sets are important when you study optimization. But, we will not put all our effort in studying the properties of Polyhedral Sets at this moment, but rather concentrate on studying having a broad view of Convex Sets and Convex functions and a specific property of Convex Sets like Polyhedral which would be discussed when an when it is actually required.

(Refer Slide Time: 02:50)



So, now I will start talking about convex functions. Now, I told you in the last but one lecture that you can define a convex function over whole R n or you can define a convex function over a Convex Set c. So, it means there that certain Convex functions which cannot be defined over the whole space cannot be defined over some parts of the set can be defined only on a particular Convex Set belonging to R n. So, which, but how do I unify this in a common frame work, and how do I say that I will define every function; every Convex function on whole of R n rather than defining some for on R n, some on a particular Convex Set c. So, this will force us to introduce the notion of an extended valued function. So, R bar is nothing, but the standard real line R union to elements plus infinity and minus infinity, of course if you take any a as a real number you have a plus infinity is equal to infinity plus a, this is one rule which is quiet natural.

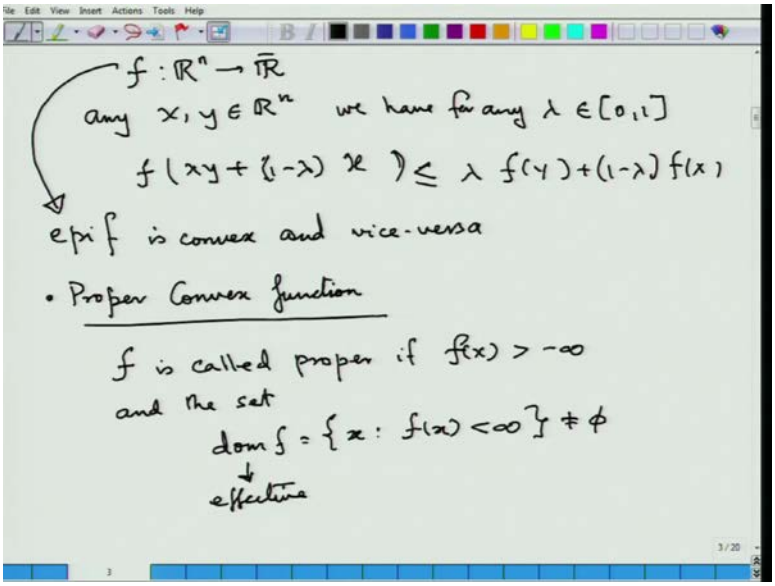
a is element of R and a minus infinity, minus infinity plus a is equal to infinity, sorry minus infinity. So, these are standard rules which you can make out this adding, if I keep on adding something very big to a number it will only grow, or if I subtract a large huge amount form a number it will only keep on reducing. So of course, the Questions are with multiplication, so if a is element of R, a into infinity is equal to infinity into a then what, of course, it is equal to a if a is not equal to zero. See when I am taking a is a is element of R means I am not considering these two numbers, right, it is only here.

Now, the Question is if a is equal to zero what would happen when Convex analysis, among the Convex analysis there are controversies in what to take, what is the meaning of this thing zero into infinity, what does is what does it mean. Now you must be wondering that can we put it zero, some say no, for convex analytic point of view possibility infinity is more meaningful. But we would follow the assumption or the rule taken by Rockefeller and wets in their famous book variation analysis and we will consider it to be infinity. So, this convention So, these are convention, this is not a mathematical certainty, its convection, is a conventions sorry. So, this is taken from the book variation analysis by the famous (( )) Rockefeller and R j b wets, a leader in stochastic optimization, one of the world's greatest these two are; obviously one of the world's greatest people in this area.

Now, of course, this you might ask me a publisher, its Springer nineteen ninety eight and there is a corrected their lot of errors, some pending errors which is corrected in two thousand four print. Now, the Question comes what do you mean by this, it is not very difficult to think it is just infinity, but what do I mean by this, in most cases it would be undefined in most books, you will see that there are no definition, all these are undefined. we When we study convexity we will hardly have any chance to face this situation.

But in convexity we cannot just forgot about this aspect the reason is very simple, if we want to define Convex functions in the traditional way as the definition of Jensen then in order to maintain the inequality we need to have some convention for this, again infinity minus infinity is defined as plus infinity this called the Rockefeller or wets convention, have you seen.

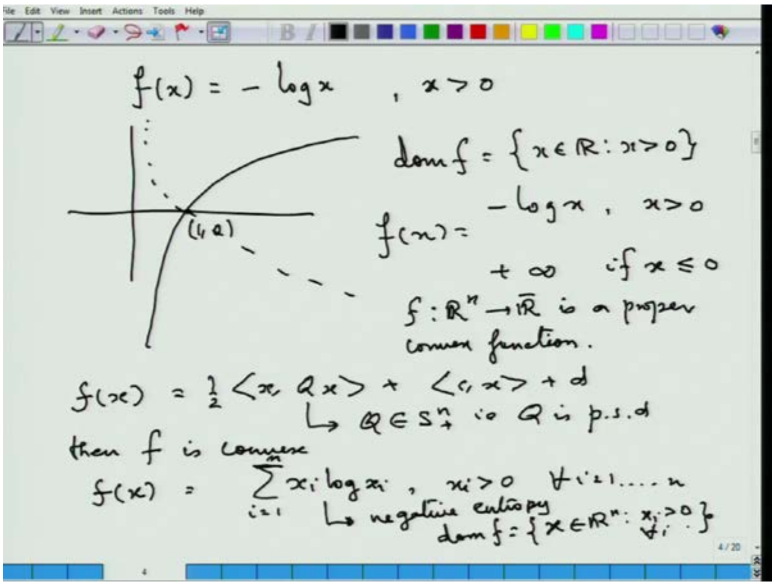
(Refer Slide Time: 08:42)



So, I can now take a function from R n to R bar and define it in this fashion, so for any x y in R n we have for any lambda belonging to zero one f of lambda y plus one minus lambda f of x, sorry one minus lambda x to be less than or equal to lambda f of y plus one minus lambda f of x. So this definition we already know, but you see this convention, so this is not a t this is plus. See what would happen, suppose if f y becomes plus infinity and this becomes, f x becomes minus infinity, say f y is plus infinity in f x becomes minus infinity then in order to maintain this inequality this side should always be bigger than this side you have to invoke this Convex Rockefeller wets convention.

So, but this is a technical issue we need not be too much bother about it, of course epi graph of f for if this function needs to be Convex, epi graph of f is Convex and vice versa. So, you might also define for such a class of functions epi graph of f to be is the definition of convexity, I do not want to tell you what is epi graph of f, if you already know this. Now, there is a important definition of a proper Convex function. f is called proper if f of x is always strictly bigger than minus infinity, and the set which is called effective domain of f that is set of all x as a f x is strictly less than infinity, that is set of x where f x is finite. So, there is at least one finite point, this cannot be non-empty, this is called the effective domain. So, a convex function is called proper if this happens, so we will largely deal with proper convex functions. Now, you see how this definition would be effective, now this is this is unified the way we can look at Convex functions.

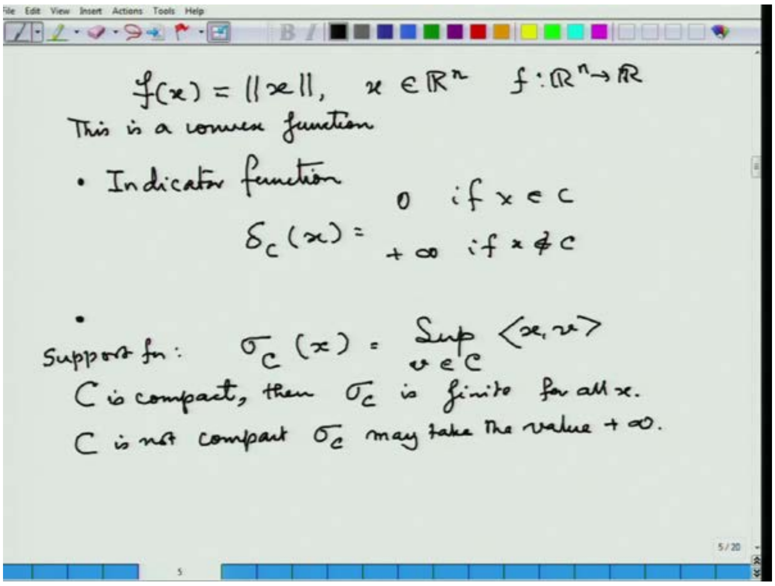
(Refer Slide Time: 11:47)



For example, if I define, look at the function f x is equal to minus log of x where x is greater than zero, now you just look at log x, if you look at log x you will observe one very simple thing that if you take the epi graph it is not convex, but if you just take the negative so it will be something like this, then the epigraph would be convex. So, this function f x is a Convex function, but it is not defined over whole of R n. So, in this case I can write dom of f is a set of all x in R such that x is strictly bigger than zero. So, I can define the function like this, f x is equal to minus log x when x is strictly bigger than zero and I can define this as plus infinity if x is less than equal to zero.

So, you see I can define log function in this specific fashion and this is the convex function, so f from R n to R bar is a proper convex function. Of course, there are Convex functions which are defined from R n to R, they take only finite values, for example a very important class of function is the following, so this is the quadratic problem, and this is the quadratic function. Now assume that this Q belongs to s n plus that is Q is p s d; positive semi definite, then f is convex. But I am mind you that if you want to directly prove it by the definition of convexity it will be very difficult to prove this to be Convex, in a certain ways, I think in different ways one can prove different function to be Convex (( )) if you take this function. So, now look at this function, this is an important Convex function called the negative entropy function, so again its domain is R n; the interior of R n plus. So, domain of f set of all x in R n such that x i strictly greater than zero for all I, may be some more examples.

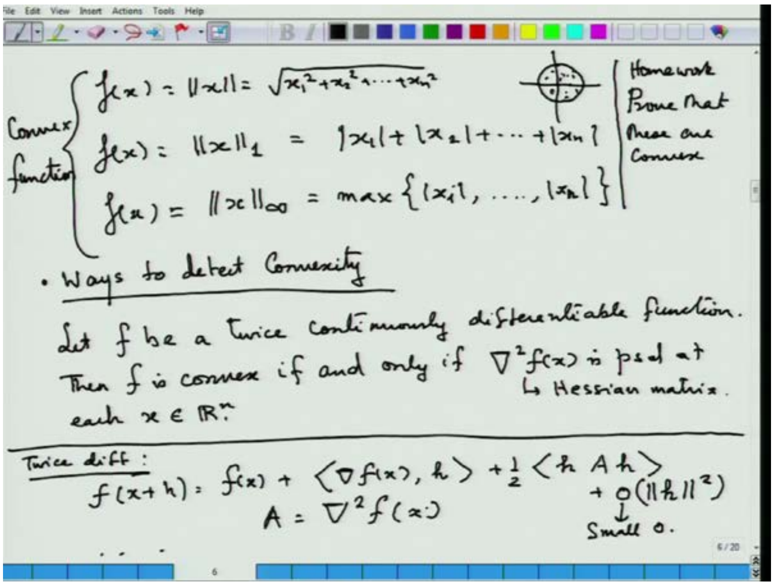
(Refer Slide Time: 15:17)



For example, this standard function f x is equal to norm of x which measure the distance of the vector x from zero, so this is of course a function from R n to R and this is Convex. Another important class of convex functions which is very very useful when we do optimality conditions is an indicator function. So, the how Indicator function, please be careful to note that this is not the characteristic function of a set which takes the value one if the element x is in the set, When it takes the value zero if the element x is not in the set. Here, it is like a penalty function which tells you that if x is in the set I do not charge you any money or any penalty, but if x is not in the set I penalize you heavily by telling that your value becomes infinity, that this is zero if x is in c, of course c is a Convex Set and this is plus infinity if x is not in c.

Another class of functions for which are very important in Convex and optimization is the support function of the Convex Set c, our set c is always Convex, we are not going to repeat every time this is a Convex Set Convex Set Convex Set, because this Convex optimization function is Convex Set is Convex, there is nothing else. Support function over set c is supremom x v, so your minimizing a linear functional over the set c, a Convex Set c. Now if c is compact, that is closed and bounded then just note a very simple fact then because this is a continuous function of a compact set then this would become finite because you will have finite value, the supremum would be finitely attend, this is a very fundamental fact from optimization. C is compact then, or what happens if c is not finite that is Question, sorry c is not compact very very sorry, if c is not compact sigma c may take the value plus infinity.

(Refer Slide Time: 19:05)

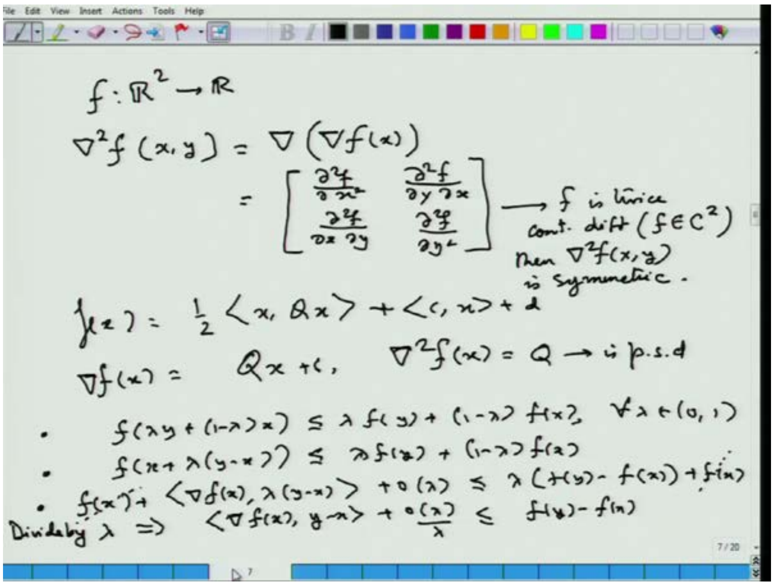


Now let me give you some other examples which could be interesting, see apart from the norm when we have define this norm, when we say there is a Convex function we are basically expecting into write this function, basically it is a Euclidean. Now in this case the unit ball define by this norm, it is norm of find everything x as a norm of x is less than equal to one would be this round ball which we view one norm. Now, this is called the one norm which is defined as the addition of absolute value of the components, so its also called the taxi cap norm on the Manhattan norm, another all of these are by the way Convex functions, all of these are Convex functions and very very important Convex function, it is called the infinity norm which says of course, homework prove that these functions are Convex.

Now, note this fact that how do you detect whether when a function is Convex, basically for the coordinate case I told very difficult to immediately pin point and say that these the Convex function or other if when you try to calculate it out in standard method trying to prove the Convex in equality it is not such easy game. So, here let us find ways to detect convexity, a very important way is as follows; let f be a twice continuously differentiable function then f is Convex if and only if the hessian matrix is p s d or positive semi definite at each x. So, this is I am talking about a finite value function only or if you want over the Dom f.

Now, of course I have not told you what is the meaning of twice differentiability and what is the meaning of this term which I am calling as the hessian matrix. So, what you mean by twice differentiability, let us segregate out and talk about twice differentiability. So, function is twice differentiable if you have this sort of an expansion, there is a matrix a such that this happens plus small o of norm h square; small o, that is if I divide by norm x square, norm x square goes to zero, this one that ratio goes to zero. Now, a can be proved by simple calculations to be the hessian matrix at x, now what is this, what does this hessian matrix look like.

(Refer Slide Time: 23:15)

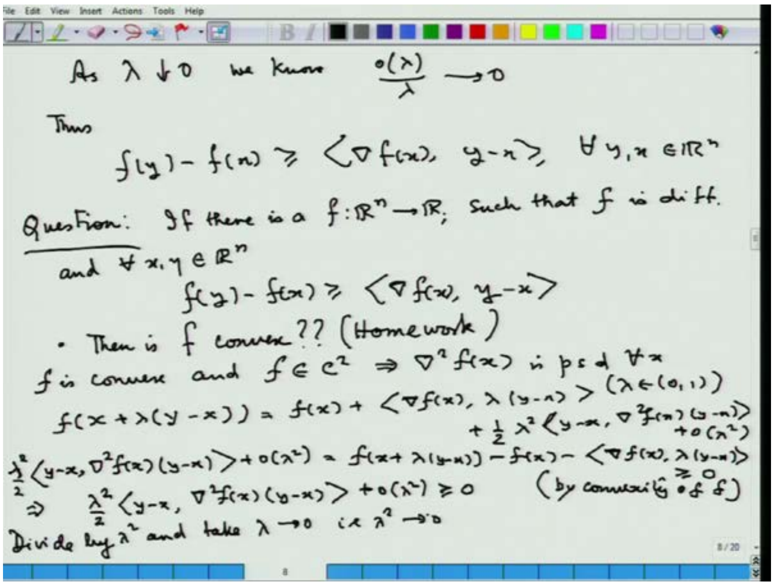


So, if f was a function for simplicity from R two to R then f the hessian matrix at x y, basically is defined in this way for those who know Jacobean matrix is the Jacobean of the vector value function grad f. So, this can be written as follows, del square f del x to del square f del y del x del square f del x del y del square f del y. You see if this all of these things are continuous that is when we say that the function is twice continuously differentiable we mean that all of these are continuous, all this second order mixed partial derivative. So, which means that if that happens then this must be equal to this by young’s theorem and then this matrix become symmetry. So, as in most cases when we are having twice continuously differentiable functions this matrix for f, when f is twice continuously differentiable which we write in short as f element of c two, then grad square f x y is symmetric and hence we can talk about p s d positive.

Now, you see when I am talking about this I am talking about also the gradient naturally because this tell as the expansion, but remember one thing now if I need to talk about this characterization, we need to talk about the characterization; the first order characterization in terms of the derivative for a Convex function because you see if I use this characterization then what would happen if you take a quadratic function f x. This half is not really required this is only for you to know making a making things looks good. when you take the derivative let me first do the gradient, the gradient of f x in this case is Q x plus c and grad square of f x is nothing but the matrix Q, so if Q is p s d if this is p s d positive semi then this thing actually holds that it is positive semi definite for each x in R n and hence that will show that if Q is p s d the quadratic function is convex, but if, but if you really want to know the proof of this then we need to talk about how the gradient of a Convex function is related to the Convex function itself.

So, suppose we want to tell something more about the Convex function, when the function is differentiable, in that case what we will do is that we will do take this following thing, let us look at this. So, let us look at f of lambda y plus one minus lambda x is less than lambda f y plus one minus lambda f x; this is obviously true for all lambda in the open interval zero one, actually because the problem is convex. I can rearrange this a bit and write this as f of x plus lambda y minus x is less than lambda f y plus one minus lambda f x. But again assuming that this function is differentiable just once, I do not bother about twice at this moment, we can write this as f of x plus the gradient of f of x times lambda y minus x plus order of lambda is less than equal to lambda f y minus f x; I am rearranging this part, plus f x and you know I can cancel out this part and now divide both sides by lambda, so I will cancel of this f x with this f x and now divide by lambda. So, lambda is not zero, between zero and one to get grad f x y minus x plus o lambda y lambda is less than equal to f y minus f x.

(Refer Slide Time: 30:20)

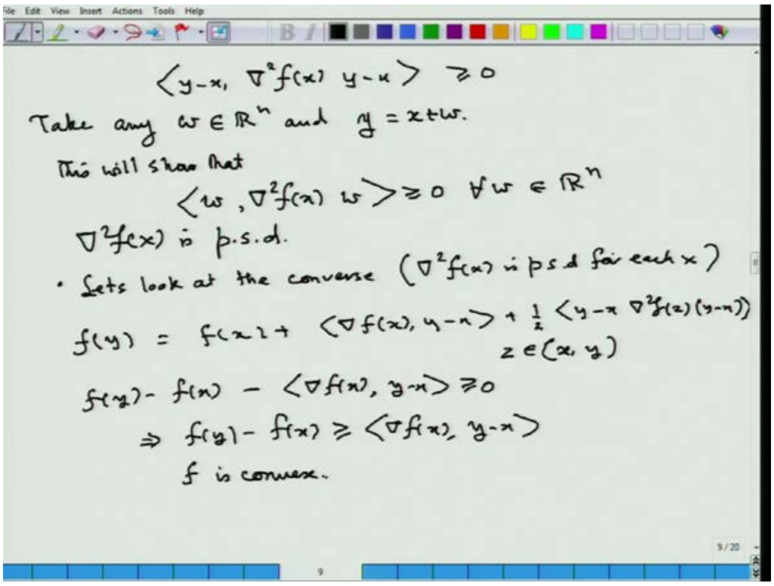


Now you know what we had done we just have to take lambda going to zero , so, as lambda move to zero, that is lambda becomes smaller and smaller we know that lambda is positive in order to zero, o lambda by lambda is also going to zero of course, those you do not remember here more categorically which should be o lambda of y minus x o, but does not matter because if you have a lambda term, multiplier term, scalar multiplier then it becomes the o term of the lambda itself.

So, this should show thus f of y minus f of x is grad of f x into y minus x, and this is true for all y x in R n, of course I am assuming the differentiable function, I am assuming it to be over R n to R. Now what about the converse, if there is the function Question now if there is a function from R n to R such that f is differentiable, which I am writing in short as d I f f and for algorithm all x y element of R n we have f y minus f x, the Question is then is f convex. So this is homework, I will give you the answer tomorrow, for the timing let me try to prove that if I have a Convex function then the easy which is twice continuously differentiable, that is if f is Convex and c two, this implies that the hessian matrix this is p s d for all x. So, now we are going to prove this statement, so we know that f is convex, so let us just write it down.

Take two points x and y, now instead of writing the convexity definition here we will assume that we will make use of this one that f is in c two and write this as f x plus the gradient of f x lambda times y minus x plus half lambda square y minus x grad square f x into y minus x plus small o of lambda square. Now look at this fact, now here is where we will use convexity, so I can write this whole thing as half of sorry, or lambda square by two times y minus x grad square of f x y minus x plus o of lambda square is equal to f of x plus lambda y minus x minus f x minus grad f x into lambda y minus x. But if you look at this, just from here you can immediately know that this whole thing by the convexity of f is greater than zero by convexity of f. So, what remains here is lambda square by two y minus x grad square f x y minus x plus o lambda square is greater than equal to zero. Now, what do I do with this, divide by lambda square, let me we choose a lambda of course between zero and one, this is the standard trick, divide by lambda square and take lambda going to zero, that is lambda square also going to zero because this is just positive quantity.

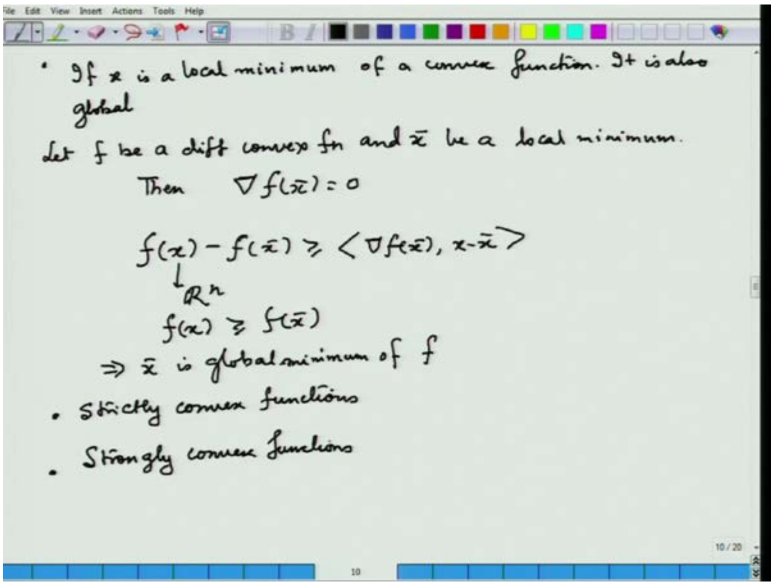
(Refer Slide Time: 35:26)



And this would imply y minus x grad square f x y minus x, now take any w in R n and set, because this y is any R n, y is equal to x plus w, this will show that this thing, this inner product is greater than equal to zero for all w in R n. This second shows that grad square f x is p s d by definition, so this is just p s d by definition, and you see if you observe here we have done the whole thing simply, in fact instead of this if you get you could have write it f of x plus lambda w y minus x is w and do the whole thing, it would have been the same. We did this because it much more easy to view, and because it is linked with convexity.

Now let us look at the converse, now you take any x and y and then let us see what happens because this is given to be p s d for each x, so if you those who know about tell us expansion for them it will be much simpler to understand f y is equal to f x plus the gradient of f x into y minus x. So, if the order term that comes after the second order expansion can be pulled into the second order expansion by writing this as half of y minus x grad square f z y minus x, where z is a element strictly inside the line segment connecting x and y, right. So it is not the interval x and y, but the line segment x and y, so z is written as lambda x plus one minus lambda y where lambda is strictly between zero and one. So, what I can now write is f y minus f x minus grad f x y minus x, this becomes greater than equal to zero because this is greater than equal to zero where z is any element in R n and this is positive semi definite, and y minus x is an element in R n, so this is just greater than equal to zero. So, this whole thing is greater than equal to zero showing that f y minus f x is greater than equal to grad f x into y minus x and y and x are arbitrary pairs, just arbitrary element, so this is true for every y and x. You see this is we have we have asked you the converse that if this happens whether it is Convex, answer is easiest and you really have to figure it out. So, you have to figure it out how is it convex, so once if I have function which satisfies this you can be sure it is Convex, so immediately you have the fact that f is Convex.

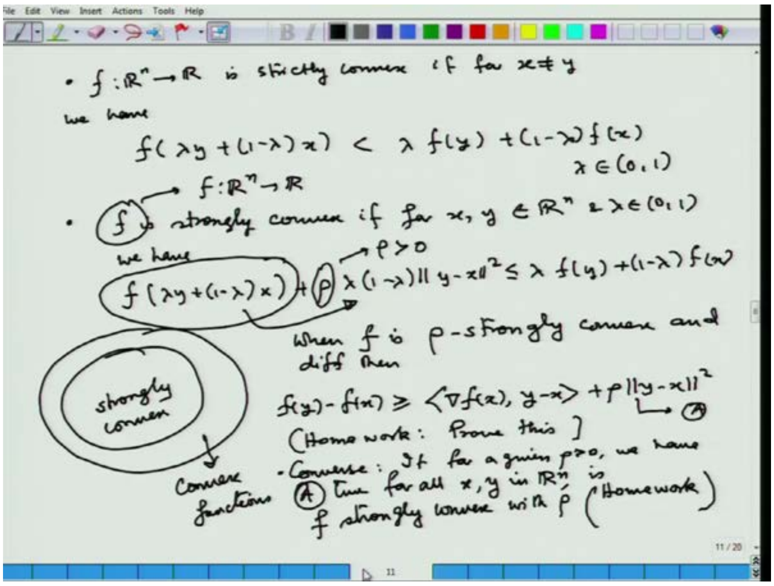
(Refer Slide Time: 39:26)



Now, very important issue from the optimization point of view, now we know that if x is a local minimum of a convex problem Convex function it is also global. Now let f be a differentiable Convex function and x bar be x bar be a local minimum, now we have already said that if this happens there is a local minima and the function is differentiable and the gradient of f at x bar is equal to zero. Now, what happens when we have a point that satisfies grad of f x bar equal to zero which is, the local minimum point will satisfy this one, but if the function is Convex for any other x in R n we can write the following thing. So for any other x in R n, whatever R n you take or whatever x in R n you take, not whatever R n sorry whatever R n, whatever element in R n you take this result is true because of problem is Convex.

But you know that grad of f x bar is equal to zero, so you can put in zero here to obtain this inequality which proves that x bar is global. So, we have f x greater than equal to f x bar showing that x bar is a global minimum of f. So, for Convex function every local minimum is a global minimum which we have proved already for a general case, now we prove it simply for the differentiable case. Now, there are two important classes of Convex functions which play an extremely important role in the solution aspects of Convex optimization problem, they are mainly the strictly Convex problem and strongly Convex problems or strictly Convex functions. So, we have to know about strictly convex functions and strongly convex functions, and by defining these two things we will stop our this course here for today.

(Refer Slide Time: 42:16)



So f from R n to R is strictly Convex, if for x not equal to y we have f of lambda y plus one minus lambda x strictly less than lambda of f y plus one minus lambda of f x, but lambda has to be now restricted between zero and one. So, what is strongly convex, we will not go into much of a discussion on this function because this is what we will need, if for x y in R n, so these are all function from R n to r. So, if you want I can just write f is from R n to R and f is strongly convex for x y x y in R n and lambda in zero one. We have f of lambda y plus one minus lambda x plus rho, so there is a rho which is fixed, so we have this expression true that this thing is not only less than equal to lambda f y plus one minus lambda f x, something more is less than because you know these thing, because we non-negative this is bigger than this part. So, which means that this part anyway it is smaller than this part on this whole part, so which means that every Convex function sorry, every strongly convex function is automatically convex, so you have a larger class of functions strongly sorry, smaller class strongly convex functions their contained in the class of Convex functions. In fact, we will show that there they are actually contained in the class of strictly convex function; we will leave it to you to prove this, so because when x is not equal to y this becomes strictly inequality.

Important feature of this is that when this rho, of course has to be strictly greater than zero, when f is, rho strongly Convex, this is called the modulus of strong convexity, rho strongly Convex and differentiable then f of y minus f of x is greater than equal to grad of f x y minus x plus rho times norm y minus x square. So it is homework, prove this. Now once you have this, now I have the have the reverse Question, converse, if I have a differentiable function which satisfies this property with some rho greater than zero satisfies every pair y and x, I have this expression true then is the function convex. So, if for a given rho sorry, if the function strongly Convex or rho strongly Convex if not convex; obviously, it is what if the function rho strongly Convex, if for given rho greater than zero we have, if I mark this as a, we have a true for all x y in R n is f strongly Convex with rho, with rho that is where there is f is rho strongly convex.

So, these these two things would be your homework and its good exercise to try it out. So, let me tell you that we have now got a very broad idea about Convex functions, their behavior, how to detect convexity for functions in c two. In fact, if this function, this strongly Convex function is twice continuously differentiable I will put on some addition of home works that you have to prove that the hessian matrix is also positive definite; it is not just positive semi definite it is positive definite. There are many aspects of convexity which can be put in homework form and which I will supply very soon.

But let me tell you that tomorrow we are going into going to very important aspect called the separation theorems for Convex Sets, that is given two Convex Sets in R n or in say R three which are disjoint that is they have no intersection that is disjoint, then you can draw a plane or put a plane in between them so that one Convex Set is in one part one side of the plane and another Convex Set is in the another set of plane, this whole thing is essentially what optimality is all about. The optimality conditions are applications of separation theorem and that is what is very very important, and optimality conditions are the soul of algorithms. So with this homework and with this little introduction to what we are going to do tomorrow I will tell you good night and good bye.

.