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

Module No # 01

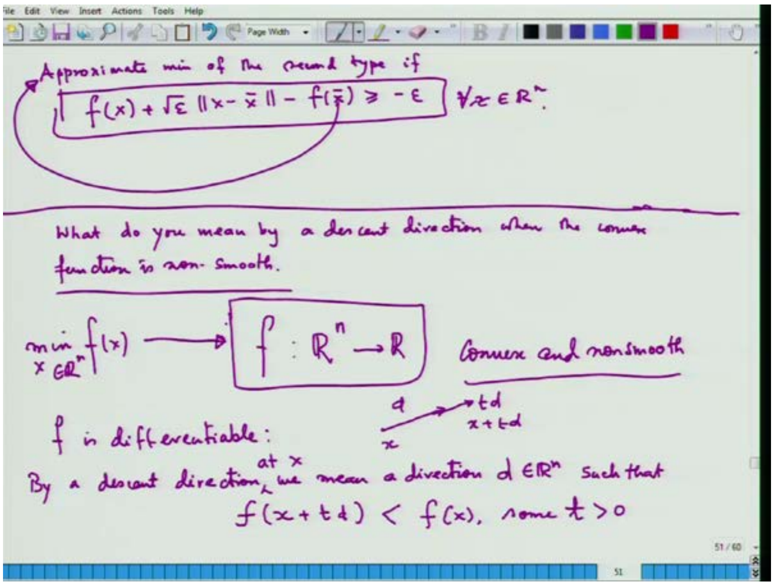
Lecture No # 37

So, we are gradually inching towards the end of the course, possibly three more Lectures in which the course would be ended taking into this lecture taking into account this lecture. So I would be having this and two more Lectures to go, and the last one would be an overall summary of what we done and what you can do from them showing you some references again. So what is important here to note that at the end of this long journey we are taking some detours. Here, we are not following a very fixed pattern or syllabus of some institute or university, so that you you can do well in the exams. This course is not for those who have some course in convex optimization or they have a course in optimization in general, but this course is for those who would use convex optimization in their work, could be engineers, and also for those who would possibly like to take of this subject as their future choice or as a hobby.

Whatever you do, when you take a trip anywhere in the world then is very very important that you also take a detour, you just do not go and see a place, and you also go and see places in the neighborhood. So what we are doing now is like seeing places in the neighborhood taking detours doing Ekaland variation principle trying to understand how to characterize approximate minima then trying to understand what do you mean by descent direction of descent when the function is non-smooth. So once we have an idea of direction of descent we will go to something more one, basically they are lot of sweets or nice food or we just taste a bit of each. So, what happens is the following that after this we are going to look at how the tools of convex analysis can be used for non-convex functions to study non-convex functions which are representable in some sense by convex functions.

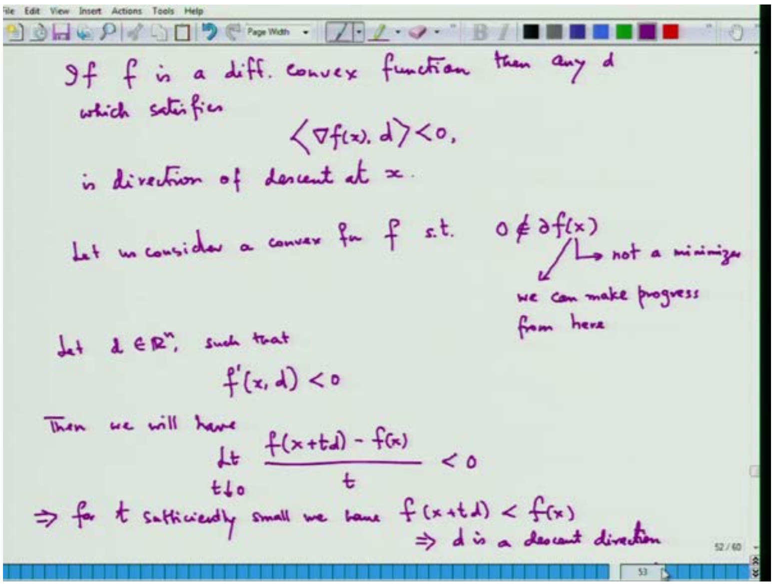
For example we will talk about the maximization of a convex function which is actually a non-convex problem. It is a minimization of a convex function which is a convex problem, the maximization of a of a convex function over a convex set is a hard non-convex problem. Because for a convex function the global maxima and local maxima are not the same we will show by very simple examples, and that immediately it makes it a difficult problem

(Refer Slide Time: 03:21)



So now what do we mean by descent here, so if f is differentiable then, let us recall what would happen by a descent direction, we mean a direction d in R n such that f of x plus t d, so it is a descent direction at x that is we are moving from the point x is strictly less than f of x for t greater than zero and sufficiently small. In the sense that if I move a little bit in that direction, this is not really a correct way to say it for t greater than zero, for some t greater than zero that would be a better way to say. So what I am doing is that if I have x here, I try to move in a direction d and I not only move in a direction d, suppose I move in a direction t d then here is my x then this point would become my x plus t d, and I really want to know that function values, I want to know that the function value here is strictly smaller than the function value here. So I have by moving along this direction I have made an improvement in the function value because we are essentially concerned about minimization. So essentially we are looking for minimizing a convex function f over R n, at this moment unconstraint minimization where f is of course, (( ))

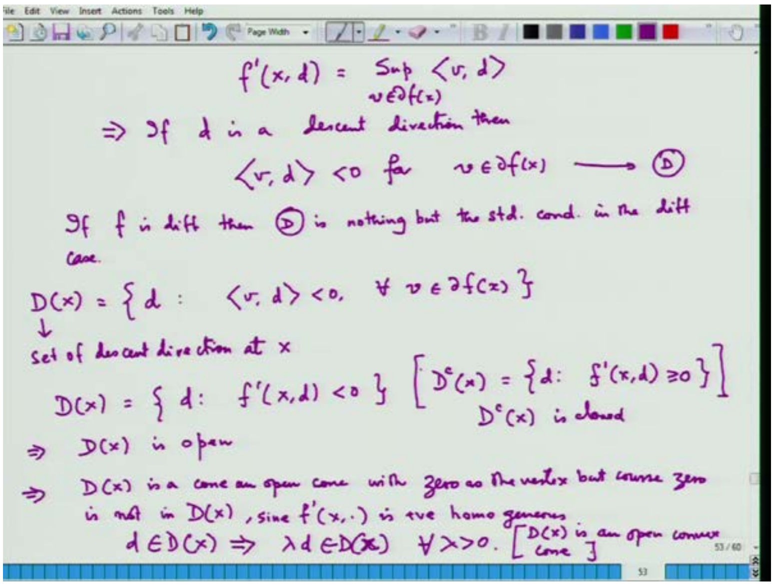
(Refer Slide Time: 05:53)



So we have already seen that if f is differentiable then there could be a verifiable condition; there could be a verifiable condition in the sense that f is differentiable differentiable convex function I will use this sort of shortcuts, so I am, you would get habituated to it or you have already got habituated to it. So if f is a differentiable convex function then any d which satisfies strictly less than zero any d which satisfies this is a direction of descent at x. Now look at the scenario when it comes to the non-smooth case, here we do not have a derivative so we cannot speak of things like this, but we can always possibly try to use the directional derivative because if convex function is differentiable the directional derivative is equal to this. Now let us consider a convex function f such that zero is not element of del f x, so x is not a minimum so we can make progress from here.

So how do you do that, one way is to assume that, to immediate this fact let d be element of R n such that the directional derivative of x in the direction d is strictly less than zero, then we will have by definition of directional derivative limit t going to zero from positive side; one sided directional derivative, f of x plus t d minus f of x by t strictly less than zero. This simply shows that for t sufficiently small; I am not going into the details of the argument, those who are mathematicians who can can immediately figure out what I am telling, those who are engineers just go back to your basic calculus course and then you would recall recognize there is basic fact that there is if limit f x goes to a is equal to l and l is strictly less than zero, then in a neighborhood of a without the point a there is a function neighborhood of a in which for every x the function value of f x is strictly less than zero two. Similarly in the same way we are writing it in an informal fashion which is done in advance mathematics for t sufficiently small we have f of x plus t d, so this would imply that d is a descent direction.

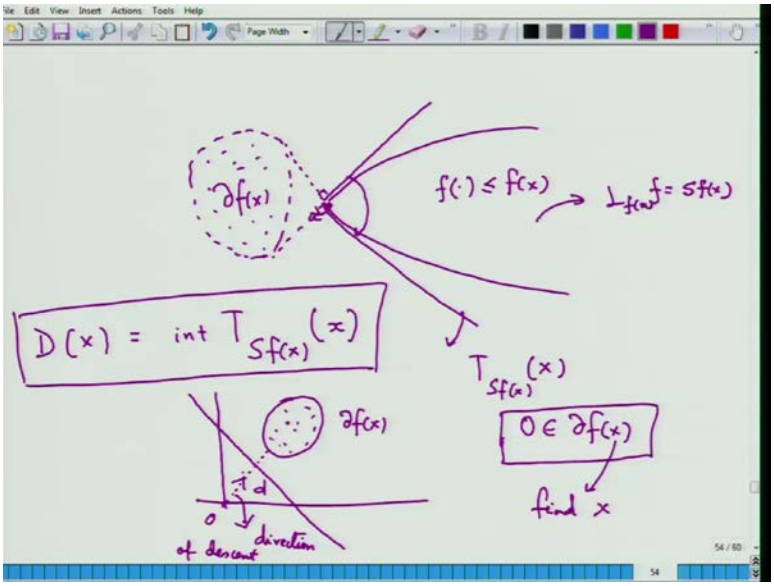
(Refer Slide Time: 10:30)



So if d is a descent direction then there is also because of the fact that the there is a relation between the sub differential and the directional derivative which says in fact it is max. This would simply imply that if d is a descent direction d is a descent direction then v of d would also be strictly less than zero for all v element of del f x, of course this is possible because of the factness of del f x, this is the compact convex set so there would be a v here such that v of d would be this, and this could be strictly less than and hence the rest of them. Now this is quite different from grad f is strictly less than zero, you observe that if f is differentiable then this condition of descent condition d in d is nothing but the standard condition in the differentiable case. Once this is done let us draw some little pictures, we have an idea of what does a descent direction means so let me collect collect the set d such that v d is strictly less than zero for all v element of del f x. So this is called the descent direction, set of descent direction set x. Of course, we can also write D x is a set of d such that f dash x d is strictly less than zero, these are the same thing.

Now what does this tell me that what sort of a set is this, it is an open set because these are the continuous function. Since this is the continuous function if I take the complement of this set this is this is the complement of this set if or which this is not satisfied. But, this x is closed because this is a sub linear function in d, a convex function in d the directional derivative which we have learnt earlier. So this would imply the first property that this D of x is open, number two D of x is a cone and open cone with zero as the vertex as the vertex but of course, zero is not an element of that cone because if zero is an element of this cone then this will not be satisfied. So this is because this happens because f dash x d is positively homogeneous in d since f dash x d is plus positively homogeneous, that is d element of D x would imply lambda d element of D x for all lambda greater than zero. But you should also remember the D x is open is a convex set also, what we have essentially is that D x is an open convex cone.

(Refer Slide Time: 15:44)

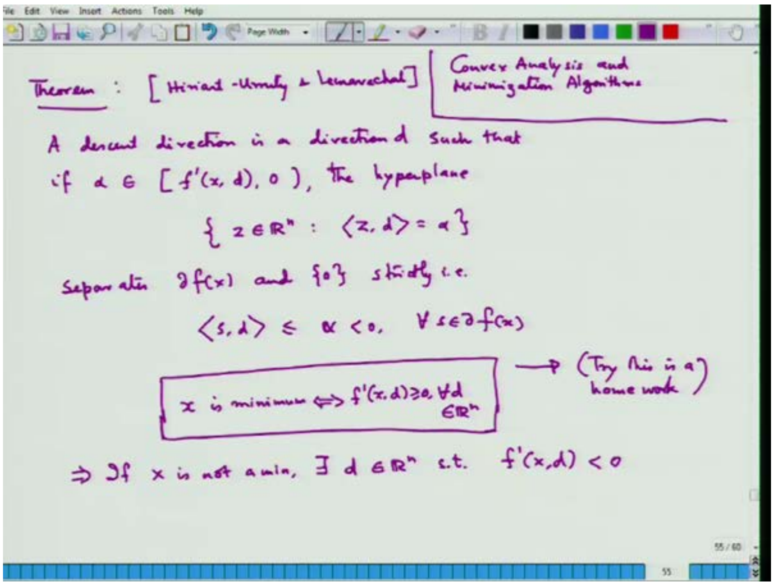


Let us draw picture about a point x which is not optimal and we are basically looking at the sub level set of that. This is not optimal then we are trying to look at x for which f of x is less than this, so the optimal I is here so it is not optimal. So this set is a sub level set, sub level set of f so we will denote it instead of this combustion notation we will denote it as s of f x for the moment or just l if you want. Now what would happen is if you have the del f x here, assume the del f x looks like this, so basically draw perpendicular here. So first draw perpendiculars at this point, draw tangent sort of thing, draw the tangent cone at this point, so this would be nothing but the tangent cone to the level set at x, and once you have drawn the tangent cone the idea is that on this you draw the perpendicular, product with this line is zero, so the del of x is the compact convex set like this. Now you observe very clearly that if you take any direction d in the interior of this particular cone, and if you take any direction here because this perpendicular line is the separating lines, so if you take any direction angle they will make is always optives, so the direction of D x is also interior of the tangent cone to the lower level set lower level set S f x at the point x. So this is also another interesting representation or a geometrical representation of the set of descent direction.

So, here is the geometrical picture, so if x is not an optimal solution this is this is the way this is how the sub level set would be related to the sub differential set. So this is the important picture and one has to keep that in mind but what, there is another way of looking at it also because you see what is happening is that basically what you want is to find a point, so the goal is to find an x such that this happens. But it is not so easy to find an x, suppose you find an x for which this is your d of x, this is your del of x, but 0 is not zero is here. So what you have to do is for example, here you draw in this case which is a nice looking del f x, you find the distance between zero and del f x, and you observe that take this direction and take any element, take the inner product with any element here in the del f x there will be the angle would be optives, say inner product would be strictly negative, and so this itself this distance can be also viewed as a direction of descent.

So what you what is this basically this is how this direction is generated that you take a hyper plane which is strictly separating zero, and basically you take a hyper plane which is perpendicular cutting this distance, then that distance right is the that that hyper plane will generate this direction d because that hyper plane would be would generated this direction d because this direction d is perpendicular to this hyper plane. So this is so this is this direction d is also is similarly, a direction of descent. So unlike the case where you have a smooth situation order then in that sort of situation there is only one way to talk about descent direction. Here there are many many many very different geometrical approaches many different approaches to talk about the descent direction, and this is the very very important way to talk about the descent direction because here for example, there is a very simple result which says the following:

(Refer Slide Time: 21:42)

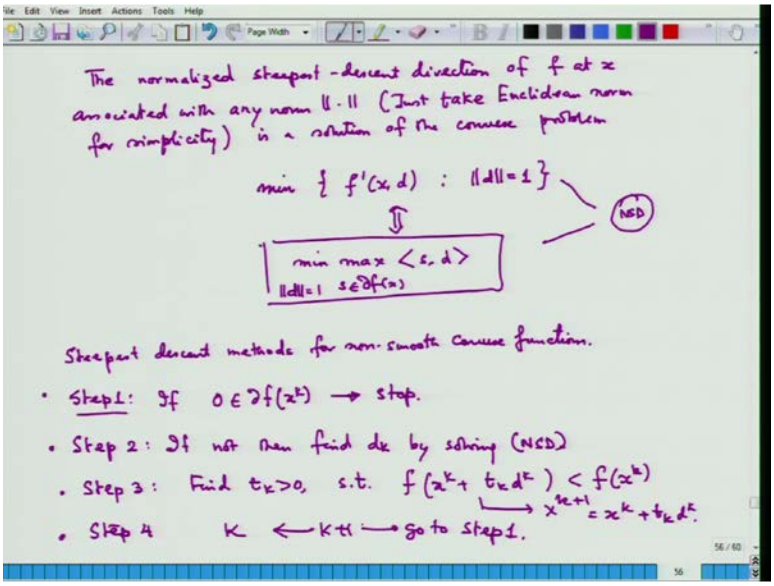


These are Theorem for this from this famous book by Hiriart-urruty and Lemarchal, it is the classic actually. Convex analysis and minimization algorithms, Springer nineteen ninety three, if I am not wrong it is nineteen ninety three, I will just check it out and tell you in a minute. Sorry its second printing is nineteen ninety six, the first printing was 1993. So these are the authors and this is Theorem from their book which says a descent direction is a direction d such that if alpha is element of, you have to figure out that this is non-empty, obviously because there is a descent direction f dash x d is strictly less than 0, so I am basically then taking this interval.

The hyper plane, this hyper plane separates del f x, so obviously here x is not a minimum thing, del f x is not equal, does not coordinate zero and zero strictly. That is s of d is less than equal to alpha, strictly less than zero for all s element of del f x which is very natural because from the definition definition of the directional derivative or the descent direction if d is the descent direction this would happen. And then if I define a separation hyper plane like this, and this is the strictly separating hyper plane, just if you know the language of strict separation. Now we are trying to solve a convex problem, basically device an algorithm we want that if I am at x, I should be able to move in a direction d till my function value keeps on decreasing, but I cannot move very little, I should be I should try to make quite a good move so that I move quite farer from x, more towards the optimal point, and my function value decreases. So this is the standard optimality condition which we will use, x is minimum if and only if f dash x d is greater than equal to zero for all d in R n. This is the standard optimality condition if you have not forgotten and if have forgotten it just go and try this as homework.

Now, once this is done what we have is the following, this would imply that if x is not minimum there exists d element of R n such that f dash x d is strictly less than zero, so there exists a direction d for which this is strictly less than zero. So in effect, what I have to do is find such a d for which the value of this is minimum among all such f dash x d I will take the minimum value. So that difference that I will get between f of x plus t d and f x would be quite large, that is f of x plus t d x plus (( )) would provide a very heavily improved value of f that it will be truly a descent. Now thus in order to do now we define what is called the normalized steepest descent direction problem.

(Refer Slide Time: 27:43)



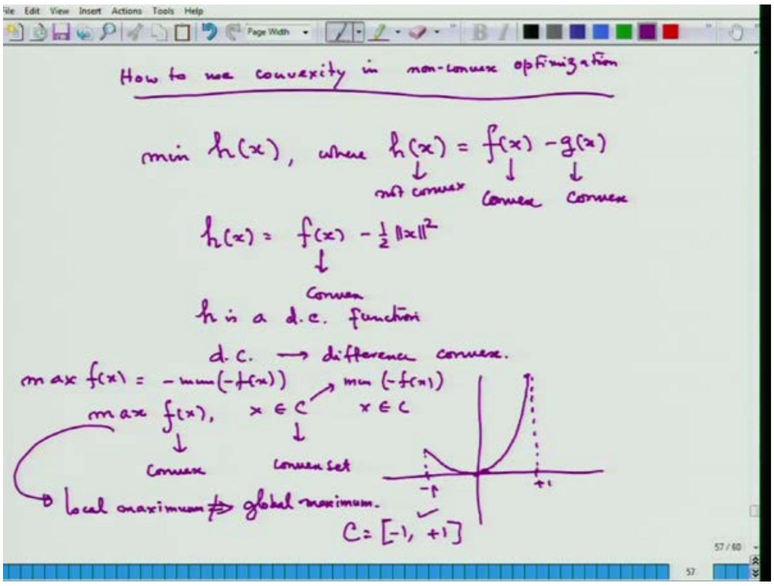
The normalized steepest descent direction of f at x associated with or any norm; just take Euclidean norm for simplicity, is a solution of the problem of a convex problem. So in order to solve a convex problem we are actually solving another convex problem, and we can of course take, because if normalize norm equal to one. So this will be a compact set over which we solve minimizing a continuous function, we can also take norm d less than equal to one, does not matter. So you can basically this problem is equivalent to the problem, so this and this, and these two problems are same, so it can be also thought about as a min max problem. Now this would this idea can lead us to what is called the steepest-descent method for non-smooth convex functions, steepest-descent method for non-smooth convex functions so let us go do this algorithm Step by Step; this is also from Hiriart-urruty and Lemarchal.

Step one is the basic stopping criteria, so we start from the k th iteration itself becoming slightly more writing in a more general way which you can understand, so all this is to give you the flavor of research in convex optimization because the basic idea of the course really ends with a basic idea of semi definite programming and that of interior point methods in linear programming. So, here what we are doing just taking this little detours is to give you the flavor of research in this area, if then stop then the solution, if not then do something. If not then find d k by I can call the n s problem normalize steepest descent or NSD, by solving NSD these are the same problems.

Now you have to find a Step size t k. (( )) Step three, find t k greater than zero, so you have to find the descent direction which gives the minimum value of this. So along that descent direction I will have to move, because I am taking the minimum value of this one, so it will be the minimum negative number, so it will be quite far from zero. So the difference between the two, this f of x plus t d minus f x would be highly negative, then actually f x would be quite bigger than f of x plus t d so you have to made a good movement. So f of x dash d strictly, descent direction does not mean that this will be, means it is enough to have a descent direction, it is good to have a normalized steepest descent steepest descent direction that is where the drop of the function value would be just steepest. such that f of

Now this is nothing but, what we write as s k plus one oh sorry I have made a blunder, not a blunder but, blunder it says that I always mentioning that whenever there is a vector I should put the subscript from the top and whenever there is a scalar the subscript is on the bottom, so this is s k plus one, this whole thing is written as s k plus one, s k plus one is s k plus d k. Step four, replace k plus one by k, go to Step one. So here so what we did today, gave a fairly good idea about developing a small algorithm, whether it runs good or not all these things are very very different issue, right. If it is not doing so, whether it is running good or whether it is running fast, quite fast or quite slow that sort of things we are not giving, we are telling that even in this difficult situation where we do not have the derivative we can still build up an algorithm which is meaningful. You can obviously develop convergence rules for this which we will not get into. So our job now would be to give you an idea of what we would discuss in the next Lecture is how how to use convex functions in non-convex optimization.

(Refer Slide Time: 35:28)



How to use convexity in non-convex optimization, there could be a convex optimization problems made out of non-convex optimization problems made out of convex functions. For example, you consider the problem to minimize x where h x is equal to f x minus g x, this is a d c function where this is called difference of convex function if f x is convex and g x is convex, there are many problems which are of this type; f x minus g x. For example, if you have any any problem of this sort, if suppose h x is of the form, if this is convex then again h is a d c h is a d c or difference convex problem; d c function, so in this case h is called d c, dc is the short form of difference convex.

So, here the components are convex functions and so I can use ideas from convexity to tell something about this problem which by itself is a non-convex problem, because the difference of two convex functions is not convex, h is called the d c function; note that h is not convex. Of course, our idea tomorrow would be to would show you what what sort of area is d c functions, arise some one or two example, and then we are also going to discuss a very important class of problems. People can ask oh! We were talking about minimization of convex function, minimization of convex function what is the great thing of minimization why not maximization, aren't they the same thing, aren't max of f x is minus mean or minus f x as a d x, that is fine. But, there are cracks because nondifferentiability brings in a lot of issues, and the nature of convex function also put puts into us a lot of problem.

So, if you take max of f x over x element of C, so here I am not talking about minimization but I am talking about maximization, this is the convex function, this is the convex set. What I have here, look look at the problem like this which will be minus one plus one and we get the problem say here, so continuous convex function. But you see here plus one, if C is equal to minus one plus one, minus one is a local maxima, plus one is where the global maxima is attained, so global maxima is not equal to a local maxima, Local maxima does not imply global minima a global maxima, local maxima for this problem does not imply global maxima, thus it is a hard problem. Any problem which where have local minima is a global minima or a local maxima is a global maxima is a soft problem or a simpler problem or a tractable problem.

It is a convex minimization where you have local minima as global minima, but for convex maximization local maxima is not a global maxima because here essentially you have what is convex function. you are talking What is max of max of f x, it is minus mean of minus f x, so this problem can be stated as mean of minus f x x element of C, but minus f x is not a convex function, it is a concave function, you're talking about the minimization of a concave function which is the non-convex function and hence the max of maximizing a convex function over a convex set is a highly non-convex and difficult problem. The problem is that, the interesting part that we will discuss tomorrow is that we can handle these class of problems using the tools of convex analysis and that will be a very very beautiful thing to see.

Thank you very much