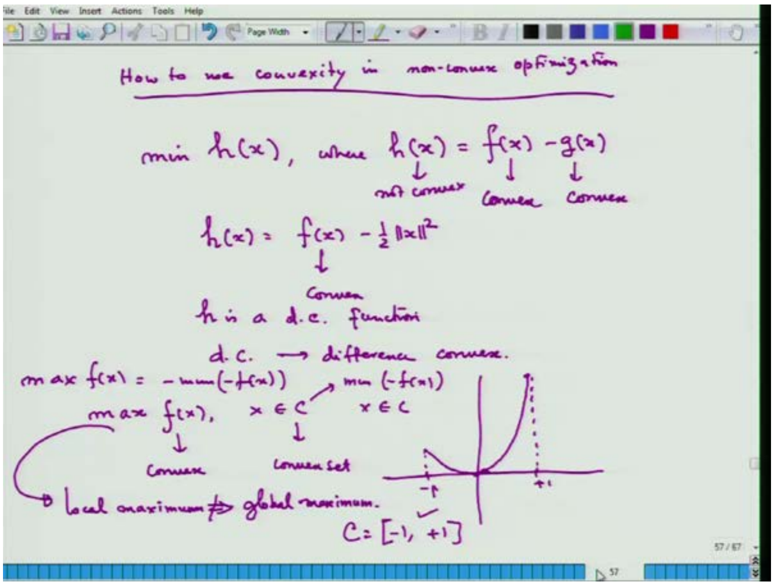
Convex Optimization

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

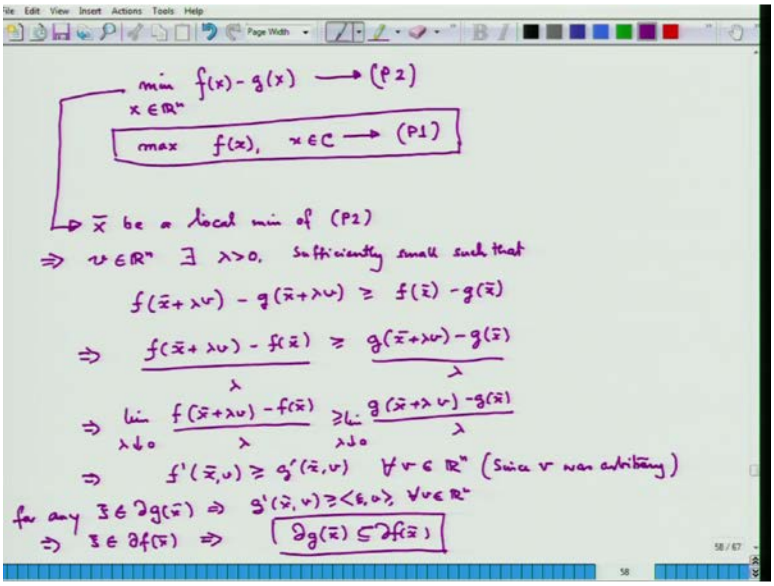
Lecture No. # 38

(Refer Slide Time: 00:24)



So, yesterday we started at the end, trying to describe what we are supposed to do today is to look at problems which are actually non-convex in nature. But they can be tackled by using the properties of convex functions, because if you observed this fact that for example, this minimization of the difference of two convex function over say x a x element of R n. Then you will have observed that here we have two convex functions. So, you can use some properties about convexity, and we have also talking about the maximization of two maximization of convex function which we have showed to be actually a non-convex problem.

(Refer Slide Time: 01:09)



So, if you look at… So, we have to basically two problems, minimization of f(x) minus g(x) for the time mean let it e e over x element of R n, and the maximization of f(x) x element of C. In the previous problem, in the problem I would call may be I should say two, two things, let me call this as p 1 this is this is what we are going to discuss in the beginning, and this let me call as p 2, because we will discuss this the next. Because… Observe there from p 1 can follow from p 2, because in p 1 if f is 0 throughout, then it is nothing but maximization minimization of a concave function, because this minus g(x) or it is maximization of a convex function.

So now, let us go to this problem of maximization of a convex function. We have shown yesterday by an example very simple drawing that a local minima local maxima of a convex function which is which is minus 1 in this case, I will I will just redraw it, make it look nicer. So, if it is minus which was minus 1 in this case is not the global maximum which is the plus which is plus 1 in the case when I am restricting C it over minus 1 and plus 1, so this is something which is very important to realize. And then see since even problem p 1 can be post as a problem p 2 by writing this as mean of minus of f(x) and then writing psi x minus f(x) or psi x is 0. So, these two problems are very strongly related or rather interchangeable in some sense.

Now, how would we do anything about them? Let us may be you are temptate to think about p 2 first, because we have already thought about minimization. So, just taking again a little bit of (( )) and listening to the dictates of may be your heart and my heart too, I would like to just show how do I do I; find the optimality condition for this problem.

Now, let us speak very clear that the optimality condition that I find for this problem is essentially of necessary condition and not a sufficient one, because this is a non-convex problem, so you cannot expect optimality condition to be both necessary and sufficient. So, let x bar be a mean of p 2, let me call it local mean that is better, we have local mean of p 2. This implies that for v element of R n that exists lambda greater than 0 sufficiently small such that f of x plus x bar plus lambda v minus g of x bar plus lambda v is greater than equal to f of x bar minus g of x bar. This would immediately bring to the (( )) the following fact.

(No audio from 04:58 to 05:12)

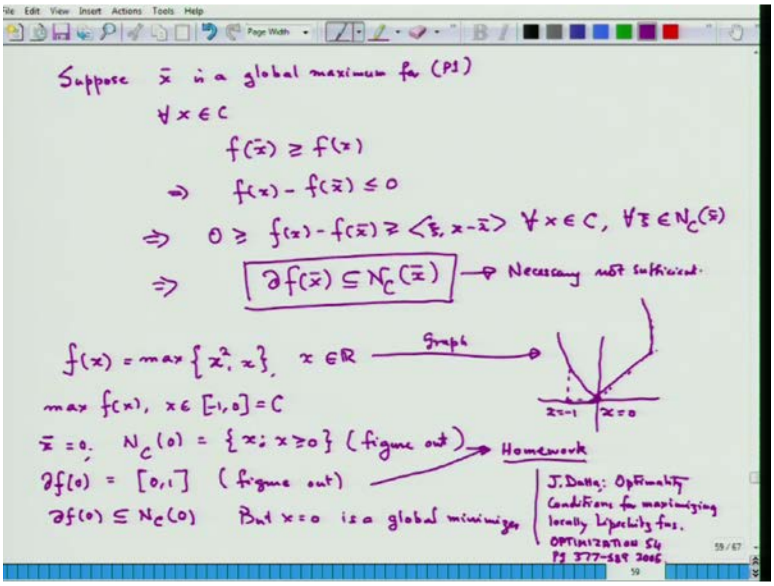
Now, dividing my lambda on both sides, because lambda is sufficiently small, and taking the limit as lambda tends to 0.

(No audio from 05:21 to 05:42)

Now, both this limits exist because these are convex function, so see we are now transfer our difficulty of non convexity to the efficiency of convexity. Here would immediately imply, and this would be true for all v, because this v was chosen to be arbitrary, so since v was arbitrary. Now, if you look at this kind of thing, this is also an optimality condition, but in the minimality form. But this would… But you see from here you cannot go back to this; you cannot show that x bar is a local minima impossible.

Now, what would what would this imply? Can I transfer it in to a sub differential condition? Of course, for any psi element of del g(x bar), it would imply that g dash(x bar,v) is greater than equal to psi v for all v element of R n. So, this would imply that psi is element… Because of this condition psi is also element of del of f of x bar which simply implies that del g of x bar is actually a subset of del f of x bar. So, this is the sub differential based optimality condition for the problem p2. So, resisting for the temptations we will just look in to this slightly you know more bothersome looking problem, a maximization one; may be our habit has become so what this 38 or 35, 37 or 38 lectures that we are seen that we have always talking about minimization and there is nothing about maximization and so things are very different. Now, how do we write down an optimality condition for maximum now?

(Refer Slide Time: 08:13)



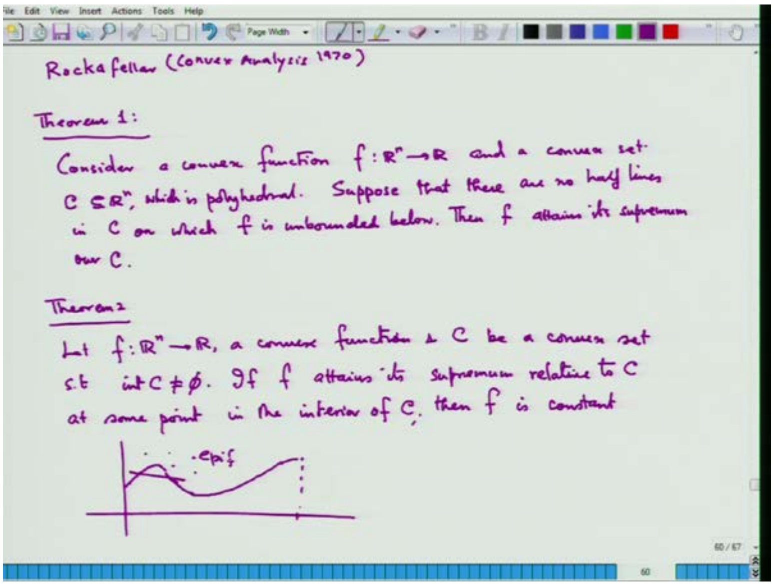
Suppose x bar is a global maximum is a global maximum for p, so global maximum for p1. Then for all x in C, f of x is bigger than f of x bar, and that would imply that f of x minus f of x bar is bigger than sorry I am making a mistake, it is maximum, so it will be opposite. So, it will imply that f of x minus f of x bar is negative. So, now again applying the definition of the sub gradient for all x element of C and all psi element of the normal contour C at x bar. This would simply imply that del of f(x bar) is N C x bar is in a subset of the normal contour C at x bar. So, this is the condition for x bar to be a necessary condition for x bar to be a global maximum of a convex function. Note that this is a necessary condition and not sufficient.

I will show you an example which will tell you that. In this example is from a paper of mind which was published in 19 sorry in 2006 which dealt with maximize this problem of maximization for convex and non-convex problems, so and and the study of their optimality conditions. So, if you look at here, if you look at this thing now, consider this problem max of this which is a convex function, I want to maximize this, x is in R and my problem now is to max this f(x) over x element of minus [1,0] which is my C. So, I am showing that I will show a point where this condition would be satisfied and that point will not be a global maximizer. So, let x bar equal to 0, so N C (0) is a set of all x such that x is greater than equal to 0. You figure it out how, figure out. You have to figure out how I have done this, I will leave it to you, but it is very simple, because it is on the real line, it is very simple to see, we just have to apply that notion of projection.

Now, if you calculate the sub differential at 0, a sub differential at 0 is on one side it is x square, other side it is x, so it is [0,1]. This also I will leave it out to figure out. I do not do this and I put both of these together as a homework. And if you look at this then it is clear the del f of 0, but if you look at the graph of this function, this is my y equal to x and here between the positive part and then it goes till 1 and goes up x square. So, the graph of the function is not (( )) like this, the graph of the function looks here and then it goes up again. So, here if you look at this point x equal to 0 is a local is a global minimizer. Here x equal to 0, this is the graph graph and x equal to 0 over the set 0, minus 1 sorry minus 1, 0, and x take x equal to 0 and take x equal to minus 1. So, on this particular set, it is clear that 0 is the global minimizer of the function. So, you can actually over the whole R it is a global minimizer. So, you can really see that this condition is satisfied, but x equal to 0 is a global minimizer.

So, this example appeared in… I will just take a second to tell you, the example appeared in a paper by myself which says which to be the title optimality condition conditions for maximizing locally Lipschitz functions is published in a journal optimization in a volume 54 pages 377 to 389 in 2005 sorry not 2006 2005. So, forgetting this, let us look at some result due to Strekalovsky which says the following. Let us tell you something which is more much more interesting.

(Refer Slide Time: 15:51)



So, we will now say a we will now show a result or mention a result by Rockefeller, it is on the book convex analysis by Rockefeller that is chapter on maximization of convex function. So, the results says the following that a convex function will never attain its minimum - local or global whatever it is, related to the set C which is a close convex set. It will never attain its local or global minimum in the interior; it will be always in the boundary, if not that function will become constant. So, here is the major result, so I will write it as theorem 1.

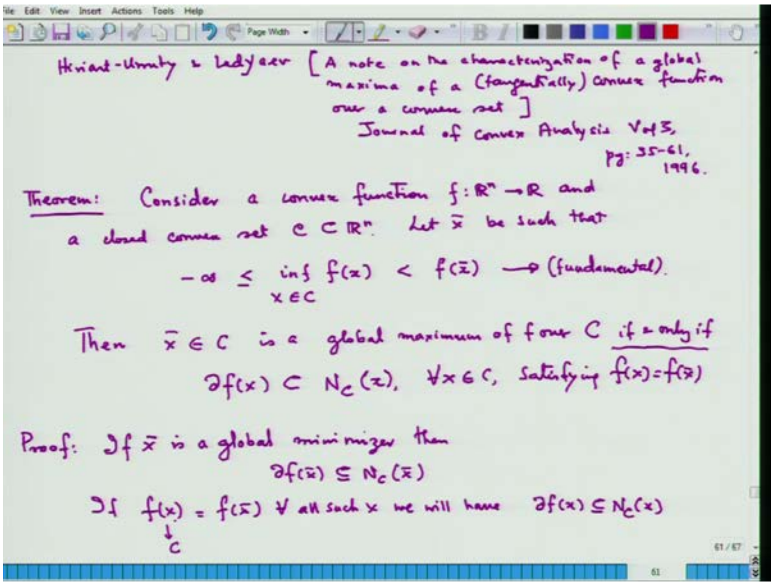
(No audio from 16:49 to 17:11)

Consider a function f from R n to R, and a convex set C - subset of R n that is polyhedral sorry not I said closed set it should be polyhedral. So, this is a reason, the… For example, here in a first example this set minus 1 to plus 1 is a polyhedral set, and that is why for linear programming problem you always have the solutions on the boundary where the linear function is about convex and concave, convex set which is polyhedral which is polyhedral. Suppose that there are no half lines in C on which f is unbounded below. These are very, very important result. Then f attains its supremum over C, so this is a condition under which it is the supremum is achieved. So, basically the you will the you you can find the point of maxima of the convex function on a polyhedral set, if this condition holds. Another result which is also from Rockefeller is the following. It is a very, very fundamental results about maximization of convex functions, is that if you have a convex function on a convex set and if if attains a point in the interior then the function must be constant.

So, let f be R n to R n, a convex function and C be a convex set such that interior of C or int of C is not equal to phi. If f attains its if attains its supremum or the maximum value whichever you want to call it relative to C at some point, it could be a local minimum also, in the interior of C then f is constant. Basically you cannot have a scenario like this, suppose you have a set like this and your convex function is minimizing, and it is ok let the convex function is having a something having a local maxima here and then of course, it the function has to come down something like this and a global maxima here. See even if it, it cannot have a local minima, once it has a local maxima you see just for a function from R to R, if the function has to raise on one side, drop on another side. Once you do that you lose the convexity of the epigraph, and hence the function would not remain to be convex. So, if the function is convex it cannot attain its supremum relative to C in the interior of C if the C has an interior. So that is the very, very important conclusion and this conclusion should be always kept in mind.

Now, of course, you can ask that can you improve this condition in some way; so that our problem, our condition becomes both necessary and sufficient that is can you guarantee a necessary and sufficient condition for local minima. We can do something with there are many, many research steps here, but we are not going to go through any of this research steps, but what we are going to do is we are going to try to mention the major result that is essentially due to Heriant-Urruty and Ledyaer, and a and a different proof was provided by myself in the paper which had mentioned in a different approach was taken. And it was done for certain different class of functions, not exactly convex, but some slightly generalized version of a convex function. So, let me mention this result due to Heriant-Urruty and Ledyaer.

(Refer Slide Time: 22:29)



So, I will also, this is the paper which they published long back in. So, both of them are leading optimizers, and those who really love mathematics, and want to become convex analyst to mathematic optimizer should read papers by Heriant-urruty and his books. So, again give me a second so that I will let you know exactly what is the details of the paper, so the this paper the title is a note on the characterization on the characterization of of global maximum maxima of a… You do not to have bother much, this is only for people who want to go and read this papers.

See when a course like this which is a quite important for application, but also mathematically very interesting is done at the end we should give in some research flavors. It is not just some course material done, so that you pass exams. There should be some courses in the this NPTEL category which would also give you some sort of research flavors. I think most courses would give you some sort of research flavor. So that is very importance. At the end of this course we are trying to give you some research research flavor, and that is why we have started mentioning papers here rather than this giving names of books.

Convex function over a convex set; now, this paper was published in journal of convex analysis in volume 3. So, every journal has a volume and every volume has some numbers, and then papers are published in each of this numbers on a particular volume. So, I am not writing the numbers, but but just the volume, because I did the traditional way to write, so pages 55 to 61 and it was published in 1996. So, my paper was published almost 10 years later, but there will be slightly different things. So, the theorem is as follows. Of course, I will try to show you how to prove this result, and that could be an interesting way to end our discussion for today. So, consider a convex function f from R n to R and a closed convex function and a closed convex set sorry convex set. Let x bar be such…

(No audio from 26:04 to 26:22)

This is what I have that the infimum there is an… Let x bar be such that it is strictly bigger than infimum; of course, if I am trying to (( )) as a candidate of maximum as a maximum that I (( )) maximum is achieved then this should at least be true alright. Then x bar element of C is a maximizer; I am not talking about a global maximizer, because I want to get a global maximization condition that is very, very important, because at the end I want a global maximizer. I will go back and again put a homework on you. I am giving to many homeworks. So, if I have a if I had a if x bar was a local maximum, can you device an optimality condition?

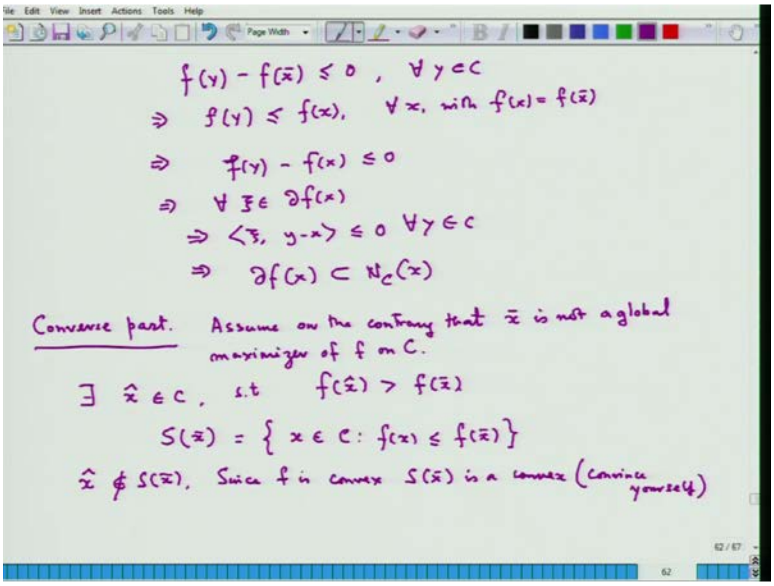
(No audio from 27:24 to 27:46)

Sorry, but try it out, it will be fun. Those who do not want to listen to all these things as I am talking about flavors of research and you feel that you have a got some idea of certain things which you can apply in your work, you may just hang around and see something and may not even bother to have a bother to concentrate much. But it is not harmful to have a look at things. C is a maximize of is a global maximum of f over C if and only if. So, it is a now it is an (( )) and sufficient condition; del of f(x), now the condition would look quite tough, but there is no other way it seems, for all x element of C satisfying f(x), but is only point which satisfies this then also this will be true. So, what it says? If there is only one x for which f(x) equal to f(x bar) when there is no other x other than x bar.

Then also if we have del f(x bar) subset of n x N C x bar, then also f(x bar) would be strictly greater than f(x bar) would be true. Then then if this condition satisfied when x bar is the only point for which this is holding true and there is no other x for in C for which f(x) equal to f(x bar), then also we will get x bar as a global maximizer. If you look at looked back you will try to see this condition that x equal to 0 does not satisfy this basic result. So, this basic condition required. So, this is this is important, this condition is fundamental. So now, maybe I will start doing the way I had proofed it proved it. So, I will go through the proof step by step. In mathematical courses, it is quite instructional to go through the proofs. So, with the with this proof I will be possibly ending today's talk and come to the minimization of d c functions tomorrow.

Now, you have observed in a statement that this statement says it is if and only if, so it is necessary and sufficient. So, if x bar is a global minimizer, of course this condition is obviously true right, unless a function is constant this is obviously true. Then what we have already proved that del of f(x bar) is subset of N C x bar. So, if x f(x) is equal to f(x bar) then for all all such x we will have, of course x is in C x is in C. See f(x) is equal to f(x bar) does not mean that f(x) is equal to f(x bar) does not mean that x is equal to x bar.

(Refer Slide Time: 31:42)



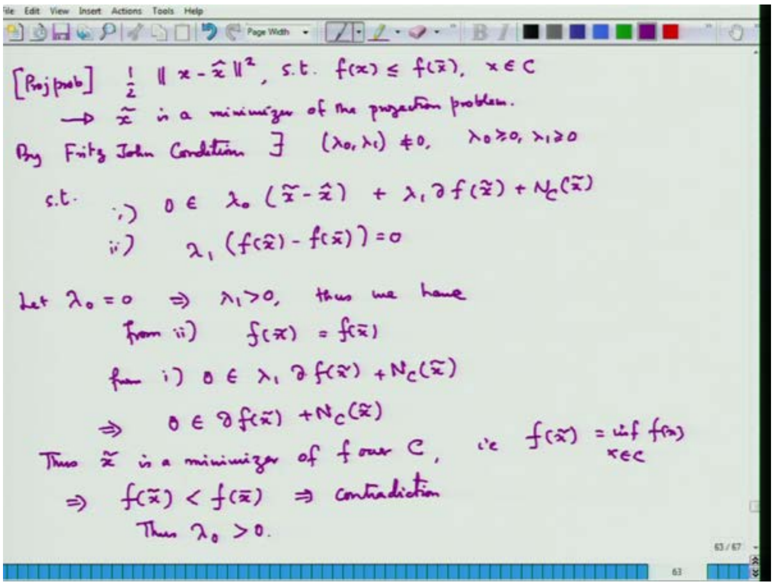
So, what we have said that now, if f(x) is equal to f(x bar), so basically if I have f(y) minus f(x bar). So, my x bar is a global minimizer, so this is what I have to show, so what I will do is, is that ok f. So, f(y) minus f(x bar) must be less than equal to 0 for all y element of C, so f(y) is less than equal to f(x), I have replace statement f(x bar) with f(x); whenever x is equal to f(x bar) this is true. Then again repeat the same set of argument I would had with f(x bar) and then you get this simple proof. This I could have left it as a homework, but ok.

So, now we have to look at the converse part. So, again we will use one of the biggest weapons of proving things in mathematics which is called proof by contradiction. So, which in this case we will assume that let assume on the contrary that x bar is not a global maximizer f on C, which means there exist x hat element of C such that f of x hat is strictly bigger than f(x). Now… Sorry f of x bar. Consider the following level set; s level set, so you can also write it as level f(x bar) whatever, but I am just writing in this simple form, this is usually many research papers they would use this sort of symbol other than the lab of something. So, the level set is a set of all x element of C here, because here only restricted as over C. So, one thing is clear that x hat is not an element of s(x bar). Since f is convex s(x bar) is a convex set. I am sure you remember this is very, very important statement, but statement which you have mentioned long back in the very beginning of the course, if you… I am not sure convince yourself.

So now, once this is done, let us see how much at what distance is x hat from this level set right. See these are all the values of f(x), so why I am doing this? That is the whole question. So, I know that. So, these are all the values of x which are below f(x bar). Now, I really want to find the distance of x hat from x bar, so the maximizer is definitely outside s(x bar); x hat could be also the maximizer, it may not be the maximizer something I do not know.

So, it is very important that it is important for us to know that… Of course, the minimizer of the function is lying here, and so x hat could be the maximizer, I have no idea, but I but x hat actually breaks this, because x bar we have assume not to be a maximizer. So, it is very important for us to know and estimate of how far is x hat from x bar. Is x hat is also in C, is the distance 0? Here of course, sorry Here we have shown that x hat is in C, but x hat is not in s(x bar), so what is the distance that is very important - means how far is this point x bar from x from the actual minimizer. So, we are trying to estimate that distance. So, in doing so we are trying to estimate, we are trying to solve this projection problem which you know that.

(Refer Slide Time: 36:27)



I am trying to find the distance of those set - the level set subject to f(x) is less than equal to f(x bar), x bar element of C. So, I am trying to solve this convex optimization problem, and this is a strongly convex problem over closed convex set which has a unique solution, so this problem - this projection problem projection problem. So, basically I am trying to see what is the distance between x hat and s(x bar) - the level set. So, of course, I have f f(x) the infimum has to be in s(x bar) and the infimum is of course strictly, the if the infimum is achieved the the minimizer has to be in s(x bar) and that is strictly less than f(x bar). So, it is very important to know at what distance the real minimizer might maximizer might be that is what we are going to show. So, we will get in to some contradiction when we try to estimate this distance. When we estimate this distance, we show that some contradiction will arise. So, there cannot be a point outside s(x x hat).

So, if if there is x hat is really outside C we will get a proper estimation of the distance between x hat and C, and we will reach no no contradiction. If I reach a contradiction, well I am trying to see how far is x hat from this particular level set that is how far the minimizer might be from my x bar. If I reach a contradiction there, then my basic assumption that x bar is not a maximizer is true. So, I have see the idea is like this, I have assumed that x bar is a maximizer and now I am trying to I found a point x hat, there must be a point x hat which would be strictly bigger than f(x bar).

Now, I am looking at all the values of x such that f(x) is less than f(x) f(x bar). Now, x hat is outside it I am try, so x hat the minimizer might be somewhere so outside s(x bar), so I am trying to at least estimate the distance between x bar and the set s(x bar) and x hat. If I am unable to make, if I make the estimate and run in to a contradiction means x hat could not be outside. See x hat has to be in C and hence proving that x bar is a maximizer. So, let us try to do this and see if we can run in to any contradiction.

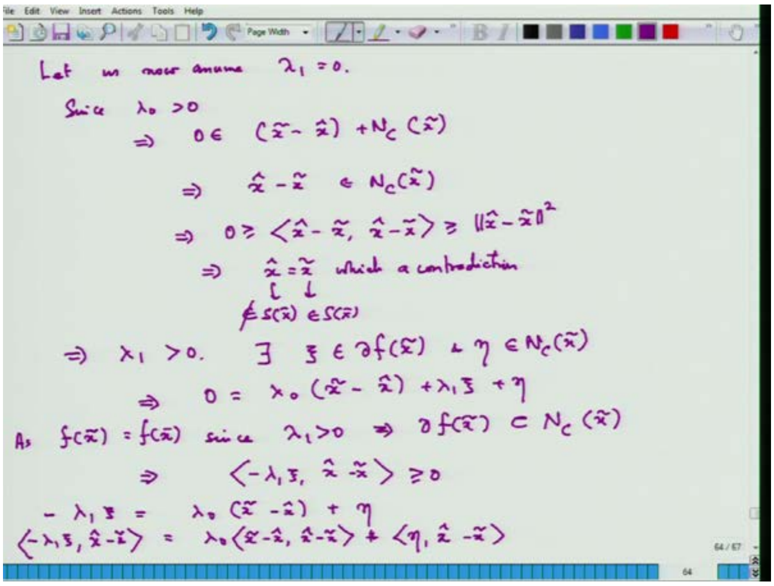
Now, again you might ask me how do I start doing this. So, let us assume we will go by using our standard Fritz John optimality condition. To do this we will apply first the Fritz John conditions. You see we have we are not aware whether there is any I do not know whether Slater condition is holding here. Slater condition holds and I can directly apply, if the Slater condition does not hold then I do not know. So, there exist a minimizer of this problem, say x tilde is a minimizer of the projection problem. Then what would happen is that I will apply the Fritz John condition or the John conditions whatever you want to say.

By Fritz John conditions, there would exist lambda naught, lambda 1 not equal to 0, lambda naught greater than 0, lambda 1 greater than 0, such that 0 is element of lambda naught x tilde minus x hat plus lambda del f lambda 1 del f(x tilde) plus N C (x tilde). See the problem is that I be cannot immediately talk about Karush Kuhn Tucker conditions, because I do not know whether if the Slater condition holds here. If I know that there is Slater condition holds here then it is direct, but here I do not know whether whether Slater condition holds and let us see what we can do further.

Now, once we do this and also it would this is a first condition and the second condition is a complimentary slackness condition which will say that lambda 1 f(x tilde) minus f(x bar) is equal to 0. Assume that lambda naught is equal to 0. Let this would imply that lambda 1 must be strictly bigger than 0, thus we have from this expression, the number (ii) expression, from (ii), f(x tilde) must be equal to f(x bar), absolutely it has to be like that, because lambda is strictly bigger than 0, this cannot be strictly less than 0 right, then it will be strictly less than 0.

And from (i), we will simply have the condition lambda 1 del f(x tilde) plus N C (x tilde), 0 belonging to this. So, this would imply, because lambda 1 is positive, I can divide by lambda 1 on both sides and the lambda 1 would be observed in the normal cone, because it is a cone. So, 0 would belong to del f(x tilde) plus N C (x tilde), so this is nothing but a necessary and sufficient condition for a point to be a minimizer of the convex function over C. So, x bar is the point. So, you see how beautifully interesting we have brought in what we have learnt in about convex minimization in the study of convex maximization. So, thus x bar x bar is a minimizer of f over C that is f(x bar) is inf of f(x). But this implies that by our first condition in the result, by this condition, this condition will immediately imply the first condition that f(x tilde) is strictly less than f(x bar) which is a contradiction. So, this would imply a contradiction, thus lambda naught is strictly greater than 0. Now, what would happen if lambda 1 is equal to 0?

(Refer Slide Time: 43:55)

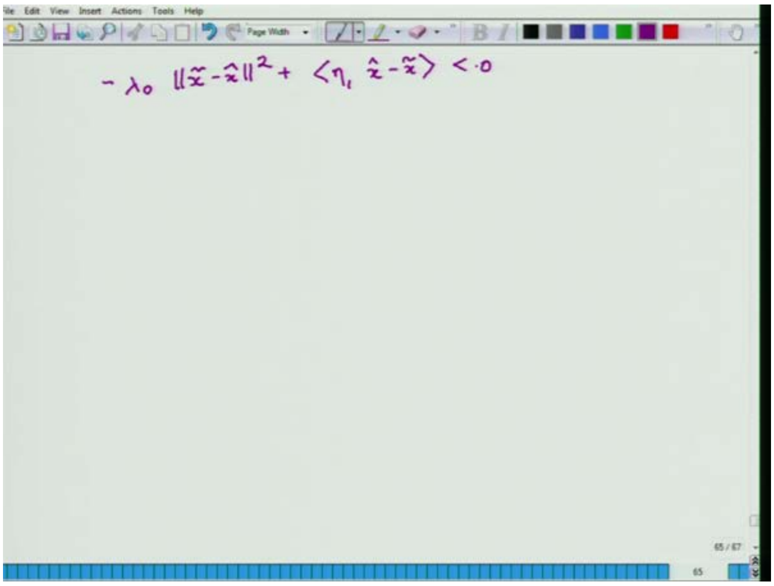


Let us now assume lambda 1 is equal to 0, so because lambda naught is a lambda naught is strictly greater than 0… Since lambda naught is strictly greater than 0, this would imply that 0 would become would be x would be contained in this set. This would imply that x hat minus x tilde is element of N C (x tilde), which would imply that x hat minus x tilde, x hat minus x tilde, x hat minus x tilde. So, this would immediately imply that x hat is equal to x tilde which is a contradiction which is a contradiction, because this is in s(x bar) sorry this is not in s(x bar) and this thing is in s(x bar), so there is a contradiction.

So, this would implies for us that lambda 1 is also strictly bigger than 0. That is great. So, what happens? So, there exists a psi element of del f(x tilde) and eta element of N C (x tilde) by our first optimality condition, it would imply that 0 must be equal to lambda naught x tilde minus x hat plus lambda 1 psi plus eta. Now, what we have said is that since f(x tilde) is equal to f(x bar) that is what we have got, because lambda 1 is now strictly bigger than 0. So, it has to be since as… So, this is now the so as you I wanted to show lambda 1 strictly greater than 0, because this is what I wanted to show.

Now, once I have this, this would imply that del of f(x tilde) is subset of N C (x tilde). So, this psi that you have is also lying in N C (x tilde). So, it would imply that minus lambda 1 does not matter lambda 1 into psi x hat minus x tilde would be greater than equal to 0. So, this is would be anyway less than equal to 0, so lambda 1 is positive. So, you take the minus it will strictly greater than 0. Now, again it is simple to see that if I multiply, now what what I did was basically I took lambda naught psi on this side, and now if I multiply on both sides, what I did was lambda 1 psi x hat minus x tilde. Actually what I did was I took minus lambda had psi I wrote this to be x tilde minus x hat plus eta. Now, I am multiplying x hat minus x tilde, so what I did was, I did multiply this with x hat minus x tilde, and when I multiplied this with lambda naught it became x tilde minus x hat, with x hat x minus x tilde plus eta times x hat minus x tilde. Now, this is less than equal to 0 this thing, we have observed that this part is strictly bigger than 0.

(Refer Slide Time: 48:32)



And now lambda naught into what you would have is this, x tilde minus x hat they cannot be equal, whole square plus eta into x hat tilde minus this. So, this is less than equal to 0, because eta is in the normal cone, and this is strictly less than 0, this is strictly bigger than 0, lambda is strictly positive, so this strictly less than 0. So, this whole thing is strictly less than 0.

So, this part is strictly greater than 0, and this part becomes strictly less than 0 that that runs in to a runs in to a contradiction. So, whatever I have assumed in the beginning that x bar is not a maximizer is wrong and x bar is a global maximizer of the problem. And hence we end our discussion tomorrow and we will today and we go tomorrow to the discussion of minimization of a d c function over a set C or over the whole R n whatever is simple for us. Thank you and hoping that you would look back in to the thing tomorrow.