Convex Optimization

Prof. Joydeep Dutta

Department of Mathematics and Statistics

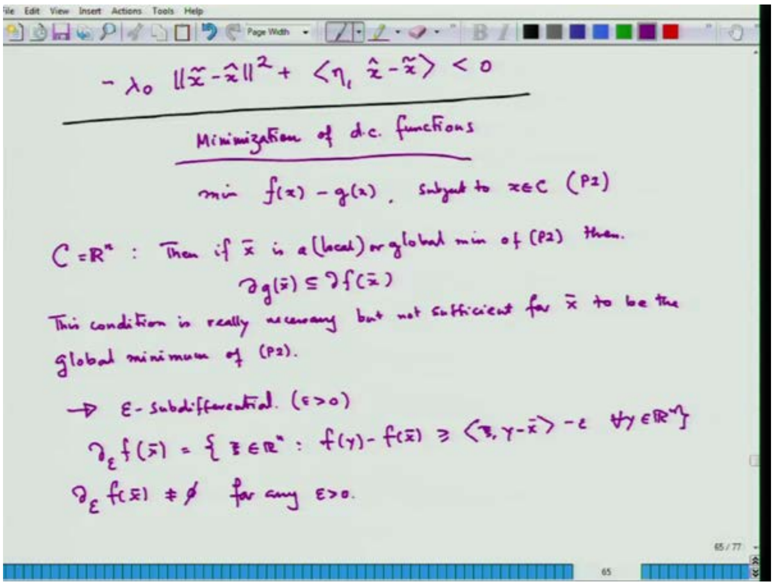
Indian Institute of Technology, Kanpur

Lecture No. # 39

So, it is almost time to say good bye to all of you, because this is the so called last truly the sorry the last lecture of this course. This course is supposed to be 40 lectures, 39 lectures are for the real course and the 40th lecture is supposedly a summarization of what is done, but I know that if I summarize things it will bore you. So, in the last lecture, I will give you what is called an entertainment lecture which would be of completely different flavor, but doing optimization at bit level which is higher then what we have been doing.

But let us mentioning that at this point what we are doing is really not a under grade level thing. It is a grade level thing largely for more for those students who are involved in some research using optimization techniques especially convex optimization techniques. So, we are going to start by talking about minimization of d c functions and minimization of this form right.

(Refer Slide Time: 01:24)

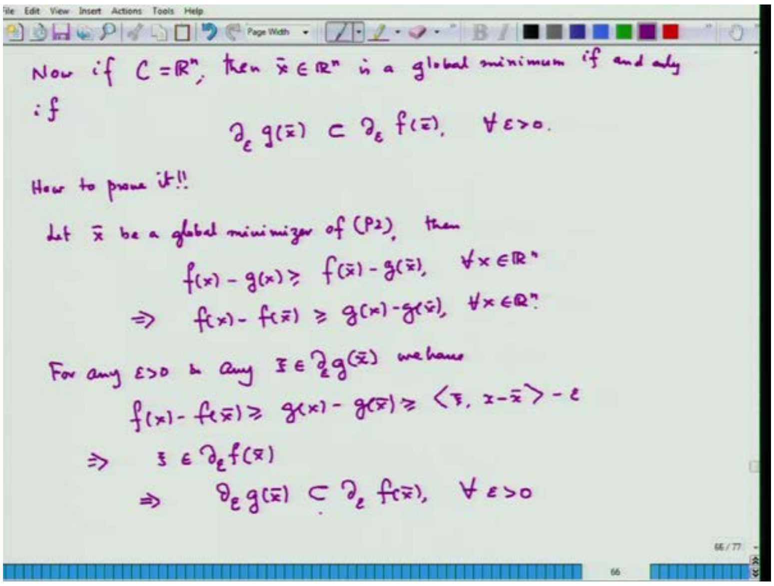


So, first we had considered the case C is equal to R n, then if f is a local or global min of this problem which we called p 2, then we know that the sorry if x bar is a local, del g of x bar. However, this condition is really necessary, is not sufficient. This condition is really necessary and not sufficient, but for convex optimization problems we have necessary and sufficient condition for global minimization. So, it is quite natural to ask that this even though these problems are non-convex, because if you have a twice contiguous differentiable function, you can write it as a d c function.

So, even if this function is non-convex, can I tell something more about it that is it possible to give at least mathematically, theoretically if an only if condition for existence of a global minimum. So, a necessary and sufficient condition for existence of global minimum, but not sufficient for x bar to be the global minimum to be the global minimum of p 2. So, what you really need to do is the ask a question, is it possible to give such a condition - a necessary and sufficient condition for x bar to be a global minimum. In that we and the answer surprisingly is easiest and our tool is a epsilon sub-differential.

And I would again like to recall you that if f is a convex function from R n to R, and the epsilon sub-differential at x given an epsilon greater than 0 is the set of all vector psi in R n which satisfies a following inequality. Psi inner product y minus x bar minus epsilon for all y in R n. It will be important to note that del e f(x bar) is always non-empty for any epsilon greater than 0. Though it has some drawbacks like even if the function is differentiable we cannot say much about the derivative here, much about that that (( )) only a single (( )) set; no, it is not necessarily. So however, it helps us in studying this kind of problems.

(Refer Slide Time: 05:08)



Now, if C is equal to R n then x bar element of R n is a global minimum of p 2. Now, we are doing only for the unconstraint case; we are not done anything for the constraint case. See we will see most of our problem is for the unconstraint case, so you might be quite surprise that we have not given any condition for the constraint case. Here giving condition for the constraint case is supposedly slightly difficult, and but it is not so difficult as we think so; we will try to see what we can do. Let us have a look at things.

First let us wait for this R n case and then try to apply. So, then if x bar is a global minimum if and if only if.

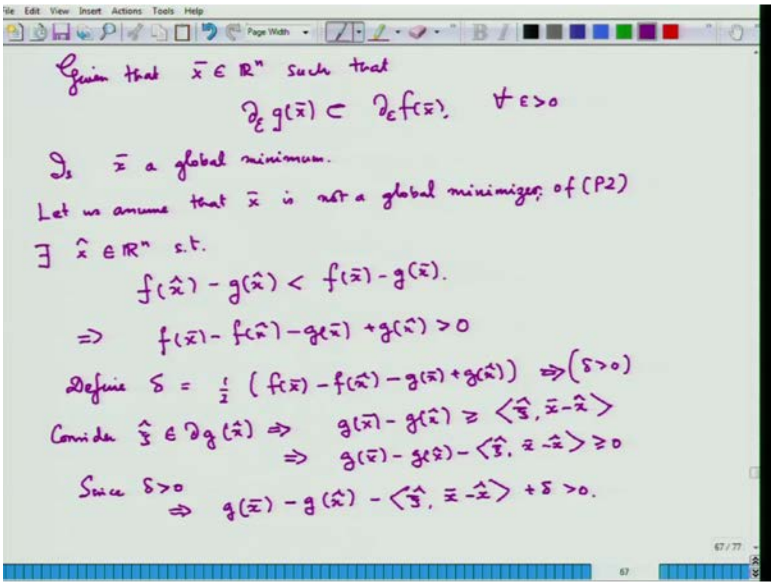
(No audio from 06:14 to 06:31)

This, now this problem as (( )), this has to be true for all epsilon greater than 0 right, for all epsilon greater than if 0 if this is true then this x bar is a global minimum of the d c problem unconstraint d c problem. So, let us write to prove this right. So, the question is how to prove it. So, let us start with the fact that - let x bar be a global minimizer of p2.

(No audio from 07:11 to 07:22)

Then f(x) minus g(x) is greater than equal to f(x bar) minus g(x bar) for all x in R n, which implies that f(x) minus f(x bar) is greater than equal to g(x) minus g(x bar) for all x in R n. For any epsilon greater than 0 and any psi element of, we have f(x) minus f(x bar) bigger than g(x) minus g(x bar) then applying the optimality condition. This is exactly same as writing now; I will just apply the definition of the epsilon sub differential which is this, which will give me a following. This would imply that psi is element of del e f(x bar). So, it will imply that del e g(x bar) is subset of del e f(x bar) for all epsilon del epsilon sub differential of g at x bar is containing the epsilon sub differential of f at x bar for all epsilon greater than 0.

(Refer Slide Time: 09:17)

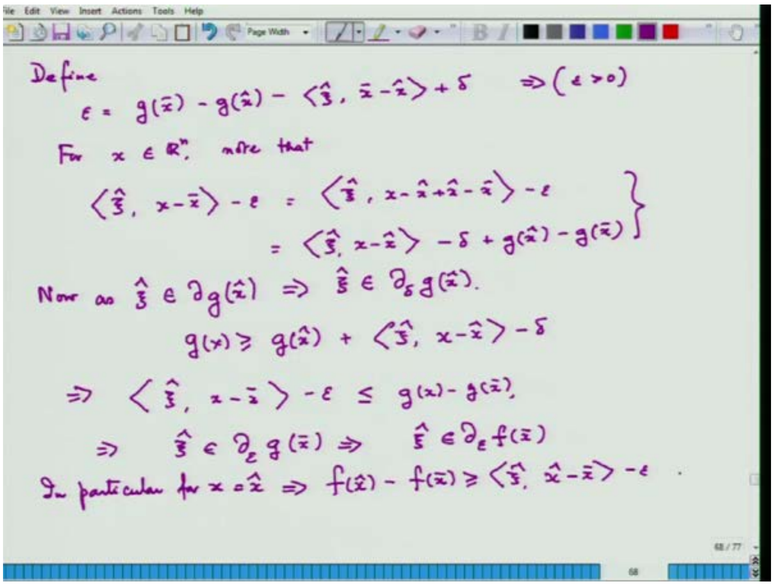


Now, the question is, given that x bar is element of R n such that del epsilon of g of x bar subset of del epsilon of f of x bar for all epsilon greater than 0. Question is, is x bar a global minimum? So, you have the (( )) becomes slightly interesting and let us see how do we do it. So, let us assume that x bar is not a global minimizer, so they there exists x at element of R n such that f of x hat minus g of x hat is strictly less than f of x bar minus g of x bar. So, now we will have this. Now, what we have done is said to be assume that x bar is not a global minimum, so we will get this. Now, how do I prove it? This is again a proof by contradiction which means that if x bar is not a global minimizer of the problem p2, then I have to show some contradiction arising. So, let me write this as f(x bar) minus f(x star) minus g(x bar) plus g(x star) strictly bigger than 0, and define…

(No audio from 11:15 to 11:22)

So, what will show? We define this positive number delta and will show that delta is strictly less than delta which is impossible. I said delta to be this, so naturally delta is greater than 0. Now, consider psi had element of del g(x hat), note it is not the epsilon sub differential just a sub differential. So, this would imply that g(x bar) minus g(x hat) is greater than equal to psi hat into x bar minus x hat. So that implies that g(x bar) minus g(x hat) minus of this part. This is clear. Now, since delta is greater than 0, it would imply that if I add delta to this quantity it will become strictly greater than 0, because to a non-negative quantity of added up strictly positive quantity, so that would be strictly greater than 0.

(Refer Slide Time: 13:06)



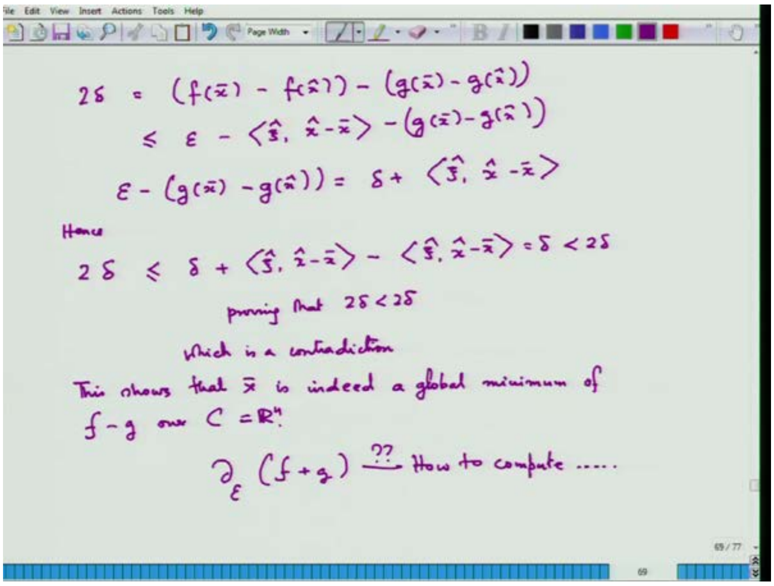
Now, this is again a positive quantity and will define an epsilon. So, for that epsilon the given hypothesis will hold. So, I will define epsilon equal to…

(No audio from 13:17 to 13:27)

Psi hat x bar minus x tilde x hat plus delta, so naturally epsilon, this should again imply that epsilon is greater than 0. This is very simple and straight for and clear. Now, for any x in R n, note that psi hat x minus x bar minus epsilon can be written as psi hat x minus x hat plus x hat minus x bar minus epsilon, which when I write down the definition of epsilon would lead to the fact that psi hat x minus x hat minus delta plus g(x hat). So, here psi hat x hat minus x bar will cancel with when I put minus will cancel with psi hat x bar minus x hat. So, what will get is this one.

Now, as psi hat is element of del g(x hat), it implies very simply that psi hat is also element of del delta g(x hat). Now, once you know this sorry you will immediately write g of x is bigger than g of x hat plus epsilon hat x minus x hat minus delta. So, this would immediately show from this above expression from this expression here from this expression that psi hat x minus x bar minus epsilon is less than g(x) minus g(x bar). This is true for any x, because x was arbitrary for x element of R n any x element of R n which implies that psi hat is element of del epsilon g(x bar). This also implies by a hypothesis that psi hat is element of del epsilon f of x bar. So… So, in particular for x equal to x hat it implies that f of x hat minus f of x bar is bigger than psi hat times x hat minus x bar minus epsilon.

(Refer Slide Time: 17:02)

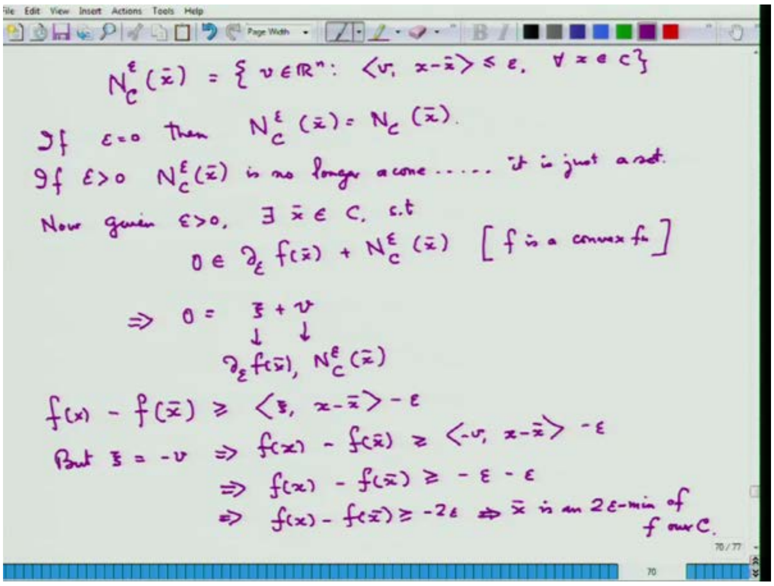


Now, if you go back to the definition of delta, then from there you can write 2 delta is equal to f of x bar minus f of x hat minus g of x bar minus g of x hat. Now, this from here, you know this is nothing but epsilon minus psi hat into x hat minus x bar, just what you got before you are writing them now is minus of g(x bar) minus of g(x hat) in bracket, so minus g(x bar) plus g(x hat).

Now, once you know that epsilon minus g(x bar) the way epsilon is defined, if you will go back to the definition of epsilon then what you get, epsilon minus this is delta minus this. So, is delta plus epsilon hat into not not it will be slightly will just (( )) of help just putting the correct sign, putting an better sign, x hat minus x bar rather than x bar minus x hat. Hence from this here 2 delta is less than equal to delta epsilon minus this is delta plus epsilon hat x hat minus x bar minus epsilon hat x hat minus x bar is equal to delta which is strictly less than 2 delta; proving that 2 delta is strictly less than 2 delta which is a contradiction. This shows that x bar is indeed a global minimum of f minus g over C is equal to R n. It is important to realize at the very instant that in this last lecture. You may now ask me the question why I am always about putting C is equal to R n for this case. If I am not putting C 2, we just a closed convex set - subset of R n - a proper subset of R n.

The answer is that yes, it can be done, it can be done and the problem is that even if it can be done, one has to realize that they we would get into certain technical it is specially with the sum of or if you take the sum of two function, basically now if I want to do for the case when C is a close convex set, I need to answer you this question, what is this, how to compute this. So, this would be getting into too much of technicality keeping in view the level of this course, such a technicality is really not required at this stage. For I can show you something quite interesting that you can use this idea what we have just proved to devise a very simple looking necessary and sufficient optimality condition for maximization of the convex function over a convex set.

(Refer Slide Time: 21:07)

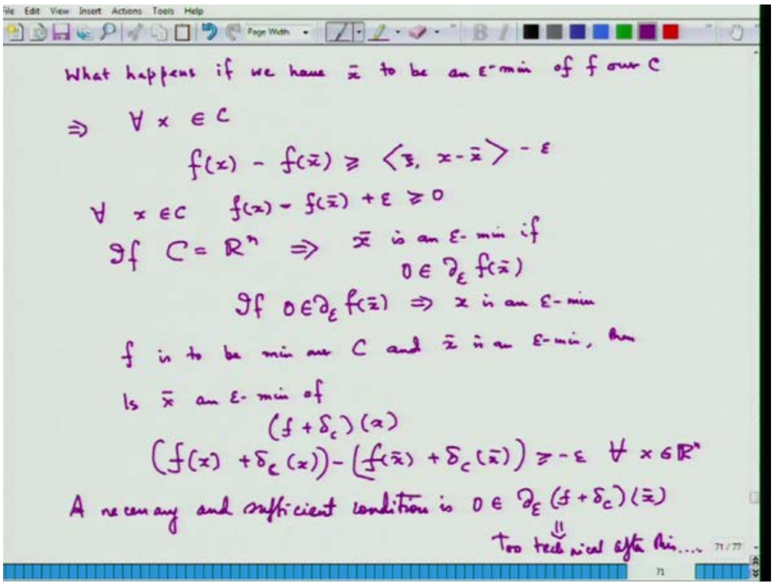


In order to do so we will introduce this idea of the epsilon normal set to C at x bar, it is a set of all v in R n such that v times x minus x bar is less than epsilon for all x in C. So, if epsilon is equal to 0, then this is nothing but the standard normal cone that we know in C x. But if epsilon is greater than 0, this is no longer no longer a cone, it is just a set it is just a set. Now, there is some interesting about it, is that suppose minus of psi, now there is suppose 0 suppose there is suppose I have the following condition; suppose now given psi greater than 0, there exist x bar element of C such that you see f is always a convex function such that 0 belongs to del e f(x bar) plus N; f is a convex function, C is a convex set, and C is a convex set it is standardized. Then let us see from here what can we get. It would imply that 0 is equal to psi plus v where this belongs to…

(No audio from 23:17 to 23:25)

So, this imply f of x bar f of x minus x bar is greater than equal to psi times x minus x bar minus epsilon. But psi is equal to minus of v, so this would imply that f of x minus f of x bar is greater than equal to minus of v inner product x minus x bar minus epsilon. But this this shows that this thing by the fact that v belongs to this; this is greater than equal to epsilon. So, this would immediately imply that f(x) minus f(x bar) is greater than equal to minus epsilon right. So, minus epsilon minus epsilon, so this would imply that f(x) minus f(x bar) is greater than equal to minus 2 epsilon; prove it implying that x bar is a is an 2 epsilon min of f over C. So, if you have both the same epsilons here what you get is 2 epsilon min. It is not the same epsilon min right. So, you see things what do you know about is very handle (( )) handle epsilon sub differentials and approximate minimum epsilon min things are slightly different.

(Refer Slide Time: 25:27)

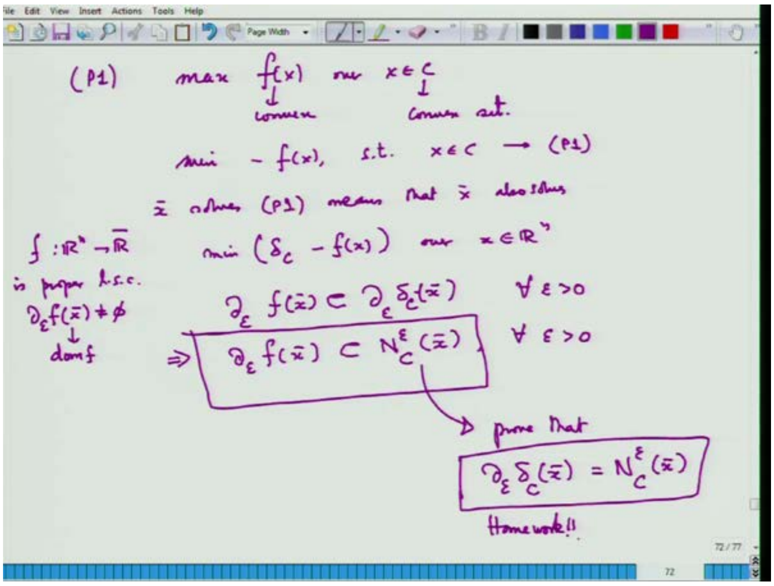


Now, what happens if we have x bar to be an epsilon min of f over C, so this would imply for all x element of C, f of x minus f of x bar is greater than equal to psi times x minus x bar minus epsilon. So, for all x it implies that f(x) minus f(x bar) plus epsilon is bigger than equal to 0 for all x in C sorry. So, this is known to be greater than 0, so if 0 is element of del e. So now, what I am having is the following. I am having that psi times x minus x bar, this one; if I put a 0 here, if 0 belongs to this you see. So, if we now from there how do I figure out some necessary condition. Suppose this is true for all x, then if I put psi equal to 0 the condition is satisfied. But then if I put psi equal to 0 this condition is not true for all elements of C, psi so when 0 element of psi they would 0. This is only true for this fact is only true for all x element of C. So, for all x element of R n, this fact might not become true.

So, if x is equal to R if C is equal to R n, now now if I choose if C is equal to R n, then it would imply that x bar is an epsilon min, if 0 is element of… Again if 0 is element of del e f(x bar) it implies that x bar is an epsilon min. Now, the interesting part is that this fact this fact is not so easy when you make the problem constraint. This analyzing epsilon minimum that is exactly what is I shown in. So, in unconstraint case, it is much more simple. So, you can say ok, then what I can do is, if f is to be minimized over C and epsilon and x bar is an epsilon min, then is x bar an epsilon min of f plus del C x; answer looks like yes, because you see if I take f of x plus del C x plus f of x minus plus f of x bar sorry minus f of x bar plus del C x bar, then observe that when this is 0.

When C is in x, this is always greater than minus infinity, minus epsilon, and when C is when sorry when x is in C this is 0, so f(x) minus f(x bar) is any way bigger than minus epsilon that is given to you. And if C is not in x, then this becomes plus infinity, so anyway that whole thing is bigger than minus infinity epsilon. So, it implies that this is greater than equal to minus epsilon for all x element of R n. So, a necessary and sufficient condition is 0 belonging to del epsilon of f plus del C x bar. Now, if I want to expand it again this question of some rule will come and here we have all lower semi continuous function and so there will lot of too technical after this.

(Refer Slide Time: 31:33)



So, we will go back to our problem of maximizing or the problem p1 of maximizing f(x) over x element of C which give again this is a convex function and this is a convex set. Let us handle this situation bit interestingly, I can pose this problem as minimize minus f(x) such that x is belonging to C. Now, if I put the phi x as the 0 function, then x x x bar solves the same as p1, p1 it means that x bar also solves min of del C minus f(x) over x element of R n. And for this optimality condition is known to us; optimality condition is del epsilon of f (x).

Now, you can say why why are you talking about del C now? Now, here the optimality condition that we have given, if you look slightly carefully can be extended to the case where this can be extended to the case where you have an extended valued lower semi continuous functions. So, if even if f from R n to R is actually lower semi continuous is R n to R bar is proper l.s.c which is the case of this function. You can similarly, for every x bar in the domain of f, you can define del epsilon x and for any element in the domain of f you will have this. Actually what you need to know is that you can now basically, if you have x bar in the domain of f does not matter. So, this this will be not equal to phi. So, it does not matter even if you have lower semi continuous function. So, even if your one function continuous, suppose your f minus g f is a lower semi continuous and g is a continuous and f is lower semi continuous, then also you will have the similar relation del e f(x) sub in bracket of del e is subset of… So, you will have this relation, this relation will hold del e, if I have f minus g del e g subset of del e f in this particular case, this the lower semi continuous function would be on the other side. So, you have just applied what you have learned.

So, what you have known, so whatever we had this this result, what we have learned here that if this happen then x bar is a global minima, when f and g are both finite valued function, but f is a proper lower semi continuous convex function. Then this is always non-empty, if x bar is an element of domain, so for any element x bar in the domain in this case C x bar is in C which is in the domain. This result will also hold true, if g is a finite valued function and f is lower semi continuous and convex, this result will also hold in a very, very straight forward fashion. The proof can be very simply modified, so I am not mention this.

So, now I will have this this thing, so for all epsilon greater than 0. So, this would imply and it will be your home work to figure out that this is nothing but this set. So, here is a neat looking necessary and sufficient optimality condition for maximization or finding a global maxima of a convex function, but you really that this is not so easy to prove, because… But is not easy to sorry it is easy to prove, but it is not so easy to compute and really verify, because the problem, this sort of conditions run into difficulties, because we have for all epsilon greater than equal to 0. Can this condition we improved, so that they are much more hand able in actual computation. So, this is a hard non-convex problem and this to improve these conditions is a challenge which is still not even answer. Sorry you can see.

So, this you prove as homework. This is very, very simple, just the definition. With this I would like to end today's talk and this course, basically ends here. Thank you for listening to this a long 39 session journey. The last talk, the 48th one, instead of summarizing I will give you a different flavor what I would say I will give you an entertainment lecture, and where all show you some very festinating aspect of quadratic optimization, and I show you how semi definite programming would play a role there. Thank you very much.