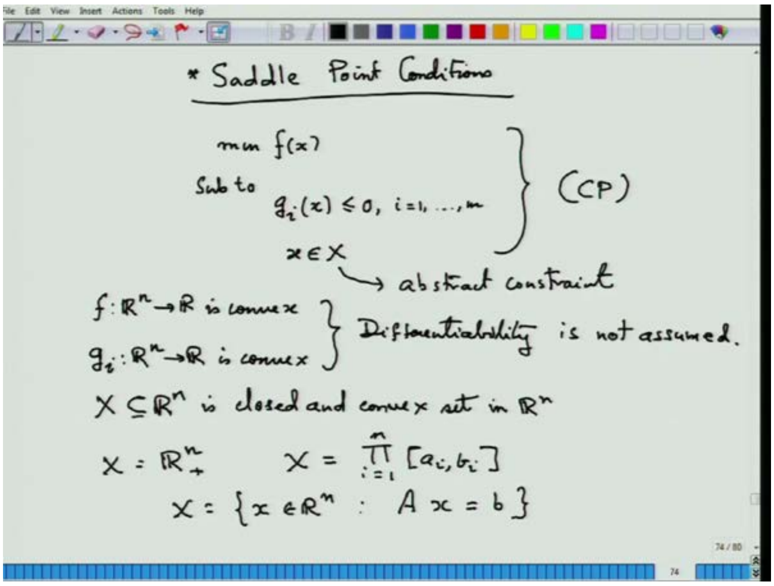
Convex Optimization

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

Lecture No. #15

So, welcome once again to convex optimization, and in this journey that we are taking together, we are taking going on the main routes as well as changing little bit of detours, which is fun I guess, and we will continue doing so. So today, we will talk about saddle point conditions as I promised in the lasts class, but also something, I will do if I have time by the end of this lecture, is to give you a list of functions and the conjugates; you can try to figure out whether, what I write down is correct or not could be your home work.

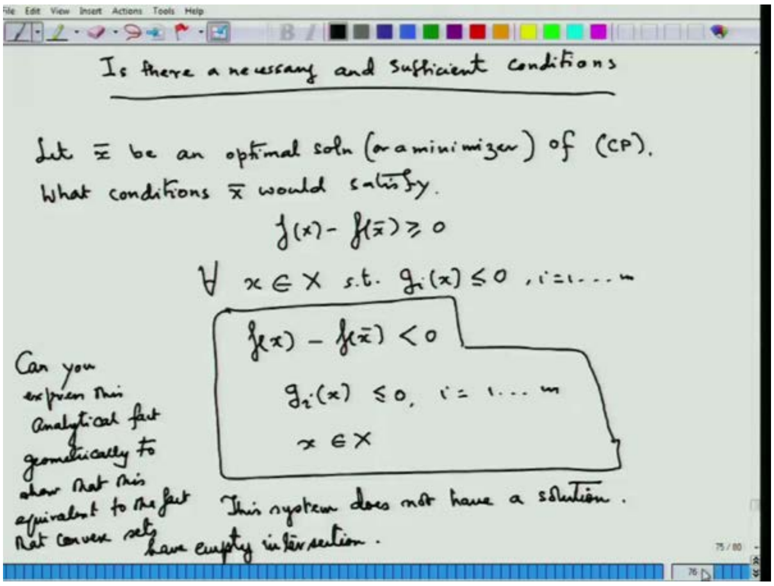
(Refer Slide Time: 01:00)



So, here again I am considering this problem minimize the function f, subject to inequality constraints and x been equal to x be an element of x, where capital X is an closed and convex set in R n; and these are convex functions f and g is, so essentially we are talking about convex programming problem CP; now these functions are not assumed to be differentiable. So, they could be just, but they are sub differentiable; obviously, they have all have a sub gradient at every point and so, this is quite a general form of convex function; now what could be x? So, x is, this is called an abstract constraint, this is you know, this is a inequality constraints; constraints given by inequalities, and this is an this is called an abstract, this constraint is called an abstract constraint, this called an abstract constraint.

So, what will be the form of this, x could be anything, x could be for example, R n plus that is it has a non negativity restriction on the variable, but x could be the Cartesian product of say n intervals in R or x could even be… So, you got a many forms of x. So, that is why it is called an abstract constraint. So, all these forms can be written at this thing, it could be additional inequality constraints, also which you just want to hide for some reason; now, so that is why we do not have any particular structure on x, when we discussed this.

(Refer Slide Time: 03:14)



Now, the question is, is there a necessary in sufficient optimality condition? (No audio from 03:13 to 03:22) Now it is okay, what is the problem replace gradients with sub differentials and say 0 belonging to something, something; some sub 0 belonging to del f x naught plus lambda del g x naught and all those things, but the question is that suppose we do not know anything about sub differentiability either, is there more fundamental way of looking at it, just looking at what means a point is about to be a minimum of such a problem, and then try to figure from there, if you come to any conditions.

Now let me tell you one thing that when a theorem gets discovered in mathematics it is done through a lot of guess and test means it is it is not that somebody comes in the morning takes a cup of coffee and write down such a writes a theorem and gives the proof, no, no no no no; it it is not like that its largely done by guess and test, someone trying to play around with something doing this, doing that, then trying to gradually put some idea in, then later on define by people and gradually a refine version comes and which we read in text book, and which gives the impression that somebody had really taken a good cup of coffee in the morning, and just came and wrote down the theorem with the proof.

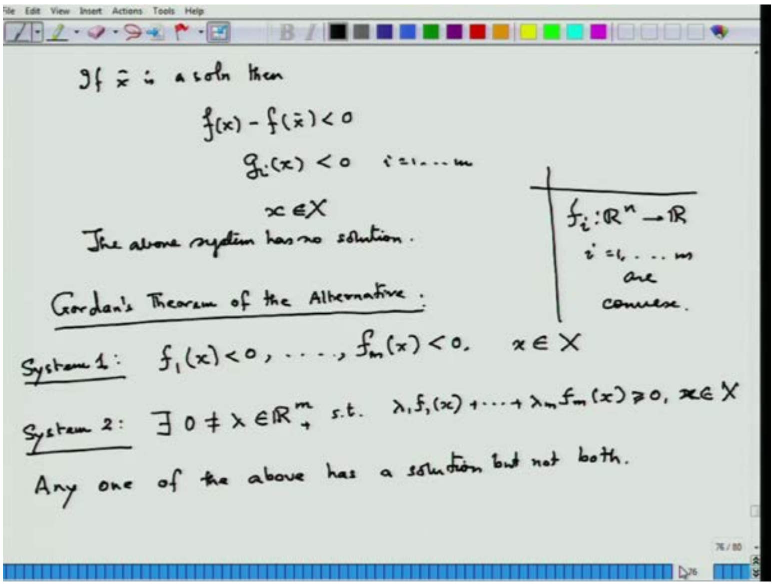
Now, suppose I am to look at this problem, and I have been given this problem, I really do not know, what to do about it; I am told that let x bar be an optimal solution, be an optimal solution or minimize whatever, because you are minimizing what condition x bar would satisfy; (No audio from 05:30 to 05:38) I only know that they are not differentiable, and I suppose, I do not know anything about sub differentiability; then is there any way, I can start off; if I am to figure such things, then possibly I would think in the following way.

How do I go about it; see, what can I say; what does it means that x bar is a solution. So, I look at the very, very fundamental things; since I have no work when in my hand to directly attack it, and look into the very, very basic thing; it is always good to tackle the basics (( )) sometimes. So, this means that (No audio from 06:26 to 06:40) this system there would for every x of this form, this would, this is what is going to happen. So, f of x minus f x bar, if this is strictly less than 0, and you have that is u, will for every x satisfying this would happen; but any x, which is satisfying this; any x which satisfies this, cannot satisfy this, because x bar is a solution of the problem, because if there is an x, which is feasible to this original problem CP and also satisfies this means, which it has a value which is better than x - f of x bar.

So, then you cannot say that x bar is a minimum, but we have guaranteed that x bar is a minimum. So, x bar minimizes the function, so over the constraints. So, this system does not have a solution; (No audio from 07:58 to 08:19) now, what is the meaning of writing that, why why I have written it like this; see the I will large issue - the central issue in optimization as per as optimality and all these things are concerned revolves around the separation theorems.

So, now I want to put everything in a form, when separation takes place, what I am going to try to show that optimality means, that some intersection of some sets are empty or intersection of two convex set is an empty set, they do not intersect. So, if optimality is reached, so x using x bar, I can define certain sets, for which their intersection for those two sets could not intersect, and as a result, which you can you can apply a separation theorem. So, instead of going in to the details, I want to figure you, I want you to figure out, can you express this analytical fact geometrically; it will be a very, very good exercise to show that this is equivalent to the fact that two convex sets have empty intersection means, basically you have to find those convex sets.

(Refer Slide Time: 10:42)



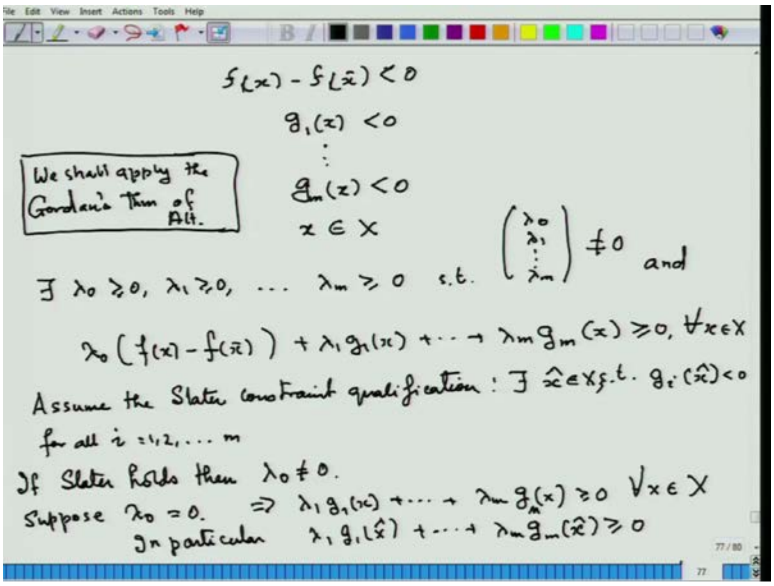
Now, in optimization theory, a lot of such systems come; will you keep on applying these systems? right what I want to look in to this further, I want to say that if x bar is a solution, then (No audio from 10:44 to 11:03) this system this system has a solution. So, you see if this system has a solution, then it is obvious that this system has a solution. So, why I have taken taken off the equality here, and made it strict inequality, because that allows us to apply the separation theorem in much easier form. So, instead of really applying the separation theorem, whatever we need to apply there can put into the form of result called gardens theorem of the alternative.

(No audio from 11:51 to 12:08)

The proof of this result is promised for the next class, not now; the fact is that what is the theorem of the alternative? You term the alternative means there are that is two things. So, there will be two systems; system one - system one of equalities and inequalities. So, what does system one say? System one says for example, for the gardens case is that I have m convex functions from R n to R; so this system means there find N x, for which all of these are strictly less than 0, all of these convex functions and there exists 0 not equal to lambda element of R m plus such that lambda 1 f 1 x plus lambda n. So, this theorem is a consequence of separation principles.

So, what does it says; there are two systems; system 1 and system 2; system one says that there is a x, for which this will have a solution; now this there will be a x, for which all these strict inequalities are valid; and the system 2 says that there exists a non-zero lambda in R n plus that is all the components has to be greater than or equal to 0 with at least one non-zero, such that this non negative linear combination of f1, f 2, f x, f m x is greater than equal to 0, for every x that you choose; now the result says any one of the above has a solution, but not both. So, once this result is given to you, use say (( )) now I have an immediate approach, I can tackle; it you see these f 1, f 2, f ns just to keep you posted in this fact that we are always using convex functions. So, this means any of the system, any, if this system there is no x, for which this is true, then this must be true.

(Refer Slide Time: 15:13)



So, in our system what we have is f (x) minus f x bar; f x bar is a fixed quantity right; this is a convex function and g 1 x strictly… So, if x bar is a solution this system, we know does not have a solution. So, this corresponds to something like system 1 here, in the gardens theorem of the alternative. So, now which means, so there would exist. So, the this this does not have a solution then the second must have the solution only any one of the our systems, any one of the above have solutions, but not both; the interesting fact is that one of them would have the solution, if this is not having the solution, then this will have; it cannot be the both are not having the solution; if one does not have the other has that is the interesting one, that is that is what it means by alternative; one of the two always has a solution, but both of them cannot have a solution at the same time.

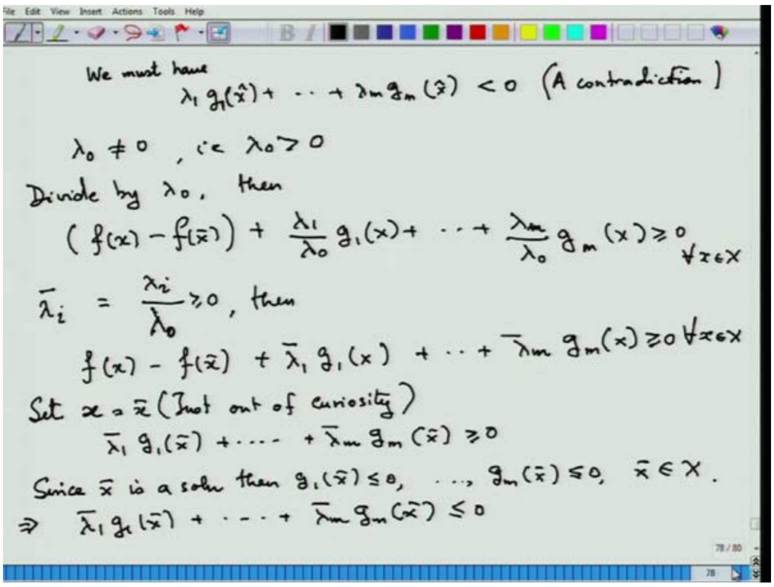
Now, so, again I want to reassert, these two systems both cannot be unsolvable at the same time too. So, if this is unsolvable, then immediately this will have a solution. So, you see here x is the variable - unknown variable and here lambda is unknown variable. So, you have to be very careful, so if this does not have a solution, this will immediately, will have a solution. So, one of them would always have a solution, but not both neither both would be unsolvable at the same time. So, there would exists lambda naught, this lambda naught is corresponding to this set, greater than equal to 0, lambda 1 greater than equal to 0 dot, dot, dot lambda m greater than equal to 0, such that the vector lambda naught lambda 1 lambda m, this vector is a non 0 vector, and such that this is non-zero, and lambda naught, (No audio from 17:27 to 17:52) so as applied the gardens alternative theorem. So, here we shall apply the gardens theorem of alternative.(No audio from 18:00 to 18:18).

So, we have essentially applying this as of always gardens theorem of alternative. So, once I have this, I do not know suppose lambda naught could be 0 here at this point, because then of course, f would go out of the description, and that would not be a fair thing to do. So, the interesting part is as follows; let me do one thing; now let us impose a condition, because we say this system does not have a solution means, if there is an x which satisfies all this, then that x cannot that x cannot satisfy this it has to be that would be greater than equal to 0 that inequality.

So, essentially somewhere we are assuming some condition like the Slater condition. So, let us assume the Slater condition or the Slater constraint qualification; now once I know this fact, then what happens that is there exists an x hat, such that g i x hat is strictly less than 0 for all i; then if this is the case, if this is the scenario, then what would happen; let us see; thus do I do a of some addition information if this happens, if the Slater condition is true, if Slater holds, then lambda naught is not equal to 0; why question is lambda naught is not equal to 0.

Now, suppose lambda naught is equal to 0, then this part will vanish; this would imply, so there exists x hat into x; (No audio from 21:09 to 21:24) now if I take, so this for every x, so for in particular for the x hat. So, in particular, (No audio from 21:35 to 21:54) now you can see what is the game, there will be a contradiction, because each of g 1 x g 2 x g m x hat is all strictly less than 0, where x hat is element of x and lambda 1 lambda 1 lambda 2 lambda n all are greater than equal to 0, but since lambda naught is equal to 0, and this vector cannot be equal to 0 competitive one of the lambda 1 or lambda 2 or lambda m must be non-zero.

(Refer Slide Time: 22:29)

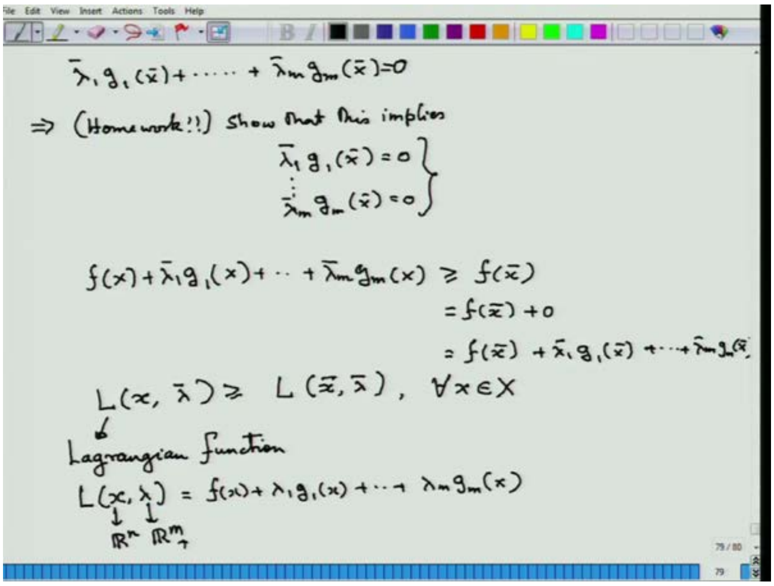


So, which means what we actually have is that we must have sorry lambda 1 g 1 x hat g 1 x hat. (No audio from 22:43 to 22:54) So, this contradicts this one. So, finally, we have a contradiction. So, lambda naught is not equal to 0, so what I now do; divide by lambda naught, lambda naught this fact, this equation. So, what I get is the following; then this lambda naught, now has to strictly bigger than 0, that is lambda naught is not equal to 0 means lambda naught strictly bigger than 0; then what we have now is, (No audio from 23:45 to 24:11) so this is what we will have.

Now, what I will do is that write lambda i bar is equal to lambda i divided by lambda naught bar right I will do this. So, then we have f (x) minus f x bar plus lambda 1 bar g 1 x plus lambda m bar g m x is greater than equal to 0, for all x in x; now said x equal to x bar, let us see what happens if I do that; because this true for every x in x, so just out of curiosity, just root x is equal to x bar; this is quiet natural to have such a curiosity, then if that is so, then this part will go.

So, I will have lambda 1 bar (No audio from 25:34 to 24:48) now because x bar is a solution, all of these are actually less than equal to 0, since x bar is a solution, then g 1 x bar less than equal to 0, because is feasible, g m x bar is less than equal to 0, and x bar is naturally element of x. So, this would imply knowing that these are, so this this would also be greater than equal to 0. So, not lambda 1 bar, its lambda naught; so you divide by lambda naught, which is non 0 strictly bigger than 0, for all these greater than equal to 0 quantities. So, it would imply that lambda 1 bar.

(Refer Slide Time: 27:23)



So, I have these two equations to compare, and this would result into a fact that lambda 1 bar, g 1 x bar equals to 0; now the home work; show that this implies,

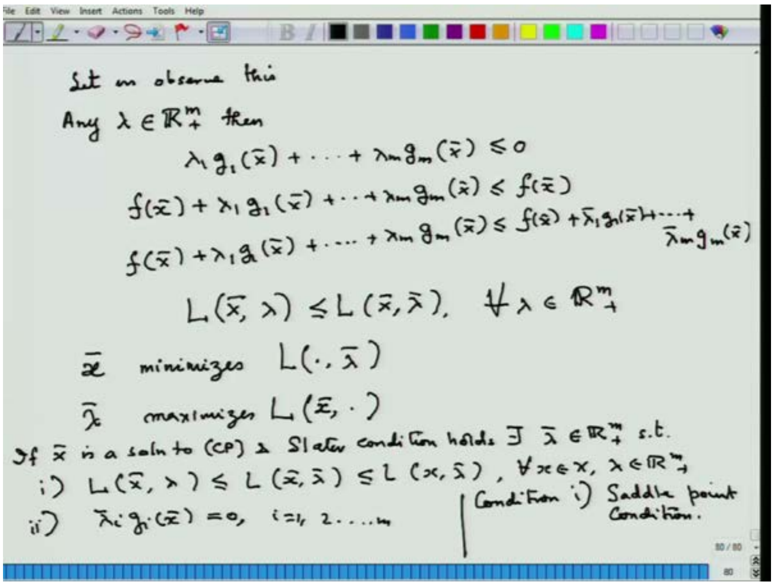
(No audio from 27:46 to 28:11)

So now, I have got the complimentary slackness condition, which I have already discussed earlier, clarifying the optimality condition. So, you now think that we are actually on the right track, because we have got something, which we know and which is, which you know is very fundamental. So, we must be on the right track. So, knowing something knowing that you are on the right track gives you certain enthusiasm. So, let us go back and go back to the formal equation again, so what would you have is now f (x) plus lambda 1 bar g 1 x lambda m bar g m x is greater than f of x bar; but I can, because this is I can write this as as f of x bar plus 0, what the what the why do I write this, because you see these part is 0, and these part has is quiet similar looking like this part.

So, I can write this as (No audio from 29:22 to 29:37) so now, there is a commonality in the appearance of these two expressions, which has this is bigger than this, now this would simply say and this can be shorted shorthanded, and we can write this as L x lambda bar, lambda bar is lambda 1 bar lambda 2 bar lambda m bar is bigger than lambda L of x bar lambda bar, means we are defining, so this is called the Lagrangian function, which you also have studied, when you have done your basic mathematics at your first year level in calculus Lagrangian function, but this is extremely fundamental to optimization.

For for our problem CP, the Lagrangian function is it just takes the non abstract constant inequality or equality constants and adds adds them to the constraint, that is an attempt to make a constant problem on constant, but here the abstract constant cannot be handled or pulled in to the Lagrangian, it can be, but it would involve a bit more technicalities, which we need not go, and because without that technicalities, we can get what we want; for the Lagrangian function is L x lambda, for any lambda, that this is in R n, and this is in R m plus so is defined as f (x). So, this now would be defined like this, and this obvious is defined like this. So, this is true for all xs in x. So, because I have not chosen that this only for x, which is feasible, so this is for any x in x this is true. So, I have shown the existence of a lambda bar such that this holds, so this and where this is what is called the lagrangian function; but also looks at this fact, fact is following.

(Refer Slide Time: 31:54)



That let so observe the fact, observed this; now take take any lambda in R m plus, lambda any lambda whatever you want, then lambda 1 g 1 x bar lambda m g m x bar is less than 0, naturally this is, all of these are lambda 1 lambda 2 lambda m are in R m plus. So, they are all greater than equal to 0 and g 1 x bar g m x bar, they are all less than equal to 0, because x bar is a solution. So, this would imply f of x bar plus lambda 1 g 1 x bar, we add an f x bar on both sides; right then you will say, how do I make them look look alike. So, I can add 0 again and put whatever you know as 0.

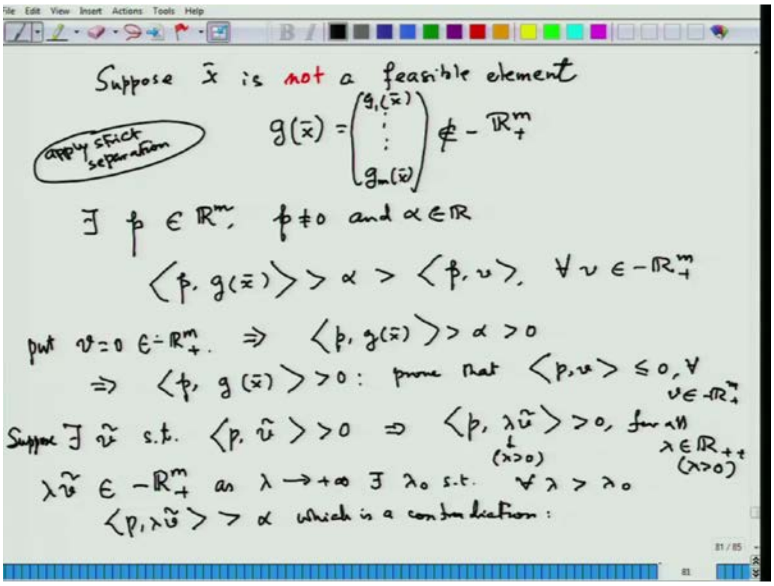
(No audio from 33:14 to 33:42)

So, what I have proved here is, L of x bar of lambda is less than L of x bar of lambda bar for all lambda in R m plus. So, what I have shown that if x bar is a solution, then there exists a lambda bar, where all the lambda 1 bar lambda 2 bar lambda m bar are bigger than equal to 0; and once I fix the lambda bar, the Lagrangian becomes a function of x, and in that state, it minimizes the Lagrangian function is minimized over the set capital X. So, once I fix up the lambda bar you see, x bar minimizes this function; while if you look at this one, it says that if I fix up the lambda, fix up my x bar, and I vary my lambda, so if x bar is a solution, I can always get a lambda bar, such that if I fix up the x bar and vary my lambda, this lambda bar is maximizing L x bar lambda. So, x bar, what lessons learnt is x bar minimizes L lambda bar, lambda bar maximizes.

So all this story that we were telling can now be summarized that if x bar is a solution to CP and Slater condition holds, there exists lambda bar in R m plus such that number (1) L of x bar lambda is less than equal to L of x bar lambda bar is less than equal to L of x lambda bar for all x in x, and lambda is element of R m plus and of course, we have to mention the complimentary slackness condition. So, together this is called the saddle point condition. So, this is called the saddle point condition, this is a complimentary slackness condition.

So, this condition, condition (1) for saddle point condition, this is the most fundamental necessary condition for a convex programming problem; the reverse question has to be asked that if I have an x bar and a lambda bar, for which these two conditions are satisfied. So, then remember here I have a lambda i bar is important not lambda i. So, if these two conditions are satisfied I have an x bar in x for which these two condition holds, can is x bar solution to the original problem that is exactly what we need to prove now. Now, let us go for the converse of what I just said that that if there are points x bar and lambda bar, which satisfies this what would happen; is x bar a solution of the original problem; now x bar here is a element of x, I do not know whether it satisfies g I x bar less than equal to 0 or not.

(Refer Slide Time: 37:49)



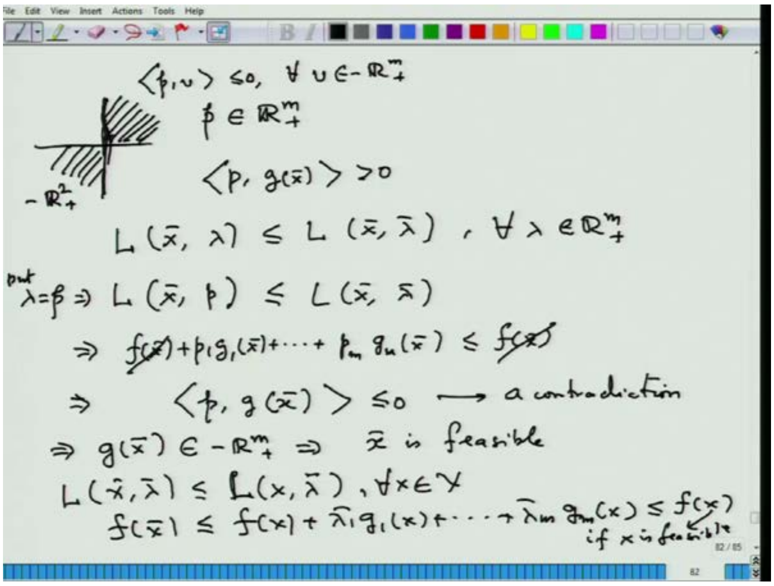
So, suppose x bar is not of feasible element; now this means that if I write down g 1 x bar, g 2 x bar, g m x bar there is one of them, which is strictly bigger than 0. So, if I am writing this, then one of this quantity is strictly bigger than 0, any one of them. So, that means that this vector cannot belong to the negative of R m plus; in the sense that all of them are not less than equal to 0, what R m plus is a close convex cone, and this is the point outside it. So, there exists p element of R m, p not equal to 0 such that, now here we are applying the streak separation theorem, such that p of g x bar, p and alpha sorry I just write and alpha element of R, such that P of g x bar is strictly greater than alpha, is strictly greater than equal to p of v for all v element of minus R m plus we had been using this when we had proved K naught naught equal to K, in the last class, we had also used the similar result.

Now let us see, here you can put a equality also does not matter much. So, if this is my strict inequality, so basically p x is equal to alpha is the fiber plain, which is strictly separating this point g X bar with v, what I claim is that see we put v is equal to 0, because R m plus as 0, this would imply that p of g x bar is strictly bigger than alpha, strictly bigger than 0. So, what you proved here is p of g x bar is strictly bigger than 0.

Now, what I want to prove is that to prove that p of v is less than equal to 0 for all v in minus or in plus now this means what; now suppose p of v is… So, they are suppose suppose suppose there exist suppose there exist v tilde, such that p of v tilde is strictly bigger than 0; suppose suppose this is not true; then this implies p of lambda v tilde is strictly bigger than 0, for all lambda is element of R plus, that is lambda for all lambda strictly bigger than 0, R plus plus; let you take any lambda strictly bigger than 0, this will be true of course

Now, lambda of V tilde is also an element of minus R m plus, this is quiet natural, all the all the vectors would remain, all the components would remain negative. So, I can make the lambda bigger and bigger and bigger. So, remember yesterday's arguments, so, it will be bigger and bigger and bigger; and so if this becomes bigger and bigger and bigger, so this value can then for certain threshold value of lambda would cross the value of alpha would be become strictly bigger than alpha. So, as lambda tends to plus infinity, there exists lambda naught such that, for all lambda bigger than lambda naught p of lambda v tilde is strictly bigger than alpha, which is which is not, which is a contradiction, because it breaks this fact.

(Refer Slide Time: 43:50)



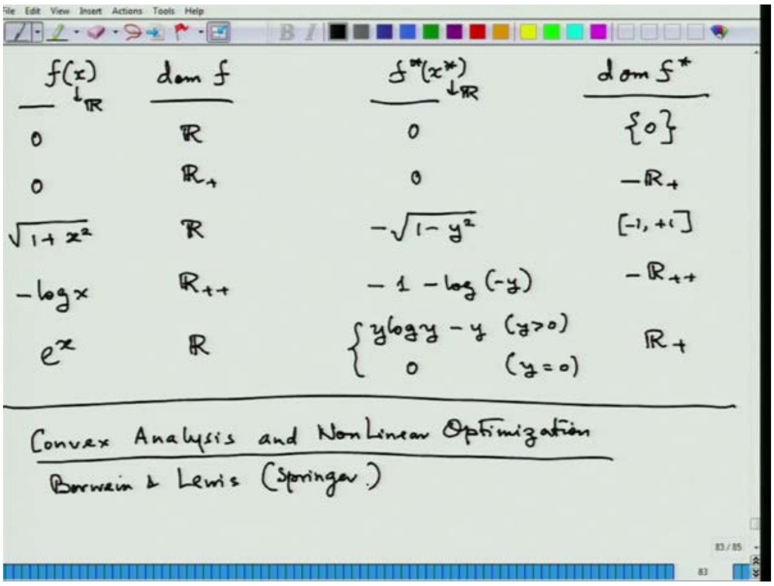
Now, once I broken that, we have a contradiction. So, we have that now that P of V is less than equal to 0 for all V element of minus R m plus. So, for every negative number this p of V has to be less than equal to 0; basically, find all the vectors P this is your R 2 minus R 2 plus. So, find all the vectors p, which for and obtuse angle with elements here, which means all the vectors here; basically, so, this implies P is a element of R m plus. Now what do I have? So, that now shows that p is element of R m plus and P of g x bar is strictly bigger than 0.

But I have that L of x bar of lambda is less than equal to L of x bar of lambda bar. Now, so, this is true for all lambda, so L of x bar of p, because p is also for all lambda in R N plus this is true. So, since p is in R m plus, so putting putting lambda equal to put lambda equal to p, implies that this is less than L x bar lambda bar. So, this could imply f of x bar plus p 1 g 1 x bar p 2 sorry p m g m x bar less than equal to f of x bar, what lambda 1 bar plus lambda 2 bar plus lambda 3 bar lambda m bar, this lambda 1 bar g 1 x bar lambda 2 bar g x g 2 x bar lambda m bar g x bar that is equal to 0. So, that part I am not writing.

So, this would imply, this cancels from both sides. So, this would be less or equal to 0, which shows that p of g (x), this can be put into this form naturally p of g x bar is less than equal to 0. So, this is a contradiction, so we come to a contradiction. So, proves by contradiction or sometimes 1, you know gives you the answers immediately. So, it implies that what we have assumed is not correct. So, g of x bar is an element of minus R m plus, which implies that x bar is feasible. Once you know that x bar is feasible, then immediately you also know this equation L of x bar lambda bar is less than equal to L of x lambda sorry L of x lambda bar. So, finally, you will have f of x bar less than equal to f (x) plus lambda 1 bar g 1 x; what do you see, if I take x to be feasible, then all of this x. So, this is true for all x in x.

So, if x is feasible, then apart from being x in x, they should also be satisfying g 1 x less than equal to 0, g 2 x less than equal to 0, g m x less than equal to 0. So, this whole part becomes negative, because lambda 1 bar, lambda 2 bar, lambda m bar is non negative; if they are greater than equal to 0. So, this part is, this is true, if x is feasible and that gives you the answer. So, f x bar is less than f (x) f or all feasible x. So, that is what we have proved. Now, the lot of things that come out of this, but we will not do this thing now; as I promised I will give you a table, we have spoken about f star of a function, the conjugate of a function, but we have given some little examples, we will give some more examples today, which so that you can figure it out nicely.

(Refer Slide Time; 48:08)



So, I will have f, I will have… I will write down dom of f, I will write down the f star, and I will write down the an f star x star basically, f star x and dom of x star. So, I will write down few functions; and just so that you can have a listing, so 0, the dom of f is R, that the 0 function, this is also 0 function, but the dom is other than 0, it is plus infinity;

you can you have you figure out, this would be your home work to figure out what I am writing is true or not. So, the dom f is R plus. So, this are we are only talking about proper convex functions, say if it is R plus means anything else other than R plus, the value of the function is 0; value of the function is plus infinity. So, x is in R; our x is in R; (No audio from 49:14 to 49:24) where domain is R, (No audio from 49:27 to 49:46) the domain is R plus plus; obviously, otherwise you cannot define log x, domain is necessary minus r plus plus.

So, these what I am giving is a collection, these results are from I will just write down name of the book, which will be very important, for your studies; those who want to really get involved into convex analysis and optimization, this is the book we should really start with at this moment; this is a very modern, very well written fascinating book; I will just, and I am actually listing them from this book; and of course, this is one of my my own favorites. (No audio from 50:36 to 50:49) So, these are some examples what what are where what is this listing is given from the book convex analysis, (No audio from 50:58 to 51:21) Borwein and Lewis, Jam Borwein and Andiron Lewis, they are very famous optimization theories.

Brownie, basically is a poly math; and this published by Springer, the first edition in 2000 and now we have second edition, if I can just show you this book, if you really want to look into the title once again; this book should be the starter for anyone, who wants to graduate in this area of convex analysis and optimization. So, this should be the starter book, and this is also one of my very personal favorites; trying the problems of this book is by itself getting into research, really doing this as some problems in this book can actually lead to some research. So, with this I end up my talk today; thank you very much, and tomorrow we will see the consequences of the saddle point here.