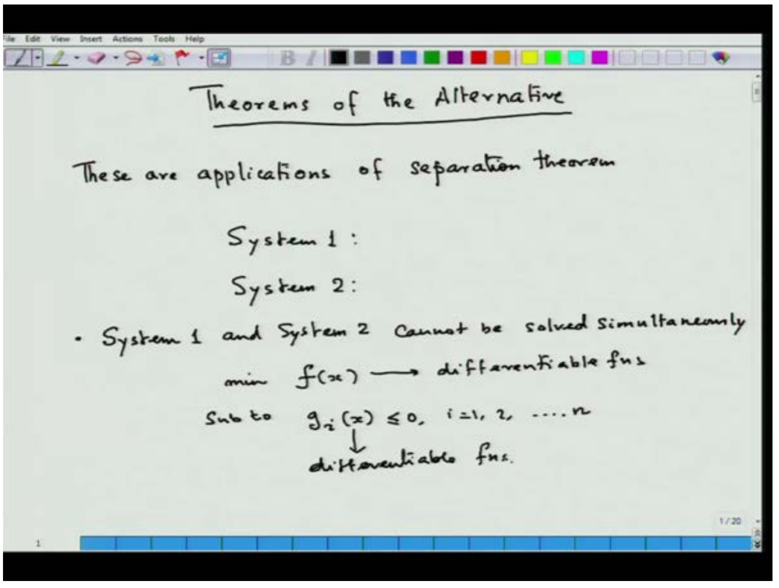
Convex Optimization

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

Lecture No. # 06

Welcome again to this course on convex optimization. Today, we are going to study theorems of the alternative. Now, theorems of the alternative are applications of separation theorem and we have already studied separation theorems in the last lecture.

(Refer Slide Time: 00:29)

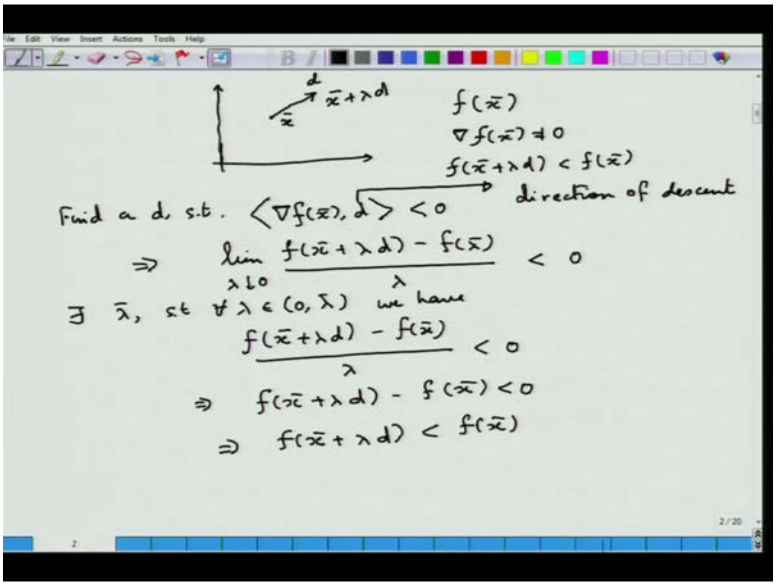


(No audio from 00:32 to 00:48)

Now, what is a theorem of the alternative? So, in theorem of the alternative there are two systems; system 1 and system 2. So, system 1 there is some equality or inequalities; system 2 there are some inequalities or equalities. So, what we say is that if system 1 has a solution, system 2 cannot have a solution; and if system 2 has a solution, then system 1 does not have a solution. Similarly, system 1 does not have a solution, system 2 has a solution; if system 2 has a solution, system 1 does not have a solution. In the sense that if one of them can be solved, the other cannot be solved; both cannot solved simultaneously.

So, the mode goal is to show that system 1 and system 2 cannot be solved simultaneously. (No audio from 01:58 to 02:08) Now, why we are interested in such systems? That is a major question. This thing comes from the very notion of optimality. Now, let me consider a very, very simple problem: minimize f(x) subject to certain constrains. Let us just take inequality constraints. Now, let me assume that this f and the g, each of the g i are differentiable functions, and you already know the definition of differentiability. (No audio from 2:58 to 3:07).

(Refer Slide Time: 03:13)



(No audio from 03:53 to 03:06)

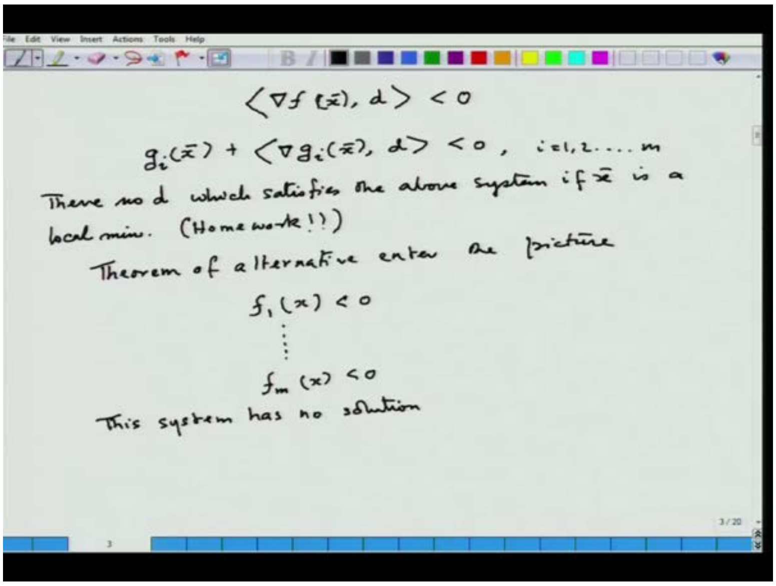
Now, let us consider a point f(x) bar here; let us consider a point x bar and suppose we know we calculate f(x) bar. And suppose this x bar is not a solution of the problem - that x bar is not solving. May be… Let us look at the picture when we have no constraints. Then basically we are looking at the scenario where grad of f(x) bar is not equal to 0. So, I must now improve. So, I must move from x bar in such a way, such that in such a detection I should move, I should move in this direction d. Such that if I come to a new point say x bar plus some lambda times d, f(x) bar plus lambda times d should be strictly less than f(x) bar, that you should be able to find such a lambda. That is exactly what is done in the unconstraint optimization algorithms. This is exactly what we have to do; find the lambda, find the d. d is the direction of the descent.

So, along the if you move along the direction d and if you move within the scale limit of lambda then your function value actually decreases. You go to of value which is lesser than your current functional value and you check whether the minimum is attended this point or not at least if this is satisfied and then move down again. So, this procedure how do I get such a d on how how will I guaranty such a d.

Observe that if I have this condition true, if this if I can find a d find the d such that this is true, then by definition you can write this as f(x) bar plus lambda d minus f(x) bar. Just you know writing down Taylor's theorem and nothing else; you can figure this out very easily. This thing is strictly less than 0. While lambda is running down arrow to 0, and you must observe that once I know this by get the very definition of a limit I can show that there would exists lambda bar such that for all lambda between 0 and lambda bar with both of these not included. We have f(x) bar plus lambda d minus of f(x) bar by lambda is strictly less than 0.

So, this would imply at f(x) bar plus lambda d minus f(x) bar it strictly less than 0, because lambda is positive. This would imply that f(x) bar plus lambda d strictly less than f(x) bar, a thing that we exactly wanted. So, any d which satisfies this relationship, this sort of d is called a direction of descent. Now, if I am now considering this problem where I have constraints, then what I can show is the following that If I consider this system…

(Refer Slide Time: 06:49)



(No audio from 07:52 to 07:21)

Sorry my mistake. This system has no solution in d that is there is no d which satisfies the above system if x bar is the local minimum. (No audio from 08:51 to 08:03) So, there is no such d which will solve this. You see you can figure out this quiet well and I would ask you to figure out this as homework. That if x bar is a local minima to this problem which you can call as p then this system has no solution if x bar is a local minima to p.

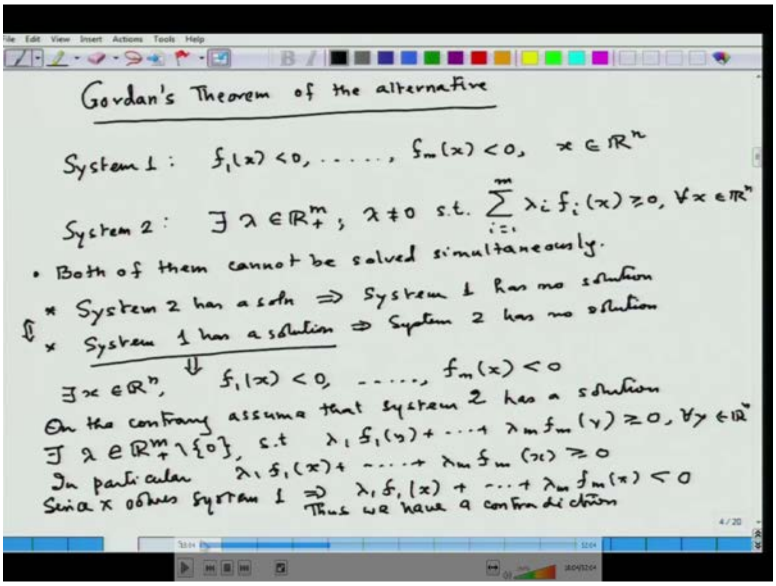
Now, the question is, if this is not solvable is there something else which is solvable, is there some other system which is solvable. So, it will allow us to characterize the necessary condition for a local minimum which is fundamental, because that characterization would allow us to compute the local minimum. This is exactly or precisely the point where the theorem of alternative arises. So, this is the place where theorem of alternatives enter the picture.

(No audio from 09:03 to 12:14)

Look what is what is this system telling? This system telling that this is a linear inequality, this is a linear function and this is an affine function right in term in d. Now, this system of linear and affine functions with strict inequalities they do not have a solution, which means that in general, I can I am looking at this system. That ok. I have a system of say convex inequality; I am just generalizing, this because these are all sub class of convex function. See if I have m convex inequalities and I am saying… This system has no solution.

So, what is the certificate means how do I say that this system has no this system does not have a solution; if this system does not have a solution something else have have a solution. That ok. If something is solvable I can say that this system does not have a solution. That is that that particular something is called the certificate of solvability or un-solvability of the system. And this is precisely what optimality's all about. This is exactly optimality of the point x bar - local optimality. So, we have generalized this thing into this frame work. We will show how this will be applied.

(Refer Slide Time: 10:52)



Now, this leads to a first optimality theorem sorry first theorem of the alternative which I am which is called the Gordon’s theorem of the alternative.

(No audio from 11:56 to 11:11)

Now, I have two systems; system 1, system 2. Let us see, let us write down the two systems. System 1 is exactly what we are just discussing that you have a chain of m inequalities with convex functions. These are all convex functions - m convex functions. And we are looking for an x in R n which will satisfy the all this or else the system 2 says that there exists that I am trying to find a lambda in R m plus that is all the components of this vector in R n is non-negative and lambda is not the 0 vector in R n. Such that an individual components of lambda multiplied with this f i’s and this sum is always greater than equal to 0 for all x, you want take.

You know, you could actually remove this R n by some convex - close convex of set of R n. So, that will also work. So now, if I want to prove this what I want to show, is there whenever system 1 is solvable, system 2 is not and vice versa. In the system, both of them cannot be solved simultaneously. So, both the theorems conclusion is that both of them… (No audio from 13:55 to 13:13) This is the conclusion.

Now, if I prove for example, system 2 has a solution, it should imply system 1 has no solution. So, this is the statement - p implies q. So, then this fact is equivalent to this, it is same; this is equivalent to this that system 1 has a solution implies system 2 has no solution. See if I prove this, the second thing, I have actually prove both of them. So, let me just show you that if I take system one has a solution, see if I I am assuming this part. So, if this is assumed. So, this would imply that they are exists and x in R n for which each of this inequalities hold. Now, I have to show the system 2 has no solution. So, on the contrary…

(No audio from 14:50 to 15:04)

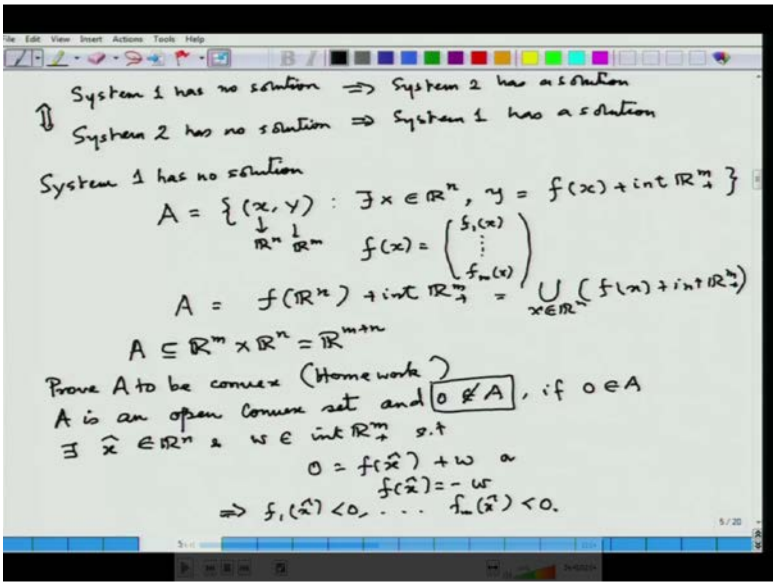
On the contrary, assume that the system 1 sorry on the contrary assume that system 2 has a solution. So, I have to prove that it has no solution. So, I am taking the contrary indication; I am taking the opposite assumption and then proving that there will be a contradiction. Assume that a system 2 has a solution. So, means there exists a lambda in R m plus, but is not the 0 vector such that lambda 1 f 1(y) plus lambda m f m(y) is bigger than equal to 0 for all y in R n. Now... So, hence in particular for the given x, in particular lambda 1 f 1(x) plus lambda m f m(x) is greater than equal to 0, because x is one of one of this element.

Now, where x is again a solution of this system; now, I know for this particular x, f 1(x) is strictly less than 0, f m(x), f 2(x) is strictly less than 0 and dot, dot, dot f m x is strictly less than 0. Now, all these lambda 1, lambda 2, dot, dot, dot, lambda m are greater than equal to 0 and one of them is at least is non-zero which means that there is there is at least one element where lambda say lambda k, lambda k is strictly bigger than 0 and f 1(x) is strictly less than 0; f f k(x) is strictly less than 0 anyway. So, f k(x) is anyway strictly less than 0. So, there will be at least one k for which lambda k f k should be strictly less than 0. So, on a whole, since x solves system 1, because this happens; it implies that lambda 1 f 1(x), this finally must be strictly less than 0. So, which means now we have a contradiction. So, this is thus we have a contradiction.

(No audio from 17:36 to 17:42)

So, we have proved two things; we have proved system 1 has the solution, system 2 has no solution. So, once we have so… Because we have a contradiction, so, our assumption that the system 2 has the solution is wrong. So, whenever system 1 has a solution system 2 does not have a solution. So, simultaneously we have proved that system if system 2 has a solution, system 1 has no solution. Now, we… Now, have we will go for the remaining one, I will say if system 1 has no solution, see look what I have done here. May be I am just crossed the page. What I have done here? I have said that system 2 has the solution implies system 1 has no solution; system 1 has the solution implies system 2 has no solution. But I have not said if system 1 has no solution what would happen right.

(Refer Slide Time: 18:40)



So, here I am trying to prove that if system 1 has no solution.

(No audio from 18:41 to 18:53)

Then it should imply that system 2 has a solution. This would say that if system 2 has no solution, this is same, this system, this is equivalent, because a statement p implies the statement q the negative of the statement q implies the negative of the statement p. System 2 has no solution implies system 1 has a solution. Now, if this is what I have to prove. So, I just have to prove any one of them; I will prove the first one.

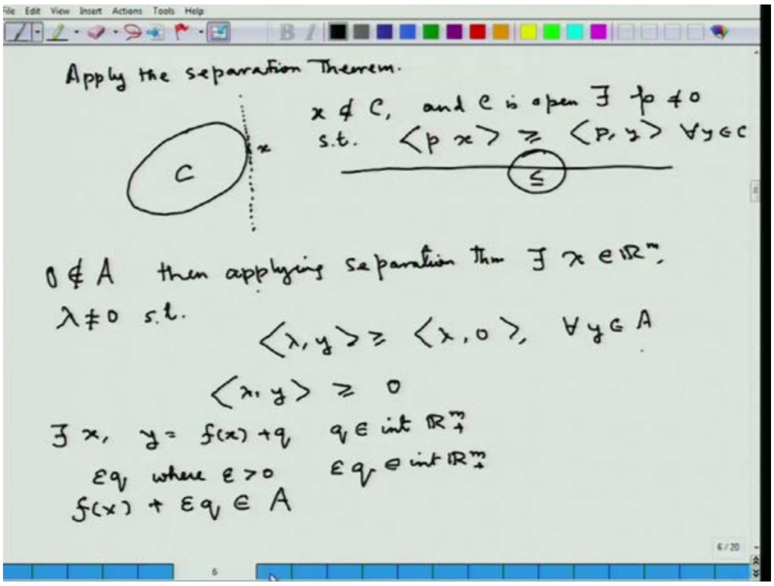
So, let me assume the system 1 has no solution. Now, let me write down a set, let me construct the set A which is of the form, (No audio from 20:07 to 20:13) there exist x in R n such that y can be written as f(x) I will write, what tell you what is f(x), as the interior of R m plus that this f(x) is the vector consisting of these real numbers f 1(x), f 2(x), f m(x). Because these functions are all convex functions from R n to R, which is so obvious and I am not repeating this fact, but here I have written everything in the form of vector. So, what I am doing, I am taking an x and computing this. And adding to it some element from the interior of R m plus that that is I am adding to it another vector whose all components are negative sorry whose all components are positive.

So, we have a f(x) I am taking a x in R n. So, how is my y constructed? I am taking x in R n computing this f(x) that is making this vector and then with it I am adding a vector whose all components are positive. In f at A can be written as f of R n plus int of R n plus. This is the way A can be written. Now, if you look at it carefully. So, this is in R n and this is in R m. So, A is a subset of the Cartesian product of R m cross R n. That is it is in the space R m plus n. Now, what I need to do here is to tell you that these set is a convex set. So, your homework is prove A to be convex. So, this is your homework. Do not mind it.

So, exactly the convexity of the function is needed to prove that this set A is convex. So, once you prove this you have to observe another fact that A is a open convex set. So, A is an open convex set, it is not a closed convex set, because it is we are talking about the interior R m plus. So, basically if you take this interior R m plus, you construct a set like this f(x) plus interior interior of R m plus then basically you have translated an open set. You have made this as the origin. So, then this would remain to be an open set. So, why… So, these sort of sets whose union is actually or you can write it like this; union of f(x) plus int R m plus x element of R m. This is the set A.

Now, these are open sets. So, arbitrary union of open sets is again open. So, this is an open convex set. And 0 is not an element of A, because if 0 is an element of A then there would exists x hat in R n, and w element of interior of R m plus such that 0 is equal to f(x) hat plus w or f(x) hat is equal to minus w. Now, every component of minus w is strictly negative. So, every component f 1(x), f 2(x), f m this all of these; so, this would immediately imply that f 1 x hat is strictly less than 0 to f m x hat is strictly less than 0 which shows that x hat is as solution to the system 1, but I have said that x 1 the system 1 has no solution. So, it means that this conclusion is correct that 0 cannot be an element, element of A. Now, I ask you to apply the separation theorem.

(Refer Slide Time: 24:52)



(No audio from 24:51 to 25:02)

You might ask me a question now. You are asking me to apply the separation theorem, but how can I do so. Because you have just told me that you take a close set and point out it you can strictly separate it. But here you are telling that the set is open and then you have a point outside it and you asking me to apply the separation theorem. The problem is this, that ok. May be the set - convex set is open that I have a convex set like this. A set C which is open that is I do not have the boundary. Now, I have a point just on the boundary. Then also this point says x is actually outside c, it is not in c. So, I can actually I will not have a strict separation, but I can have a hyper plane passing through this point.

So, I will have a separation. So, if I if x is not in some sets c and c is open, there exist a p not equal to 0 such that p of x is always bigger than p of y for all y in c or p of y is always bigger than p of x for all y. It could have be a reverse equation also you could change the inequality that is the matter. So, here we have A to B open set and 0 is outside A. Now, here actually 0 means (0,0), basically this is the 0 of R n and this is the 0 of R n. Now, if that is so then how do I go about applying the separation theorem. So, we will just have to look into this thing. So, there would exists some lambda mu not equal to 0, and lambda mu is element of… I am applying the standard thing R m cross R n such that lambda mu times (y,z) is bigger than equal to lambda mu times (0,0) for all (y,z) in a or (x,z) what would I detect show sorry, sorry I think I just have to go back and look into this thing all (x,y)

(No audio from 27:58 to 28:08)

For every (x,y) that you have… There is a mistake, the mistake in the set contraction. So, we will first prove that system one has no solution and let us construct this set A which is given this way that it consists of all y in R m for which there would be an x such that y can be written as f(x) plus in some element from the interior of R m plus that is. This vector would have all its components strictly greater than 0 and this f(x) is actually a vectorial representation of the system of functions we have f 1(x), dot, dot, dot, f m(x).

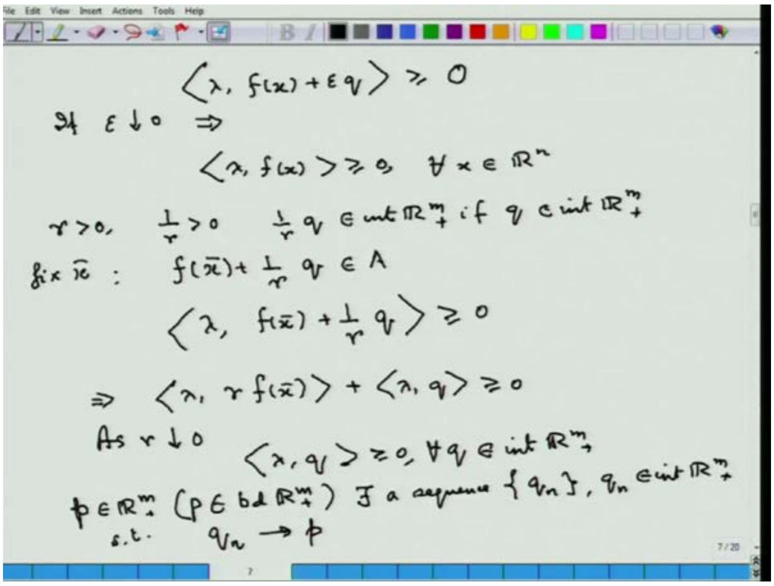
Now, A can also be written in this form which is same as writing this. If you this set interior R m plus is an open set and when you translate it by f(x) this remains an open set. So, for this each of this x, this is an open set and when you take union this would also be an open set. And A is an element of R m - is a subset of R m. So, you have to prove that A is a convex set which is of course an open convex set, because we have proved the openness. A is a convex set which is your homework and that is observed by in observing that each of this f 1, f 2, f m are convex and 0 is not an element of A, because if 0 is element of A then you see there would be an x hat and w for which this would be true, and then I can write take w to the other side. And now each of the components of minus w is negative showing that this is negative, and that is x hat the solution of system 1 which is faults, because system 1 is assume to have no solution.

Now, we have to apply the separation theorem. Now, you might ask me that ok. In the previous separation theorem story, we had talked about closed set and a point outside it, and you can have a strict separation. But when you have an open set like this, if you consider this convex set when you everything inside, but you do not have the boundary. Then any point on the boundary we can also be a point which is not contained in c. But through such a point also you can draw hyper plane containing x which puts x on the hyper plane and the whole set c on the other part of the hyper plane. You see there is open set c is actually strictly inside the hyper plane.

So, what you can say is that this result will always hold true, irrespective of whether if the set is open or closed. So, this sort of separation the standard separation will always hold true. Of course, this inequality can be reverse also sub to you. So now, here we do not have 0 in A. So, then applying separation theorem, (No audio from 31:00 to 31:10) there exists lambda - element of R m and lambda not equal to 0. Such that lambda of y is greater than equal to lambda of 0 for all y in A. This is the standard separation theorem. I have just reversed the inequality, does not matter.

So, lambda of y is greater than equal to 0. Now, how does the y look like? So, there exists an x. So, take a y. So, there exists an x such that y is equal to f(x) plus some q where q is element of interior of R m plus. Now, you take any element in the interior of R m plus say q and you construct the element epsilon q where epsilon is greater than 0 right. This is what you can do. So, then this epsilon q will also have all its components to be positive. So, this will be also in interior of R m plus. So, if I construct this element f(x), this x plus epsilon q then this is also in the set A.

(Refer Slide Time: 33:06)



So now, I can put this here to show that lambda times f(x) plus epsilon q is greater than equal to 0 right. Now, if epsilon goes to 0, it immediately shows that lambda f(x), because in the limit this will what is going to happen. Now, this x was arbitrary, you could choose any x and construct elements like this. So, you could choose of one x and construct the element and prove that for that particular x, this will be true. You could do another x and do the same thing. So, this is true for every x.

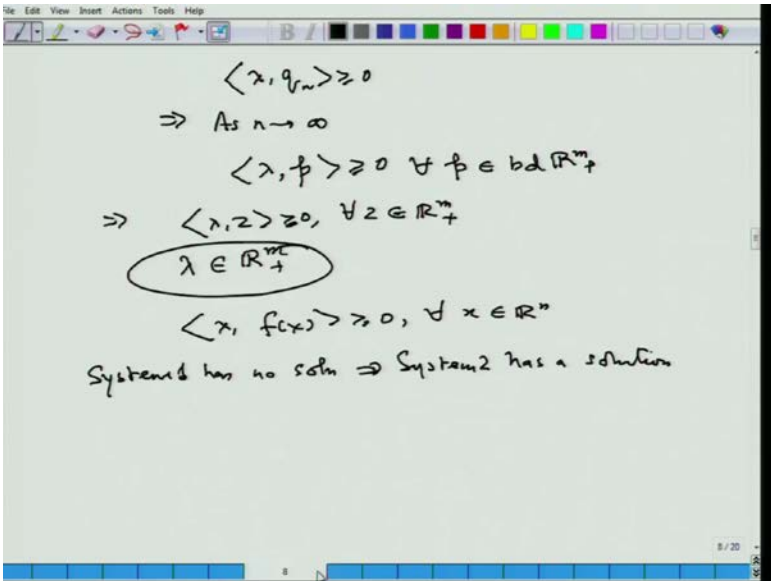
Now, take r to be greater than equal to 0 and 1 by r is greater than equal to 0. So, 1 by r into q is element of interior of R m plus, if q is in the in the interior of R m plus.

Actually, you know interior of R m plus is the cone n. So, it is a conic structure. So, this obvious, because if you take a positive number and multiply by positive number is the positive number. Now, construct this element take an x bar fix the x bar fix x bar, and construct the element f(x) bar plus 1 by r times q. Now, this q could change, take any q, I have taken a q constructed this 1 by r, fixed up the x and constructed this element and these element is again in A.

So, again by the separation theorems lambda of f(x) bar plus 1 by r time's q is greater than equal to 0. So, it means lambda times r f(x) bar plus lambda times q, it is just inner product, they are multiplied by R n, is greater than equal to 0. Now, as r goes to 0, this will go to 0, because this is fixed number. So, this will go to 0 and the inner product that will go to 0. This will imply lambda q is greater than equal to 0 or q element of interior of R m plus. Now, this is this q is arbitrary. So, this is true for all q in the interior or R m plus.

Now, take any q naught or take any such p in R m plus. Now, because any because R m plus is closed any such p is an limit point. So, for every element in the interior you are proved this is true, now take p in the in R m plus and just take p to be in the boundary of R m plus. So, there exists a sequence q n, with q n in the interior of R m plus such that q n converges to p. Then what I have is again that lambda of q n is greater than equal to 0.

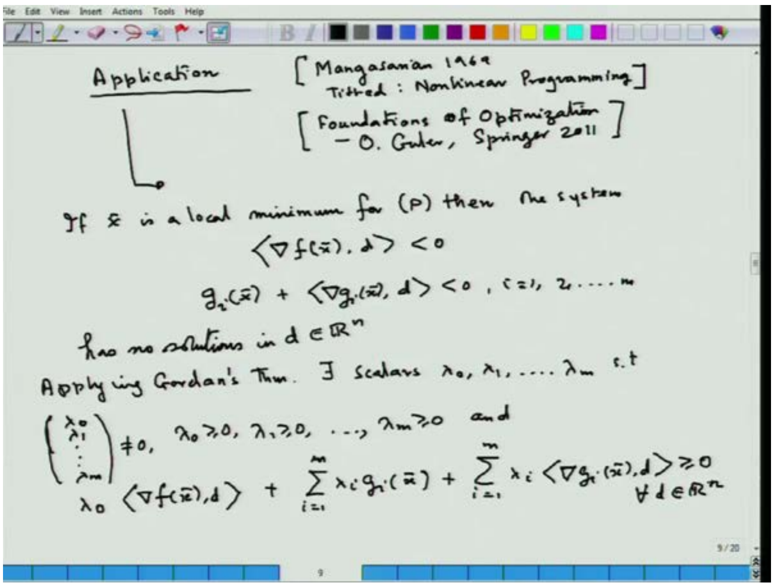
(Refer Slide Time: 37:02)



So, this would immediately show as n tends to infinity, lambda of p is greater than equal to 0. But this p was an arbitrary element. So, what I have shown at the end is lambda time z is greater than equal to 0 for all z in R m plus. So, this is true, first I have proved that this is true for all p in the boundary of R m plus and we have showed that everything is true, this is true for all q in interior R m plus. So, it is true for every z in R m plus.

And hence lambda must also be an element of R m plus which is obvious. Because it is for every non-negative vector, it is giving greater than 0. So, it must. So, it is making an acute angle. So, it must itself be in R m plus. So, that is exactly what we have proved, we have proved that lambda is in R m plus and also we have proved that lambda of f(x) is greater than equal to 0 for all x in R n. So, what we have showed that system 1 has no solution would imply system 2 having a solution. So… So, we have we have proved this part and hence we also proved this part.

(Refer Slide Time: 38:47)



So, let us see, how do we apply this result.

(No audio from 38:42 to 38:54)

So, here we will apply it to the optimality condition to find the optimality condition for a local minima and get a result which we will soon discuss. So, Gordon's alternative theorem is not the only alternative theorem and the many, many alternative theorems like Motzkin's alternative theorem, Tucker's alternative theorem, there is also an alternative lemma or alternative theorem due to forecast for the forecast alternative theorem. So, all these are very well represented in the book by Mangasarian. So, Mangasarian is the author he wrote a book in 1969 titled non-linear programming and it has all this beautiful results.

But you see let me tell you all those results are given in terms of matrices that is instead of convex functions that we have used in the Gordon's theorem. They have used linear function. Does not matter, because a lot of these things are only for linear function and it is not so easy to put them into the convex frame work and here we have put in into the convex frame work to make it to more general.

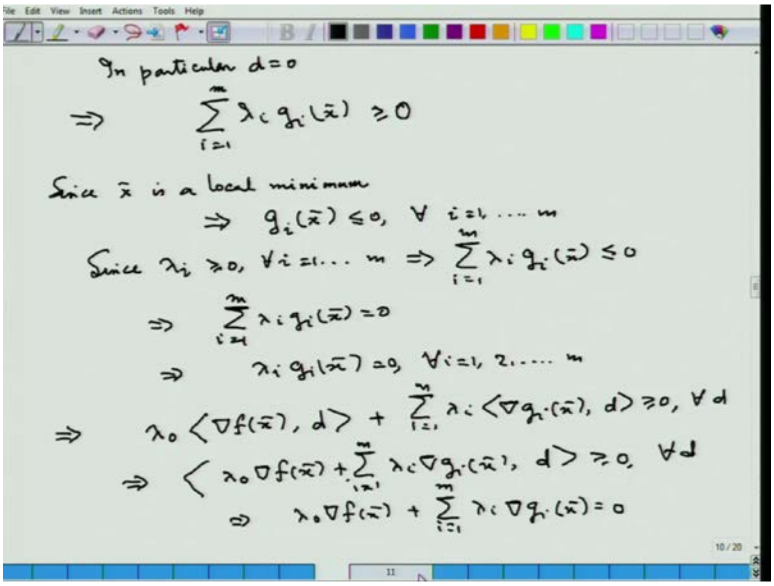
Another book is titled foundations. So, this book is role on the regular form its original publisher. It is now available through (( )) in their classics in applied mathematics series. Foundations of optimization, this book also is the fabulous discussion of theorems of alternative by Osman guler, it is just come out from Springer in 2011. It is a lovely book with very, very, very good discussion of optimality conditions and theorems of the alternative. So, let us go ahead and do this application.

So, as an application of the Gordon's alternative theorem that we have just learnt, we would go back and try to apply it to find the necessary an optimality condition for the problem p that we started in the very beginning. That ok. If this is differentiable optimization problem with inequality constraints what is my optimality condition, what is my necessary condition for the optimality that given if x bar is a local minima. Can I tell what sort of conditions it will satisfy in terms of the gradients of this function? And that is exactly what we are trying to now prove or rather find using the theorem of the alternative.

Now, as we have already discussed that our major impactors of studying this theorems of the alternative is to look into this fact that. This is exactly what happens if x bar is a local minimum of p that this system has no solution in d. Now, if this system has no solution in d, if you observe this function is linear in d and this function is affine in d. So, all of these are convex functions. So, now… So, this is this corresponds to the first system - system 1 of the Gordon's theorems of the alternative and what we are telling essentially is that optimality of x bar or local optimality of x bar is same as the first system of the Gordon's alternative theorem and that system has no solution. So, the second system would have a solution. So, it means that there would exist scalars. That is each of them there is corresponds a scalars. So, like scalar corresponding to each f i, lambda is a scalar corresponding to each f i.

Here also you have scalars corresponding to each of these functions. So, there will be a scalar corresponding to this which is lambda 0 corresponding to g 1 x bar g 1 grad g 1 x bar d lambda 1 dot, dot, dot lambda m, all of this has to be greater than equal to 0 by Gordon's alternative theorem and the full vector cannot be 0. It can as to be non-zero vector. And you keep on multiplying with this and this is exactly what you will get. This is true for every d, this has to be greater than equal to 0.

(Refer Slide Time: 43:22)

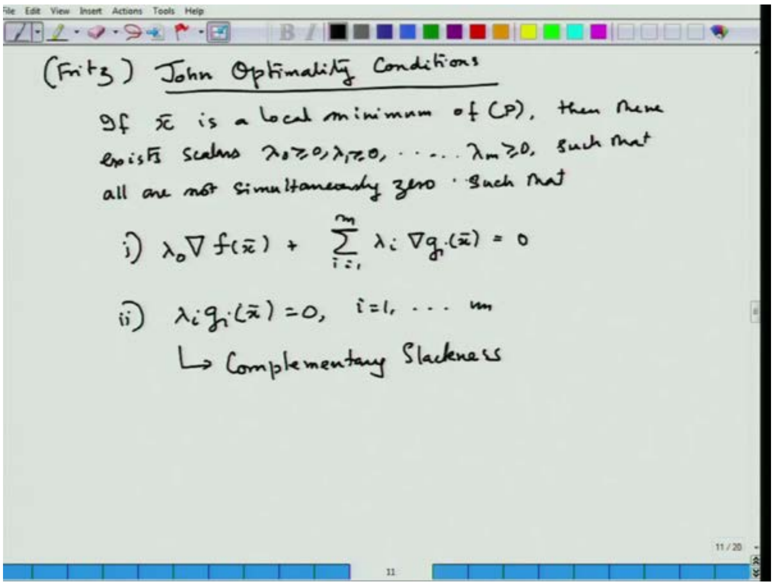


So, because it is true for every d in R n, I consider in particular d is equal to 0. That would immediately leave me just with the equation, because if I put d equal to 0, this will go and this will go. It would just leave me with the inequality this sorry is greater than equal to 0. Now, what does this mean? Observe that I have said that x bar is a local minimum to the original problem. So, which means this x bar is the feasible to the original optimization problem, which means x bar is a solution means it has to satisfy the constraints. So, all of this g i x bar is less than equal to 0. Since x bar is a local minima, it would imply that g i x bar is less than equal to 0 for all i equal to 1 to m. Now, since lambda i is greater than equal to 0 for all i equal to 1 to m it would immediately imply that the summation lambda i g i x bar, this must be less than equal to 0, because your multiplying negative and positive quantity and then your adding all non negative, non positive or negative quantity. So, that is less than equal to 0.

But then this would contradict with this. So, but… So, if both are them has to be satisfied, the only way is to have this. So, this would imply. Now, because each of them are negative quantities and all of them are add up to 0 which would imply that all of them has to be individually equal to 0. Now... So, once this I know to be, this to be 0. So, what I am left with is lambda naught grad f(x) naught f(x) bar d plus summation lambda i grad g i x bar d is greater than equal to 0 for all d. So, this can be simply summed up a nicely written as lambda naught grad of x bar plus summation i equal to 1 to m.

Now, this is true for every d. So, if I put this vector is w then I am saying that w d is greater than w inner product d is greater than equal to 0 for all d since. So, if I put instead of w if I put minus w in d as minus w. So, it will show me that the norm of w square would be less than equal to 0. And hence w would be 0. So, if this happens for all d, it essentially says that the linear function cannot always contain for every value of the variable it cannot be non-negative. It has to be both negative and positive in that sense. This would imply immediately that lambda naught.

(Refer Slide Time: 47:32)



(No audio from 47:06 to 47:15)

So, combining this and this, and the fact that all these values lambda naught dot, dot, dot lambda 1, all these are not equal to 0 simultaneously leads to the famous John optimality conditions of Fritz John. Fritz John is the full name of the person. The title is John. I do not know why they are always use Fritz John. So, you could have written like this also, this comes to the John optimality conditions given in 1948. By the way, when he first sent this result it is very, very crucial, because he used optimality condition for inequality constraints, till that time it was optimality conditions with equality constraints which were important and they were solved by the Lagrange multiplier technique.

But modern optimization and modern applications inequalities are the hallmark of constraint representation. So, thus John's John's optimality condition was the first to handle inequality and thus it is very, very important. And interesting fact is that when he first sends this paper to the dew channel of mathematics, it was rejected and was published in a conference, in fact many, many good mathematical papers when only published as conference proceedings.

So, this shows that if x bar is a local optima - local minimum I would say, because I have already said it is a minimization problem. So, local minimum of p then there exists scalars is real number basically. Lambda not greater than equal to 0, lambda 1 greater than equal to 0, lambda m greater than equal to 0 means whatever we had written earlier. Such that all are not simultaneously 0; there is lambda naught, lambda 1, lambda m all are not 0 - all are not simultaneously 0 0 such that number 1 lambda naught times…

(No audio from 49:59 to 50:18)

The second condition is very important it is called the Complementary Slackness condition. He says that both of these lambda i and g i cannot hold with strict inequality at the same time. That is lambda i strictly greater than 0 and g i was strictly less than 0 cannot hold. Because then this product would be strictly less than 0. See if g i is very strictly less than 0, lambda i would be 0. So, the… But if g i x y is equal to 0, lambda i can be equal to 0. So, all this constraints where g i x bar strictly less than 0 are called inactive constraint, because lambda i is equal to 0.

So, we end our talk today with the deduction of the famous John optimality condition. You see how in this case where f and g i is in this case are not been assumed to be convex, just they have been assume to be not just differentiable. We have been able to use convex tools to prove the optimality condition, there by showing the power of convexity itself. Thank you very much.