Convex Optimization

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

Lecture No. # 14

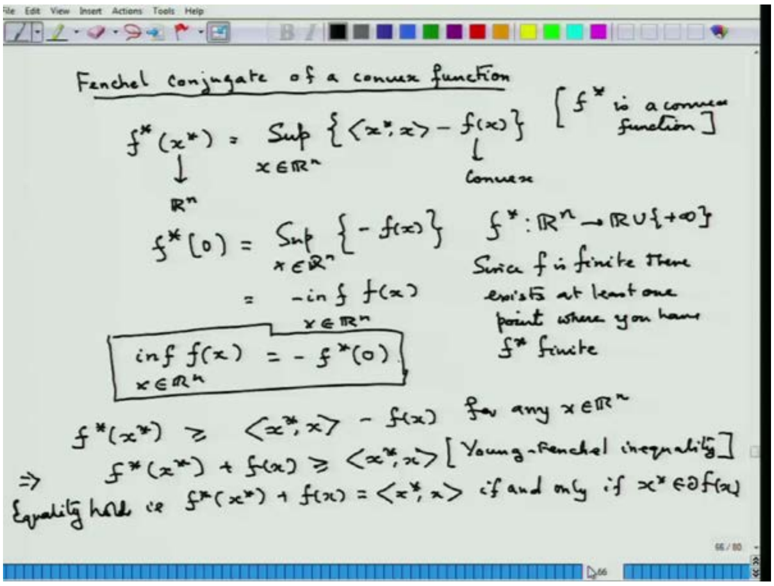
Welcome back once again to this course on convex optimization. We have over this, say 12 or 13 lectures we have indeed done a lot of things. You must be wondering possibly that, now come on just to find a maxima and minima of function, which we have done almost waving your hands in high school, when in a calculus course; you need so many stuff to do that. But the fact is that, you never realize, what we are doing in high school, because in that level, you do not differentiate between global minimizers, local minimizers, you do not even bother whether you have got the exact solution or not. So, we have lot of issues involved.

Here, we are gearing up machinery, which can take on practical problems. Here we are building up a machinery, which can answer many diverse type of queries, regarding optimization. Here we are building up a machinery, which would allow us to develop algorithms. The problem that we have seen in high school or even in your first year math, one courses, across engineering institutes or science institutes, is that what you have seen is just very, very small problems.

Real optimization problems cannot possibly be solved by hands and you know, you therefore, need a machinery, which would support you to build algorithm. And of course, we will come to the algorithms very soon; we would soon indulge ourselves in the pleasures of linear programming, which is a very, very important part of convex optimization. Our aim in this course is clear, in the sense that see you know, just I would just interrupt here, once in a while it is good to focus on the aim in a course; but it is not that, everything that you plan to tell to students, can actually be done? Always you have ambitious dreams, but you work towards that dream, but you go to a certain level.

So, the course, as I have envisaged in my mind, before I thought of delivering this lecture that if you gradually gets changed, is changing as I am moving. And I also at the end, want to see you learning some modern things with a huge amount of applications, things, which you can really apply in your own work. So, I am assuming a huge amount of audience from engineering in this course. So, lot of things would be of help to you as you as we go through the course.

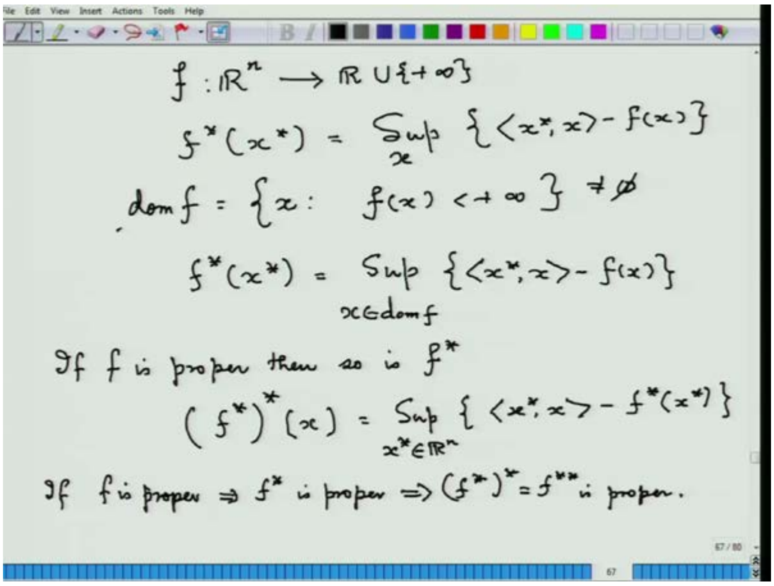
(Refer Slide Time: 03:39)



So, in the last class, we have spoken about what is called Fenchel conjugate. Now you see it is very important to know, if whether you are having fun or whether you are not having fun. If you are not having fun, then you must leave right now; you must stop seeing this course; and just forget about it. But you are having fun, you need to carry along, you will have to have for bit of lot for mathematics of course, because this is a mathematical course, and I am from the mathematics department, so I will be more mathematical in my orientation. But be with it, if you think that these are finally, you will get something, which you will need for your other studies like applications or whatever or possibly, we are just enjoying it.

So, I you do a thing, because you need it or you are just listening to a thing, because you are having good fun, just like you see a movie. So, if you are not doing either of the two, leave right now; do not see these lectures see these lectures. So now, I have defined conjugate, and also there was a promise that I would prove that the polar of… The polar cone is the cone itself.

(Refer Slide Time: 05:07)



So, I now look at what happens if my convex function is an extended valued and proper convex function. It is proper, so there is no use will bring in minus infinity. So, then also you can define the conjugate in the same way, it does not matter.

(No audio from 05:25 to 05:35)

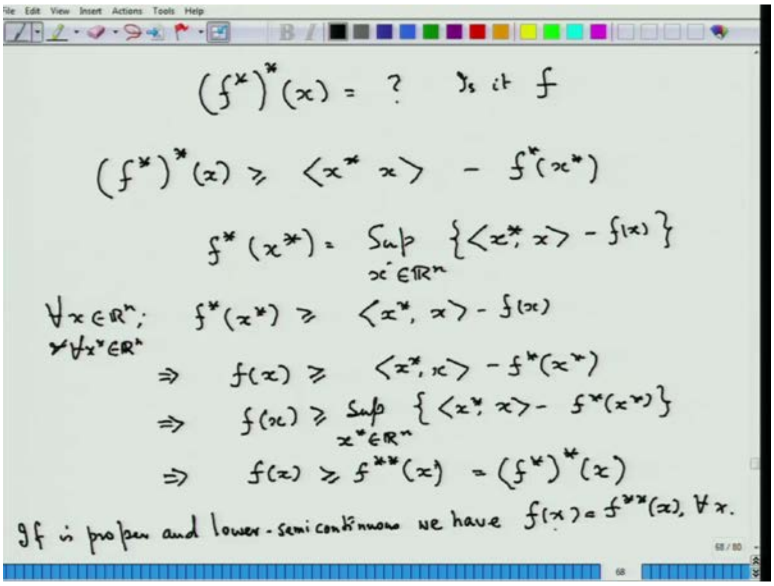
Now, observe this function would never take minus infinity value, would never take that value. The reason is as follows, because this function is proper; so, domain of f, the effective domain of f, which consists of all x, such that f of x is finite, this is non empty. So, there exists at least one point, where f (x) is finite; so in that case, that will be finite, and supremum would be that value. So, if f (x) is infinity then of course, this value would become minus infinity and of course, would not contribute to the supremum. So, you can write for this particular case, you can write this.

(No audio from 06:33 to 06:42)

So, you just have to bother about the effective domain. So, when it was finite value, this was R n; this has advantage; putting it in this framework has an advantage. A very, very fundamental result, which essentially is the structure of optimization, when a convex optimization in some sense, is the relation between f and f double star; so f star, if f is proper, and so is f star. Now, just like we have defined polar of the polar of a cone, so you can also think of defining the conjugate of the conjugant.

Now, let us see what is this thing means, you have the conjugate, and then you define the conjugate, which acts on x; and which is exactly a same definition, just for f star is now the function; now you are running over x star element of R n. Now, you see f star itself is a proper convex function; it is extended valued in general. So, here x star has been fixed, and your x has been fixed, and you are minimizing over x star. So, if f is proper, it implies f star is proper, means it has one finite value; there is one x in R n for f x is finite, which I can also write as f star star is proper. There is something bit more interesting that will come out of this f star star business.

(Refer Slide Time: 09:35)



So, you can ask that you have said that polar of the polar is the cone. So, what is f star star of x equal to, is it f? Now that is a good question of course; and we would like to see whether this is true or not. Let us go back and look at the definition of f star star. So, by the definition of course, I just remind you that this definition, I can replace that R n by domain of f star, which by the way is non empty. So, it will be x star x minus f star of …Can we infer anything from here? Does not look like, if you put it, you try to (( )) things it will get more complicated. So, what do I do with it?

So, let us go back to f star, and see what we, can we have something from there? f star (x star) is supermom of x belonging to R n. So, it is the definition of the conjugate even last class. So, there you go, f could be extended value need not be extended value, you could just think while you are reading this or when you are looking at what I am telling, you can really take f to be finite value and work. The interesting part of all these things being on the you tube possibly as I am told, and there are lot of these courses on the you tube is that you can do replace, it is like reading a book; if you do not understand a thing, you go and need it again. So, this is exactly what you are going to do; you can stop the show, and go back again, and have a look. It is very important that you go back and follow the trail of the arguments.

Now, what do I have from here; which you take any x, this is what I have. Now let us change the positions, so this would imply f of x is bigger than now this relation is not only true for all x, it is true for any x star you take. So, for all x, and for all x star e R n. So for the pair x x star, this result is true. So I can now just change my position. Once I have changed my position, I am almost in a form, where I can; I am in the form of this definition. Now, what I can do is, apply supermom over x star right. So, x because x star is absent here, so this I would not be affected by the supremum operation.

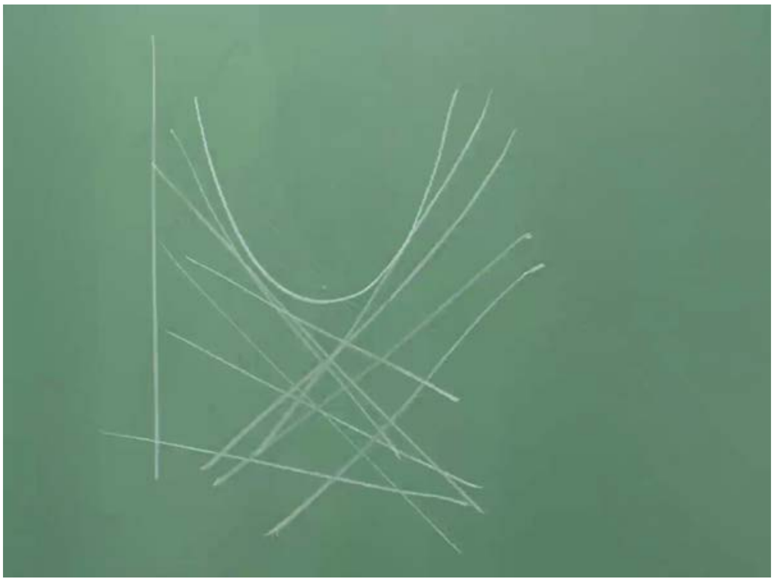
(No audio from 13:37 to 14:00)

So, these are also classes of affine functions, which are lying below f. And what you have proved, what we have proved for any proper convex function, this is nothing but the definition of f star star (x star) - f star f star, which is same as I just written it in the standard way, it expose would write in convex analysis of optimization. So, f (x) is always sorry not x star, this is x; this is the definition of f star star x. So f, so we know the inequalities goes in this way, what about the reverse; can I prove that if f is proper and lowers, am I my continuous? Then so if f is proper, and lower semi continuous, so I do not want to get into this too much technicalities rather than giving with the ideas.

(No audio from 15:19 to 15:27)

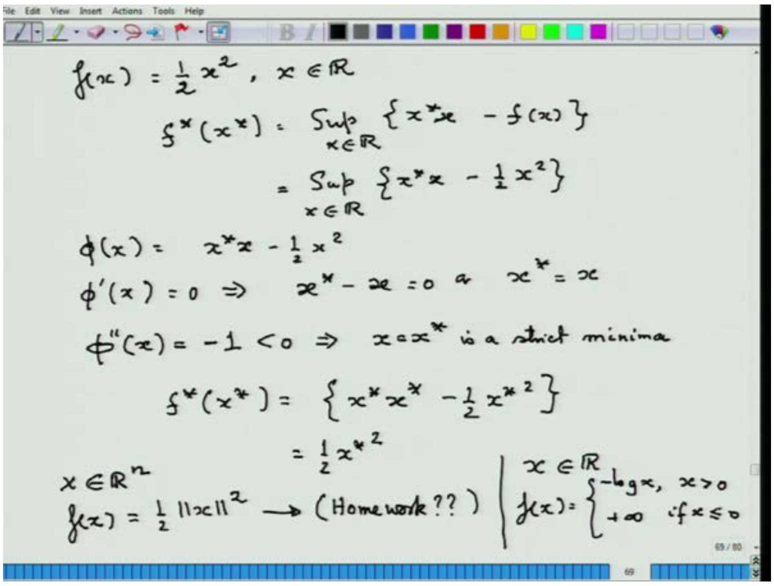
We have f of x is f star star of x for all x; so, this is a fundamental result.

(Refer Slide Time: 15:51)



So, if you take the supremum of all these affine functions lying below the convex function, then what you essentially get is your f star. So, if you take the supremum of all this affine functions, you take the envelope, the envelope is f by the very definition, what we are proving by this this thing what it shows that the upper envelope of this is nothing but f star and f star star is actually f.

(Refer Slide Time: 16:39)



So, let us try to figure out some conjugates. Let me take the function f (x) half x square x is in R, let us see what happen. So, this would be conjugate is f star (x star) supremum over x in R. Now here, the inner product is just multiplication x star into x minus f of x; so, basically supremum…

(No audio from 17:19 to 17:31)

A very important thing to observe here is the conjugate function is usually not differentiable, because it is expressed as the suprema of affine functions. Now, if you observe this little, little fact, then what I have to do is to find the maximum. So, I have gone back to high school I have gone back to high school mathematics again. So now, I have a function phi of x, that x star is some fixed number; and I have to minimize it over x. So to do that, I first have to check the first order in necessary optimality condition.

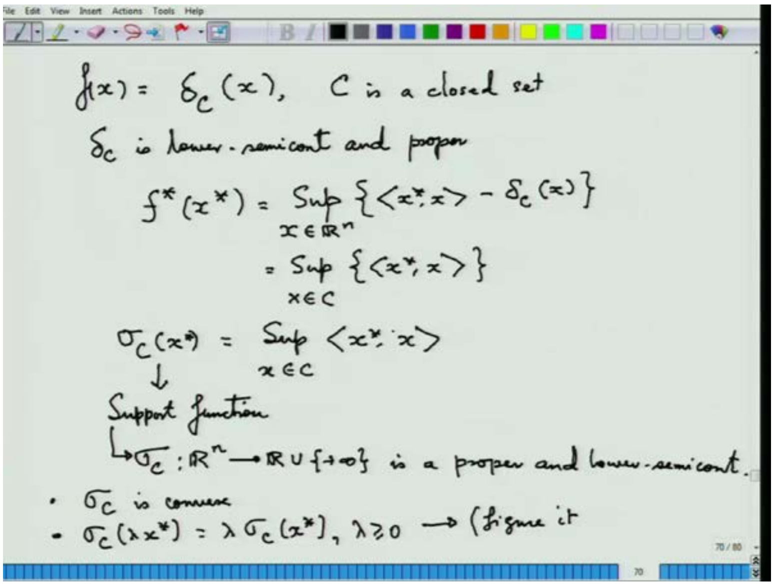
Find an x, which would satisfy phi dash of x or the derivative of phi at x is equal to 0. So, we would imply that this would imply, we are take the derivative of this one x star minus x is 0 or x star is equal to x. Now, is this the supremum? To do so, you have to take the second derivative of x. If I take the second derivative of x, what is happening? I take the second derivative, it will become minus of 1, for whatever x you choose. Now this is strictly less than 0, which implies that x star, x equal to x star is a strict maxima. Here the problem was that concave maximization problem, which is just the opposite. If f is convex, then minus f is concave or f is concave means, minus f is convex. So, this is a concave function, whose global minima or local minima are global, so you have a strict minima.

I did not much define, what is a strict minima? I will, I think we did in the very early days. So, it means that for any x naught equal to x star, f of x is strictly bigger than f of x star that is what it means. Now, so my supremum value has been evaluated at the point x star; and this is nothing but x star, x is nothing but x star. So this is, this value is x star, x star minus half x star square. So, this is exactly equal to half x star. So, so this conjugate and the function are the same, so this is the only example, where the conjugate and the function are the same. So, your homework would be, to now go to the slightly higher dimension which is N. So, now, you take x element of R n, and consider the function f (x)…

(No audio from 21:03 to 20:13)

Tell me tomorrow, what did you, what you find; is it the same thing? Is would be back to half norm square x or something else? Quite interesting. Another homework, x is in R and I define the function f (x) is log of x minus log of x convex function; if x is strictly bigger than 0, and I define this as plus infinity, if x is less than or equal to 0; your job is to find the conjugate, find f star. Here also the question is to find f star; do not need to find f star star, you you can find it anyway.

(Refer Slide Time: 22:38)



Now, let me the curious enough and ask you f (x) is equal to the indicated function of C at x, where C is the closed set; del C is lower semi continuous and proper. Let us, let me try to calculate, f star x star in this particular case.

(No audio from 23:19 to 23:33)

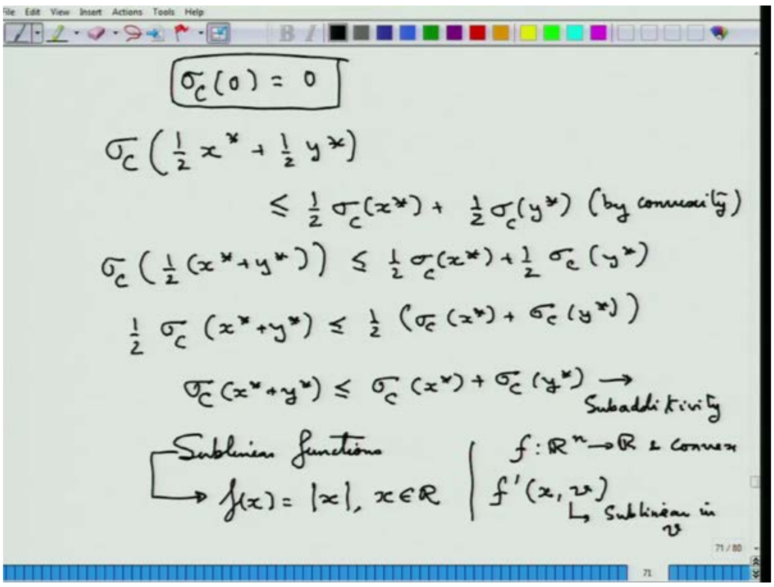
Now you know that, whenever x is not in C, this is plus infinity, and so this will giving minus infinity; and so I really have to bother about those x which are in C, and for those x which are in C, this is 0. So, this will be reduced to… Now, if you look at this expression, this expression, this thing, where you have a fixed x star, and you are varying the x over C, this is called the support function of the set C that is this is usually denoted as the support function; so it is a support function. So, sigma C is a proper and lower semi continuous convex function; why it is lower semi continuous? Ponder over this question; why I am directly writing it is lower semi continuous? Too much of math I guess, but we will get that life going slowly.

So, this sigma C, which is called the support function, has an interesting property; sigma C is convex of course, and sigma C satisfies this property; if you multiply lambda with x star, and this is nothing but lambda times sigma C x star. You might ask me hey guy, come on. What would happen if your sigma C x star is plus infinity and you take lambda to be 0? So, I claimed that this is, this must be, this is true; you you can figure this out, figure it out; but you have to be cautious.

Once you are playing with infinity, it is like playing with fire a bit. It does a lot of good things to you, but you have to be very cautious that is that is the whole thing. So, you figure it out, this is just putting things in the definition, because if you put lambda x star here, you take the supremum of linear function, lambda will just come out; it does not matter; it will be supremum of lambda or if it is lambda time supremum of f is if lambda is bigger than equal to 0; this is the standard thing for the finite value case. So, this will be true, if you in for this case.

Now, you have to be very, very careful. If lambda, this is pus infinity; if lambda is 0, this is 0. Now if lambda C is 0, and suppose sigma C x star is plus infinity, what would be this side? Now we have to invoke what we have spoken about plus infinity or when we introduced extended valued functions. We have said that we due apologies to all those who do not agree, we would consider lambda into 0 into plus infinity to be 0.

(Refer Slide Time: 27:55)



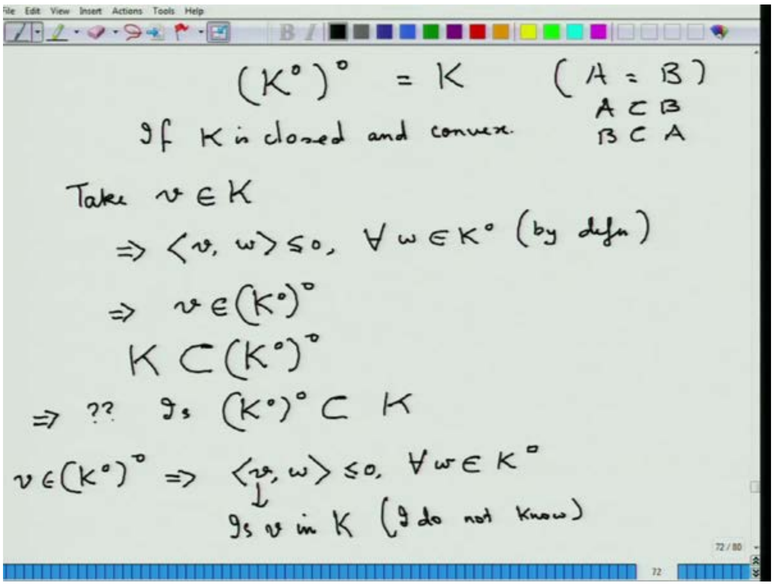
So, what we would have is that, under our convention that lambda, if lambda is equal to 0, then lambda into plus infinity is 0. This would give me the fact that sigma C of 0 is 0. So, under that convention, you get this thing. So such a function, which satisfies these two properties is called the sub linear function; not really linear, slightly linear. And of course, convexity will immediately give you using half, you will immediately get by using this and this is the following thing.

So, sigma C half of x star plus half of y star, y convexity of sigma is half of y convexity. So now, I can write this again as sigma C half of x star plus y star is less than half of sigma x star plus half of sorry sigma C sigma C y star. Now, by the rule of positive homogeneity, let this rule is called positive homogeneity, but non-negative homogeneity, but does not matter. So, this will immediately tell me half of sigma C x star plus y star is less than equal to half of sigma C x star plus sigma C y star. So, this would obviously, give me…

(No audio from 30:09 to 30:23)

So, this property is called subadditivity. So, if a convex function is positively homogeneous has this property, then it will have this property; such functions like the support functions are called sub linear functions. An important example is if f (x) is equal to the absolute value of x. Another further important example is if f is from R n to R and convex, of course; f dash the directional derivative at a fixed x, becomes a sub linear function of v, as you vary the direction v. So, it is sub linear in v. Now this property is so crucial for going beyond convexity. So, we have learnt a bit about conjugates, some interesting facts.

(Refer Slide Time: 32:10)



We will see them later on when we speak about duality theory after a few days. But now let me go back and try to full fill the promise that I would prove that K star of star is K, if K is closed and convex; that is what I promised yesterday; and I am keeping my promise; because I am speaking to you through electronic media and not face to face, so it is quite possible for me to check, how many of you have actually done the job; some of you might have, some of you might not have. But it is instructive to try; it is not that everything that you do in mathematics, you will succeed in doing your problem. But trying is fun; it is the journey is the goal and not possibly the destination always.

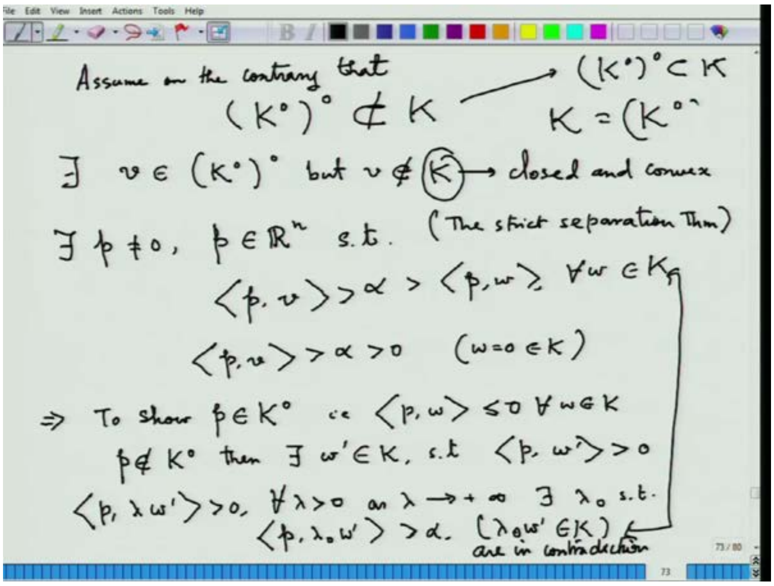
So, when you take this attempt to solve a problem, you learn a lot of things by doing that; and that would help you to understand many, many things in this subject; anything as which is mathematical. Now, how you prove that a set A is equal to B? You first prove this A belongs to B, when A is a subset of B, and then you prove that B is a subset of A. So that that means, both A and B are same; that is every member of A is a member of B, and every member of B is a member of A. So, there members must be the same members.

Now, take K, take v in K. So, this would imply by the definition of K polar, v of w is less than equal to 0 for all w in K polar; because is not K, when in w you take in K by definition w of v is less than equal to 0 for all V in K. So, for whatever v you have chosen, this would be true for any w, you taken k polar; that is by definition of K polar by definition. So, this would imply that v is element of K polar polar, again by definition of polarity of a cone. Now, this would simply mean the following that I have taken a V in K, and got shown that the same view belongs to this. So K belongs to K polar polar.

The next question of course, is the following. I am asking you this question is K polar polar, this is bit tricky. So, let us try take an element v in K polar polar. So, this would imply v of w is less than equal to 0, for all w in K polar; that is the definition. But that does not tell me that v is in K; it does not… If v is in K, then this would be true. But this no way tells me that v has got to be in K. Every element in K satisfy this; but there is no way that I can show from here that v is in K; is v in K, I do not know.

So, this approach like immediately what I have done is something, which is not going to work. So, what shall we do? So, when I direct looking approach fails, you have to take the indirect approach in mathematics. Proves by contradiction however, early they might look or possibly of the last resort that you have to take. So, let us do one thing. So, I will do a proof by contradiction.

(Refer Slide Time: 37:25)



So, I will claim, I will assume to on the contrary…

(No audio from 37:24 to 37:35)

That… assume on the contrary that K polar polar is not a subset of K. So, there exist a v in k polar polar, but v is not in k. So I am assuming the opposite thing. So, I will say that if I assume this, I will reach a contradiction that is what is called proof by contradiction. Like or dislike the fact is that we like it or dislike it the fact is that there is no other way, but to resort to this approach. Thus the direct way as we have seen was not give us any conclusion. So, let me see what can be done.

So, now v is not in K, and K is the closed convex cone. So, you can apply the separation theorem that you will learn quite a bit before, may be some 6, 7 plus early before. So, you might have forget forgotten it. So, it is possible that you have forgotten it; so let us apply it again. So, there would exist a p not equal to 0, p belonging to R n, such that p of v is strictly bigger than alpha, is strictly bigger than p of w, for all w in K; this is the strict separation theorem. So, here we have applied the strict separation theorem.

(No audio from 39:39 to 39:51)

Now, once we have known this fact, what should we do now? Now, this is p; by the way, if you get confused. Now, I can put w equal to 0. So, then I have p of v, by setting w is equal to 0, because which is in the cone K. Now, I want to prove that to show that p is in K naught that is p of w is less than equal to 0. Now, if p of w, p is not in K naught; then if p is in K naught, if p w is less than equal to 0, for all w in K; and p is in K naught right. By the very definition, if this happens, it would imply p is in K naught if and only if. So, if p is not in K naught, then there exists a w dash in K, such that p of w is w dash is strictly bigger than 0. Now w dash is an element of the cone K; so, this means p of lambda w dash is strictly greater than 0 for all lambda, strictly greater than 0.

Now, as lambda tends to infinity, this will become bigger and bigger, so big; that it can actually cross the value of alpha. So, there would exist lambda naught such that P of lambda naught w dash would be strictly bigger than alpha. In fact, and lambda naught W dash is obviously, an element of K. So, this would be in direct contradiction, this, this and this, are in contradiction. So means, I am disproving this fact, but which has to hold by separation theorem, which means what there is the contradiction means that what are assumed is not correct, and K polar polar is also a subset of K, proving that K is nothing but K polar polar.

So tomorrow, we will give some more examples of conjugate functions; talk about something called Fenchel duality; but just give you a hint. But we will tomorrow concentrate on the thing, which I am keeping and writing at this moment, saddle point conditions. And that would lead us gradually to the study of linear programming, and we will we will spend quite a bit of time delving into the pleasures of linear programming. Thank you very much.