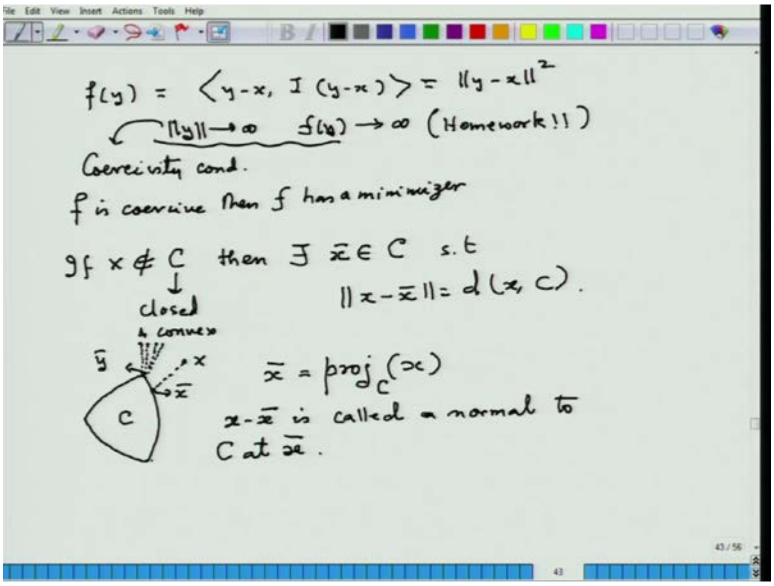
Convex Optimization

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

Lecture No. # 11

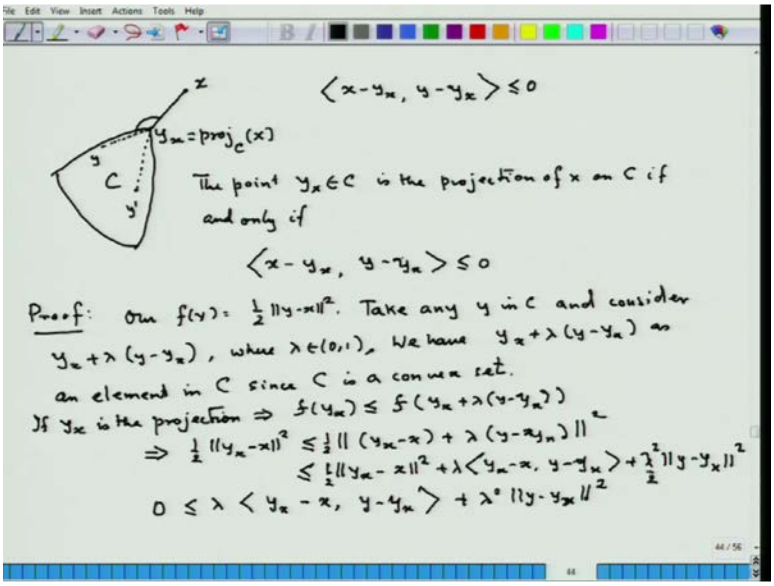
Welcome once second to this course on convex optimization. Yesterday, we ended our course talking about projections and normals. Because we had symbolized a sub differential of the indicator function as as in normal cone to set at that point. So, its very important to know what is actually that normal and normal cone.

(Refer Slide Time: 00:52)



Now, it is where we had finished in the last lecture that; for every point outside there is a to a closed convex set C, there is a point on the basically on the boundary of C, such that the distance x minus x bar. The norm of that provides the minimum distance of distance, that gives you the distance of x from the set C.

(Refer Slide Time: 01:18)



Now, let us look at certain interesting properties the normals follow. Again take this convex set; take this point here, which is a projection point of this x, which is and this is x bar, and or let may be I will call it, I will just call this as y x, because y is what we were denoting as point in C, and y x is the projection of the point x. So, we will write this as projection of on C of x on C. And now, you take any other point; say y in see and join it with y x. Then this angle, if you have observed is an obtuse angle.

So, which means that the inner product of this, and this vector is less than equal to 0. And these two for any y you take. So, again you have other y. So, y dash you will have the same thing. But interestingly this is a necessary, and sufficient condition for of the point y x to be the projection of x on C. So, let me write down a very important result. The point y x element of C is the projection of x on C, if and only if; x minus y x in a product y minus y x is less than 0. So, its a if and only if condition. So, its instructive to go through the proof of this. So, we will do the proof.

So, how do we go about doing the proof of this fact. The proof is as follows that, you have taken in our case our f (y) is of course, this f is dependent on the x naturally. Now, take any y, y in C and consider, y x plus lambda times y minus y x, where lambda is a number between 0 and 1. So, because these y x, and y both are in C; we have y x plus lambda y minus y x, as an element in C. Since, C is a convex set.

(No audio from 05:10 to 05:18)

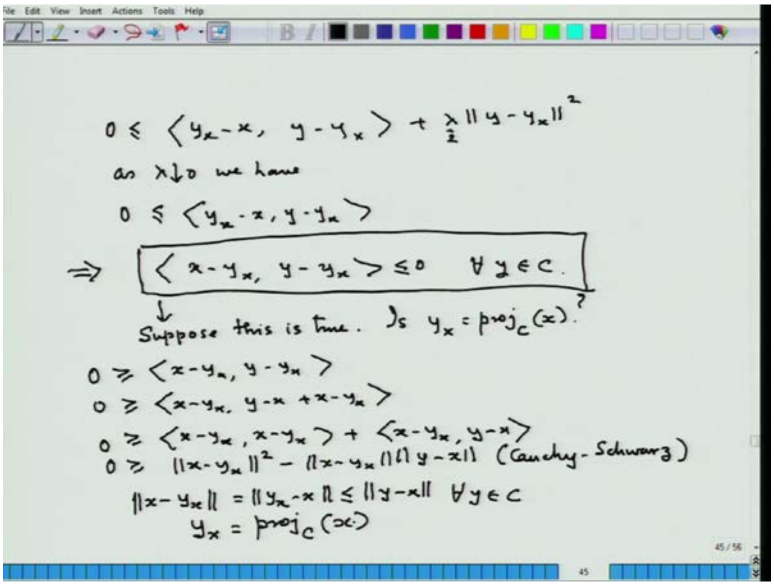
Now, what do I know, suppose I know that y x is a projection. So, let we are studying with the fact that y x is a projection. So, here we have started with a fact that y x is a projection; so, if y x is the projection.

(No audio from 05:33 to 05:40)

So, it would imply that f of y x must be the minimum; y x is a minimize. The projection bonds are obviously, minimizer of this function. So, minimum... So, because these are element in C, so it will obviously, sorry. Now, writing down this fine, the function of form this would give me half of norm y x minus x whole square less than norm of y x minus x plus lambda y minus y x.

So, this if I write down. So, this will become, if I open up the norm, norm y x minus of this was half half. So, half norm y x minus x whole square plus half into 2 into y x minus x lambda outside, and y minus y x plus lambda square times norm y minus y x whole square. Now, this will cancel off. So, you would have 0 greater than equal to lambda times, norm y x minus x and y minus y x plus lambda square. Now, if I divide. Now I can divide, because lambda is between 0 and 1; I can divide both sides by lambda.

(Refer Slide Time: 07:52)



And I will obtain this expression will obtain 0 is bigger than y x minus x into y minus y x plus lambda times norm y minus y x whole square. Now, as lambda goes to 0, we have now its positive and going down to 0; we have…

(No audio from 08:19 to 08:30)

This would imply x minus y x in a product y minus y x is less than equal to 0. Now, since y was arbitrary; it was just any element in C. So, this is true for all y in C. Now, suppose I have this result this result has is been given to me. So, suppose this is true.

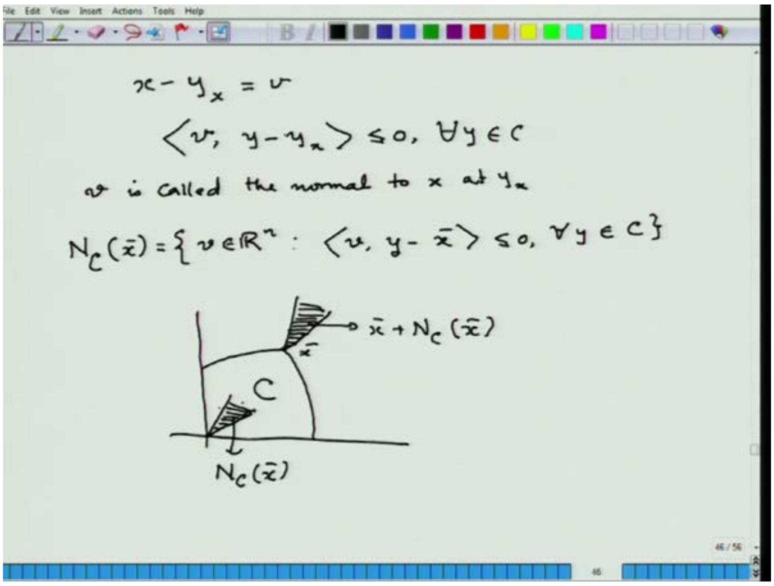
(No audio from 09:03 to 09:12)

The question is y x projection of x on C, that is the question. So, we have not proved that yes, it is; so, there is a beautiful result which as a this beautiful correspondence. So, what you have is a following - zero is bigger than x minus y x, y minus y x. So, again you can write this as y minus x plus x minus y x. So, again this would become if you do the, in a product x minus y x, x minus y x plus x minus y x into y minus x. Now, this is nothing but x minus y x whole square, and by using the Cauchy Schwarz inequality this is obviously, greater than minus of norm x minus y x.

(No audio from 10:27 to 10:40)

Cauchy Schwarz. Now, this means that, I can now take on this sign and cancel out this x minus y x, and so finally, I will get the following inequality. I will get that norm of x minus y x, which is same as norm of y x minus x is less than equal to norm of y minus x. But y was any arbitrary element in C. So, it is true for all y in C. So, this relation would hold for all y in C, because this is true for all y in C. So, this will immediately mean the following that y x is the projection, because it solves the projection problem; and that is it. So, if I talk about the normal cone: You now see, how does that idea of normal cone comes.

(Refer Slide Time: 11:52)



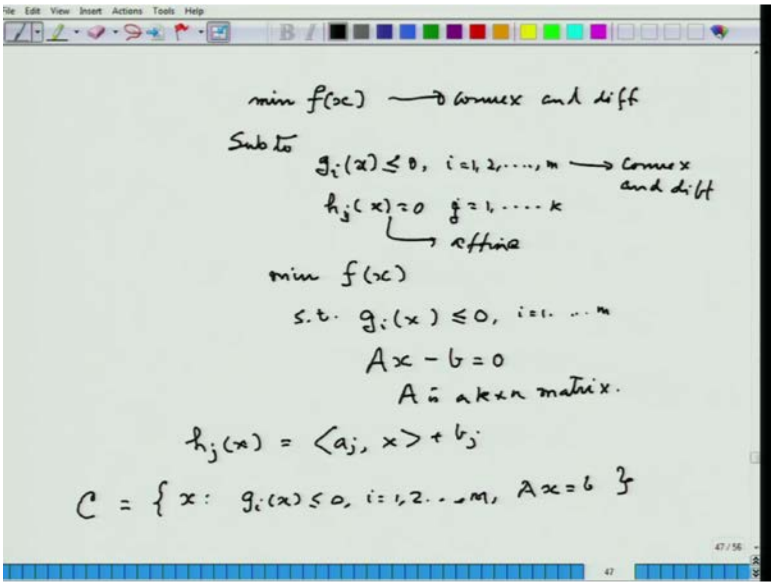
Now, this x minus y x this term, if I write it as v; then I have this expression. Now, this v is called as we have already seen earlier; v is called the normal, this x minus x bar which is v in our cases called the normal to C at x bar. So, v is called the normal to C at x bar; C at y x sorry not x bar y x. So, as you have observed from the picture, that they can be more than one 1 x to which y x is the projection. So, they can be more than one such v's for which this is true.

So, let us take the collection of all such v's in R n; such that... So, I take a point x bar. So, I take a point x bar, and see and collect take a collection of all those v and R n's as a this is true. Now, if you look at this set, this set follows the definition of a cone; if you take any v then lambda v is also an element of this set. So, this is set is called the normal cone - the cone of normal basically to the convex sets C at the point x bar. A normal cone as we have seen is a vehicle for representing optimality optimality conditions. And in you see how the sub differential notion of the indicator function is linked, directly to this geometrical thing. So, it also shows that at a certain level the sub differential is also a very geometric thing, and also brings in a (( )) so much on the geometry of the space.

And as a result of which a lot of enriching happens, and lot of interesting things get revealed, because this in the play between analysis and geometry. Now, once you have this, the question is of course, if I put 0 0 would satisfied this equation, and so zero must be in the normal cone. That's fine; suppose, I was set like this; this is my set, this is my set C. And here you draw the normal cone, see like this; this my x bar. But you might ask me then where is your 0 - 0 is here, but the x bar is not 0. Actually, this is nothing but a translation of the usual original normal cone of, to draw the normal cone taking zero as the base point or the vertex, you draw lines parallel to this one.

So, this is actually your normal cone to C at x bar; and this is nothing but a translate of this to the point x bar, that is x bar plus N c x bar. So, this tells you that; if I take the origin of this two x bar, then this is what will happen. Then the normal cone is exactly this; if the origin is now x bar. Now, once I know about this the question would, I would have some very interesting calculus about normal cone which I will show you, which will help us to do a lot of interesting things, and which would help us to do a to write downs are in optimality conditions. Let us go back, and let us slightly complicate the minimization problem - the convex minimization problem that we had studied.

(Refer Slide Time: 16:07)



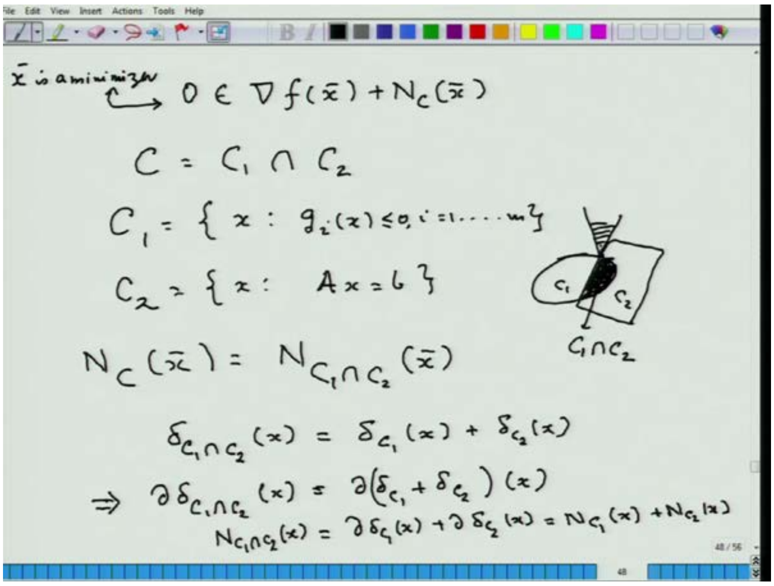
So, suppose I want to take minimize function f x, subject to m inequality constants. And remaining is affine constants h i (x), h j (x) equal to 0. So, I can now make this compact; this is convex, and differentiable; this is this functions are convex and differentiable. And affine functions are anywhere, these functions are affine. So, these are convex and differentiable. This problem can be equivalently set up like this minimize, f x subject to…

(No audio from 17:26 to 17:37)

N i, where A is a k cross n matrix; now how do I write this, because each h i, h j x can be written as a j x plus b j. Now, there are k such constraints. So, linear plus some translation, that is what the meaning of an affine functions. Now, once you have this, because otherwise you know the full feasible set won not be a convex set. So, my C here is now described like this. So, if you look at this, if I take a j to be a row of a matrix; so, a 1 into dot dot dot a k. So, they form a k cross n matrix which is a here, and b 1, b 2, b k is the number of this constant. So, they would form a vector b. Now, this C here.

(No audio from 18:45 to 19:06)

(Refer Slide Time: 19:16)



These are what we can write in my C. So, if I want to write optimality condition it becomes 0 is an element of grad of f (x), suppose x bar is the minimum. So, x bar is a minimizer, if and only if this holds. Now, you see computing the the whole question now lies how to compute, we have studied studied the Karush Kunh Tucker conditions, just in the last lecture I guess. This is the what we have studied the Karush Kunh Tucker conditions. Now, how do I reach the Karush Kunh Tucker condition here, we have come to the Karush Kunh Tucker condition by applying the max function the calculus rule.

Now suppose, I am not having in hand this calculus rule; what am I supposed to do. Now, you observe that this set C is quite a complicated set; it is not an such a easy set that, you can write you can compute the normal cone. How do I compute the normal cone for this set. Do I need certain conditions to compute the normal cone, but at the end first let me write this C as two simpler sets. So, let me write C as C 1 intersection C 2 where both of them are convex set, where C 1 only deals with the equality - inequality constraint. And C 2 deals with the equality constraint, affine constraint.

So, normal cone to C at x bar is now written as normal cone to C 1 intersection C 2. So, how can you compute the normal cone. Basically, you have two convex sets, and you have this is the intersection zone; and take any point here or say here. And you are supposed to compute the normal cone. You see, here again this relation between the sub differential of the indicator function, and the normal cone - that the normal cone is nothing but the sub differential of the indicator function would become very, very important. The indicator function of C 1 intersection C 2 is same as writing indicator function of C 1 x plus indicator function of C 2 at x.

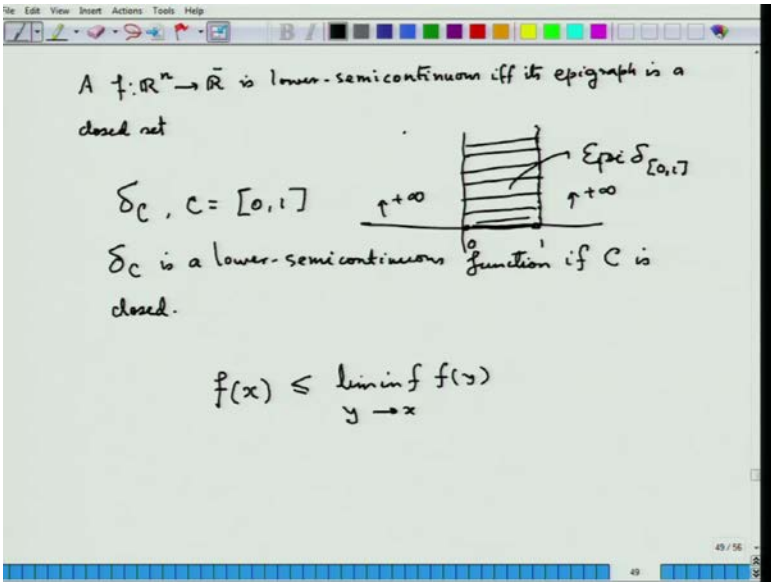
Because you see if there is an x which is (( )) C 1 and C 2; then this would be both would be 0, and hence this will be 0. If some equality holds; suppose x is in C 1, but not in C 2. Then x is in x is not in C 1 intersection C 2. So, this would be infinity. So, x is in C 1, so this would be 0, but this would be infinity, but 0 and plus infinity is infinity by our law; so, this inequality holds. Now, this means the sub differential.

(No audio from 22:59 to 23:18)

So, again these are proper lower semi continuous convex functions. And of course, I am not defined much about lower semi continuous convex functions, we just we will do it very soon in much more detail.

Let me just tell you. The sum rule that we have learned there is applicable; assume that C 1 and C 2 is having a interiors. So, an interior of C 1 intersection interior of C 2 is non empty, like this one. So, then you can apply the sum rule - the calculus rule to write; and this is nothing but… And what is this? This is nothing but the normal cone to… See how the calculus rule for sub differential, gives you a calculus rule for the for the normal cone. And that’s that’s the beauty the link between the geometry and the analysis. And, let me tell you what do I mean by lower semi continuous function, otherwise every time I use the term you might be bit worried.

(Refer Slide Time: 24:43)



Is lowers semi semi continuous, if its epigraph, if and only if, so I write double f. If and only if its epigraph is closed. So, look at the function d C for a convex set, take take an take an example take C to be 0, 1. So, that function would look like this. That function would become 0 between 0 and 1, and then it will be plus infinity otherwise. So, the... So, this is the epigraph and of course, the epigraph is closed; the epigraph is of course, closed and hence, del C is a lower semi continuous function.

So, for any closed set C, this is a lower semi continuous function. Of course, the set has to be closed; if the set is open then it won not be true.

(No audio from 26:23 to 26:33)

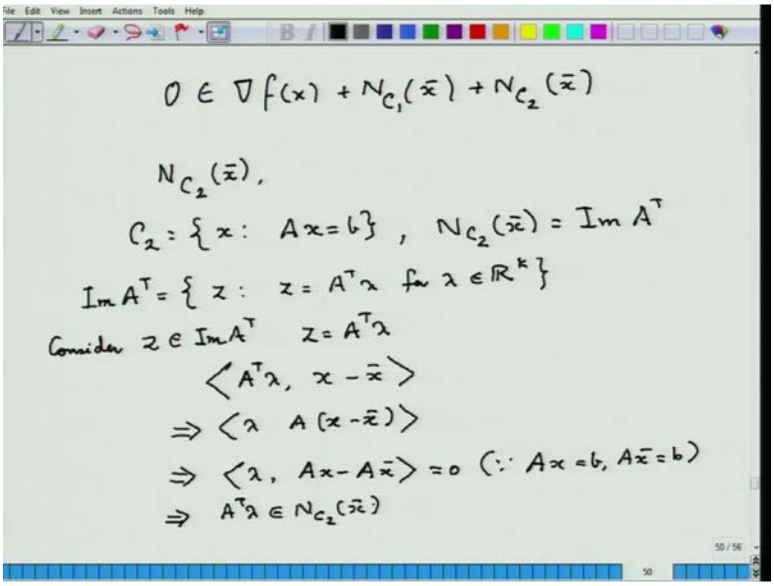
Now, I have given away geometric definition of lower semi continuity, where there is also a more mathematical definition, but looking at the audience, I would not like to put it. But for those who are mathematically oriented, just I want to remind them this simply means that. Say if it is lower semi continuous at x, this is what it means. So, this is a motion of limine, which I not be cleared to many many many students, and which we do not want to deal with.

So, for us geometrical definition is a most; easy definition, because all the examples that we will do we can hand it with this. And you need not get so much, walked up with this lower semi continuity business, because at the end we will be dealing with continuous functions - functions from R n to all which are continuous. And of course, they are all nice and helpful things - things which you understand pretty well.

Now, my problem becomes optimality condition for my problem becomes.

(No audio from 27:38 to 27:56)

(Refer Slide Time: 27:40)

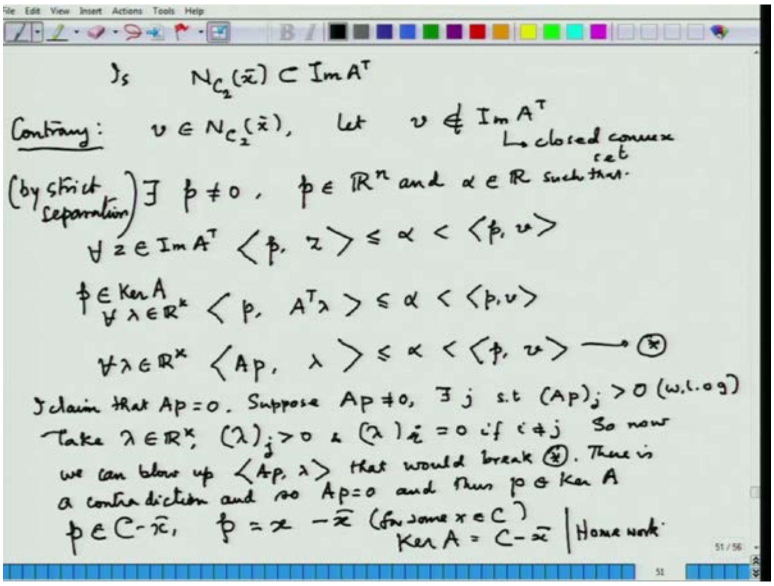


So, in order to get the Kunh Tucker conditions or Karush Kunh Tucker conditions, I have to compute this and compute this. So, what comes out is that the lagrangian multipliers or the Kunh Tucker multipler those lambda y, lambda i's. They are not just ordinary multipliers or auxiliary variables that you use, to convert our problem from constant form to unconstant form. But they are deeply linked with the geometry of the space that you are working in the feasible set. So, our task would be to compute this, as well as task could be to compute this.

So, let me first today do the easier one; that is let me compute N C 2 x; where C 2 is set of all x, such that A x equal to b. What we will show is that N C 2 x bar is nothing but the image of A transpose. A transpose, we would have linear mapping every matrix is a linear map and vice versa. And in order to do so, how would I go about doing so. So, (( )) what is what is image of A transpose? If I have the image of A transpose here, then it is the set of all z. So, as z is equal to A transpose lambda for lambda element of R k.

So, first let me take a z; consider z element of image of A transpose. Then z is equal to A transpose lambda. So, let me compute this quantity. So, if this is my v. So, the normal normal if this must be in the normal cone; that means, see whether it is in the normal cone. Then take any x in C and take the x bar; compute this. And this must be less than equal to 0 for this has to be, if this has to be in the normal cone. So, this would imply lambda times A of, but this means A x minus A x bar, but if this both are in C then both are equal to b. So, since now, this would imply that A transpose lambda is element of the normal cone to C 2 at x bar. So, now I have to look for the reverse one.

(Refer Slide Time: 31:21)



So, my next question is normal cone to C 2 at x bar, subset of image of A transpose. Let us try out suppose, I am having this problem; how will I try out this problem. That is the question. The way to try out this problem is as follows; you can it is very difficult to take a v, and prove that it is a image of A transpose. Because v in C 2 is… C 2 is a point where A x is equal to b, and nothing you know about C 2 the structure is not very clear. So, how do you go about it.

You go about it in this following pattern. Let let me take a v in N C to x bar. And let, if I am assuming on the contri contour on the I am assuming, I am assuming something which is contraring to what I want to prove. Take v here, and let v not be element of image of A transpose. The image of A transpose is a closed set, and we had already spoken about the separation theorems, in the very beginning which says that. If v is not in the image of A transpose; since image of A transpose is a closed convex set. Then what should I do, we should do the following.

So, I can now apply the separation theorem. So, there exists a p not equal to 0; p element of R n, such that p of z in the supremo over; I I will just do it much, because you are talking about streak separation. Because we can talk about streak separation by streak separation; that means, I down the principle, because you might just forget it while we are doing by streak separation, where the streak separation there is a p in R n; and alpha in R such that, this is less than equal to alpha strictly less than p of v.

Now, what does the z for and this is true for all z element of image of A transpose. Now, what I want to show from here is that p is element of kernell of A that is that is exactly what I want to show. So now, what we would do is the following; is that here let us what is z. z is an element of the form A transpose lambda, for all lambda in R k. So, this would be satisfied for all lambda in R k.

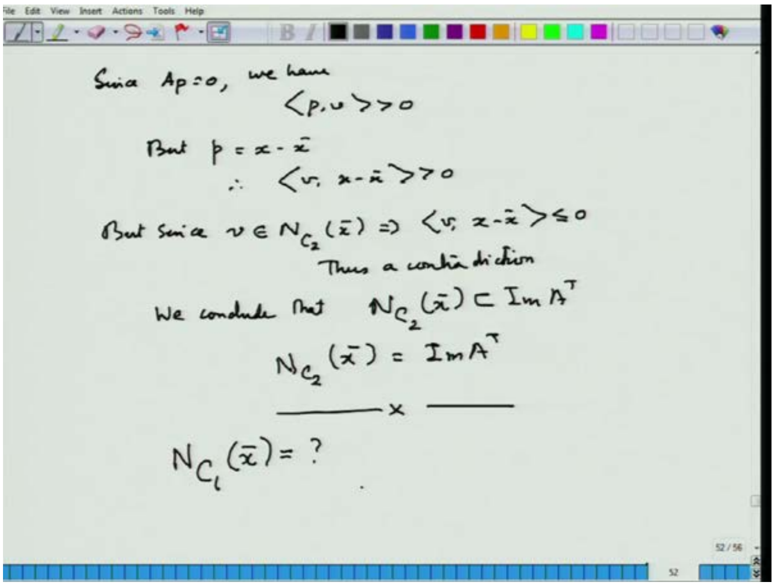
(No audio from 35:04 to 35:28)

Now, how do I claim that A p is in kernell of A, I claim A p is equal to 0. If I claim such a fact then I need to prove it. You see suppose A p is not equal to 0. So, then there exist a say j say jth component, such that the jth component of A p j is either greater than 0 or less than 0. So, let me take this to be strictly greater than 0; does not matter. This can be done without loss of generality, you can do with negative also it does not matter. This is true for all lambda. So, since it is true for all lambda in R k, take lambda in R k such that lambda j, jth component of lambda is positive and lambda any other component, lambda ith component of lambda lambda i is equal to 0, if i is not equal to j.

Then you can make this lambda, as big as you like as big as you like as big as you like. So, this will be positive number which can grow grow grow, in such a way that will blow up and it will cross this value of alpha, cross this value of p v, and break the separation inequality. So now, what we said is that we can blow up A p lambda. Since, I can blow up A p lambda. So, that would break that would break the inequality, that would break this inequality. Let us write star break. So, I cannot do that, I cannot break that. So, there is a contradiction; and so, A p is equal to 0 and thus p is element of kernell of A.

Now, I give you a home work to prove that p is actually element of C minus x bar, that is p can be expressed as some x element of C minus x bar for some x in C. In fact, you can show that kernell of A is C minus x bar, and this is independent of the x bar, this is easy exercise, say I just do it as a home work. So, what I now get is that, once I this is A p is equal to 0; I will get p of v strictly greater than 0.

(Refer Slide Time: 38:59)



Since, A p is equal to 0, we have p v strictly greater than 0, but p is equal to x minus x bar; therefore, v of x minus x bar is strictly greater than 0. But since, v is element of. So, here it will be C 2 sorry not C C 2, C 2 x, C 2. So, v is element of N C 2 x bar, and that would imply v x minus x bar is less than equal to 0, and then thus a contradiction. So, there our initial claim that there is a v in N C 2 x bar, and v is not in the image of A transpose is wrong; and hence so, we conclude that N C 2 x bar is also a subset of image of A transpose.

So, we have proved both ways; this is subset of this, and these are subset of this. And so, we will have N C 2 of x bar is equal to image of A transpose. So, with this we end the talk here, and in the next lecture we would like to compute this one, which would be very interesting, and in see the role of the status condition. Thank you very much.