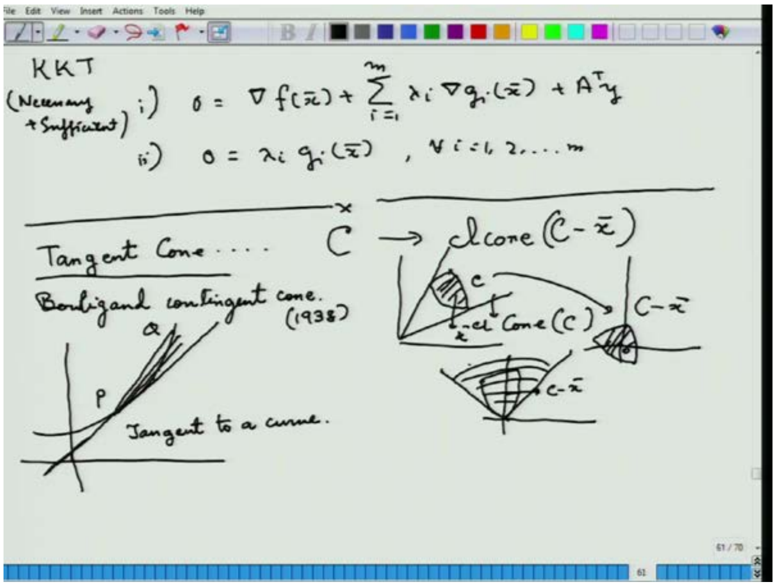
Convex Optimization
Prof. Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture No. # 13

So, welcome back once again, and as promised we would start discussing the notion of tangent cones today.

(Refer Slide Time: 00:28)



In the last class, if we go back to what I have said, then you would observe that we ended up with this sort of Karush Kuhn Tucker necessary and sufficient condition; I have proved the necessary part, and you you can prove the sufficient part your-self by employing the definitions of convexity f and g, and the affines of this class of constraints x equal to b. So, we have said that a major role here has been played by an object called closer of the cone of the convex set mine translated to point x bar. So, basically what happens is that that sort of cone has particular geometry structure, the closure of the cone of…

So, if you take a convex of C, we saw that we were, this cone had played an important role; actually, you know, you you take any set C a convex set like this, and you want to generate a cone, that is take any element, and take every element of C, and pass rays through the origin through those points. So, basically at the end, you would end up with something like this. So, if this set is closed, then of course, you will have, for example, if if this king point is not on the convex set C, then the resulting cone will not be closed. So, this is what is the cone or may be cone generator by the set C; C minus x bar means you have taken a point x bar, and then you are doing C minus x bar. So, 0 becomes a part of the cone; basically basically translate the cone, (( )) what the position of x bar is now taken up by the point 0. So, origin becomes a part of this.

So, you make this translation C minus x bar, where x bar is some element here. This particular cone or the particular cone that you will now see coming out of this, has for a convex set, has some meaning; for example, if it is C minus x bar, where x bar is not in the inside, but it is in one of the boundaries like here, then you would you would actually get a cone like this; sorry you get set like this; this is you are C minus x bar, and then, at the point 0 you are trying to draw some sort of close convex cone.

So, now this has link with the notion of tangent cone or the Bouligand tangent cone, Bouligand contingent cone.

(No audio from 03:24 to 03:32)

So, it was discovered in 1938 by introducing, 1938 by Bouligand for very different purpose not for optimization, but it has certainly come to play a important role in a optimization. See, if you look at the very basic notion of a tangent, which we have learned in high school is that, if you have a curve like this on a Cartesian plane, then you take two points P and Q, and if you want draw tangent to this curve at P, so, tangent is a line that touches the curve only at the point P, the tangent P, and does not touch at any other point. So, what do you essentially do, is to join this chords PQ, and then take this succession of points coming towards P. So, you basically then, start joining P with those succession of points, and finally, you end up with the line like this. That is a limit; so that is your tangent.

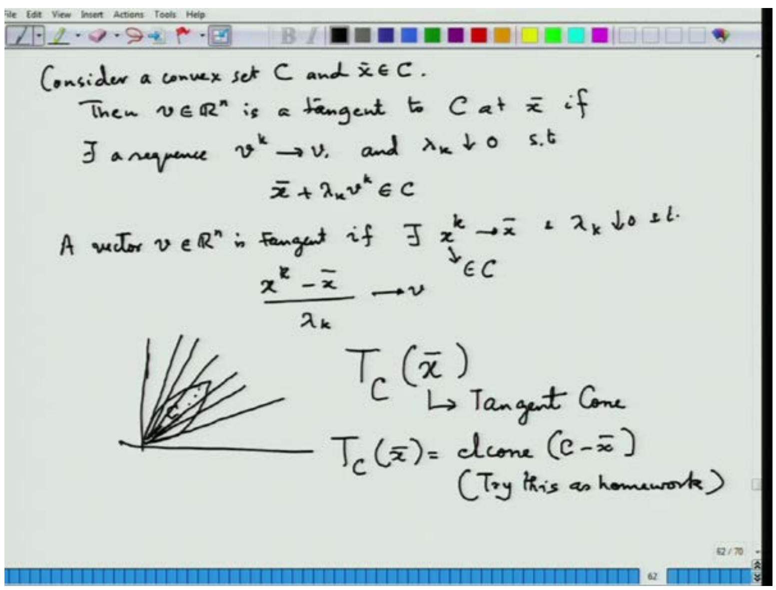
Now, this is the tangent to a curve.

(No audio from 04:34 to 04:44)

What about the tangent to a convex set? So, what do you mean by tangent to a convex set? So, in this case, the notion of tangent C, here you see, here you are talking about a smooth movement from Q to P, but what about if the set is discrete, or means, just some few points, can you talk about a tangent to it? Of course, here you considering convex set, we need not bother about all the things.

So, if you look at this picture once again of this C minus x bar, and I am trying to draw a cone generated by this C minus x bar, if you see all these lines are some sort of tangents. So, the tangent cone to C at x bar.

(Refer Slide Time: 05:34)



So, consider a convex at C, and then, V element of R N is a tangent to C at x bar, if there exists a sequence, this symbol is a short form for the word - there exists. A sequence V k of vectors going to V, and lambda K which is real numbers going down to 0, that is they are all positive going down to 0, such that x bar plus lambda k V k. So, any vector V, which satisfies this property would be called a tangent to the set C at a x bar.

Now observe that here this, when I am having a real sequence, my sequencing, that is my indexing is on the bottom, and when I am talking about a vector sequence here, my indexing is at an unknown on the top. So, this can have an equivalent definition that, of vector V element of R n is tangent if there exists x k, sequence x k converging to x bar, and lambda k converging to 0, such that x k minus x bar by lambda k converges to V. See, this calls comes from this fact, on both of these are equivalent, you can check it out.

If you look at this diagram, the ideas are actually built in here, the tangent to a chord, that , when the difference when x k the sequence is coming to x bar, so, this is a x bar, the sequence is coming down to x bar, then they would be a lambda k which controls it. So it does not just become this, thus vector does not become 0, but it is controlled. So, is a deduction, it is control by lambda k, such a way that ultimately the limit of this vectors are pointing in a certain direction.

So, if you have a tangent, if you any set here, I will assume that this set for example, this nice looking convex set C. So, then what is a tangent? Of course, they are two curves, you draw these two curves, but I can talk about any sequence, coming to this point in C. So, what what is important this this x K must be in C. And I am looking at this differences x K minus x bar x K minus x bar, and then modulating it with lambda. So, these vectors that will generated by other sequences in C, they can be now thought of as a tangent vector; may not necessarily be thought of as a tangent vector; those who are more ((taking)) on, more habituated with the school geometry, coordinate geometry, might be having little bit of problem, but let me tell you that in convex geometry or the geometry of optimization, everything is one-sided, because convexity is one-sided, because in inequality and as a result of which your geometry is not both sided, like the tangent here, your geometry here is also one sided. So, this if you look at it, I can call the set of all these vectors, they form a cone, and that is called the tangent cone V at x bar to C at x bar, the tangent cone.

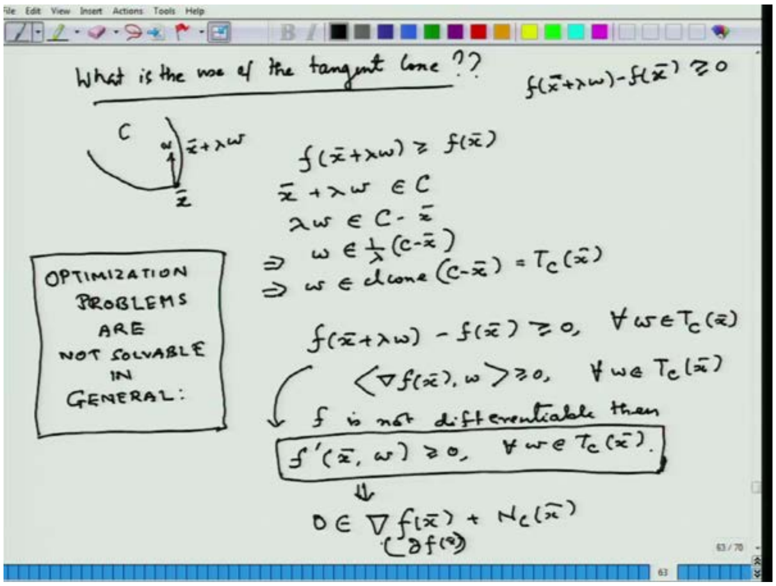
And, so, set of all V(s) form this set of set which is called, which is a cone, and which we call the tangent cone. Now even C is convex; see, this can be defined for any set C, but here we are gluing ourselves to a convex set, because essentially we are talking about convex optimization. So, rather than complicating the facts of life, if it is essential that we just concentrate on the things at hand, the things that we really are interested in, that is minimizing a convex function of a convex set. Now, if you look at this thing, the question is and if you see the picture that we have drawn of tangents, this tangent here, this can be proved that, (No audio from 10:44 to

10:56) try this as homework or I will show it is to tomorrow. So, if C is a convex set, the tangent cone has a beautiful and clear expression.

So, what is this tangent cone use full for…

(No audio from 11:18 to 11:34)

(Refer Slide Time: 11:21)



There is the first question. So, tangent cone is useful possible in, using its for its use in optimality condition; you might say other how many optimality conditions here going to use, because you know optimality conditions, we have been studied from the beginning, but in optimization largely story of finding optimality condition, that would lead you to compute the points. Now, you have already see in the normal cone. Of course, a tangent cone has some relationship with the normal cone, and the optimality conditions gets you know re dressed, as you change from normal to tangent cone.

If you go back to over story, you will see that with the normal cone, the sub differentially is involved or the sub differentially itself is involved with the normal cone ; zero belongs to del f(x) not plus the normal cone or zero element of grad affects plus normal cone. The directional derivative is associated with the tangent cone. So, the tangent cone geometry is expressed with the directional derivative, and the normal cone geometry is expressed with the gradient or sub differentially itself. So, there is some (( )) duality between this two things, normal cone and tangent cone. So, the duality involved in this expression, is also expressed by the tangent cone or normal cone, and this dual relations will come very soon. But let us see what the tangent cone can do.

Now, there is another interesting way to look at the tangent - the notion of tangency how can you arrives such a notion of tangency. So, we are studying optimization. So, we are essentially talking about variation, that is if I want to say that some x bar is actually a minima, then I should be able to say, how it is varying with the respect to other points; that is if I am talked having a point x bar here, and I want to say that this is a minimum; at then, if I move of bit away from x bar and steel be in C, say I have this direction w, and this new point is a x bar plus some lambda w; if this is a minimum, I should always for a function f, I guess.

Then this most always with true; now, the important thing is that the direction that have choosing, has to be of this formed that x bar plus lambda w must be in the set C. So, lambda of w must be in the set C minus x bar; shall I take lambda to be positive of course, some moving along this direction. So, its positive. So, because lambda is positive, this is element of one by lambda times. So, this simply says that w is element of closure of the cone, of occurs in the cone enhance in the closure. So, any w for which I can measure this change this variation, that is f of x bar plus lambda w minus f(x) bar, I am measuring this variation.

So, if x bar is a minimum. So, moving along those direction which will keep these new points, these points in the set C would actually should give me, this difference to be greater than equal to 0. So, this is the variation, that we are trying to measure. So, but over which points in along which direction, I should measure the variation; those direction should belong to the tangent cone. So, the very notion of variation which is very very fundamental optimization, that is the very basic definition of what you mean by optima is intimately linked with the notion of tangent cone. And hence, it is important that we have some time to see, how this idea is linked with optimization; in the sense how its linked to the optimality conditions. Now, suppose I have chosen this w over this for this convex function this w, from the tangent.

And then, what I am having is that if x bar is minimum, then…

(No audio from 16:34 to 16:47)

Actually, if you observe, again remove this by x minus x bar again take any x here do x minus x bar, and if a move along those directions x minus x bar or lambda x minus x bar this will be actually true. So, what would happen, if I apply differentiable the property of this, I will simply get.

(No audio from 17:05 to 17:14)

So, this is my optimality condition. And in fact, if the function f is not differentiable.

(No audio from 17:21 to 17:30)

Then, the directional derivative, in the direction x bar, in the at the point x bar in the direction w was be greater than 0 for every w. Now, this is a very very important fact, that we must note this is a important fact. So here you see, how the tangent cone is playing a role; then what is it link with the normal cone; so, are we having two different optimality conditions; can we move from one of this to the other, can we from here, can I come from here to this condition. These are this replaced with sub differential.

So, can I connect, can I come to such conditions. See, this thing would require some more more tours into the land of convex analysis; convexities of fundamental what what so, intrinsically beautiful; that never mind even if you genius, who are in the audience or may be (( )) from management sciences whose just respond from plug in get the answered, and declare non-optimal point is optimal one. Can also take some pleasure from the beauty of this subject; I would like to emphasize that, when I say that you are pushing a bottom, and tell pull telling an non-optimal point is optimal; it is not, with an intension to ridicule any one.

The intension is cleared, that most people who practice optimization at least India in my country, hardly really know the mathematics behind optimization, and unless you aware of the mathematics behind optimization is not really possible to appreciate or really know, what algorithm to use on, what sort of problem and when I would stop an algorithm.

See, its very very important to know at the very outside, because what happens is that; you would just have a problem, you would have some mat lab, f mean cone something would put in a problem, and they will through an answered. And you would accept it, and possible declare it has a solution, optimal point; the very important fact, that you should learned, when you study optimization - when you stress it again, if you read (( )) book introductory lectures on convex optimization; this is what you rights in the very beginning that. Optimizations problems in general are not solvable; of course, you can solve toy problems like minimizing x x square our whole are, minimizing f(x) equal to x when x is are between minus one in plus one, these are toy problem.

But real problems when I am talking about optimization problems, I am talking about real problems. So, quadratic optimization problem, you (( )) solve it. If a real quadratic optimization problem with large number of variables are in large number constant, is not so easy to do, we cannot just give just can solve it; if even if it is a convex problem, you what you get is in an approximate answer; unless there are very special structures, very very special situations. So, they are not solvable in general.

(No audio from 21:29 to 21:41)

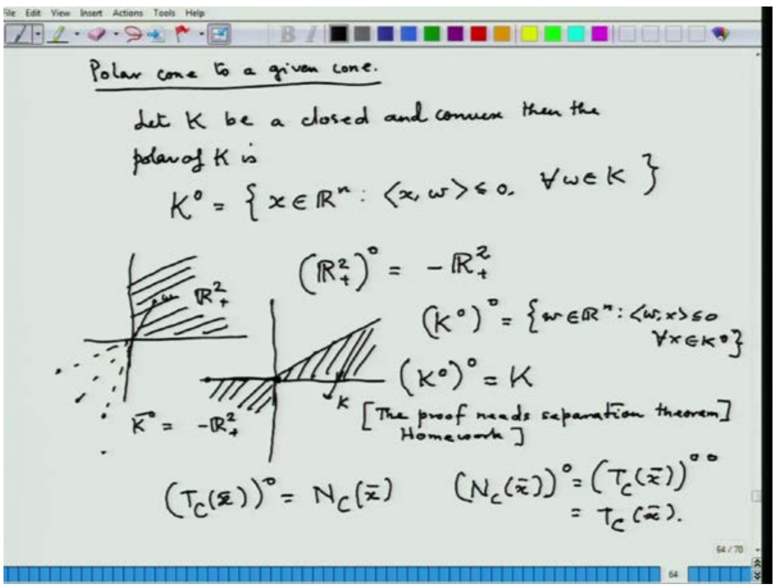
Then what is then needed to study optimization; I am in not to find the actual minimum; then what you how do I how do I solve my problem, you said minimize this this this, and you just say that are you cannot find the minimizer. So, why are you wasting all this times, we have lot of optimality condition etcetera. Theory by the ways always nice in the sense that, there are lot of assumption; we assume the perfect world for perfect theory, but the real world is not super effects. So, theory is a only a guiding tool to tell you, how good is the approximation that you have taken.

So, what you take at the end of running algorithms, in is is an approximation; how good is your approximation is what theory will tell you, how does your points satisfy the Karush Khun Tucker condition, atleast how how for is it from the Karush Khun Tucker, actual Karush Khun Tucker solution. So, these are the things, which makes the makes learning of the theory important, whether your engineer in management science are something else or of course, for mathematics or you are you are in computer science.

So, we will come into algorithms very soon (( )) a very soon, but may be in few more lectures. But we will given extensive time on algorithm, but it is very very important understand at the very outside, the theories always guiding you, to do to generate those algorithms. Now, once I have done this, let us see see in order that you go back and four between two different type of optimality condition; one expression terms of tangent cone and directional derivative, another expression terms of the gradient dot the sub differential, and the normal cone; you need to know update about notion of polar cone.

(No audio from 23:43 to 23:55)

(Refer Slide Time: 23:44)



So, let K be a close convex cone, I do not need a close convex cone to give a general definition of polar, but over cones are essentially close and convex most cases; so, we just assume that fact. Then the polar of K.

(No audio from 24:19 to 24:29)

Set of all X in R n, such that x times w is less than equal to 0, for all w in K. So, if that is my definition, let me take the simplest example R 2 plus. So, take any w in K, R 2 plus take any w here, I want all those vectors v which will make an obtuse angle with this, that is angle has to be more than 90 degrees. So, this all points here, will all the vectors here, in this third quadrant will formed an obtuse angle; if you take some vector here, it will form an acute angle any vector here, will form an acute angle. So, essentially if I right R 2 plus the positive (( )) non-negative (( )); and take it spooler, it is nothing but minus of R 2 plus, suppose you take any cone like this.

So, you have to finds point which will make obtuse angle, it cannot make acute angle; if take any point here, it is making a obtuse obtuse angle; take any point here is being obtuse angle. Here again, but this particular cone, where I half the, I have just drawn a line and rub the other part, it is steel again minus R 2 plus. So, if you take this as K and this is again K polar; then line must just look and ask this question, what is the polar of the polar. I just curious question nothing much two why about. So, by if I define it, this consist all w in R n, such that w times x is less than equal to 0 for all x in K polar that is a definition, I am just repeating the definition moving K v K polar that is all; that means, that I am asking the question what is this? Interesting fact is that K polar polar is equal to K, K polar polar is equal to K; that is interesting thing. So, how would you do the proof - the proof needs separation theorem.

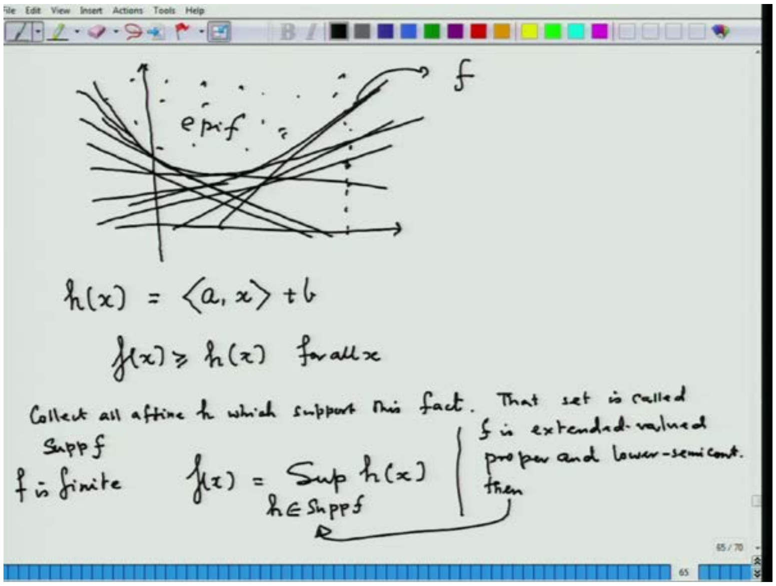
(No audio from 27:33 to 27:52)

So, try this is homework, tomorrow I am almost going to take an assignment class by trying to show that the tangent cone is that, an also try to proof this one. Know I am telling you, all this story of polar and all this things; does it have any relationship with the our subjected hand, tangent cone, normal cones and optimization. The interesting fact is this.

If you take the polar, take the polar of the tangent cone; you come up with the normal cone. That is an fascinating, that you just the geometry; you just have to look at into the look into the geometry the… So, these two things also geometrically linked; that this is a set of all vectors v which makes obtuse angle with elements of the tangent cone. And now, you take the normal cone, you take the polar of this; which is polar of the closed and convex cone T C x bar.

So, you take the polar of the polar, and then by using this thing, we get this is nothing but the tangent cone. This is quite interesting, but now my question would be how do I get back between two optimality conditions. For that we really have to wait a bit. Now, you have a once we have a discuss all this optimality conditions this sort of geometry, that is involved in the whole discussion.

(Refer Slide Time: 29:56)



Let us get up bit bold, and look in to the structure of the convex functions, still more in more detail in more detail. You look at a convex function, you see you look at this epigraph basically epi if; now you have at a every point of this close convex set, there are drawn a continuous convex nice look in convex function; and so, its finite valued. So, it is continuous and so, at every point you are having are tangent hypo plane, a supporting hypo plane, and you see what are this supporting hypo planes; the supporting hypo planes are define by affine functions.

So, basically you take any affine function which is lying below the graph of the convex function, and then take some sort of a upper upper envelope of those collection of affine functions, they would actually generate you the epigraph - itself the convex function itself, they will generate you the curve the graph of the function. So, what happens is that; take take a convex function f this is f ,and then take affine functions of this form; such that f of x is equal to is greater than equal to h x for all x. So, here we are talking about affine functions, we are not talking about something vertical. You are not talking about line like this, which we are talks about the in affine function x equal to 0. (( )) in the R 2 spaces the affine function x equal to 0. So, it is 1, 0 into x y is equal to 0. So, you are not talking about these sort of, in I am I am not talking about line which tells which as gives you x equal to 0.

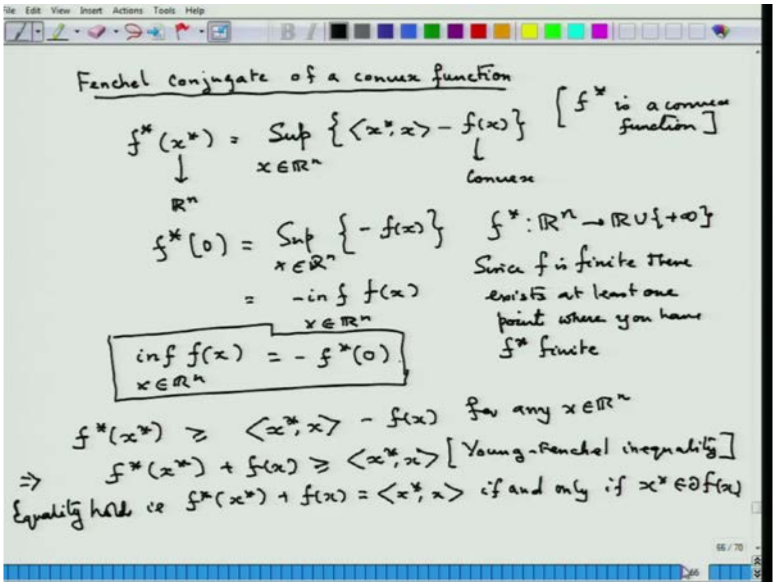
I am talking about, non vertical hyper planes. (( )) vertical hyper plane, then you disk cutting through the epigraph. So, once your cutting through the epigraph, it is impossible to make this sort of assertions; see you cannot, cut through the epigraph, you have to be low the epigraph has a result of which we have to slant down, because if you do not slant down; if your non vertical, you cannot maintain this thing. So, you see what I have drawn is all this affine function has there functional values at any point, you take this affine function. The functional values of f is here, and the functional value of the affine one is here, the functional value f is here. So, this is always holding for this class.

So, what I am trying to says that, you collect the set of wall h which supports this. So, collect all affine h which support this; that set is called support set of f; is so… It's a collection of all affixing functions h which is minor, which is major i is by f or which minor i is in the f in the sense that, f is always bigger.

Now, if you look at now the forereaching result in convex analysis, which actually gives you a global structure of convex functions is the following; it tells you that of function if it is a continuous function like this, if r if f is finite; you always have with h belonging to support of f. So, what do you taken x, and you computed that x the values of all the h as which are in the set; and then take the supreme of all those values and that will be exactly give you f x, because you see if we take a value here; a one of the h x, another h x, another h x, and you go up and you finally, reach the graph. So, that is exactly what is is the story.

So, this is not only talking talks about the case one is finite infect; if f is extended valued proper, and lower semi continuous; I have already told your lower semi continuous is - lower semi continuous. Then this holds. But we are not going to do is so much, bother about this lower semi continuous business, because that is is essentially technical thing, but is a very help full technical thing, but for our study is we can just we happy with continuous convex function, because that is more less than thing that will handle in practice. This useful device that we had just spoken about this is result has interesting consequences; now, let us look at the this newly new cones convex functions which we construct out of this very idea, that if you take the supermom of functions measures by an convex functions; that you get back to the convex function itself. So, if you take the envelope, upper envelope (( )) family of affine functions; what you get is a convex function.

(Refer Slide Time: 36:44)



So, we are going to now define the Fenchel conjugate or a convex function.

(No audio from 36:45 to 37:08)

So, define a function f star, which takes any element x star in R n x star in R n, and you compute the value of this function as follows; takes supermum of all x in R n. Now, you see that two ways to prove that this is a convex function. So, my first job is of f is a convex function, you can prove that (( )) without even f is an convex over case is just convex. Now, if you all observe this. So, what I am having is the family of affine functions; every fixed x this is an affine function in an x star, and by what we have just seen by what we have just seen, the upper envelope would provide us with a convex functions. So, this is a convex function, but you can directly try compute it. So, f star is a convex function; one was you wondering why all this operations. Now, there is something interesting about this operation; let me, look at f star 0, its supermom extra 0. So, its 0, I do not use this.

So, this means the negative of the infimum of f x over R n.

(No audio from 39:25 to 39:43)

So, this is a very very fundamental fact, which means that if I have a finite infimum, then x star is a proper function; see f star can never take the value infinity, because if you if f is the finite convex function which we are taking, any put any x star, then they will be if forgiven x; this values always finite. So, it will never take the superimum cannot minus infinity, R n (( )) is an non-empty. So, a superimum could be plus infinity, it could blow up the superimum it. So, if the function f is a finite infimum, then you are garneted that this is a proper function; there will at least one point which is 0 at the function is finite. But it can be prove that, if you have a convex; if you have finite function f it, would be any way of proper function.

So, this f star is actually in a extended valued function. But since f is finite, there exists at least one point; where you have f star finite. So, f is a nice a quasi-function - nice function which has a minimum over x is square then f star 0 is 0. We will try to find the usefulness of this, and in fact tomorrow we will come with the raptor of examples of f and f star; and its implications of in optimization, and this function has bought in a new area- call computational convex analysis which is are thriving area, which one might get interested, somebody in mathematics point just get interested by this whole thing or somebody you, once to work in the interface math, and its associated computations. So, if you look at this, I am (( )) to give examples, I am not going to be examples today for this particular case f, and extra appear.

So, if you observed if you observed f star x, by very definition is bigger than x star x minus f(x) for any x in R n. This would employee f star x star plus f of x is bigger than x star x. So, this is called the young Fenchel inequality; you might thing that has is no use giving this, such triviality a name, because is just obvious, from this definition. Yeah its obvious or what is not obvious is the fact that, equality holds that is f star x star plus f of x, this x star of x; if and only if, x star belongs to the sub differential of f at x. So, if f is differentiable then, x star is exactly equal to the gradient of f at x. So, with this very basic introduction to the notion of conjugate - conjugates play a very important role, because when I have define a convex function f, over a convex minimize minimization problem or a minimizing a convex function over a convex set, there is some other convex problem which is defined or another cone K f problem which is define in interms of the negative of f star; which is call the dual problem, which comes into play which will soon talk about little later.

And then, this is exactly of very very very basic thing what we have done. So, tomorrow we have going to examples, and even prove what we had said that, we have going to prove K not not is K, and let us see what we are given even the home work; we have given you to prove this one. So, this is what the tangent cone when C is convex, this is what the tangent cone is, (( )), this definition. And of course, we want go to prove this fact, but before we do so, we would give a raptor of examples from appear of f and x star, but now the question is what would happen if f is not just finite; it is extended value, but proper; at it is finite at least one point.

So, then can you define this, answer is a yes; you can define it is similar way, and you can show that even f is proper, that it has at least for finite value and is never taking the minus infinity value; then f star is also proper function. Because once we for example, the interesting question, what happens this is the indicated function of the set C. Then, what is the f star. So, we will try to figure out of those things, and (( )) interesting. So, and has lot of linkages with optimality condition, that we also figure out tomorrow. So, with this we end today's talk, thank you very much.