Convex Optimization

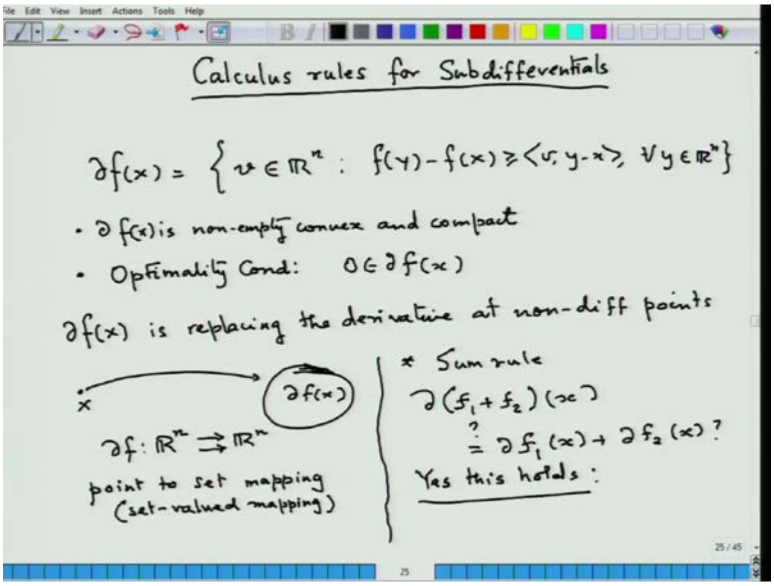
Prof. Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Module No. # 01

Lecture No. # 09

Good evening once again, and once again welcome to this course on convex optimization; in the last lecture, I had spoken about how can we surmount the difficulty on non-differentiability, which you can take as the bane or a boon in convex optimization, possibly convex optimization is richer, because of the presence of non-differentiability of convex functions and the fact that optimalize precisely at those points.

(Refer Slide Time: 00:49)

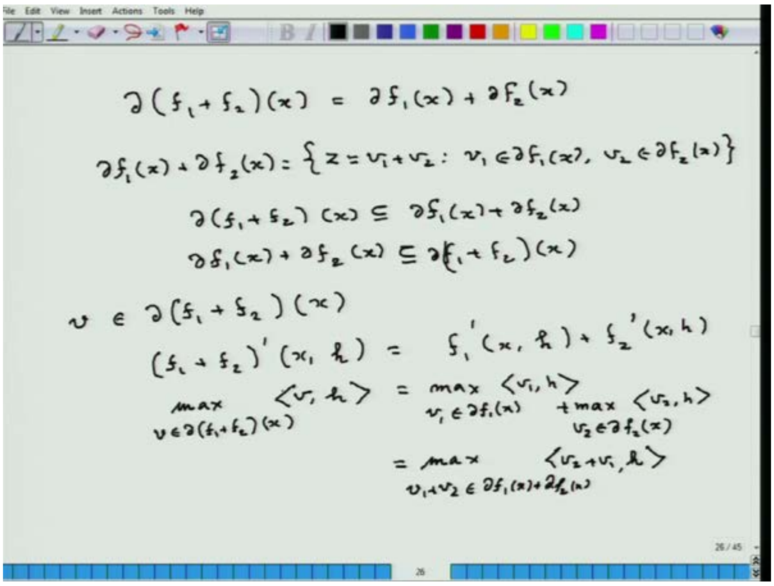


Now, what I want to reiterate is that we are introduced the notion of a sub differentiable of a convex function; given a function from R into R, the sub differentiable is the collection of all v in R n, such that this condition holds for every y. Now, what we had known about this case is that del f is del f x is non-empty convex and compact and further, we had written down an optimality condition in the unconstrained case, the necessary and sufficient optimality condition of formers rule in the convex case, which is the famous condition 0 belonging to del of f x, this is the necessary and sufficient for a point x to be a global minimum of the convex function over R, the whole space R n.

Now, if I claim that del f (x) is replacing the derivative at non-differentiable points, then I need to show that it exhibits a calculus, just like a ordinary derivative exhibits a calculus that if you take the sum of two derivatives, which I take the sum of two functions and take the derivative, it is nothing but the derivative of the sum of the individual functions. So, there is also composition of two functions and you take taking the derivative of that, whether such rules, do work in this case; it is very important to know of course, we can give a lot of examples, which we will come very soon, but it is very important to know at that, at this outset that this mapping the sub-differential takes an element X in R n and puts it into a convex compact set.

So, del f symbolically is written is a mapping from R n to R n, but it is a special type of mapping, it is called a point to set mapping or set valued mapping anyway whatever you want to use; now remember, so when we are trying to build a calculus for this class, this sort of things, we have to really appreciate the fact that every time we are dealing with sets, we are no longer dealing with numbers or we are no longer dealing with vectors. So, these are sets and so it is very important that we take utmost care, when we are talking about sets. So, if I want the calculus rule; so what sort of calculus rule I can think about; let me think about one rule, which is the sum rule. So, you take two convex functions, f 1 and f 2; and take their sub-differential at a point x to you have this rule, (( )) this do you have something like this; there is a question. (No audio from 05:03 to 05:11)

(Refer Slide Time: 05:40)



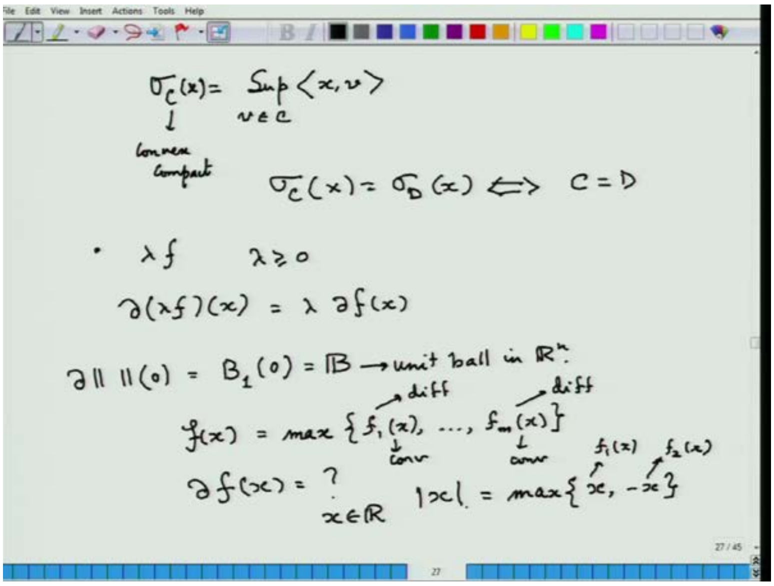
The answer amazingly is yes this holds, but let me tell you that here this is a set, and this is a set, see if I go back and try to understand this writing; then here we are talking about the equality of two sets. So, here for example, what is the meaning of this set; this set means is a collection of all z, which is written as v 1 plus v 2, such that v 1 belongs to… And v 2 belongs to del f 2 x. Now, if I want to show the relation back and forth. So, what I really have to show is that first I have to show that (No audio from 06:39 to 06:51) and then of course, I have to show that… So, a a subset of b and b subset of a. So, a is a is equal to b; now all these are quiet easy to… This is quiet easy to prove you have to just write down the definition.

This would need a bit of little bit of hard work, see if we if I take a v from what would happen. So, can I do something with it; all idea would be to use the directional derivative and the fact that the directional derivative acts as a support function, that is the better idea would be to use this fact, that if I take the directional derivative of x, directional derivative of f 1 plus f 2, any direction h and this is nothing but this is the simple calculation, which you can easily do. So, once I have this, what does this mean; what would this mean; it would mean that this is nothing but max of all v elements of del f 1 plus f 2 x v of h is equal to max of… Max because is a compact set. So, v 1 of del f 1 x v 1 h plus max of I am just writing down this definition, which you this relation, which you already know, we have studied in the last class.

(No audio from 09:08 to 09:20)

So, form this, can we conclude that this is equal to this; we can actually conclude that, because this fact, in fact, one can write this whole thing in a much more compact way (No audio from 09:44 to 09:57) or v 1 plus v 2 h (( )) So, if the support functions of two sets are equal. So, this is called the support function.

(Refer Slide Time: 10:08)



So, this is called the support function if you have a two, I have a convex, take a convex compact set, in that, it works for any convex set, but let us just talk about convex compact set; take a convex compact set, and take this v, some can calculating at any, I will may be I will remove any v and make it look like x; from computing this at any x in R n. So, how do I compute I compute and it not right super max is same thing we will consists of all v element of C, such that I want to find the minimum of this problem or the maximum of this function x of v, x is fixed over all element of C.

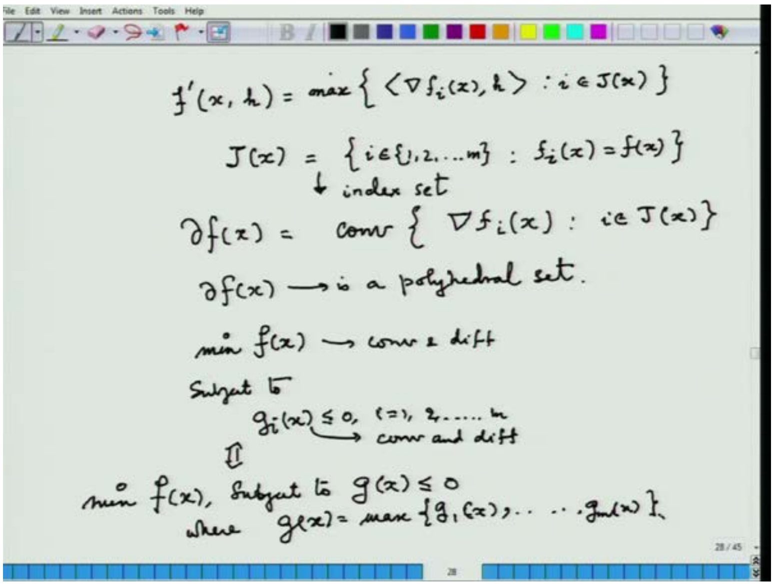
So, what you can prove is that if you have two convex compact sets, and you have this. So, what you can prove that if you have convex compact sets, then this is equal to this, if and only if C is equal to D, this fact would lead us from to prove, would lead us to conclude this fact this one. (No audio from 11:27 to 11:37) So, one of the major calculus rules will looks like it works for sub-differentials. Now, I would like to know that,if you have multiply a function by a constant, and then you take its derivative and the constant comes out and you new derivative of the new function is the constant time, the derivative of the old function. So, if I take lambda of f, where lambda is greater than or equal to 0, you see if I am take a negative lambda then, lambda of f need not be convex anymore. So, now I ask myself the question, what how would you I compute this, sub differentially; if I know the sub differential of f, then by the way, you can understand it is not so easy to always compute sub differential.

Now, for example, if you take the function norm of x; then the norm - the euclidean norm is only non-differentiable at the point 0; and I leave it to you as a homework to prove that this is nothing, but the ball of radius 1, centered at 0 or more and D this notation B is called the unit ball in R n. You can also find a composition rule and all sort of stuff, but there is something, which actually differentiates convex calculus from the standard ordinary differential calculus, and that is the notion of a max function, that is here we are going to talk about a convex function given as a maximum of say m, other finite value would convex functions.

Now, suppose all these are even differentiable, all this convex functions are differentiable; my query would be to find, now here is a point where we are absolutely coming into a different paradigm, because whenever you have a function created out of maximization or minimization of some other class of other finite number of functions, then the resulting function is not differentiable in general. So, there is nothing, you cannot define the derivative of grad f, there is no way you can do. So, if you have these are these are differentiable convex functions, then this is the convex function, but need not be differentiable And also I want to emphasize, why this function is important, because most of the non-spoon is an non-differentiability in convex optimization and optimization particular arises not in some arbitrary manner, but in some very ordered fashion by taking maximization or minimization of few functions.

For example, if you look at the standard well known example, which we had been giving in the last lecture, the absolute value of x, when x is R; so, this can be written as max of x and minus x, these two functions. So, this is my f 1 (x) and this is my f 2 (x), you see even the most standard well known thing can be expressed as a max function, so how do I find its sub differential. So in fact, I am just now going to write both the directional derivative as well as the sub differential.

(Refer Slide Time: 15:50)



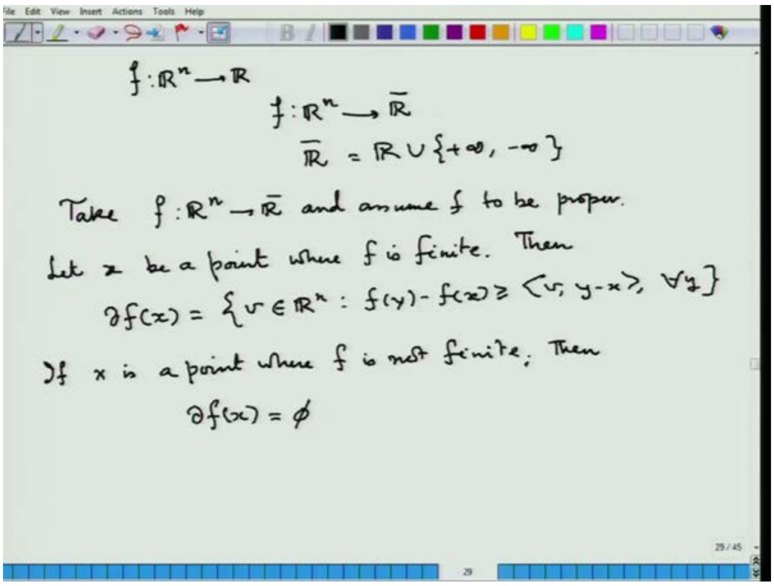
So, let me first write down for that class of function the directional derivative, (No audio from 15:51 to 16:00) this is max of the gradient of f i (x) in the direction h, that is the inner product of grad f i into h, where i belongs to J of x bar sorry J of x. So, what is this J of x J of x is some sort of indexing of where the maximum is attained? The J of x is the set of all indexes i from 1 to m. So, among those indexes, find the one such that f i (x) is equal to f (x), because if you put an x, there would be one function at least for which, where the maximum value would be attained, because we just have finite number of functions, there could be more than one. So, for given x, there could be two indexes, here four indexes, for another m there would be say only one index.

So, it all depends, in fact, on the choice of your x. So, we have to make this set depend on x, this is called the indexing set or index set; now if you look into this very carefully, you will observe something that what I have done, I have taken maximum of these things over this. So, I am taking maximum or some real numbers and that is giving me this stuff right. Now, this if I now compute the support function or compute the support like compute the set whose support function is this, then the sub differential of f at x is the convex value of all elements of the form grad f j (x), where sorry grad f i x, where i again belongs to J (x). So, you take few elements and take a convex value.

So, this set del f x of a maximum function is a polyhedral set; polyhedral set which is quiet interesting in (( )). So, we will see that how this idea can be used to derive the optimality conditions for differentiable convex programming problem; we will show that a lot of any optimization problem, you take a convex optimization problem, minimize f (x) subject to g i (x) less than equal to 0. Now, I can always express this problem as a convex optimization problem with only one convex constraint; let me assume that f is differentiable and all the g i's are differentiable. So, this is see, convex and differentiable; I will not always write convex, because we are only talking about convexity, we are not talking about any other class of functions

And then you can write this problem is equivalent to minimize f (x), subject to g (x) less than equal to 0, where g is a single constraint, where g (x) is max of sorry max of g 1 (x) dot dot dot g m (x). So, you see any convex function with many constraints can be written as a convex function with a single constraints, but there is one lopsidedness of this whole issue that you lose differentiability of the constraint function, but that does not mean that we have lost everything, because our mathematical structure is very well done, because we already have a clearly done explicit formula, how to compute the sub differential of this class of functions; so this sort of functions.

(Refer Slide Time: 21:10)

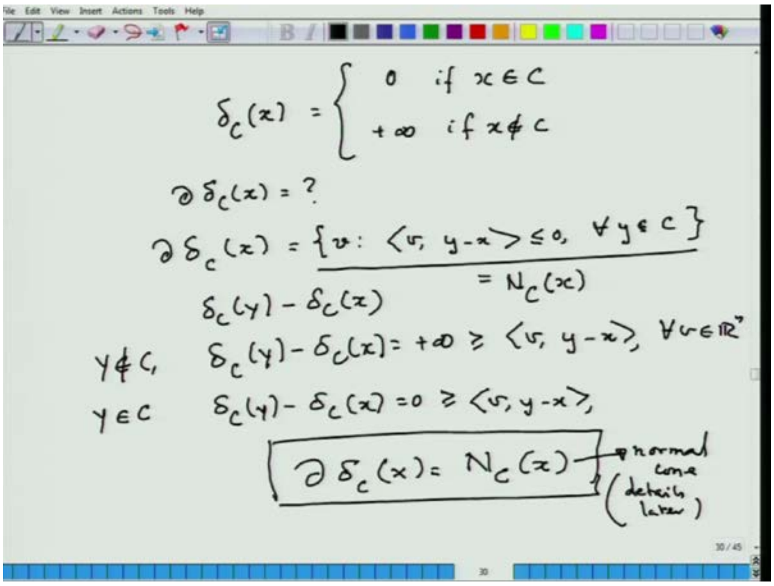


So, let me go ahead again; we had or now only been talking about convex functions from R n to R, but in the beginning we said that in convex analysis and convex optimization, it is natural to have convex functions from R n to R bar, where R bar is R union the two infinities and we wrote down some rules for how to handle the infinities. Now, can I if I have the convex function of this form; can I define a sub differential of this function; the answer comes out to be increasingly interesting and it is yes. So, take f from R n to R bar and assume f to be proper.

So, there is at least one point, where the function is finite. So, let x be a point, where f is finite; then del of f (x) is defined just in the same way; (No audio from 22:42 to 22:54) note that if there is a y, for which f i is plus infinity that that does not break this equation at all. So, this y has to be over all the R n, all y in R n. So, if, but x cannot take f (x) cannot take the value plus infinity, then it would become meaningless. So, f (y) can be plus infinity that does not harm the equation, (( )) and we will just given example of that.

Now, what happens if x is a point, where f is not finite, (No audio from 23:26 to 23:33) then del of f (x) is defined to be the empty set, but there is the what if question, it does not yet all mean that if del of f (x) that if f is that del f of f (x) is non-empty for every point x, where f x is finite; no, they are problems where at the point, where f is finite this set becomes empty. So, the story that we had for R n to R is now you see its getting complicated.

(Refer Slide Time: 24:28)



Now, let us do a little bit of example hunting for this class of functions; now a very important class of functions in convex optimization is the use of the indicator function of a convex set, which takes on the value 0, if x belongs to C and takes on the value plus infinity, if x does not belong to C, we have already mentioned this earlier. Now, this is a convex function from R n to R and it is a proper convex function, because it does not take the value minus infinity and over whole of C it take the value 0, which is the finite number the question is what is this is something very interesting. So, del of del C x here is a set of all v, such that v into y minus x is less than equal to 0, for all y element of i, sorry all y element of C.

Now, how would you prove this; this proof is quite simple, because if I write down del C y minus del C x, you see x has to be in C, because otherwise it is not finite. So, if x is in C, this is always 0, and here it is either 0 or infinity. So, suppose I y is such, y is not in C then del C Y minus del C x is nothing but plus infinity, which is greater than v times y minus x, for all v in R n. So, whatever v you take, this is anyway true; these are number and this is obviously, less than plus infinity.

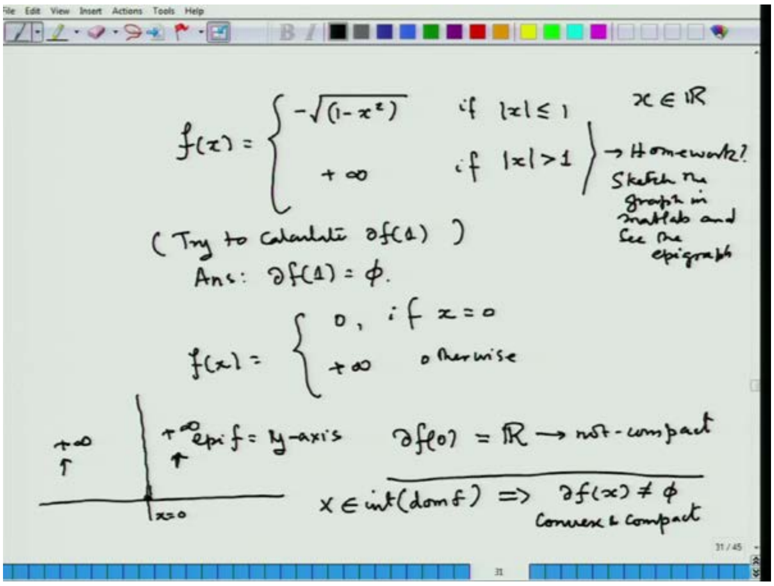
Now, if Y is sorry if Y is now in C, then del C Y minus del C x is equal to 0, and any v which has to be a sub gradient satisfies, has to satisfy this sorry for has to satisfy this sorry, this is sorry, because we have taken particular y. Now, so we have covered the whole all the y in R n. So, the common v, which satisfies both is the v, which satisfies this and exactly that is what we have written down there; now this class, this particular type of set has a particular name, and if you observe that very carefully this is the convex cone, and this has a particular name called the normal cone to the convex set C at x.

We will come to the normal cone business very, very soon, but before that let us we will do something else, we will try to write down some optimality conditions; now we will not bother much about this at this moment, the geometry etcetera will come later on, and which has which has (( )) geometry, which has lot of interesting things. The sub differential so what we have, the sub differential of the indicator function is the normal cone to C at x, this is the fundamental result in convex analysis and this result has led to the development of sub differential calculus for non-convex function, this is truly a fundamental result and this is called the normal cone, which we will talk later details later.

Now, you will again come and ask me a very important question; now you have defined sub differential for a class of functions, which is extended valued; now can you give an example where the function is finite at those points, but does not have a derivative at the

(( )), does not have a sub differential and that point; in the sense that the sub differential is empty at that point. So, there is the example that we will show.

(Refer Slide Time: 29:01)



Now let us look at this function f (x) is minus root over 1 minus x square, when x is line between minus 1 and plus 1. So, x is obviously, in R in this case; and it is plus infinity, if it is outside. Now this function is a convex function; this you can try to sketch the graph in matlab homework, you sketch the graph in matlab, and then check the epigraph and see the epigraph. Also you observe that if I take put x equal to 1 or x equal to minus 1 at that point the function of a value is 0, so it is finite; but it does not have at that point, non-empty sub differential; sub differential at that point is empty. So, your homework is to find try to calculate del f 1, the answer to this question is del f 1 is equal to phi.

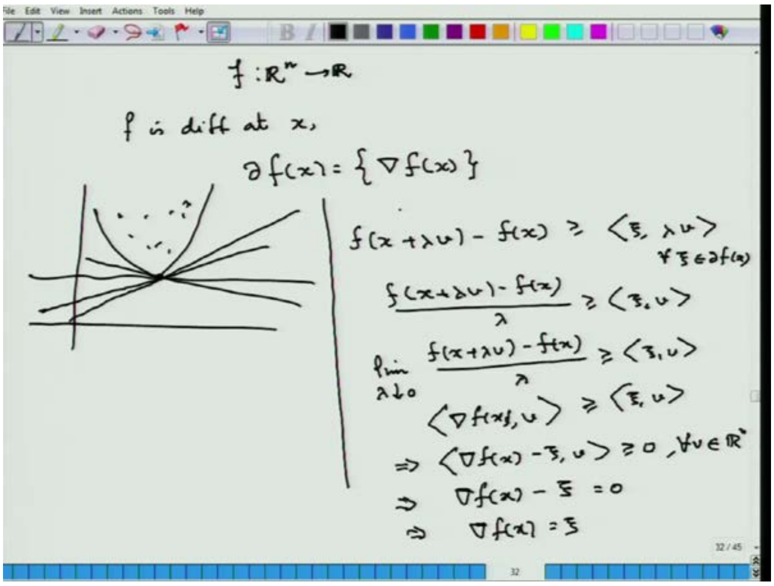
Now, there are lot of issues, which you come up when you take this extended valued function. So, they are very, very important issues are this now we have said that the sub differential for a function from R n to R is convex, compact and bla bla nice look, nice thing, but what happens when I have this sort of an extended valued function. Let me just take a very simple case; so f (x) is equal to 0, if x is 0 is equal to plus infinity otherwise; if you look at the graph of this function, so this is the graph of this function, x equal to 0; 0, 0.

So, the epigraph is nothing but the whole y axis. So, epi of such this function is the y axis; now here the function value is plus infinity, here the function value is plus infinity;

now if you look at the sub differential at 0, because that is the only point where the function is finite; del of f 0 of course, the function not differential, there is no question for differentiability for this extended value stuff. So, del f 0 here is R and R is of course, not compact. So, it is convex and closed.

So, a sub differential is always convex set and a closed set, but it need not be bounded always. So, this is the very important example, which shows that; when does a sub differential become closed convex and bounded, even when the function is extended value the answer is as follows. So, if x is in the interior of domain of f and you know what is the domain of f already is domain of f is the set of all x where the function is finite if you take the interior of the domain of f, then it would imply that x that the sub differential of f at that point is always non-empty and convex and compact; that is of course, the dom f has to have interior, we are expecting the dom f to be full dimensional; if it is not full dimensional nothing can be said. So, if here the dom f here was full dimensional, but they are interior they did not exits right.

(Refer Slide Time: 33:26)



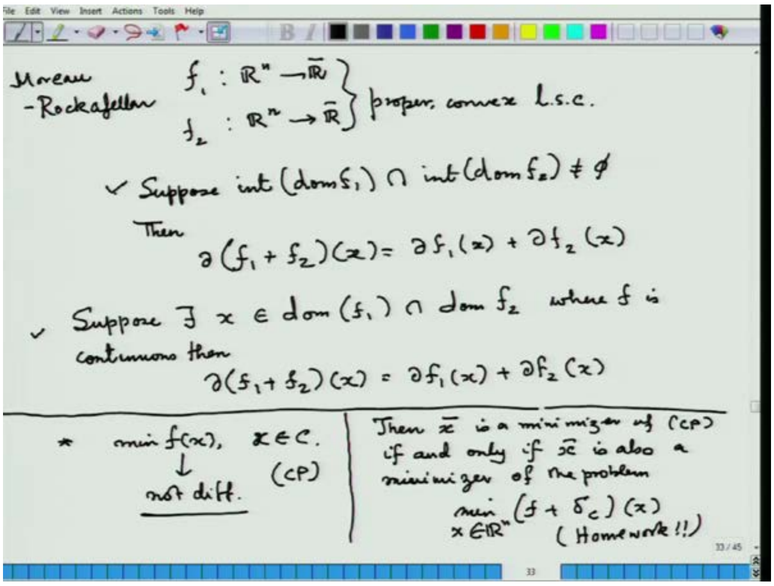
So, as whatever neighborhood you take, it is outside the set. So, del f (x) is non-empty convex and compact; you will now ask me a question; take the function from R n to R, code was simple case of course, all many optimization problems would be of this class. Now, if f is differentiable at x, do you have this fact that the sub differential contains only the gradient the answer turns out to be yes; the sub differential has only the gradient and that is why the sub differential idea is a true generalization; if one might ask why cannot you have some other idea to define something, which can imitate the derivative, but other ideas like using the notion of weak derivative from distributions, did not work well and it is the intrinsic property of the epigraph, which is the convex set came into been that we are that at the point of non-differentiability of a convex function the epigraph has a infinite number that supporting hyper plane and that is the that is the fundamental idea that builds into the making of the sub differential, that is if you have a convex function, which is not differentiable, see you have a king point here.

Then if you look at the epigraph, which is the convex set then at this point, there not on that there is not only one tangent hyper plane, there is many, many supporting hyper plane. So, this is the fundamental idea that is built in that the this is just a convex way of bringing the fact that slope of the tangent in the is the derivative, that is bought in through, through the convex language, which is the language of the supporting hyper planes and which brings us to a set rather than just a vector, which is a curious an important point.

Now, suppose I have a point, where the function is differentiable say suppose it is x, then let us see f x plus lambda v minus f (x) is psi times lambda v for all psi element of del f (x) unless writing the sub in inequality, if I do so, now bring out the lambda, so f of x plus lambda v minus f x by lambda. Now, the function is differentiable when use the Taylor's theorem or the differentiability definition to come up with the situation that if I now take the limit has lambda tends to 0 from the positive side, then what you will finally get is this; and this by differentiability you know, this is nothing but grad f x into v grad f x in a product v is (( )).

So, the directional derivative is nothing but the gradient; gradient into the direction of taking the derivative. So, directional derivative is gradient in the direction v. So, this would simply mean grad f (x) minus psi v is greater than equal to 0, but this v was arbitrary. So, it is for all v element of R n. So, now, what we are getting that I am having a linear function, which is non-negative everywhere it cannot happen. So, which it cannot happen unless grad f (x) minus psi is equal to 0, see you have a v. So, you could take psi minus grad f (x) as one those vs and put it here to get the norm of grad f (x) minus i whole square less than 0. So, grad f (x) would be psi. So, for whatever psi, you take grad f (x) would be psi and that is exactly what you want to show here. The question, which is of immediate important to us is the following.

(Refer Slide Time: 37:53)



Now if you have two functions proper f 1 and f 2; proper and convex; proper convex and we will add a little bit of thing, it is lower semi continuous - l.s.c; I hope all of you lower semi continuous means that the epigraph of these two functions are closed, the epigraph is the close set that is the meaning of lower semi lower semi continuity; it is an only if and if condition, it is something beyond continuity, we have already known that if our function is from R n to R, it is a convex function and it is always continuous; but if it is from R n to R bar, we lose continuity, because of this add of adding this infinities, but we have something more, which is lower semi continuity, because continuity cannot be… This lower semi continuity cannot (( )) can be characterized by epigraph.

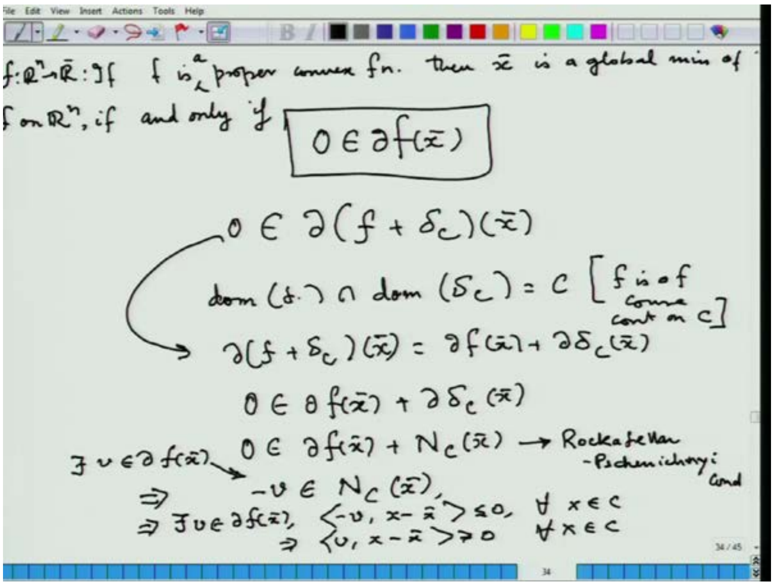
So, even if continuity is lost at the boundary of the domain, but still we can have lower semi continuity, see if I have this and if I have the following result, which is extremely important that I will consider that my domain of f 1 and domain of f 2 are full dimensional sets in the sense that they have interiors the nice sets; suppose interior of the domain of f 1, intersection interior of the domain of f 2 is not equal to phi, then you have this fabulous fascinating result; some rule that for a class of functions, which is extended value this is the very big advancement in convex analysis, this is called the Moreau Rockafellar theorem; both of them one of the greatest convex analyst and optimizes about time, he is a French, he is an American; Rockefeller is famous for his book convex analysis, he is mostly worked in mechanics Moreau john john jack John Jack Moreau from France and (( )) actually; in France and Rockefeller is was in Washington (( )) this is one thing.

Another is that you can do it like this; suppose, there exits an x in so, this is one one condition, under which this is true, I am giving an another condition, suppose there exits x in dom of f 1 intersection dom of f 2, where f 1 or f 2 does not matter, where f 1 is continuous, then del of f 1 plus f 2 x is equal to del of f 1 x plus del of f 2 x. So, this is again a fabulous result. So, now, what is the use of such a result; why we are so interested in this extended valued convex functions and what we want to do with it; we want to find out the optimality condition that we had been looking for so long; what have we been doing we have been looking into this aspect, we have been looking into this following fact that I have a convex function f (x) to be minimized and x is belonging to c; I want to find a necessary and sufficient optimality condition, when f is no longer differentiable.

Now, if that is so, how do I find an result we had already done something exactly that we have shown that if x bar is a global minimum of C p, then for each x this is true; of course, if this holds then x bar is also global minimum of C p. So, then what we can do is that this C p, then this holds and this result the converse is also true; the converse is also true. Now, what I want to say is that here there was a problem, my psi x was changing with every x, but psi x was belonging to this, but it was changing with every x, because of course, your using this compactness issue and all this things.

Now, what I want to assert is that I do not want the psi to change for every x. So, that is exactly where we need to look into our story; now what I will do is the following; if I call this program problem as CP, then x bar is a minimum minimizer rather minimizer of C p, if and only if, if and only if x bar is also a minimizer of the problem mean of f plus this, this is the problem over x x element of the whole R n. So, now we are talking about the extended valued function; you can prove this very easily this is homework this is too easy to prove you might ask what is the big deal of course, x has to be in C naturally and so, what is the very big deal this happens. So, how do I get an optimality condition, this is as follows.

(Refer Slide Time: 45:22)



If f is proper, if f is all proper convex function, then x bar is a global mean of course, x bar has to be a global mean means if f of x bar is finite, otherwise it has no meaning, x bar is a global mean, a global minimize, if and only if 0 belongs to del f x bar. So, this forma's rule remains relevant even when we are talking about proper extended valued convex function of course, here we are taking f from R n to R bar. So, this remains relevant even if this holds, when a even if this function is extended values this fundamental result remains relevant and remains true.

So, once I know this fact that if I solve a… If I have an x bar, which is solving this x bar is solving this unconstraint problem. So, this is result is for unconstraint problem say x bar is a global minimum of may be of f on R n, if and only if this holds. So, x bar in here is a global minimum of f plus the indicator function over whole of R n. So, what I will get; I will get 0 belonging to del of f plus now what is the domain of the function f, it is whole of R n. So, and what is the domain of the function del c, it is whole of c. So, domain of f 1 intersection domain sorry domain of f domain of f intersection domain of the function del C is nothing but c, but the function f, which is f 1; f 1 is taking, f is taking the place of f 1 here, this function f 1 sorry f 1 this function f 1 is continuous. So, this function f 1 is continuous over whole of R n. So, it is naturally continuous over c. So, this condition is satisfied for this particular case. So, now, I can write the some rule this would imply its 0, because this is equal to del of f at x plus del C at x sorry at x bar, because I have said x bar is the minimum right x bar is a minimum.

So, we will now have, so what I will have what I will have, because and f is of course, continuous over C continuous on c. So, this will lead us to the fact that del of f plus del C of x is del of f at x plus del of del C at x. So, what we will conclude from this fact is that 0 actually, belongs to del of x plus del of del C of x and 0 belongs to del of x plus this is what we have symbolically written as normal going to C at sorry at x bar x bar, x bar, x bar.

So, this is one of the famous necessary and sufficient optimality condition for a convex programming problem and it is called the Rockefeller Pschenicheryi condition; Russian Pschenicheryi condition; now observe what I have from here, this says that this implies that they are must exists of v in del f (x) such that minus of v must belong to this for some v element of for some sorry or they are exists or I should write nicely, their exits v in del f x bar, such that this happens. So, this translates to this, which is very simple because there must be a v in del f x bar and w here such that v plus w is 0. So, minus v is equal to w, which also belongs to normal cones. So, this is what we have.

So, what does this means; it means there exists v in del f x bar such that minus v into x minus x bar because that is the definition of a normal cone which is same as of course, the del of this and we have already seen, what is the formula. So, for all x in c, so now, we have improved upon our initial understanding, initial optimality condition, we have improved and showed that I can have a fixed v, which will work for all the xs. So, I have a I will have a fixed v such that v into x minus x bar is greater than equal to 0 for all x. So, the Rockefeller Pschenicheryi condition can be written now.

So, we have strengthened the optimality condition, which we have done in the last lecture. So, with this we stop our lecture today; and then in next tomorrows class, we go into some more relevant issues in convex optimization and convex analysis called the conjugate of a convex function; and that has deep links with optimality and the optima itself; and we should exploit this, because this conjugate convex function would help us in many, many ways as we will go along. So, for today let me tell you a good night and thank you very much.