Difference Convex Functions and Optimization

Prof. Juan Enrique Martinez - Leqaz
Department of Economic Theory
Indian Institute of Technology, Kanpur

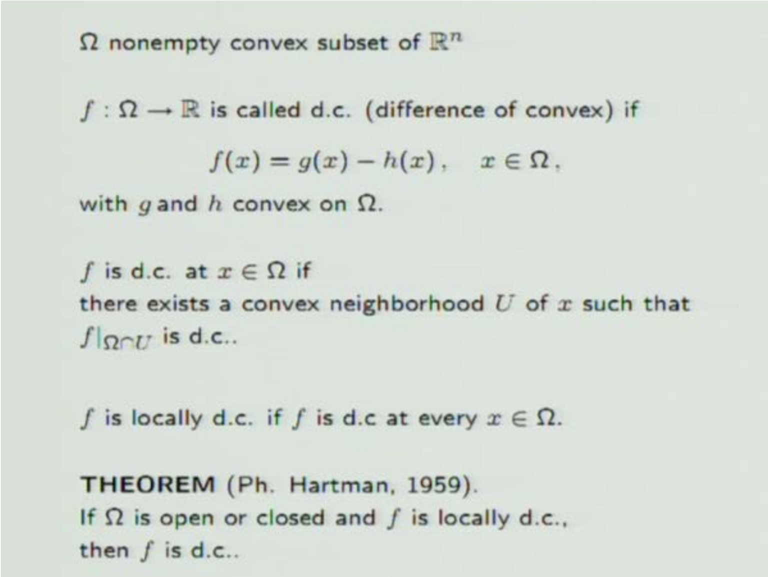
Lecture No. # 41
Special Lecture – 01

I am going to give a survey lecture on d c analysis, and its applications to optimization. D c analysis - what does d c? D c stands for difference of convex. So, d c functions are functions which can be expressed as a difference of two convex functions, and they are generally non convex, as we are going to see there are lot of non convex functions, which can be expressed as a difference of two convex functions. So, they use of d c functions, d c analysis is a non convex analysis; it means, looking at non convex problems with convex high; using convex tools in a non convexity setting.

This is a very typical attitude in mathematics differential calculus, what is it? You are linearising functions. So, you are applying linear methods in a non-linear setting look at non smooth analysis. What is it? You have non differentiable functions, but you are using sort of derivatives substitutes for derivatives. So, again you are looking at the non differentiable well with differentiablize. So, here we are looking at the non convex well with convex size.

So, you can see here the title is basically an outline of my lecture, I will first present some general results on d c functions, where elementary, then some sub differential analysis, some applications to duality in optimization. And finally, I will present our result related to lipschitz continuity of d c functions, and I will conclude by presenting the main references, I have used to prepare this lecture.

(Refer Slide Time: 02:20)



Well, in the first part of my talk we will be dealing with a functions defined on our end. In the last part I will consider the more general case, when the space is (( )) space, but here to simplify representation we just consider that, we have a function f defined on a non empty convex subset of R n and we call it d c, as I said before difference of convex, if it can be written on omega on the domain as a difference of two convex functions.

I have a small question I want to know that it is it is clear that the number I want to d c functions is quite abundant.

Yes we will see this later of course, you have taken omega as a sub set of R n.

YesWould be a convex naturally a convex subset of R n suppose I take f from R n to R n means both g and h are from R n to R n is that class of functions also abundant

yes yes yes I think. So, you are looking at vector functions, which components are (( )) since it’s since, for scalar functions we have this abundance. In fact, it consulates in to.

I am not talking about that if you are talking of some instead of omega if you have R n there means f is from R n to r.

Is then also you have lot of d c function is abundant.

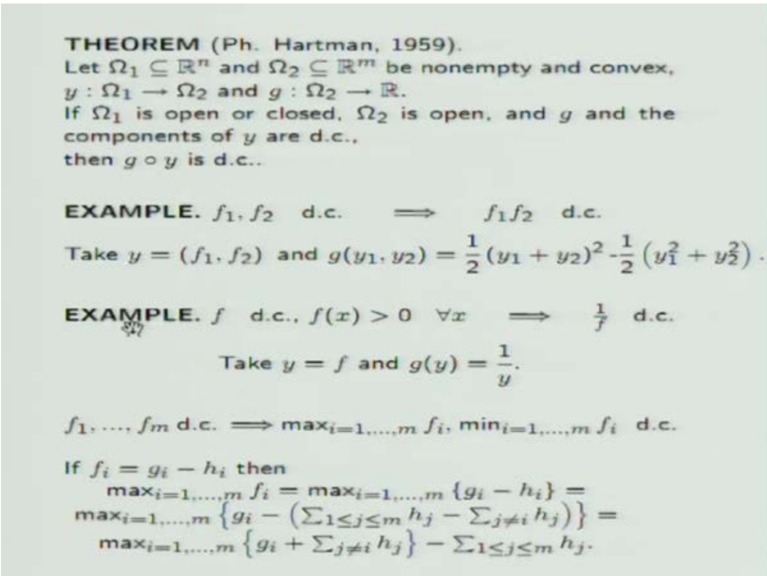
Yeah in the space of finite valued functions on R n.

Yeah it is a abundant.)

Yes, again from the same must be here in the general case, where my (( )) of non convex functions I mean non convex functions, which had finite value on omega in particular omega can be R n(( )) and then the abundance is relative to the larger set of functions you are considering well. This is the notion we are going to use all the time, but there is also a local motion we say that the function f is d c at a point in the domain. If it is this d c on some convex neighborhood of the point must specifically, on the intersection of some convex neighborhood of the point with a, with a domain and we say that the function is locally d c. If it is d c at every point which means, at every neighborhood there is a decomposition, but in principle there may not be a common decomposition for all the neighborhoods.

Nevertheless there are two important theorems by Hartman published in 1959, will give their précised reference at the end the first theorem says that in the case, when the domain of the function is either open or closed and convex of course, then there is no difference between locally d c and d c. Globally d c I mean whenever, you are sure that at every point there is a neighborhood with a d c decomposition, you can also be sure that there is a global decomposition although the proof is not constructive. So, it this result doesn't tell you how to obtain a global d c decomposition out of the local d c decompositions of the functions.

(Refer Slide Time: 05:46)



The same applies to the next, theorem also due to Hartman in the same paper which basically, says that when you compose d c functions you get another d c function. Here is the precise result look at victor(( )) function y which is d c component wise and look at another function g which is a scalar value and it is d c assuming, that the domain of the vector function is either open or closed and the domain of the scalar function is open then the composition is d c. Again this result is not constructive, it assures you that the composition is d c, but doesn't tell you how to get a decomposition out of the composition of the functions y and g.

By using this result you can prove that a number of functions are d c. For instance you have here two examples the product of two d c functions is d c. Here is the proof apply the theorem to the vector function y whose components are f 1 and f 2 and consider g be the function y one times, y two which is this d c as clearly shown by this decomposition then applying the theorem, you get that the product is d c the reciprocal of a positive function f is a d c (( )) by applying that theorem to the case. When y is the vector function which in this case, it is a scalar and coincides with f and g is the function one of our y which is convex for a y positive.

So, we see that the class of d c functions is rather rich its closed under well. Obviously, under the operation of addition, subtraction, multiplication by real numbers and now, we have just seen is closed under multiplication also the reciprocal of our d c function is d c and we see here that the maximum of our finite collection of d c functions is a d c and they they showed an easy proof. You can see here is in this case constructive is very c if our (( )) function f i you have a decomposition as a difference of two convex functions g i and j i then, here you have at the end a decomposition of the maximum of those functions as a difference of two convex functions.

This function and that function are convex because what we are doing is just adding convex functions, which preserve convexity and they can maximum of a collection of functions, which preserves convexity to. So, here we know how to obtain a decomposition of the maximum out of decompositions of the individual functions that entering to this maximum.

(Refer Slide Time: 09:10)

Similarly,

$\min_{i=1,\ldots,m} f_i = \sum_{1\le j\le m} g_j - \max_{i=1,\ldots,m}\left\{h_i + \sum_{j\ne i} g_j\right\}.$

In particular, if $f = g - h$ then

$f^+ = \max\{g, h\} - h$ and $f^- = \max\{g, h\} - g.$

φ_g, φ_h affine, $g \ge \varphi_g$, $h \ge \varphi_h$

$$f = ((\max\{g, h\} - \varphi_h) + (g - \varphi_g)) - ((\max\{g, h\} - \varphi_g) + (h - \varphi_h))$$

$f_1 = g_1 - h_1$, $f_2 = g_2 - h_2$, with g_1, g_2, h_1, h_2 convex and nonnegative. Then

$$f_1 f_2 = g_1 g_2 - g_1 h_2 - h_1 g_2 + h_1 h_2.$$

$g_1 g_2 = \frac{1}{2}(g_1 + g_2)^2 - \frac{1}{2}\left(g_1^2 + g_2^2\right)$

EXAMPLE. $f(x, y) = \frac{x}{y}$, with $y > 0$

$\frac{x}{y} = \frac{1}{2}\left(\left(x^+ + \frac{1}{y}\right)^2 + \left(x^-\right)^2\right) - \frac{1}{2}\left(\left(x^- + \frac{1}{y}\right)^2 + \left(x^+\right)^2\right)$

One can obviously do the same for the point wise minimum of a finite collection of functions. I stress the fact that is point wise maximum or point wise minimum of a finite collection because the result fails to be true(( )) not only the proof, but the result is you would say; obviously, wrong if we consider an infinite collection of functions supermom of an infinite collection of d c functions may not be d c. Well in particular, we have specific decomposition for the positive part and the negative part of a d c function.

Positive part is the maximum and negative part is the maximum of minus function and 0 because we are taking, the maximum of two functions then apply in their precedent result, we get this decompositions for the positive part and for the negative part and of course, the functions appearing in d c composition are convex because we are taking the maximum of two convex functions.

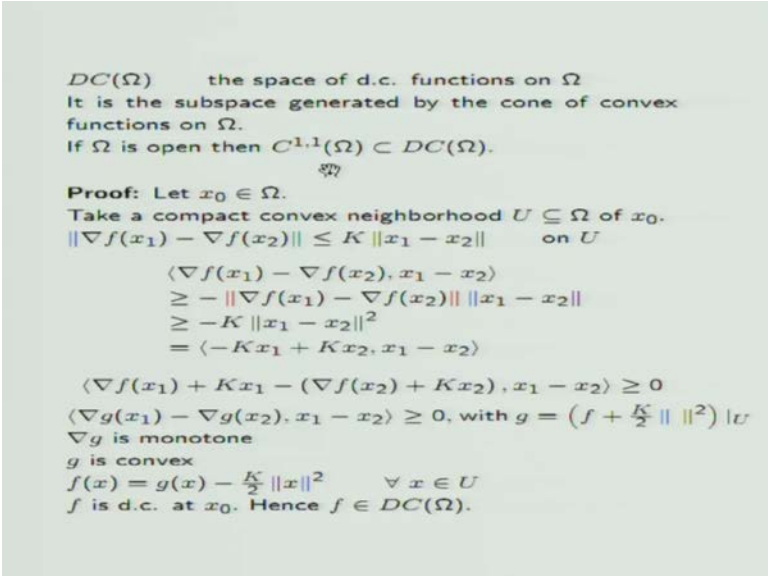
Well to be very soon, in some cases its very useful to have a decomposition which is not only a difference of two convex functions, but a difference of two convex functions which are also non negative and it is shown here, that you can always achieve that I mean, if you get me two convex functions. I can find two convex functions, which are moreover non negative and whose difference is the same as the original function. So, to this just you simply need to consider an affine minorum for each of the two functions it always succeeds. If a function is proper for a proper convex functions you all (( )) and affine minorum.

And then just rewrite g minus h in this way, it is clear that after simplification this is equivalent to g minus h that is to (( )) with, but all the functions in this expression which are written between parenthesis are convex because. What are we doing? We are taking the maximum of two convex functions and the only subtractions, which you can see here the function which is subtracting is affine and subtracting an affine function does not destroy a convexity. Why is this decomposition with the extra condition of non negativity useful? Because using such decompositions in some cases as in the one, we will see here you can construct explicit decompositions for some operations with functions, when instance is product.

Suppose, you have here two functions f 1 and f 2 which are the c and the decompositions, which is for them f decides the property of non negativity. Then we can expand the product in this way and in this algebraic sum, every term is the product of two non negative convex functions since, the n non negative its product for instance the first ten (( )) can be expressed in this way as a difference of two convex functions explicitly, but it's essential that g 1 and g 2 are non negative because otherwise raising to the square would destroy convexity.

So, you have here a situation in which a (( )) having decomposition with not only convex, but also non negative functions is a useful. Here you have n is a, here you have an example of application of the preceding a technique the function of two variables x y, x over y with y with y positive this is the product of two functions. X is 1 of the factors and the other is 1 over y both are convex, but x is not non negative. So, we have to decompose it as a difference of two non negative functions and a using this decomposition. The technique I have show here then after some obvious calculations, we end that with this expression for x over y as a difference of two convex functions. As I said before, the space of d c functions is rather large. It contains the class of c 2 functions even more class of all.

(Refer Slide Time: 14:08)



So, called c 1 1 functions c 1 1, stands for functions which are locally lipschtz and help a functions, which are differentiable and help a locally lipschitz gradient in particular c 2 functions are c 1 1. So, the (( )) space of c 1 1 function is contained in the space of defect c functions on omega.

Yeah actually is c 2 functions there is a proof for the c 1 1 functions right.

Yes every comment on the proof now.

I think I saw order c 2 functions I learnt it from (( )) they said that c 2 functions are always d c.

Yes

But this is this is from the omega right?

For, the for the class of c 2 functions. In fact, the proof is simpler than, even simpler than what I have here the proof is as follows take a compact neighborhood of the, of a point at a given point you take a convex neighborhood, thank you very much. Then on that neighborhood the secondary (( )) is bounded from below.

Yeah, compact

By a, by a positive number then.

Yeah.

You come use this to express, the function as a difference of a convex function minus a multiple of the square of (( )) and then you have that the function is locally d c and by Hartman theorem it is locally d c. So, this is a very simple proof for c 2 functions, but for c 1 1 functions, the proof is not difficult and it's here given a point in the domain you come take a compact convex neighborhood of that point. Then you have the Lipchitz condition you have a common constant k which works for all points in this neighborhood and then look at this inequalities (( )) one is yes (( )) the second is the Lipchitz condition and then we end up with this an equality, which can written like this.

That what we have here is the gradient of g this function f plus k halves, this square of (()) on U then this inequality. What is telling us is that the gradient of g is monotone, but this is equivalent to saying that the function is convex. So, this function f which is g minus k halves the square of (( )) on U is d c because both g and the square of norm are d c. So, we have the d c condition at the point is 0 and then a because of Harman theorem the function is globally d c.

No I have one comment to make.

Yes.

I am just trying to figure out that kappa that k that you have used that k is positive of course, because that

Yes.

You know this is a interesting thing and this f is a class of functions called weakly convex functions introduced by

Yes.

J b (( )).

I know.

In 1983.

So, what we are proving that every c 1 1 function is actually weakly convex.

Locally.

Locally yes.

Locally and then Harman theorem doesn't tell you that if you have this weak convexity condition locally

Yeah

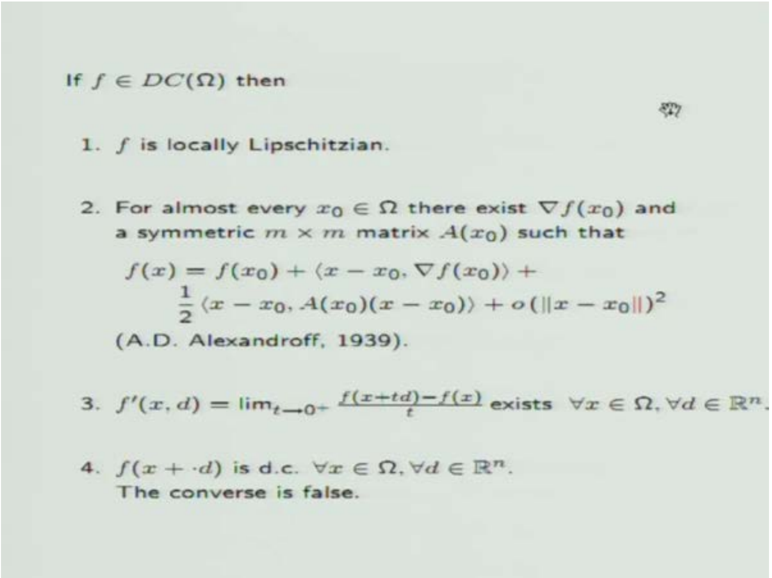
You get it globally.

that is the interesting point yeah because that is the interesting point I want to know yeah.

This is an important difference

So, we have here the proof also we have here the statement that the space of d c functions is the smallest, space of functions which contains all convex functions. So, it contains all convex functions all concave functions and is the smallest subspace which contains those functions.

(Refer Slide Time: 18:01)



Then It is clear that this set of d c functions on omega is very large, but not as large as I could, we could imagine every d c function must (( )) several properties, which are listed here first of all they must be locally Lipchitz because every convex function is locally Lipchitz and when you take differences this locally Lipchitz character is preserved.

Second there is a no (( )) by Alexandroff which says that every convex function at every path almost every point, in the (( )) of (( )) is a second order Taylor expansion. We know by (( )) theorem that convex functions are being locally Lipchitz are differentiable almost everywhere, but the Reci (()) by Alexandroff says more there is not only a gradient everywhere almost everywhere, but also almost everywhere there is a matrix which plays the role of the Hessian matrix.

9 But it is.

This matrix.

But it is not with hessian.

If it may not be hessian in particular because the function may even fail to be differentiable on our neighborhood. So, even if the gradient does not exceeds in on that particular neighborhood. You can be sure that there is a matrix which would play the role of the Hessian matrix.

I would like to (( )) slightly note this is in rockefellers book or (( )), but I would like to be I understand it slightly more if I take a can I take a minute.

Of course,

Yeah for almost every x naught there exist a there exist grad of f x naught which is of course, true where this (( )) theorem for convex functions.

Yeah what will?

And symmetric m cross m matrix such that this holds, but I cannot prove that this matrix because if I define second order differentiation of a function.

If the Hessian exists, it must coincide with this matrix.

Yeah that.

But the Hessian needs not restrict even more the function need not be differentiable around the point.

Around the point x naught.

Need not be or it may be, but not be twice differentiable. So, in this (( )) there is no Hessian matrix, but there is one matrix.

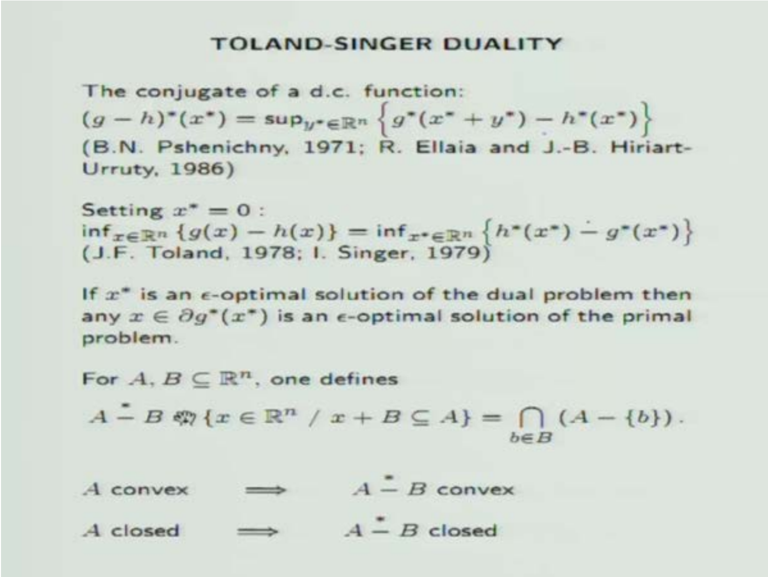
If it is twice differentiable then it is hessian. So, you are just not right right. So, then it is a powerful result right.

Yeah yeah. So, this statement was for convex function, but of course, it consulates two differences is because I mean, one can easily see that this almost every (( )) property is also satisfied in that case well then, we have a (( )) property that since, convex functions have one sided directional derivatives. They require the same must happen for every d c functions, we are just subtracting two convex functions and the last result on this page says that every d c function is d c on every straight line on every segment, let's say contain in the domain of the function, but the converse is wrong there is a counter example showing that the function.

The converse is doing only the convex case the converse is doing.

The converse is doing the convex case. So, for convex is if and only if, but for d c no there is some example showing that a function, may have a older restrictions to straight lines d c without doing d c by the way, the result type presented before about the inclusion of the c 1 1 set, in to the d c set is doing finite dimensions that proved that proof does not apply, in the infinite dimensional case and there is a counter example also showing that the result doesn't go through in then in the infinite dimensional case, these are the general properties about d c functions.

(Refer Slide Time: 22:22)



Now, I move to the next topic which is duality, first for unconstraint optimization problems and then we will consider a constraint optimization. I start by presenting the formula which was obtained by Pshenichny long time ago forty one years ago and were discovered by Elleia and Hiriart urruty sometime later. In fact, it is very easy to prove this formula, one can leave a very simple algebraic proof of this formula for the conjugate of a difference of two functions.

In terms of the conjugate of the functions which are involved in this difference g and h look at this dot below the minus (( )) what it means, is that in case that you subtract plus infinity minus plus infinity by definition you take, take the difference to be equal to plus to minus infinity sorry, on the other hand when the dot is above the sign which means, the opposite convention that plus infinity minus plus infinity is equal to plus infinity.

Why why have you put x star equal to 0 where are you putting this.

I will come to this in a few seconds, first I am comments on the on the first formula this is an expression for the conjugate of a d c function, but I will tell you to prove it I said there is a very simple purely algebraic proof. You only need h to be convex g can be an arbitrary function, but for h you have to assume that it is a convex proper and lower semi continuous with this assumption, the proof is an exercise a very easy exercise. Now, in this particular formula now, I come to your question take x star to be equal to 0 then you have the conjugate of our d c function at 0 and the conjugate of a functional 0 is minus the minimum of the function.

So, for x star is equal to 0 this left hand side here reduces to minus, this infimum then from your (( )) x star equal to 0 in this expression you get this infimum up to a minus sign. So, this result which is known as to land singer duality because they obtained the result independently essentially at the same time, but we more complicities methods this result is an immediate consequence of the formula for the conjugate of our d c function.

In fact, we are over the time considering finite valued functions, but we consider also extended real valued functions. In this case the formula still goes to if you put the dot over the minus sign, here as it is on the right hand side this is a duality result yes.

x star equal to 0 is put on the right hand side of the formula on the first line.

Yes

Then I would have (( )) y star element of R n g of y star minus x star of 0

Yeah, this with from minus sign.

Which comes that (( )) supermom with.

no. So, it should why shouldn't it have x star of 0 because you are putting x star is equal to 0.

Now, I have made that change of variables you would obtain here wait a moment, well thank you very much for noticing this because this formula is not correct x star should be y star.

Exactly it should be y star.

X star should be y star. Thank you, for pointing this out x star is y star yeah, otherwise this term would be would not be depending on y star no this that, the this is an important correction the real formula is with y star.

Then the why is the y star is dummy. So, you will.

Exactly. So, thank you for pointing out this, this type this is a duality result very well known in in non convex optimization, but it's not is a non standard duality result because in convex optimization who are used to (( )) which infimum of the primal programming is equal to the supremum of the dual problem minimization, maximization here both are minimization problems. So, the use is very different as in classical convex duality, but nevertheless solving the dual it helps solving the primal and moreover as we will see now and moreover this is an evolution in the sense that the dual of the dual is the primal.

So, look for computing the dual you simply take conjugates and exchange reverse the order. Then if you use the separation to the dual then you get the primal because g star star is g and h star, star is h. If you have (( )) and both g, h and h are proper convex and lower semi continuous well as I said before. So, then that we have problem faced helps to solve the primal problem because if, you look at an approximate optimal solution of the dual problem epsilon optimal solution x star you just need to take a sub gradient of the conjugate of g star at this epsilon optimal point to get an epsilon optimal solution of the primal problem.

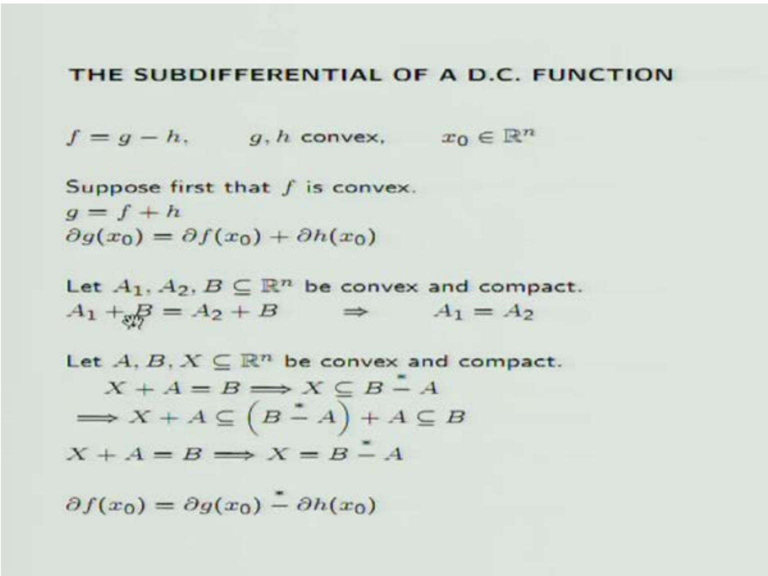
We have a formula for the conjugate in a, we also want to obtain a formula for the sub differential for this g function. One combination are there first we are taking the conjugate of a function which is not necessarily convex, we can the definition is applicable to any function no matter that its convex though their properties are not. So, nice if the function is not convex same with the sub differential. The (( )) of sub differential does not need the function to be convex, but there are some problems is your, if your function is not convex first problem is that this sub differential may be empty at many, many points.

And the second problem is that looking at the different definition of the (( )) sub differential, the you say that it's a global definition. You you need to know the function everywhere, but when the function is convex the sub differential is closely related to the direction of the (( )) for the computation of which you only need local information. So, for convex function the definition is just local this is now, the case in general for non convex functions. So, if you consider the sub differential of a d c function take in to account that it may be empty at many points, first thing and second that you need full information of the function not just local to to compute it. Well to give a formula for the sub differential of a d c function one needs to consider, this operation star difference is called sometimes of two subsets of R n, it was in previous (( )) in that same (( )) which is mentioned here. So, it's a set of points such that if you translate the seconds at B following this vector x then that runs late is a subset of A. So, the set of all those translations is this star difference equivalently.

We can easily see that the star difference is this intersection, this star difference may be empty very frequently because for the non emptiness, we need the assistance of A translate of B which is a subset of A and very often such a translate will not exist or from another point of view, we are making a key A here intersection with a large collection of functions of sets one set for each element in B. So, we are intersecting. So, many sets that we may have an empty set, but nevertheless we don't care this difference may be empty, but we easily see looking at this expression is that this star difference is (( )) properties. Every property of A which is preserved and their translations and intersections will be held by the star difference because we are just doing translations and intersections.

So, if A is convex this this star difference will be convex no matter how regular or irregular B is the same with closeness. If A is closed then the star difference with whatever B will be closed or if A is bounded the star difference with B will be bounded no matter how.

(Refer Slide Time: 31:49)



Now, we are in a position to study the sub differential of our d c function, but first let us make a conjuncture and let us see if it works, but the expression for the sub differential of our d c function should be to get some inspiration. Let us consider first the easy case when the difference of the two convex functions turns out to be convex. We find the formula for the difference in that case and then, we will see if this formula extends to the general case. So, assuming that f is convex then, we have this relation between three convex functions g, f and h and because of the (( )) of this sub differential we have this equality.

But we are assuming that we know, the sub differential of g and the sub differential of h and with this data, we want to compute the sub differential of f in other words from this equality we from this equality, we want to deduce the value of this term we want to solve this equation with this unknown this is an equation of this type with three sets x is the unknown a and b are data.

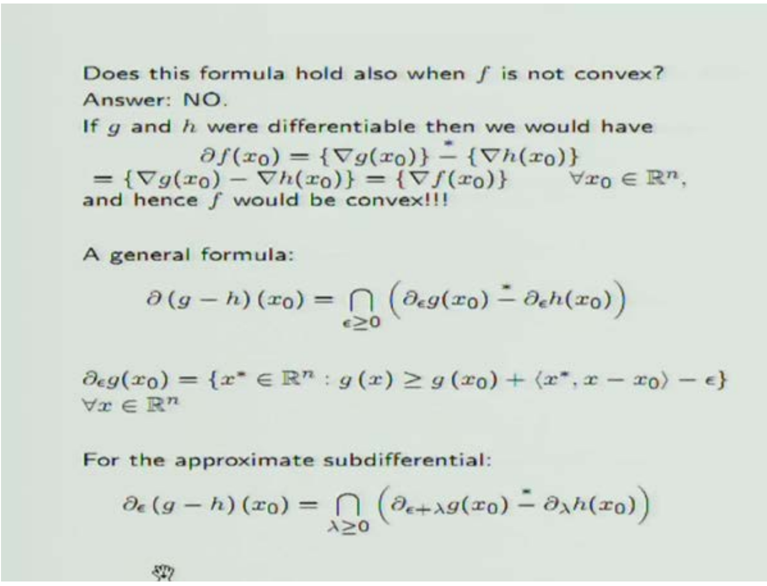
So, how can we solve this equation, we are dealing with convex convex compact sets sub differentials are convex and compact? So, first of all we have to recall the. So, called consolation property for for a closed convex sets. Convex compact sets if you look at this equality a one plus b equal to a two plus b then we can consulate b and deduce that a one is equal to a 2. One can prove this consolation (( )) very easily using dissipation theorem even more directly taking support functions because the support functions of this sub is the sum of support functions.

And since, our sets are compact this support functions are finite value to every (( )) then you can consul out just using, the standard arithmetic and then you end up with the equality of the support functions of A 1 and A 2, which is equivalent to the equivalence between the sets if the sets are convex and closed. Now, once we know this property, consider the equation we want to show for a the very definition of the star difference, we get that if x is a solution we are assuming that our solution exists.

So, if we have a solution for this equation, this solution X must be a subset of the star difference. Now, from this inclusion we get this one simply by adding A to both sides X plus A is contained in this set plus A which is a subset of A because of the definition of the star difference, but we are assuming that X is a solution of the equation. So, X plus a must be equal to b all the inclusions here are actually, equalities in particular (( ))first one this equality, then we use the consolation property to consul a and then we conclude that the solution of (( )) equation is the difference, but the star difference mean we proceed like in standard arithmetic's moving a to the right hand side with different, but with this special difference.

So, in this particular case assuming that f is convex, we have that the sub differential of the difference is the star difference of the sub differential, but convexity was essential in this proof at this step we need convexity to use the additivity of this sub differential. Now, the question is does this formula hold in the general d c case of course, the proof doesn't hold, but may the formula holds true in the more general case.

(Refer Slide Time: 35:52)

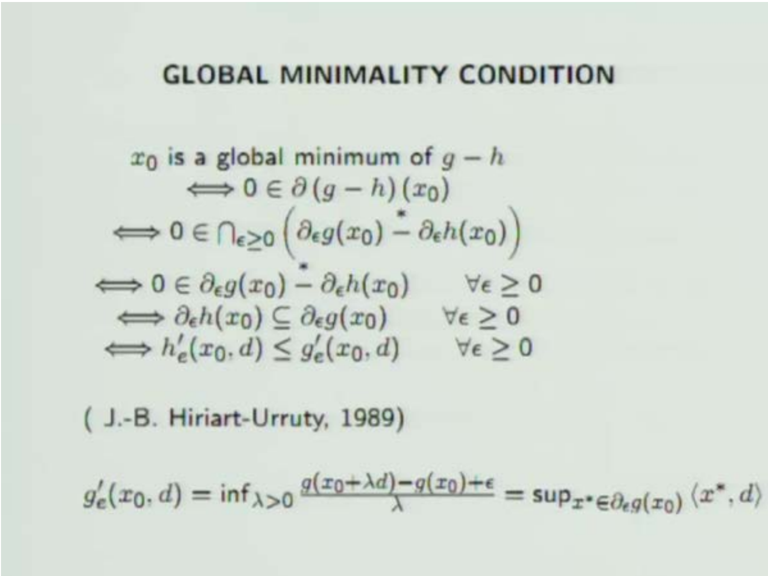


But the answer is no consider for example, the case when the two functions are differentiable in this case the sub differential reduced to the gradient. If that formula were through then we would have that this sub differential of f is the star difference of this single terms. But for single terms there is no difference between the star difference and the ordinary difference they coincide. So, we would have the single term of the difference of the variants, which is the gradient of the difference, but in particular it would be non empty everywhere and non emptiness of the sub differential everywhere means, the function is convex. So, the conclusion would be if that formula would hold true in the general case, every function which is the difference of two differentiable convex functions would be convex and this is; obviously, true; obviously, not true.

There is also another argument, may be when more elementary look, this expression again in the right hand side we are dealing with convex functions. So, to compute this sub differentials we only need local information, but on the left hand side if the function is d c we cannot do it we just local information well. So, the true formula is this one which is more complicated, you need to consider epsilon sub differential. So, the definition for the epsilon sub differential is here and its worth pointing out that epsilon sub differentials, even for convex functions require global information with just local knowledge we cannot compute them.

So, you have to take a star difference of the epsilon sub differentials and then the intersection of well all positive or non negative epsilon. And also this formula is a more general version for epsilon sub differentials it basically, says if here you write the epsilon prime sub differential, you just have epsilon prime here and that it this is for this formula this again.

(Refer Slide Time: 38:07)



Using that formula for the epsilon sub differential, we can immediately deduce the global optimality condition for a d c functions which was obtained by Hiriart urruty in nineteen eighty nine. Although since, the formula for the sub differential of a difference was not known at that time, the derivation of optimality condition is not the simple in their in the initial paper by Hiriart urruty.

The derivation using the formula is as simple as you can see here x 0 is a global minimum of the of the difference. If 0 is a sub gradient this is just an algebraic fact follows immediately from the definition of the sub differential. Now, the next step is using the formula we have just obtained 0 belongs to the intersection means, 0 belongs to every set in the family and 0 belongs to a star difference means, the second set is contained in the first set or equivalently we here in equality between the support functions. The support function of the epsilon sub differential which is. So, called epsilon directional derivative.

Which can be also expressed by means, of this infimum like in the case of convex functions. I mean like in the case of the exact directional derivative, epsilon equal to 0. If we take here x equal to 0 then, we get an expression directional derivative because for convex function this question without epsilon is a increasing with lambda. So, the limit is equal to the infimum but.

I have a comment here.

Wait a second please, but here we have infimum, but it's not the limit. So, again this is not a local notion even if it is called derivative, we need global information to compute it, it provides us with equivalent information as the epsilon sub differential, which we have seen is not a local notion, even for global for for globally convex functions you please.

You referred this about j b Hiriart Urruty but.

Yes

As far as I have seen are you sure you is that this notion has already, it was in in Rockefellers 1970 book.

Which notion?

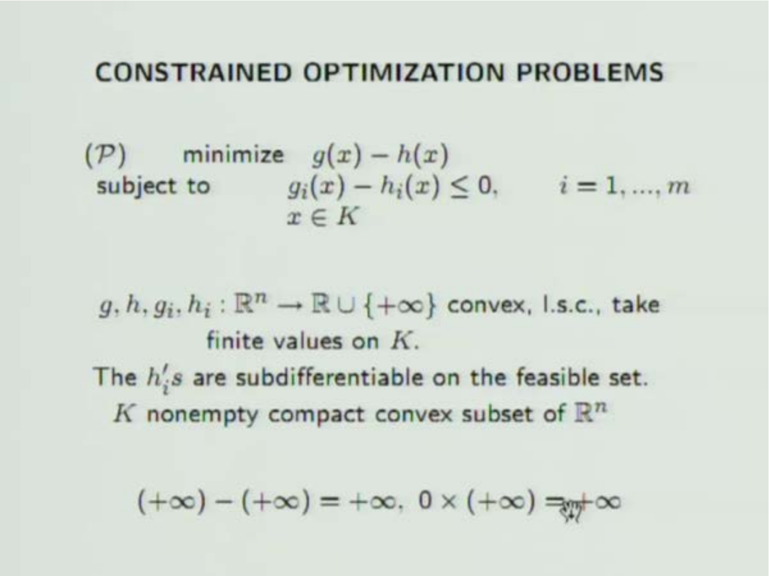
This g dash epsilon x naught d.

No, no I am referring to Hiriart Urruty for the optimality conditions.

Oh optimality conditions.

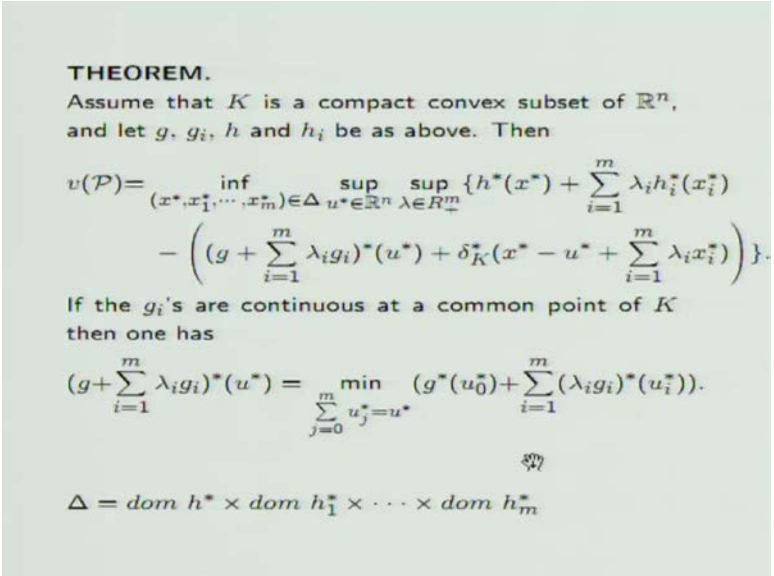
So, this result that x 0 is a global minimum of the difference if and only if the epsilon sub differential of x is contained in the epsilon sub differential of g for every epsilon, this is due to Hiriart Urruty and this this situation here is for the first part not not for the second part. This equality here is well known for epsilon equal to 0, the directional derivative is the support function of the sub differential, here we have an expansion well

(Refer Slide Time: 41:15)



Now, let us consider duality for a d c optimization problems with constraints. We are assuming here that both the objective function and the constraints are d c. But to avoid very technical discussion on a constraint qualifications, I am going to avoid them by a considering an extra constraint x belong to k. If it is k is a non empty compact convex set thanks to it, we can avoid completely constraint qualifications here you have the precise functions on the functions, we are considering in the conjunction which will be used in the formulas we are going to see.

(Refer Slide Time: 42:04)

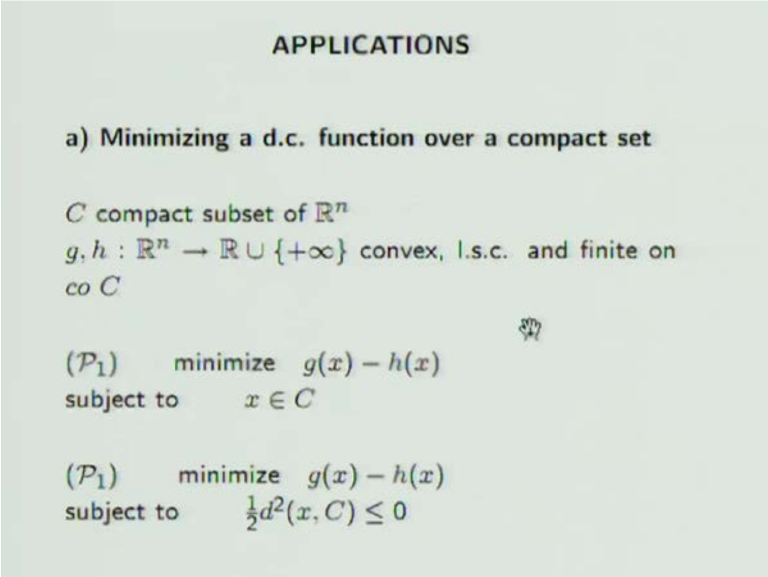


Then we have this complicated expression which is the duality result, we have (( )) here the optimal value of the primal problem and here you have other problem, which is expressed in terms of the conjugates of the delta.

Well here is not just the conjugate of delta, but conjugate of linear combination, but you seen an extra assumption for instance, that all this functions are continuous at a common point of the domain, then we can replace with that which is really in terms of the conjugates of the original functions. This is a very complicated formula and. In fact, the real problem is mixture of minimization and maximization, you have an (( )) expression the original problem was minimization.

But the interest of this problem is that it unifies, the most classical non convex duality theorem with the most classical convex duality theorem namely, Toland Singer for the non convex case and Lagrange, Lagrangian duality for the convex case, if there are no constraints it's very sad to say that this formula reduces to Toland singer duality. If there are constraints, but all the functions involved are are convex which means, that x can be h sub I's are all equal to 0. Then this formula reduces to the standard Lagrangian dual problem.

(Refer Slide Time: 43:41)



Just two theoretical applications of this general formula, first you consider the minimization of a d c function over a compact set, compact set which need not be convex. So, this is our our problem yeah.

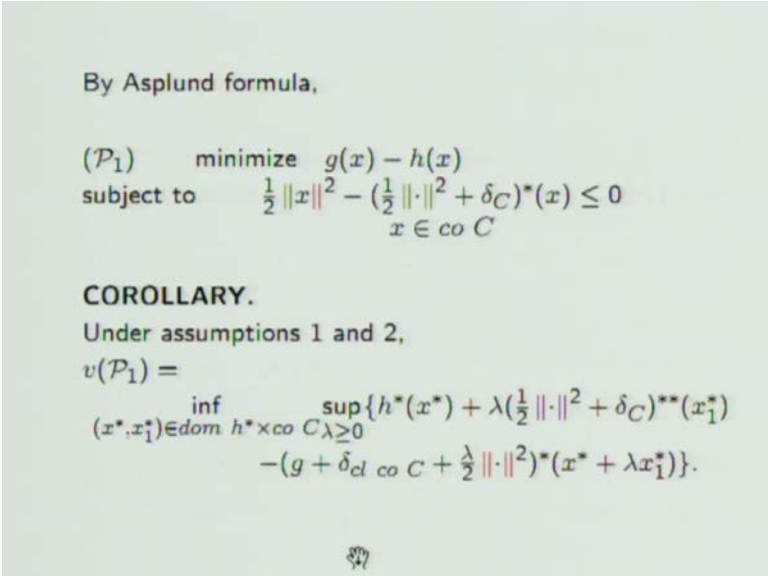
The set c in p 1 is not convex anymore it need not be.

The set c is compacting not convex.

Just compact ok

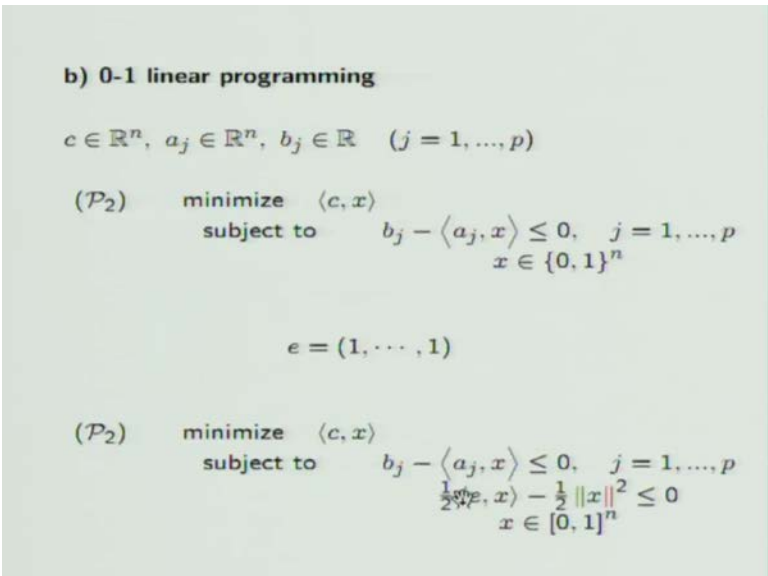
So, we are considering a very very general problem. In fact, it can be seen that the minimization of an arbitrary lower semi continuous function over a compact set can be reformulated in this way.

(Refer Slide Time: 44:32)



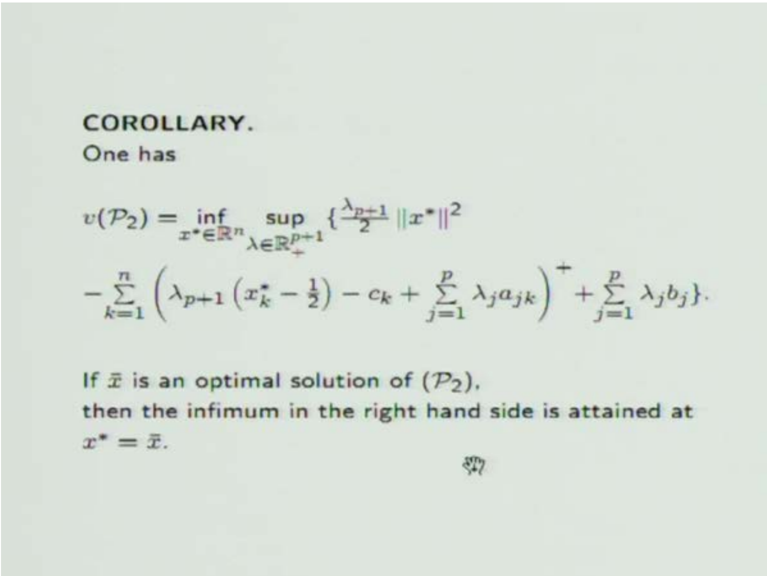
But then we can replace this condition x belongs to c by this inequality, which is the c because by the well known formula Asplund one half the distance to the set can be expressed as this difference of convex functions then. Now, to give a (( )) in which there was the extra condition, that x belongs to a compact convex set we have to consider a here this extra condition x belongs to the convexly of c, c was compact. So, the convexly of c is now, compact and convex. Now, we just apply the formula. We have seen before and we have these expression, which doesn't look as complicated as the one for the general case for this very general problem of minimizing a d c function over compact set.

(Refer Slide Time: 45:20)



Second theoretical application is the problem of a linear programming with binary variables 0, 1 variables, this is an important program in application roughly speaking every (( )) optimization problem can be reformulated under this problem. We are minimizing a linear function subset to in linear inequalities, but all variables must be 0, 1. Then this problem can be reformulated in this way, one can see that if instead of saying that its variable is either 0 or one we consider that its well our belongs to the close interval 0 and 1, but then we are the (( )) condition which is a d c inequality linear minus convex. So, actually concave then these two constraints are equivalent to this one and then we have the problem reformulated with our format.

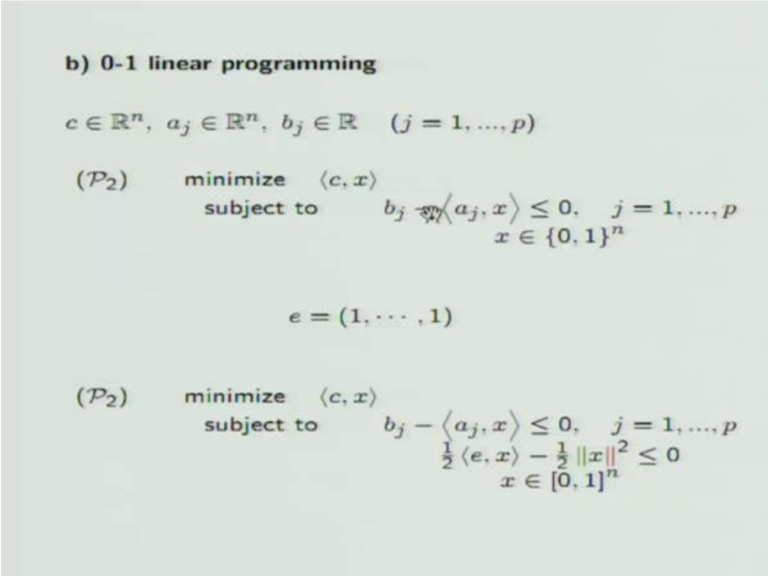
(Refer Slide Time: 46:21)



Then we can immediately apply, the general formula and obtain this exact duality result for a linear programming problems with 0,1 variables of course the right hand side is non convex, one cannot expect miracles the 0,1 linear problem is non convex. So, it is a non convex dual, but nevertheless the expression here look that complicated and moreover if you succeed to solve the problem, then you can really give a solution of the primal problem because if we have an optimal solution of our problem.

Then this is one among all the optimal solutions of the minimization part of the dual problem. So, you succeed to solve the dual problem. You obtain a collection of optimal dual solutions and then you know that, this set of optimal dual solution contains at least one solution of the primal problem. In particular if the solution set of dual problem is a single term that single term, is the single term of the optimal solution of the primal problem yes.

(Refer Slide Time: 47:38)



Can I go back to the previous slide now this very interesting actually to me these are things I am? So, interested in now the p 2 that you are forming here p 2 by adding this constraint by which you are replacing.

Geometrically its very simple.

Half, but this is e right e x.

E is the vector of once a once.

Then that constraint would.

That is only true if.

In two variables.

0 and 1.

In two variables, this is the unit square this is the feasible set the vertices when, what we are doing is replacing this set of four points with the intersection of this set with this circle with the compliment of this circle and the intersection is the set of vertices.

You you can also prove this algebraically in that very very.

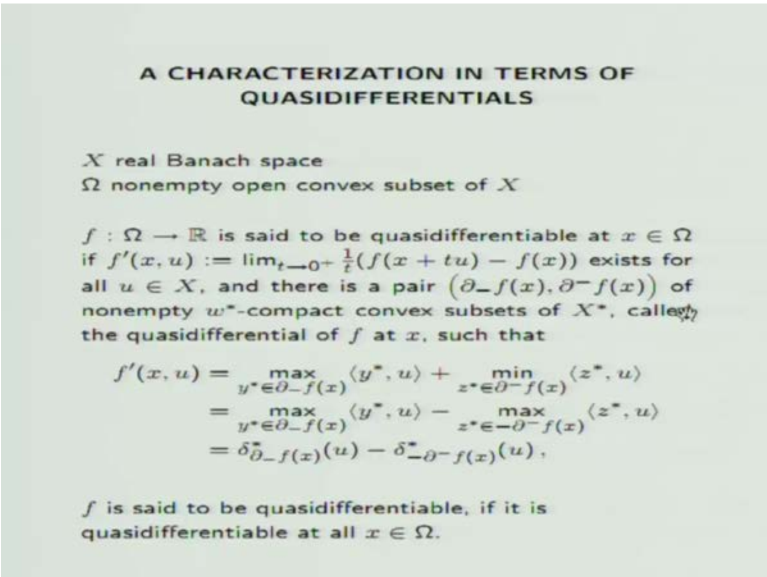
No no I one at the second that constant is true only if if and only if x is in 0 and n.

Yes

That is clear.

This together with the fact that the variables are between 0 and 1 well. Now, about to a more theoretical part which is a characterizations of d c functions, there are two ways of looking at d c functions. First as (( )) before d c functions are locally lipschitz. So, its natural, natural to look for characterizations of d c functions, within the set of locally lipschitz functions and the proof for this would be the (( )), but another point of view is we also know that this functions are directionally differentiable. Then how to characterize this function in the larger set of directional differentiable functions and to answer this question the most reasonable tool is the quasidifferential define long time ago by denviano (( )).

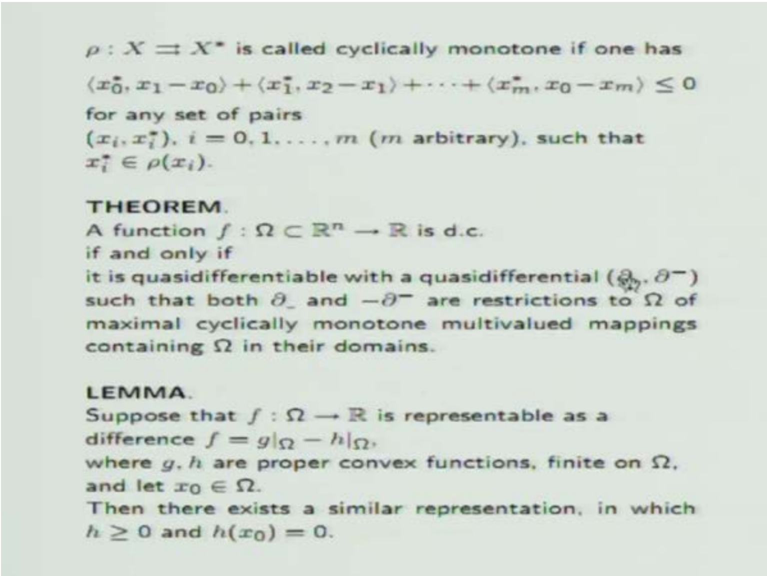
(Refer Slide Time: 49:53)



I will present some results in the setting of (( )) spaces and I will mention the differences with finite dimensional case, where first of all I recall here the definition of quasi-differentiability which basically, says that the one sided the rational derivative, which is always positively homogeneous can be expressed a sub difference of two convex and positively homogeneous that is to say sub linear functions. Then those functions has to (( )) support of compact convex sets in this case with weak star topology.

And then these two sets of which the support functions, appear in the expression for the derivative are called the pair formed by this two sets is called the quasidifferential of the function at the point and function is said to be quasi differentiable if it is, so at every point in the domain. The first set in the quasi differential would be called a sub, the sub differential and the second set the super differential.

(Refer Slide Time: 51:03)

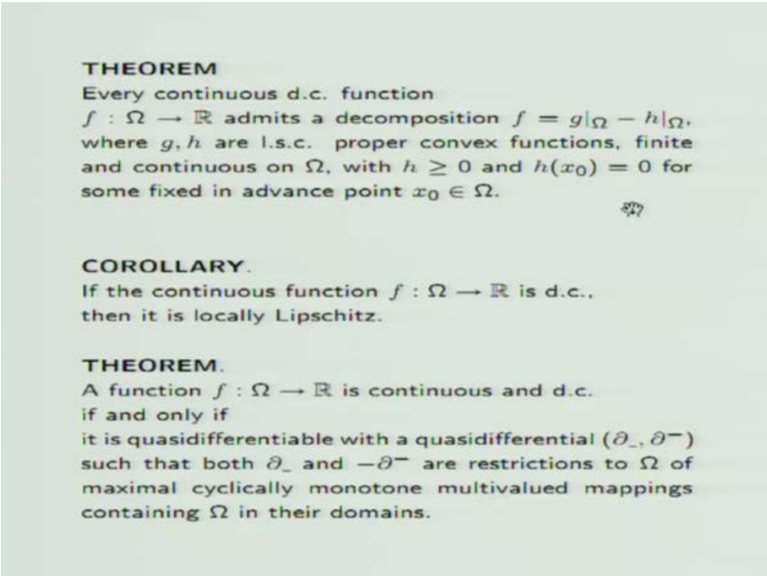


Now, the set of quasi differentiable functions is very large it contains all d c functions its quite clear, but it also contains all differentiable functions.

Then it contains many functions which are not d c, so how to characterize this functions in terms of the quasidifferential to obtain this characterization we need. So, we recall the notion of cyclic monotonicity, which is (( )) in convex analysis because it used to characterize sub differentials, then in the finite dimensional case. So, for functions defined on subset of our n we have this characterization a function is d c if and only if, it is quasi differentiable everywhere and the quasidifferential is such that both the sub differential and the super differential are restrictions to the domain of the function of maximal cyclically, monotone multi valued functions that contain omega in the domains this is if and only if. So, it is a characterization.

But its true only in the finite dimensional case, to see what happen in the general (( )) space case when need first this lemma, which can be proved easily using hammanac theorem, which says that if we have a difference of two proper convex functions then you can get a similar representations which is. So, called normalized, normalized means that the second function is no negative and vanishes at (( )) specified point then, using this theorem we connecting this important result, which says that whenever you have a d c function which is continuous, then in the d c decomposition you can create the two times continuous (( )).

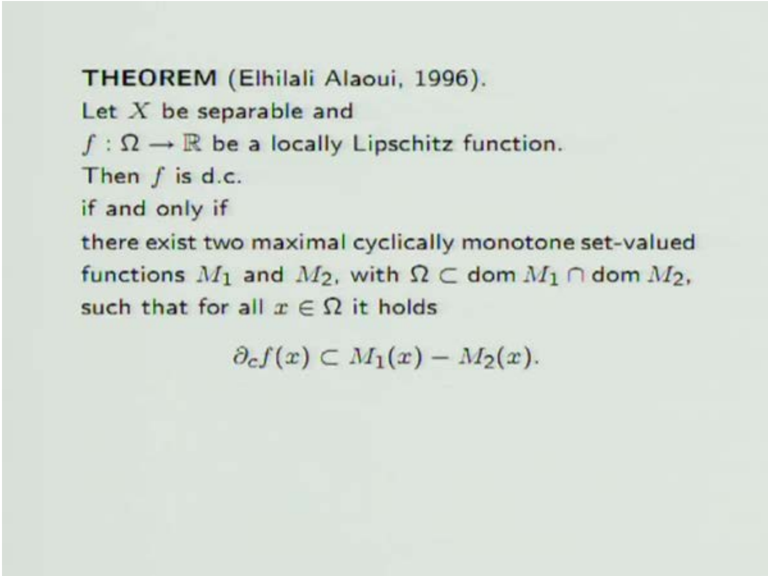
(Refer Slide Time: 53:13)



As a consequence continuous d c functions are locally lipschitz because continuous convex functions are locally lipschitz and then we have here the infinite dimensional (( )) of the result we showed below, which says the same the same, but with the difference that you need here f to be continuous not only d c, but continuous and a under this extra assumption, you have exactly the same as before, but you need continuity yes.

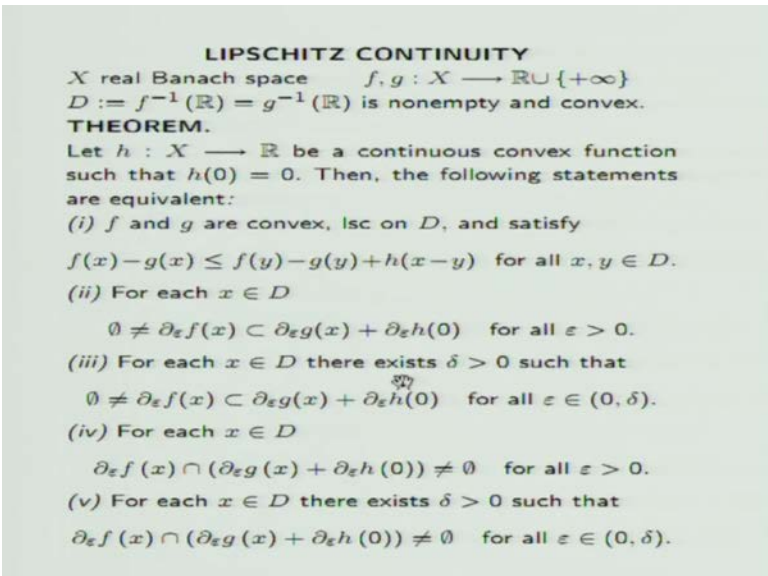
I am just reading them.

(Refer Slide Time: 53:49)



Yeah is the simply (( )) we showed before, the only different is continuous here with our continuity we cannot do anything and this is to be compared with much older result by obtained by Elhilali Alaoui long time ago, in which d c functions are characterized in the set of locally lipschitz functions. Using the Clarke generalized gradient one needs, the space to be separable and the characterization is this one f is d c. If and only if the Clarke gradient is contained in the difference algebraic difference mean (( )) difference of two maximize cyclically monotone mappings, which contain omega determinant of function in their domains.

(Refer Slide Time: 54:26)

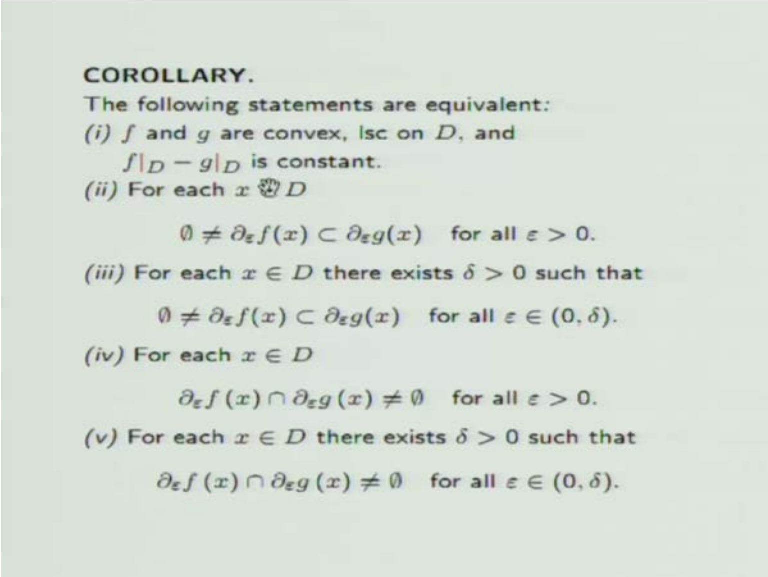


Now, the last part of my talk will be about lipschitz continuity. Everything will follow from I need to write a result, which I find very interesting not by the result itself, but because of the proof which fortunately I have no time to tell you now, but the proof the difficult part of the proof, you see are very interesting technique invented by a young Czech student (( )) well this result says that all this five statements are equivalent.

The only (( )) assumption is that the two functions, we are dealing with are define the (( )) domain which is non empty and convex no other assumptions. Then we have that using the next function h which is assumed to be a continuous and convex and vanishing at the origin, then the two functions are convex and lower semi continuous on their domain and satisfy this inequality if and only if each of this and then all of them of this results hold, well all this results or this statements are expressed in terms of the approximate sub differentials, the epsilon sub differentials of the involved functions.

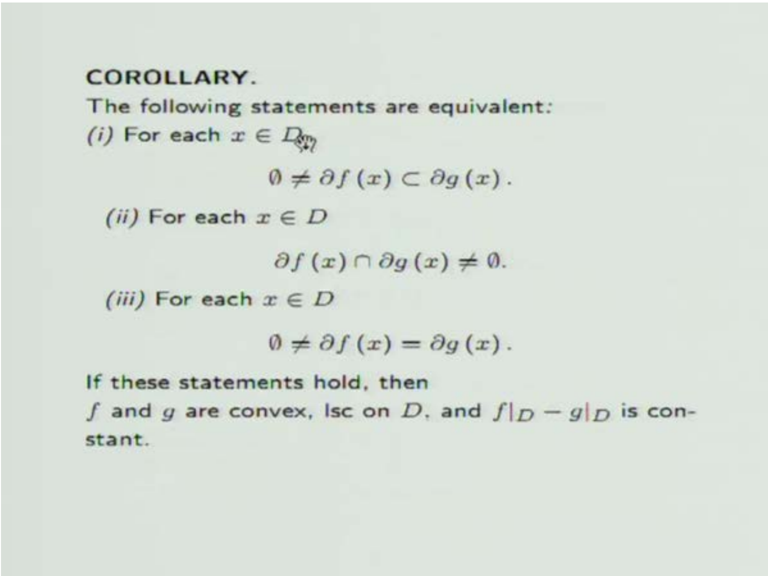
And most of the implications are easy one can make a circular proof, some implications of (( )) two implies three is (( )) because two and three are the same statement one for every epsilon the second only for sufficiently more epsilon also three implies five is obvious because this intersection would be the smaller set, which is assumed to be non empty in the same way two implies four is obvious and four implies five is obvious. So, to complete the proof we only need to imply one implies two, which is a quite you see convex analytic proof and five implies one which is where the technique by (( )) that the job.

(Refer Slide Time: 57:01)



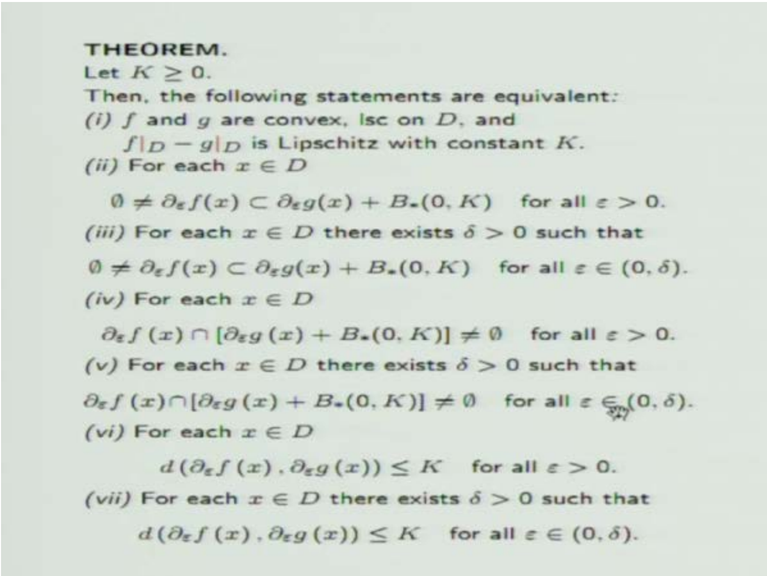
H is any convex continuous function vanishing at 0. If we take this function to be identically equal to 0 then we get this equivalences, the two functions are convex and lower semi continuous and coincide up to an additive constant, if and only if we have two or we have three or we have four or we have five the interesting implication, here is five implies one because just assuming that the epsilon sub differentials always intersect for sufficiently more epsilon. We conclude with some (( )) result first of all that the functions are convex and lower semi continuous and then that they are the same at two analytic constant.

(Refer Slide Time: 57:34)



Also if you want a result with just sub differentials not approximate sub differentials, then here is the result. Again this is an integration result because assuming one of this conditions, we get that the functions are convex lower semi continuous and their difference is constant here, the interesting implication is two implies three because just assuming that the intersection of this sub differential is non empty everywhere, you get equality or you get the fact that the two functions are identical up to analytive constant, this follows immediately from the first result. We are talking about the Lipchitz condition.

(Refer Slide Time: 58:13)



The Lipchitz condition appears, when we take h equal to k times below A is the Lipchitz constant, then we have all this equivalences this follows immediately from the general theorem by just computing the epsilon sub differential of the norm at the origin, which turns out to be the (( )) in the dual space, well not the (( )) the (( )) with various k and center of the origin. So, this is just up to the statement five and immediate application of the general theorem, but we have here two more statements which appear to be weaker than the previous ones.

If you look at five for instant five says that, there is a point here and a point there such that the difference is the is in (( )) that is to say such that, the difference is a normal less than or equal to k then the distance between the two sets is less than or equal to k, but this is apparently weaker because the distance need not be attained nevertheless, it can ((

)) to be equivalent. So, it’s interesting to notice that you just need the distance between the epsilon sub differentials to be less than or equal to k for sufficient small epsilon to be sure that your d c function f minus j is Lipchitz with constant k.

Another, interesting observation is that some of the statements here are symmetric for instance, one if minus g is Lipchitz if and only if g minus f is Lipchitz, the same with many others, but not this is not the case with a statement two and a statement three, but because they are equivalent then we get the symmetry which means, we can interchange f and g here and here everywhere.