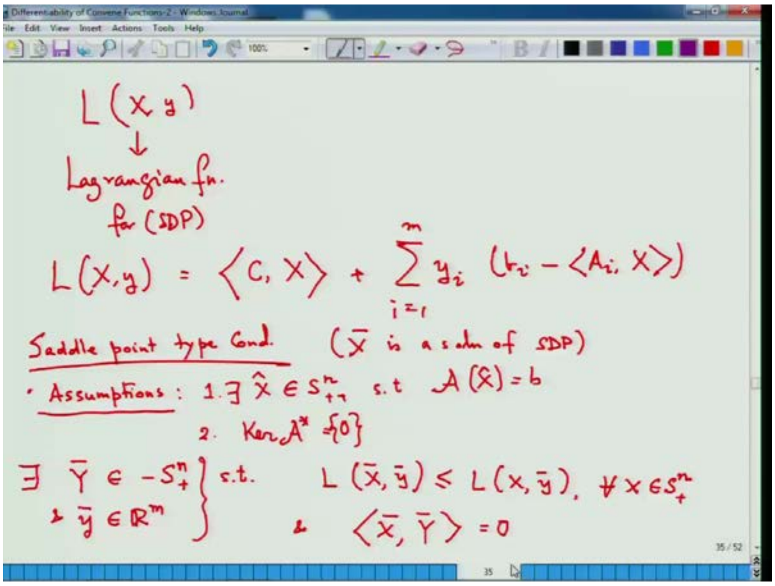
Convex Optimization

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

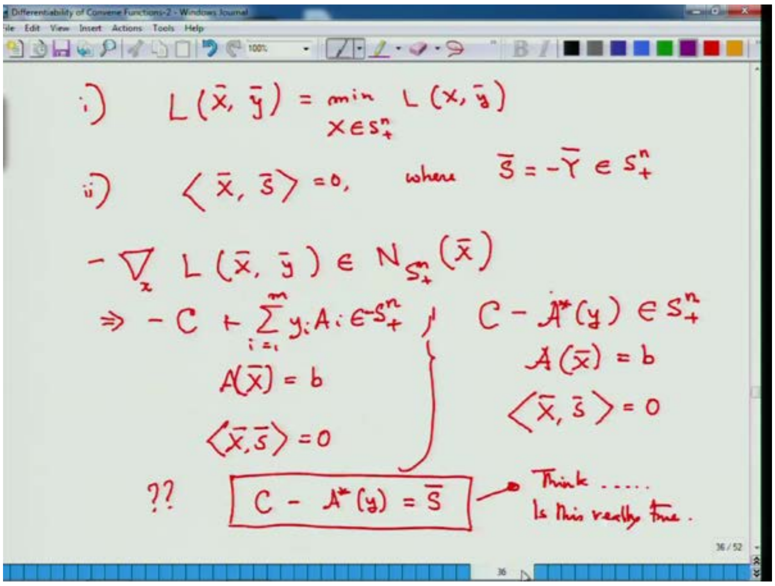
Lecture No. # 35

(Refer Slide Time: 00:31)



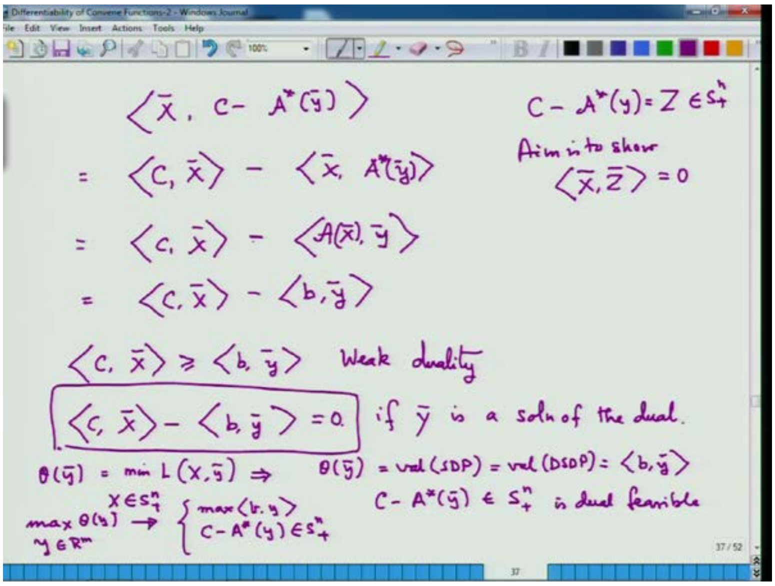
Let us recall what we had said in the last class. We had actually constructed the Lagrangian of the semi definite programming problem SDP. And then, we prove that if the following to assumption one in to that you see on the screen, if these two are satisfy that (( )) S n plus plus, and kernel of A star is equal to 0. Then if you take these two conditions; then what would you have is the following, that they would be y bar element of minus S n plus, so S bar element of S n plus as such that this thing hold; it is like a saddle point condition that you know in standard convex optimization, but we are looking at the whole thing from the perspective of SDP as a problem and trying to handle it.

(Refer Slide Time: 00:59)



Now, what you have here is the following; then this would tell me that this is happening, and this what you get where I am setting S bar equal to this. Now, then for this you apply a standard the optimality condition for convex optimization, and that would lead to this conditions. Now, is it the same S bar that you will give this difference, that is the whole question or this as S bar something else, the thing is this S bar that you see here is an extra condition, that we get when we studies SDP. It is not that something which is associated with this.

(Refer Slide Time: 01:47)



So, what we can show is that, if you look at X bar and C minus A star y, then you would really see one thing that this is nothing C X bar minus X bar into A star y, and by the definition of my adjoint, this is nothing but A X bar; this this a matrices. So, A A X bar y. So X bar is matrix here. Here what you see is a multiplication between 2 matrices; that has been converting into multiplication between vectors. In a product between vectors; multiplication in higher dimension is in a product. So, this is nothing but C of X bar minus b of y; now, if you observe what we have now this, because strong duality holds, because in X bar is the solution. So C of X bar is minus b of y bar is greater than equal to 0. So, you will always for from the duality point of view, you always have this week duality result.

(No audio from 03:26 to 03:37)

Now, what happens is the following. This is always true, this is what you get. Now, C X bar is the primal one, and b y bar is the dual one right. Now, what I have to show is that this product is 0, so if I right C minus A star y as some Z element of S n plus, basically my aim is to show - X bar Z bar is equal to 0. Now, what we have got - we have we know that the strong duality holds, that is C X bar the solution; so these greater than 0, so C X bar minus b y would be equal to 0, if y is a solution of the dual right. Because y is the solution of the dual strong duality will give me this equal to 0, which we have already have an idea of.

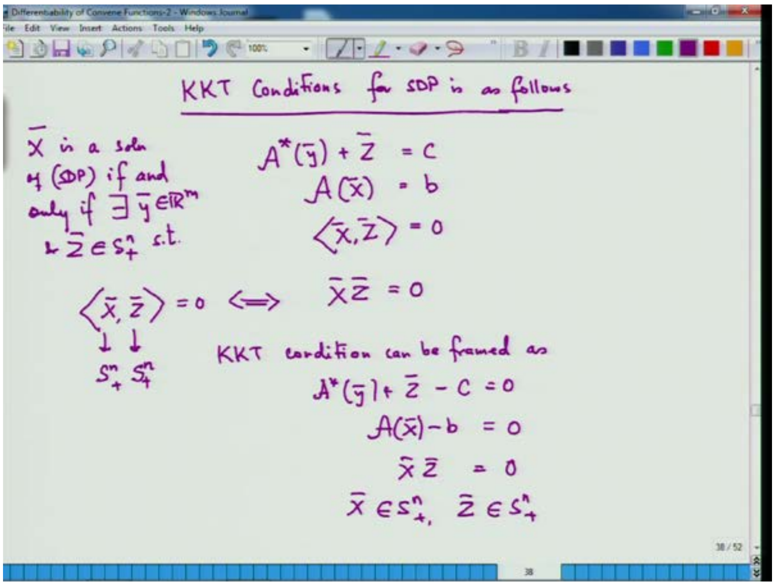
So I miss the bar sign here on y should be y bar here, and this is just I am sure, you can figure out this mistake. Now, if I calculate theta y bar which is equal to mean of L x y bar, where x is element of S n plus, then it implies that this is nothing that theta y bar is nothing but the value of SDP. And that is same as value of the dual of SDP which is from what is theta y bar - theta y bar value of dual of SDP, this is the value which will be nothing but b y bar that will be the value; the theta y bar is value of SDP value of DSDP.

So, if the value of the dual problem would be given by b y bar. So in fact, you can prove that y bar is actually the value of SDP, which if you look at theta y bar what is a the dual problem - it is maximize theta y bar over y bar element R n right. Basically is the dual problem in our case, basically this is same as writing maximize sorry, not theta y y. C minus A star y is a element a S n plus, y is a element of R n . So theta y, so the minimum of this is equal to b y bar, if this is true, but this is already true. So, I already know that C minus A star y bar from our previous results, this result.

This is already element of S n plus plus, and theta y bar which corresponds to the objective value of the dual variable which is same as b y bar, that is the theta y bar is the maximum dual objective value. The dual objective value theta y bar is actually nothing is same as has a same value as b y bar. So that is why b y bar is equal to theta y bar, and that is equal to 0. So that is sorry that is why C X bar minus b y bar this part is equal to 0. What we have done we have shown, that y bar is dual feasible and theta y bar is actually the value of the dual. And what is theta y bar - theta y bar is b y bar is this holds true. So, if this is holding true; so theta y bar must be y bar.

So, if these result wholes true this fact wholes true, and theta y bar is b y bar. And we know already that this holds true, so theta y bar is v y bar, and we have proved very simply from the where basic fact here, by using this fact theta y bar again is nothing but the value of SDP which by duality, because all the conditions are duality satisfy it is value of DSDP and this is this; and so, this is equal to 0. So what we know get get is that if I put this equal to Z bar X bar Z bar is equal to 0.

(Refer Slide Time: 08:39)



So, my KKT conditions.

(No audio from 08:39 to 08:46)

For SDP is as follows. So I will have that if X bar is a solution of SDP, X bar is the solution of SDP; if and only if is a convex problem, so it is an if and only if, you know this very well; KKT conditions of both necessary, and sufficient. X bar is sufficient an SDP if and only if, there exists y bar element of R m, and Z bar element of S n plus such that. So such that A star y bar plus Z is equal to C, A of X is equal to b, and X, Z is equal to 0. But, there is a very interesting result which says that if X, and S, X and Z you take any matrices X and Z.

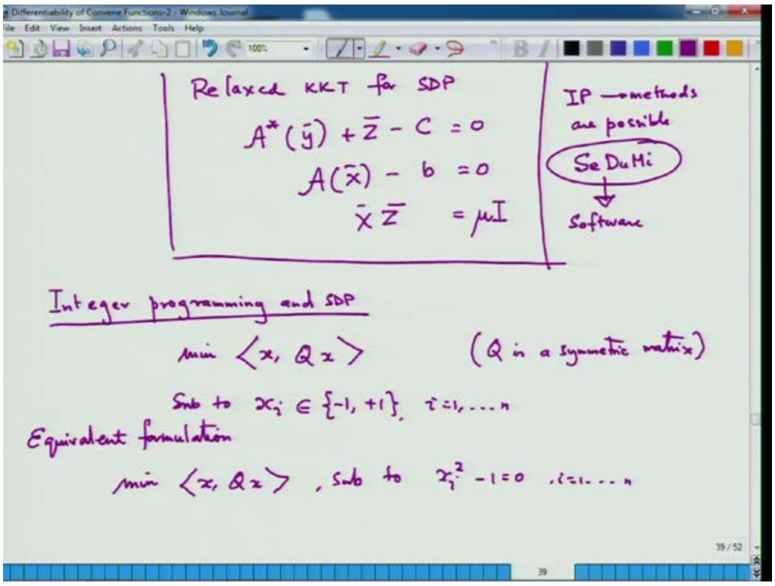
(No audio from 10:11 to 10:20)

So this is a standard result. Since, this is in a S n plus, and this is in a S n plus both are in a S n plus, then what would happen is the following; it would be like this. That this equal to 0, if and only if X bar Z bar this product is equal to 0; so, your KKT condition can now be framed for a SDP can be framed as.

(No audio from 10:59 to 11:10)

The KKT condition can now be framed as, that if X bar is solution of SDP, if and if this x such that A star y bar plus Z bar minus C is equal to 0, A X bar minus b is equal to 0, and X bar Z bar is equal to 0 - where X bar is element of S n plus, and Z bar is element of S n plus.

(Refer Slide Time: 11:54)



So, if anyone to develop the interior point method, the relax KKT condition that you use is the following is a takes a thing of this found.

(No audio from 12:03 to 12:23)

So instead of zero, you put some mu into I. So this is the relax KKT condition for this SDP problem, and deriving doing something with this trying to develop Newton method for this, can also lead to interior point methods for SDP. An IP methods for SDP exist, when we will and they have polynomial time, and we will try to discuss bit towards the end of this our discussion, now but will not give too much stress to, so (( )) going a path by which you can do, IP methods for a SDP.

(No audio from 12:56 to 13:06)

There is a software calls Se Du Mi for semi definite programming, so there is a software which you can use and download it from the net, and try to solve semi definite programming problems, may not be (( )) very large scale what written standard scale you can do it. Now, in before jumping in to this (( )) of what is how can we really solves semi definite programming problems, it is having it should be useful to know why semi definite programming is so exiting. Is not there, because I just found that I can right linear programming in semi definite programming problem, and for SDP I can right IP method which are polynomial time, but we have also just discuss that for linear programming problem, we can right methods which are polynomial time.

So, it is not that SDP is only relation between SDP earlier SDP is only use is for may be trying to solve a linear programming problems, you see we will talk about now the relation between integer programming in a SDP.

(No audio from 14:17 to 14:30)

So, we will consider a very standard quadratic integer programming problem, and x i for each i has to be either minus 1 or plus 1. There are 2 options. So, this a very important class of problems call integer programming problem - the Q is a symmetric matrix, you could assuming to the positive semi definite also does not matter. Now, how do I integer programming's are very difficult by very nature, because if n is very large then it is one might say what is problem, just you find out the variables which have this minus 1 or plus

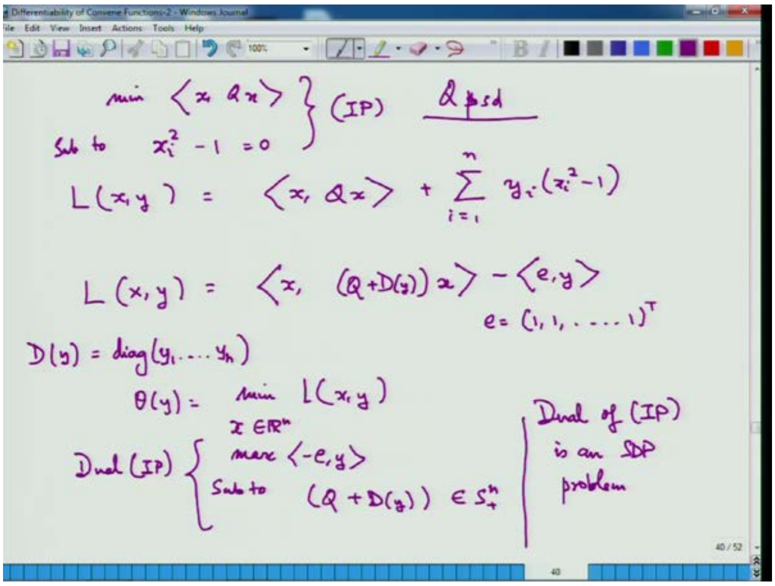
1 values, that there are actually if you take in two dimensions. Then it is fine, you know it is there are 4 possible cases; and then you just figure out with the 4 possible cases; what is the function value, but the question is if there, if I am talking about much much higher dimensions - an n is very large, so that huge number of value then its not possible to evaluate the computable heavily slow down, if you want to make a direct enumeration.

So, how do you solve such difficult problems? So, integer programming, if you could have solved discrete optimization problem, there would have been no study of optimization theory, an it would would not have been not only an an an exciting game. It is an exciting game, because we know lot about continues programming, will no about lot about continuous optimization nothings that we are discuss right, they was nothing no discreetness of this nature, in the feasible set. This discreetness is very important, because this is practical - practicle problems demand discreetness.

And so from from the practice point of view, this integer programming problem, one discrete optimization problems are very very important. But how do we, but on the other hand we have much better solution methods, and very clear cut techniques techniques or other technologies to solve convex continues optimization problems. So, how can I convert this sort of problems in to a continuous problem, and try to solve this. So, one equivalent formulation in the continuous form is as follows.

So, I am writing down into equivalent formulation which is given as follows. So is minimize, you can put half X, Q x also does not matter. Subject to here it was... So, I have now what - n constraints. This is g 1 x, g 2 x , g 3 x for all of them are equality constraints. So, now I have converted this problem into a problem with a quadratic problem with quadratic equality constraints.

(Refer Slide Time: 17:59)



Now, if I construct the Lagrangian of the this function, let me just write write let me a write down the problem again, you are remembrance. So I am having a problem of this sort, minimize this subject to… Now, what you do is let us write down the Lagrangian of the problem which we already know how to write down.

(No audio from 18:16 to 18:26)

See, y i as know sign here, because it is in equality constraints. This is my Lagrangian function. Now, I can write down the Lagrangian function in a much more C compact way is as following. So, L(x,y) can be written as, so this is my IP problem - integer programming problem.

(No audio from 18:51 to 19:10)

Where, D y is a diagonal matrix.

(No audio from 19:16 to 19:28)

Now if I one, now if I one to find the dual of Lagrangian dual of this problem, what I have to do is find theta y which is to minimize L(x,y) or x element of R n. Now, this minimum value if you observe that if, this is the always greater than equal to 0; then this is greater than equal to 0. So this is if this is always greater than equal to 0 for every x means, this is positive semi definite. So which means, my dual problem - dual IP is given us problem of maximizing minis e y subject to the liner matrix inequality; so this is also semi definite programming, this is semi definite programming problem, because it is in the form of a dual of SDP problem, and dual of a SDP problem is also an SDP problem.

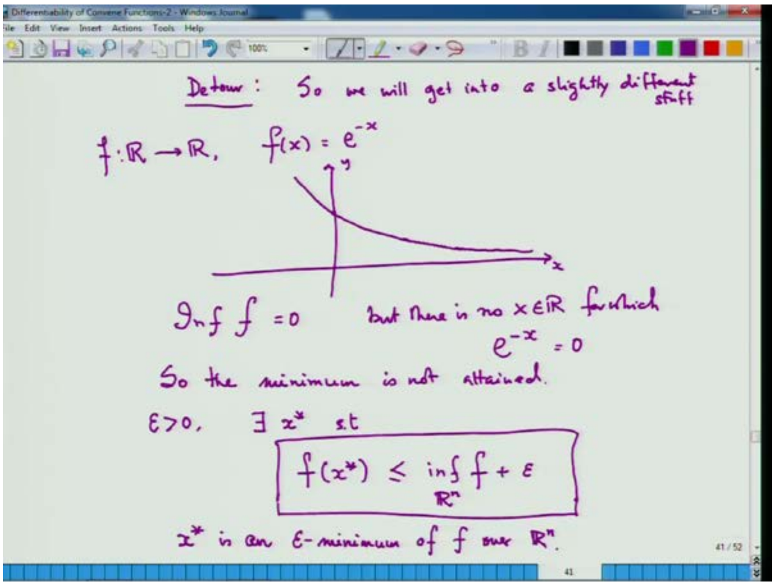
Dual of IP is an SDP problem, you see a very important class. So, now I can solve the dual the SDP; there may need not be, because you non convex problem then that need not be strong duality holding between these two the original problem and the dual problem, but by solving the dual which can be done by using semi definite techniques from semi definite programming. We can immediately tell you, what is the lower bound to the original integer programming problem - An integer programming problem, these are the problems we are actually looking for bounds. We are looking for lower bounds on the minimum value, we are not immediately wanting to find the exact solution. So this is very one very very useful area, where semi definite programming can come in.

In fact, if you write down the dual of the semi definite programming (( )) this problem as you take the dual of this problem, because then you will see that it will be it will look like a exactly a semi definite programming problem. And it will come in terms of the problem data, where this vector e i, I am not writing every time e is this. So here we see, how a problem of quite high level difficulty, you have as an integer programming problem can be approximated very fast in terms of much more simpler, and tractable problem the semi definite programming problem.

So this is something very important, there is a non-convex problem, you are actually formalizing it is in such way at they become convex optimization problems. They that that you look look it through the i of convex optimization problem. Even if you have taken Q to be Psd; that does not matter; Q Psd has no saying Q to be Psd has no saying, because the way by the very simple fact that, if Q is Psd does not matter - this ultimately this is a non-convex set. So, minimizing convex functions over a non-convex set is non-convex optimization. So non-convex optimization is been approximated fast, and some information can be generated vertex solution by solving a convex optimization problem; which is in the form away in SDP problem, and that is why is semi definite programming problems are so important; this is a actually called a semi definite relaxation of the original problem.

So, now you have a fairly nice idea what, semi definite programming problem is, and how it could be useful. So, and I have already given a hint how can any conditions are developed and how, they can be put in to the frame work use for solve, use in the IP methods for solving linear programming problems. Now, because we are taking a tour of convex optimization.

(Refer Slide Time: 23:50)



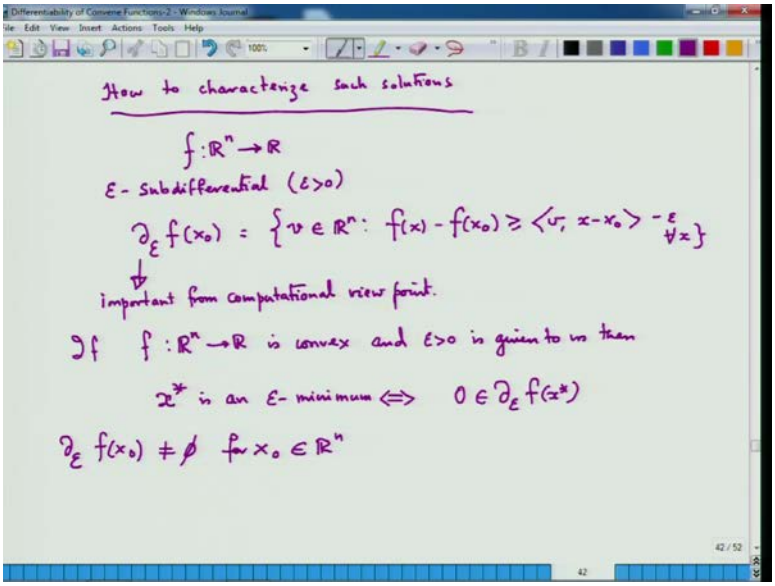
Let us check a little bit of detour. So, what is this detour meaning? So we will get into a slightly different stuff; so we had been talking about linear programming an SDP almost looks similar like it, the whole all the functioning, but it is not; it is very power full, because its convex optimization problem, which you are almost talking as if you are doing a linear programming problem. So we will get into slightly different stuff. In actual optimization, it is not always possible to get the minimum. For example, if I take a simple function from R to R, and it is a convex function f(x) is equal to e to the power of minus x.

So what is happening is that, you see if I take this function, an if I just try to draw the graph of this function. This is x, and this is y axis, this is f x. So I had 0, it is 1, so this 0, 1; so this 0, 1 point this thing is coming, and going down, and down, and down, and down. It goes as near as possible to the x axis, but never touches the x axis. So the infimum value of f is 0 that, it is lower bound - not the greatest lower bound is 0, but there is no X in R for which; so the minimum is not attend, minimum or infimum in this case is not attend . So what can be say of such situation, but if the infimum or the lower bound exit, if you give me any epsilon greater than 0. Then there would exist by definition, there would exist an x star; such that, f of x star is less than equal to inf of f; say our R n if you do not mind plus epsilon, you change the epsilon the x star will get change.

So, once I given give the epsilon, you are always able to find this x star. So, any x star which given an epsilon which if it is satisfy this property, it is called x star, it is called an epsilon minimum of f over R n. In real applications, we are essentially looking for epsilon minimum, we do not want to find the exact minimum; as we have seen in when we are studying the complexity of interior point methods for linear programming problem. We wanted to stop, when the duality measure comes down below thresh hold epsilon, we want stop the algorithm.

So, basically what you get is some sort of approximation solutions, and not the solutions that may not the exact 1; in most cases you do not find exact solutions.

(Refer Slide Time: 27:55)



So now, How to characterize such solutions.

(No audio from 27:57 to 28:15)

In case of convex functions which we are interested in… There is an notion call epsilons of differential, we have already know, we already have a studied the notion of the sub differential; here we are looking at the notion of epsilon sub differential. The epsilon sub differential which for which we define as follows at any point x naught is given as set of all x i in R n or say let us just change it v in R n, such that f(x) minus f(x) naught is bigger than v times x minus x naught minus epsilon for all x. Now, this del epsilon x naught is very important from the point of view of algorithms; important from computational view point.

(No audio from 29:21 to 29:38)

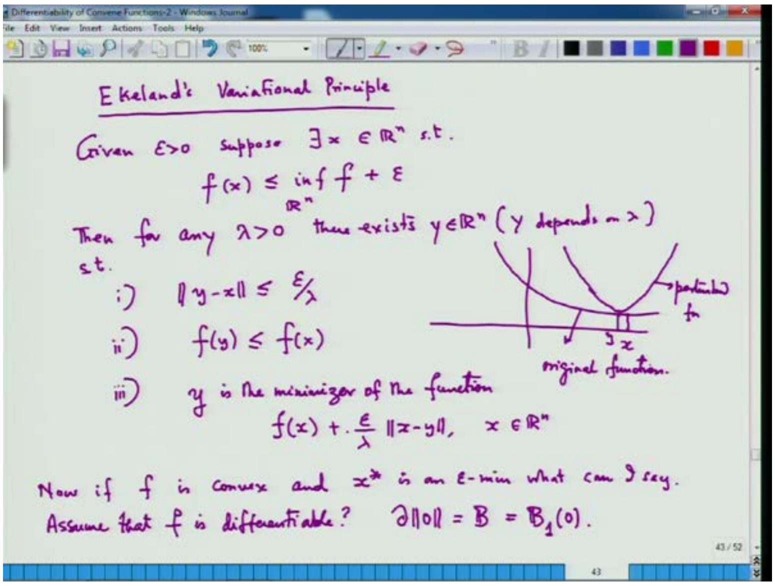
So, what do I get? What do I get here is as follows. See here, if any if this function is differentiable, it does not tell me that this epsilon sub differential which is convex set, it need not be, it need not become a single term; that is very very important. So, if f from R n to R is convex, and epsilon greater than 0 is given to us; then x star is an epsilon minimum; if and only if which is usually use this sign is used, if and only if 0 is belonging to. Now, very sorry zero is belonging to del f e x star. Now, one important thing to note that this function this one del f e x star, so del f e x naught in general del epsilon f(x) naught is not equal to phi, for all x naught in R n.

So, this is a very very important aspect of this sub differential, that is never empty empty right. Now, what would happen if my function is (( )) f is differentiable. What, because I cannot say anything about the sub differential about the epsilon sub differential, that it will be nothing but grad f(x) on a grad f(x) naught. So what can I say if the function is differentiable, f(x) is differentiable.

This would lead us to a study of the Ekaland Variational principle

(No audio from 32:00 to 32:15)

(Refer Slide Time: 32:02)



They cannot very short principle, which is the for reaching result in the theory of the optimization is a very very powerful one, and it tells us how and approximate solution actually behave. So, we are now going write down the Ekaland's Variational principle. We will show how Ekaland Variational principle can be used; see Ekaland Variational principle says the following. So, given epsilon greater than zero, suppose there exists x in R n; such that f of x is less than in plus epsilon. Then for any lambda greater than zero, there exists y element of R n such that number 1.

So far any lambda strictly greater than 0, there exists y lambda y; so y of course, depends on lambda which we let me just write y depends on lambda. Such that norm of y minus x less than epsilon by lambda, number 2 f of y is less than equal to f of x, number 3 y is the minimizer of the function.

(No audio from 34:50 to 35:02)

So, if I take this function, y is a minimizer of this function; that is the whole thing. So, this is something which is important, and we should keep in mind. So what it says, that let your function - original function need not have a minimum; and it has a epsilon minimum, and let this be the epsilon minimum. Then, I can find some epsilon another epsilon point near it, so that that point is the part is the exact minimum, other unique minimum a part of function. So this is my part of function, this is a exactly the part of function. And this is my original function Now, if f is convex and, x x star is an epsilon minimum what can I say.

(No audio from 36:25 to 36:33)

That and assume f is differentiable.

(No audio from 36:41 to 36:57)

So tomorrow, we are going to compute this facts; and we are going to show what happens if we take this question. A very interesting fact we will immerge, and you will see that, even if the function does not have a exact minima, and if the convex function is differentiable, we can tailor lot about the characterization of its in exact minimum. So, so basically you we are having an exact in exact minimum or a epsilon minima, and we are finding another epsilon minima right, because if this is true then again put this is less than f y, the so y is also epsilon minima which is a unique minimizer of some other function. So, this part of function place a fundamental role in this story.

So, I would just like you to take a look at this very very carefully, because this is the very very important result, and its very very important to its of great importance to know these are, if you are going to study of optimization theory. Because in actual practice, you are going to look at only epsilon minimizes, you are not going to look at true minimizes. So, tomorrow we are going to discuss starting of from here, what happens when x star is epsilon minimum what can I say that if it is if it is f is further differentiable. And for that, you would have to know a very very important result which says, if you take the non-function. And find the derivative at zero; this is nothing but the unit ball that is the ball of radius 1 center at 0. And this fact would be used in tomorrow's, we have we have to use it here, in tomorrow's discussion, thank you.