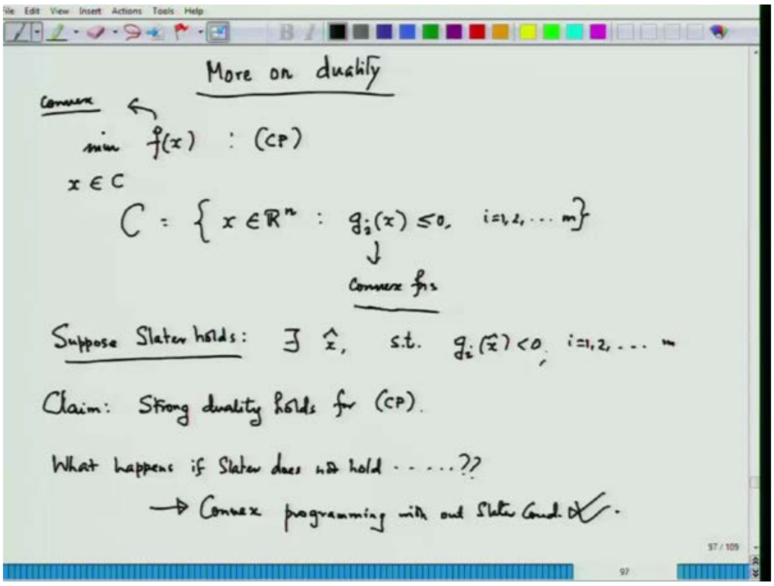
Convex Optimization

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

Lecture No. # 18

So, as promise in the last class, we had two proofs; one is the proof for strong duality for convex programming problems, another is a slightly different thing, it relates to linear programming problems. So, our main issue was to prove the strong duality for convex programming problem, and show that if the Slater condition fails, in case of the convex programming problem, then a finite duality gap may arise. So, the duality gap, which is the difference between the minimum value of the primal and the maximum value of the dual is 0. When Slater condition holds, and could be nonzero could be finitely nonzero, if the Slater condition fails.

(Refer Slide Time: 01:20)



So, let us first consider the problem. Let us just go back to inequality constraints, how we will do it, if the equality constraint would be homework. So, this is my problem, CP where for our convenience C is that consist of all x in R n, which satisfies a following m inequalities. And each of these are convex functions; and this naturally is a convex function; no differentiability assumption has been assumed that all.

And you see; now, suppose Slater holds, that is there exists x hat. Such that, g i x hat is strictly less than 0 for every i. My claim is the following, strong duality holds for CP. Now, the issue is a Slater condition must hold, that is we are assuming that this set C has a non-empty interior.

Now, you might ask me the question, what happens if Slater does not hold? Does that mean that my strong duality can fail even for the convex programming. In case, the answer is yes, we will show by an example. That strong duality will fail even if Slater holds, and the problem is convex a very, very simple looking problem.

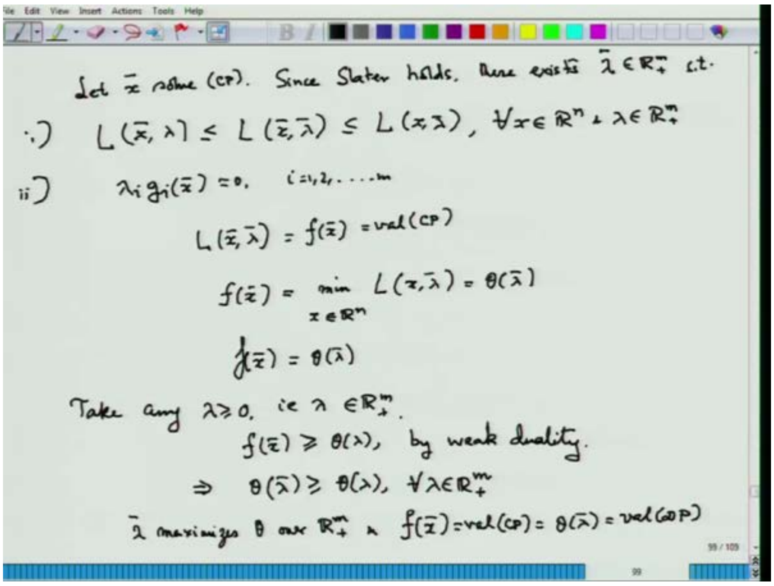
But what can also happen; that you might observe that the Slater does not hold, but strong duality is still holding. The question is why the Slater is holding, but the strong duality results are not holding. In such a case, you really have to note the following that. If you find that strong duality holds, even if the even when Slater is not holding, and there is something else, which is slightly more general than Slater such statement, such conditions are holding on the constraints. So, there could be a constraint qualification and for a convex programming problem more general than Slater. And that condition holds that give rise to the strong duality results.

Also Slater condition is not the only condition on which the convex programming is dependent. If you have listened to the lecture, you would have observed Slater, Slater, Slater, Slater, and Slater. So, Slater is not the only condition on which this is dependent, and one can actually go beyond this. But keeping view the level of the course that we are going to speak about, we are not going to get into those things. That is the area of current research, which has started from the late form the may be just from 1992 and 2000. So, it quiet recent research.

So, those who does not is not satiated by the answer, and once to have more specially, those who are actually in the field of optimization post and PhD students in optimization or listening to this lecture. They could just go to the internet and search for convex problems, convex programming; without Slater condition. That is the huge literature on it. Also if you look at into this book; in which, I am also one of the author’s optimality conditions in convex optimization. A book which at already mentioned.

So, in this book published by tile and Francis, it is jointly with professor Anulekha Dhara. This book also contains a lot on material concerning to this fact. So, let us coming to our own world was Slater holds, which is the good problem. And most of the standard optimization problem Slater holds; standard convex problem Slater holds.

(Refer Slide Time: 06:41)



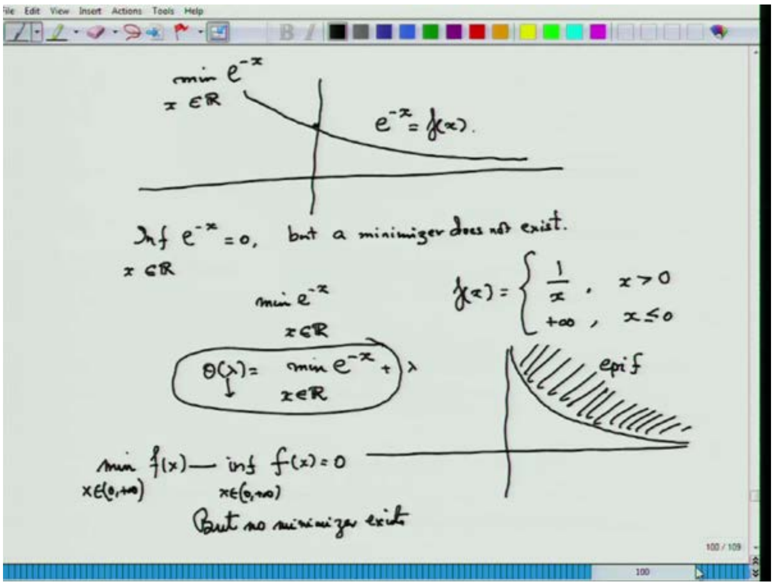
So, now, let x bar solve CP. Since, Slater holds (no audio from 06:52 to 07:02) there exists lambda bar element of R m plus. Such that, L x bar lambda is less than L x bar lambda bar less than L x lambda bar. This is one condition that holds true for every x in R n, and lambda element of R m plus. Let me just tell you that, I do not tell mentioned what a Lagrangian is again, because you have already by know seeing the term for quitter long time, to really immediately know your mind. That this is nothing but f (x) plus lambda 1 bar, lambda 1 g x, lambda 2 g 2 x, lambda 3 g 3 x, plus dot; dot; dot. And the second condition, which is called the complementary slackness condition, tells us this. That is both of lambda is an g i x bar cannot have strict hold with strict inequality, at the same time.

Now, what do I have? I have the following, I know that L of x bar lambda bar, it is actually f of x bar, and which is nothing but the value of the problem CP. So, you see our saddle point condition leads your strong duality, once you know that. f of x bar is equal to minimum of L x lambda bar x element of R n. This is a standard thing that you know from saddle point condition, this fact means this. So, this condition on the right hand side, this one leads to this condition. Now what is by definition, this is nothing but theta time's lambda bar. So, what I have is f of x bar is equal to theta times lambda bar.

Now, take any lambda greater than equal to 0, that is lambda element of R m plus. Then what would happen is the following? You will immediately know that f of x bar is bigger than equal to theta of lambda, by weak duality. See once you know this fact, then you will immediately see this would imply taking into view this fact. That theta lambda bar is bigger than equal to theta lambda, for all lambda in R m plus. So, what does that show? It shows lambda bar maximizes theta over R m plus just by the definition of maximum.

And so, the value of the dual problem is theta lambda bar. So f (x) bar, which is value of the primal problem CP is same as theta lambda bar, which is the value of the dual problem DP, and that is exactly what strong duality means.

(Refer Slide Time: 11:34)



Now, every convex programming problem did not have a point, where its minimum value is achieved, because if I take C to the power minus x. So, if I consider the program, and if you draw the graph of its function and 0, it is 1. And then it is functionally goes down to a 0, very bad looking map drawing. So, this is the functional value e to the power minus x is f (x).

So, now it goes down towards 0 very fast, it assumes totally approaches the x axis. But there is no point on the x axis or the real line, where e to the power minus x actually becomes 0. So, the minimum value, so the infimum value of e to the power minus x is 0, but a minimizer does not exist. So, this can be a situation.

So, if I have such a problem, then what is my dual problem? So, for example, in what happens to my dual problem. For example, in this case; if I look at this problem, my dual problem is that is not g i x. So… (No audio from 13:18 to 13:29) Now, x element of R t is system of constraints. Obliviously; there is an x for which x is strictly less than 0, Slater condition actually holds in that sense.

So, in this case if I write theta lambda, this unconstraint case. So, can I write a theta lambda, my question? See in this unconstraint case, if I try to write something like this, and so my Lagrangian is this there is no g. So, there is no lambda. Then I cannot define something like a theta lambda, because there is no such lambda here. That is why? The writing of Lagrangian, and the writing of the dual problem only make sense, when you have constraint problem. You might ask me, you have given example of an unconstraint problem. Where, there is no the convex problem and there is no point, where it reaches the minimum.

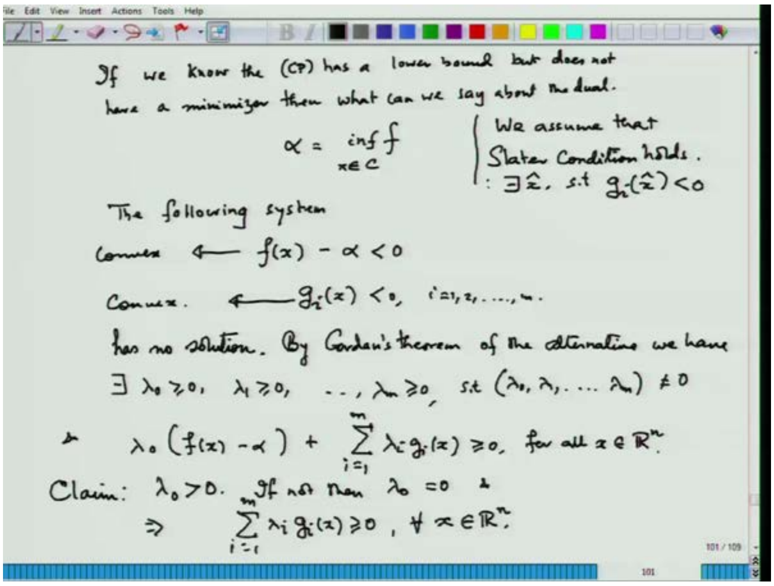
Can you give me an example, where constraint problem where to such a thing is happening? I will give you an example.

(No audio from 14:37 to 14:52)

So, here I have a convex problem, which is extended valued. And then this convex problem is actually this, the epigraph is this. This is your epi f. Now, I consider minimum of f (x) over x element of 0, 1. Then of course, you observe the same thing happens that this function have x becomes larger, the function value rapidly goes towards 0, but it does not never reach 0. There is no point x for which 1 by x become 0. It rapidly to declines, so 0 what there is no x for which it goes to goes there.

So, what happens here again is in f of f x over not 0, 2. I mean for your convenience I will make this or plus, but no minimizer exits. So, here is the constraint case; of course, I would have taken like this, and have a whole thing on the closed set, but does not matter. Let me now go back an answer ask this question.

(Refer Slide Time: 16:46)



If we know that CP has a lower bound, but does not have a minimizer.

(No audio from 16:56 to 17:15)

Then my question is what happens to the dual problem? Can we say anything about the dual? (No audio from 17:22 to 17:35) So, let us now argue step by step. So, let us see what would be the steps of that argument.

The argument is as follows. Now, let because this problem has a lower bound, the problem has an infimum. So, let alpha; this is what you know. Now, once I know this the following is true.

(No audio from 18:01 to 18:28)

This system has no solution. Now, this has no solution; then what can we further say? So, before we get into any other thing just again, let us; iterate we assume that Slater condition holds. (No audio from 18:59 to 19:13)

Now, this has no solution, by separation arguments or by Gordan's theorems of the alternative, which we have already worked through. So, my Gordan's theorem of the alternatives, we have that there exits lambda naught bigger than equal to 0, lambda 1 bigger than equal to 0, lambda m bigger than equal to 0. Such that, lambda naught, lambda 1, lambda m collectively this is not 0. And lambda naught… and you reach this expression.

So, this is what you get, because this above system is no solution, and each of these functions are convex. This is very clear, because this comes out from the data of the problem. Now, if this I have to put i equal 1 to m.

Now, let us see, what we can do. If I put Slater condition, let me see what happens? Now, you see I will assume that lambda naught is 0, because of the Slater condition, I am trying to prove that lambda naught is strictly greater than 0. But we will go by the reductive add of certain procedure; a mathematics of proof by contradiction. So, let me may be is better to writing this way is much simpler to understand.

First let me write what I claim. The claim is that lambda naught is strictly greater than 0, because Slater holds. Interesting part of mathematics has the many few a lecture, and many do a proof of a thing. You possibly do it step by step, and it appears so structure, people would wonder; how did one get into one main such a structure, how could one figure out that this would be the structure of a prove.

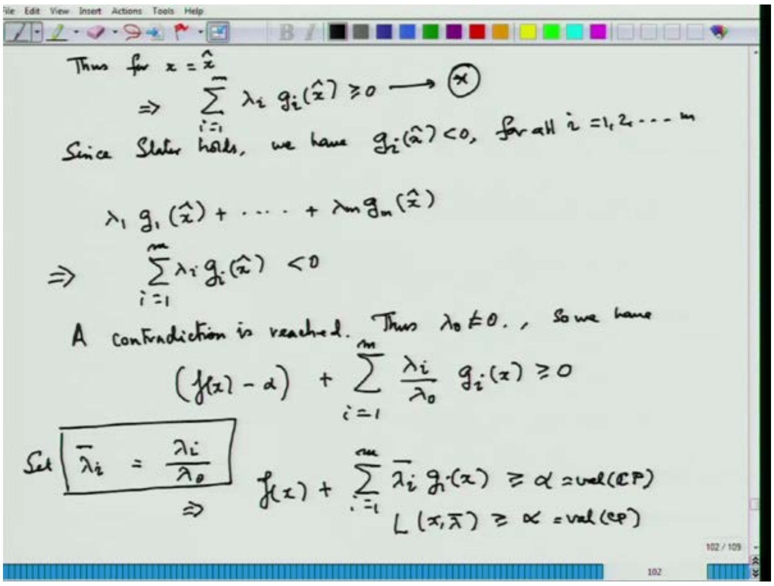
But know let me assert you that all this, where done earlier with guess, and test, and getting ideas from numerical examples. Later on these things, where more formalize and unified and put into a form which can be readily accessible by which is readily accessible by many people. And that is why, you would see that most mathematics proves a highly structure; unless you know, some it is a issue of research proof. Our proof of completely new thing mean given where there can be lot of hand waving, and you know lot of things, which you are not very clear.

So, when we were talking about standard things in mathematics, you will see the prove is very structure very well written, because it is not, because somebody certainly came and one day started writing all those, because it a lot of peoples idea gone into refining the idea improving the theorem.

So, there is a very famous statement by a mathematician called Geon, called routa an MIT. Who speaks about F. Riesz. Riesz representation theorem is absolutely fundamental to functional analysis, and that is by Riesz representation theorem, we can show that every linear map on R n is nothing but the inner product, the dot product, which is very very important thing. So, when Riesz used to write a paper, he never wrote it down immediately. He wrote a paper publish in oxford general, and kept on refining it with the new idea has been pushed in, and published in some still better channel, while kept on marinating on it. And then at much latest stage you would publish a final version of what you wanted.

So, this is the thing that person like a F. Riesz is to do. So, refinement of a theorem comes after many, many steps. What you see the very define version, you might think that I am just giving step by step, step by step, as if by magic I have come to know the rules. No these are all by guess and test, and then it has been refined, and written in this form. So, now I had my claim is lambda naught is strictly bigger than 0. If not then lambda naught is 0, and this will imply that this expression is now greater than equal to 0, this expression vanishing, because you put lambda naught equal to 0.

(Refer Slide Time: 25:20)



Now, this expression holds for all x, because this expression had been true for all x. Now, in particular it is true for the x hat for which at every i, g i (x hat) satisfied with a strict inequality. Thus for x equal to x hat, it would imply that summation i equal to 1 to m. (No audio from 25:39 to 25:42) This what we have from the other equation, just we in the last page this one.

Now, since x hat, since Slater holds sorry not x axis, it is Slater holds. We have… Now if I write down this expression. This expression would be this expression would look like this, but now look at this expression forget this one.

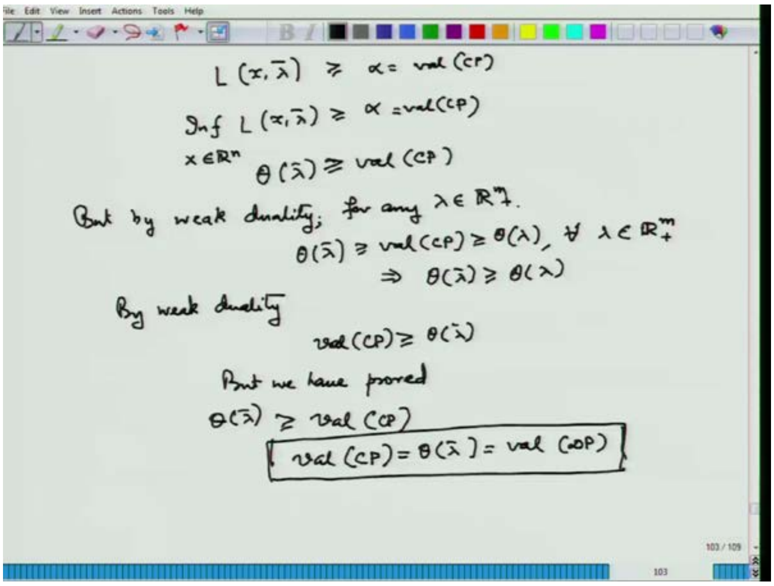
Now, for each of these are strictly less than 0. Lambda 1 to lambda m all are bigger than equal to 0, but since lambda naught is always has been assumed to be 0. Then the hope and we know that the whole vector lambda naught, lambda 1, lambda 2, dot, dot, dot, lambda m is not 0. So, among this lambda 1, lambda 2, lambda m, one nonzero quantity is there, because of the fact that we have chosen lambda naught equal to 0.

So, this would imply that summation in this expression, because these are all, because there is one element of them positive, say lambda one is positive. So, this will become as strictly negative quantity. So, finally, this… So, what you have finally is this expression, but I also had this expression form the other one.

So, putting x equal to x hat, because this true for all x putting x equal to x hat in particular by choosing the particular x hat, with for which Slater is holding. I will get this, but actually what I should get, is this. So, there is a contradiction. So, this implies thus lambda naught is not equal to 0. So, we have f of x minus alpha plus summation lambda i by lambda naught. So, we divide both side by lambda not, because lambda naught is positive.

So, set lambda i bar is equal to lambda i by lambda naught. So, this implies now, f of x plus summation i equal to 1 to m lambda i bar alpha which is the infimum, or which is the value of P, this is also the value of P. The infimum or the minimum value of the problem CP sorry not CP. Now what is this? If you look at it very carefully, what is this? What is this expression? It is nothing but L x lambda bar, and that is greater than equal to alpha, which is the value of CP. See, for the original problem, I just know that it has a lower bound, I do not know whether there is a minimizer for which the lower bound is attained, or the infimum is attained.

(Refer Slide Time: 29:52)



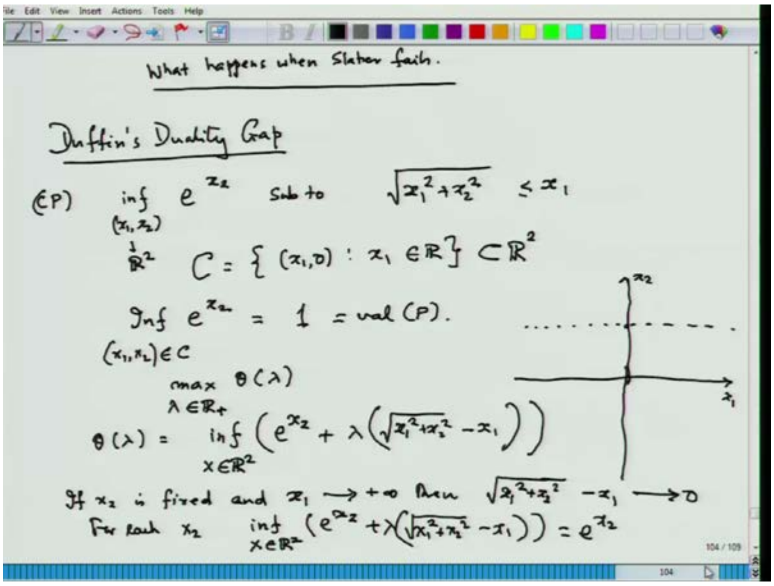
So, let me rewrite, L x of lambda bar is bigger than alpha is equal to val of CP. So now, if I take the infimum over all x in R n of this expression sorry x lambda bar, this becomes bigger than alpha, equal to value of CP. And this is nothing but by the very definition theta lambda bar is greater than equal to value of CP. But by weak duality (no audio from 30:39 to 30:50) for any lambda element of R m plus theta of lambda bar, is bigger than value of CP, is bigger than value of theta of lambda.

But again by weak duality now, value of CP is actually bigger than theta. So, this would imply from here that theta lambda bar is greater than equal to theta lambda. So, you have a lambda bar which maximizes theta. So, you know the dual upper bound exits, the dual maximizer exits. There is a lambda bar for which the dual function is maximized. But you also have here that by weak duality this is holding, value of CP is greater than theta time theta of lambda bar. But we have proved that value of CP is obviously, less than theta lambda bar.

So, combining this, I have value of CP is equal to theta lambda bar. So, what is theta lambda bar, it is nothing but value of DP. So, interesting part is that, even if the original problem does not have a lower bound. Sorry, I am again mistake, I take that. Even, if the original problem does not have a minimizer. It can have a lower bound, it has lower bound. That is what we have assumed. But it does not have a minimizer. There is no x, for which that infimum value is achieved, that alpha value is achieved. There is no x such that f (x) bar is alpha. Then it does not matter. If the dual problem is feasible it is all right. If the dual problem is feasible which it is, because dual problem is always feasible here, because lambda is greater than equal to 0, lambda is in R m plus which is all right.

So, then the strong duality, not only hold the under Slater condition, but the dual maximum value is achieved. That there is a lambda bar for which, the dual value is achieved. Now, we are going to give an example, what would happen, even in the case of convex programming when Slater fails?

(Refer Slide Time: 33:35)



So, what happens when Slater fails? So, let me write down, this is what? This example is famous in optimization literature as Duffin's duality gap. Here is the problem; original problem P or CP is to find the infimum over x 1, x 2. Both x 1 and x 2 is in R, from this is in R 2, find this subject to root over x 1 square plus x 2 square is less than x 1.

Now, let me figure out what is the feasible set of this problem. It is only telling that x 1 square, whatever we take for x 1 if I take x 2 to be 0, and then my feasibility is guaranty. So, what will what would it happen, it will immediately tell me that whatever x 1 I take, I will have x 1 square plus x 2 square is less than x 1 square. So, I will have x 2 square is equal to is less than equal to 0. But x 2 square is equal to 0, then because x 2 square is also greater than equal to 0. So, I will get x 2 equal to 0.

So, my feasible set consist elements of this form. This feasible set is a subset, in fact a proper subset. So, these are proper constraint optimization problem. And you have to observe that Slater does not hold, Slater condition does not hold that is if I put 0 here, and if I take any x 1, it will become x 1 equal to x 1. So, there will be no Slater condition holding. There is no strict in strict inequalities are never strict, at all point this inequalities active.

Now, x 2 is 0. So inf of in C, now is nothing but e x 2 is 0 is1, and that is the value of P. So, what about the dual problem? Dual problem is max of theta lambda, lambda element of in this case I have one constraint. So, it is in R plus. So, here m is actually 1. Now, this will be something very simple observe that, I can write theta lambda as inf of x in R 2, I am constructing the Lagrangian L. (no audio from 36:42 to 36:52)… this minus this will be less than equal to 0. So, this is my theta lambda.

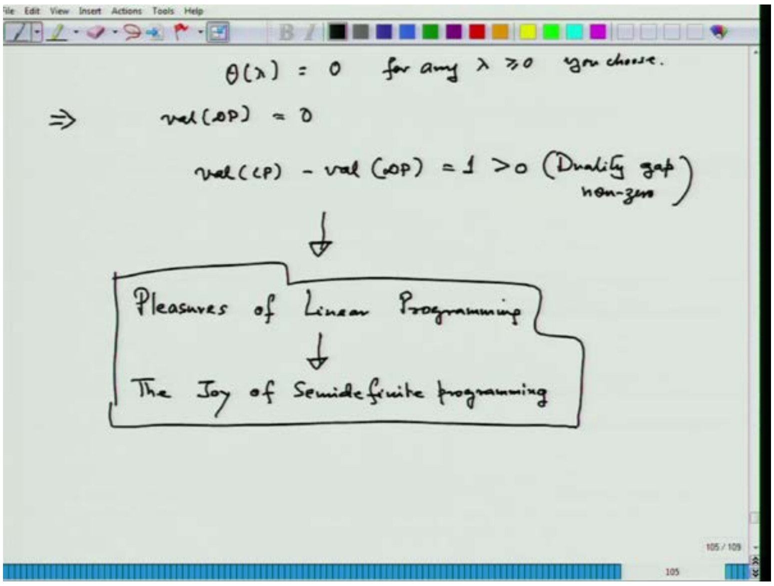
Now, how do I compute that theta lambda? How do I know, what is theta lambda? We will show that theta lambda is 0, for whatever lambda we take. And so, the maximum value would be 0, and I say duality gap. Slater does not hold.

Let us see, how I compute this. The interesting part is that you see what would happen, if I fix up the x 2. So, this is x 1 and this is my x 2. If I fix my x 2; and then, I vary my x 1. So, these are the points I will work through, then as x 1 becomes larger and larger. So, as if x 2 is fixed, and x 1 I ran them to plus infinity, then does not matter even if goes to. So, what would happen if I ran this to plus infinity? In the difference between these two keeps on shrinking. So then x 1 square, because x 2 is fixed, because finally the infinity the larger number will don in it. This is going to 0.

So, on this line means once x 2 is fixed. Then the minimum value over x 1 is nothing but x 2, so infimum for a fixed x 2 for each x 2, infimum x over R 2. So, once I fix the x 2, I can only move along the line which is passing through x 2, which is parallel to x 1 axis. So, basically infimum over x 1, which is same as infimum over x 1, infimum over x element of R 2 is seem as infimum over x 1. x 2 plus lambda root x 1 square plus x 2 square minus x 1, this thing is nothing but e to the power x 2.

So, I know the minimum value of the function over this line when x 2 is held fixed. But if I change the x 2, again I will know in the same way the minimum value, which will be e x 2 of that particular value of x 2. So, on each of the lines; the function value the minimum value of the function when minimized over x 1 is just e to the power x 2. Now, if I vary the x 2 then I am varying the function over the whole plane, and thus I get the minimum over the whole plane and e to the power x 2 you know goes to 0, the minimum is 0.

(Refer Slide Time: 40:02)



So, theta of lambda is actually 0, for any lambda you choose; any lambda greater than equal to 0 you choose. This implies that value of DP is 0, 0 function. So, value of CP minus value of DP is 1 and which is a positive duality gap. So, even for convex problem, in the case of non-satisfaction of Slater condition duality that may arise. So, there is the if the when the general condition that might satisfy that, which is the something at the level of research, which we had just spoken of. That might not even hold for this particular case. In fact, does not hold for this particular case.

So, with this interesting example, I stop here today. And from the next class onwards we start a fascinating journey, into what I call the pleasures of linear programming. And after we finish this pleasures of linear programming, my next sort of lectures would be I would like to call it, the joy of semi definite programming. You see how much fun, we will have here so much things can be said. A lot of difficult problems can be played along with this semi definite programming.

So, this will be our theme for quite some time, now after this you already have a quiet wide idea about that theory of convex programming. And now we are going into very particular type of convex program, and these are the two mostly important classes of convex programs. So, with this I end today's lecture, thank you very much. And hope to see you tomorrow.