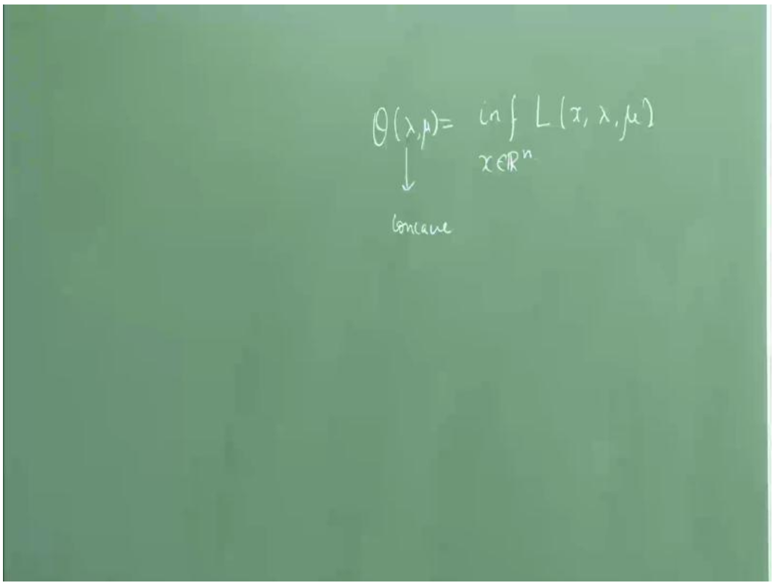
# Foundation of Optimization
**Prof. Dr. Joydeep Dutta**
**Department of Mathematics and Statistics**
**Indian Institute of Technology, Kanpur**

**Lecture - 33**

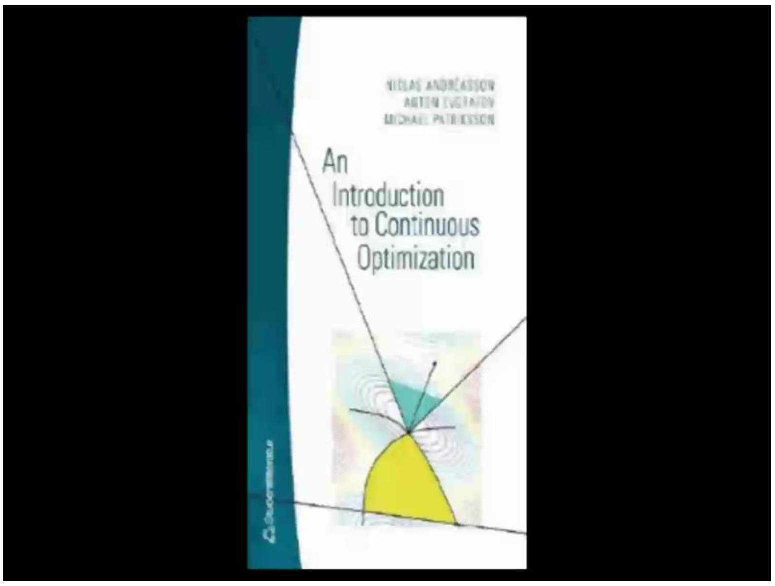
Let us continue our discussion on duality. We will try to wrap it up today or maybe in the next lecture.

(Refer Slide Time: 00:33)



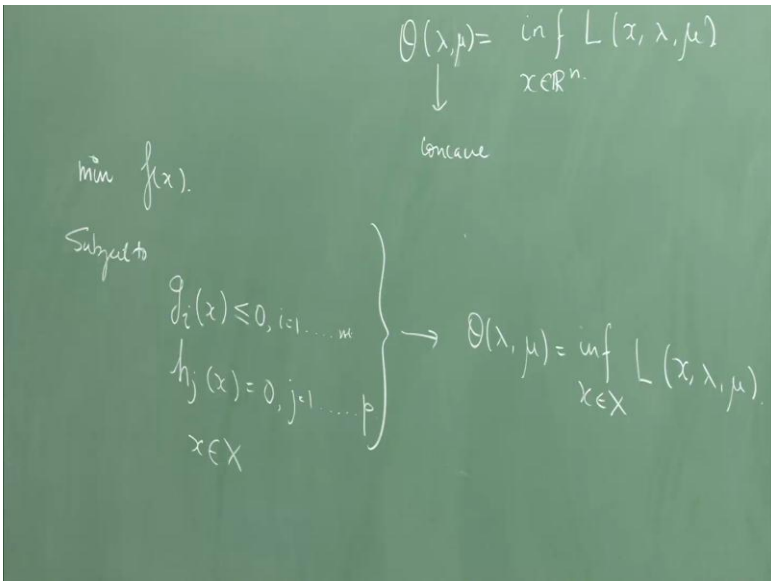
We were talking about the fact, there is a dual function… This is a concave function in lambda mu; jointly, in lambda mu, it is concave. So, we need to now look into the dual properties of this function. If it is concave, then it is negative; if it is convex, then can I find a subdifferential and all those things? And, when it is differentiable? That is also a question.

(Refer Slide Time: 01:37)



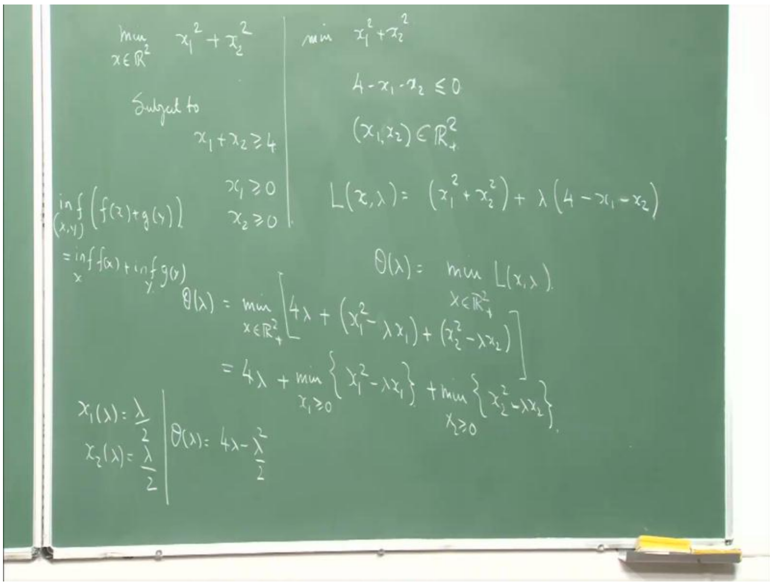

We will start with one example from the text by Andreasson, Evgrafov and Michael Patricksson. This text is called An Introduction to Continuous Optimization. This is available in India. And, in fact, it is an Indian print that I am trying to show you – An Introduction to Continuous Optimization. It is published by Overseas Press in India. And, it is by Michael Andreasson, Evgrafov and Michael Patricksson; Michael Patricksson is a very renowned optimization researcher. And, let us tell you something before I actually give that example from that book that, showing that, there can be a situation, where theta lambda is actually differentiable, not just non-differentiable.

(Refer Slide Time: 02:16)



The interesting part that I want to tell you that if I make the problem slightly more complex, we add an additional constraints. So, these constraints are hard constraints like bounce on the variables. So, in this particular, if this is the situation, then the Lagrangian is lagrangian only consists of the functional constraint; and, this is a non-functional constraint as such as it is represented. Of course, if you have x i between say retaining all the x i's there between minus 1 and plus 1; then, of course, you can write (( )) them as inequality constraints. But you basically would increase the number of constraints. See in order… And, you make the Lagrangian look quiet complex. See in order not to do that, one can put it in this form and then you can again write the Lagrangian as the same Lagrangian. But, in this particular case, when you write a dual function, the infimum of x over x is no longer over R n, but would be restricted to this particular set X.

(Refer Slide Time: 03:59)



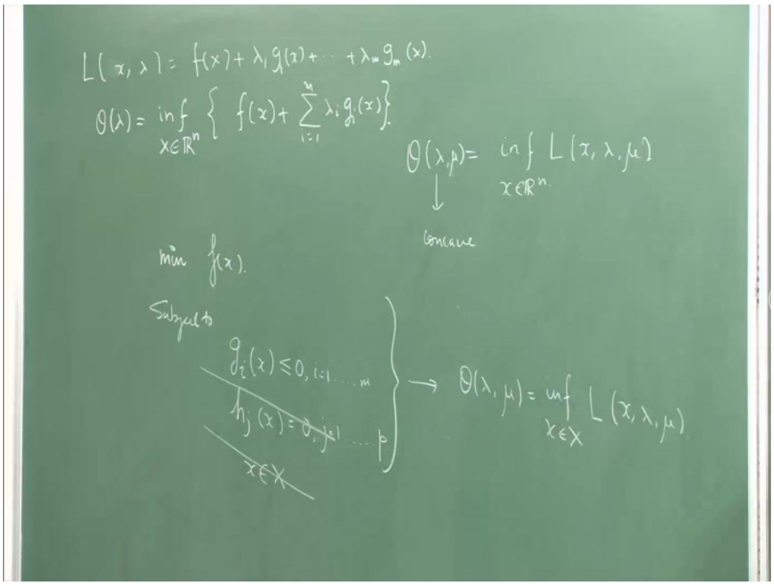

Now, let me provide this example. The problem is to minimize… So, this is a convex function – x 1 square plus x 2 square. So, x 1 and x 2 are in R. So, this is a problem in R 2. Basically, I should write x is in R 2. So, I am minimizing over R 2 subject to a linear constraint, which is x 1 plus x 2 greater than equal to 4. And, x 1 is greater than equal to 0 and x 2 is greater than equal to 0. Now, these constraints are hard constraints; means even if you in real application, means, when you are actually applying the algorithm, if there is even a slight violation of this, this might not be of so big harm to the problem. But, any violation of this would actually change the problem; this cannot be violated. This is called a soft constraint and this is called a hard constraint. I told you quiet often. So, this I can write minimum of… I can write this as… So, there is one inequality constraint. And I can write x x variable, which is x 1, x 2; that is like this, is in R 2 plus; that is the non-negative orthant. So, that is essentially the problem. I should… I made a mistake in writing this; I should have written it like this, instead of putting it like that. I can write it as 4 minus x 1 minus x 2.

Now, if I found the Lagrangian function, I will only club these two in the Lagrangian; I will not keep that in the Lagrangian. So, here I have one constraint say L x, lambda, would be... Now, theta lambda may be known; not mu, there is a mistake; should be lambda. Now, 4 lambda is a constant, because I am fixing a lambda and then minimizing. So, theta lambda in this particular case is to minimize for x element of R 2 plus. So, here I am restricted of these Lagrangian function. Now, observe that, here I have made this restriction quiet similar to the one here.

Now, once that is done, here you see lambda is fixed. So, 4 lambda is something which I can take out. So, if I want to now find theta lambda, basically, I am calling for the minimization of… Let me now write the thing in a proper way. So, 4 lambda is something I need not bother and even take it outside the minimization; plus… So, you see these variables have got decoupled from each other. So, when you want to minimize such a decoupled function, that is… Please remember this very simple formula, which you can prove. See you want to minimize this over x, y. And, this is simply nothing but inf of f x over x plus inf of g y over y.

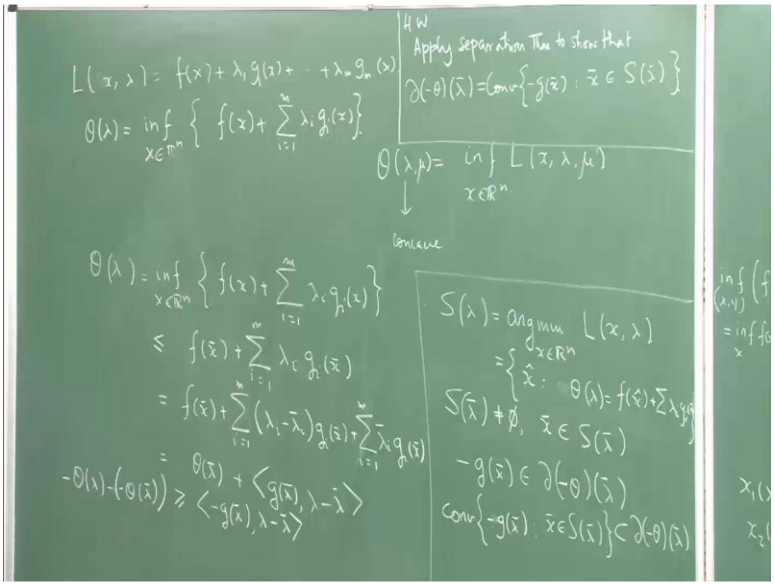
So, now, if you want to do this, this and this… These are decoupled. So, this is acting like f x and this (( )) like some g y. So, 4 lambda can be taken out. So, I can write this as minimum over x 1 greater than equal to 0 x 1 square minus lambda x 1 plus minimum… Infimum actually, but you can actually figure out the minimum of this – x 2 square minus lambda x 2. Now, let me tell you that, the minimum of this is obviously depending on lambda; and, that is given by lambda by 2 and this one is as same as that. So, that is, it looks just the variables are different; instead of x 1, x 2, it is same as lambda by 2. And, theta lambda you can calculate up to with this. So, usually, this is a differentiable function. Now, do not tell me how to calculate; ask me how to calculate the minimum of this one, because this is a real variable thing; you can just figure out by yourself.

(Refer Slide Time: 09:45)



Now, the question would be, if it is not differentiable, if it is concave; its negative is convex, can you find the subdifferential? Can you compute the subdifferential? These are very important thing that one has to know that… Suppose I now take a problem this one, but I cancel the equality constraint and cancel this one. I just keep inequality constraint. In that case, the Lagrangian would look like… It will look like this; it will absolutely be of that form. Now, theta lambda is inf of f x plus… Now, depending on the lambda, the solution set will change; of course, I am just in (( )) whole x, all the x. Now, the question is I really do not know whether a minimizer of this exists or not. If a minimizer of this does not exist, you cannot tell… you can compute the subdifferential, but it would be too too complicated; it will be extremely extremely complicated. Now, if I know about the minimum set, that is, given a lambda, I have an x hat, which is in the solution set. Then, I can tell something about the subdifferential set; hat at given lambda. So, let me just take this part and describe you this fact.

(Refer Slide Time: 12:06)



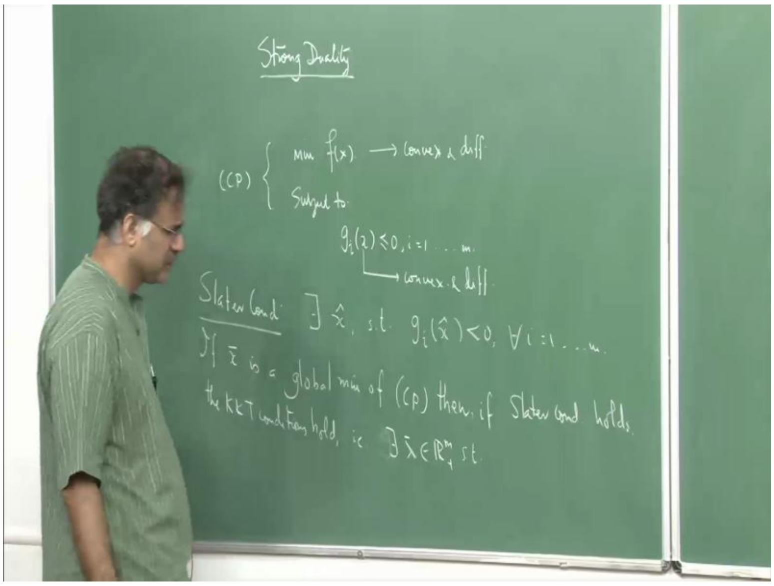

Now, let me look into this simple fact. Now, say S lambda is a argmin set; that is, any… This consists of all x hat such that theta lambda is equal to f of x hat plus summation lambda i g i x hat. That is what it means; as a set of minimizers of this problem; that is all. Now suppose S lambda – let us take S lambda bar is not… phi. And let us take x bar is element of S lambda bar. So, let me just have this information. Once I have this information, what can I do? Let me take any other lambda, other than lambda bar. And then, let me now try to deduce something. So, take some lambda; not lambda bar, take theta lambda. Now, theta lambda is that inf of f x. Now, this one is less than obviously, because (( )) particular choice x bar, this infimum is less than equal to this. So, what I now do is, I add and subtract from each of these lambda i bar g i x bar. So, this will give me a simple algebraic manipulation.

Now, this simple algebraic manipulation will immediately show you what you actually want. So, this… I can combine this to write this as theta lambda bar, because x bar is element of S lambda bar. So, theta lambda bar is exactly f of x bar plus lambda bar g i x bar. Now, these two combines to give me theta lambda bar. And, here I can write it more in a compact fashion. I am writing this as lambda minus lambda bar; where g x is the vector g 1, g 2 dot dot dot g m x. Now, what does this show me? This show me that, if I take a minus sign, because then I will have minus theta lambda minus theta lambda bar, because this minus will be greater; minus of minus theta lambda bar plus theta lambda bar basically; I will take it to the other side. This will become minus here and then go to the other side. It is greater than equal to minus g of x bar lambda minus lambda bar. So, what does it show? This immediately shows me that, minus of g x bar is a subset of the subdifferential of theta at lambda bar. This is what you have. This is this.

Now, say for any x bar, which is here, this is in the subdifferential. So, I can write finally, because this is a convex compact set. Finally, I can write that, the convex hull of minus g x bar is where x bar belongs to S lambda bar is a subset of the subdifferential of minus theta at lambda bar. But now, I will give you as a homework for those who have taken a look at the separation theorems from the books just knowing the statement will do. So, if you know about the separation theorem… This is a homework. Now, apply separation theorem to show that, del of minus theta lambda bar is convex hull of minus g x such that g x bar such that x bar is element of S of lambda bar is actually there is an equality. So, these are very very fundamental formula. Now, if I say that, instead of R n here, I have capital X; and, if that X is compact, then such a thing is anyway guaranteed. So, that S lambda bar is nonempty. So, here the formula that you see is computed only under the assumption that, this is nonempty; which is not always easy to guarantee. But, in many cases, yes.

Now, for example, here whatever we have lambda, these are the solution. So, theta lambda is always this. I know a set x's solutions. Now, what I am trying to tell you is the following. We have discussed about the differentiability property of the dual function. We have also spoken earlier about something called weak duality, which says that, the minimum value of the primal problem is always bigger than the maximum value of the dual problem. So, the dual problem provides a lower bound to the optimal value of the upper to the other primal problem. Now, the question does not lie like this. The question is the following. When are these two equal? That is, when strong duality will hold? And, that is exactly what we are going to discuss at this moment.

(Refer Slide Time: 20:58)



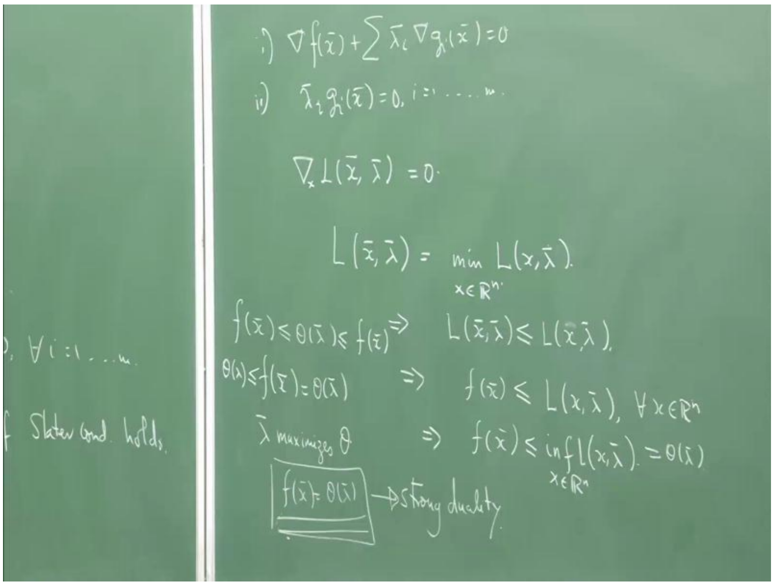
We will now talk about strong duality as we had spoken. Strong duality – that is, the equality of the primal value and the primal optimal value and the dual optimal value. So, what does strong duality tell us? And how do we prove strong duality? Now, here let me tell you that, strong duality is always guaranteed for a linear programming problem. Strong duality is not always guaranteed for a convex programming problem. And, for a non-convex problem, it is never guaranteed. Under some conditions for a convex problem (( )) simple conditions. You can always be sure that the strong duality holds. So, the problem we had spoken of is as the projection problem; we showed that, the geometrically strong duality holds is a case, where all the conditions are satisfied.

Now, let me tell you following thing. Suppose I have a differentiable convex optimization problem. Just we need inequality constraints. So, this is convex. So, when strong duality comes, we are actually in the fold of convex programming. Assume that, these functions are also differentiable. Assume that, this n plus 1 differentiable convex function. Now, once I know that, I would… Let me assume an important condition – a constant qualification called the slater condition of which you already know; that is, there exists x hat such that…

Now, once you have that and if x bar is a local minimum – not a local minimum, it is a global minimum convexity; there is no local minimality. If x bar is a global min of this problem, which I can call for the time being CP; we prefer convex programming; then…

So, if this holds, if problem bar is a global min… then, if slater holds – slater condition holds, we know that the Karush-Kuhn-Tucker condition holds; the KKT conditions hold; that is, there exists lambda bar element of R m plus such that…

(Refer Slide Time: 25:42)




f of x bar plus summation lambda i bar – gradient – gradient of g i x bar – this is 0. And, number 2 comes the complimentary slackness condition, which we have mentioned many many times. This is 0. Now, look at the first equation. What does the first equation say? The first equation says that, if I fix the lambda bar, then this is 0 and the derivative been taken over x. The partial derivative is only over x; which means… But the Lagrangian function f x plus summation lambda i g i x bar is a convex function now for a fixed lambda; lambda positive, non-negative.

So, obviously, this lambda is non-negative. And, the convex function x bar is a critical point of that convex function; which means L x bar, lambda bar – this is unconstraint problem, is the minimum of L x, lambda over all x element of R n. And this immediately tells me L x bar, lambda bar. So, you see we have recovered one side of the saddle point conditions from the Karush-Kuhn-Tucker conditions such that reverse we are looking at the thing.

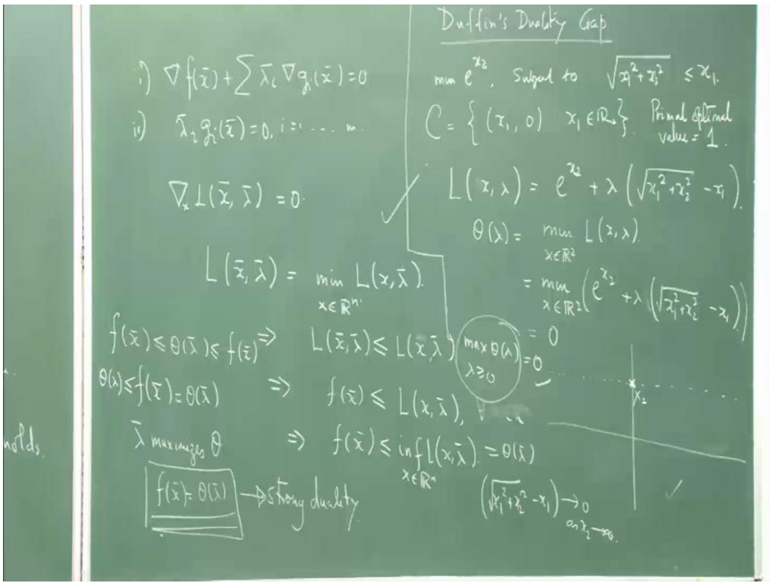
Now, if you look at this, if I break it up knowing that, lambda bar g i x bar is 0; I will simply have here is f of x bar, which is now, is the minimum value of the (( )) problem of the primal problem, is less than L x, lambda bar for all x in R n. Now, this is the fixed quantity. Here for whatever x you choose, this would be bigger. So, this would imply that, f of x bar is less than equal to infimum of L x bar. And, this is nothing but theta lambda bar.

So, what we have obtained from there is f x bar is less than equal to theta lambda bar. But, if I go back and look at the weak duality, then I would have theta lambda bar to be less than f x bar, because x bar is feasible to the primal and lambda bar is feasible to the dual; lambda bar is assessed to be greater than equal to 0. That is a dual feasibility condition. Dual feasible set is always nonempty in this particular case of Lagrangian duality; which means that, f of x bar is equal to theta of lambda bar.

Now, if you take any other lambda, which is feasible to the dual; that is, lambda bigger than 0; then, by weak duality, again you will have this; which means lambda bar maximizes theta; which now means… So, lambda bar maximizes theta. And, at the same time, we also have f x bar is theta lambda bar. You also have f x bar is theta lambda bar. So, what does it say? It says the maximum value of the dual problem is same as the minimum (( )) is equal to the minimum value of the primal problem. And, this is exactly what is strong duality. This is exactly what is strong duality.

Now, there is an interesting thing, which I would not analyse here, because it will be beyond the scope of the class, beyond the scope of this set of very basic lectures, is that, if I do not know about what is the x bar, which minimizes the problem; I only know that, this problem has a lower bound; then, I can still show that, the dual maximum is achieved at some lambda bar and the lower infimum of the primal problem is same as the maximum of the dual problem. But, that we will not discuss here at all. What I will tell you is that, that is, show you now is that, this slater condition is a very important condition. If I take off a slater condition; if I have a convex programming problem at the slater condition fails, this strong duality will fail to hold; that is, f x bar would be strictly bigger than theta lambda bar.

(Refer Slide Time: 31:36)




And, that is what we are going to now discuss through what is called Duffin's duality gap. It is the very famous reserve examplity to Richard Duffin. Now, what I will do is the following. I will write down this very famous problem. Problem is to minimize subject to… The first thing that you have to observe that, when I write root over x 1 square plus x 2 square is less than equal to x 1, it simply means x 1 square plus x 2 square is less than equal to x 1 square; which means x 2 is equal to 0. See if I put x 2 is equal to 0 and take any x 1, then the feasibility is guaranteed. So, the feasible set C in this particular case is set of all x 1, 0. Of course, if I have to maintain this one, because here if I even take a negative number – if I take minus 2; minus 2 square plus 0 plus 4; root over 4 is equal to minus 2. So, minus 2 cannot come here. So, what has to come is R plus; x 1 has to be greater than equal to 0. So, I cannot take a number by the very expression – very expression on this unequality, I cannot take x 1 to be non-negative. So, that is the feasible set.

Now, let me write down the Lagrangian function, which is… Now, look at the whole thing very very carefully. Now, if x 2 is 0, then the primal optimal value – this is a primal problem – primal optimal value is equal to 1. Now, let us look at the dual optimal value. How do I compute the dual optimal value? Now, theta lambda is equal to… is same as… Now, let me look at this whole thing. Now, what I will do, let me fix up a value of x 2 and now vary x 1. So, once I fix the value of x 2, this is a function of x 1. So, on this line, for a fixed… – this fixed x 2 – on this line, I need to find what is the minimum of this function. And then I will vary the x 2. So, basically, I am sweeping this horizontally and then sweeping this vertically; means I am covering all of R 2.

Now, look at this. When x 2 is fixed; as I increase the value of x 1; as I make x 1 go to infinity, make it bigger and bigger and bigger and bigger and bigger; this part goes down to… The difference between this and this continuously decreases; that is, x 2 gets fixed and x 1 becomes so large that it dominates. And, x 1 so the value of this is almost same as this. So, there is difference between root over x 1 plus x 2 square and x 1. That continuously diminishes. And hence now, on this line, the minimum value of this problem – on this particular line for this fixed x 2; say this is x 2. This value is e to the power x 2. Now hence now, I try to vary x 2 over this line and see what is the minimum of… So, basically, what I am doing? Minimizimg over x 1 and then minimizing over x 2.

Now, if I vary x 2 along this line; so, as x 2 goes to at minus infinity (( )) x 2 goes to 0. So, the theta lambda value is actually 0. So, max of theta lambda when lambda greater than equal to 0 is also 0. So, you will see there is a duality gap; optimal value of the dual is 0; optimal value of the primal is 1. And so, there is a nonzero duality above 1. So, this is a very very famous example.

And, I think your optimization education, even in the fundamentals, would not be complete unless you really understand this example. And, it came out in a very famous journal of math programming. And, see again, I would like you to notice how I have done it. I have first fixed up x 2 and minimized this function over x 1. So, I have covered this part. So, what I know is the minimum value of this function over this line. So, for all such lines (( )) have fixed x 2 from here, I go down; I know what is the minimum value. So, the minimum value of the function over the whole x 1, x 2 plane is what I get as I make x 2 go to infinity minus infinity once I know the minimum value over this line when x 2 is fixed. And, that is e to the power x 2. And, that is how we obtain our problem.

So, here we have almost done a lot about duality. We have done various properties of duality. I will give you a small example when I begin tomorrow’s lecture as to how I can make this duality useful. First of all, you should notice that, how can I actually apply this duality, this idea to real problems. Of course, I will give you some homework not to really do, but go and try to find in the internet, find in the text; that shows that, if you have a linear programming problem, then to get strong duality, you do not require any constraint qualification; no slater condition, nothing. And, you will get the same… You will always get strong duality.

Now, just go and find out from the internet. But I will still want you to tell me what is the proof. The proof again would be based on the approach that I have taken through Kuhn-Tucker conditions. So, maybe we will do it tomorrow a bit of duality for half of the class and then we will show how can we use… Again, the idea of the projected – not the gradient, but projected subgradient method to actually make the dual problem computable. Actually, we can do some computation with the dual problem. That is what I am going to demonstrate in the next class. So, I think this is almost towards the end of the course and we will have…

After tomorrow's class, we will have 3-4 more lectures. Of the three lectures, one we will be trying to describe to you how to apply Newton's method to constraint problem. And probably, if we have time, we will talk about sequential quadratic problem. And, at the end, we will give you a brief idea about a new approaches – not a new approach, a old approach now given in a new box, because now, it has revealed a lot more properties called the direct search method, because we now have convergence analysis for that.

See when you do optimization algorithms, it is very very important that you just do not bother about writing an algorithm and say whatever I have got is fine; you have to show that, whatever you have got is fine and you have to do it mathematically; you have to show under what conditions that the sequence at an algorithm would generate would either go to critical point or hit the actual solution. So, these are very very important aspects that you want to do when you want to design optimizational algorithms. So, thank you for listening to me. And we will continue our discussion on duality and bring in some subgradient time methods (( )) use of subgradients to solve non-differentiable convex function, convex problems in tomorrow's class.

But it is very very important for you to realize that, non-differentiability is a very, very important thing in optimization, because when you want to use this dual problem for a convex problem or any non-convex problem also, the dual function is always a non-differentiable usually a non-differentiable convex function. So, it is very very important to know that non-differentiability is something, which you cannot get away within optimization. Though when you had started learning basic maxima, minima, you had really bothered only about differentiable functions. So, math optimization is really not about differentiability, optimization is essentially about non-differentiability about non-smoothness. And that is something you have to realize if you want to have your foundations quiet strong in this area.

Thank you very much.