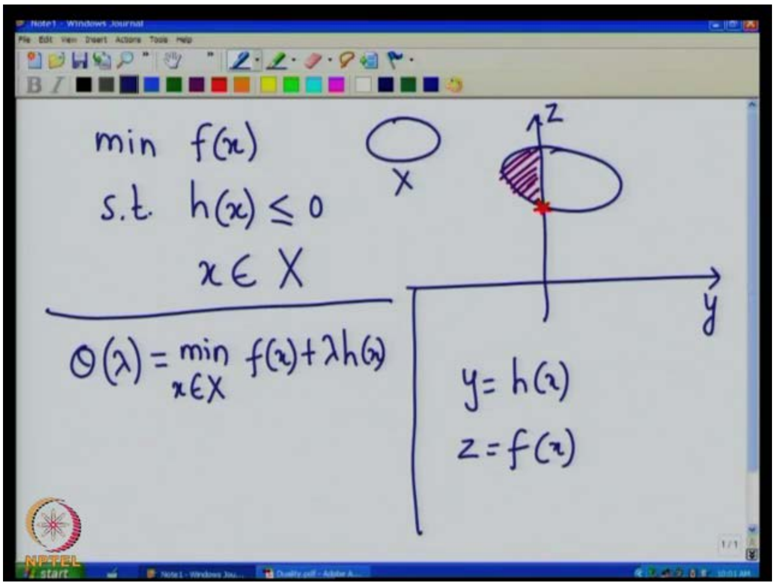
Numerical Optimization
Prof. Shirish K. Shevade
Department of Computer Science and Automation
Indian Institute of Science, Bangalore

Lecture - 29
Lagrangian Saddle Point and Wolfe Dual

(Refer Slide Time: 00:34)

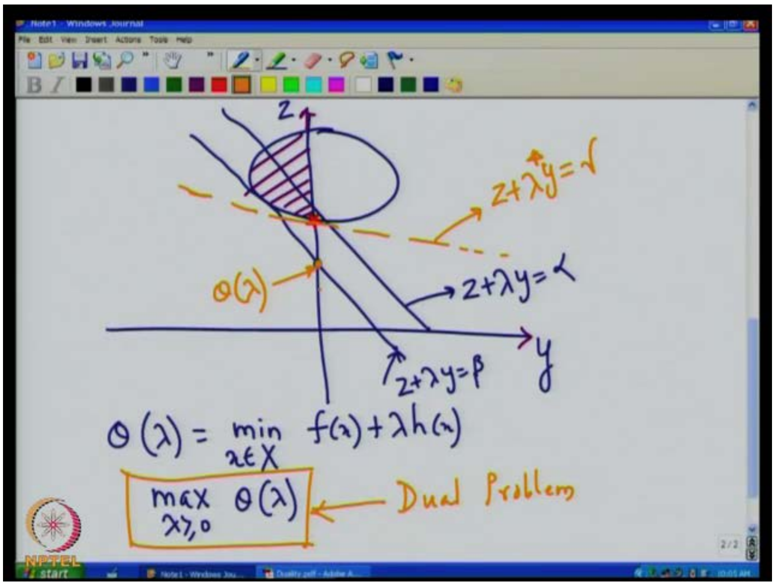


Hello, welcome back to this series of lectures on numerical optimization. In the last class we started discussing about the duality theory. And, in particular we looked at a simple problem where we want to minimize a function f(x) subject to a single inequality constraint, which is h(x) less than or equal to 0 and x belongs to a set X. So, we saw that in order to solve this problem; we transformed it to a another space which we called it as y z space; where y is equal to h(x), and z is equal to f(x).

So, every point in the set X is transformed to the y z space; and the corresponding set of points in the y z space, suppose the points in the on or the interior of this surface. Then, we are looking at the feasible points; and then the set of feasible points is the set where y is less than or equal to 0. So, this is the set that we are interested in it. And, what we want to do is that we want to find the minimum of this problem; and that minimum of this problem is the point which is. So, this is the optimal solution to the original problem.

And, then what we did was we defined a function called theta lambda; theta lambda is a function f(x) plus lambda h(x). So, theta lambda is minimize x belongs to X f(x) plus lambda h(x), and this f(x) plus lambda h(x) corresponds to a line in the y z space; the line z plus lambda y. Now, let us assume that lambda is nonnegative. So, it is a line with slope minus lambda.

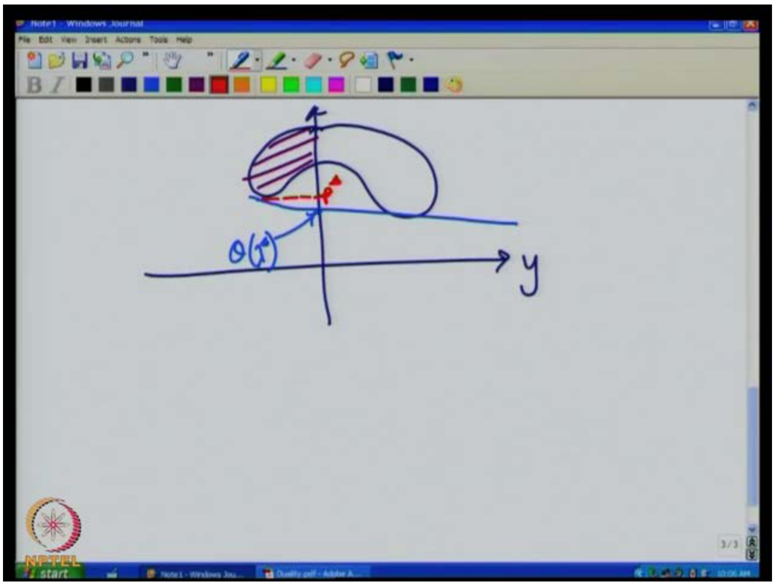
(Refer Slide Time: 03:13)



And, let us look at the picture again. So, z plus lambda y equal to constant is a line; in this space z plus lambda y is equal to alpha is a line with the slope minus lambda. Now, if you want to find out theta lambda. So, theta of lambda is nothing but minimum of f(x) plus lambda h(x); x belongs to X. So, for given lambda we are in we are interested in finding out the line; which supports this feasible set from below and also touches that feasible set. So, for a given lambda we are interested in getting a line which is having the same slope; and this is z plus lambda of y is equal to beta say. So, the so the minimum value is theta lambda. Now, we have already seen that optimal objective functional value is here.

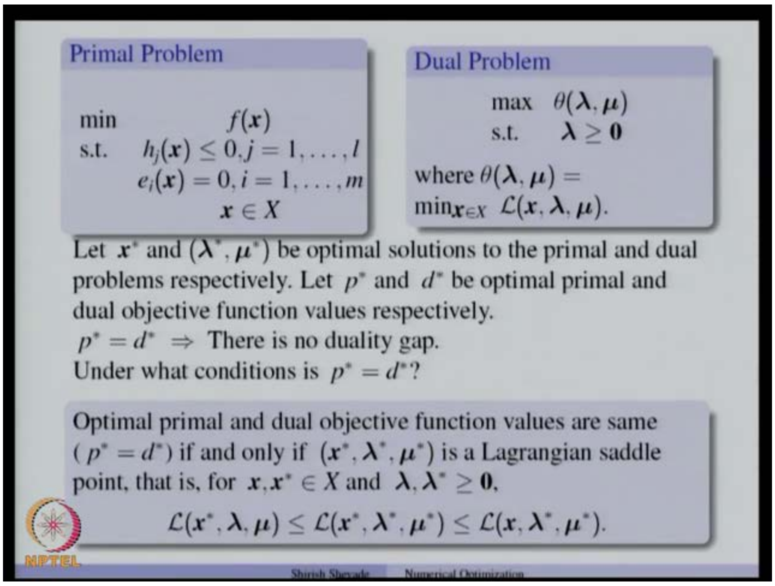
So, for a given lambda the theta lambda that we get is somewhere here; which is different from the optimal objective functional value. So, what we are in interested in is finding out that lambda which is nonnegative such that the theta of lambda is maximized. So, in other words we will be interested in looking at a line, and this is z plus lambda star y is equal to gap. So, this is the line with slope minus lambda star and that is obtained by solving this problem. And, this problem is called the dual problem. So, we are interested in finding out the dual problem. And, the reason for doing this is that many a times the dual problem is easier to solve. We will some examples in today is class about this, but it not always true that one has to always the dual problem. Because sometimes one can get a duality gap then one solves also given dual problem.

(Refer Slide Time: 06:37)



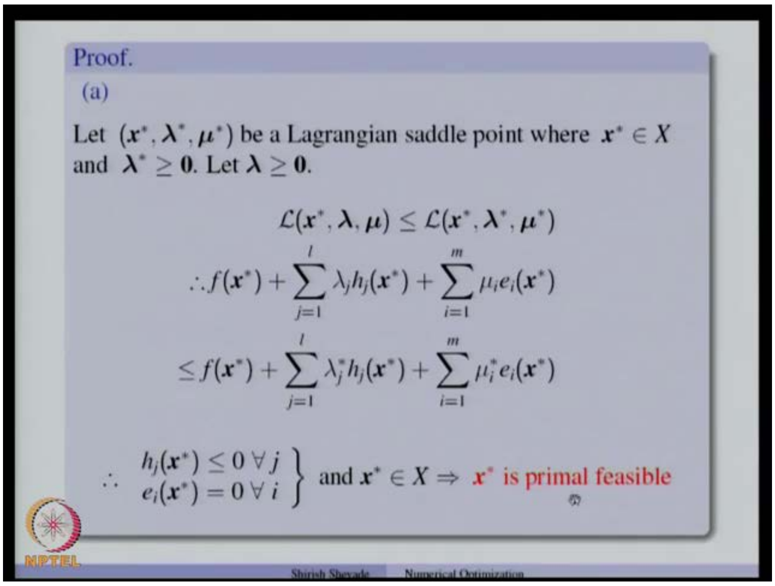
So, for example we saw another example where in the y z space. If the set g is something like this and the feasible region is. So, then we will see that the minimum of this is at this point. And, if we solve the dual problem so we would get a hyper plane like this or line like this. So, this is the theta of lambda star and this value is p star. So, you will see that there is some duality gap. And, therefore it is not always advisable to solve the dual problem. But in some cases it may be a good idea to solve the dual problem because it may be easier to solve. And, not only that the optimal dual value will be equal to optimal primal value. So, what are the conditions under which the duality gap is 0? So, we started looking at this question.

(Refer Slide Time: 08:02)



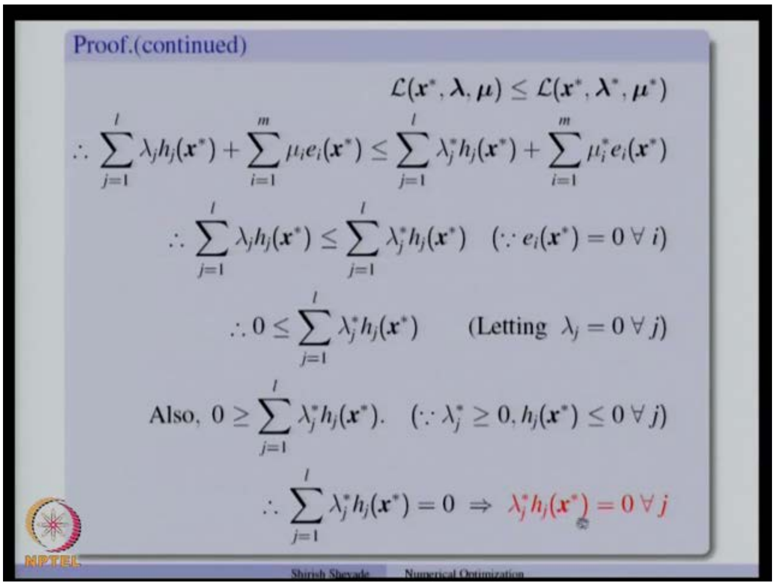
And, we saw the first part of this theorem which says that optimal primal and dual objective function values for this problems are same; if and only if x star lambda star mu star is a Lagrangian saddle point. That is the conditions for the saddle point holds; when L is the Lagrangian is the function under consideration.

(Refer Slide Time: 08:36)



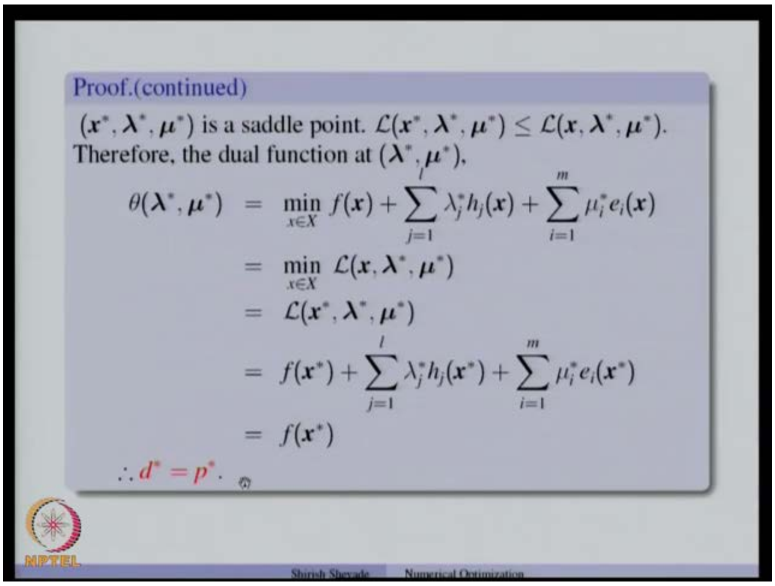
And, we saw this proof in the last class that if x star lambda star mu star is Lagrangian saddle point.

(Refer Slide Time: 08:46)



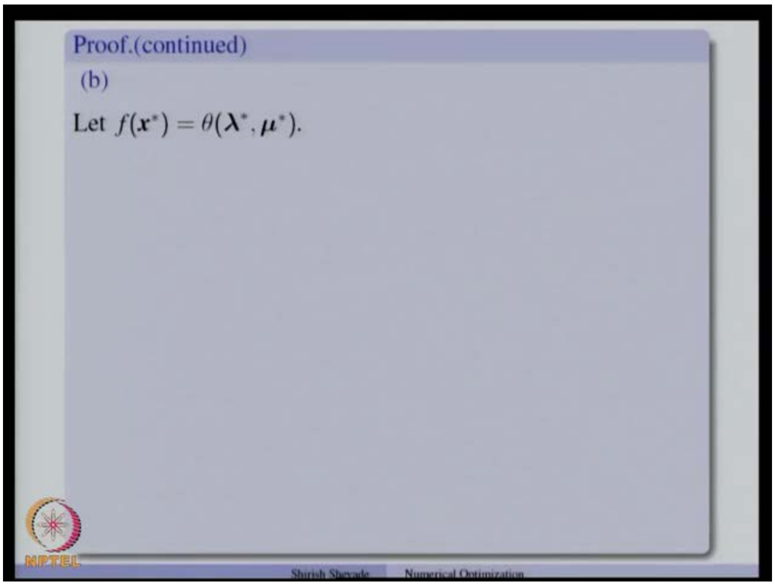
Then, we showed that x star is primal feasible and then lambda j star of x star is equal to 0.

(Refer Slide Time: 08:52)



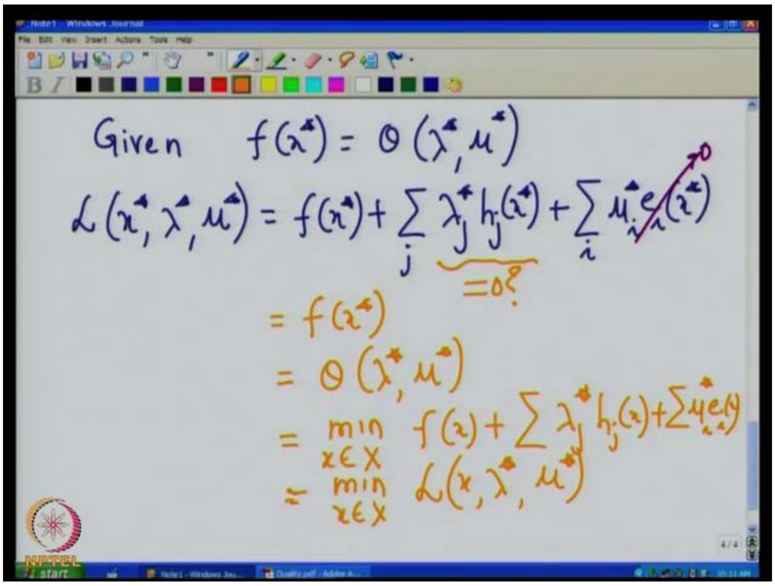
And, therefore what we get is theta of lambda star mu star is equal to f of x star. So, which means that there is no duality gap.

(Refer Slide Time: 09:01)



Now, we will see that the other part of the proof. Now, to prove this…

(Refer Slide Time: 09:12)

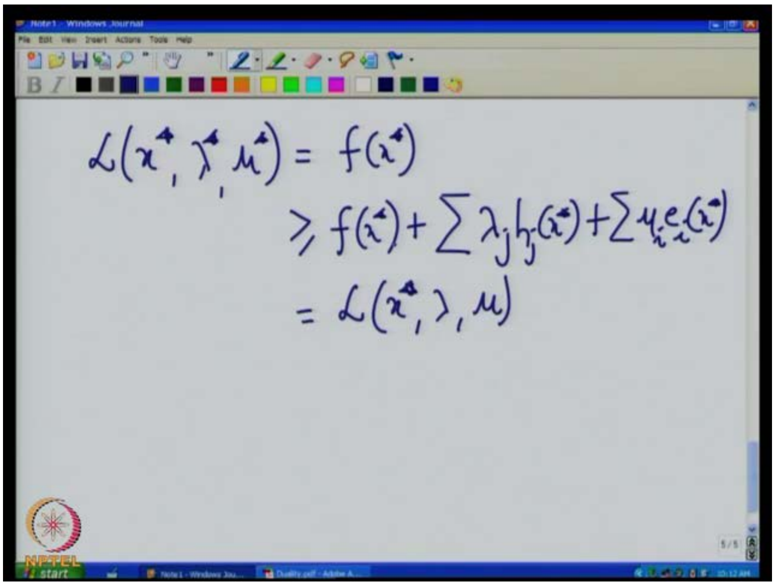


So, what is given to is that f(x) star is equal to theta of lambda star mu star. Of course, we assume that x star primal feasible such as satisfies all the primal constraints; and lambda star mu star satisfy all the dual constraints.

So, this is given to us. Now, we want to show that x star lambda star mu star is a Lagrangian saddle point. Now, if we look at L of x star lambda star mu star; that is nothing but f(x) star plus sigma lambda j star h j of x star plus sigma mu i star e i of x star. Now, since x star is primal feasible we can say that this quantity is 0. Now, what about this quantity? So, suppose this quantity is 0 we are not sure as of now but suppose this quantity is 0. Then, we can write this as f(x) star and what is given to us is theta of lambda star mu star. And, theta lambda star mu star is by definition minimize our x belongs to X f(x) plus sigma lambda j star h j(x) plus sigma mu i star e i(x).

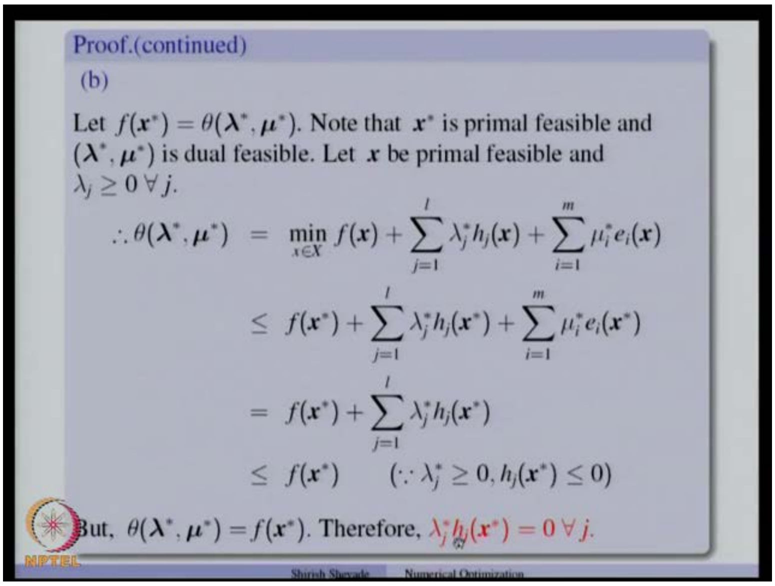
Now, this quantity is nothing but minimize x belongs to X the Lagarngian of x lambda star mu star. And, therefore we can say that L of x star lambda star mu star which is a minimum of L of x lambda star mu star over x belongs to X. So, therefore L of x lambda star mu star is less than or equal to L of x lambda star mu star for all feasible x. So, the question is that can we show that lambda j star of x star is equal to 0? So, if we show that then one part of the saddle point conditions are satisfied. Now, similarly once we show that lambda j star of x star is equal to 0; then what we get is L of x star lambda star mu star is equal to f of x star.

(Refer Slide Time: 12:28)



So, L of x star lambda star mu star is equal to f of x star. And, if we choose lambda’s to be nonnegative then we can write this as this is nothing but which is greater than or equal to f of x star plus sigma lambda j h j of x star plus sigma mu i e i of x star. Now, this quantity is 0, h j of x star is less than or equal to 0 and lambda’s are nonnegative. So, f of x star; so this quantity is always less than or equal to 0. And therefore f of x star is greater than or equal to this and this is nothing but L of x star lambda mu. So, L x star lambda star mu star is greater than or equal L x star lambda mu. So, both the saddle point conditions both the conditions related to the saddle point are satisfied. And, therefore we can say that under the 0 duality gap condition there exists a saddle point. So, the important thing we have to show is that lambda j star h j of x star is equal to 0; and that is what we will show now.

(Refer Slide Time: 14:00)

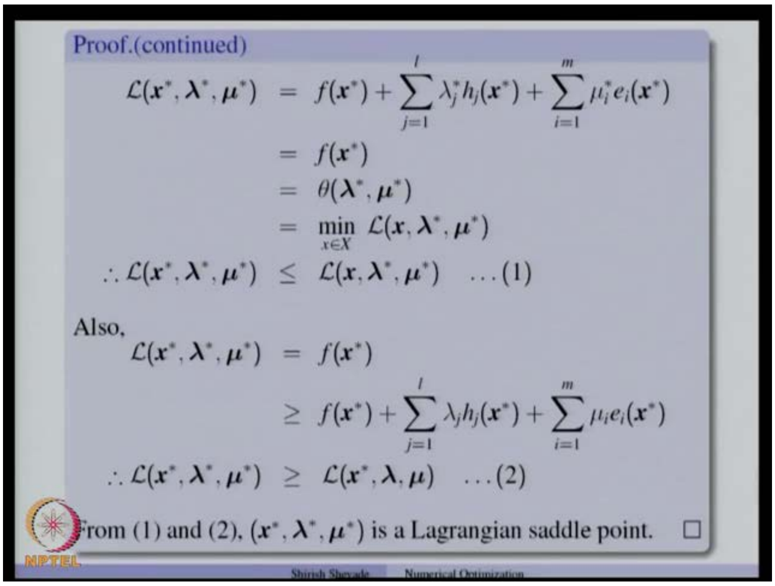


So, note that x star is primal feasible and lambda star mu star is dual feasible. And, also let us assume that x is a primal feasible point and lambda j's are nonnegative for all j. Now, theta of lambda star mu star is nothing but minimize effects subject to lambda j star; subject to x belongs to X minimize f(x) plus sigma lambda j star h j(x) plus sigma mu i star e i(x).

And, this is less than or equal to f of x star plus lambda j star h j of x star plus sigma mu i star e i of x star. Now, x star is primal feasible so this quantity is 0. And, lambda j star h j of x star is summed up for all the constraints and added to f of x star. So, this is what we get. Now, what is given to us f of x star is equal to theta of lambda star mu star. So, there is a strict equality in this case. So, now since lambda j star are greater than or equal to 0 because they are dual feasible; and h j of x star is less than or equal to 0. This quantity is less than or equal to 0 and therefore, this is less than or equal to f of x star. Because lambda j star is nonnegative h j of x star is less than or equal to 0 for all j. But we are given that theta of lambda star mu star is equal to f of x star. So, this quantity has to be 0.

And, therefore what we get is lambda j star h j of x star is equal to 0 for all j. And, this is what we wanted to use.

(Refer Slide Time: 16:04)

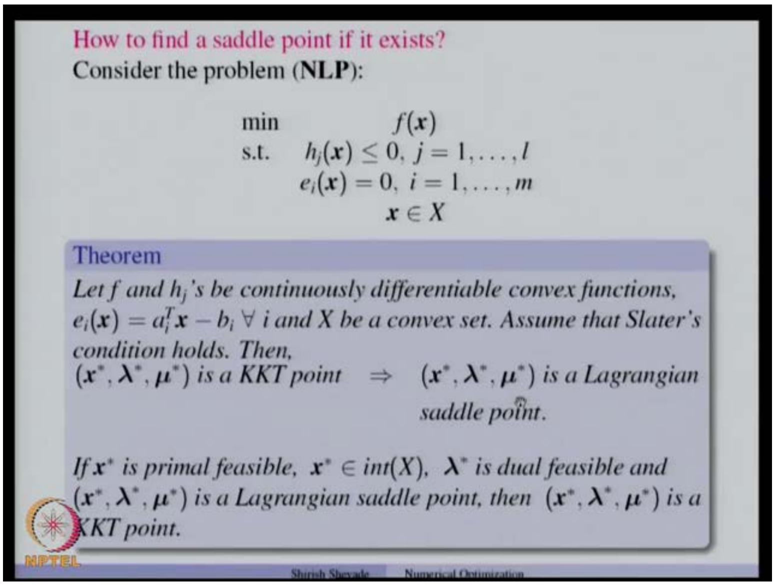


Now, one can write the Lagrangian of at x star lambda star mu star to be this sum and as I mentioned earlier that e i of x star is 0, lambda j x star h j of x star is 0. And, this quantity is nothing but f of x star. And, therefore we can write this as theta of lambda star mu star because there is no duality gap. So, theta f of x star is nothing but theta lambda star mu star.

And, by the definition of theta of lambda star mu star we have minimum theta of lambda star mu star to be minimum of x belongs to X, L of x star lambda star mu star. And, therefore what we have what is that L of x star lambda star mu star is less than are equal to L of x star lambda star mu star; that is our first condition. Now, we again look at L of x star lambda star mu star. Now, we have already shown that lambda j x star h j of x star is equal to 0 and we know that x star is primal feasible; so this quantity is also 0. So, L of x star lambda star mu star is nothing but f of x star that we have already shown. Now, that quantity is greater than or equal to f of x star plus this quantity because this quantity is anyway 0.

And, remember that lambda's are nonnegative h j of x star less than or equal to 0. So, f of x star is always greater than or equal to this quantity; and this quantity is nothing but L of x star lambda mu. So, we have got another condition we show that L of x star lambda star mu star is greater than or equal to L of x star lambda mu. Now, if you combine 1 and 2; we can see that x star lambda star mu star is a Lagrangian saddle point.

(Refer Slide Time: 17:54)



Now, so for we have shown that there is no duality gap if and only if there exists a Lagrangian saddle point. But how do we find out saddle a saddle point if such a point exists?

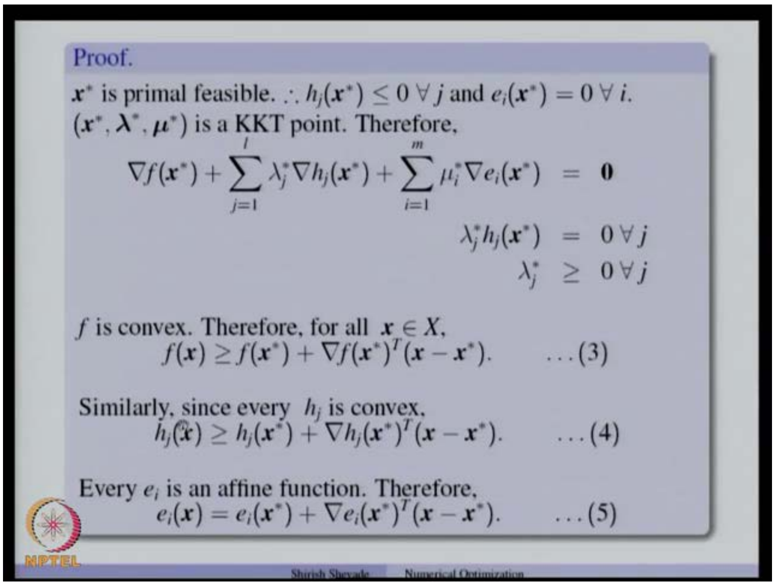
Because it is very difficult to check the saddle point conditions. So, is there any better way of ensuring the existence of saddle and finding it out. So, let us consider a general non-linear programming problem; there be a minimize effects subject to the set of inequality constraints and the set of equality constraints. Now, under certain convexity assumptions one can show that the KKT point is a Lagrangian saddle point. And, KKT points as we saw in one of the earlier classes there easy to check. So, here we have another important theorem which states if f and h j x are continuously differentiable convex functions.

So, the objective function is convex the functions associated with the inequality constraints of the type h j(x) less than or equal to 0. So, the function h j(x) are all convex and the function e i(x) are affine functions. So, in other words e i(x) is nothing but a i transpose x minus b i. And, assume that X is also convex set. So, we have a convex programming problem. So this is no longer a general non-linear program, but it is a convex programming problem. There we want to minimize a convex function subject to a convex set. And, it is clear that this set is convex set because it is a intersection of all convex sets.

Now, we also assume that latest condition holds. So, that means that it the constraint set has nonempty interior; the constraint set is convex and has nonempty interior. Then, the first result says that if x star lambda star is a KKT point; then x star lambda star mu star is a Lagrangian saddle point. And, the seconds result says that if x star is primal feasible and it belongs to the interior of the set X and lambda star is dual feasible. And, x star lambda star mu star is a Lagrangian saddle point then x star lambda star mu star is a KKT point.

So, under certain convexity assumptions and the Slater's constraint qualification condition a KKT point is a Lagrangian saddle point. And, therefore for such problems the duality gap does not exist. And, it is sometimes easy to solve the dual problem rather than the primal problem. So, we will first study the proof of this theorem and then later or we will see some examples related to this conditions. So, the first part of the proof is about proving that a KKT point under the convexity assumptions is a saddle point. So, let us look at that condition that proof.

(Refer Slide Time: 21:36)



So, we assume that x star is primal feasible. Therefore, satisfies all the inequality and equality constraints. And, of course, x x star belongs to the set X. Now, we are given that x star lambda star mu star is a KKT point. So, it has to satisfy 3 conditions along with the feasibility conditions. So, the first condition is that the gradient of the Lagrangian should vanish at x star lambda star mu star the gradient which is evaluated with respect to x. So, gradient of the Lagrangian is gradient f of x star plus sigma lambda j star gradient h j of x star plus sigma mu i star gradient ei x star.

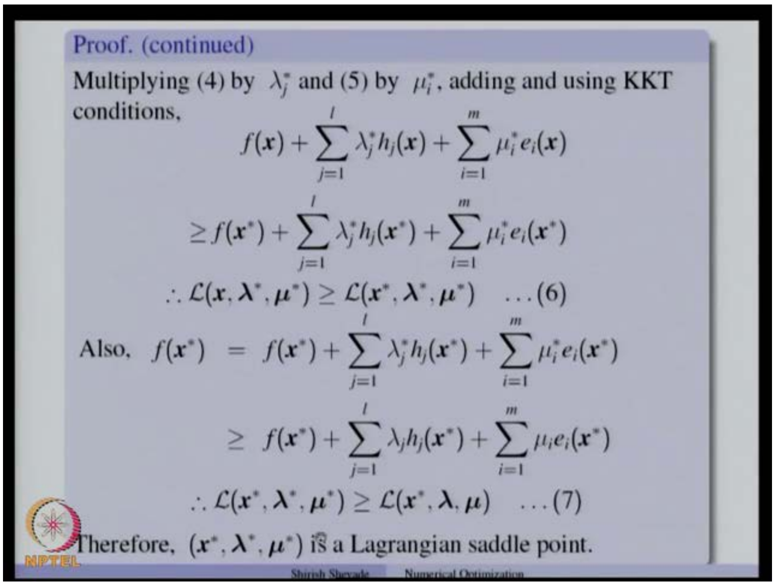
And, that should be equal to 0. And, second condition is about the complimentary slackness condition. So, lambda j star h j of x star equal to 0 for all j. And, other condition is that all the Lagrangian multipliers lambda j star are nonnegative. Now, these conditions along with the feasibility ensure that the KKT's conditions are satisfies by x star lambda star mu star. Of course, we are assuming that x star also belongs to set X; that is part of the primal feasibility. Now, we make use of the convexity of the functions f h and the affineness of the function e i. And, use those conditions to derive that x star lambda star mu star is indeed a saddle point.

So, let us look at the function f. Now, f being a convex function we have already seen in one of the earlier classes; that an affine approximation of a convex function at any point does not over estimate the function. So, since f is convex we saw this result earlier that f(x) for every x in the feasible set X f(x) is greater than or equal f of x star plus gradient f of x star transpose x minus x star. So, the right side is an affine approximation of the function at x star. And, we know that that does not over estimate the function. Now, the same result can be applied to the convex functions related to the inequality constraint. That is the function h j x less than or equal to the constraints h j x less than or equal to 0 and the associated functions are h j x. So, for every h j x which is a x which is a convex function h j(x) is again greater than or equal to h j of x star plus gradient h j of x star transpose x minus x star. Now, e as affine function. Therefore, we can write ei(x) equal to ei of x star plus gradient ei of x star transpose ei of x minus x star.

Now, remember that we want to find out the conditions associated with the Lagrangian saddle point. So, the Lagrangian saddle point is nothing but I am sorry the Lagrangian function L x lambda mu is nothing but f x plus sigma lambda j x h j x plus sigma mu i e i x. So, in order to get towards those conditions what we need to is that we need to multiply this equation; equation 4 by lambda j and equation 5 by mu i. Then, sum them up and the same exercise has to be done in the right side. Since, lambda j's are nonnegative; the inequality sign inequality direction does not change. So, even if you multiply lambda j throughout the inequality same inequality holds. And, this is a equality and mu i ei is an affine function and mu i is unrestricted in sign. So, we can multiply mu i by mu i both sides and then we add up the left hand side as well as the right hand side.

Now, when we add the right side in the first term what we get is f of x star plus sigma j h j of x star plus sigma mu i e i of x star. Now, if we multiply by lambda star and mu star. Then, what we get is gradient f of x star plus sigma lambda j star gradient h j of x star plus sigma mu i star gradient e i of x star; and that quantity is 0 because of this condition. So, if you multiply second by lambda j star and equation 4 by lambda j star and equation 5 by mu i star. Then, this quantity the second terms vanish.

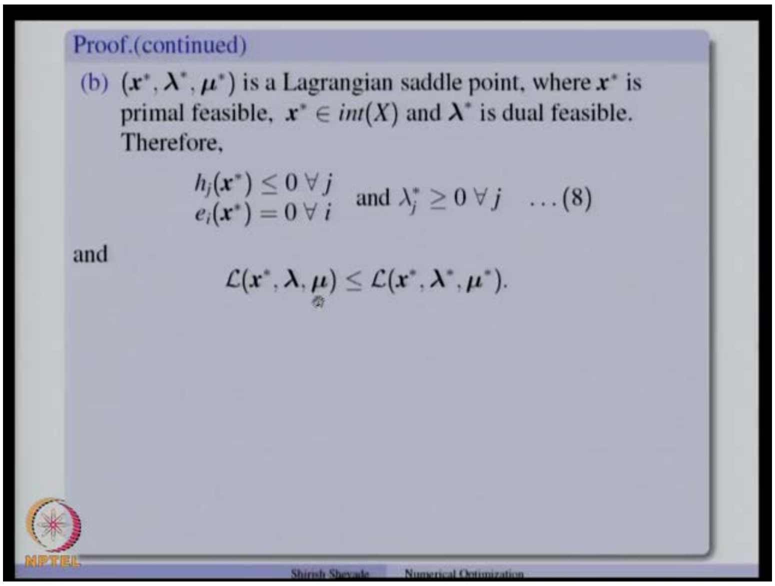
(Refer Slide Time: 27:24)



And, then what we are left with is the following f(x) plus sigma lambda j star h j of x star plus sigma mu i e i of x star. So, that is the first part and that is greater than or equal to f of x star plus sigma lambda j star h j of x star plus sigma mu i star e i of x star. And, then the quantity corresponding to the second term vanishes because of this conditions. Now, we can use this condition further to remove this quantity and the feasibility of x star to remove this quantity. So, what we are left with is L f x star lambda star mu star is greater than or equal to mu star. Now, f x star is nothing but f of x star plus sigma lambda j star h j of x star plus mu i star e i star because this is 0. And, the complimentary slackness condition give this quantity as 0. And, this quantity is greater than or equal to f(x) plus lambda j h j of x star plus sigma mu i e i of x star.

And, therefore what we have is L x star lambda star mu star is greater than or equal to L x star mu. And, using these 2 conditions we can say that x star lambda star mu star is a Lagrangian saddle point. So, this was the first part of the theorem. Now, the second part says that if we are given a Lagrangian saddle point in the x star is in interior then the KKT conditions are satisfied. So, in other words we have to show that this hold.

(Refer Slide Time: 29:11)

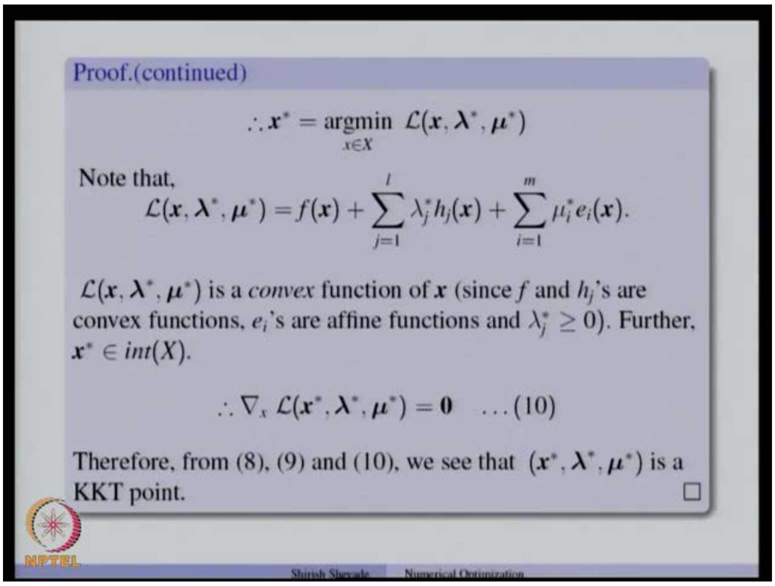


So, we assume that x star lambda star mu star is saddle point and x star is primal feasible. So, that means h j x star less than or equal to 0, e i x is equal to 0, for all j's and i then x star belongs to interior of the set X.

This is the assumption that we are making and lambda star is dual feasible. Now, x star is primal feasible. So, the feasibility conditions for the primal are satisfied lambda star is dual feasible. So, feasibility conditions for the dual are satisfied. Remember that the Lagrangian multipliers corresponding to the equality constraints are unrestricted in sign. Now, x star lambda star mu star is also Lagrangian saddle point. And, therefore one of the conditions of the saddle point is L x star lambda mu is less than or equal to l x star lambda star mu star. Now, if you expand this so, the and cancelling the f x star term; what we get is something like this. Now, using the same logic that we use in the previous proof we will we can show that lambda h j (x) star equal to 0. And, therefore 1 of the KKT conditions is satisfied.

So, lambda j star is greater than or equal to 0 because it is dual feasible; primal feasibility is satisfied x star belongs to the interior (X). And, we also shown the complimentary slackness condition. Now, L x star lambda star mu star is less than or equal to L x star lambda star mu star because of the saddle point conditions. And, therefore x star is obtained by minimizing the right side function L x star lambda star mu star with respect to x belongs to X. So, the x star is nothing but argmin mean of x and the objective function is the Lagrangian where lambda star mu star are kept fixed.

(Refer Slide Time: 31:25)

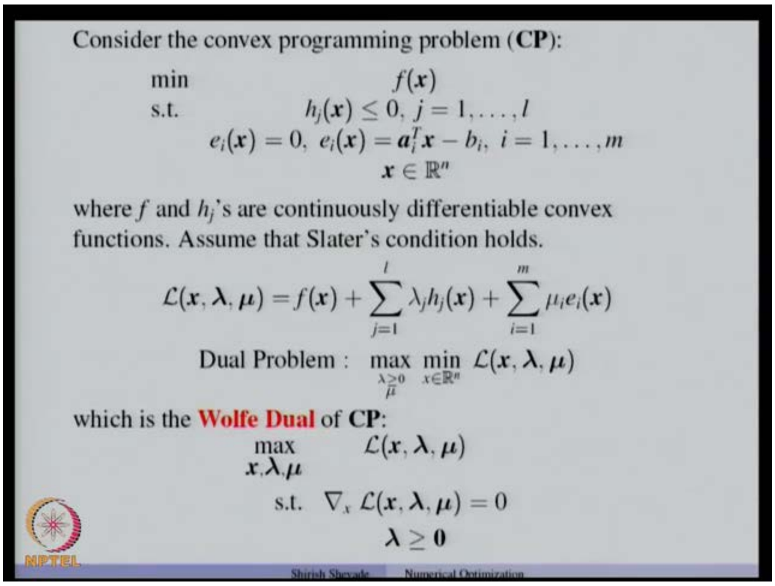


Now, L x star lambda star mu star this quantity if we expand it so, we have f (x) plus sigma lambda j star h j (x) plus sigma mu i star e i (x). Now, f (x) is a convex function, h j (x) is a convex function, lambda j star are nonnegative, e i (x) is an affine function. That is the problem that we are considering and mu stars are unrestricted in sign. So, since e i (x) is affine function f(x) is convex and h j(x) is also a convex function lambda j star nonnegative; this function is a convex function. So, L x star lambda star mu star is a convex function.

And, x star is the minimum of that convex function over the set x and we have assume that x star lies in the interior. So, the minimum of a convex function lies in the interior means that the gradient of the convex function L should vanish at x star. So, in other words gradient of L evaluated with respect to x and evaluated at x star lambda star mu star is equal to 0. And, this was 1 of the KKT conditions that we wanted to prove because we had already proven the complimentary slackness conditions. And, we knew that lambda j star is we assume that lambda j stars are dual feasible. So, they are nonnegative and x star is primal feasible. And, therefore all the conditions KKT conditions are satisfied. And, therefore x star lambda star mu star is a KKT point.

And, this was possible because of the assumption that x star belongs to the interior of the set x and L is the convex function over the convex set x. So, this theorem has important implication. And, that this theorem gives us an idea about how to write the dual problem of a given convex programming problem?

(Refer Slide Time: 34:01)



So, let us consider a convex programming problem; where f is a convex function. The objective function is convex the constraint the inequality constraints are h j (x) less than or equal to 0. And, the function h j (x) corresponding to these inequality constraints is also a convex function; the equality constraints are given by a i transpose x minus b i equal to 0. And, let us consider the set x to be entire set r n.

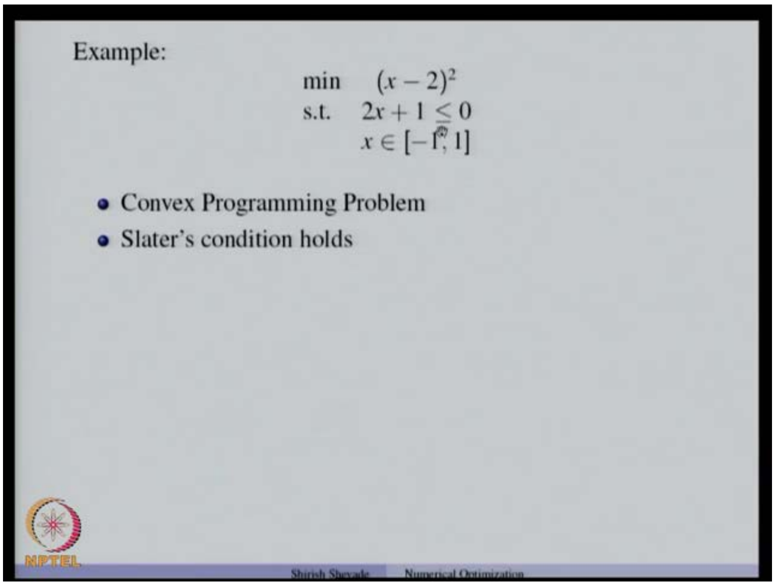
Let us also assume the differentiability of f and h and they are also convex. Let us assume that the constraint set finally that we get is also such that it has nonempty interior. So, which means that Slater's condition holds. So, the Lagrangian which is a convex function (x) is f (x) plus sigma lambda h j (x) plus sigma mu i e (x). And, the dual problem is a max min problem; where we minimize the Lagrangian with respect to x belongs to r n. And, then that function is maximize with respect to lambda's which are nonnegative and mu which are unrestricted in sign. Now, if you look at this problem minimize x belongs to r n L x lambda mu.

Since, L is a convex function which we saw earlier. So, the minimum at the minimum of this gradient of the Lagrangian with respect to x should vanish. Therefore, the dual problem becomes maximum maximize L x lambda mu subject to the constraint that the gradient of the Lagrangian evaluated with respect to x vanishes; and lambda's are dual feasible. So, this problem is called the Wolfe dual problem. Wolfe was the first person to show that for convex programming problems under this constraint qualification condition. 1 can write the dual in a simple form which is maximize the Lagrangian subject to the constraint at the gradient of the Lagrangian with respect to the primal variable vanishes.

And, the lambda's the Lagrangian multipliers are associated with the inequality constraint of the type h j (x) less than or equal to 0 are nonnegative. Now, these 2 problems the original CP and the dual problem CP under the assumption of constraint qualification or Slater's condition have the same optimal solution. So, in other words these 2 problems are equivalent to each other. So, one can either solve this problem or this problem and 1 would get the same optimal solution or in other words there is no duality gap.

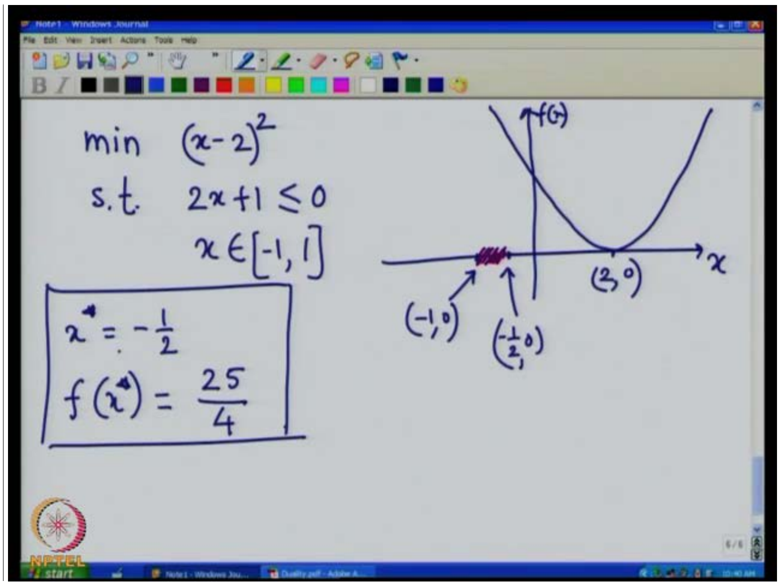
But then 1 may wonder that this was the problem with respect to x; here we have introduced more variables lambda and mu. And, therefore this problem may be more difficult to solve compared to the original problem. But as we will see some examples you will realize that the dual problem is many a times here the Wolfe dual problem is many a times easier to solve compared to the primal problem. Although there is no rule that every time one has to solve the dual problem; but sometimes the it becomes easier to solve the dual problem than the primal. Because some of these variables get eliminated and some of the constraints also get eliminated. And, we are left with the simple problem than the original primal problem. So, we will look at some examples.

(Refer Slide Time: 38:15)



So, this is a an example to minimize a 1 dimensional function x minus 2 square subject to the constraint that 2 x plus 1 less than or equal to 0 and x belongs to minus 1 to 1. So, this interval minus 1 to 1 becomes our constraint set x which is the constraint h(x) less than or equal to 0 and this is the objective function. Now, you will see that the objective function is the convex function. This constraint set is a convex set and this constraint set is also convex set. So, the intersection of the convex sets is also convex set. So, we have a convex programming problem and this also satisfies Slater's condition. Now, let us solve this problem.

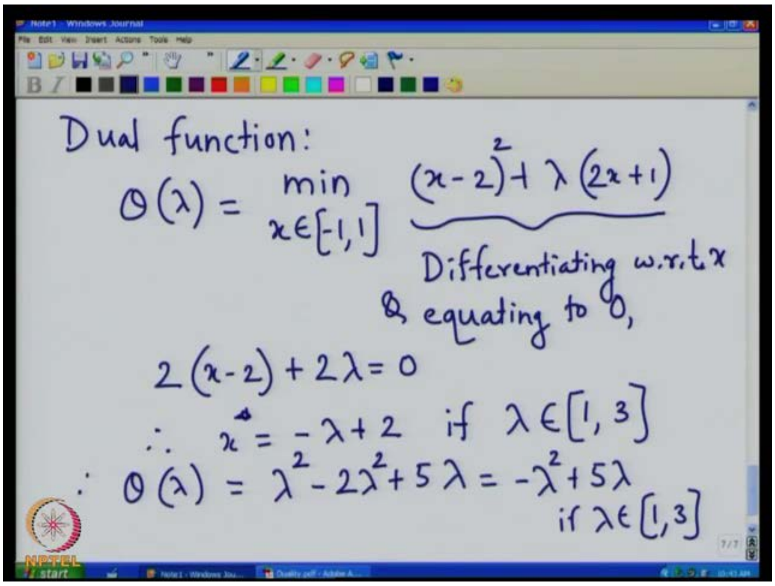
(Refer Slide Time 39:29)



So, we have the problem minimize x minus 2 square subject to 2 x plus 1 less than or equal to 0 and x belongs to minus 1 to 1. Now, let us see the graph of the function. So, the objective function is a function like this which has 0 at this point; the function value 0 at this point. Now, let us look at the constraints. So, the first constraint says that x is less than or equal to minus half. And, that would be somewhere here and x belongs to minus 1 to 1. So, in other words we are interested in this constraint set which is the interval from minus 1 to half.

So, we are interested in finding the minimum of this function over this interval which is shown here by shaded lines. Now, it is clear from this figure that minimum of the function will occur at x star equal to minus half. And, the corresponding value of the objective function f(x) star will be minus half minus 2. So, which is equal to 25 by 4. So, this our primal objective function and this is the solution to this problem. Now, let us write down the dual problem. And, see how the dual can be solved?

(Refer Slide Time 42:09)

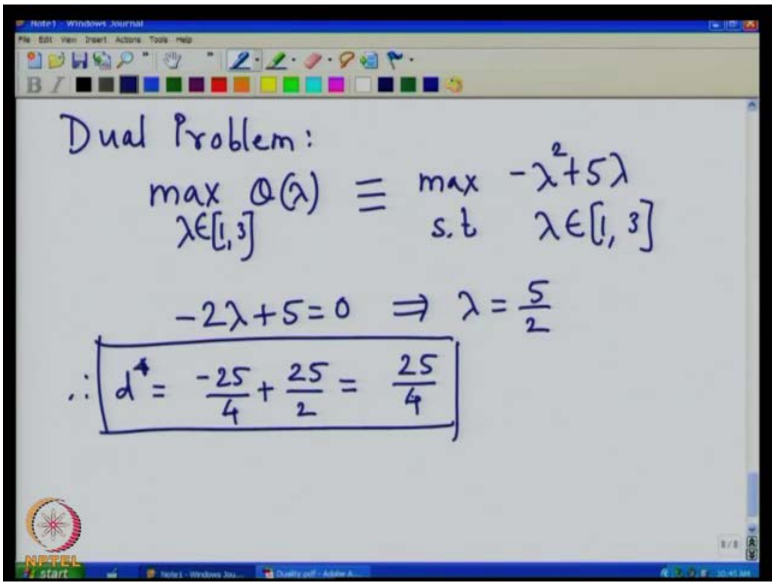


So, let us look at the dual function. So, dual function is a function theta lambda which minimizes x which is in the interval minus 1 to 1 x minus 2 square. That is the objective function and lambda into the constraint and the constraint this 2 x plus 1. Now, this function is a quadratic function in x. In fact it is convex because the coefficient of x square is positive here. So, we differentiate this function with respect to x and equate it to 0. So, differentiating with respect to x and equating to 0, what we get is 2 into x minus

2 plus 2 lambda is equal to 0 and from that we can get x star. And, therefore x star to be minus lambda plus 2.

Now, remember that x has to be in the interval minus 1 to 1. So, in order that x has to be in the minus interval minus 1 to 1. So, which means that x star also has to be in this interval we have to set a range for lambda. So, the lambda should belong to since x star belongs to minus 1 minus 1 to 1 lambda should belong to the interval 1 to 3. So, this will make sure that x star also lies in this interval. And, if we substitute this x star here what we get is so therefore, theta lambda theta lambda will be equal to substitute this value here. So, what we get is lambda square and then minus 2 lambda square plus 5 lambda. And, this is nothing but minus lambda square plus 5 lambda. If lambda belongs to the interval 1 to 3 so we have got theta lambda. So, now the dual problem.

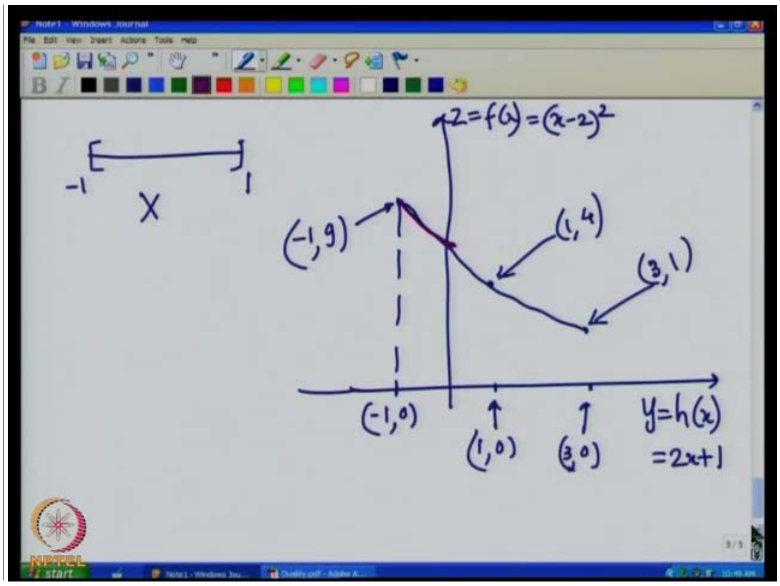
(Refer Slide Time 45:25)



The dual problem is maximize theta lambda with respect to lambda which is in the interval 1 to 3, and this nothing but maximize minus lambda square plus 5 lambda; subject to lambda belongs to 1 to 3. Now, this is a concave function. So, again we differentiate this function with respect to lambda equate it to 0. And, what we get is minus 2 lambda plus 5 equal to 0 and which gives lambda is equal to 5 by 2.

So, therefore, the dual objective function value d star which should be we plug-in this 5 by 2 in the objective function. And, what we get is minus 25 by 4 plus 25 by 2 and which is nothing but 25 by 4. Now, if we compare this dual objective function value with the primal objective function value that we got which is nothing but f (x) star. We see that this dual objective function value and primal objective function values are the same. Now, with respect to the y z space what happens? So, let us look at that.

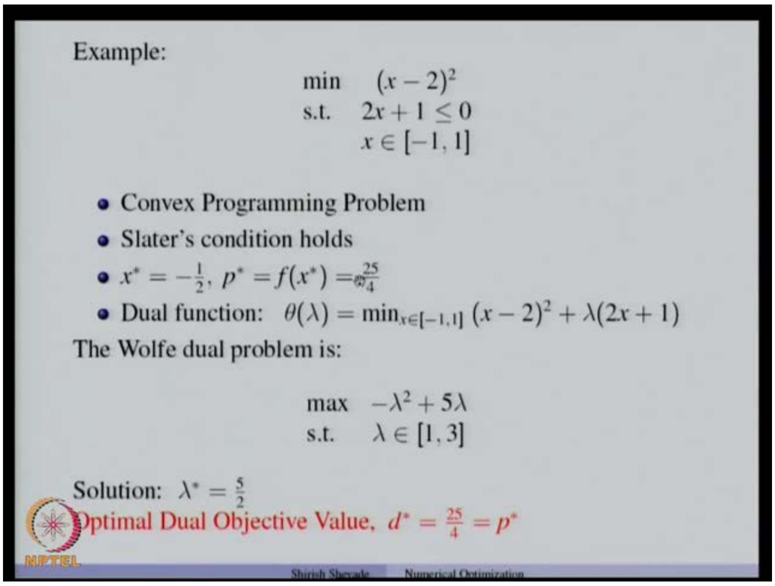
(Refer Slide Time 47:21)



So, this is the y z space that we have. So, z is nothing but f (x) and y is nothing but h (x). Now, h(x) is nothing but 2 x plus 1 and f(x) is nothing but the given objective function which is x minus 2 square. So, set x is the interval minus 1 to 1. Now, this interval we take each value of x from this interval and map it to the y z space. So, we get a curve like this. So, in other words when x equal to minus 1. So, this that is the quantity which is here minus 1. And, then when x equal to minus 1 what we get is 9. So, this point which corresponds to (No Audio From 49:01 to 49:09) so this point is minus 1, 9. When x equal to 0 y is 1. So, this is this quantity this point in the y z space; and when x is 0 this is 4. So, this is 1 4; and when x is 1 y is 3 and z is 1. So, we get a function like this and what we are interested in is y less than or equal to 0.

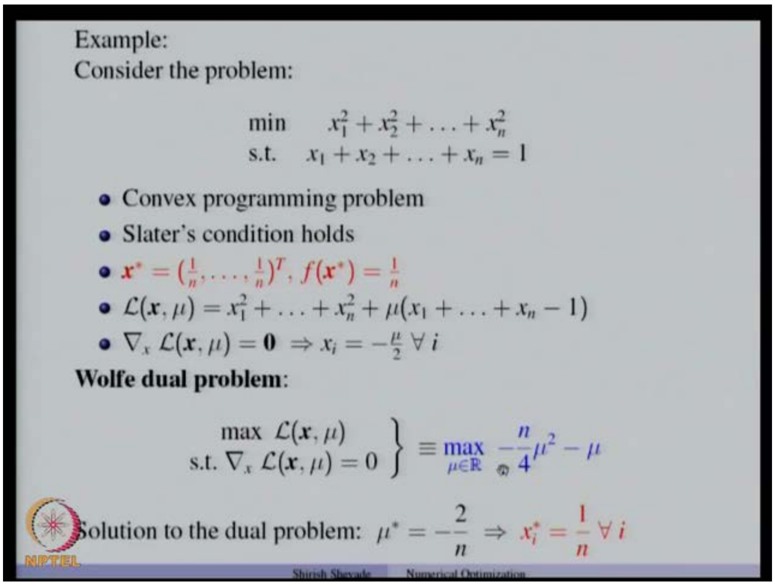
So, we are interested in only this part of the function. Now, if we take a line which supports this feasible set from below. Then, that line will have slope which is minus 5 by 2. And, moreover as we saw earlier that the optimal primal and dual objective function values are same. And, therefore there is no duality gap.

(Refer Slide Time 50:57)



So, the dual function is minimize x minus 2 square plus lambda 2 x plus 1. And, therefore the Wolfe dual problem is maximize minus lambda square plus 5 lambda and lambda belongs to the close interval 1 to 3. And, therefore by solving this we get lambda star to be 5 by 2. And, the optimal objective function and for the dual problem; which is 25 by 4 is same as the optimal objective function value for the primal problem. Now, look at another example.

(Refer Slide Time 51:33)



So, we have a problem to minimize norm x square subject to the constraint that x 1 plus x 2 up to x n equal to 1. This is a convex programming problem because the objective function is convex the constraint set it is affine. So, Slater’s conditions are automatically satisfied. Now, we can see that the solution to the this problem is that all values of the coordinates are same and they are 1 by n. And, they are optimal objective function value is 1 by n.

Now, let us look at the Lagrangian. So, Lagrangian is nothing but the objective function plus mu times the constraint; this mu is associated with this equality constraint. Now, we can write the Wolfe dual of this problem; so one of the conditions that is used in the Wolfe dual is that the gradient of the Lagrangian is respect to the primal variable 0. And, this implies that x i equal to minus mu by 2. Now, if we plug-in x i equal to mu by 2 in this Lagrangian. Then, we can write the dual problem Wolfe dual problem as maximize the Lagrangian with respect to the constraint that the gradient of the Lagrangian with respect to x vanishes.

Now, note that mu is associated with the equality constraint. Therefore, there are no sign restrictions on mu. Now, by plugging-in in this value of x i which is equal to mu minus mu by 2; in this Lagrangian we get a function which is minus n by 4 mu square minus mu. Now, what is the advantage of writing the Wolfe dual? So, you will see that the Wolfe dual problem is a 1 dimensional optimization problem in this case. In fact it is a unconstraint optimization problem. So, the original n dimensional constraint optimization problem which had some nice properties that like it is a convex programming problem and Slater’s condition are satisfied. So, such a n dimensional optimization problem we were able to convert it to a 1 dimensional unconstraint optimization problem.

And, not only is that because of the condition because of the results that we saw earlier for such problems the duality gap 0. Because the KKT point is a Lagrangian saddle point and therefore, the duality gap is 0. We can solve this problem and get a solution to this problem. Now, this is a unconstraint optimization problem in mu. So, if we differentiate the objective function is convex in terms of mu. So, if we differentiate this objective function with respect to mu and equate it to 0; what we get is mu star to be minus 2 by n. Now, if if you plug-in this value of mu star in this what we get is x star to be 1 by n.

And, that is same as this. So, this is a very important example in the sense that it gives us some idea about the advantage of solving the dual problem. In some practical situations the original problem could be an infinite dimension problem; where the norm is defined in that infinite dimensional space. And, solving an infinite dimensional problem is very difficult but in that case if we can convert the problem into a dual problem. It may so happen like in this case that the dual problem is still with respect to number of variables in dual problem is still very small and manageable. And, in that process we will not lose anything because the optimal primal and dual objective function values are same. Because of the existence of Lagrangian saddle point or for the convex programming problems the KKT point and the Lagrangian saddle points are the same. If the Slater's condition qualification holds.

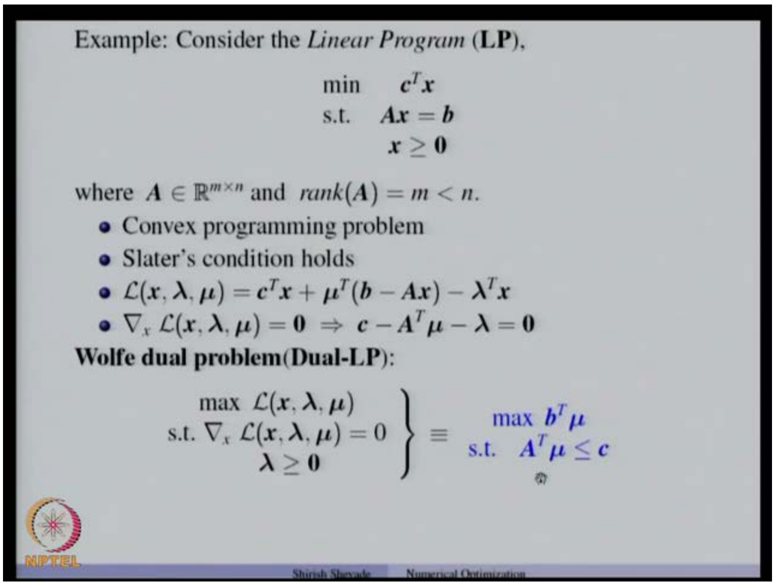
So, from this example we can see that there is some advantage in solving the in writing the dual problem of the convex programming problem. Of course, most of this results for convex programming problem. For general non-linear programming problem but dual problem may not be equal to the primal problem and there could be a duality gap. And, therefore, when 1 applies duality ideas 1 has to be careful about the convexity and the constraint qualification conditions that we discussed today. So, we will see some examples in the next class. In the last class we were discussing about Wolfe dual. And, in particular we consider this problem that we want to minimize x 1 square plus x 2 square up to x n square subject to the constraint that x 1 plus x 2 of x equal to 1.

So, this was the problem that we were considering last time. And, we saw that this is also a convex programming problem and Slater's condition holds good. And, we usually see that the solution is a point where every coordinate is 1 by n and the optimal objective function value is 1 by n. So, to write the Wolfe dual we need a Lagrangian. So, which is defined as the objective function plus mu time the constraint function; and the gradient of the Lagrangian with respect to x equal to 0 implies x i equal to minus mu by 2. So, plugging-in in this value of x i in this Lagrangian we can write the Wolfe dual as maximize Lagrangian function; subject to the constraint that the gradient Lagrangian with respect to x is 0.

And, using this value of x i which is minus mu by 2 we can write the Wolfe dual as max of minus m by 4 mu square minus mu. So, you will see that the dual problem here is unconstraint problem in terms of only 1 variable. the primal problem had n variables and

1 equality constraint the dual problem has a only 1 variable and it is a unconstraint problem. This problem is much easier to solve and as we saw last time mu star is minus 2 by n and that gives x i star to be 1 by n which is same as what we saw earlier. So, many a times the dual problem is easier to solve than the primal problem. Also note that if we have a very large value of n then solving this primal will be difficult. Instead it will be easy to solve this dual problem. Now, let us look at another program.

(Refer Slide Time 59:40)



This program is popularly known as linear program. We will discuss about linear problem sometime later but the problem formulation is like this. We want to minimize a function c transpose x; subject to the Subject to the constraint x equal to b x greater than or equal to 0; where A is m by n matrix, and the rank of the matrix A is m which is less than n. The reason why this is called the linear program is that the objective function is linear and the constraints are linear in terms of the variables. Now, we can clearly see that this constraint set is a convex set and the objective function is also a convex function. So, minimizing a convex function convex constraint set is a convex programming problem.

Also, let us assume that the feasible set that we have is a nonempty set and moreover it is also a non singleton set. So, which means that there exist a point which lies in the interior of the constraint set. So, that implies that the Slater is condition holds. Now, with these 2 properties that the linear program is convex linear problem and assuming that the Slater is condition holds; we can write the Wolfe dual of this problem. So, to write the Wolfe dual we need a Lagrangian. Now, there are some equality constraints and some inequality constraints. So, there will be a Lagrangian multipliers corresponding to these equality constraints which we are going to call them as mu. And, the Lagrangian multipliers corresponding to the inequality constraints; we are going to call them as lambda.

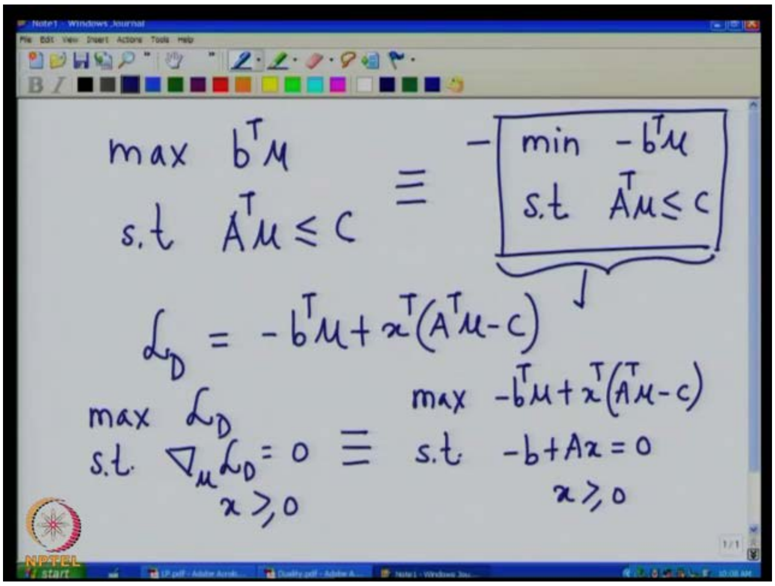
So, the Lagrangian is the objective function plus mu transpose b minus Ax minus lambda transpose x. And, the gradient of the Lagrangian with respect to x equal to 0 implies that c minus A transpose mu minus lambda equal to 0. We need this condition when we write the Wolfe dual. So, that is why we have calculated beforehand, and then we are in a position to write the Wolfe dual. So, Wolfe dual of this l p the formulation which is given here is maximize the Lagrangian subject to the constraint; that the gradient of the Lagrangian with respect to x is 0. And, then the Lagrangian multipliers corresponding to the inequality constraints are non negative.

So, if you consider this Lagrangian and consider the fact that the we need to satisfy this constraint that the gradient of the Lagrangian should vanish. So, c minus a transpose will be lambda minus equal to 0. So, let us substitute this in the Lagrangian expression and what we will get is only the term involving mu and b. Because the other terms c transpose x minus mu transpose Ax minus lambda transpose Ax will get cancelled because of this condition. So, the Wolfe dual of this problem is maximize b transpose mu subject to the constraint; that A transpose mu less than or equal to c.

So, this condition is arrived at by using this expression. So, we have this constraint that the gradient of the Lagrangian should vanish. So, which means that a minus a transpose minus lambda equal to 0. Therefore, c minus mu equal to lambda and lambda is are nonnegative. So, c minus a transpose mu is greater than or equal to 0 or in other words a transpose mu is less than or equal to c. Now, this Wolfe dual you just only the variable mu, the variable lambda does not appear anywhere in this dual.

So, you can see that is in the original problem the matrix of A size is m by n. So, the number of mu is that we here will be equal to m. So, this becomes a m dimensional optimization problem. And, moreover the variables mu is are unconstraint; unlike the original primal variables are which a non negativity constraint in them. Also, you will see that there are no terms in the dual which involve lambda. So, the number of constraints in this problem, the number equality constraints in this original linear program is equal to the number of variables in the Wolfe dual. Therefore, if m is the very much less than n; then it may be a good idea to solve this problem instead of the original linear program. Now, let us take this dual problem and see what happens if we write the dual of this problem? So, let us take.

(Refer Slide Time: 1:05:40)



So, we have maximize b transpose mu subject to the constraint at A transpose mu less than or equal to c. So, we want to write down the dual of this problem. So, this problem let us first bring it to the minimization form. And this problem is nothing but minimize minus b transpose mu subject to A transpose mu less than or equal to c. So, let us 1st write down the dual of the problem which is given in this box and then use the negative sign. So, let us look at only this problem 1st**.** Now, to write the dual we need to write the Lagrangian. So, let us first write down the Lagrangian of this we will call it as LD because d stands for the dual.

So, Lagrangian of this dual problem that we are going to write and that is nothing but minus b transpose mu plus x transpose A transpose mu minus c this is the Lagrangian. And note that this is also a linear programming problem. So, it is also a convex programming problem and let us assumes that the Slater is constraint qualification holds. So, we can write the Wolfe dual of this. So, Wolfe dual of this problem will be maximize

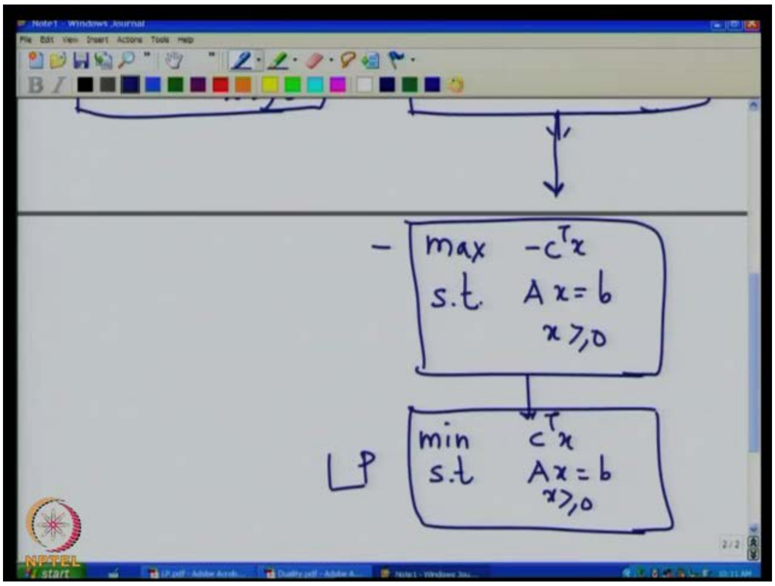
LD subject to the constraint that gradient of LD with respect to the primal variable which is mu is equal to 0; and the Lagrangian multipliers corresponding to the inequality constraints are nonnegative.

So, here x are the Lagrangian multipliers. So, x greater than or equal to 0 or which is same as maximize minus b transpose mu plus x transpose A transpose mu minus c constraint qualification holds. So, we can write the Wolfe dual of this. So, Wolfe dual of this problem will be maximize LD subject to the constraint that gradient of LD with respect to the primal variable which is mu is equal to 0; and the Lagrangian multipliers corresponding to the inequality constraints are nonnegative.

So, here x are the Lagrangian multipliers. So, x greater than or equal to 0 or which is same as maximize minus b transpose mu plus x transpose A transpose mu minus c subject. So, let us calculate the gradient of LD with respect to mu. So, which will be minus b plus Ax equal to 0 and x greater than or equal to 0.

So, let us take this condition what we have here minus b x plus c equal to 0 which is same as Ax plus b. And if we substitute this condition in the objective function these 2 terms gets cancelled.

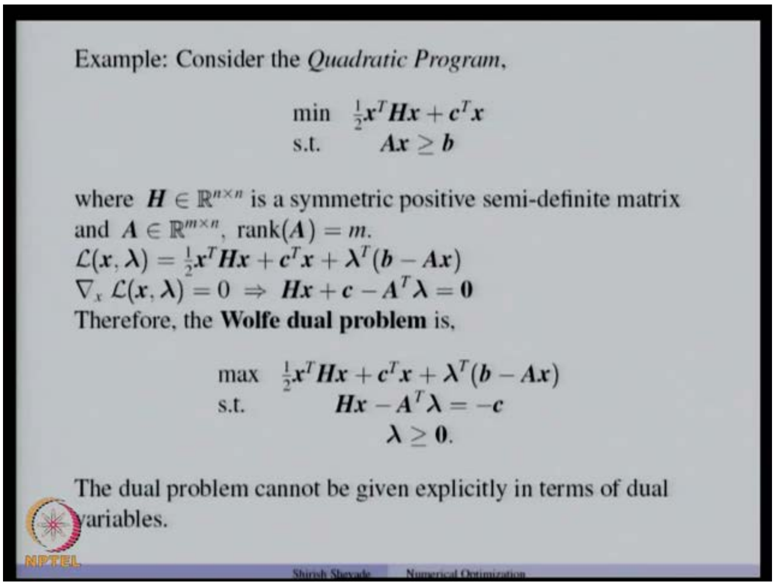
(Refer Slide Time: 1:09:54)



And, what we are left with is minus c transpose x where maximize minus c transpose x subject to Ax equal to b x greater than or equal to 0. So, the this problem which is the dual of the problem which is given in the box is. So, the dual of this problem we can write this as maximize minus c transpose x subject to Ax equal to b x greater than or equal to 0.

And, remember that there was a minus sign which we had not considered. So, we will consider that now so, we have the minus sign. So, here also we will use the minus sign and we will carry that minus sign here. And, now even we consider the equivalent problem from the negative of max of this quantity is same as minimize c transpose x subject to Ax equal to b x greater than or equal to 0. And, this was nothing but the linear program that we started with. So, if you write down the dual of the linear program which is this. And, the if we rewrite the dual of this dual it is same as the original linear program. So, this is where interesting property associated with linear programs. Now, so for we were able to write the dual problem in terms of the Lagrangian multipliers associated with equality and the inequality constraints; but that may not always be the case.

(Refer Slide Time: 1:11:49)

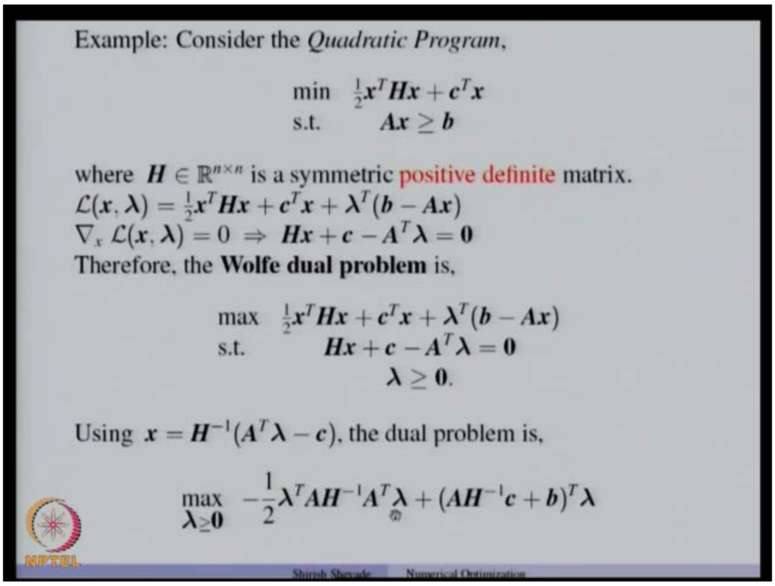


Let us consider a simple example. Let us consider a quadratic program to minimize x transpose H x plus c transpose x subject to the constraint that Ax equal to b. Now, the H matrix is a n by n positive symmetric semi definite matrix and A is n by m matrix and the rank of A is m. Now, we assume that the Slater is constraint qualification holds; this is a convex programming problem. So, we can write the Lagrangian as the objective function plus lambda transpose b minus x b minus Ax. Now, the gradient of the Lagrangian with respect to x equal to 0 implies that H x plus c minus A transpose lambda is 0.

And, therefore Wolfe dual of this original quadratic program is like this where we maximize half of x transpose H x plus c transpose x plus lambda transpose mu minus x; subject to the constraint that H x minus a transpose lambda equal to minus c and lambda is nonnegative. Now, you will see here that the number of variables in this dual problem is more than the number of variables that we have in the original primal problem. Here, the number of variables where n associated with the variable x or here the number of variables will be n plus m associated with x and lambda. So, in this case we really do not gain anything by writing the Wolfe dual. Because we have simply increase the number of variables and the problem also does not get simplified. Because we still have the quadratic programming problem with very linear constraints; same as what we had in the primal problem where the objective function was quadratic and the constraints were linear.

So, we do not gain we do not really gain anything in this case. So, this is an example where the dual problem cannot be given explicitly in terms of dual variables. Here, the variables associated with the dual problem are both x as well as lambda this has to be kept in mind. But things can be simpler if we assume that the matrix H is positive definite matrix instead of the semi definite matrix. Because if H is positive definite then H is invertible. So, we can write x in terms of lambda is by using the fact that H is invertible. And, then 1 can write the dual problem in terms of only the dual variables which are lambda.

(Refer Slide Time: 1:15:03)

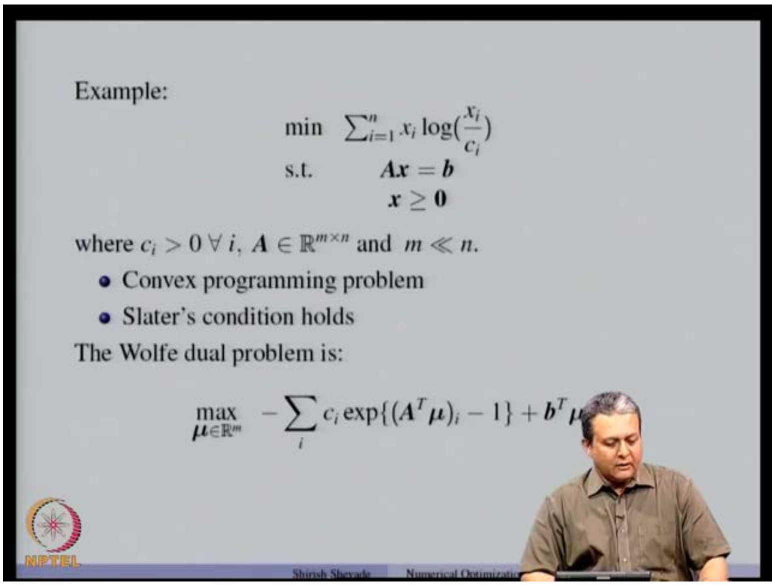


So, let us see how to do that? So, as in the previous case we write the Lagrangian and then said that the gradient of the Lagrangian with respect to x to 0; which gives us this equation H x plus c minus a transpose lambda equal to 0.

And, then we can write the Wolfe dual by using x equal to H inverse A transpose lambda minus c. So, we use this equation and write x in terms of lambda. And, then the dual problem becomes maximize lambda with respect to lambda; the quantity minus half lambda transpose AH inverse transpose lambda plus AH inverse c plus b transpose lambda. So, you will see that by using this transformation we were able to eliminate x from the dual problem. And, this dual problem is now in terms of only the dual variables which are lambda. And, typically these will be less in number compared to the number of variables in the original primal problem.

Further this problem has the simple constraint of the type lambda greater than or equal to 0 compared to the constraint of the type Ax greater than or equal to b in the original primal problem. So, this problem will be easier to solve compared to the original primal problem. So, this an example where we 1st saw that the dual problem cannot be written explicitly in terms of dual variables. But then if the (( )) matrix in the original quadratic program is symmetric positive definite matrix. Then, one can write the primal variables in terms of the dual variable by using the fact that H is invertible. And, then 1 can write the dual problem only in terms of dual variables.

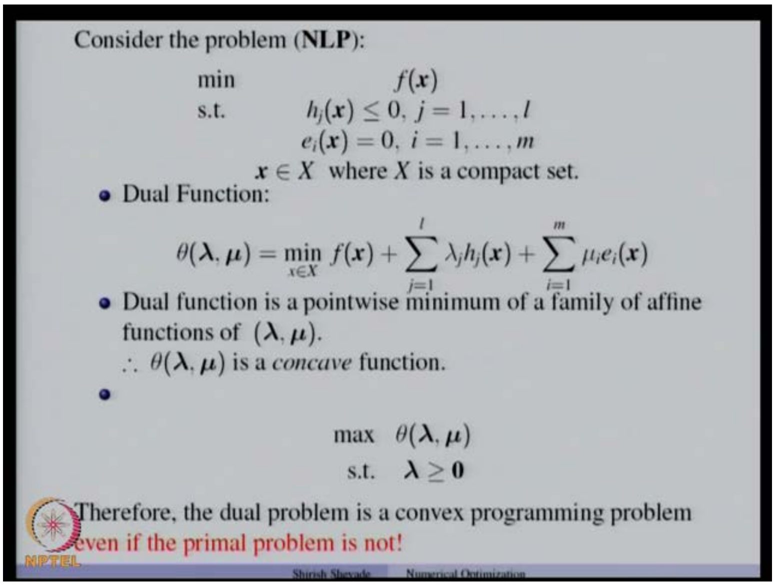
(Refer Slide Time: 1:17:06)



Let us see 1 more example where we minimize sigma Ax i log x i by c I; where c i is are positive constants and the constraints are x equal to b x greater than or equal to 0; and A is n by m matrix where m is much less than n. Now, you can verify that the objective function is a convex function of x. Now, we have already seen that the constraint of this type x equal to b x greater than or equal to 0 which we saw in the linear program case these constraints form a convex set. So, this is a example of a convex programming problem. And assuming that the Slater is condition holds; one can write the Wolfe dual of this problem. So, turns out that the Wolfe dual of this problem is very easy to write.

And, you will see that the Wolfe dual is a maximization problem with respect to mu and the mu is unrestricted in sign. So, the dual of this problem is an unconstraint optimization problem which is easier to solve. So, I will leave it as an exercise to write the dual of this problem and compare that with the 1 which is given here. Now, before we conclude our discussion on duality theory; I would like to make 1 important remark about the dual problem. We have already seen that in many cases the dual problem is easier to solve than the primal problem. But there is 1 important property of dual problems which makes them very attractive. So, let us see that property.

(Refer Slide Time: 1:19:09)



So, let us consider a general non-linear programming problem of the type minimize f x subject to the constraint that h j x less than or equal to 0 and e i x equal to 0; where x is a compact set and x belongs to the compact set. So, x lies in the intersection of the set set of constraints which are represented as x less than or equal to 0, e i x equal to 0 and then the compact set x. Now, let us look at the dual function which was defined as theta lambda mu to be minimize x over x, f x plus sigma lambda j h j x plus sigma mu i e i x. Note that as i mentioned earlier truly speaking this is should be Infimum of x over x. But we assume that the minimum exists so, if the minimum does not exist 1 has to write the Infimum.

Now, the important point that the we should note that this dual function theta lambda mu is a point wise minimum of family of affine functions of lambda and mu. So, the dual function is a function of lambda and mu and it is obtained by a point wise minimum of different affine functions. And, we already know that such a function is a concave function. So, the point wise minimum of family of affine functions is a concave function. Therefore, the dual function is a concave function. And, the dual problem which is maximize theta lambda mu subject to lambda greater than or equal to 0.

So, we maximize the concave function subject to this convex set lambda greater than or equal to 0. And, maximization of a concave function can be written as minimization of a convex corresponding convex function and which is a convex programming problem. So, irrespective of what the primal problem is or what is what the nature of the problem is the dual problem is always a convex programming problem. And, therefore it becomes very attractive to solve this dual problem. Because there is no question of local minima as far as dual problems are concerned. So, therefore many applications typically are based on the dual problems because which are convex programming problems. And, then the solution of the primal problem can be obtained by after obtaining solution of the dual problem. So, with this we conclude our discussion on duality theory.

Thank you.