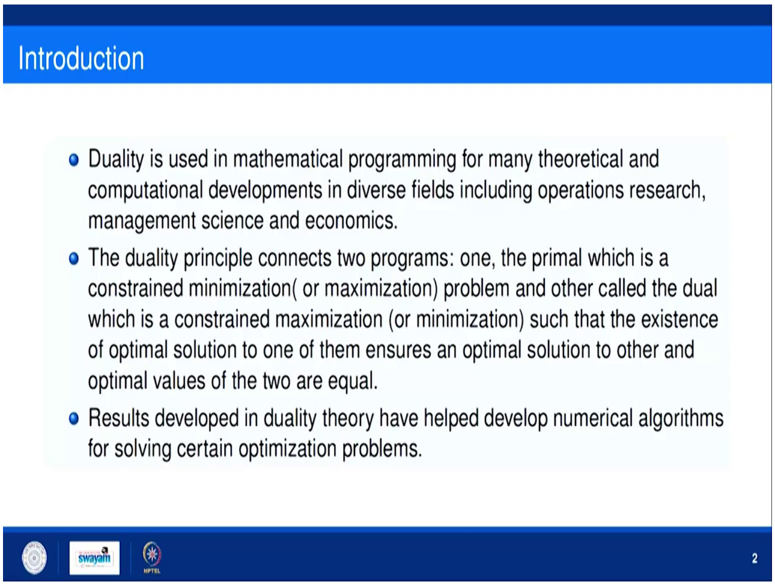
Essential Mathematics for Machine Learning
Prof. S. K. Gupta
Department of Mathematics
Indian Institute of Technology, Roorkee

Lecture - 38
Concepts of Duality

Hello friends. Welcome to lecture series on Essential Mathematics form Machine Learning. In the last lecture, we have seen few concepts of ah machine learning; fundamentals of support vector machines. Now, in this lecture we will see the Concepts of Duality which we will use in soft margin or hard margin classifiers.

The concept of duality is important to find out Lagrange function to find out the dual of a given problem so that, it will be computationally the problem will become computationally easy to solve.

(Refer Slide Time: 01:01)



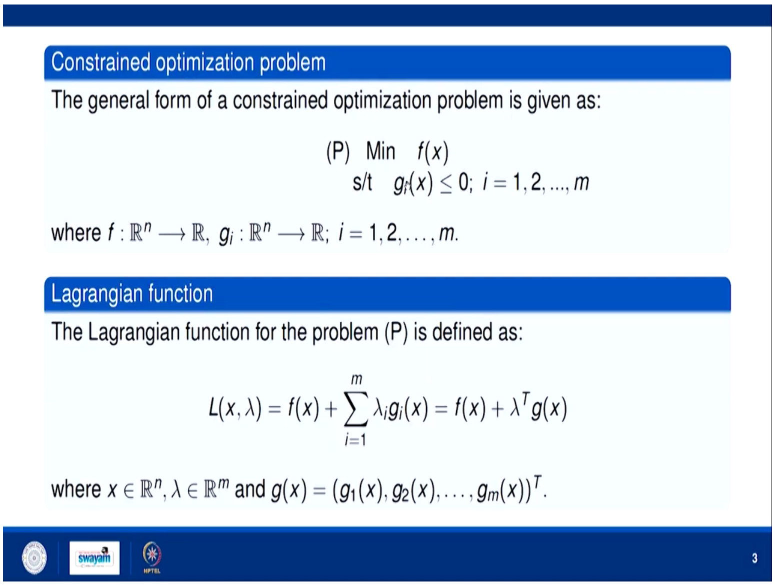
So, what do you mean by duality and how can you find out dual of a general non-linear problem? So, first it is some introduction. So, duality is used in mathematical programming for many theoretical and computational developments in diverse fields including operation research, management science and economics.

The duality principle what a duality principle is? The duality principle connects two programs: one, the primal which is the constraint minimization or maximization problem and the other is called the dual for which is a constraint maximization or minimization. Such that the existence of optimal solution to one of them ensures an optimal solution of the other and the optimal values of the two are.

So, the main important concept behind duality is that, we have 2 problems. Problem 1, problem 2; is connect it connect the 2 problem in such a way that if you are having the optimal solution of the 1 problem say 1, if you have the optimal solution of a problem 1; then it will ensure the optimal solution of the other problem 2 also, and the objective value of both the problems are equal.

So, that is the main concept of duality principle. Now, why duality? So, results developed in duality theory have helped develop various numerical algorithms for solving certain optimization problems. So, it is required to in developing various numerical optimization algorithms.

(Refer Slide Time: 02:35)



So, now the question is how can we write the dual of a given non-linear optimization problem? So, suppose we have this non-linear optimization problem; minimization of a function f x subject to g i x less than equal to 0; i from 1 to m, so there are m number of constraints and we have a single objective f which is to be minimized.

So, this problem we are calling as problem P. So, where this function f is from R n to R; that means, there are n number of unknowns g i are the function from R n to R; for all i. Now, to find out the dual of this problem P, let us first write the Lagrange function of this problem. How we define a Lagrange function? So, Lagrange function is basically function of two parameters x and lambda, x is R n and lambda is in R m.

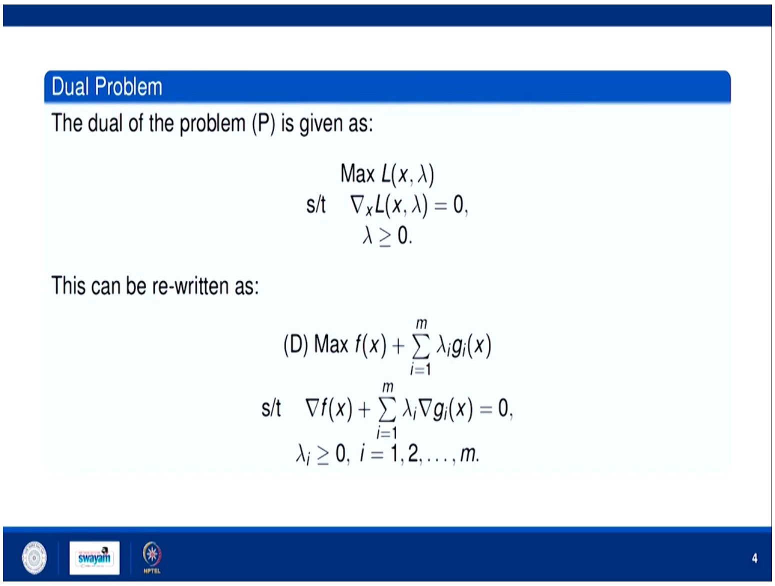
So, lambda are the Lagrange; a lambda are called Lagrange multiplier corresponding to the constraints. How many constraints we are having in this problem P? We are having m number of constraints. So, corresponding where we will be having m number of Lagrange multipliers in the Lagrange function.

So, the Lagrange function is a function of two unknowns x and lambda, which may be defined as the objective function; the objective function is f x here you see, objective function comes here. And the sum of lambda i g i; i is from 1 to m because, there are m number of constraints.

So, this lambda 1, lambda 2, lambda 3 up to lambda n; these are called Lagrange multipliers. So, in j we can also write this problem as f x plus lambda transpose g x; lambda transpose see lambda is basically, lambda 1, lambda 2 up to lambda m, and g is basically g 1, g 2 up to g m.

So, when we multiply these two then we get, g 1 lambda; lambda 1 g 1 plus lambda 2 g 2 and so on up to lambda m g m. So, in this way we can write we can write the Lagrange of a given non-linear programming problem.

(Refer Slide Time: 04:47)

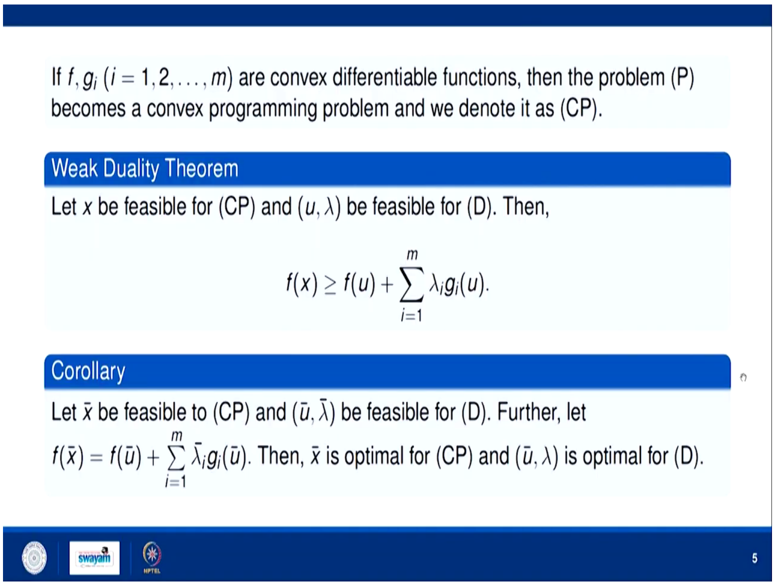


Now, how can you write a dual of this problem? So, dual may be as, the dual of this problem P is given as we write maximizing of Lagrange function; subject to gradient respect to x of this Lagrange equal to 0; lambda greater than equal to 0.

So, this is how we can defined Lagrange we can define the dual of this problem P ok. Now, this dual this problem can be rewritten as see what is the Lagrange function? Lagrange function we have already seen. The Lagrange function is basically f x plus summation i from 1 to m lambda i g i ok.

So, this is Lagrange, this Lagrange will come here maximization of this subject to; now the gradient of this function this Lagrange respect to x. So, gradient of this respect to x it will come here. The gradient of this respect to x will come here equal to 0 and lambda i greater than equal to 0 for all i. So, this is how we can write the dual of this problem P; then this dual is also called Wolfe phase dual ok.

(Refer Slide Time: 05:54)



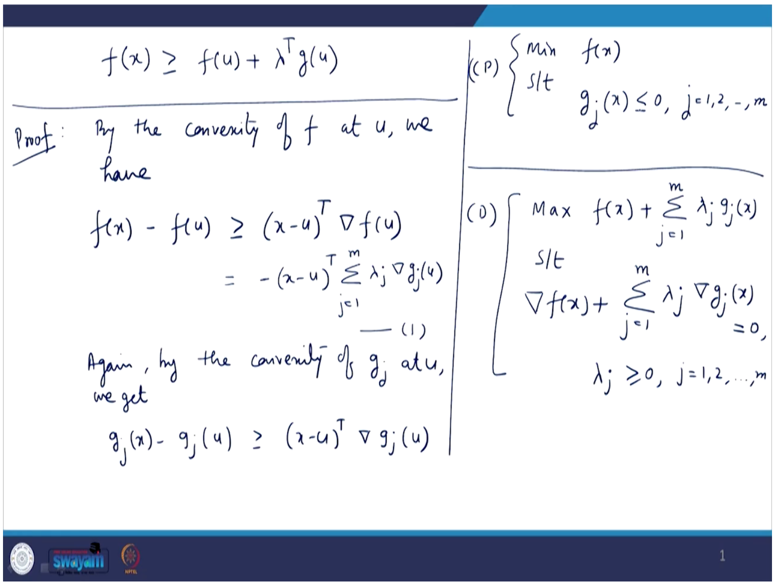
So, if we are calling two problem as a; if one problem the initial the first problem is called primal problem, and the second problem is called the dual problem. Now, there is some relation between these two, as I already told you that the optimality, condition optimality of one of the problem guarantee is the optimality of the other problem and vice versa.

So, there are some relations between these two and these are called duality relations. There are certain duality theorems you see; so what are those results? Let us first try to understand those results. So first is if f and g i; g i for all i are convex differentiable functions, in this problem.

See, in this problem P if this f and all g i's are convex then, this problem is called convex programming problem. And now we denote this problem as CP ok. The problem P we are calling as CP. So, the first result is which we are calling as weak duality result it states that, if x is a feasible solution of CP and u, lambda is feasible for problem D, which is a dual of this problem P of this problem CP then, this inequality always hold.

This inequality means the minimum value of the objective function of the primal problem P or CP is always greater than equal to the maximum value of the objective function I mean the value of the objective function of problem D.

(Refer Slide Time: 07:42)



So, how we got this result? So, let us try to understand. So, we are having two problems here. So, what is our primal problem? The primal problem is minimization of f x subject to g j x less than equal to 0; j from 1 to m. And this problem if, f and g j's are convex for all j; we are calling as CP.

And the problem D which is a dual of this problem CP is given by maximization of Lagrange function which is f x plus summation j from 1 to m lambda i lambda j g j x subject to the gradient of this should be 0; that means gradient of f x plus summation j from 1 to m lambda j gradient of g j x equal to 0; and lambda j greater than equal to 0 for all j.

So, this is this problem is D and this problem is CP. So, what is weak duality result? Weak duality result says that, if x is any feasible point of the problem CP and u lambda is any feasible point of the problem D then, f x is always greater than equal to f u plus lambda transpose g u.

So, now we have we have to prove this. So, it is given to us that f is convex, and g j for all j's are convex; x is feasible for this problem and u lambda is feasible for this problem D. So, let us try to prove this result. So, since by the convexity of f at u, we have f x minus f u is greater equal to x minus u transpose gradient of f u.

This is why the convexity of f; because f is a differentiable convex functions hence, this result will hold, that we have already discussed in our in the previous lectures. Now, since u is feasible for this problem D. So, you multiply both sides by x minus u transpose so, that means, this will be equal to minus x minus u whole transpose summation i from 1 to m lambda j gradient of g u ok.

This is because you see u lambda is feasible for this problem D. So, you multiply both sides by x minus u transpose; so, this quantity will be equal to the negative of x minus u whole transpose summation lambda j gradient g j u. Now, again so say this inequality as inequality 1. So again, by the convexity of g j at u, we have we get g j x minus g j u is greater than equal to x minus u whole transpose gradient of g j u.

(Refer Slide Time: 11:57)

Multiply by λ_j both sides of the above inequality, & sum up all the m- inequalities, we get

$$\sum_j \lambda_j g_j(x) - \sum_j \lambda_j g_j(u) \geq (x-u)^T \sum_j \lambda_j \nabla g_j(u) \quad \text{— (2)}$$

from (1) & (2), we obtain

$$f(x) - f(u) \geq \sum_j \lambda_j g_j(u) - \sum_j \lambda_j g_j(x)$$

$$\geq \sum_j \lambda_j g_j(u) + 0$$

$$\Rightarrow f(x) \geq f(u) + \sum_j \lambda_j g_j(u) \quad \square$$

2

So, multiply by lambda j both sides of the above constraint of the above inequality, and sum up all the m in equations inequalities, we get; so, now what we will obtain? Summation lambda j it is lambda j g j x it summation over j minus summation over j lambda j g j u is greater than equal to minus x minus u whole transpose summation j lambda j gradient of g j u; this is suppose this is in equation 2.

Now, you can use this in equation because this side is less than equal to this side oh this is positive this is not negative; this is positive. So, now you from one and two see this side is less than equal to this side ok, and the negative of negative of this side is greater than equal to this side.

So, using 1, from 1 and 2 what we obtain? From 1 and 2 we get here we obtained this from this side it is f x minus f u is greater than equal to so, you multiply both sides with minus 1. So, this factor will come before towards ok; axis feasible for this problem.

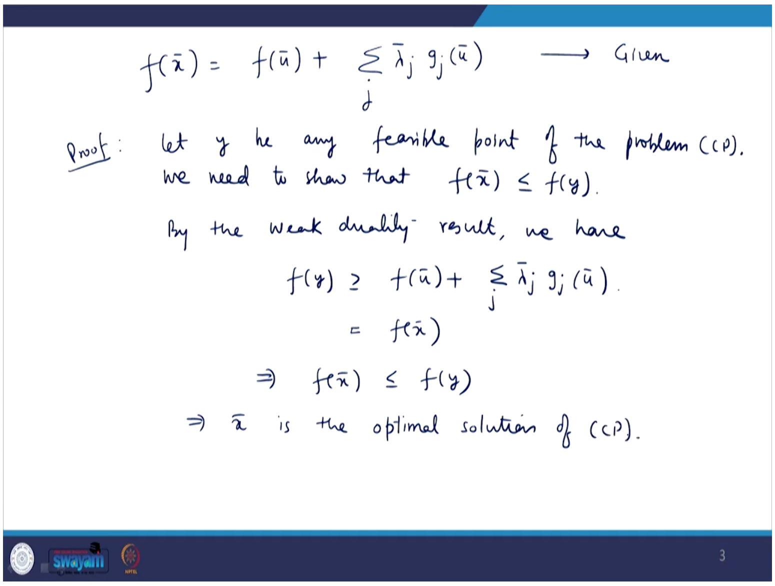
So, g j x will be less than equal to 0 for every j ok. So, negative of g j will be greater than equal to 0 for all j and lambda j's are non-negative. So, from this we will obtain that this quantity with negative sign is greater than equal to 0; so that means, it is greater than equal to g j lambda j g j u plus 0 ok, because this because this is less than equal to 0; for every j and lambda j's are non- negative.

So, when you multiply and sum it sum it up for all j; so this quantity will be less than equal to 0 with negative it will be greater than equal to 0. So, what we obtain from here now it is f x is greater than equal to f u plus summation over j lambda j g j u. So, hence we got the result. So, this is our weak duality results result. So, basically weak duality result is important to find out the bound of the problem; as the primal problem or the dual problem.

If there is any u or lambda which is feasible for the problem D then, we can easily say that this value will be finite for that u and lambda and f x will always be greater than equal to that value. So, that will be helpful in finding out the bound of a given problem. So, the primal problem or the dual problem similarly if you know x as a feasible problem for the problem CP then, you know this value some finite value; then we can say that the objective function a dual problem will always be less than equal to that value.

So, this is helpful in finding out the bounds of the primal problem or the dual problem. Now, the next is now the next result from the weak duality result that if x bar is a feasible point of CP and u bar lambda bar is feasible for D, and suppose the value of the objective function of the primal at x bar and the value of the dual problem at u bar lambda bar both are equal; both coincides. Then, as far as the optimal solution of optimal solution for CP and u bar lambda bar will be the optimal solution for problem.

(Refer Slide Time: 16:38)



So, the proof is again very easy to obtain see here what is given to us? It is given to us that f x bar is equal to f u bar plus summation over j lambda j bar g j u bar. So, it is given to us. And we have to show that x bar and x bar and u bar lambda bar are nothing but the optimal solutions of the primal or dual problems.

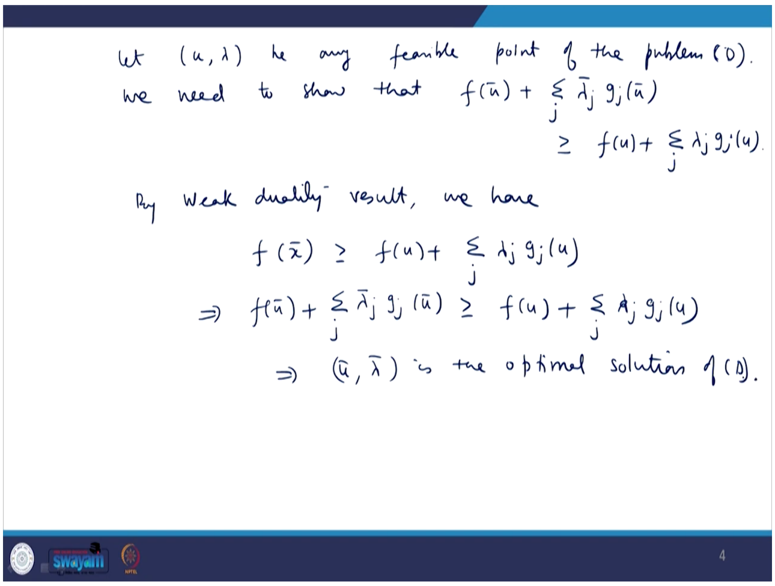
So, if you want to show that x bar is an optimal solution or dual problem then, we need to show that f x bar will be less than equal to f y for any feasible y, 2 problem CP. So, let us try to prove this result proof let y be any feasible point of the problem CP ok. So, we need to show that f x bar is less than equals to f y ok; this we need to show.

So, now by the weak duality result, we have f y is greater than equals to fu bar plus summation over j lambda j bar g j u bar; since, y is feasible point for the any feasible point of the primal problem and u bar lambda bar is also a feasible point of the problem D; then by the weak duality result this inequality will hold ok.

And this quantity is nothing but is equal to f x bar. So, it is equal to f x bar. So, this implies f x bar is always less than equals to f y and hence, we can say that x bar is nothing but the optimal solution of the primal problem because, this value is always less than equal to f y for any y; for any feasible y so; that means, x bar is the optimal solution of the primal problem.

And similarly, if you want to show that u bar lambda bar is an optimal solution or dual problem then, now so this implies x bar is the optimal solution of the CP.

(Refer Slide Time: 19:25)



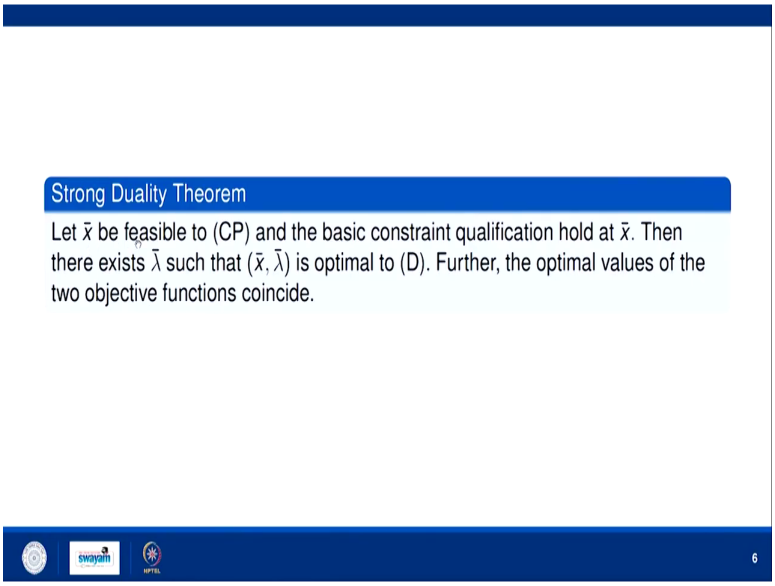
Now, let u y u lambda be any feasible point of the problem D ok. Now, we have to show that u bar lambda bar is an optimal solution. That means, so we need to show that f u bar plus summation over j lambda j bar g j u bar is greater than equal to f u plus summation over j lambda j g j u; because it is a maximization type problem.

So, again this is easy because, by the weak duality results, what we have? We have f x bar is greater than equal to f u plus summation over j lambda j g j u g j u; because x bar is the feasible point of the primal problem and u lambda is a feasible point of the dual problem.

So, this implies since, this is equal to because this is equal to this so we can replace this by this. So, this is this implies f u bar plus summation over j lambda j g j u bar is greater than equals to f u plus summation over j lambda j g j u bar g j u. So, this implies u bar lambda bar u bar lambda bar is the optimal solution of D.

So, hence we have proved the result. So, in this way we can say that, if the optimum values of the primal and dual are equal then the points where they are equal are nothing but the optimal solutions of the respective problems.

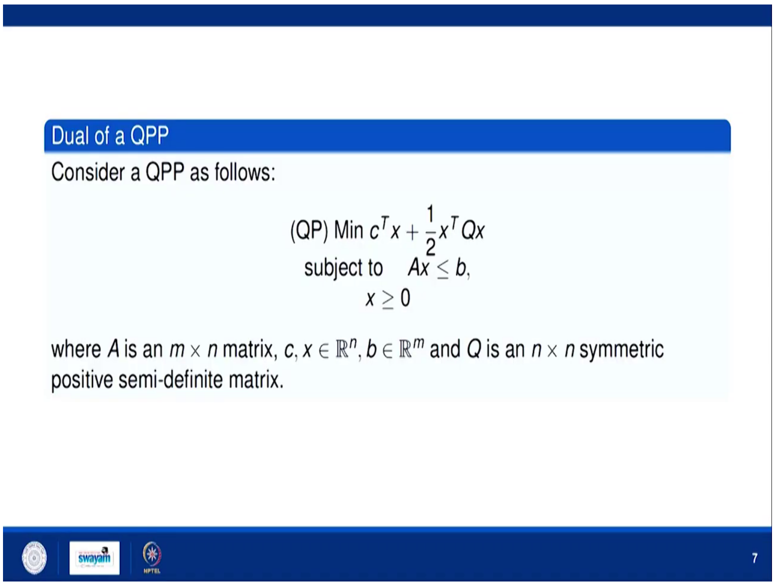
(Refer Slide Time: 21:50)



So, next here is strong duality result with the state that, if x bar is the feasible point of CP and the basic constraint qualification hold at x bar then, there exists lambda bar such that x bar lambda bar is optimal to the problem D. Further, the optimum values of the two objectives coincide.

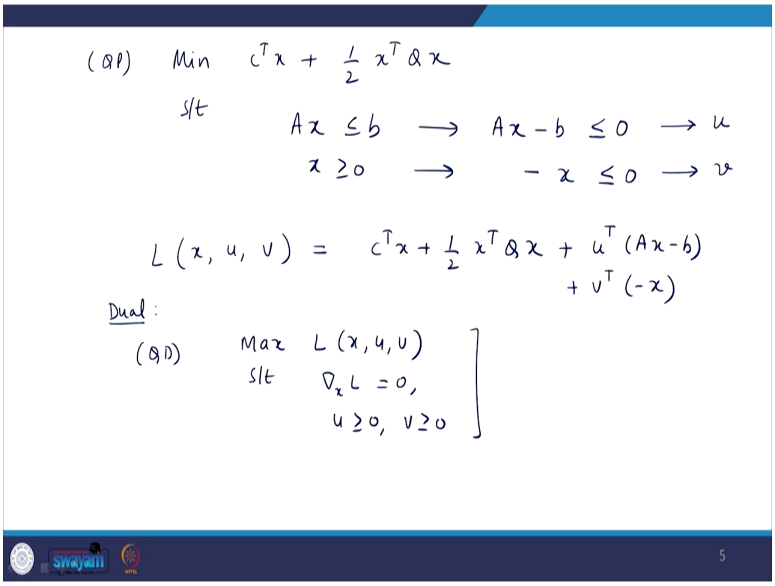
So, basically weak duality result is helpful in finding the bounds of the primal of the dual problems ok. And their strong duality result guaranties the optimal solution of the one problem if we have the optimal solution of the primal problem. If we have the optimal solution of primal problem then, we there exist an optimal solution in the dual problem and the value of the objective values of both are equal ok.

(Refer Slide Time: 22:33)



So, now let us write to try to find that how we can write a dual of a QPP, quadratic programming problem. So, this is a Quadratic Programming Problem. So, how we can write a dual we already know that this is a QPP because, objective function is quadratic in nature and all constraints are linear. So, here A is an m cross n matrix, c and x belongs to R n, b belongs to R m and Q is a n cross n symmetric positive semi definite matrix. So, again to write a dual we can first write the Lagrange of this function.

(Refer Slide Time: 23:09)



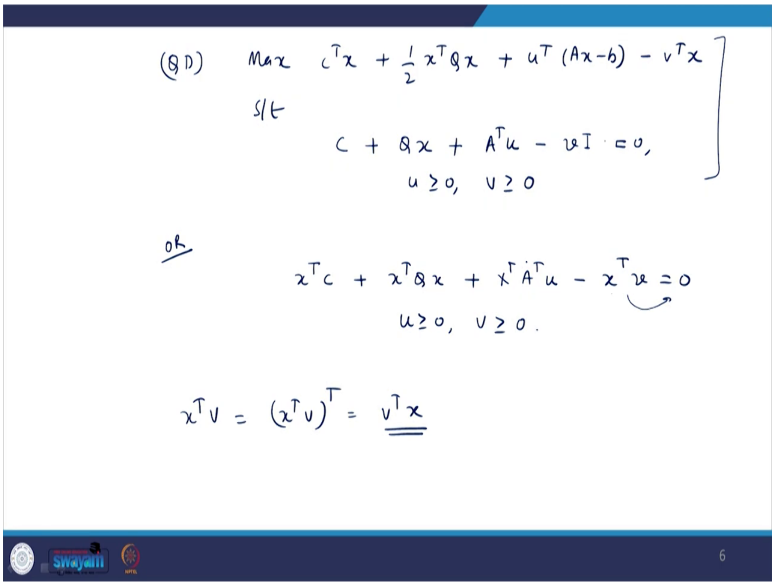
So, what is the problem we are having? Problem is minimizing c transpose x plus 1 by 2 x transpose Q x subject to Ax less than equal to b and x greater than equal to 0. Now, this constraint can be re written as A x minus b less than equal to 0 and this constraint can be written as minus x less than equal to 0. Let the that the Lagrange multipliers corresponding to these constraints are u, and let the Lagrange multiplier of these constraints are v.

So, L which is now a function of x, u and v can be written as, first is the objective function; objective function is what? c transpose x plus 1 by 2 x transpose Q x ok; plus now it is u transpose Ax minus b we are there is u transpose G x, it is u transpose Ax minus b; G x is this and this collectively we are calling as G x and this is plus v transpose minus x. So, this was the Lagrange function of this problem QP, this problem is QP.

Now, how can we write a dual of this problem? The so, a dual of this problem will be D here QD can will be now, it is maximization of L x, u subject to gradient with respect to x of L will be 0 and u greater than equal to 0 and v greater than equal to 0; as we have already discussed the prop the dual of a given a NLP given non-linear problem will be nothing but, the Wolfe phase dual is nothing but, maximizing of the Lagrange function subject to gradient respect to x of L is 0.

And the Lagrange multipliers must be non-negative. So, now let us rewrite this problem. So, what will be QD now? It will be maximization of 1 by 2 c transpose x; oh it is not 1 by 2 it is c here. It is c transpose x.

(Refer Slide Time: 25:12)



So, it will be c transpose x plus 1 by 2 x transpose Q x plus u transpose A x minus b minus v transpose x subject to what are the conditions? So, you can verify from here subject to gradient respect to x. So, what will the gradient with respect to x?

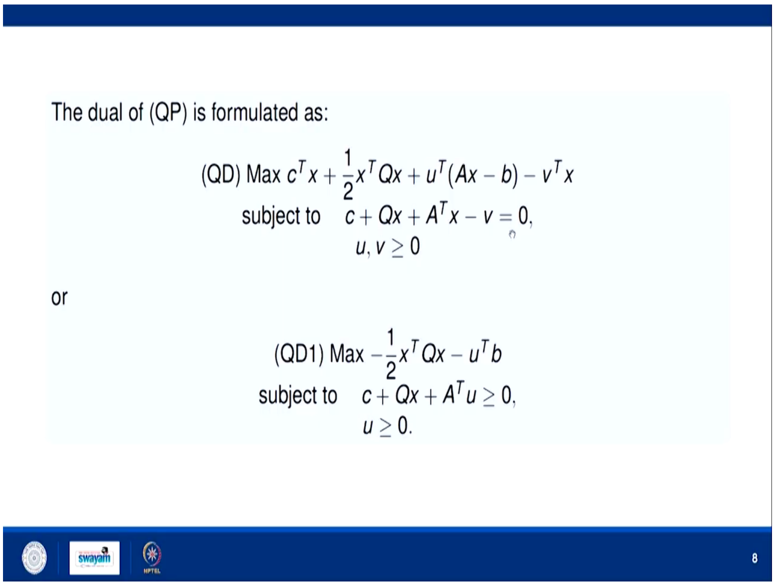
You differentiate this function with respect to x so, it will be c plus Q x because q is a symmetric matrix; otherwise it will be q transpose x, but here Q x is a symmetric matrix so Q transpose would be Q; plus it will be A transpose u and minus v equal to 0; v I equal to 0 ok.

And this u is non-negative and v is non-negative ok. Now, this can be rewritten as, now if you see if you multiply this equation by c transpose by x transpose so what will obtain? So, it will be x transpose c plus x transpose Q x plus A x transpose A transpose u minus x transpose v equal to 0; and x and u and v non negative.

So, I am trying to like rewrite these equations these this problem. So, basically dual is this; but I am trying to write this problem in a simplified way. So, I can replace this quantity this is x transpose v which is same as x transpose v whole transpose because, this is a scalar quantity. So, this is equal to v transpose x.

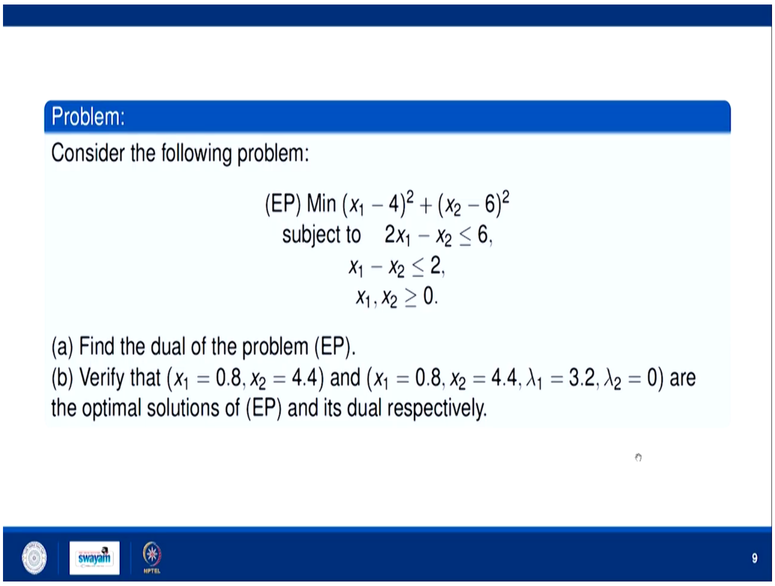
So, I can replace v transpose x from this equation to this. So, if I replace this if I replace this from here to this equation what will be what I will obtain finally? So, finally I will obtain, now you can see from here.

(Refer Slide Time: 27:41)



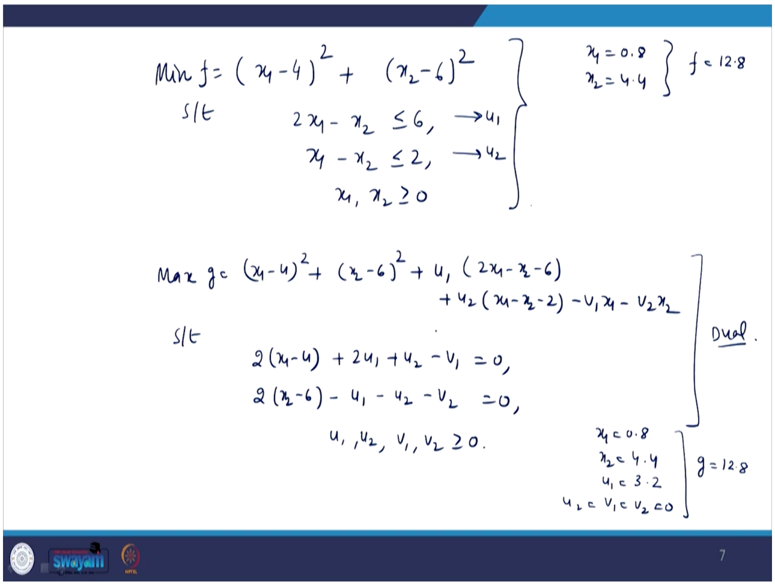
So, this is a dual which we are already obtained. So, now you can replace v from this equation to this ok. So, if I replace this finally after simplification of this simple equation I will obtain this. Now, this is equal to v; so I can write v as greater than equal to 0; so it will be greater than equal to 0. So, v is eliminated from the entire expression. So, now I can say that this is also the this is the dual of this problem QP. This problem QP ok.

(Refer Slide Time: 28:18)



So, let us take a problem now. Suppose, you are having this QPP, and you want to write a dual of this QPP. So, how can you write?

(Refer Slide Time: 29:29)



So, what is the QPP now? It is minimization of x 1 minus 6 whole square plus x 2 minus 4 whole square 4 and 6 ok; it is 4 here; 6 here. Subject to what are the equations? The first equation is 2x 1 minus x 2 less than equal to 6, and x 1 minus x 2 less than equal to 2, and x 1, x 2 non negative. So, this is the primal problem.

So, what we are dual of this problem? We have already seen, the dual will be nothing but, it will be x 1 minus 4 whole square plus or if you want to eliminate u then that will be nothing but, so you have to otherwise you have to transform; so anyway you can write in that way only; x 2 minus 6 whole square plus so that the Lagrange multiplier correspond to this is u 1, for this is u 2, for this is a v 1 and v 2.

So, u 1 2x minus x 2 minus 6 plus u 2 times x 1 minus x 2 minus 2 minus v 1 x 1 minus v 2 x 2 ok. Subject to now, take the gradient of this problem the gradient will be 2x 1 minus 4; first with respect to x 1 2 x 1 minus 4 it is plus 2u 1 plus u 2 minus v 1 equal to 0, and it is 2 x 2 minus 6 minus u 1 minus u 2 minus v 2 equal to 0 ok; u 1, u 2, v 1, v 2 non negative.

Now, so these are dual of this problem. This is a dual ok. Now, at this point 0.8, 4.4. So, if you are taking x 1 as 0.8 and x 2 as 4.4. So, it is satisfying all the constraints this point is satisfying the constraint; and the objective value of this problem is suppose this objective value is f and this is suppose g the objective value coming out to be 12.8.

Now, if we take this problem then, here we take x 1 as 0.8, x 2 as 4.4 ok; and u 1 as 1.6, v 1, v 2, u 1, u 2 both all are 0. So, this is basically so lambda 1 here given as 3.2 so it is 3.2.

Now, if you take this point, this point satisfying both the dual constraint all the dual constraints in fact, and the objective value at this point is also the same as f is also the g is also at this point is 12.8; you can verify very easily very easy you can verify because, when x 1 is 0.8 and x 2 is 4.4, so this is 0.

So, this is 0; u 2, v 1, v 2 are already 0. So, this whole quantity is 0; so we left with only this these two terms which are same as the; which is same as a two terms of f. So, this is nothing but, a same objective value.

So, hence we can say that since the value of the objective function at these two points coincide; so these are nothing but, the optimal solutions of the prog this problem and this problem respectively ok. So, in this way if you want to write out a dual of a given NLP; here for illustration I have taken quadratic a programming problem.

In general, if you are having any non-linear problem which is a convex programming problem then, though you can write the Wolfe phase dual in this manner. You have a given non-linear problem, minimization type subject to g j's less than equal to 0. You know you can add a dual as the maximization of a Lagrange function subject to gradient of Lagrange respect to x equal to 0 and all Lagrange multipliers greater than equal to 0. So, that is a general format of writing a dual of a given non-linear programming problem. In the next lecture we will see that how we can find out hard margin classifier using duality theory so.

Thank you very much.