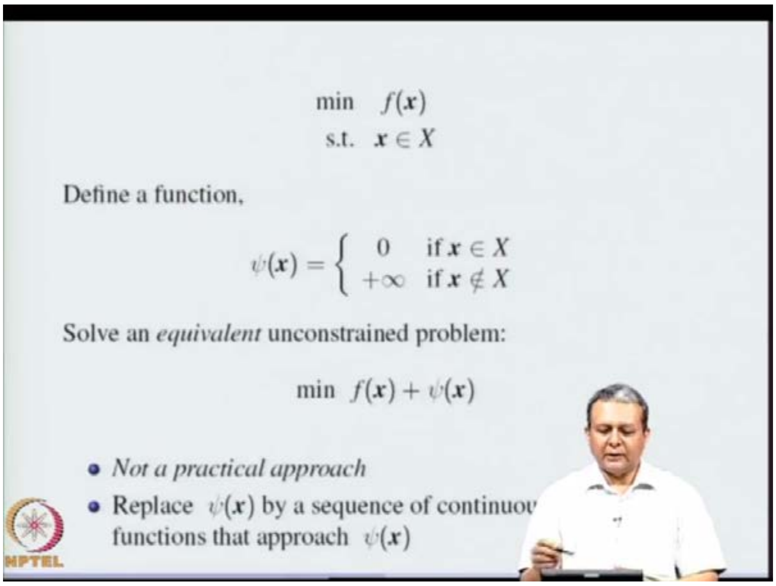
**Numerical Optimization**
**Prof. Shirish K. Shevade**
**Department of Computer Science and Automation**
**Indian Institute of Science, Bangalore**

**Lecture - 40**
# Barrier and Penalty Methods, Augmented Lagrangian Method and Cutting plane Method

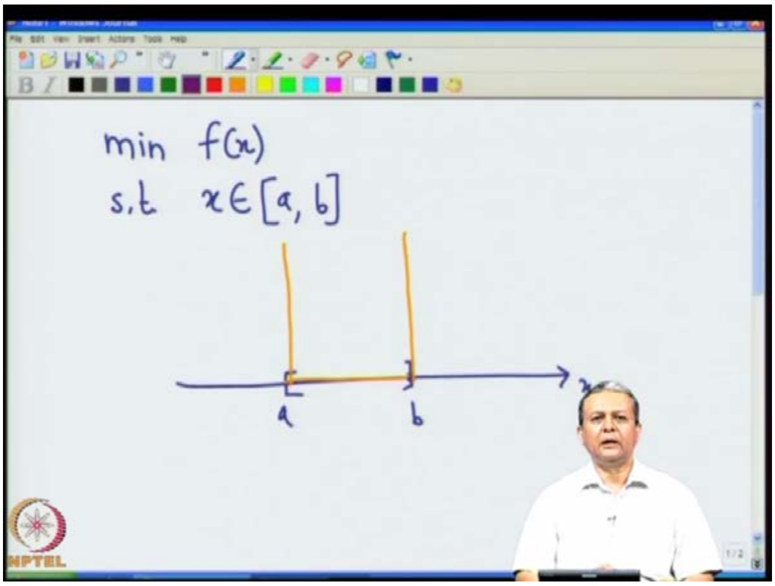
Welcome back. In the last lecture, we started discussing about penalty function methods, and the idea was the following that, if you want to minimize the function f x.

(Refer Slide Time: 00:30)



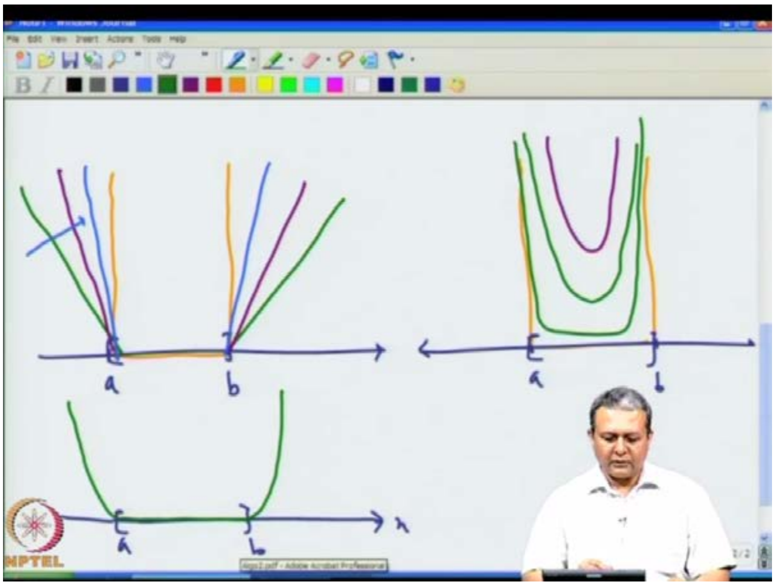

Subject to the concern that x belongs to the set x. The idea of penalty function is to define a function psi x, which takes the value 0 when x belongs to x and, which takes the value plus infinity, when x does not belongs to x. And this next step is to solve, an unconstrained, which is to minimize f x plus psi x. So, you will see that this problem is equivalent to this problem and other solution. If x star is solution to this problem, then psi x star will be 0, because x star belongs to x, and we will get f of x star as the optimal objective to functional value.

(Refer Slide Time: 01:27)




So, the idea was the following that if you want to minimize some function subject to some interval, closed interval a b. So, up function psi would however the following from where it takes the value 0 in this interval and infinity elsewhere. So, the moment we go move away from this interval the function was 2 plus infinity. Now, this is this function is not continuous, so it cannot be used directly to solve an unconstrained optimization problem, because most of our techniques that we studied were based on derivatives of the function.

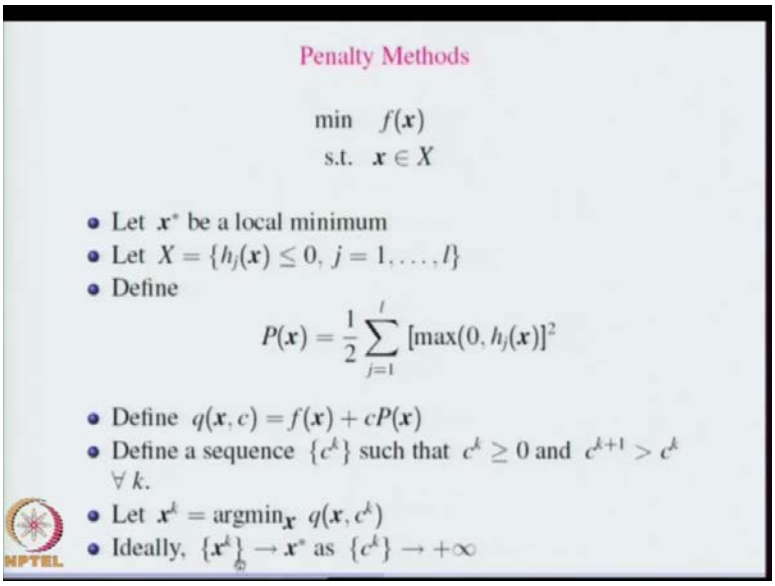
(Refer Slide Time: 02:15)




So, so it may be a good idea to approximate the function by function like this, and then make use of a series of functions. So, for example, we could have another function which is like this and then another function this is like this. So, as the parameter value changes the function slowly moves towards the desired function, one can also have a function like this. So, in this case one could change the parameter value to get a function like this or change the parameter value, to get a function which is like this.

So, if you use this set of functions they, they are called the penalty function methods, and if you use these set of functions the methods derived using them are called barrier function methods. So, this method is also similar to interior point method that we saw earlier, that a point which is feasible at the beginning. The next iteration that point will not, be allowed to go outside the feasible region, because of this barrier so they are also similar to the interior point methods. So, instead of using psi x, we defined a sequence of continuous non negative functions that approach psi only at the limit.

(Refer Slide Time: 04:01)




So, let us look at the penalty function methods so let us consider the problem to minimize f x subject to the constraint that x belongs to x. And let x star be a local minimum let us denote the constraint set x, which is formed using the inequality constraint x j is less than or equal to 0. So, all let us assume that all the constraints are of the h j x less than or equal to 0 later on we will see the extension of this to a general non-linear program.

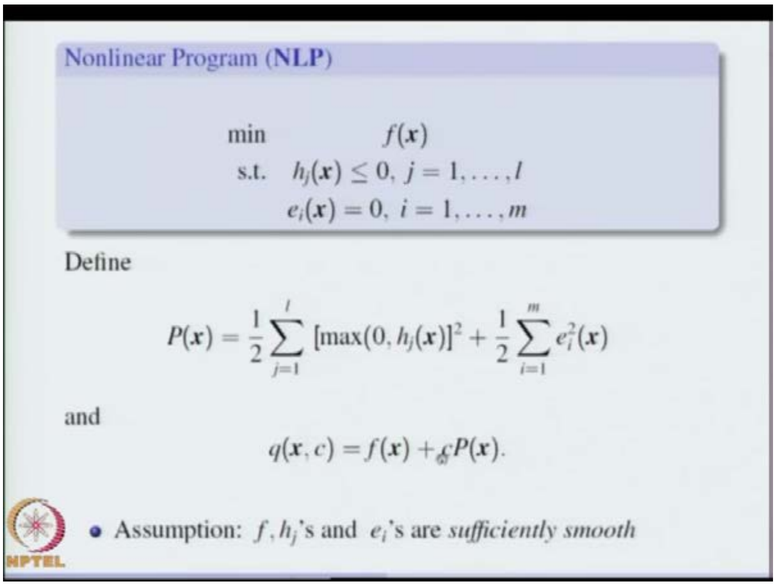
Now, the way, we have defined the penalty function here, we will see that the function is continuous but, it is not differentiable at the boundary points. So, it may be a good idea to have a function, which is continuously differentiable, so that we can use our derivative based methods, to solve those, problem. So, one possibility is to have a function so suppose this is the interval a b so one possibility is to have a penalty function, which is sufficiently smooth.

So, instead of having only the continuous function we can have sufficiently smooth function, so that we can use the derivative based approaches to solve this problem. So, one such function is a function, where which takes the value 0 if h j x is less than 0 and if h j x is greater than 0 then it will take the value h j x. So, the idea is to penalize, whenever x crosses the feasible region, and the penalization is can be obtain using this penalty function.

Note that this is not only the penalty function one can have different penalty functions, but, suppose if we choose this function. Then let us define function q of x and c to be the function f x plus seen to P x, where c is a positive constant and P x is the function that we have defined here. Now, have a sequence c k such that every c k is non negative and c k plus 1 is greater than c k for all k.

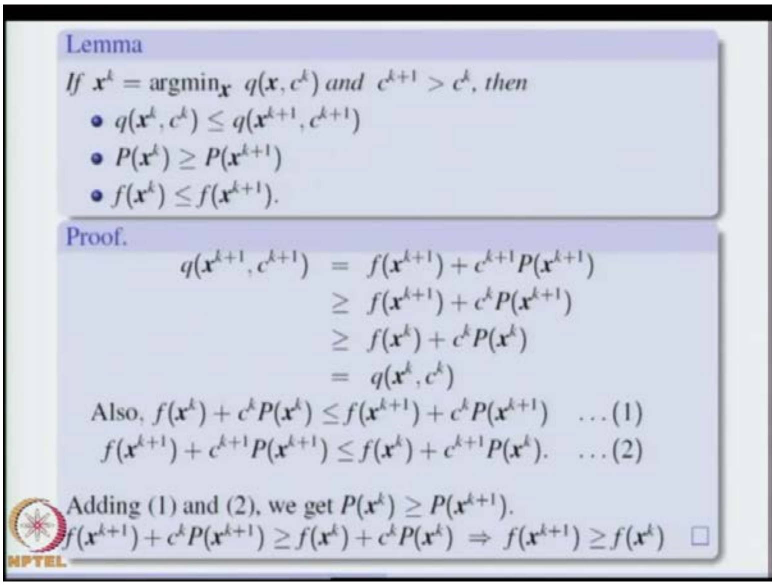
Then suppose for a given value of c k, let x k be the minimum value or optimal value of x, which minimizes q x and given c k. So, if we have a sequence as c k, then q x c k or s k is will give a such sequence of optimal values for different values of c k. Remember that c k plus 1 is greater than, c k so it is monotonically increasing sequence. So, c would go to infinity as k increases and, we have a sequence of s k for every value of c k. Now ideally what should happen is that the s k tends to infinity as a c k tense to infinity.

(Refer Slide Time: 07:45)




Now, for a general nonlinear program, we have minimize f x subject to h j x less than or equal to 0 and e x equal to 0. I going from 1 to m and j going from 1 to l, we can define the penalty function in a similar way. So, for inequality constraints we have max of 0 h j x square for equality constraints, we have e i x square. Now, let us assume that, this f h and e are sufficiently smooth, and let us define the q of x c to be f x plus c P x where c is again a objective constant. Now, when we have a sequence c k does the sequence s k generated by minimizing q x c k converges to x star, so we would like to answer this question.

(Refer Slide Time: 08:41)

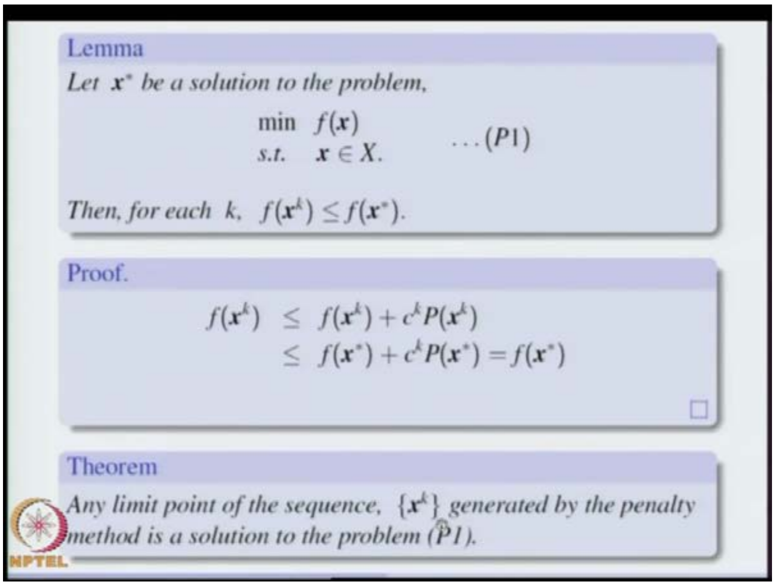



Now, here in an important lemma, which says that if k is the optimal value of 2 x c k and c k plus 1 is greater than c k then the following 3 conditions hold, so that is q x k c k is less than or equal to u x k plus 1 c k plus 1. Then P x k is greater than or equal to P x k plus 1 and f s k less than or equal to f x k plus 1.So, you will see that the value of f does increase, at every iteration, now we will show that this value will converge to f of x star as c tends to infinity. So, let us first prove this now by definition q x k plus 1 c k plus 1 is nothing but, f x k plus 1 plus c k plus 1 P k P x k plus 1. Now c k plus 1 is greater than c k, so we can write this as this quantity, get equal to this quantity, because c k plus 1 is greater than c k. Now, at c k x k is the optimal value, so this quantity will be greater than or equal to f s k plus c k P s k and that is nothing but, q x k c k.

So, the auxiliary function q increases after every iteration that is the first result. Now, the second result says that the penalty function value at x k is at least the penalty function value at x k plus 1. So that means that as k increases, the penalty function value decreases. Now, let us see how to prove this, now if we fix c k f k is the optimal value, so f of x k plus c k into P p of x k will be less than or equal to this quantity. On the other hand if you fix c k plus 1 then f x k x k plus 1 is the optimal value for any feasible x. So, this second inequality holds, now if you add this 2 inequalities than the quantities involving f get canceled. And we use the fact that c k plus 1 is greater than c k to show that P x k is greater or equal to P s k plus 1. So that proves the second part that the penalty function value decreases.

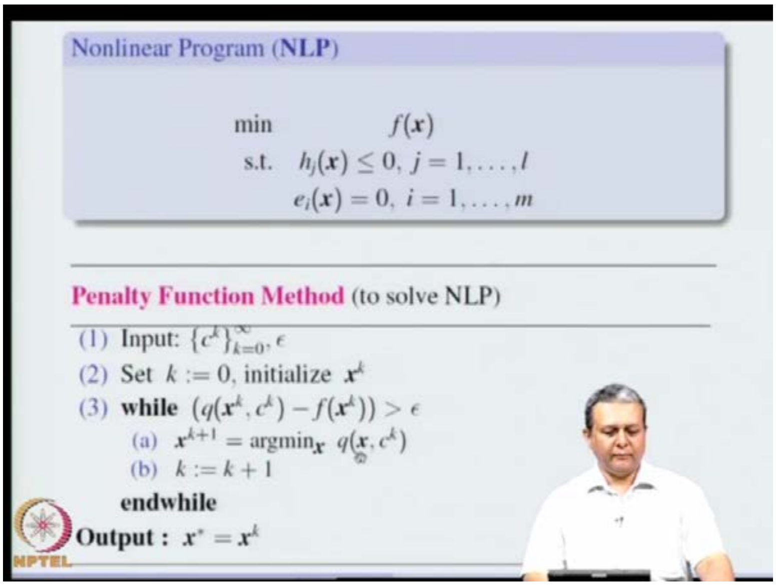
Now, what happens to the objective function f, so we know that for a fixed c k s k is the minimum value. So, f of x k plus c k P of x k s k will be less than or equal to this quantity. And this quantity is nothing but, it implies that f of x k plus 1 is greater or equal to f of x k. So, that means the objective function value increases after the every iteration of optimization of the auxiliary functions.

(Refer Slide Time: 11:53)




Now, if x star is a solution to this problem, then f of x k is less than or equal to f of x star for every k. So, let us prove this which is very easy to show, now c k is a positive quantity P of x k is also a positive quantity, so for the f of x k is less than or equal to f x k plus c k into P of x k. And this is nothing but since s k is optimal for every feasible x, so this quantity is less than or equal f x star plus c k P of x star and by definition of penalty function P of star in 0. And therefore, this quantity is nothing f x star, so f of x k becomes less than or equal to f of x star and now using the 2 lemmas. We can prove the theorem that any limit point of the sequence x k generated by the penalty method is a solution to the problem P 1.

(Refer Slide Time: 12:53)

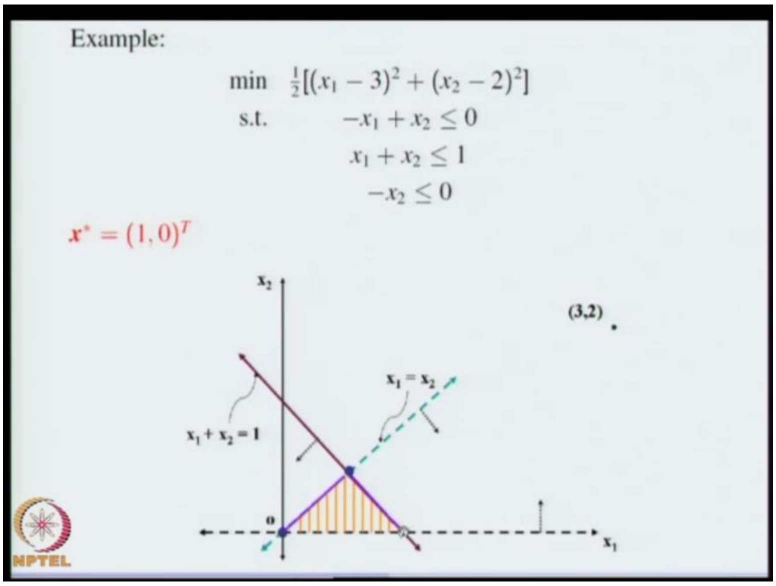



Now, let us see the penalty function method to solve general none linear program, now the input to this penalty function method is the sequence c k, where k goes from 0 to infinity and a positive epsilon value which is used for stopping the program. So, the iteration counter is set to 0, and the initial value x k is obtained. Now, how do we get this x k, it may be a good idea to get this x 0 or the initial point by solving an unconstrained minimization problem to minimize f x.

Because in this case we are not ensuring that every feasible point has to be in the interior of the feasible set. But, as the iterations progress, it will move towards the interior of the feasible set. So, we initialize the x k and while the difference between the auxiliary function and the actual objective function is greater than epsilon. We solve an unconstrained minimization problem to minimize u of x given c k. The interaction counter is increased by 1 and this procedure is repeated.

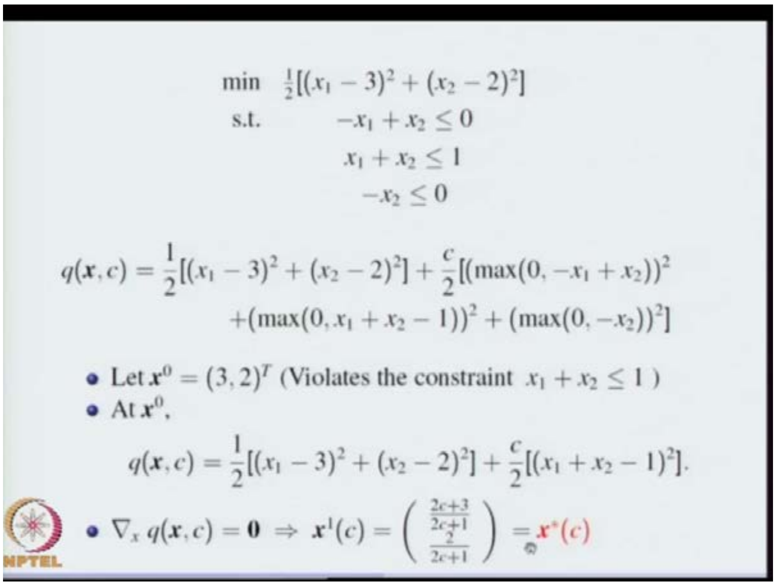
So, you will see that, what we need is just a sequence c k, which is a monotonically increasing sequence and which tends to infinity as a, tends to infinity. And every time, we need to solve only a constraint optimization. Unconstrained optimization problem, and finally, at the end we will get x star which is equal to x k. So, very easy way of solving, general non-linear program but, the only thing is that we need to define the auxiliary function appropriately.

(Refer Slide Time: 14:56)




So, let us see an example, we have seen this example earlier to minimize x 1 minus 3 square plus 2 minus 2 square subject to this constraints. And we have already seen that the minimum of this problem occurs at 1 0 or in other words you will see that if we had use the active set method and started with this point. We would followed this path to reach the solution whose x 1 co-ordinate is 1 and the x 2 co-ordinate is 0. So, it is the circle of minimum radius centered at 3 2 which touches this feasible region and the optimal point is this point.

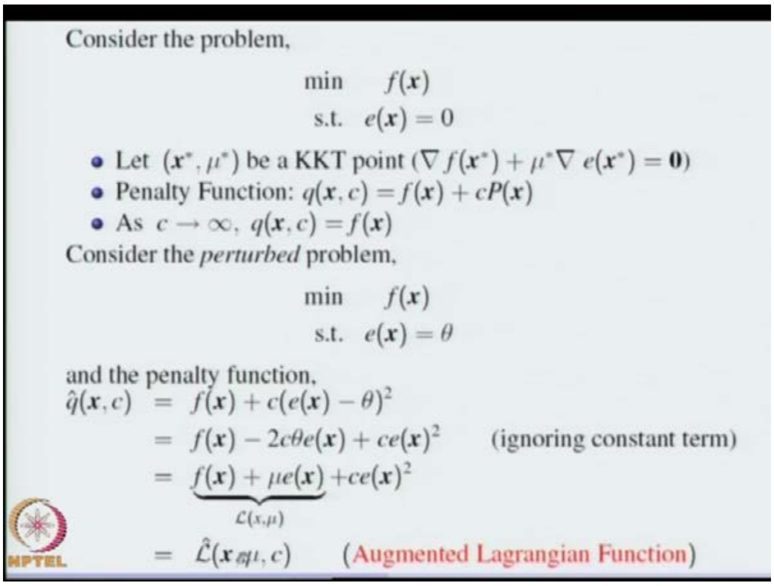
(Refer Slide Time: 15:48)




Now, let us see how to solve it using penalty function method, so we define the auxiliary function to be the objective function plus max of 0 coma h j x square. So, we have 3 constraints, so there will be 3 terms corresponding to the 3 inequality constraints, now we have to start with some point. So, we take the unconstraint minimization of this objective function, and that minimum occurs at point whose x 1 co ordinate is 3 and x 2 co-ordinate is 2. So, let us take us, take that as our initial point, now this initial point you can check that it violates the second constraint x 1 plus x 2 is less than or equal to 1. So, the other constraints are satisfied, so if we consider max of 0 minus x 1 plus x 2 that will be 0 and similarly, for the third constraint. So, only this constant will be active so let us rewrite our auxiliary function with respect 2 only the violating constraint.

So, at x 0 the auxiliary function will have a form like this, now for a fixed c this is the convex function. And if we equate the gradient of this function to 0 so we get x 1 c to be 2 c plus 3 by 2 c plus 1 and 2 by 2 c plus 1. So, since this is a convex function this is also x star c. And therefore, if we take the limit as c tends to infinity, because in the penalty function method we have to take c to infinity. So, if we take the limit of this as c tends to infinity what we get is start to be 1 0 as desired. So, you will see that penalty function method generates sequence of x k c, where the c also varies for every iteration and finally, in the limit we will approach x star. So, a lot depend on, the lot depends on taking the limit of this x k c as c tend to to infinity. Now, the problem is that as c approaches towards infinity, the optimization of this function will cause some numerical difficulties, as c approaches infinity, so is there a way out.

(Refer Slide Time: 18:50)

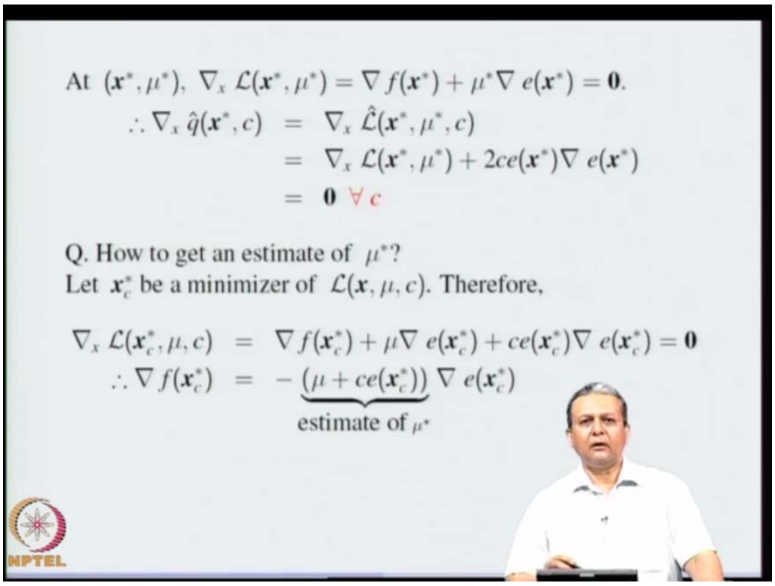



So, so let us consider a simple problem to minimize the f x subject to the equality constraint e x equal to 0. And we have seen that the first order necessary conditions in that there exist x star and mu star, which is the KKT point that is the gradient of f x star plus mu star into 2 gradient e x star equal to 0. Now, penalty function, in the penalty function method, the auxiliary function is defined as the some of f x and c into P x. And it depends on the optimal value depends on, c being taken to infinity. And as I said that at the c equal to infinity or and c moves towards infinity, there can be some numerical difficulties associated with, this optimization problem. Although at c tends to infinity q x c will be equal to f x, but optimization of q x c will become difficult.

Now, let us consider the per term problem, where instead of 0, let us take some constant theta, and minimize f x subject to constraint that e x equal to theta. Now, will this part of problem, help us in getting rid off our dependence on c, especially as c approaches infinity. So, let us consider the penalty function let us call it has q hat x c and that is nothing but, f x plus c into e x minus theta square. So, instead of f x plus c into x square we have a new penalty function, c into e x theta minus square. Now, if we expand this and ignore the constant term, what we get is f x minus 2 c theta into e x plus c into 2 e x square.

Now, if you consider minus 2 c into theta to be mu, then this can be written as f x plus mu into e x plus c into e x square. And f x plus mu into e x is nothing but the Lagrangian function using x and mu as the variables. And to that Lagrangian function there is a extra quantity, which is added and therefore, it is called the augmented Lagrangian function. And let us denote this augmented Lagrangian function by L hat now, this function L hat is a function of x mu and c.

(Refer Slide Time: 21:34)

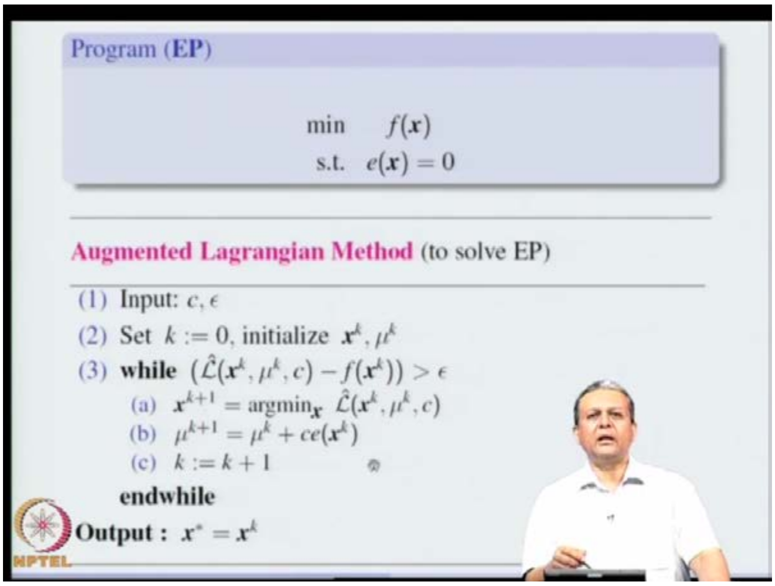



Now, we know that the, if x star is a local minimum, then there exists mu star, such that the gradient of l with respect to x evaluated at x star mu star is 0. And that is gradient f x star plus mu star into gradient e x star is 0. Now, if we take the gradient of the augmented lagraingian with respect to x. And evaluate it at x star mu star and c, so what we get is the gradient of the original Lagrangian with respect to x evaluated at x star mu star plus 2 c into e x star into gradient of e x star.

And quantity is 0, because x star mu star is a KKT point, and this quantity is 0, because x star is a feasible point. So, the important thing to note here, is that the gradient of the Lagrangian, the augmented lagarangian with respect to x evaluated at x star mu star at a given point of at the given value of c is 0 for every value of c. So, this is the important observation that this gradient vanishes irrespective of the value of c. So, the, we really do not have to worry about getting a sequence c and ensuring that c turns to infinity, which might cause numerical difficulties. Now, the question is that how do, we get this estimate mu star, now if x star let us denote by x star c, the minimizer of the Lagrangian.

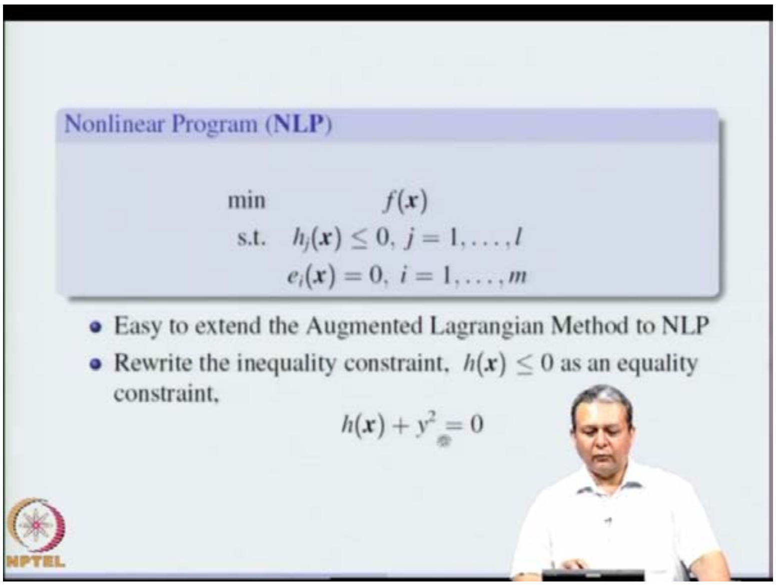
So, the gradient of the Lagrangian vanishes so that is expanded here, and these gives us gradient of f x star to be minus of mu plus c e x star c into 2 gradient d x star c. Now compare this with this equation, so we will see that the, this quantity which is in the parenthesis it is a good estimate of mu star. So, the idea is that every iteration we get an estimate of mu star and repeat the procedure.

(Refer Slide Time: 23:50)



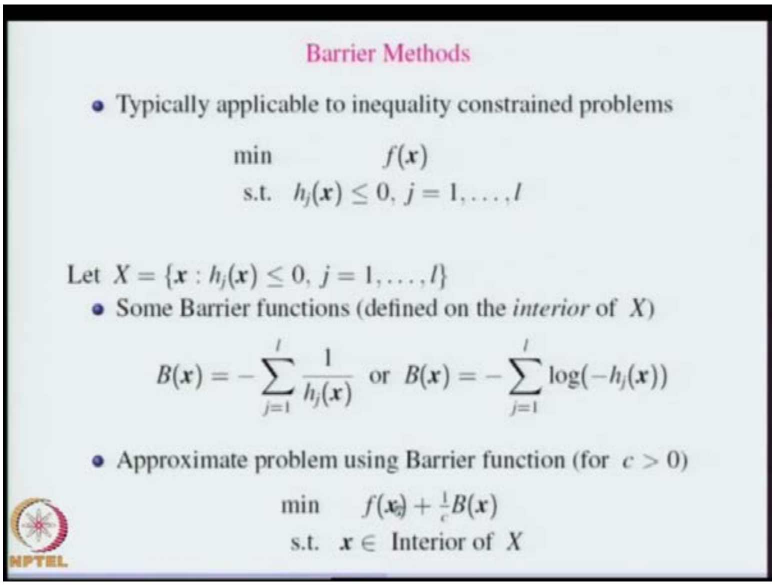

So, let us see the algorithm which is called the augmented Lagrangian method, so the problem is to minimize f x subject to x equal to 0. And you will see that the, at every iteration we solve unconstrained optimization problem to minimize L hat and then update mu update mu k plus 1 using mu k and e of x k. So, what you need is only some input number some, some number c which is positive number and epsilon which is the tolerance parameter for stopping. So, this algorithm becomes very simple, and that can be used to solve.

(Refer Slide Time: 24:34)




General non-linear program also, if we have multiple equality constraints, so there will be a mu's associated mu star associated with each of the constraint. So, those can be estimated easily and an inequality constraint can be written as h x plus y square is equal to 0 by introducing new variable y. So, once we have equality constraint problem again the augmented Lagrangian method can be used.

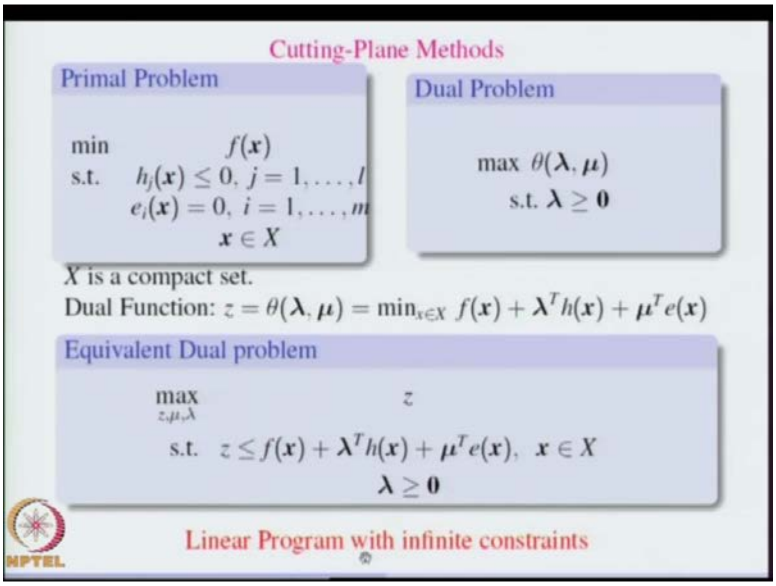
(Refer Slide Time: 25:01)




Now, similar to penalty methods there exists barrier methods, so which are typically applicable to inequality constraint problem. Now, consider this inequality constraint problem and let us denote the set constraint, set by the set x. Now, different bearer functions are used, so for example, one can use barrier functions of the type minus 1 over h j x or minus log of minus h j x. Now, the idea is that approximate the given problem using the barer function. So, if we choose the constant c which is greater than 0 and write the barrier function, write the auxiliary function as f x plus 1 over c b x.

Remember that c is greater than 0 and if we have a sequence c which tends to infinity, and then solve this auxiliary problem to minimize f x plus 1 over c b x subject to x belongs 2 by integral of x. Now, on the face it you may find that, this is the constraint problem and which is difficult to solve, but if you really take the gradient of the objective function. You will see that this constraint is typically not necessary, because this gradient of the objective function. Especially the barrier function will make sure that the point does not cross the feasible region. So, the new point will always been the integer of the feasible region and the, if carefully implemented this function, this method also works very well. The analysis of this method is similar to the penalty function method, so we will not repeat it here

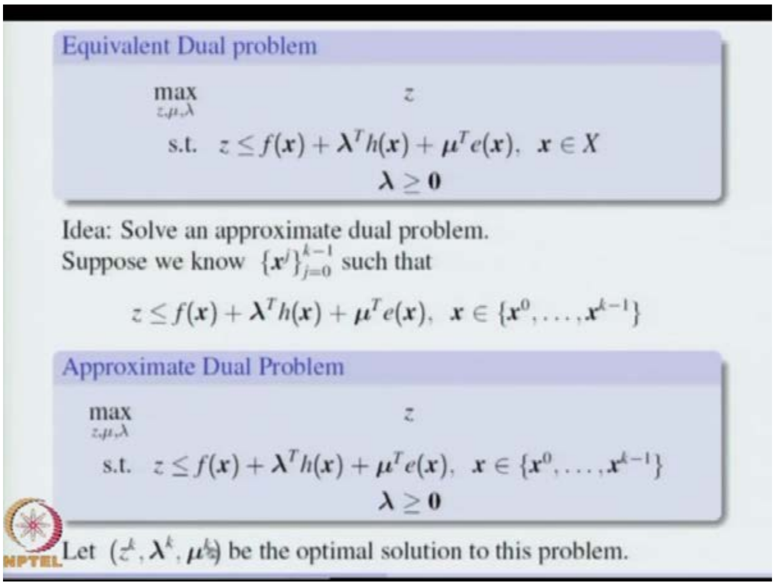
(Refer Slide Time: 26:50)




Now there exists a, another class of methods, which also depend on the Lagrangian dual these are called the cutting plane methods. So, let us see the general non-linear program to minimize the f x subject to the inequality and equality constraints and x belongs to the set x where x is a compact set. And the dual problem of this problem is maximize theta lambda mu where lambda is nonnegative, so lambda are the Lagrangian multipliers corresponding to the inequality constraints. Now, the dual function theta lambda mu is nothing but, minimum in fact it should be infimum, but since x is a compact set. We are assuming that, there exists a minimum, so minimize f x plus lambda transpose h x plus mu transpose a x and let us denote this as z.

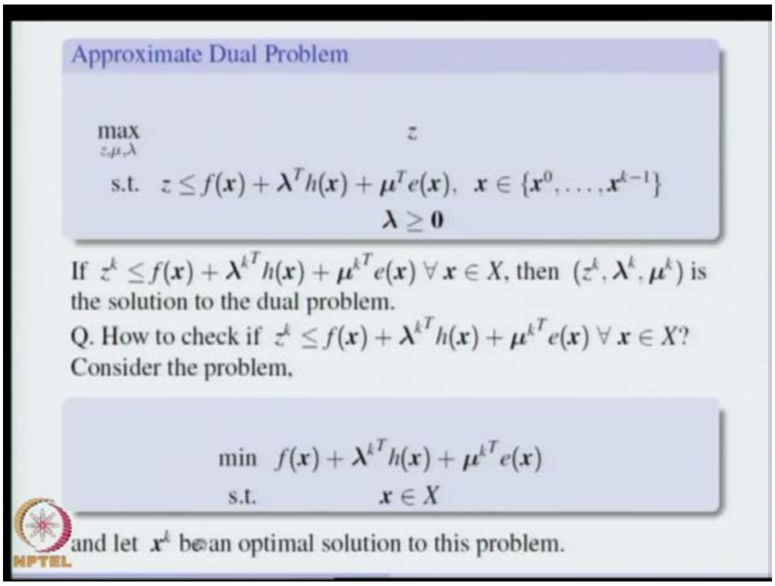
So, the equivalent dual problem can be written as maximize z subject to z less than or equal to this quantity and x belongs to z x. Now, you will see that this problem is a linear programming problem, so maximize the linear function subject to this, constraints and the variables are z mu and lambda. Now, the only problem, with this linear program, that the number of constraints is infinite. Because we have to make sure that x belongs to the set z x so such problem is difficult to solve.

(Refer Slide Time: 28:43)



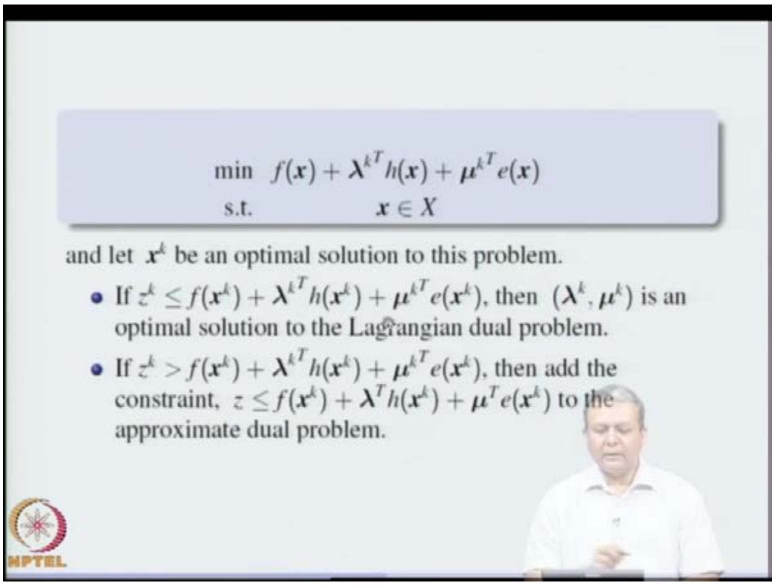

So, instead may be a good idea to solve an approximate dual problem, now suppose we know that there is the sequence of x j is going from 0 to k minus 1. Such that z satisfies the inequality of all x in this sequence, then we can write an approximate dual program as to maximize j x subject to this constraint. But x belongs only to x 0 to 0 x k minus 1, so let us assume that z k lambda k mu k is an optimal solution to this problem. Now, we have to check whether this is an optimal solution to the original problem or not.

(Refer Slide Time: 29:31)



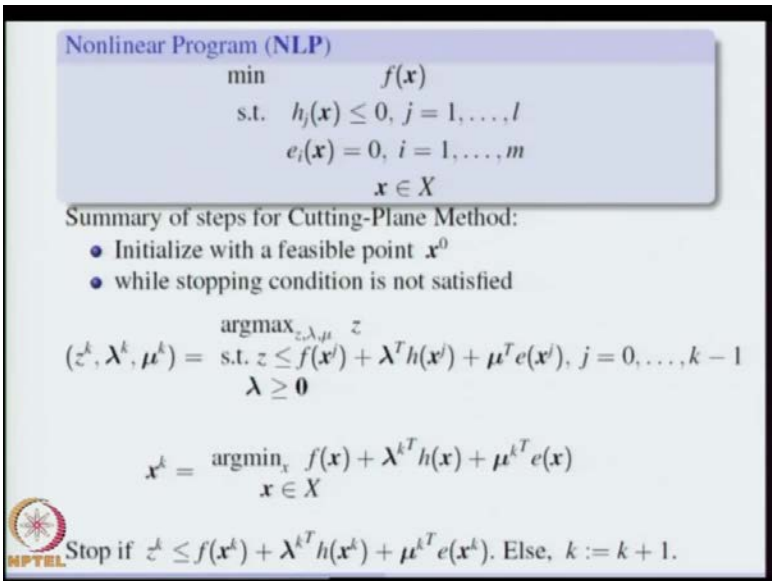

And one way to check this, is to consider the problem to minimize so fix lambda k and mu x and minimize f x plus lambda k transpose h x plus mu k transpose e x subject to x belongs to x. And get an solution x k and at x k you find out this objective function this value and, then see whether z k is less than or equal to that.

(Refer Slide Time: 29:58)




So, if z k is less than or equal to that quantity then lambda k mu k is an optimal solution to the Lagrangian dual problem. And if z k is greater than that f of x k plus lambda k transpose h x k plus mu k transpose a x k. Then this constraint is added to the approximate dual problem, so for the next iteration, there will be a extra x k which would appear in the approximate dual problem.

(Refer Slide Time: 30:29)




So, for a general non-linear program the cutting plane method works like this that we initialize with a feasible point x 0. Remember that all the cutting plane methods start with the initial feasible point x 0, and then solve a linear program to get z k lambda k mu k. But this linear program solves the approximate dual problem were the x is take the value form the set x 0 x 1 to x k minus 1. Now, after having obtain the, this solution we fix lambda k mu k, and find out x which minimizes f x plus lambda k transpose h x plus mu k transpose e x, subject to the constraint at x belongs to x. And then check whether this, the z k that we have got satisfies this, so if z k is less than or equal to f x k plus lambda k transpose h x k plus mu k transpose a x k.

Then we stop, otherwise go to the next iteration. So this cutting plane method tries to add extra constraints, extra affined constraints as and when needed to improve the objective function; and it is very popularly used in many applications. Now this completes our discussion on some of the constraint optimization algorithms. As I mentioned earlier it may be difficult to cover each and every method in this course. So, we have tried here to select some of the commonly used methods, and solve and use them to solve constraint optimization problems.