Essential Mathematics for Machine Learning

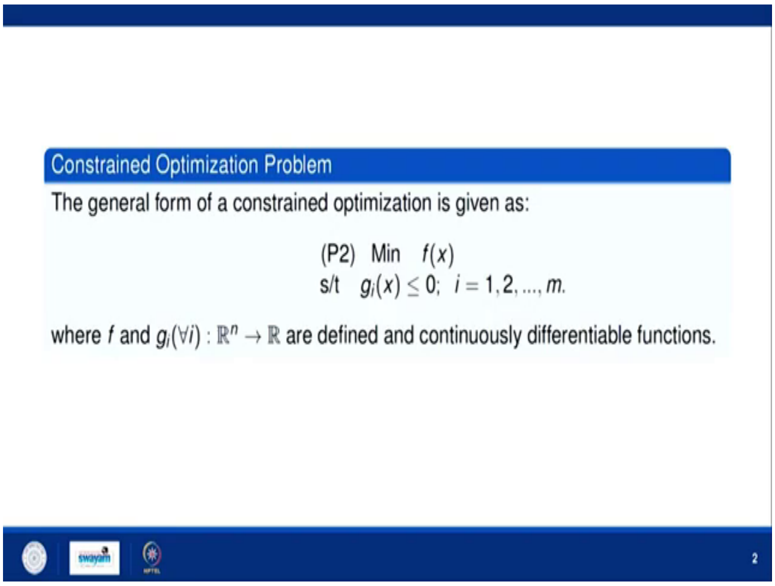
Prof. S. K. Gupta
Department of Mathematics
Indian Institute of Technology, Roorkee

Lecture - 28
Constrained Optimization - II

Hello friends, welcome to lecture series on Essential Mathematics for Machine Learning. In the last lecture we have seen that if we have a if we are having a Constrained Optimization problem in which we have to optimize a function subject to certain set of conditions or constraints, then that problem is our quadratic programming problem if the objective function is quadratic and all constraints are linear. That will be a convex programming problem if objective function is convex and all constraints are convex.

And we have solved few problems based on this also in the previous lecture. In this lecture we will see that if a general non-linear problem is given to you then how we can solve that problem. If it is a convex programming problem? Ok.

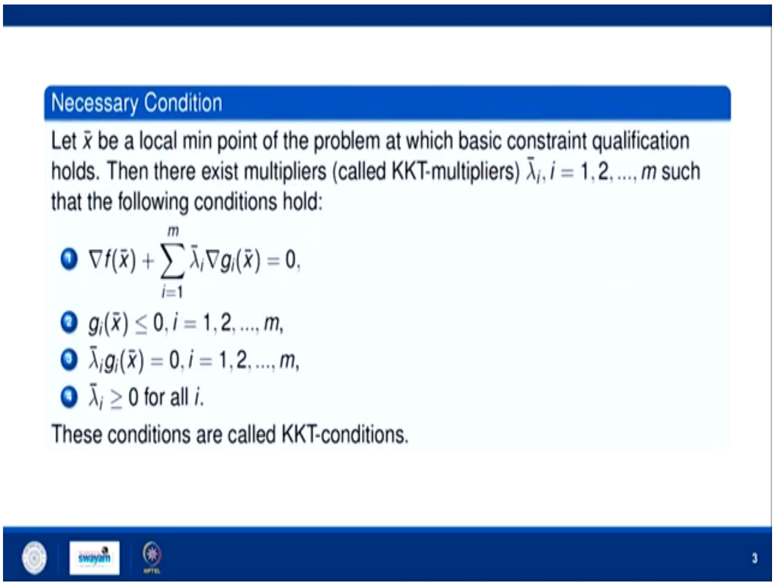
(Refer Slide Time: 01:16)



So, this is the general constrained optimization problem which we have already discussed, we in which we have to minimize a function f which is objective functions subject to g i x less than equal to 0 i from 1 to m. So, there are m number of constraints ok.

So, here we are assuming that f and g i, for all i are defined and continuously differentiable functions ok.

(Refer Slide Time: 01:39)



Now the necessary condition. The necessary condition means that suppose a point x bar, suppose x bar is a point which is feasible. Feasible means satisfy all the constraints and is a local optimal point for this problem P 2. Then what are the conditions that x bar will satisfy? So, those conditions are called necessary conditions ok. And here these conditions are called Karush Kuhn Tucker conditions KKT conditions ok. And a point with satisfy all these conditions is called KKT point.

So, what is a theorem let us read out the statement. Let x bar be a local minimum point of the problem P 2 at which the basic constraint qualification holds. What do you mean by basic constraint qualification? Basically these are the additional requirement which are required in the problem, so that the multiplier corresponding to the objective function is nonzero ok.

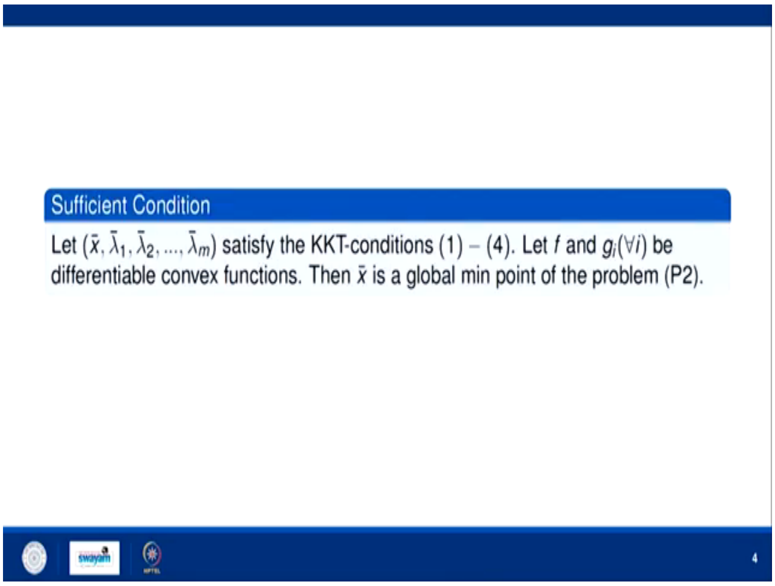
We want to multiplier. Here it is one. This one is coming only because we are applied some constraint qualification on the problem P 2. So, basically constraint qualification means that these are the additional requirement on the problem P 2, so that the Lagrange multiplier corresponding to the objective function is nonzero ok. So, holds.

Then there exist multipliers also called KKT multipliers. KKT means Karush Kuhn Tucker multipliers, lambda i bar i from 1 to m such that the following conditions hold. So, these are the following conditions which must be hold and these conditions are called KKT conditions.

So, if x bar is a local minimum point of this problem P 2 of this problem P 2 then these conditions hold, but these conditions are only necessary condition, these may not be sufficient. So, what the additional requirement?

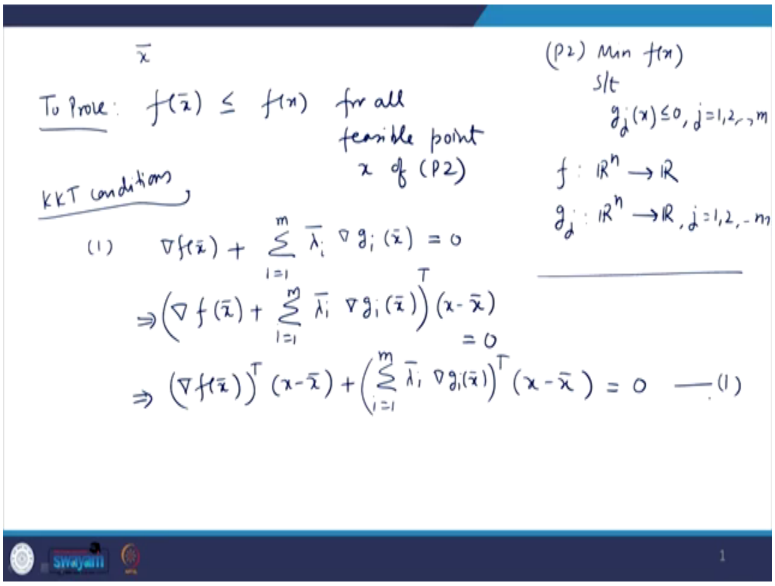
What is the additional assumptions which must be required on f or g i's, so that these condition becomes sufficient. Sufficient means, that if you solve these conditions, if you solve a point, if you solve these inequalities and a point x bar will satisfy these inequalities that will be a local optimal point of the problem P 2.

(Refer Slide Time: 04:18)



So, those conditions are basically if we called as sufficient conditions and the additional assumption is, that f and g i for all i must be a differentiable convex functions. So, basically is a problem P 2 is a convex programming problem then these conditions, then a point we satisfy these conditions will be a local optimal point or the problem P 2. That means, that these conditions become sufficient. Then x bar is basically not only local, but it will be a global minimum point of the problem P 2.

(Refer Slide Time: 05:02)



So, the proof is very easy. See if you want to show the proof. See what are the KKT conditions? What is the problem P 2? The problem P 2 is; problem P 2 is basically minimization of f x subject to g j x less than equal to 0 j from 1 to m ok. f is a function from R n to R and g j is other function from R n to R j from 1 to m ok.

Now, we have to show that a point which is a KKT point, that means satisfy the KKT conditions Karush Kuhn Tucker conditions, then the point will be a global minimum point of the problem P 2 and of course, problem is convex programming problem. So, let us try to prove this result, the proof is easy.

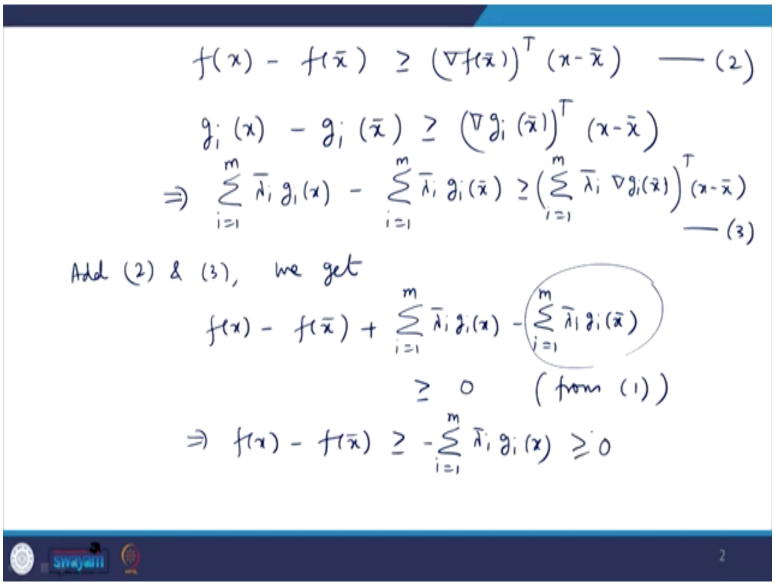
See if we want to show that x bar is a global minimum point of the problem P 2 then what we have to show? We will have to show that f x bar is less than equal to f x for all feasible point x of P 2 this to prove. Because we have to show that x bar is a global minimum point of the problem P 2. That means, you take any feasible point and for any feasible point of this problem f x bar should be less than equal to f x.

And given it is given to us that KKT conditions are satisfied and the problem is a convex programming problem. So, what are what are the KKT conditions?. So, let us go to the first KKT condition. The first KKT condition is gradient of f x bar plus summation i from 1 to m lambda i bar it is i lambda i bar gradient of g i x bar is equal to 0; this is the first condition.

So, now this implies. So, we know that x bar satisfy this condition ok. So, this implies we can use this as gradient x bar plus summation i from 1 to m lambda i bar gradient of g i x bar. So, if it is 0 then this will also be 0 ok. We are access any feasible point of the problem P 2. Now this further implies gradient f x bar whole transpose x minus x bar plus summation i from 1 to m lambda i bar gradient g i x bar whole transpose into, so into x minus x bar should be equal to 0 ok.

Now, since function and constraints are convex, suppose this is equal to this. This is equation 1.

(Refer Slide Time: 08:29)



Now, since f and g i for all i are convex. So, by the convexity definition we can say that f of x minus f x bar is greater than or equal to gradient of f x bar into x minus x bar. And also g i x minus g i x bar is greater than equal to gradient of g i x bar whole transpose into x minus x bar.

And this further implies, this constraint further implies this inequality further implies. Now since lambda i bar, see since this lambda i bar is greater than equal to 0 for all i. So, we can say, we can multiply this by lambda i and sum up all the constraints.

So, what we get; i from 1 to m lambda i bar g i x minus summation i from 1 to m lambda i bar g i x bar is greater than equal to summation i from 1 to m lambda i bar gradient of g i x bar whole transpose x minus x bar, suppose this is 3 ok.

Now you add 2 and 3 2 and 3 add 2 and 3, so what we get? We left with f x minus f x bar plus, now this is summation i from 1 to m lambda i bar g i x minus summation i from 1 to m lambda i bar g i x bar is greater than equal to.

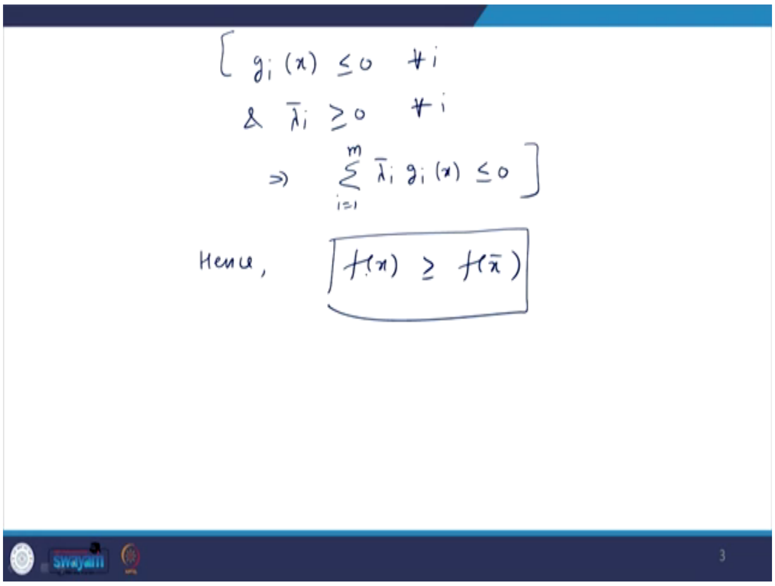
So, what we are having in the left hand side? In the left hand side we are having this transpose x minus x bar plus this transpose x minus x bar which from one, from this one is equal to 0 ok. So, we can directly write this is greater than equal to 0 from 1 ok.

Now again you go to KKT conditions lambda i g i x bar is equal to 0 for all i. If it is 0 for all i; that means, sum is also 0. So, sum is also 0 that means, this quantity this quantity is equal to 0 ok. Because it is what? It is lambda 1 g lambda 1 bar g i g 1 x bar lambda 2 bar g 2 x bar and so on up to lambda m bar g m x bar and all are 0. So, sum will also be 0.

So, we left with this finally, gives f x minus f x bar is greater than equal to summation i from 1 to m with negative sign lambda i bar g i x. Now what this x is? This x is a feasible point of this problem P 2. So, if it is a feasible point of this problem P 2 that means, this constraint are is are satisfied for the for x, and lambda i bar are greater than equal to 0 for all i.

So, that means, that this expression is less than equal to 0 and with negative sign this will be greater than equal to 0 ok. Because g i x is less than equal to 0 for all i and lambda i bar are greater than equal to 0 for all i. So, this implies summation lambda i bar g i x are less than equal to 0 ok.

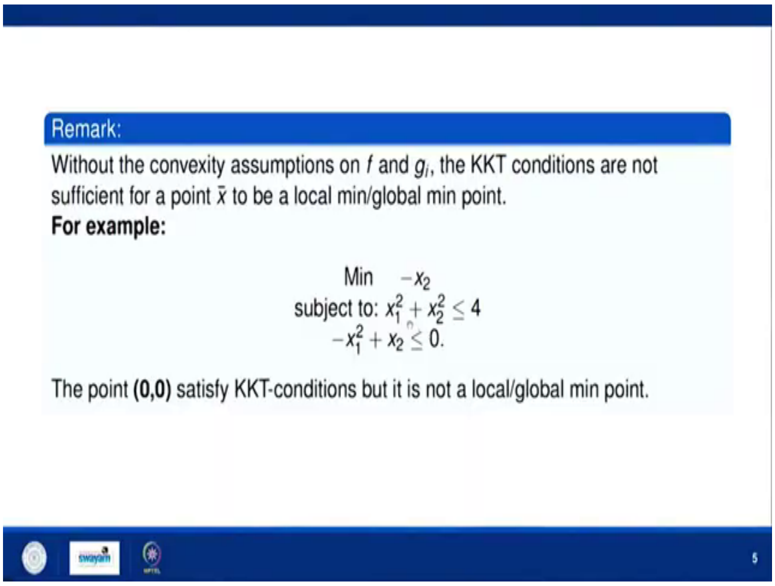
(Refer Slide Time: 12:03)



So, hence this is less than equal to 0 negative will give greater than equal to 0. So, what this implies? This implies. So, I am putting it in a bracket, so hence f x is greater than equal to f x bar for any x feasible to the problem P 2 and this implies x bar is a global minimum point of the problem P 2.

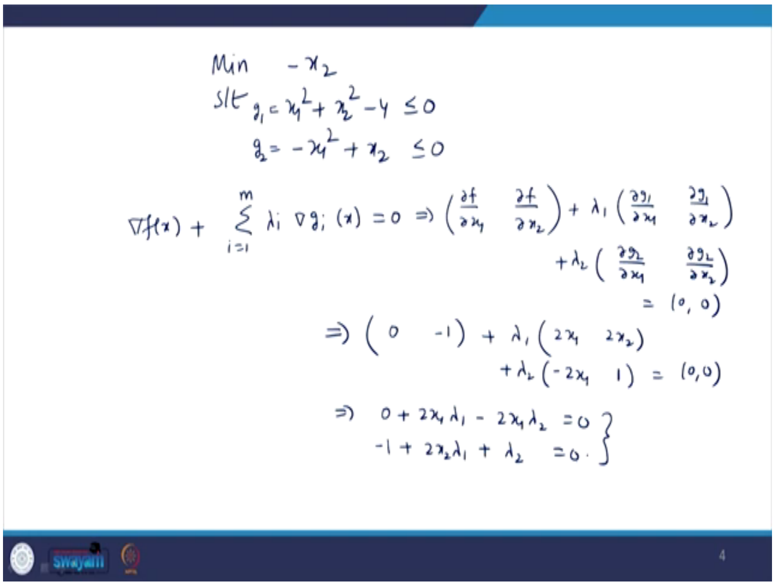
So, what we have shown we have shown that if a point is a KKT point that means, satisfy the KKT conditions and the problem is a convex programming problem then that point is the global minimum point to the problem P 2. Now convexity is a important condition, without convexity we may not say that that point is a global minimum point of the problem.

(Refer Slide Time: 13:21)



Suppose you have this problem. To understand this let us take this problem. This is minimization of minus x 2 subject to these two constraints.

(Refer Slide Time: 13:30)



So, let us see what this problem is, it is minimum of minus x 2 subject to subject to it is x 1 square plus x 2 square minus 4 less than equal to 0 and second is minus x 1 square plus x 2 less than equal to 0, this problem

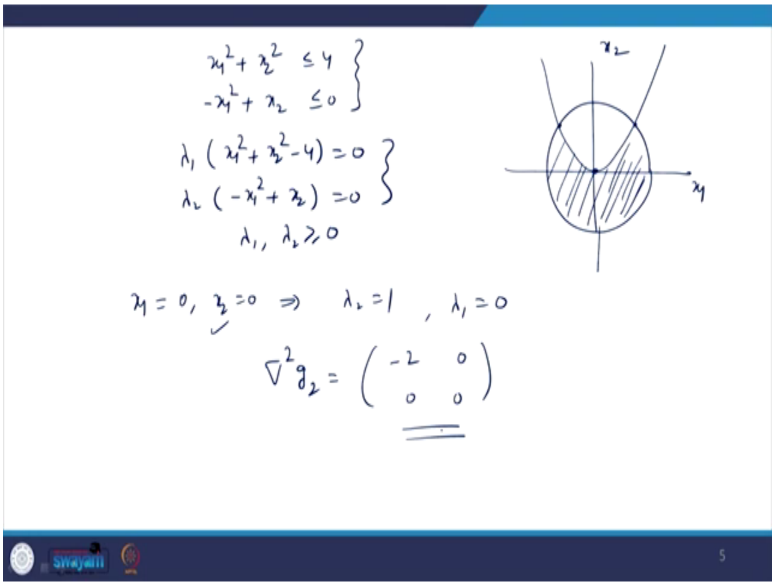
Now, let us write the KKT condition of this. So, how to write the KKT condition? The first condition is gradient of f x plus summation lambda i i from 1 to m gradient of g i x should be 0. This implies, this is g 1 this is g 2. So, gradient of how many variables? 2 variables. So, what is del f, so square is not there.

So, gradient of f x. So, first with respect to x 1 it will be del f by del x 1 del f by del x 2 plus plus will come here actually, lambda 1 times gradient of g 1, so del g 1 by del x 1 plus del. So, this is not plus this is a vector. So, this is del g 1 by del x 2 plus lambda 2 times del g 2 by del x 1 and del g 2 by del x 2 which is equal to 0 0. This is by the first constraint, first KKT condition.

Now, this further implies, why del f by del x 1 0, del f by del x 2 minus 1 plus lambda 1 times, del f del g 1 by del x 1 2 x 1 it is 2 x 2 plus lambda 2 times minus 2 x 1 and g 2 is 1 which is equal to 0 0. And this implies 0 plus 2 x 1 lambda 1 minus 2 x 1 lambda 2 is equal to 0; the first constraint first equation. The second equation is minus 1 plus 2 x 2 lambda 1 plus lambda 2 equal to 0.

So, these are the two constraint which we got from the first KKT condition.

(Refer Slide Time: 16:01)



What are other KKT conditions let us quickly see for this problem. The other KKT condition is the first, the second one is the feasibility condition that is x 1 square plus x 2 square less than equal to 4 and minus x 1 square plus x 2 less than equal to 0. This is a feasibility condition.

And next condition is a complimentary condition, that condition is called complimentary condition lambda i g i x equal to 0. So, lambda 1 g 1; g 1 is this condition equal to 0 and lambda 2 g 2 is this condition equal to 0 and lambda 1 lambda 2 greater than equal to 0. So, these are the various KKT condition for this particular problem.

Now, let us check whether 0 0 is a KKT point or not. KKT point means it satisfies the KKT conditions or not. So, if you put x 1 0 x 2 0, so 0 0 equal to 0. If it is 0 then lambda 2 comes out to be 1. So, lambda 2 is 1. So, we I am putting x 1 equal to 0 and x 2 equal to 0. So, this implies lambda 2 is 1.

So, 0 plus 0 less than 4; true, 0 0 less than equal to 0; true, 0 if it is 0 0; that means, it nonzero so; that means, lambda 1 is 0. And 0 0 this is 0, this is 1 and 0 0 equal to 0. So, and lambda 1 lambda 2 are also non negative. So, we have shown that 0 0 satisfy all the KKT conditions, but still 0 0 for this problem is not a global minimum point. What is the reason? The reason is that this problem is not a convex programming problem.

Why? Because you can see if you draw the graph of this problem, the graph of this problem is you see x 1 square plus x 2 square less than equal to 4, this is x 1, this is x 2. Less than equal to 4 means inside the circle and minus x 1 square plus x 2 square less than equal to 0 means this parabola and outside this parabola; that means, in inside the circle there is this region.

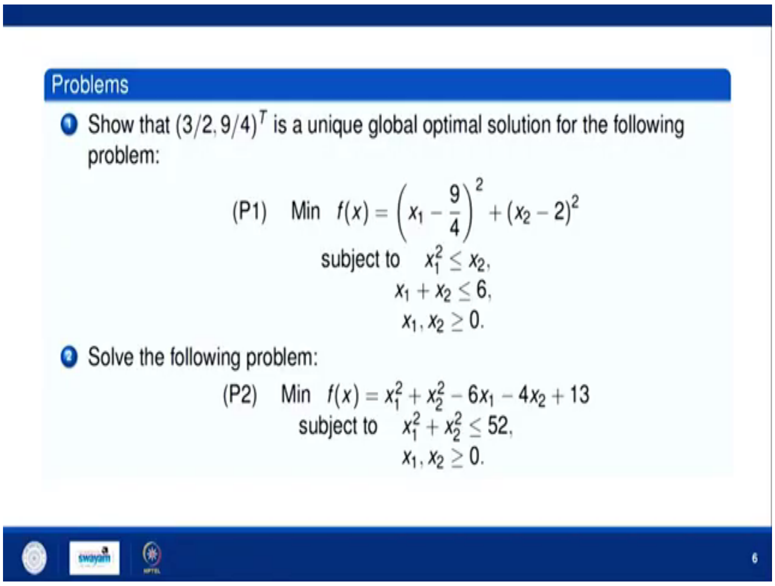
So, 0 0 is clearly not a global minimum point of this objective function minus x 2. We are you have to minimize minus x 2, you have to minimize minus x 2 means you have to maximize x 2. So, maximize x 2 will be where, maximize x 2 will be at this point or at this point.

So, this point clearly will not be a global minimum point of this function; however, it is a KKT point and the reason why? The reason is, because this feasible region is not a convex set.

Or you can say that this is all if you find the Hessian of the second constraint, Hessian of a second constraint, Hessian of a second constraint del square of g 2 is minus 2 0 0 0. So, this Hessian is not positive semi definite. That means, the constraint g 2 is not a convex function. That means, this problem is not a convex programming problem.

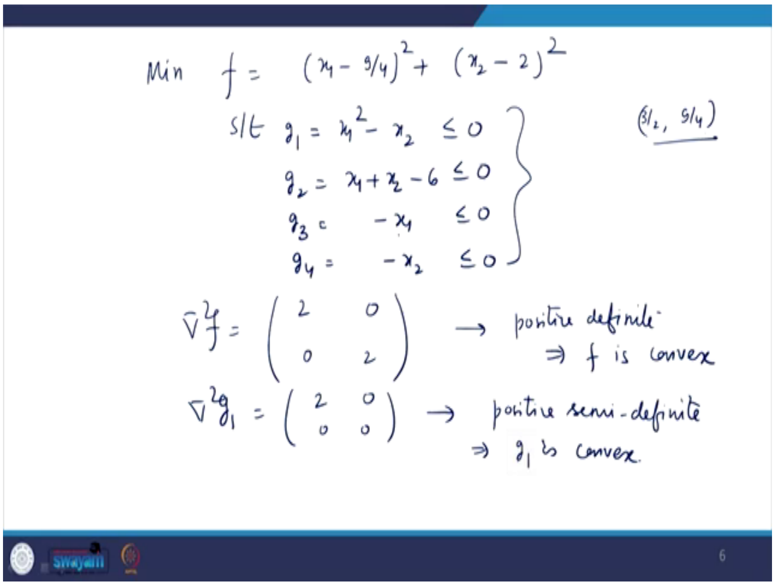
So, what we have concluded? We have concluded that KKT point will be a global minimum point of the problem P 2 only when a given problem, a given p problem P 2 is a convex programming problem. So, that is a sufficient condition.

(Refer Slide Time: 19:38)



So, here are few examples based on KKT condition, let us discuss one problem say problem 1. So, we have to show that 3 by 2 9 by 4 a unique global minimum global optimal solution the following problem. So, for first we have to show that the problem is a convex programming problem, then we have to write all the KKT conditions and then we have to show that this point 3 by 2 9 by 4 satisfy all the KKT conditions. This is sufficient to show that 3 by 2 9 by 4 is a global optimal solution of this problem.

(Refer Slide Time: 20:22)

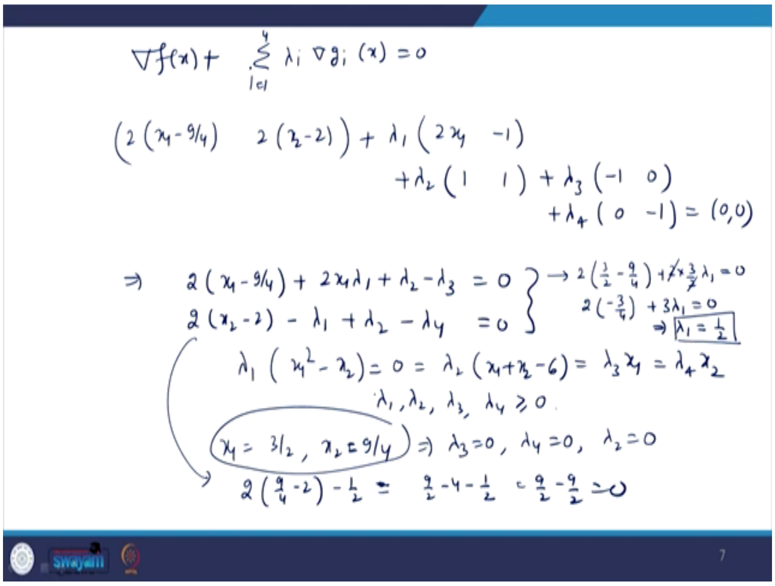


So, let us try this problem, its in the same way we can try for this problem solve this problem also. Problem is minimizing f equal to x 1 minus 9 by 4 whole square plus x 2 minus 2 whole square subject to. What are the conditions? The first condition is x 1 square minus x 2 less than equal to 0, suppose this is g 1. The second constraint is x 1 plus x 2 less than equal to 6. The third constraint is minus x 1 less than equal to 0. The fourth constraint is minus x 2 less than equal to 0.

Now, g 2 is linear so convex, g 3 is linear so convex and g 4 is linear so convex. Let us see for f and g 1. So, what are Hessian matrix of f? It is 2 0 0 2. So, this matrix is clearly positive definite matrix and hence convex. So, f is this, this is positive definite and this implies f is convex.

Now, similarly find greedy Hessian matrix of g 1, it is 2 0 0 0 which is again, which is positive semi definite. What we have seen? So, hence we can say that this problem is a convex programming problem. Now let us write the KKT conditions. So, what are the various KKT conditions?.

(Refer Slide Time: 22:10)



So, KKT conditions of the first KKT condition first KKT condition is gradient of f x plus summation lambda i gradient of g i x equal to 0. Here i is from 1 to 4, because there are four constraints. So, what is gradient of f x? Gradient of f x will be, you find the gradient of this f, so gradient will be twice x 1 minus 9 by 4 twice x 2 minus 2 plus lambda 1 times gradient of first constraint, which is 2 x 1 for x 1 and minus 1 for x 2 plus lambda 2. For lambda 2 it is 1 1, for lambda 3 it is minus 1 0, for lambda 4 it is 0 minus 1 is equal to 0 0.

So, this implies twice of x 1 minus 9 by 4 plus 2 x 1 lambda 1 plus lambda 2 minus lambda 3 equal to 0. The second will be twice of x 2 minus 2 minus lambda 1 plus lambda 2 minus lambda 4 equal to 0.

So, these are the two conditions which you obtain from the first KKT condition. Next are the feasibility conditions. So, feasibility conditions are the same conditions, these are the feasibility conditions.

Next are the complimentary condition. The complimentary condition is lambda 1 time the first constraint the first constraint is x 1 square minus x 2, this is equal to 0 which is equal to this is equal to 0 which is equal to lambda 2 times x 1 plus x 2 minus 6 which is equals to lambda 3 x 1 which is equals to lambda 4 x 2 and lambda 1, lambda 2, lambda 3, lambda 4 all must be greater than equal to 0.

Now, now what to show? We have to show that this point 3 by 2 9 by 4 is the global optimal point of this problem. That means, it must satisfy the KKT conditions ok. So, x 1 is 3 by 2 and x 2 is 9 by 4. So, you substitute first you substitute it here. So, what is x 1 square?

So, first you first you see whether it is feasible or not ok. So, first we will see the feasibility. So, 3 by 2 and 9 by 4, 3 by 2 and 9 by 4. So, you substitute 3 by 2 is greater than equal to 0, 9 by 4 is greater than equal to 0, 3 by 2 plus 9 by 4 is what? It is 6 by 4 plus, 6 by 4 means 15 by 4 minus 6 which is less than equal to 0 that is true. And x 1 square minus x 2, 9 by 4 minus 9 by 4 is 0 which is equal to 0. So; that means, first thing is clear that this point is a feasible point.

Now, the other conditions. Now this is x 1 is non 0 and lambda 3 into x 1 is equal to 0, so this implies lambda 3 equal to 0 and similarly lambda 4 is also 0. And x 1 is 1 3 by 2 and x 2 is 9 by 4. So, and this is this sum will be 15 by 4 which is not 6; that means, this is not 0. That means, lambda 2 will come out to be 0. From this we cannot say anything, lambda 1 may or may not 0 because this is 0.

So, these three we have obtained. Now lambda 3 is 0, lambda 2 is 0 and x 1 is 3 by 2. So, from here you can calculate lambda 1. So, what will be lambda 1 from this condition? It is 2 into 3 by 2 minus 9 by 4 minus 2 into 3 by 2 plus 2 into 3 by 2 into lambda 1. These two are 0 which is equal to 0.

So, what we have obtained from here? So, these two will cancel with this two. So, this is nothing, but 2 this is 6 by 4 that is minus 3 by 4 plus 3 lambda 1 equal to 0. So, this implies lambda 1 is equal to. So, 3 3 cancels out 1 by 2 ok. Now only one condition left; lambda 1 is also non negative only one condition to check. So, 9 by 4 minus 2. So, let us compute it here.

So, 2 into 9 by 4 minus 2, lambda 1 is 1 by 2, lambda 2 is 0, lambda 4 is 0. So, this is 0, this must be 0. So, let us check, so this is equal to what? This is 9 by 2 minus 4 minus 1 by 2. So, this is again 9 by 2 the 9 by 2 minus 9 by 2 which is 0. That means, this condition is also satisfied. Hence we can say that this point satisfy all the KKT conditions.

So, hence by the sufficiency KKT conditions we can say that this point is a global minimum point of this problem ok. In the same way we can solve the problem P 2 also. You write the KKT, you first check whether its a convex programming problem or not. If its a convex programming problem then you write the KKT conditions and try to find a point x which satisfy all the KKT conditions, that will be the global minimum point to the problem P 2.

So, in this lecture we have seen that if we have given a non-linear programming problem and it is a convex programming problem. Then to solve such problems you just write the KKT conditions ok. Try to solve the KKT condition and a point will satisfy the KKT conditions will be a global minimum point of the problem if it is a convex programming problem.

Thank you very much.