Essential Mathematics for Machine Learning

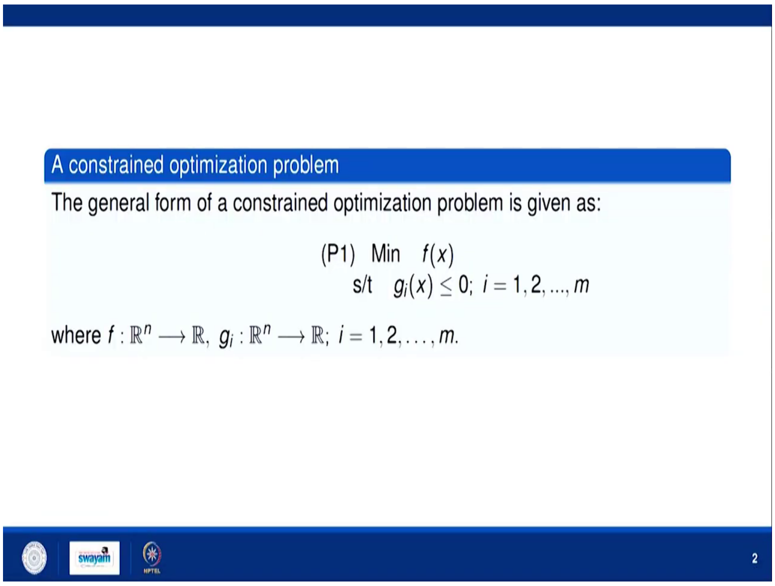
Prof. S. K. Gupta
Department of Mathematics
Indian Institute of Technology, Roorkee

Lecture - 27
Constrained Optimization - I

Hello friends. So, welcome to lecture series on Essential Mathematics for Machine Learning. In the previous lectures we have seen some of the properties of convex functions. We have seen that if Hessian matrix is positive semi definite. This implies that the function is convex on s. We have also seen about unconstrained optimization problem that how can we find the necessary and sufficient conditions for a local minimum for a given unconstrained optimization problem.

Now, what happens if a constraint is added to a optimization problem? I mean; I want to say that if it is a constrained optimization problem. What are the necessary or sufficient condition for constrained optimization problems?

(Refer Slide Time: 01:18)



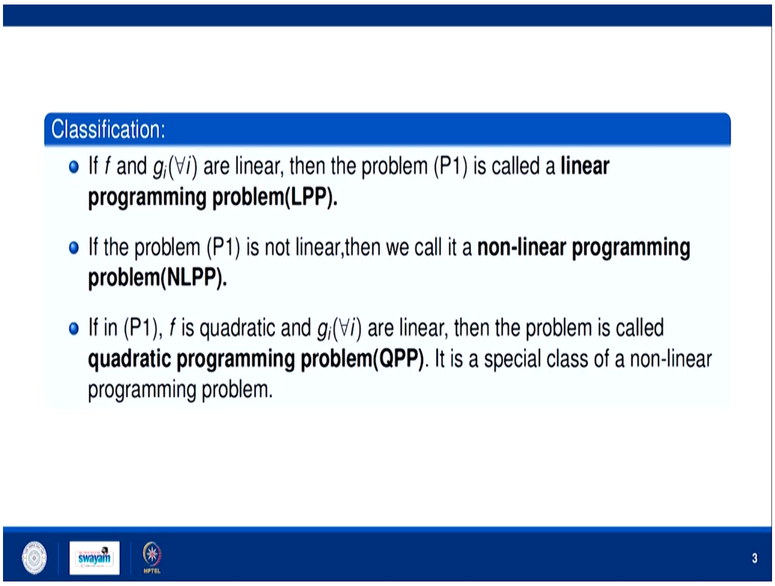
So, before going to the necessary or sufficient condition for constrained optimization problems. Let us see some of the basic concepts of constrained optimization. So, first of all what a constrained optimization problem is. A constrained optimization problem is basically a problem in which we have to maximize or minimize a given function called objective function which is f here subject to the given set of constraineds.

Here we are having minimum of f x it may be maximum also here I am taking as minimum. The same algorithm will be followed for maximum type constrained optimization problem as well.

So, if it is minimum f x subject to g i x less than equal to 0. So, this f is basically objective function subject to m number of constraineds. So, this problem P 1 here is called a constrained optimization problem ok. Now there are different type of constrained optimization problem this problem P 1 which is here depending on the choice of the function f depending on the choice of g i is it may be classified different categories.

The first of all the simplest one or the easiest one problem is linear programming problem which we have already studied in our previous classes. What are linear programming problem is? If this f and all constraineds are linear, then such problems are called linear programming problem ok.

(Refer Slide Time: 02:41)



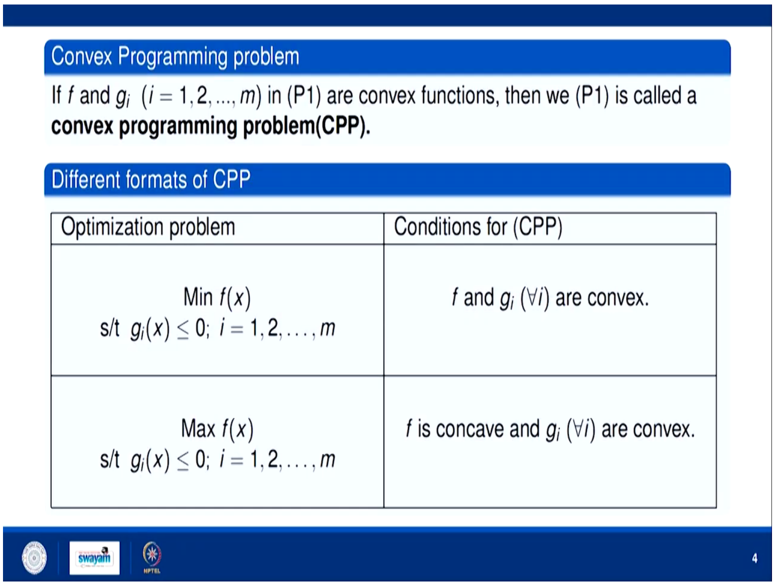
So, the first classification the simplest one is a linear programming problem if f and g i for all i are linear. Now if it is not linear; that means, that the objective function or at least one of the constrained is non-linear, then such problems are called non-linear programming problem or non-linear optimization problem ok.

Now, there is a special class of non-linear optimization problem and that special class is called quadratic programming problem. What happens in quadratic programming problem?

In this problem in this problem P 1; if this f x is quadratic that is of degree 2 power 2 variables have maximum power 2 and all constraineds are linear, then such problem is called such problems are called quadratic programming problem. It is also a non-linear programming problem with a special feature the special feature is that the objective function is quadratic and all constraineds are linear ok.

So, basically, if we are roughly speaking we are having two classes of two classification of such type of problem; one is linear programming that is non-linear and there is a special class of non-linear that is called quadratic programming problem ok.

(Refer Slide Time: 04:01)



Now what is convex programming problem? Now this problem P 1 which we have just discussed this problem P 1 is called convex programming problem. If the involved objective function f as well as all the constraineds all the m constraineds here are convex; that means, if f and all g i's are convex then such problems are called convex programming problem ok. Now there are different formats of convex optimization problem or convex programming problem what are different formats see.

If this is minimization type f x subject to g i x less than or equal to 0 i from 1 to m which is P 1 basically, then such problem is convex programming problem if f and g i for all i are convex. We have already seen this thing. But, if objective function is maximization type and g i is less than equal to 0 for all i, then this we this is called convex optimization problem or convex programming problem if function objective function is concave and g i for all i are convex ok.

(Refer Slide Time: 05:13)

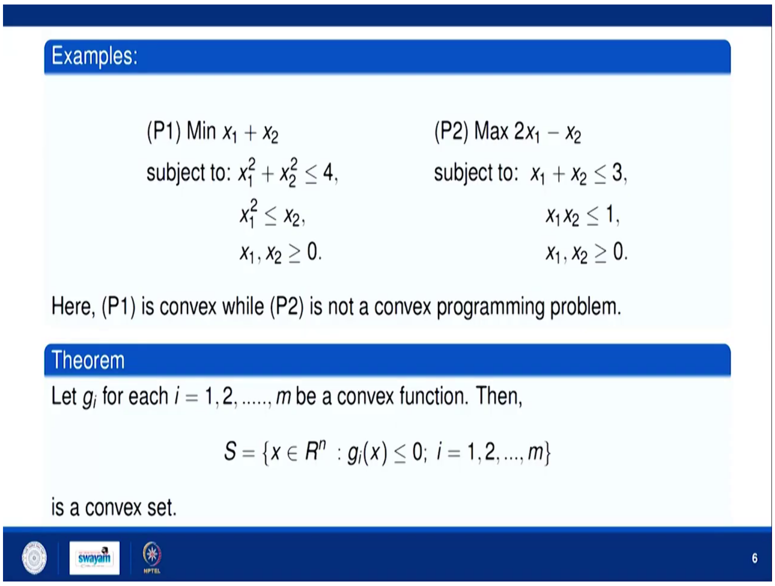
Min $f(x)$ s/t $g_i(x) \geq 0;\ i = 1, 2, \ldots, m$	f is convex and g_i $(\forall i)$ are concave.
Max $f(x)$ s/t $g_i(x) \geq 0;\ i = 1, 2, \ldots, m$	f and g_i $(\forall i)$ are concave.

Now, this now if it is minimization of f x subject to g i greater than equal to 0 for all i then this particular problem is called convex programming problem if objective function that is f is convex and all g i are concave. Because inequality is reversed here and this problem; the last one that is maximizing f x subject to g i x greater than or equal to 0 for all i is called convex programming problem; if f is concave and g i for all i are again concave ok.

So, basically, if we understand this particular format, then we can easily derive the other format also for the that when we can say such problems are convex optimization problem or convex programming problem. So, we basically have to understand this basic format this important format minimization of f x subject to g i x less than equal to 0 for all i and if objective function as well as all constraints are convex, then such problems are called convex optimization problem.

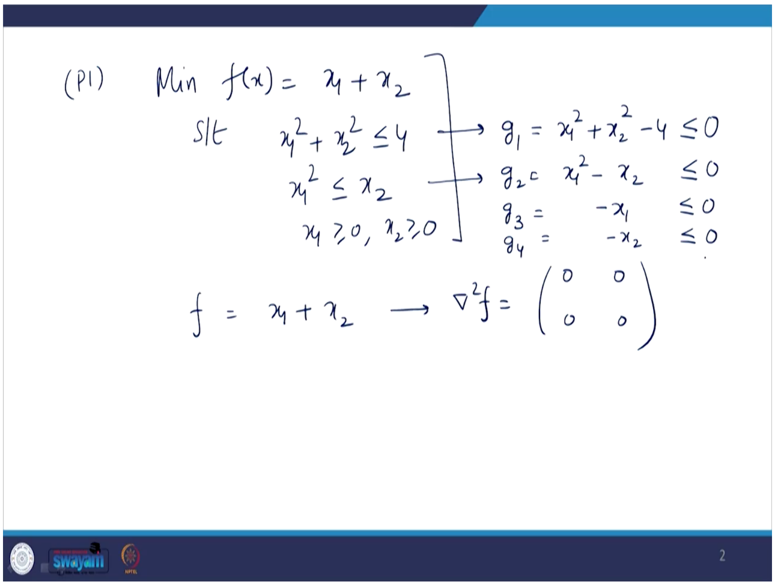
Now, linear programming problem which we have just discussed in which objective function as well as all constraints are linear that is; obviously, convex optimization problem, because all linear functions are convex. Whether, the objective function is maximizing form or minimization form. So, all the linear programming problems are basically convex programming problem. But, a non-linear programming problem may not be convex may not be a convex programming problem ok, in general.

(Refer Slide Time: 06:47)



So, let us discuss a few examples which are a convex programming problem on which may not be a convex programming problem. So, let us discuss problem P 1.

(Refer Slide Time: 06:58)



So, what is the problem P 1? It is minimization of f x which is x 1 plus x 2 that is a linear function ok. This is a linear function here subject to what are the constraineds? The first constrained is x 1 square plus x 2 square less than equal to 4 and the second constrained is x 1 square less than equal to x 2 and x 1 greater than or equal to 0 and x 2 greater than equal to 0 ok.

So, let us see that whether this problem is this problem we are calling as suppose P 2 or P 1 and we want to see that whether this problem is a convex programming problem or not. So, number 1 this problem is of course, this problem is a not a linear programming problem. It is a non-linear programming problem because objective function is linear ok, but constraineds are non-linear. So, this problem is a non-linear programming problem.

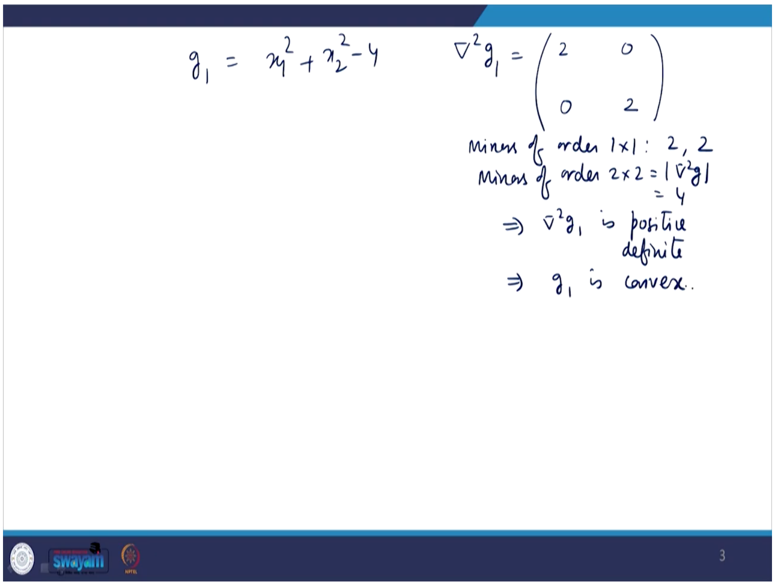
Now, in order to see that whether this function, this problem is the convex programming problem or not. We first have to check objective function and we then we have to check all the constraineds whether they are convex or not ok. So, first let us see the objective function. What is the objective function here? Objective function is x 1 plus x 2 what is the Hessian matrix of this function?.

Hessian matrix is simply 0 0 0 0 or you can directly say that since objective function is linear and the linear function is obviously convex hence f is convex. Or you find the Hessian matrix Hessian matrix comes out to be a null matrix and null matrix is basically a positive semi definite matrix. So, we can say that this function is convex ok.

Now let us see the first constrained. What is the first constrained g 1? The first constrained g 1 x 1 square plus x 2 square minus 4 less than equal to 0 ok. What a second constrained? Second constrained is x 1 square minus x 2 and less than equal to 0 ok. What a third constrained; third constrained is minus x 1 and it is as equal to 0. What a fourth constrained? Fourth constrained is minus x 2 which is less than or equal to 0 again. This x 1 greater than equal to 0 we can write as minus as well as an equal to 0.

So, how many constraineds we are having in this problem? We are having 4 constraineds and if all the 4 constraineds are convex. Objective function is at already shown that it is convex. So, if all the 4 constrained is also convex, then we can say that this problem is the convex programming problem. So, let us see the first constrained. The first constrained is g 1 and what is g 1? g 1 is x 1 square plus x 2 square minus 4.

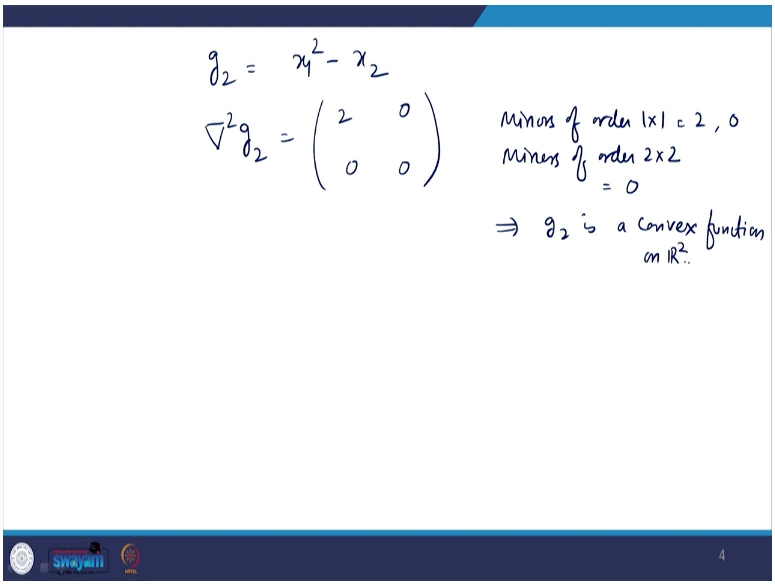
(Refer Slide Time: 09:41)



Let us find the Hessian matrix of g 1. The Hessian matrix of g 1 will be 2 x 2 because we are to do the second derivative respect to x 1. Then x 1 x 2 0 x 1 x 2 0 then it is again 2 ok. So, its eigen values are 2 and 2 ok. Either you can see eigen values or principal minus. So, principle you can find that minus also. So, minus of order 1 cross 1 are 2 and 2 and minus of order 2 cross 2 are determinant of this matrix and determinant of this matrix is 4.

So, all the minus r strictly greater than 0; that means, this matrix is positive definite and hence convex. So, this implies this Hessian matrix is positive definite and this implies a strictly convex or convex. So, we can say that g 1 is convex ok. Now, so, first constrained is convex. Now what is the second constrained? See, what is the second constrained? Second constrained x 1 square minus x 2.

(Refer Slide Time: 11:11)



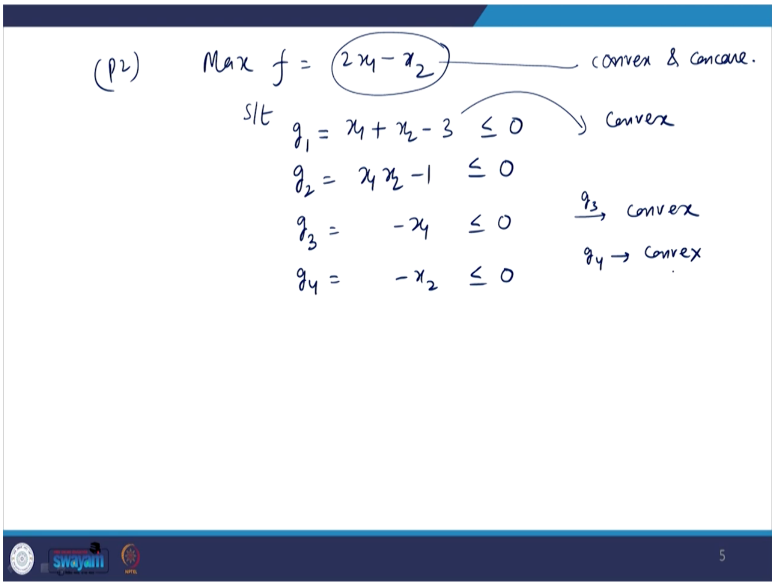
Now, similarly, let us find for the second constrained. So, second constraint is x 1 square minus x 2. So, what are Hessian matrix of g 2 it is 2 0 0 0. So, again if you use the minor test. So, minors of order 1 cross 1 are 1 cross 1 are 2 and 0 and minors of order 2 cross 2 are 2 cross 2 is, because there is only 1 minor which is a determinant of this which is 0.

So, this all the minors are greater than equal to 0 implies this Hessian matrix is positive semi definite and hence we can say that g 2 is convex. So, this implies g 2 is a convex function on R 2 ok. Now what is the; what is the third constrained? The third constrained is linear minus x 1 which is; obviously, convex and the fourth constrained minus x 2 is also linear which is also convex.

So, hence we can say that the objective function as well as all constrained for this problem P 1 are convex and hence we can say that this problem is convex programming problem ok. So, the example 1 the problem 1 is basically a convex programming problem. Now let us see for the problem 2. For the problem 2, what is the objective function here? Objective function is 2 x 1 minus x 2 which is linear which is linear. So, it is concave, because for maximizing for maximization type we should see that f should be a concave function.

Then, see we have already seen that if it is a maximization type maximization type problem, then f must be a concave for convex programming problem. So, it is concave as well as convex. So, it does not matter for this linear part otherwise we should see that this function is concave. So, that is concave. Now, the first constrained is x 1 plus x 2 minus 3 less than equal to 0 which is linear constrained hence convex ok. So, let us see this thing here.

(Refer Slide Time: 13:36)



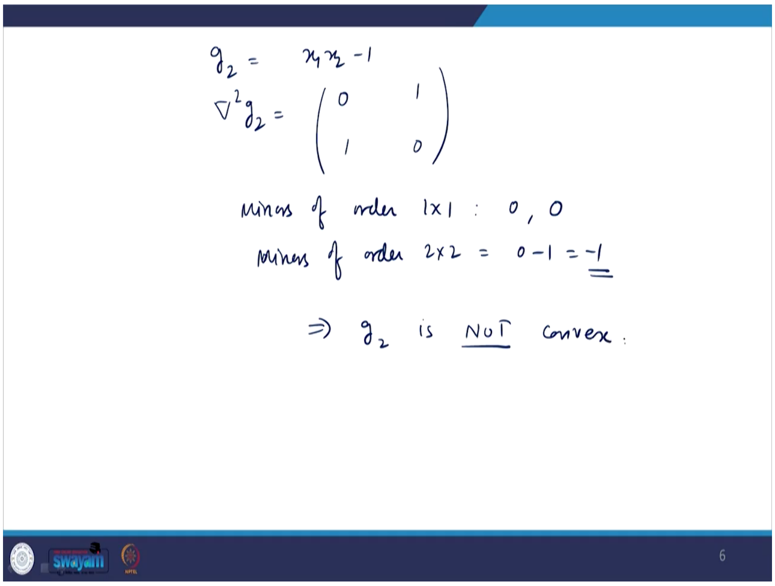
For the problem P 2; problem P 2 is maximization of f equal to 2 x 1 plus x 2 2 x 1 minus x 2 sorry. It is 2 x 1 minus x 2 subject to what are the constrained? First constraint is x 1 plus x 2 minus 3 less than or equal to 0 this is g 1 ok.

The second constrained is x 1 x 2 less than equal to 1. The second constrained is x 1 x 2 minus 1 less than equal to 0. Third constrained is minus x 1 less than equal to 0, fourth constrained is minus x 2 less than equal to 0.

So, this is linear; this is linear means convex; convex as well as concave here convex and concave. This is linear; this is linear means it is also convex, this is linear. So, convex, g 3 is also convex. And g 4 is also convex, because linear functions.

Now we need to check only for g 2. So, let us see whether g 2 is convex or not. If it is convex then it will be a convex programming problem otherwise it will not be a convex programming problem.

(Refer Slide Time: 15:02)

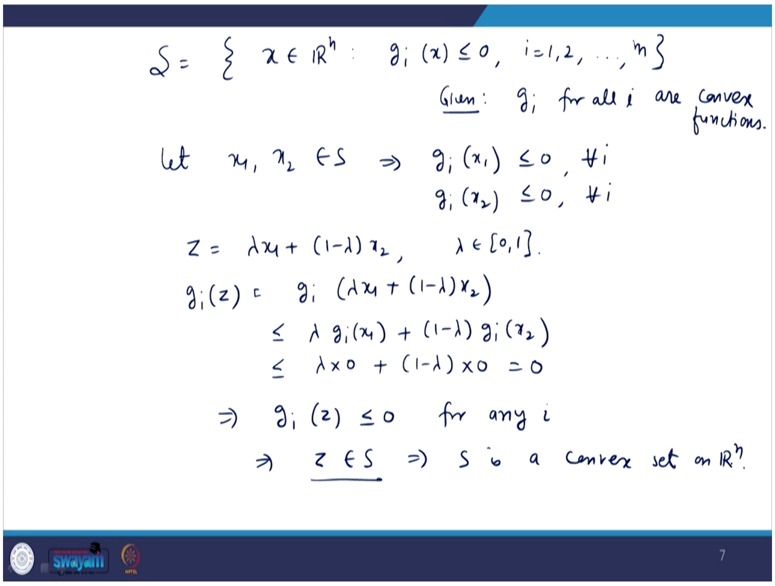


So, what is g 2? g 2 is x 1 x 2 minus 1 less than x 1 x 2 minus 1 ok. So, here it should be ok. So, what is Hessian matrix of g 2? It is second derivative second partial derivative respect to x 1 for g 2 which is 0, it is 1 1 0.

So, minors of order 1 cross 1 are 0 and 0 and minors of order 2 cross 2 are 0 minus 1 which is minus 1. So, this make it indefinite ok. So, this matrixes this Hessian is not positive semi definite and hence this function is not convex. So, this implies g 2 is not convex. So, what we have conclude it? See, in this problem; in this problem this is convex and concave both, this is convex, this is convex, g 4 is convex where g 2 is not convex.

So, this implies that this problem P 2 is not a convex programming problem. Because objective function as well as all constrained must be convex ok. For minimization type objective function is all constrained less than or equal to type. So, hence we have concluded that P 1 is this problem is convex and this problem is not a convex programming problem. Now we have a important result for this that if g i for each i are convex functions then this set is a convex set ok.

(Refer Slide Time: 16:57)



So, let us proof is very easy. You see you can see that we are constituting all those x in R n such that g i x is less than equal to 0 i from 1 to m. So, this is basically what? This is basically a feasible set or a feasible region satisfying all the constraineds. Now give it is given to us that all g i's are g i's for all i's are convex functions it is given to us. And we have to show that this set S is a convex set.

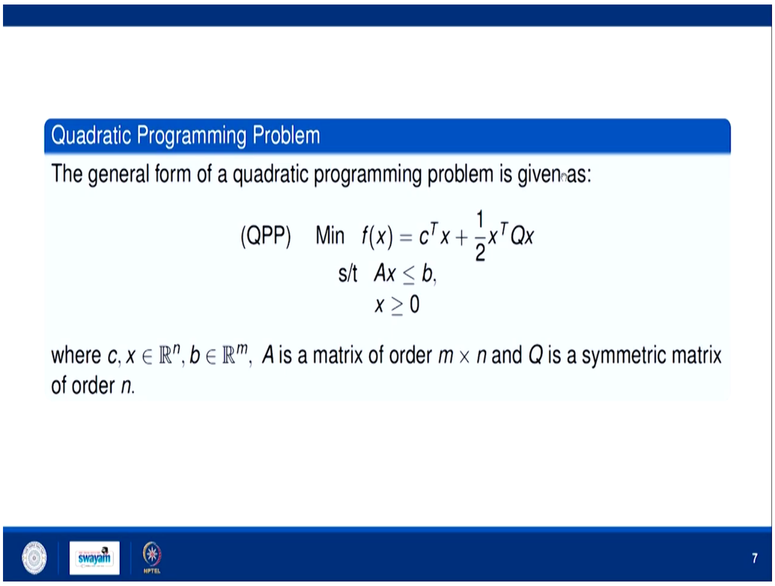
So, how we can show this set is the convex set you can take two arbitrary point in S and try to show that the convex linear combination of those two point is also in the same set. So, let us see let x 1 x 2 in S. So, this implies g i x 1 is less than equal to 0 for all i and g i x 2 is also less than equal to 0 for all i. Now let z equal to lambda x 1 plus 1 minus lambda x 2 which is convex linear combination of x 1 and x 2.

Now, what is g i z for any i? It is g i lambda x 1 plus 1 minus lambda x 2. Now these g i's a convex for all i. So, this means is less than equals to lambda g i x 1 plus 1 minus lambda g i x 2 and g i x i g i x 1 is less than equal to 0 for all i, g i x 2 less than equal to 0 for all i. So, this is less than equals to lambda into 0 plus 1 minus lambda into 0 is equal to 0.

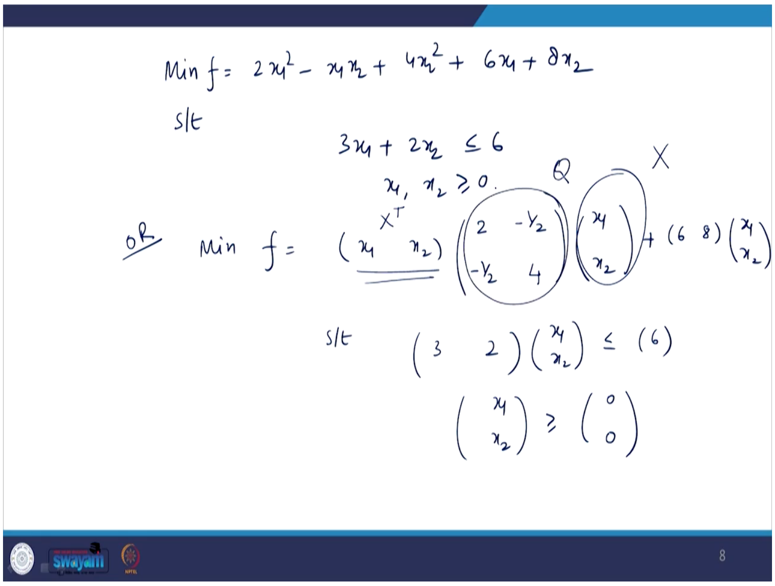
So, this implies g i z is less than equal to 0 for any i and this implies z belongs to S and this implies S is a convex set on R 2 R n convex set. So, what we have seen? We have seen that if we are taking a convex programming problem, then the region then the feasible region of a convex programming problem is always a convex set ok.

So, that is the important property see in this example; if you draw the, if you draw the graph of this. So, it will not be a convex set the region main is not a convex set that is why it is not coming out to be a convex programming problem. So, before going ahead let us first come to quadratic programming problem. So, we have seen that a quadratic programming problem is basically the objective function is quadratic and constraineds are linear ok.

(Refer Slide Time: 20:00)



(Refer Slide Time: 20:08)

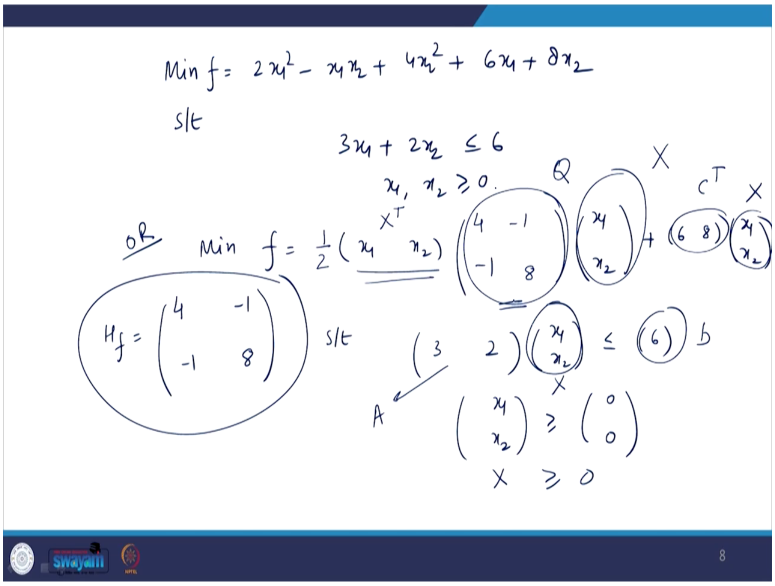


So, suppose you are having this type of problem minimizing 2 x 1 square minus x 1 x 2 plus 4 x 2 square plus 6 x 1 plus 8 x 2 subject to 3 x 1 plus 2 x 2 less than equal to 6 and x 1 x 2 non negative. A simple problem I have taken to understand. It is a non-linear programming problem and this non-linear programming problem, basically a quadratic programming problem because objective function is quadratic and all constraineds are linear.

Now, if you take this problem f. So, you can write this problem as x 1 x 2. This 2 will come here this 4 will come here this is minus 1 by 2, this is minus 1 by 2, this is x 1 x 2 will come here this is 6 and 8 and this is x 1 x 2 and subject to this is 3 2 x 1 x 2 less than equal to 6 and this is x 1 x 2 greater than equal to 0 0.

So, if you take; if you take this as capital X ok. So, this will be X transpose and this you can take as Q. So, here I have taken half ok. If I have taken it half here. So, if I have taken a half here. So, I can take it half here and I multiply this by 2.

(Refer Slide Time: 21:55)



So, in that case it will be, in that case it will be 4 8 minus 1 minus 1. So, it is 1 by 2 X transpose 2 X plus this is c transpose this is X. This is matrix A, this is X, this is b and this is X greater than equal to 0 this is 0.

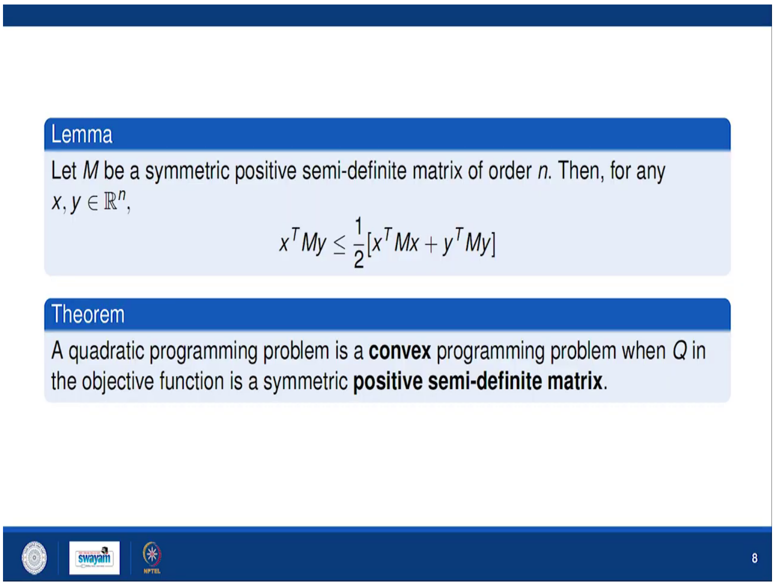
So, what I want to show, I want to illustrate you that any problem of this type which is having quadratic objective functions subject to linear constraineds can be written in this format which is c transpose x plus half of x transpose 2 x subject to A x less than equal to b x greater than equal to 0 ok. So, here this Q is what? If you carefully see what Q is this Q what is the Hessian of this f? If you compute the Hessian of this particular f.

So, Hessian matrix of this f will be; you have to differentiate respect to x 1 twice. So, that will give 4 minus 1 minus 1 8. So, which is here. So, this Q is nothing, but Hessian matrix of f for this type of QPP ok. Now the interesting question is that when this problem is convex programming problem.

See, constraineds all are linear. So, we are not worrying about the constrained constraineds are linear. So, it these are; obviously, convex, but what should be the restriction on f on c or Q. So, that we can say that the given problem is a convex programming problem. Now why we are focusing on convex programming problem so much? That we will discuss in the next lecture ok.

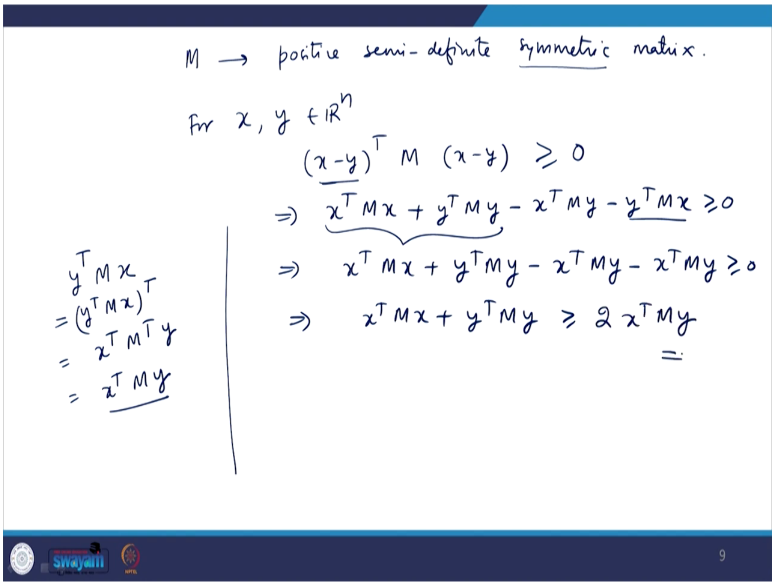
So, first we will see that when we can see when we can say that this problem is a convex programming problem.

(Refer Slide Time: 23:59)



So, to understand this we have two important results. The first result is; if this M is a symmetric positive semi definite matrix of order n, then for any x y in R n, this inequality hold. The proof is very easy; see the proof is easy see here what is given to us.

(Refer Slide Time: 24:19)



We have to given that M is a positive semi definite symmetric matrix. So, we have to we have to show that x transpose M y for any x y in R n is always less than equal to 1 by 2 times x transpose M x plus y transpose M y. If it is a positive semi definite matrix. So, for any x y in R n x minus y transpose M into x minus y. See x minus y is also a vector in R n and if it is a positive semi definite matrix then x transpose M x is always greater than equal to 0 for any x in R n and this x minus y is also a vector in R n.

So, for this also this is greater than equal to 0, because M is positive semi definite matrix. So, what does this implies further? This further implies x transpose M x plus y transpose M y minus x transpose M y minus y transpose M x greater than equal to 0 ok. Now let us so, this is required in the proof; if you see the proof this is required here.

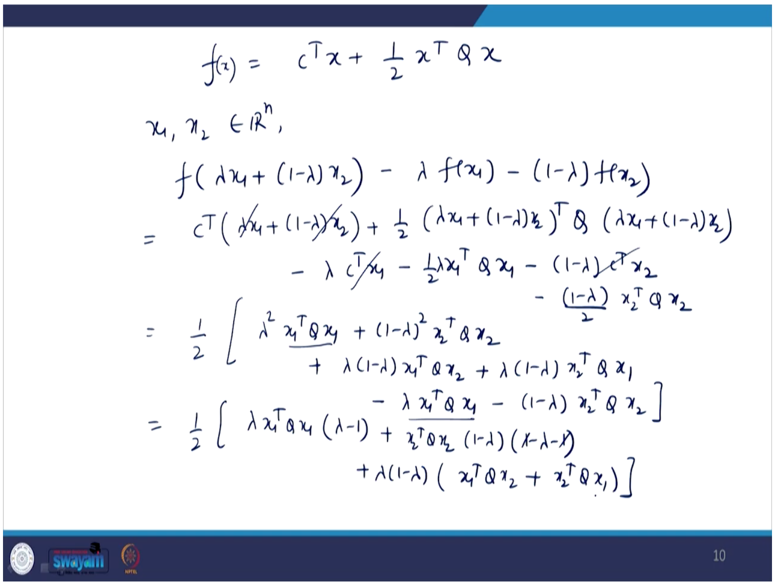
So, now, let us focus on this part. So, what is y transpose M x. Now let us see; let us see here. What is y transpose M x? y transpose M x is basically a real number a scalar number it is in R if it is in R. So, it will be equal to its transpose ok.

So, what is equal to x transpose M transpose y and M is a symmetric matrix also. So, symmetry means x M transpose is equal to M. So, this is equals to x transpose M y so; that means, this is equal to this.

So, this further implies using this x transpose M x plus y transpose M y minus x transpose M y minus x transpose M y is greater than equal to 0, because this is equal to this we have just shown. So, this implies x transpose M x plus y transpose M y is greater than equal to 2 times x transpose M y. So, hence we have done. Hence that we if we take this 2 this side. So, hence we have proved this result.

Now this proof result is required to prove the next result. The next result is a quadratic programming problem is a convex programming problem if Q in the objective function is positive semi definite matrix. If this Q here is positive semi definite matrix then this problem QPP will be a convex programming problem. So, how can we show this? How can we show this result? So, the again we will proof is easy.

(Refer Slide Time: 27:38)



So, what is the objective function here? Objective function is c transpose x plus 1 by 2 x transpose Q x to showed at this function is a convex functions. Let us take 2 point x 1 x 2 in R n and take the c l c of these 2 points and take the c l c of these 2 point the c l c will be equal to this and will be minus lambda. See what I am trying to do here.

We have to show that is this f is convex. So, to show that this f is convex if we apply the basic definition of; basic definition of convexity, then we have to show that this quantity is less than equal to lambda f x 1 plus 1 minus lambda f x 2. So; that means, I have to show that this quantity is less than equal to 0 ok. So, I have taken this expression I will try to show that this is less than equal to 0 if Q is positive semi definite matrix.

So, now let us quickly substitute all these values from given f this is the f x. So, this is lambda x 1 plus 1 minus lambda x 2 and then it is plus 1 by 2 lambda x 1 plus 1 minus lambda x 2 whole transpose Q into lambda x 1 plus 1 minus lambda x 2 minus lambda times x 1 minus 1 by 2 x 1 transpose Q x 1 minus 1 minus lambda times c transpose x 2 minus.

So, this lambda will come here also. So, it is 1 minus lambda times 1 by 2 it is 1 minus lambda f x 2. So, it is x 2 transpose Q x 2 ok. So, this lambda (Refer Time: 29:35) scalar quantity can be taken out. So, this will comes out to be c transpose x 1 lambda times. So, this will be cancelled with this term. This 1 minus lambda again will come out. So, which is c transpose x 2.

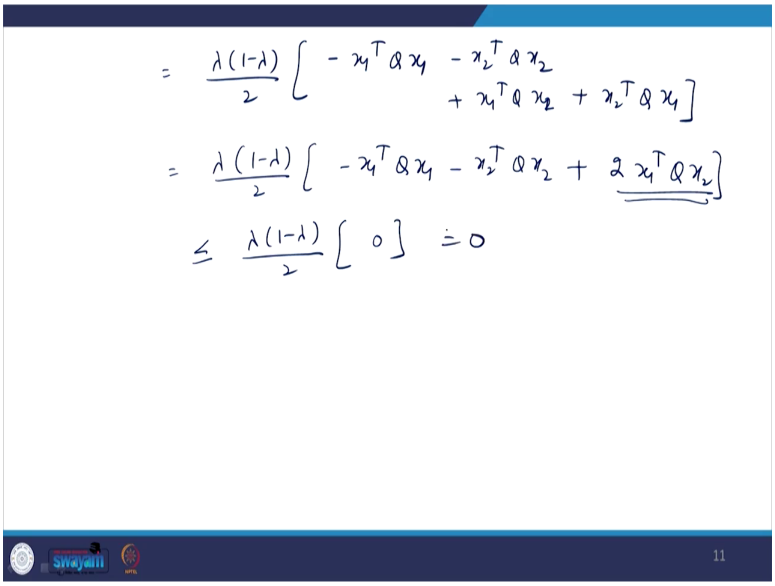
So, this will cancel with this term ok. We are left with these terms. So, 1 by 2 can be put common in all the terms and let us simplify this expression. So, it is lambda square x 1 transpose Q x 1, because lambda is a scalar quantity plus 1 minus lambda square x 2 transpose Q x 2 plus this is lambda into 1 minus lambda x 1 transpose Q x 2 and plus lambda into 1 minus lambda x 2 lambda x 2 transpose Q x 1 minus lambda times x 1 transpose Q x 1 and minus 1 minus lambda times x 2 transpose Q x 2.

So, these are final expression we left with. We any how we have to show that this is less than equal to 0 in order to showed at these function is a convex function. And we have to use that Q is symmetric positive definite matrix, positive semi definite matrix this we have to use.

So, now let us see now from this and this what is common? From this and this lambda is common, x 1 transpose Q x 1 is common, this will left with lambda minus 1. Again from these 2 x 2 transpose Q x 2 is common and this is 1 minus lambda is also common. So, it is 1 minus lambda minus 1 and leave these 2 term as it is; lambda 1 minus lambda you can take common x 1 transpose Q x 1 plus x 2 transpose Q x 2 ok.

So, this 1 1 cancels out. So, lambda into 1 minus lambda you can take common from all the 3 all the 4.

(Refer Slide Time: 31:48)



So, what we get? We get with we left with lambda into 1 minus lambda upon 2 ok. It is what? Negative can come out. So, it is minus x 1 transpose Q x 1. What is the second term? Second term will come from here which is minus x 2 transpose Q x 2. And then from here it is sorry it is x 1 transpose Q x 2. So, this will be; this will be x 2 and it is x 2 transpose Q x 1. So, x 2 transpose to Q x 1 it is x 1.

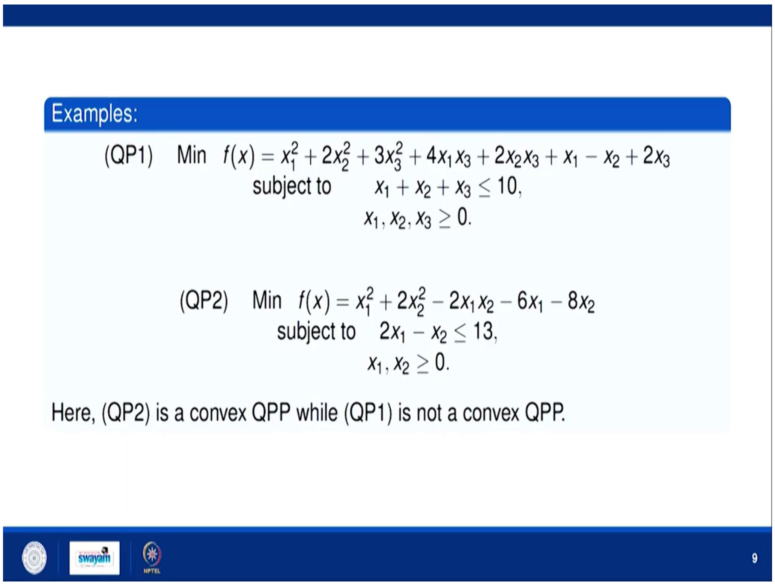
See here, it is x 1 transpose to Q x 2 it is Q x 2 and it is Q x 1 ok. So, it is plus x 1 transpose Q x 1 Q x 2 plus x 2 transpose Q x 1. Now you can make use of that lemma that result which we have just discussed this result; that x transpose thus this twice of this is less than equal to this.

Because, you see here because these two are equal we have just shown these two are equal these two are equal means this can be written as further can be written as lambda into 1 minus lambda upon 2 times minus x 1 transpose Q x 1 minus f 2 transpose Q x 2 plus 2 times x 1 transpose Q x 2 ok.

Because, these two are equal. So, this expression from that lemma will be less than or equal to lambda into 1 minus lambda by 2. This whole quantity is 0 less than or equal to 0. So, it is 0 by this result whole quantities less than equal to 0. So, hence we can say that this function is convex if Q is symmetric positive semi definite matrix. So, whenever we deal with quadratic programming problem in solving in developing algorithm for machine learning.

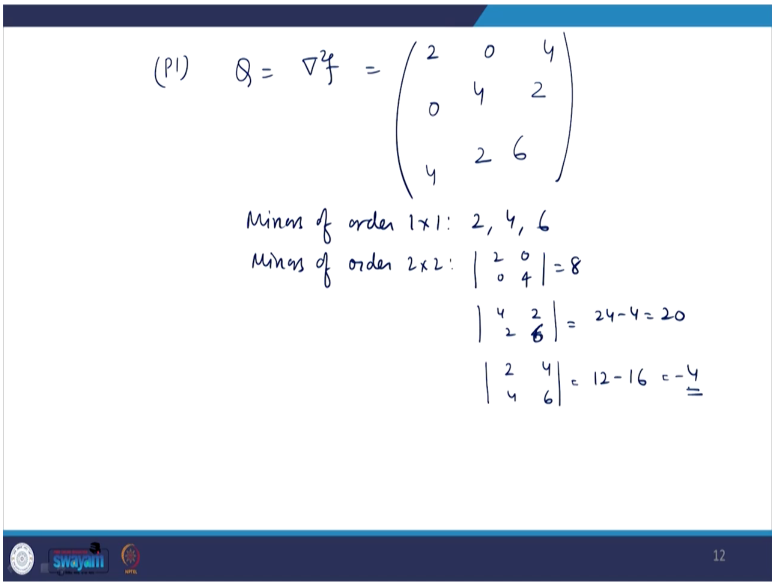
Then, if we have to see that whether a given QPP is convex or not. We have to simply work out for Q; you simply try to find out the Q and try to see whether that Q is positive semi definite or not if it is positive semi definite then the given QPP will be a convex programming problem.

(Refer Slide Time: 34:23)



So, for illustration here are examples. So, let us discuss the first example. So, constraineds are linear. So, now, we are not worrying about the constraineds we have to find the Hessian matrix of this Q only.

(Refer Slide Time: 34:45)



So, what is a Hessian matrix here for the Q? So for the first example what is Q? Q is nothing, but Hessian matrix of f. So, what is Hessian matrix here? Diagonal element will be 2 4 9, 2 4 6 sorry, it is 2 4 6 and it is 4 x 1 x 3, 4 x 1 x 2 means 4 will come here 4 will come here, then it is 2 x 2 x 3, 2 x 2 x 3; x 2 x 3 means 2 will come here 2 will come here this is 0, x 1 x 2 is 0 ok.

Because, the other part is linear. So, Hessian will be 0 and from x 1 x 3 x 2 x 3 that only these term these terms will be there. Now to show that it is symmetric and we have to just showed at this Q is positive semi definite matrix.

So, find the minors of this Q. So, minors of order 1 cross 1 are simply 2 4 6, minors of order 2 cross 2 are determinant of 2 0 0 4 which is 8. Determinant of 4 2 4 2 2 6 ok; which is 24 minus 4 is 20 which is positive again and it is 2 4 4 6, which is 12 minus 16.

So, this row and this column will be deleted 12 minus 16, which is negative. So, this is not; this is not a positive semi definite matrix and this problem hence this problem is not a convex programming problem. So, similarly, one can show that so, this problem is not a convex programming problem; however, this if you find the Hessian matrix of this.

So, one can easily show that this is convex programming problem. Because, the Q will be positive semi definite matrix. If you are having a quadratic programming problem and you want to show that it is a convex programming problem or not. You simply find out the Hessian matrix of that f that objective function that is nothing but Q ok. And if it has come out to be if it is come out to be positive semi definite that implies that is sufficient to say that the given QPP is a convex optimization problem.

So, in the next lecture we will see that if it is a general non-linear problem, then how we can say then, how the convex programming problem are important.

Thank you.