Distributed Optimization and Machine Learning

Prof. Mayank Baranwal

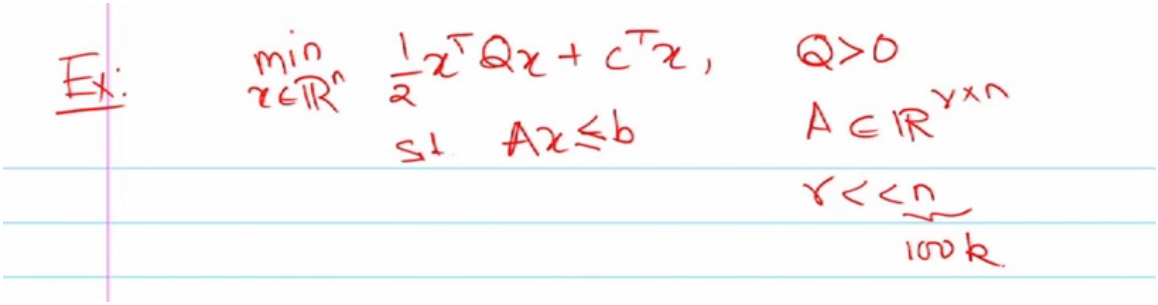
Computer Science & Engineering, Electrical Engineering, Mathematics

Indian Institute of Technology Bombay

Week-2

Lecture - 8: Primal and dual optimization problems

In the last class, we looked at this particular example on quadratic programs right, which was to minimize this quadratic function x in Rn x transpose Qx plus c transpose x. let us say Q is positive definite, subject to some inequality constraint of the form or let us say inequality constraint of this form Ax less than equal to b, where A is some matrix in R r cross n. Alright so in this case let's say r turns out to be much much smaller than n and can be in the range let's say 100k or a million-dimensional sort of variable right. Now let's say different agents are trying to solve this problem in a distributed manner. So what could be a potential challenge that you can see in solving such problems especially in this case. So let's say we have a network of agents and every agent has access to a part of this A matrix.



So everyone is trying to solve this global problem but they only have access to certain rows of this A matrix. Number of such rows are this little r. So everyone would try and come up with their own sort of estimate of this global x. Now what is the particular challenge that you see in solving this problem in a distributed manner? especially when you have r much much smaller than n or n happens to be very large.

So again in distributed optimization whenever an agent solves for a particular problem they are also going to be exchanging information with their neighbors. And exchanging a variable which is 100K dimensional, that is going to pose a serious requirement on the communication channel. The bandwidth that is required to share that large dimensional variable. And this is not a very viable approach to go about solving this problem. Because every time you, I mean everyone like any agent solves this problem locally, they would be exchanging their information with their neighbors and trying to get a sense of how the final sort of global estimate looks like.

And this is going to be seriously communication intensive. So, when n is very large. So,

it is going to pose a computational or bandwidth requirement on communication channel right? and this may be prohibitive and n is very large. The other problem that we see is well I mean it is not too difficult, but at the same time working with these kind of inequality constraints may not be as easy right. So, the question is can we try to come up try to like reformulate this problem in a manner such that it reduces the bandwidth requirement and they are much simpler constraints to work with.

Is that clear? So that, so our objective is to be able to convert this optimization problem into an equivalent sort of optimization problem such that, so can we come up with a simpler optimization problem and when I say simpler, so something which has lesser bandwidth requirement So really the two sort of varied like two different dimensions of concern are r and n right. We know that r is much much smaller than n. So something that scales with r and not with n right. So bandwidth requirement scales with r. So that would be a simpler version right.

So what does Ax less than equal to b represent here? So these are r inequality constraints. right and if something if I come up with a problem which has fewer inequality constraints to work with and but it scales with the number of inequality constraints. So, that would be a simpler problem to work with right. So, it scales with r and not n and has simpler so, when I say simpler optimization problem the other objective is instead of working with this kind of inequality constraints Ax less than equal to b can the constraint be simplified further. And the answer to this particular question is yes and that is one of the reasons why we are studying this right.

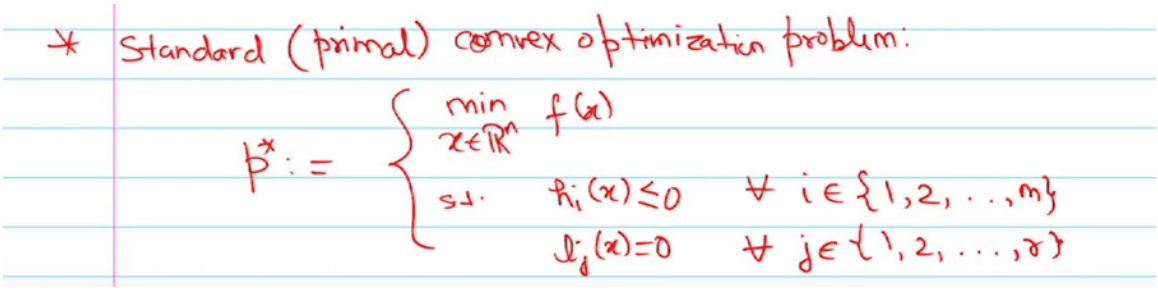
But then and the way we sort of approaches we convert this. So, this is called a primal optimization problem. In its original form it is a primal optimization problem and we are going to look at its. So, we are going to convert this primal form to something called dual optimization problem. we are going to change this to a dual optimization problem.

So, and we would see that it is much easier to work with dual optimization problem in certain cases than working with the primal form. We are also going to look at things like weak duality. then conditions under which weak duality becomes strong duality ok. And finally, we are going to look at something called Lagrangian dual function. So, again all of this are going to be related to be able to pose this problem, pose this optimization problem as an equivalent dual optimization problem and try to study conditions under which these forms are equivalent.

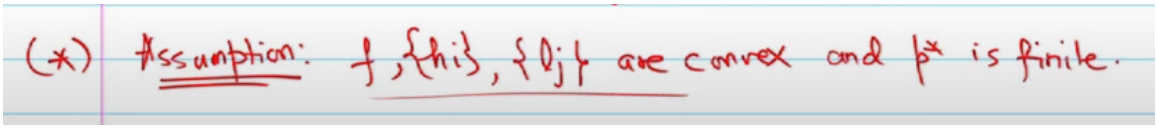
So, that would be the emphasis of today's lecture. Is that clear? Any questions so far at least from the point of view of motivating why we need to study dual optimization problem. So in a primal optimization problem things scale with x right and x can be large dimensional here and you wouldn't want to be exchanging very large amount of information through a communication like channel because it would have larger bandwidth requirement. Again we have inequality constraints to work with. So we would want to come up with an equivalent problem which possibly returns the same solution, but something which scales with r the number of inequality constraints that you have and instead of working with these any like complex inequality constraint maybe we can

simplify this further.

So, that would be the objective and that that will be achieved using a dual form and we are going to look at dual forms of optimization problem. So, what is this like let us start with the standard primal form or primal convex optimization problem ok. So, a most sort of general form of convex optimization problem is. minimize some function f of x subject to you have a bunch of you have bunch of inequality constraints. So, hi x less than equal to 0 or and then you have some equality constraints lj x equal to zero.

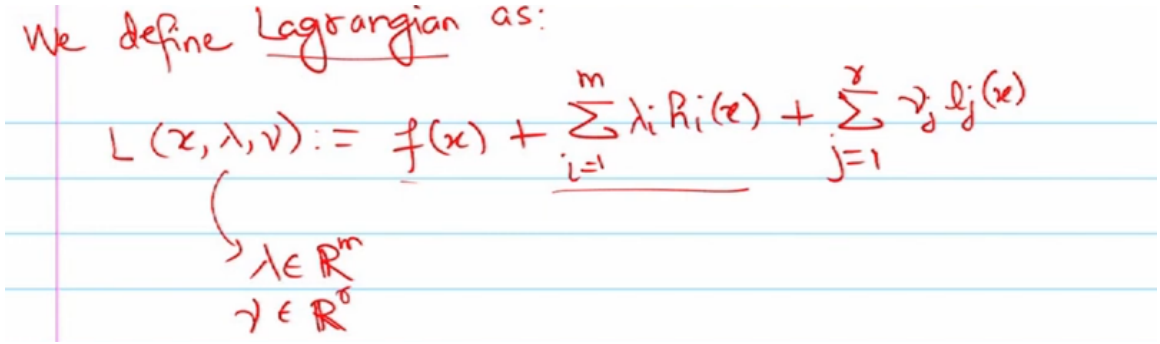


and we are going to be assuming. So, all the functions f, hi, lj these are convex ok. So, these are convex functions and let us say the optimal solution to this problem happens to be p star So, p star is the minimum value of f of x subject to these constraints. So, why p star? Because this is the primal form. So, I am using the term p to denote that the primal objective value is p star.



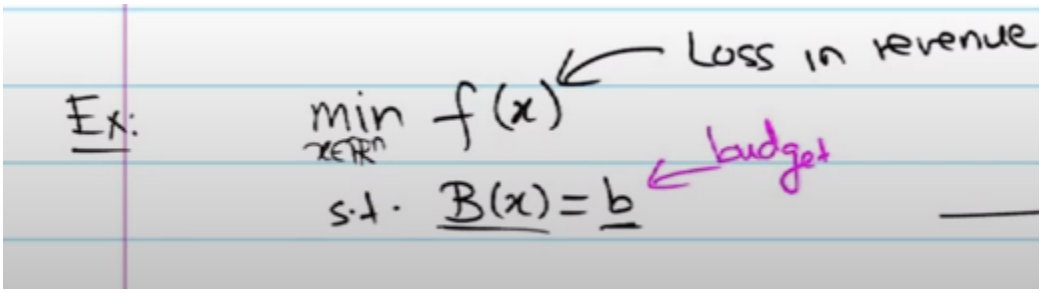
So, we are going to be assuming that these functions are convex. In fact, in most cases this function lj, the equality constraints we work with linear equality constraints and we assume that p star is finite. So, throughout the lecture we are going to be working under this assumption and the question, and then we will try to come up with the dual form of this primal optimization problem ok. So, for this primal optimization problem. So, we define Lagrangian finite value yeah p star is less than infinity like I mean between minus infinity all right.

So, we define Lagrangian for this particular problem is f of x plus summation lambda i hi plus summation nu j lj. let us first look at the definition and then we will see we will basically try to study the consequences of this studying this particular object. So, again to start with we have this primal form right, this primal object optimization problem which is to minimize f of x subject to these inequality constraints and in the these bunch of equality constraints. And depending on the number of inequality and equality constraints we have, we define something called Lagrangian of this particular problem. And Lagrangian is defined to be f x plus this particular term.



So, essentially to say that lambda is in Rm and nu is in Rr. Is this clear? All right. So why do we care about this particular Lagrangian? So let us motivate this with an example. Suppose you are a company and you are trying to minimize total loss in revenue. So let us say you are trying to minimize a function.

Think of this function as loss in revenue subject to some budgeting constraints. So let us say the budget is B x equal to b ok. So, the total budget that you have a budget function. So, depending on the number of quantities you produce. So, x is the number of quantity right.

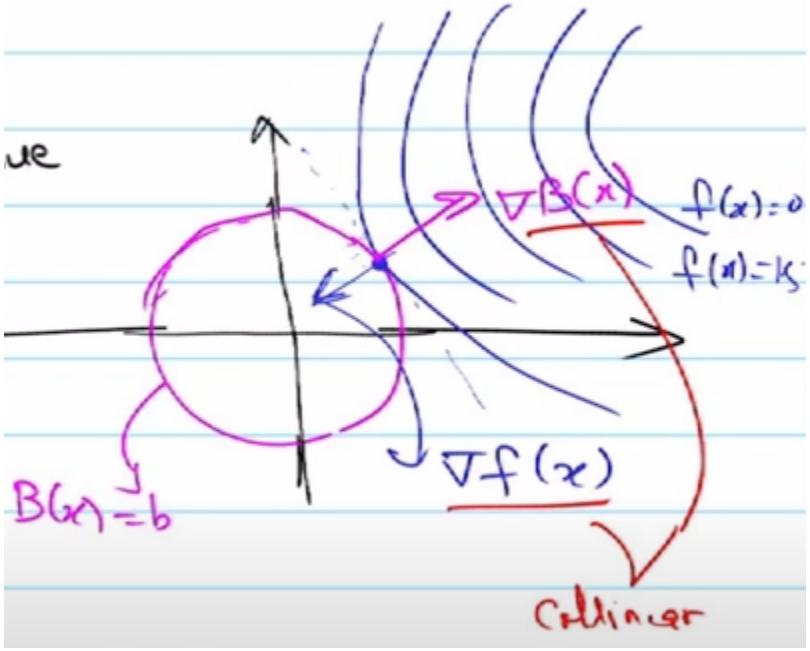


So, you are trying to minimize total loss in revenue based on how much quantity you produce. So, you want to optimize the number of quantities you produce. So, as to minimize the total loss in revenue and then you are I mean you have some budgeting constraint. So, little b is your budget that you want to operate under ok. So, if I try to analyze this pictorially.

So, again your functions or let us say your budget set or constraint set is something like this. So, this can be. I mean you can operate like let's say this is the total budget that has been given to you. I mean in general it can be less than equal to b, but I am trying to study a simpler case. So, for now just ignore this particular part.

So, we are just looking at what happens to Lagrangian when we just work with this Lagrangian under equality constraint. So, let's say this is the equality constraint that you are asked to operate under. So, this is your budget, the total money that you have and then you want to spend that much money, but at the same time you want to minimize the loss

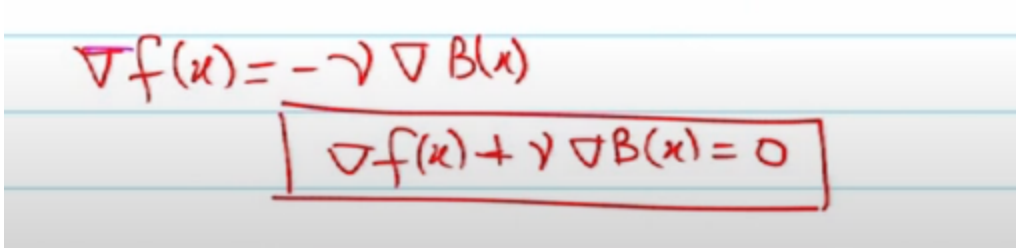
in revenue. And your function again we can draw the level sets of the function. So, this can be for instance f x equal to 0, f x equal to some value let us say 1.



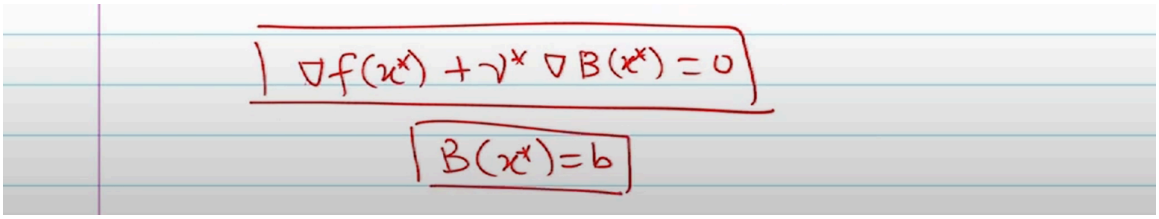
5 so and so on. So, f x increases this way right. And when does this function get minimized subject to this particular equality constraint? Yeah, you are right in the sense that maybe I should have drawn this differently. So, interior is not included. So, let us say in this case here. So, Bx equal to b is this constraint, just this constraint right.

So, when does this function get minimized? If it is basically gets minimized at the point where the function touches this for the first right. So, at this point here ok. At this point what is the direction like basically you have at the since it touches at this point right. So, essentially if I if I draw a line which is orthogonal to this particular direction it is it is like a tangent to this constraint set as well right. So, the gradient of this this would be the gradient of f of x and if I look at this constraint set B x equal to b what is the normal or the gradient of this constraint set it would be pointing in this direction right ok.

and all we know is that these vectors either it can be pointing in this direction or in the opposite direction depending on whether it increases in this direction or decreases, but we know that these vectors are going to be collinear right. So, in this so basically we know that at least pictorially we know that gradient of f of x is some constant nu times gradient of B right because these vectors are collinear. So, vectors here and here these are collinear ok. I mean they may be pointing in the same direction or in the opposite direction, but they would still be pointing along the same line. So, same direction right or another way to write this is gradient of f of x plus nu times gradient of B of x equal to zero.

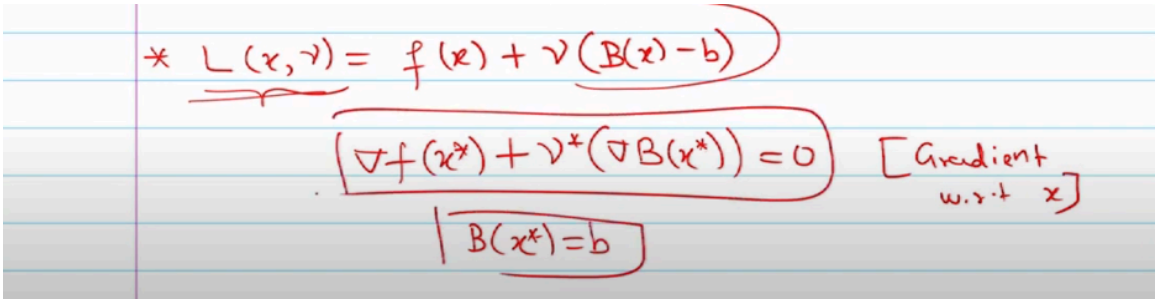


So, this is one of the constraints that we have for optimality. So, if now let us define this optimal point to be x star and so one of the constraints that we have for optimality is that gradient of f of x star is some new star times gradient B x star. So, this is equal to 0 right. that is one of the constraints. What is the other constraint? B x star equal to b right.



I mean of course, x star has to be a feasible point. So, B x star equal to b is another constraint. So, now for this problem, this is first of I mean by the way this is a primal optimization And let us look at the Lagrangian for this right. So, there are no inequality constraints here.

So, we would not even include this. I mean this is just to motivate the need for Lagrangian like this. So, there is no equality constraint here or inequality constraint. So, the corresponding Lagrangian would be for this problem L x nu. There is just one equality constraint. So, the Lagrange I mean so the dimension of nu is just 1.



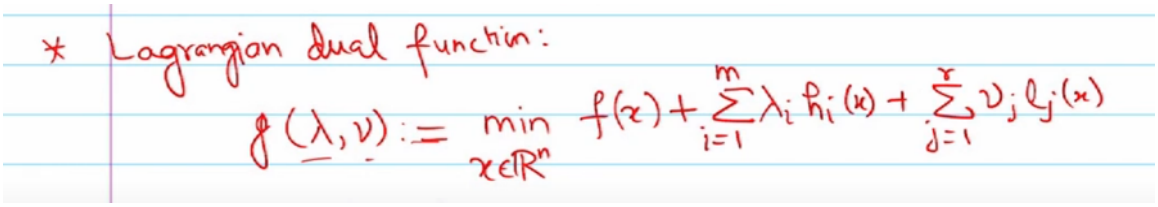
So, this would be f of x plus nu times B x minus b. Now, suppose now treat this particular Lagrangian. So, remember we started with the constrained optimization problem. Now, think of this constrained optimization problem as an unconstrained optimization problem with this being your objective function. So, when the objective function is unconstrained or the optimization problem is unconstrained, what is the condition for optimality? for unconstrained optimization minimization gradient vanishes right.

So, in this case there are two variables x and nu. So, the gradient with respect to both the variables must vanish. So, when I say gradient with respect to x. So, that would be gradient of f of x at the optimal point x star right. At the optimal point x star gradient x plus nu times gradient of B at x star this is equal to 0.

So, this is gradient with respect to x and you assume that x star, nu star are the optimal, x star nu star is the optimal solution and the gradient with respect to nu must also vanish right because it is a function of two variables x and nu. So, the gradient with respect to nu should also vanish and when you take the gradient with respect to nu you are left with just this equality constraint which anyway needs to be satisfied. So, B x star equal to b is another constraint. and that is what we had obtained even pictorially right. So, in some sense Lagrangian helps you convert a constrained optimization problem into an unconstrained optimization problem.

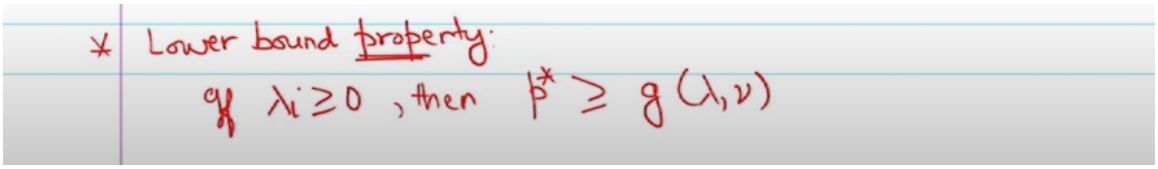
Is this clear? Any questions on this? Yeah. So, with inequality, it is not that straightforward, but we will come to that, but for the equality-constrained optimization problem is this clear? So, both these constraints need to be satisfied and I mean even analytically and as well as like pictorially that sort of make sense. So, these quantities lambda and nu these are called Lagrange multipliers. and they would also help you dualize your primal objective optimization problem. So, they are also called dual variables ok. Is this clear? So, lambda and nu are your Lagrange multipliers or the dual variables and they play a role when we try to come up with a dual formulation of your primal optimization problem.

But really you should see Lagrangian form or sort of the Lagrangian as a way to convert a constrained optimization problem to an unconstrained optimization problem, especially for equality constrained optimization problem. Yeah, I mean we have not come to that yet. I mean you can define the Lagrangian even for lambda just for any lambda, but I mean when you will come to that part. but for now just assume that lambda lambda is m dimensional and basically as many as the number of inequality constraints and nu is the number of equality like the dimension of nu is the number of equality constraints all right. So, with this we define something called Lagrange dual function or Lagrangian dual function.



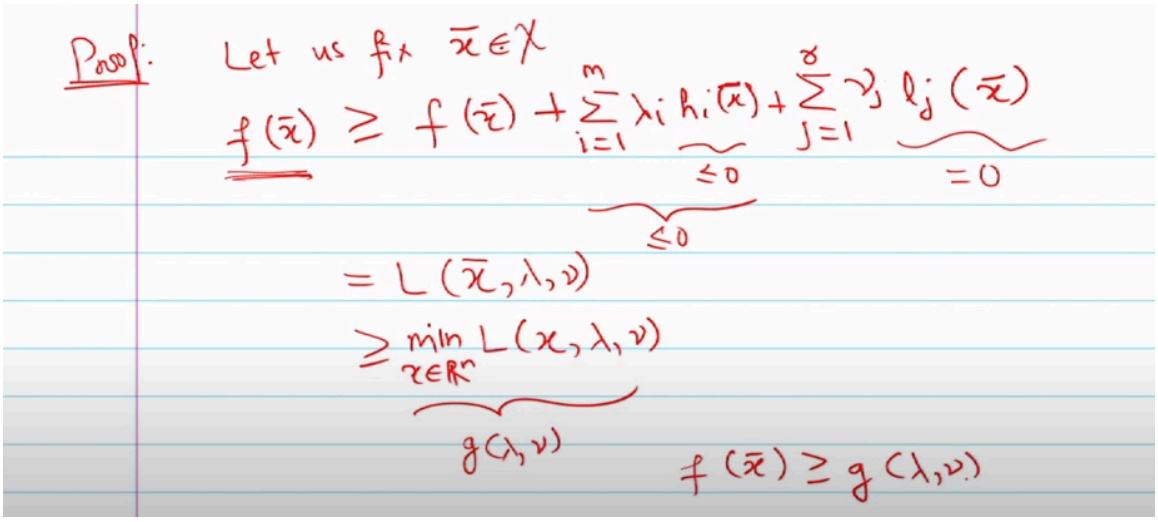
So, this is a function in terms of your Lagrange multipliers lambda nu and it is defined to minimum over x of the Lagrangian. So as I said, so this is nothing but minimize the Lagrangian with respect to x, just with respect to x. So for each lambda and you fix a lambda and mu and you minimize the Lagrangian with respect to x and that gives you one particular value of g, that Lagrangian dual, right? So g lambda, again it is a function

of two variables and you try to get rid of this x variable by minimizing it over x, ok? Is this clear? Alright. So this has a nice lower bound property. So, this says that if your lambda is and that basically comes like answers your question.



So, if your the Lagrange multiplies for the inequality constraints if they turn out to be positive, then your primal objective value always over approximates your Lagrangian dual for any here. for all lambda nu as long as lambda is a positive ok. And this gives you an idea as to I mean we will come to the duality gap and things later, but this so as long as this is true p star is always in sort of the upper bound on this Lagrangian dual function.

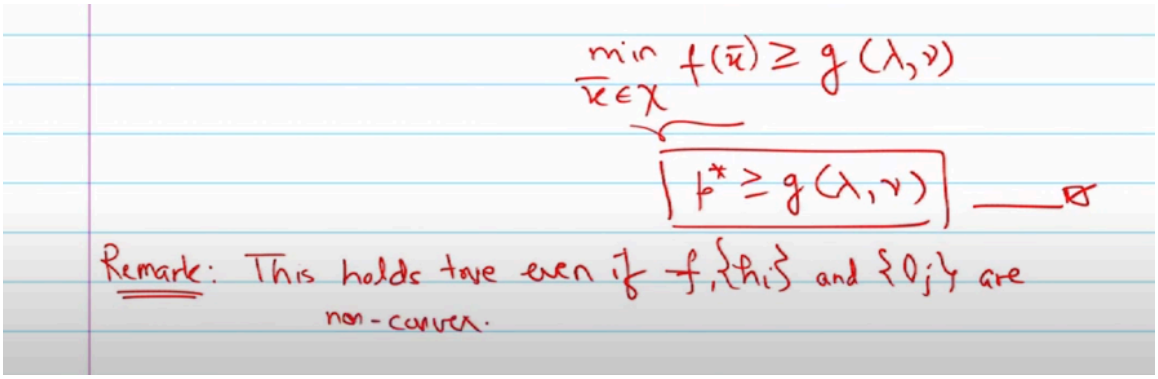
And let us look at a quick proof for this. yeah ok. So, let us let us fix some x right some x bar. So, f of x bar we know it is going to be greater than equal to f of x bar plus summation i 1 through m lambda i h i of x bar plus j 1 through r nu j l j of x bar. Why is that? We choose an x bar which is a feasible point. So, why is that? So if x bar is a feasible point then lg of x bar is 0, hi of x bar is less than equal to 0 and if lambda is a greater than equal to 0 then this whole term is less than equal to 0 right. So therefore the left hand side is always going to supersede the right hand side right.



Is this clear? So, what is the right hand side? It is nothing but Lagrangian evaluated at x bar lambda nu. And this is true for a specific x bar. So, if I choose the minimum value of Lagrangian, that is also going to be true. and what is this value? The Lagrangian dual function. So, what do we get from here? f of x bar is greater than equal to g lambda nu and this is true for every x bar in your feasible set.

So, if I try to minimize x over all x bar in x f of x bar this should also be true and this value is nothing but p star right. So, we get that p star is greater than equal to ok. Is this

clear? So, p star is always an upper bound on the Lagrangian dual function. So, what is the best that you can do? By the way this a quick remark. So, this up this lower bound property this holds true even if the functions f g sorry f h i and l j are non convex.



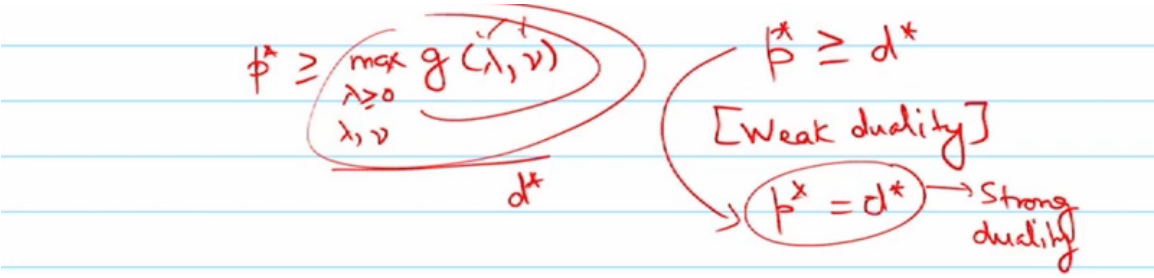
because nowhere we used a convexity argument. Even if they are non-convex, your primal objective value, optimal value is always going to exceed the Lagrangian dual for any lambda nu. Is this clear? Is this clear to everyone? So, what does this suggest? This suggests that like if even if I try to maximize this, maximize this Lagrangian dual because this is true for any lambda and nu right as long as lambda is greater than equal to 0. So even if I try and maximize this with respect to lambda and nu, the best optimal value that I can obtain which is the maximum of this particular thing that is again going to be upper bounded by p star. So p star is the best estimate of the optimal like the maximum value of this Lagrangian ok and your lambda and nu are your dual variables right.

So, this is an optimization problem in terms of lambda and nu. So, this kind of suggest that p star is also going to be if I try to maximize as long as lambda is greater than equal to 0 maximize lambda. So, this thing this is still true right because this is true for any lambda and nu and if I try to look at any equivalent problem which looks something like this. which is now defined in terms of lambda and nu right. So, this brings us back to the first problem that we looked at for the example that we started with which is this particular problem right. Now, lambda corresponds to inequality constraints or number of inequality constraints right the size of lambda and number of inequality constraints here are just r.

So, if r is much much smaller than n and if I work with the dual problem which is in terms of lambda and nu. then I need to work with much smaller size problem right and this makes things much easier and what are the constraints that we have just that lambda is greater than equal to 0 again much easier constraints to work with than something like A x less than equal to b. So, you see the like advantage of working with. So, in, but this time instead of working with the minimization problem it will be a maximization problem. So, maximize g lambda nu subject to lambda greater than equal to 0 and that is going to be your dual problem.

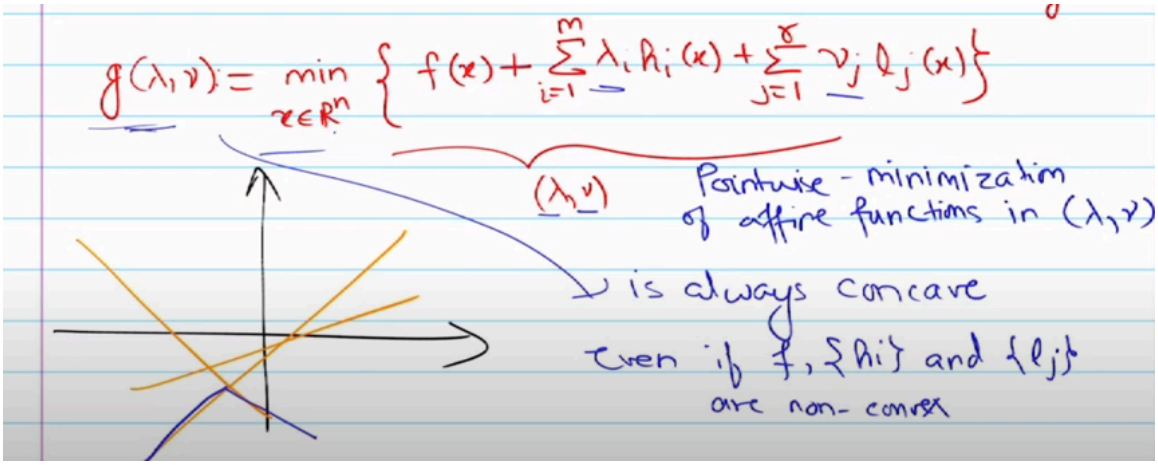
Is this clear? Any questions on dualizing your primal problem? So, again it is done in terms it is done using the Lagrangian dual which is g lambda nu and we know that this is

the best estimate that you can get even if you try to maximize this function is p star right. if it turns out that p star is equal to this particular thing and we are going to look at conditions and that is called strong duality. So, in general p star is so let us say this optimal value is d star. So, in general p star is always greater than equal to d star and this is called weak duality ok. If there are if you can show that p star is going to be equal to d star I can very well work with this equivalent problem right the dual optimization problem and by the way this thing is called strong duality and we can guarantee strong duality under certain assumptions and then we are going to look at those assumptions as well, but the idea is if strong duality holds then it may make sense to work with the dual optimization problem in certain cases then work with the primal one ok.



It is it is I mean in general it is not tight I mean it is I mean. Right right. So, yeah. So, one thing that like one of the properties that we should look at for this particular problem. So, again.

So, when we talk about minimization we think of convex functions right. When we talk about maximization then we should consider concave function. So, can we say something about this g? Is g concave? So, what is the definition of g? Minimize x in Rn f of x plus summation i 1 through m lambda i h i of x plus j 1 through r nu j l j of x. So, what can we say about this g? Is it convex, concave? When is it convex, when is it concave? So, first of all g is a function of lambda and nu, right? It is not a function of x, it is a function of lambda and nu. And what is, I mean how are we constructing g? through point wise minimization right.



For each x we are doing this minimization. So, for each lambda nu, we are doing

minimization over x. So, we are doing point wise minimization right. So, remember like in one of the lectures we looked we had looked at operations that preserve convexity and that one of them was point wise maximization. So, let us see how point wise minimization works. Suppose you have functions like this and so on right and let us look at the point wise minimization of these functions.

So, here maybe I will try to get rid of this something like this right like. So, what would be the point wise minimization of this function these set of functions. So, this function gets minimized at this point here the minimum would be something like this right. So, this looks like a concave function. So, just as point wise maximization is convex, point wise minimization is concave. So, and what about functions? So, g lambda nu, it is basically affine function of lambda nu.

So, pretty much like these functions right, affine functions are linear functions. So, point wise maximization of affine functions is always or point wise minimization of affine functions is always going to be concave. So, this g lambda nu is always concave even if f, hi and lj are non-convex. why because it is point wise minimization of affine functions in lambda nu okay.

So, this g lambda nu of always turns out to be a concave problem. So, maximization of g lambda nu it is always a concave optimization problem or concave problem and we know that we can solve I mean it is easier to work with convex just as it I mean it is easier to work with convex functions minimization of convex functions it is as easy to as I mean as easy as maximizing the concave function right. So, even if your original problem is non-convex dual problem is always concave that is that is one thing that you should keep in mind it is always concave in lambda nu is this clear. Yeah, the dual problem here which is going to be called, but the thing is like this particular minimization, if you are able to come like get rid of x and be able to present it in a closed analytical form, then you have an expression for G lambda nu right, otherwise the expression for G lambda nu need not be known. So, then it becomes I mean then it becomes difficult right. So, as long as you are able to get a closed-form expression of this and for simpler objective functions simpler functions f, h, i and l, j you would be able to then it is much easier to work with dual problems.

It is always a concave problem. Yeah, I mean I am saying that like you want to be able to solve this and let us see you are trying to maximize g lambda nu. I mean you know the functional form of f x h i and l j, but in order to maximize this function you also need to know the functional form of g right. So, at least this particular optimization problem either you are able to solve this analytically even if it is non-convex even let us say f and h and l are non-convex if you are able to solve this analytically then it is fine. If they are non-convex and you are not able to solve this analytically then you would have to run an algorithm to solve this first and then for a given lambda you get this expression right. Now there because if it is non-convex you may get some I mean you may get stuck at a local minima right.

So, you do not know if you have solved this correctly. So, in convexity even if you are

not able to solve it in a closed like in a closed form because it is a convex problem we know that we are going to get like we are going to converge to a global minima and therefore, even if you arrive at it numerically it would still be a correct estimate whereas, if it is a non-convex problem then it becomes challenging. Thank you very much.