Distributed Optimization and Machine Learning

Prof. Mayank Baranwal

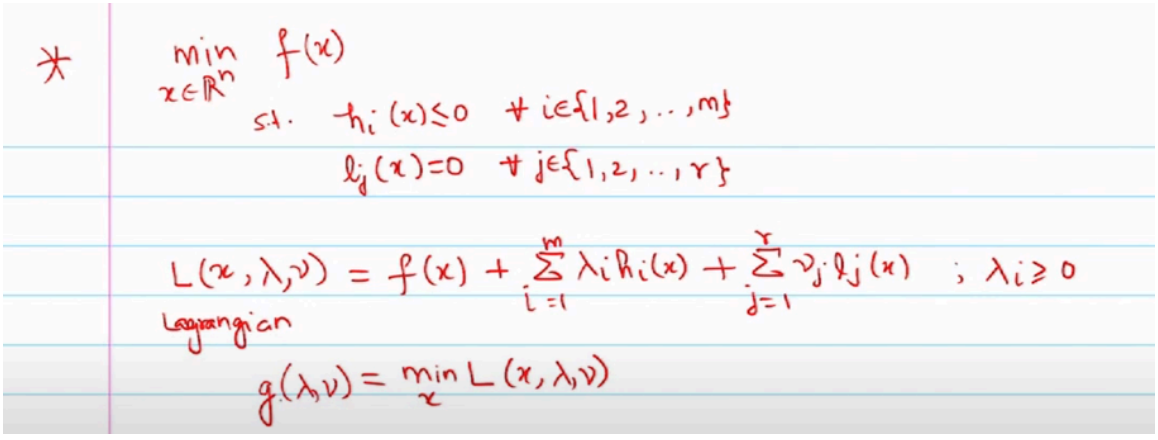
Computer Science & Engineering, Electrical Engineering, Mathematics

Indian Institute of Technology Bombay

Week-6

Lecture - 21: Constrained Optimization Problem

So, in today's lecture we are going to sort of focus more on constrained optimization problem. So, how can we potentially convert a constrained optimization problems to unconstrained optimization problem using Lagrangian right. We know that like if this is the optimization problem that we are going to be working with, the primal optimization problem subject to some inequality constraints of this form and equality constraints So you can basically formulate the Lagrangian for this optimization problem in terms of the primal variable x and the dual variables lambda and nu and this looks something like and this for lambda 0, okay. So, this is a Lagrangian and how do we define Lagrange dual for this? So, first of all it is a function of lambda nu, right, your dual variables lambda nu and this is defined as x Lagrangian, right, okay. So, if you were to let us say there is no optimality gap or there is no duality gap right, then you can very well maximize this Lagrange dual and the optimal solution in that case would turn out to be or optimal value of G star would be the same as the primal objective value right. So, in today's lecture, we are going to extend the FxTs-GF that we have looked at in the previous lecture.

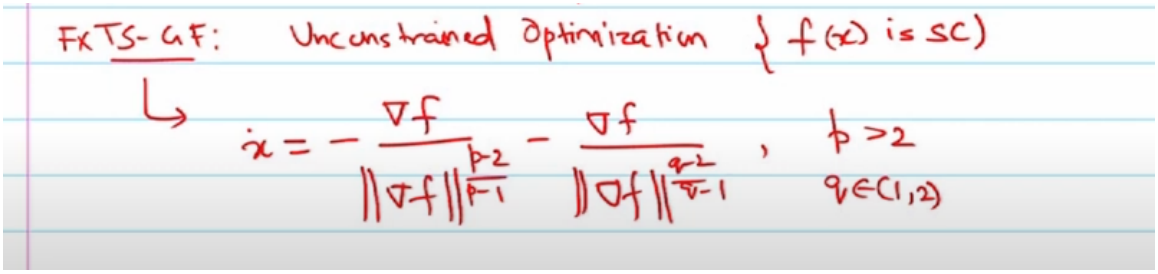


And again like this is not specific to just the fixed time convergent algorithms, but we are basically going to extend gradient flows for constraint optimization. So, the idea for today's lecture, we would be focusing on extending FxTs-GF for equality constraint optimization problems. and also for saddle point problems. So, eventually as I said

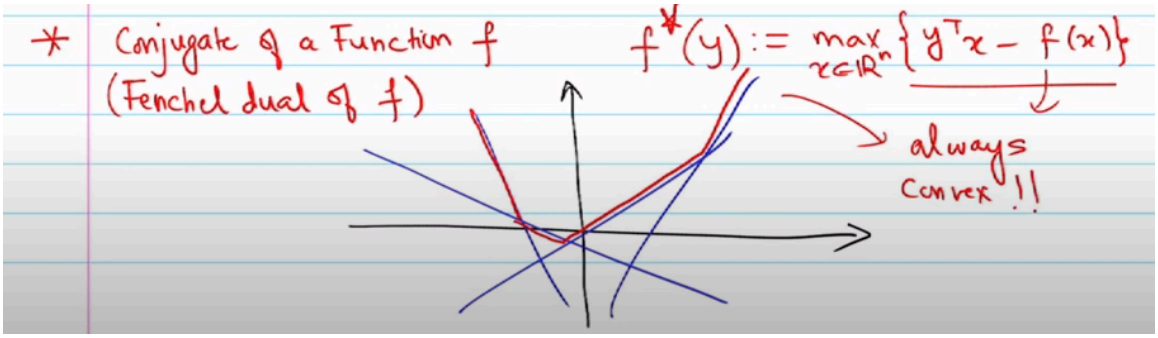
meanwhile you can analyze and maybe also formulate new algorithms in continuous time everything needs to be discretized right the implementation needs to be discretized.

So, in the so, in that spirit we are going to be looking at something called augmented Lagrangian methods also known as the method of multipliers. And without going too much into the theory of method of multipliers, we would be looking at the mechanism of convergence for method of multipliers. And in the subsequent lecture, this would also give rise to something called ADMM, which is Alternating Direction Method of Multipliers that is. So, Alternating Direction Method of Multipliers. So, the way this ADMM algorithm is set up, this would actually give you the glimpse of the first version of the distributed optimization algorithm or distributed optimization problem that we are going to be looking at.

So, we would start with something called at least motivate distributed ADMN and that would be a nice departure for a general distributed optimization problem. The next 2 to 3 lectures will focus around augmented Lagrangian methods, ADMN, distributed ADMN and then we would sort of start with something called distributed optimization, okay. So, just to briefly recap, if we are looking to minimize an like if it is an unconstrained optimization problem. let us say your f of x is strongly convex. So, we design something called FxTs-GF which was given as minus ok something like this, but what if you have a constrained version of this thing right.



So, let us say you want to minimize f of x, but subject to some equality constraint and how can we use or how can we design something equivalent of the unconstrained optimization like unconstrained optimization algorithm or the FxTs-GF that we looked at right. So we are going to look at something called conjugate of a function, so definition conjugate, it is also called Fenchel dual of the function and it is usually denoted by this f star, So, I am going to be using this type of star for dual and the usual f asterisk for the optimal value of f. So, in just be aware of the notations. So, I am going to be using this kind of star for the French dual or the conjugate of function and it is defined to be So, maximize this particular quantity, so and y sort of comes in here. So, what kind of function is this? Is this a convex function or a concave function or we cannot say anything.



So, first of all f star is a function of x or y. y right. So, what kind of in the on the right hand side what kind of for a given x what kind of functions in y do you have and affine function in y right. So, if you have affine functions in y that that may look something like these right and so on and I am doing what point-wise maximization and point-wise maximization gives me basically create basically gives you convex function right. So, point for if I look at the point wise maximization of this, it would look something like this right and this is a convex function.

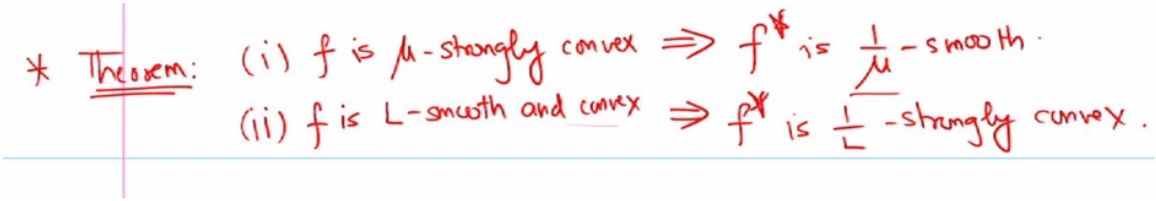
So, regardless of what your f is, I mean f may be convex may not be convex, your eventual dual of f or the conjugate of f that is always convex ok. So, this is always convex Is this clear? So, for a different f of x you will have different like hyperplane right. So, I am just checking the point-wise maximum. So, these different straight lines is basically given by different values of x and therefore, you will have different offset right f of x. for different values.

So, we fix our y right. So, f star y is what we are trying to find. Now, different values of x will give me different hyperpanes and I am doing point wise maximization and that is how I am evaluating right. So, point wise maximization is what I am evaluating. So, the bottom line is this particular conjugate of f that is always convex.

I mean even if your f is I mean not convex to start with right. And this is something that we had already seen in the context of Lagrange dual as well So this was always a concave problem or concave problem because this time it is a minimization over like if I look at the look at the definition of this function right it is a fine in lambda nu. So the same idea holds but this time it is minimization. So point wise minimization is basically gives you a concave function. So this is always a concave function even if your original function f is original problem is not convex to start with right.

So, therefore, you can that is why we look at the maximization of the Lagrangian dual because it is a concave problem. So, for concave problem we look at the maximization as an objective and for convex problem we look at minimization as the objective. So, I am

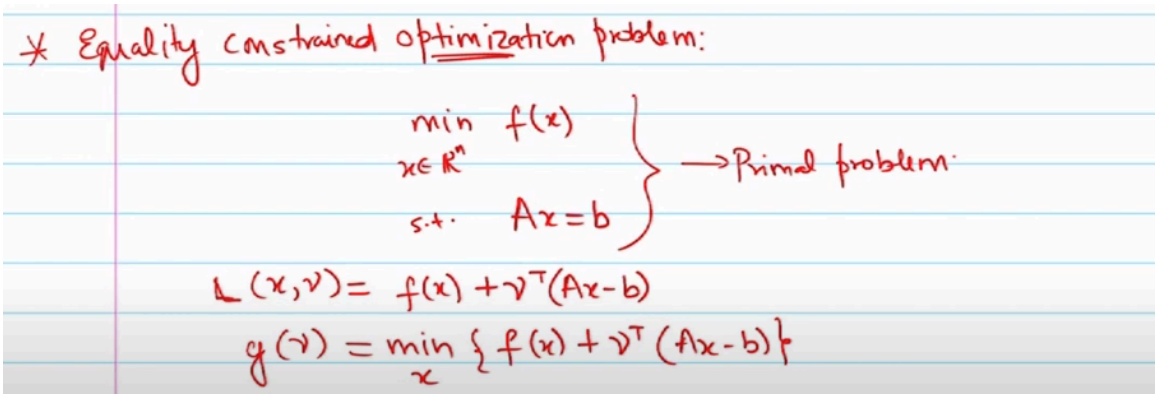
going to state a theorem again without most likely this will be part of your homework So if your function f is mu strongly convex, then the conjugate of f, so f star is 1 over mu smooth, okay. So like we had L-smoothness when we talk about a function being mu strongly convex and l So, if f is just strongly convex then f star is 1 over mu smooth. So, the smoothness modulus is going to be 1 over mu and f star is in any way we know f star is always going to be convex.



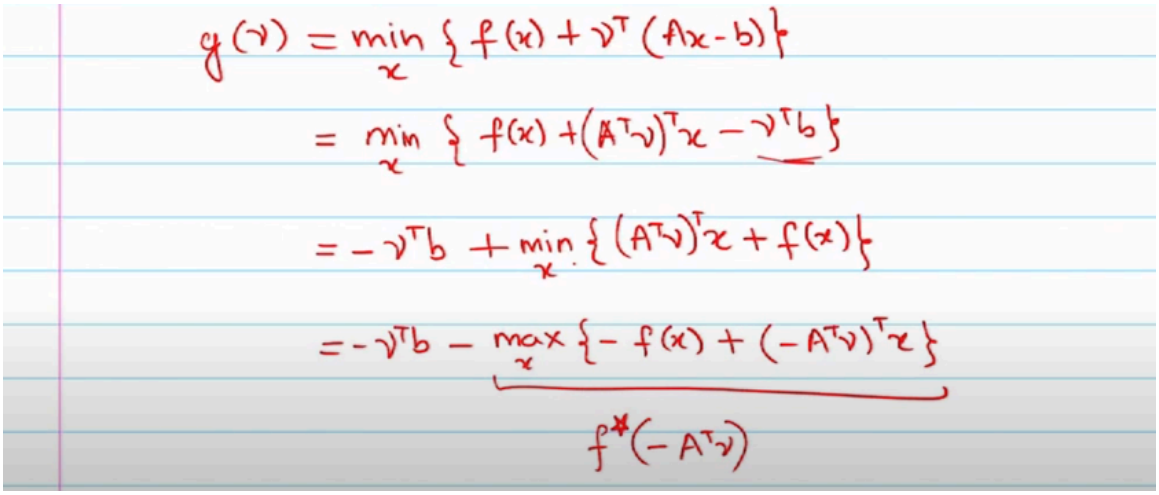
So, it is not just smooth it is also going to be convex, but it is all it is going to be 1 over mu smooth ok. The second result is if f is L-smooth and convex then f star is 1 over L strongly convex. So, if I do not include the convexity of f, f star is going to be convex no matter what right, but by including this convexity and L-smoothness we can guarantee that the f star is going to be strongly convex. In fact, the modulus of strong convexity is going to be 1 over L. So, it I mean basically the corresponding coefficients kind of they become reciprocal of the original model ok.

So, the smoothness becomes strong convexity and strong convexity becomes smoothness and so on. So, that this is one this particular result I will probably have you guys prove this in your homework, but that is the idea. So, let us look at the equality constraint optimization problem. in fact, linear equality constraint. And let us try to motivate why we are looking at something like a frenchel dual or the conjugate of a function.

So, we consider this problem minimize x in Rn f of x subject to equality constraint of this form A x equal to b. So, this is your primal problem. So, what is the Lagrangian for this problem? L x nu that is going to be f of x plus nu transpose Ax minus b. And if I try to define the Lagrangian dual which is g nu that is going to be minimum over x f of x. So, I can expand this and I can write this as f of x plus A transpose nu transpose x minus nu transpose B.



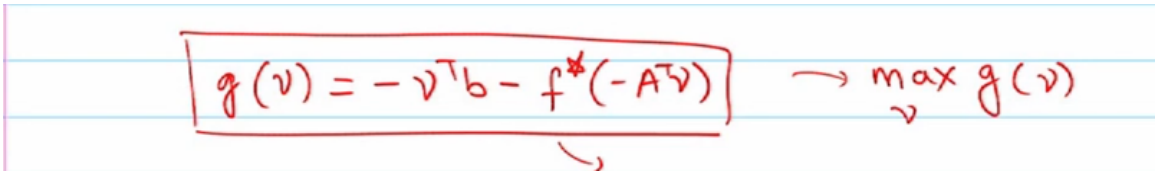
Is nu transpose B a function of x? No, right. So, I can just simply write this as minus nu transpose B plus minimum over So minimum over x is same as, minimum of a function is same as maximum over the negative of the function, right. So this, okay. and what is this quantity by definition? This is conjugate of f evaluated at minus a transpose okay, is this clear? I think it was right before wait let us see minus yeah here. So, this is your, so you are going to be getting a negative value here right first and then you are going to be making negative of it.



So, that the same the same value appears. So, this is fine. No, you can take it the argmin would be argmin or argmax would be the same, but then like let us say this evaluates to 3. negative of this if I try to maximize it becomes negative 3 right and I have to change it back to 3, I mean because it is a value that I am trying to add right. So, g nu becomes negative nu transpose b minus f star a transpose b, f star is what type of function? right, negative of convexes and g we know is always a concave function right.

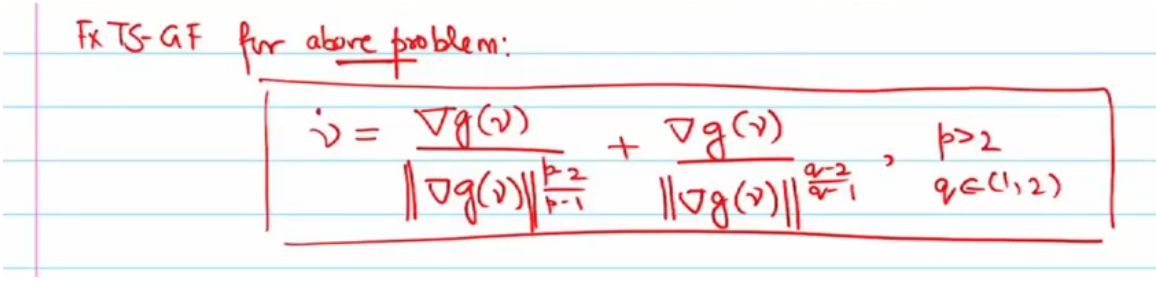
So, now you see that there is a connection between the Lagrange dual and the conjugate of the function right. So, let us say if I give you an f and I mean if I give you an f and you

know how to evaluate its or let us say the closed-form expression for the f star is known. Then you would not have to, I mean then you can, what you can do is you can simply try and look to maximize this g nu, right. So essentially you can pretty much work with this unconstrained optimization problem which becomes a maximization problem of a concave function. Is this clear? So if let us say you happen to know the closed form expression for f star given an f.

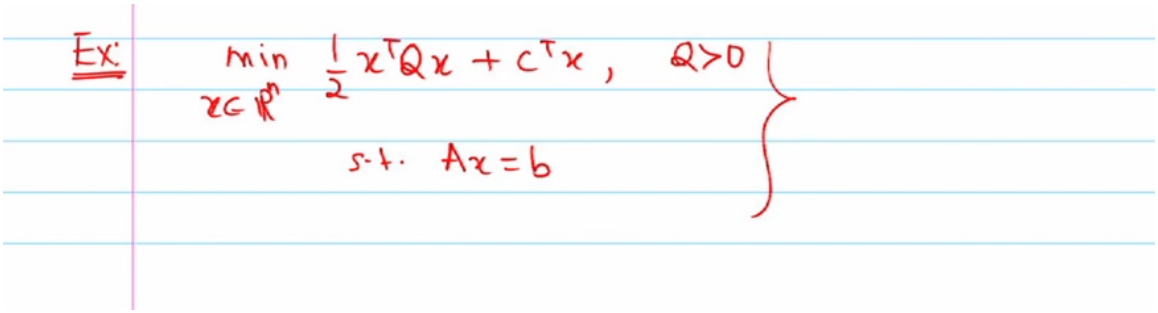


If a closed-form expression for f star is known, then you can simply evaluate g nu. and the dual problem is basically maximize with respect to new g nu right and this is an unconstrained optimization problem. So, you can pretty much treat like if you happen if you happen to know the conjugate of a function you can pretty much treat this as a as an unconstrained optimization problem and directly look to maximum like basically maximize this dual right. So, what kind of algorithm can achieve that? So, like if I were to design an FXTS kind of thing.

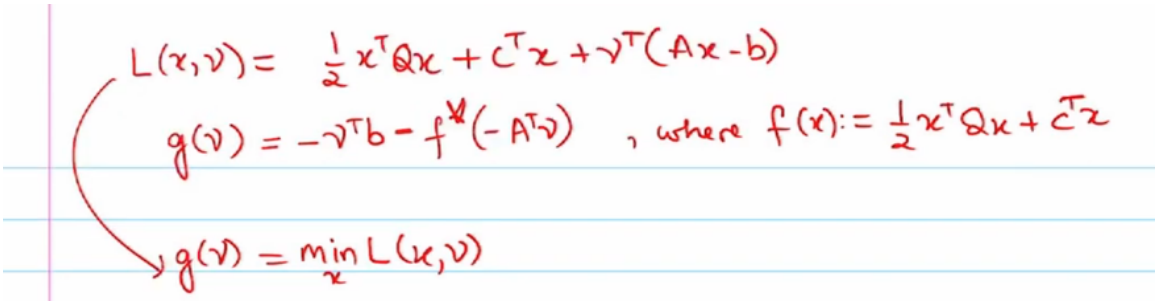
what would be my x new dot. So, first of all it is now I mean I am going to be defining in terms of new dot right. Because I am going to be running this algorithm for the dual variable. So, new dot is going to be Yeah so instead of minus gradient it will now be plus gradient right because it is a maximization problem. So you are going in the direction of maximum increase. So this would be gradient So, the equality constraint optimization problem again if you happen to know the corresponding conjugate of the function, you can treat this as an unconstrained optimization problem.



In fact, it is a maximization problem now. So, you run this particular algorithm with without the minus sign right. Is this clear? So, let us look at one example and let us try and evaluate the frenchel dual of a function f right. So, the primal optimization problem is minimize with respect to x half. So, it is a simple quadratic program Let us also assume that q is positive definite.



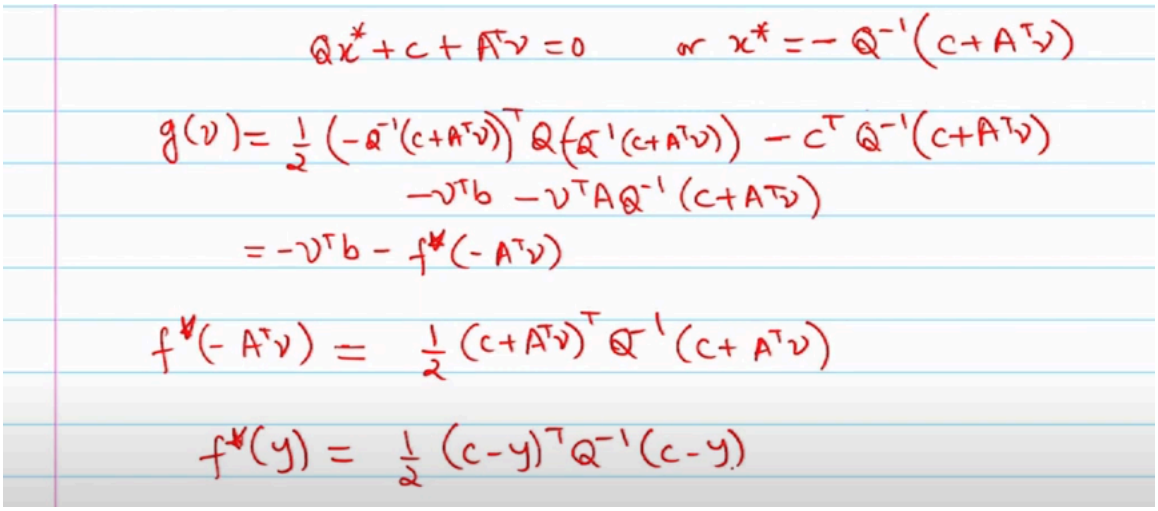
So, it is a at least strongly convex kind of quadratic program subject to Ax=b. So, can you try and evaluate the dual of this function dual of x. So, this is your f of x right. So, what is a Lagrangian here? Again, it is going to be simply half x transpose q x and g nu is going to be minimization with respect to x. So, that is going to be negative nu transpose b and we know that it is going to be minus f star of a transpose nu right, minus a transpose nu where f of x is half x transpose qx plus c transpose x.



So, if I try and find g nu directly from here which is going to be minimization with respect to x. So, that means I, so this is what I am trying to find and this is an unconstrained minimization with respect to x. So, for that I need to set the derivative of this Lagrangian with respect to x to 0 right that would be the condition for optimality. So, what is the derivative with respect to x for the first term? Let us call it q x star just to show that it is right or x star turns out to be okay. And let us substitute this back and try to find g nu.

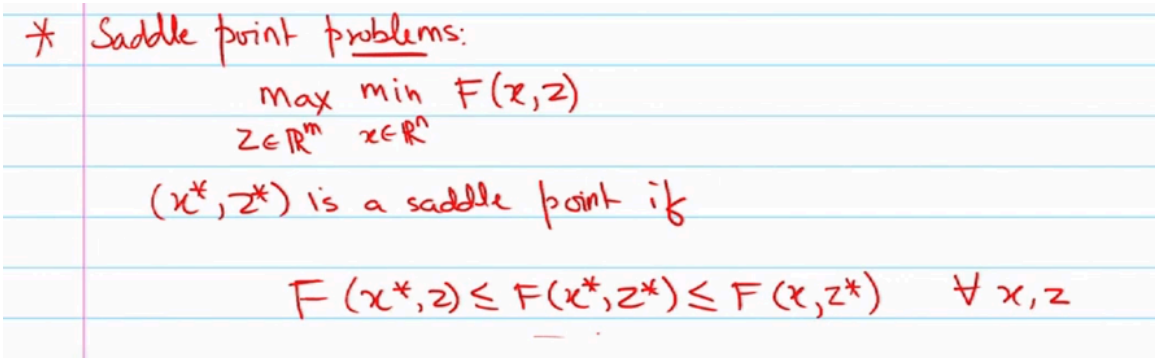
So, which is a function which is going to be a function only of nu right. So, g nu would then be half q inverse c plus So, you see a bunch of term in terms of A transpose nu right. So, all you need to do is you need to write this in a form and then you would be able to find the f star of this thing right and if you rearrange the terms let me directly write this or do you guys want to work it out, but if you rearrange the term it turns out that f star of minus a transpose nu is essentially minus half c plus a transpose nu transpose q inverse not minus half this is a half and this basically gives you what f star y would be and So f star y turns out to be half c minus y transpose q inverse c minus y for this particular like function f. So for quadratic problems or quadratic programs of this

form or for maybe similar analytical functions, sometimes you may have the closed-form expression of f star directly known. So, you can say right I mean use it right as it is and directly work with the Lagrangian dual and it becomes a gradient ascent problem on the Lagrangian dual.



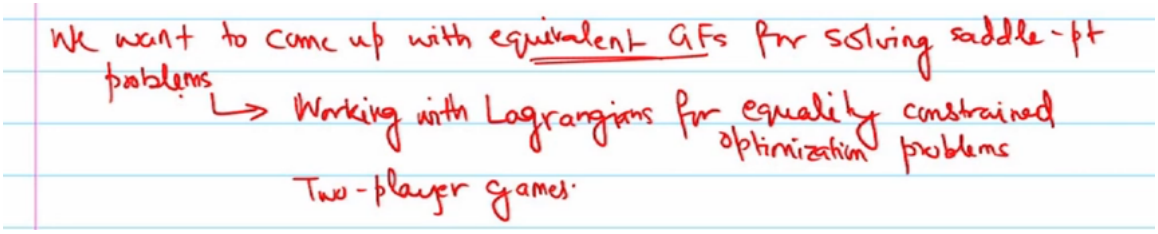
Is that clear? So, that is how you can work with constraint optimization problems, equality constraint optimization problems. We have not looked at how to sort of include inequality constraint optimization problems, but equality constraint optimization problems using conjugate of functions you can work it out. The other class of problem that we looked at was the saddle point problems. So, the saddle point problems are of this form. So, you have maxmin let us say kind of problem, you want to maximize with respect to one variable, minimize with respect to another variable a function f of x nd z.

okay and we say x star z star is a saddle point if. Then we say that x star z star is a basically a saddle point for every x and z. Do you see this kind of problem anywhere? At least in the context of the course, I mean in this course, have you come across any problem of this form? What about this Lagrangian? You are trying to minimize with respect to x and maximize with respect to lambda nu right? So it is essentially a saddle point problem on the Lagrangian directly. So instead of directly let us say I mean instead of trying to take this route of conjugate of f, I can also look at it as a saddle point problem on the original Lagrangian right where I am trying to minimize with respect to x and maximize with respect to nu. we are still focusing on the equality constraint optimization problem.



So, minimize with respect to x and maximize with respect to nu and that is what I am trying to find. So, I am trying to find x star and nu star, which is the saddle point of the Lagrangian that is one way you can look at it this look at this kind of problem. A lot of optimization problems I mean by themselves are essentially saddle point problems. So, in game theory for instance two-player games right it is a max min kind of objective. So, there you naturally I mean see the saddle point problems.

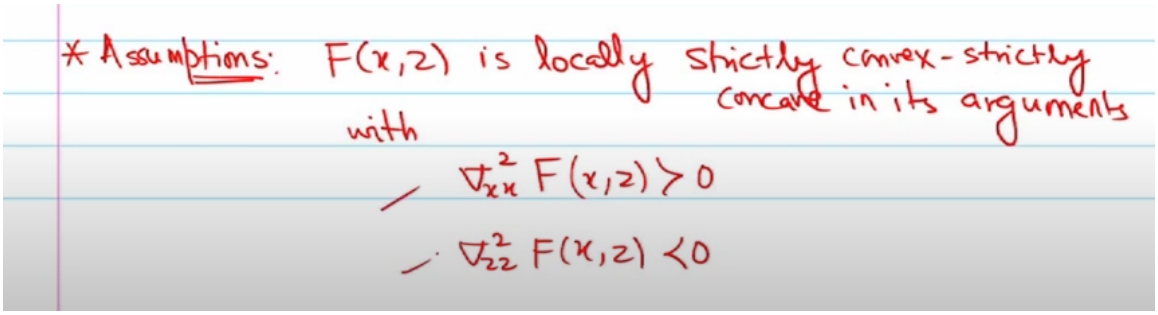
So, how like if you if you were to evaluate the saddle points. So, how can we sort of modify FxTsGF kind of thing. So, the idea is now. So, we want to come up with equivalent gradient flows for solving saddle point problems right. And in terms of the application of saddle point problems, so one example is when working with Lagrangians right.



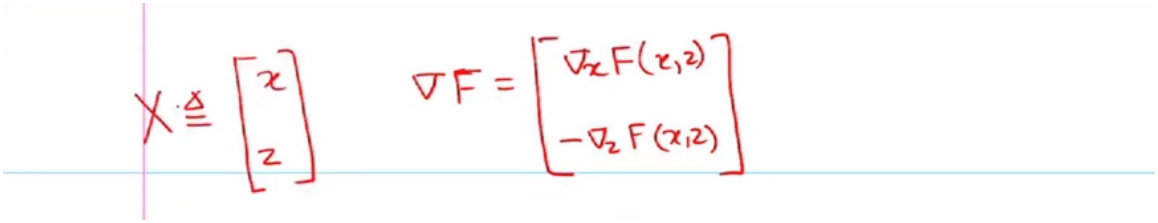
for equality constraint optimization problem let us say or another example is two two-player game where you have you naturally have this max min or min max kind of problem ok. So, if you have something like this, so how do we come up with an equivalent gradient flow for solving the saddle point? And again as I said I mean we would be looking at the fixed time converging gradient flow, but it is not specific to fixed time. As I said I mean fixed time is essentially basically reparameterization of your original gradient flow. So, if it works for the fixed time convergent gradient flow, it would also work for simple gradient flows as well, it just said convergence would be faster with this.

So, quick some assumptions. So, if you were to use fixed time convergent gradient

flows, I mean functions should either satisfy PL-inequality or they should be strongly convex or they should be strictly convex with Hessian-like positive definite rate. In case of saddle point problems f cannot be convex in both the arguments otherwise it becomes like it for the for the variable in which we are trying to minimize this function you assume f to be convex in that variable and concave in the other then this max min problem makes sense. So, we are going to be assuming f x z is locally strictly convex strictly concave concave in its arguments. So, meaning it is basically strictly convex in x and strictly concave in z with we are going to be assuming that the. So, this is going to be or rather for every x in z let us say we assume that this is positive definite and because it is strictly concave in the other argument this is going to be negative definite.

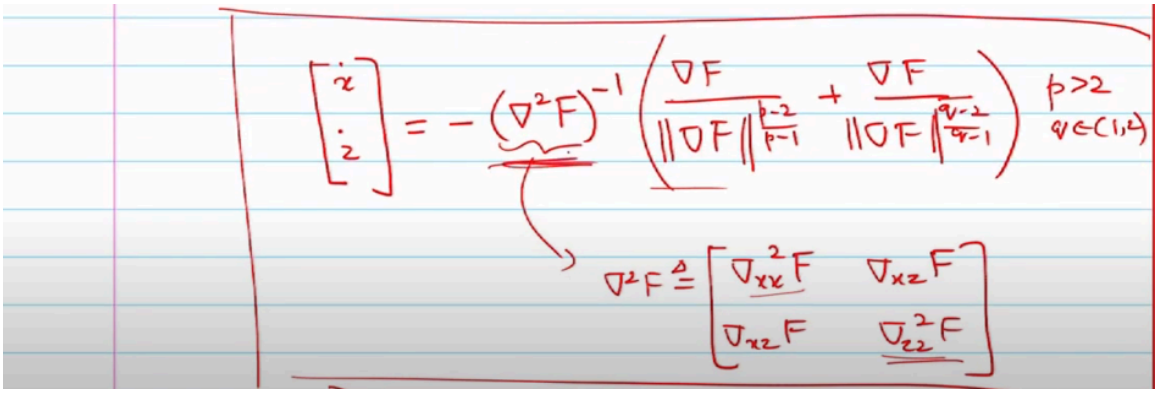


So, we assume that this is this is the setup. So, then if these two conditions are satisfied then you can say it is strictly I mean strictly convex strictly concave. I mean the other way around need not be true as we have already seen in the context of let us say x function like x to the 4, h n is not always positive definite right, but it is strictly convex. So, we assume that this is true. So, essentially the Hessian equivalent Hessian is going to be invertible. So, what do we need to do if we were to solve this problem? So, it should look like a gradient descent in x and gradient ascent in z right, because we are trying to minimize with respect to x and maximize with respect to z.



So, the kind of algorithm that you would end up designing would look something like this. So, first of all let me define this capital X which is going to be x and z and gradient of F So, we define this new gradient vector which is basically concatenation of the gradients with respect to x and z, but for the z vector we are going to have a negative of it right because we want to run gradient ascent on z. So, the equivalent for x dot is going to be where this is nothing but the Hessian the f is nothing but your Hessian which is again

with p greater than 2 and q in. But you can see because we have defined gradient here with the negative sign.



The Hessian of f is positive. Hessian of f is invertible is what we can because here it is going to be negative right. Let us say there are no cross terms. So, this is positive definite, this is negative definite, but then I mean it is invertible is all you can say with because of the assumptions that we made. So, this inverse exists and this gradient f is actually defined in terms of I mean instead of using the gradient with respect to x and z and concatenating it there is a negative sign here because we want to run a gradient ascent on the z variable.

So, this particular this particular gradient flow. So, this so, converges in fixed time. for the saddle point problem. If let us say the problem was strongly convex strongly concave to start with then you would not even need this right. So, then it would have simply been this particular term had the problem it is strongly convex strongly concave. Something that we like we have already looked at in the context strongly convex functions only that if we were to minimize strongly convex function then then you do not need the Hessian inverse right.

Hessian inverse is needed when you work with strictly concave, strictly convex kind of setting. So, again proving convergence in fixed time is going to be one of your homework problems, but this is pretty much the modification that you need to do. Yeah, so you assume that the saddle point exists. I mean if you are assuming say strict convexity, strict concavity saddle point would exist, but in general I mean like yeah you assume that saddle point exists. I mean the kind of assumptions that we are making strict convexity, strict concavity would guarantee existence of a in fact unique saddle point if you, but yeah.

So, well I mean in this case we I mean we have the global kind of definition here. Yeah, I mean this particular definition is called locally strictly convex, but I mean we assume that it is true for every x and z. So, but then it is locally strictly convex, strictly concave if this

is true in some ball around here, in some ball around here x and z. So, for now, you can just I mean for the sake of simplicity assume it is globally strictly convex, strictly concave. So what would be a good, let us say if you want to show that this particular optimization algorithm or this particular dynamical system converges to equilibrium in a fixed time, what would be a good choice for Lyapunov candidate? So what is a good choice of Lyapunov candidate when you work with strictly convex functions? Half norm gradient, right.

So even here you are going to be working with this when you solve the homework problem, this is the Lyapunov candidate that you are going to be working with. And you would need to show that V dot is less than equal to some c1 V to the alpha 1 and a c2 V to the alpha 2 with alpha 1 less than 1 or an alpha 2 greater than 1. then you would you would be able to show that this convergence in a fixed time. So, this is what you need you would need to show.