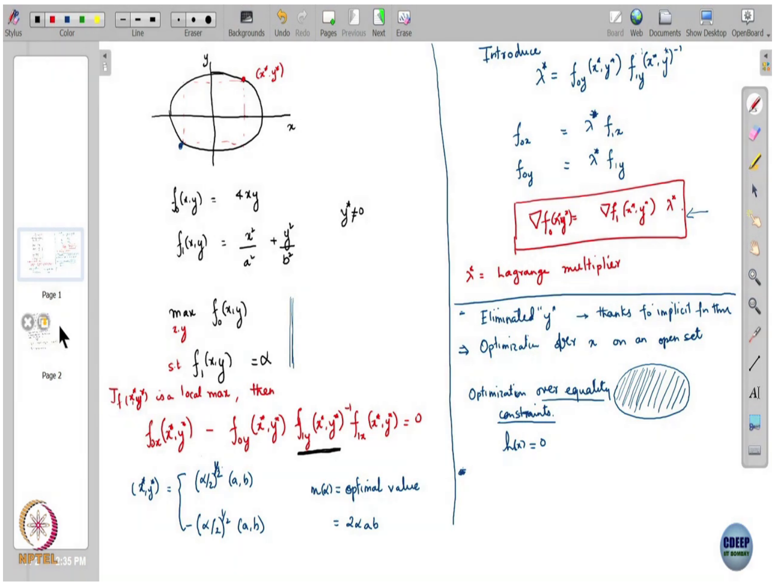
Optimization from Fundamentals
Prof. Ankur Kulkarni
Department of Systems and Control Engineering
Indian Institute of Technology, Bombay

Lecture – 6A
Optimization with equality constraints and introduction to Lagrange multipliers - I

Alright. So, let us continue where we left off in the previous class. We were discussing this problem where we had a ellipse in 2 dimensions.

(Refer Slide Time: 00:25)



So, this is a problem in r 2, then horizontal axis to be denoted by x the vertical axis is denoted by y. And so we were looking at points that were lying on the surface of this ellipse that is the black trajectory that is shown here ok.

And you can point like x star y star which lies on the surface of the ellipse and we wanted to find a rectangle with whose end points are whose corner points are on the ellipse that has the maximum area. And for simplicity we said let us take the rectangle to be aligned with the coordinate axis. So, the axis of the rectangular aligned with the coordinate axis ok.

So, the mathematical problem we posed was we have f 0 of x comma y which was the area of a rectangle whose corner point here we say x comma y, that is that you can see is it can be given by 4 x y and x comma y should lie on this ellipse; that means, x comma y should satisfy the equation f 1 of x comma equals alpha that is this equation. So, x 1 f 1 of x comma y equals alpha where f 1 is the ellipse equation; so, x square by a square plus y square by b square ok.

So, any x comma y that lies on the ellipse basically satisfies x square by a square plus y square by b square equals alpha alright. So, over all such points that satisfy this equation we wanted to find the x comma y that maximizes the area. So, our objective was to maximize f 0 of x comma y over x comma y such that or subject to the requirement at f 1 of x comma y equals alpha.

And, now assuming x star y star is a local maximum. So, if x star y star is a local maximum, we said that then it is necessary that this equation holds. The equation here says f 0 x I evaluated at x star y star minus f 0 y evaluated at x star y star times f 1 y evaluated at x star y star inverse. So, that this is divided by f 1 y of x star y star times f 1 x of x star y star this must be equal to 0.

Now we, and we said we will make the assumption here that x star that y star is not equal to 0 so consequently f 1 y of x star y star was not 0. So, it is legal to divide by f 1 y of z star y star alright. So this, so we derived we ended the class at this step that where we derived this condition.

So, now, this condition can be interpreted in another way. So, let me take off from here and introduce for you, ok before we do that actually let us evaluate also what the solution turns out to be. So, this it will turn out if you solve this equation now you get that x star y star is a local maximum then this must hold you can solve this equation you get that x star y star the values are something like this.

So, you get two solutions one of them is alpha by 2 square root of alpha by 2. You get two solutions one is square root of alpha by 2 comma times a comma b the other is negative of square root of alpha by 2 that means a comma b ok. So, these are the two solutions that you get. And for both of them the area is actually the same the area the optimal area let us call that m of alpha which is the optimal value that turns out to be equal to 2 alpha a b ok.

So, now, they you get these two solutions the reason for that the reason, but you can check that both of them are local maxima essentially what they are corresponding to a point x star y star here and its reflection on this side, this other point here which has negative x star comma negative y star ok. They are both giving you they can both either of them can be taken as the solution ok.

So, now let us write this condition this red equation here in a slightly different way. Let us introduce this introduce this notation let us call let us lambda star be denote this quantity which is f 0 y of x star y start, f 1 y of x star y star inverse at lambda star denote this. Then in that case I can write this the red equation in the following form I can simply say that f 0 x so I will drop the x star y star for simplicity f 0 x is equal to lambda star times f 1 x.

And this same equation that I have written here the blue equation that I have written here this same equation I can just I can take f 1 y f 1 y of x star y star inverse to the other side and write that in this form I get that will give me f 0 y equals lambda star of f 1 y. In other words what we have we can put these 2 together and write it like this.

If I take the gradient of f 0 and that is at x star y star; at x star comma y star that is equal to the gradient of f 1 at x star y star times lambda star; so, the earlier red equation which is kind of complicated looking can be simplified in this sort of form if we introduce this other notation other quantity lambda star.

Now, lambda star has an important you can see what this thing is doing, it is basically giving you an equation it is giving you a condition that says that the gradient of the objective evaluated at x star y star is actually just a scalar multiple lambda star of the gradient of constraint evaluated at x star y star ok. So, in particular the gradients are actually collinear one is a just a scalar multiple of the other alright.

So, this condition is what we will generalize now. The lambda star here is has a name it is called Lagrange multiplier. So, what we can do is we can the way we instead of writing a complicated equation like the earlier red one what we will do is, we will introduce this additional variable called Lagrange multiplier and write another equation that is in terms of x star y star and the Lagrange multiplier alright.

So, that is what we will be we will do now. But, before I get to doing this in more generality I want you to carefully understand what exactly we accomplished when we got when we solved this particular problem in this way ok. So, and then that will also give you motivation for why we should be considering the Lagrange multipliers and so on.

So, how did we go about solving this problem? We said we have a way of solving we have a condition which says that which is a necessary condition to get to a solution for a point to be a solution of an optimization problem over an open set. So, when your maximizing or minimizing a function over an open set we know how to address that problem.

We said we will took this particular problem which was not over an open set and we said let us address this in some way. So, what did we do? What were the steps we followed? So, the steps we followed was the first step was, we eliminated the y variable right. So, we eliminated one of these variables and this was how why were we able to do this? We were able to do this thanks to the implicit function there.

Then we will eliminate this, we eliminated this y it thanks to the implicit function theorem that then gave us as an optimization over x so this will give as an optimization over x on an open set. Now, when we applied the implicit function theorem remember what would we what did we use?

We used that around this point x star y star here we were always on the surface of this particular ellipse right. So, can you could we have applied the implicit function theorem? If we did not have this particular feasible region, but rather a feasible region that look like this, you have the ellipse and also the shell of the ellipse and also the interior.

If you have a problem like this when you have both the shell and the interior right, you it is not possible to say that you would be able to eliminate one variable in term and get it in terms of the other. You can do it provided your constrained to be on that surface because that surface gives you this additional equation which let us you solve for one variable in terms of the other.

But if you can be both on the surface as well as inside there is no definite equation that you can involve right. So, it is. So, the kind of constraint for which we can do all these steps that we followed are those constraints when we are optimizing over surfaces. So, surfaces take the forms form h of x equal to 0 or in this case f 1 of x comma y equal to 0 right.

So, this works for problems where we are optimizing optimization over equality constraints. So, when you optimizing a differentiable function with only equality constraints then we are effectively optimizing over a surface, and then that the equations of that define that surface let you eliminate one variable when get in terms of the other right.

So, if you had both the surface and its and the and its interior you know the shell of the watermelon as well as the you know the pulp of it inside all of that if you have then you have then you cannot invoke you cannot do this ok. So, this elimination of y also requires this.

we also later used that you know, later we again use the same thing, we also we after writing everything in terms of x we also saw that the we took the derivative of the constraint with respect to x. You can go back and check and put that equal to 0 that also is because we are always on the surface right. So, with that so, that again requires that you are we are we you have only an equality constraint ok.

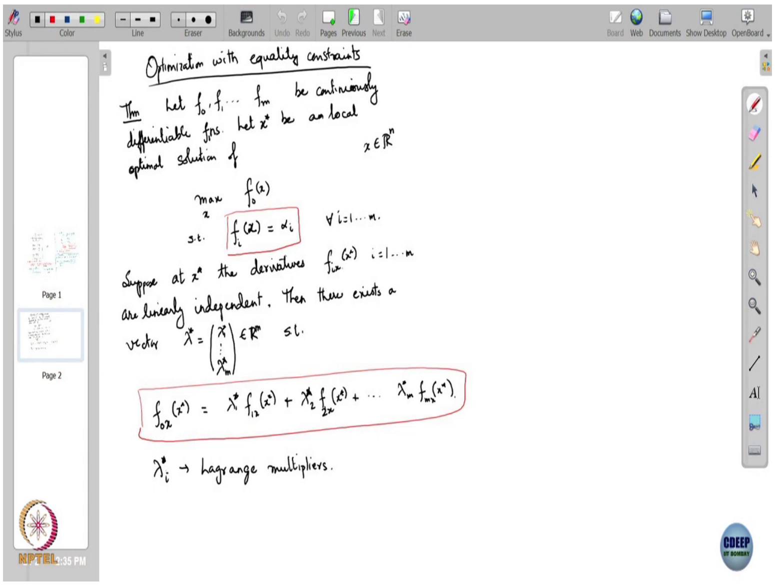
So, the thing that we what we have done here is basically we can we have if we can replicate all these kind of steps for general problems with equality constraints, we should be able to get a general purpose result for how to solve those sort of optimization problem.

That the at the last step we what we got was we solved for x star y star some of you may be wondering what how did I get this. Well, this in this case we solved for x star y star and we could get that I could get that in closed form it is not very hard, but in general you would have to do this numerically alright.

So, the point is to start the all of if all the effort in optimization is to start with a problem that is written in this sort of form and reduce it to something that looks like this, where you have now just an equation that needs to be solve and then you have some ready machinery to get to solve that equation ok.

So, all the effort is to see how problems that are posed in as this sort of in a decision making sort of form where you have to maximize some goal subject to some constraints and how do you get them down to a bunch of equations that need to be satisfied alright.

(Refer Slide Time: 15:53)



So, now; so, I will just state for you the general theorem for when you have an optimization problem with equality constraints. So, right so rather than prove it I will just give you intuition for how this relates to the previous example that we studied ok. So, let me state the theorem.

So, suppose let f 0 f 1 dot dot dot f m be continuously differentiable functions, let x star be a local optimal solution of this problem. I am writing in this in the maximization form just because of earlier problem was in maximization form. So, f 0 x is what you need to maximize over x and you have to satisfy this subject to the requirement that all these equalities must hold. So, f i of x equals alpha i for all i from 1 to m.

So, there are m different equality constraints here f 1 of x equal to alpha 1 f 2 of x equal to alpha 2 etcetera all of them should hold. So, we are optimizing over the common reason that in which all of them are satisfied ok. And over that reason we are trying to find the x that maximizes f 0 of x. Now, the theorem says the following, suppose you look at these derivations suppose at x star the derivatives f i x of x star for i equal to 1 to m not 0 to m i equal to 1 to m.

So, you are looking at the derivatives corresponding to the constraints ok, evaluated at x star. So, these are can someone tell me what length vectors are these? If x is in r n. So, x belong x we are optimizing in r n, what length vectors are these? f i of x this sorry f i x of x star yeah exactly so the derivatives right so there are n length vectors all of them row vectors ok. f i x evaluated at x star and there are m of these row vectors ok.

Now, so the n length vectors m of them suppose that these derivatives. So, they form the m of these vectors suppose these derivatives are linearly independent; suppose these are linearly independent ok; so, the ones at x star. So, the derivatives evaluated at x star these are linearly independent ok. Then there exists a vector lambda star and let us write this as a column vector lambda star is written as lambda 1 star dot dot dot lambda m star (Refer Time: 20:30) rewrite them. So, this vector is a m length vector is vector in r m.

So, there are its components it has one component for each constraint that is there in the problem there are m constraints in the problems, so there is one component for each constraint. So, lambda star is in R m such that we have this. So, I look at f 0 x evaluated at x star; that means, a derivative of the objective evaluated at x star that is can be written as lambda 1 star times f 1 x evaluated at x star plus lambda 2 stars times f 2 x evaluated at x star plus dot dot dot lambda m star f m x evaluated at x star.

So, you can see what we have done here, we said we started with a an abstract decision problem like this which has maximize this function subject to these requirements and what we have said is that well, if there if x star is a local solution then you must be able to satisfy these equations ok.

What is this, how many equations are there here? So, this is one equation, but one vector equations right. How many components are we talking of? n; n of them because these are all derivatives right. So, they are n length vectors so that are effectively n, n scalar equations here. How many unknown do we have? What are our unknown?

Student: (Refer Time: 22:27) unknown.

We x star is an unknown which we want to find in addition to that the lambda is also an unknown right. So, x star has is n lengths so there are n unknowns in x star m unknowns in lambda because lambda is of length m right. So, there are m plus n unknowns you have n equations that come from here you need some more equations, where would they, where are they coming from?

So, there are other equation there are m additional equations that we have which comes from the fact that x star must be feasible x star must be must lie in the is be must be feasible for the optimization problem. So, f i of x star should be equal to alpha i for all i so that gives you additionally m equation ok. So, all of these put together help you solve the optimization problems. So, you have these equations and you have these questions. So, there is this n plus m equations in n plus m unknowns.

So, what were done is basically then taken a problem like this and taken a problem this kind of abstract problem and said that this is it is necessary that x star satisfies x star satisfies a bunch of equations, but they need to be written in terms of additional variables not just x star not x star alone because now constraints are involved you need to introduce a few additional variables and we need exactly one additional variable per constraint

Student: (Refer Time: 24:06).

yes. So, the question here is what if the number of constraints are very large? So, what if m is very large? So, I will come I will discuss this matter. See if m is very large then you are intersect you are on the on several surfaces at once.

In fact, eventually you would the very fact that you have to satisfy m of these equations that itself can end up determining constraining x so much that you will probably even get just one solution right; if m becomes n for example, that itself determines what which point you are what your feasible reason is right.

So, it that can happen, but usually the then that is a case of a poorly formulated problem because you have constrained it so much that now there is not nothing to search really because the constants there is only one alternative for you effectively. So, that sort of problem can occur, but it is not a very interesting you know interesting things to study because it is probably not well formulated to begin with ok.

Let me so, again these here are the lambdas here are again called the lambda stars these are is a called Lagrange multiple. While we write this let us also just think I understand a bit about the role of all these assumptions here therefore that you see how this is backward compatible with what we have seen so far in that example right.

So, here the key assumption I have made here is that of course, these are all continuously differentiable that is one assumption that all these the objective constraints that these functions are continuously differentiable.

And then I said that if I look at the derivatives of the constraints look at the derivatives of the constraints they are linearly independent the derivatives of the constraints are all linearly independent.

Now, that is a hint to our previous assumption here that we had assumed that y star was remember y star was assumed as non zero y star was taken as not 0. So, we were because the reason was there we said we do not need to look at such points ok, but what y star non zero ensured was that this f 1 y at evaluated at x star y star is not 0.

The analogous think that we need here in this more general set up is that constrained derivatives are all are put together are linearly independent. Now, what that ensures is geometrically what ensures is when you are optimizing over all these surfaces. The intersection of all of these surfaces if the constraint derivatives are linearly independent then that let us you get let us you define what is the in terms of the derivatives the tangent surface to all of these surfaces ok.

So, this will become more and more evident as I as you go later into the course, but basically the tangent surface gets defined and once you have the tangent surface defined in terms of that you can, we can write out conditions like this, ok.