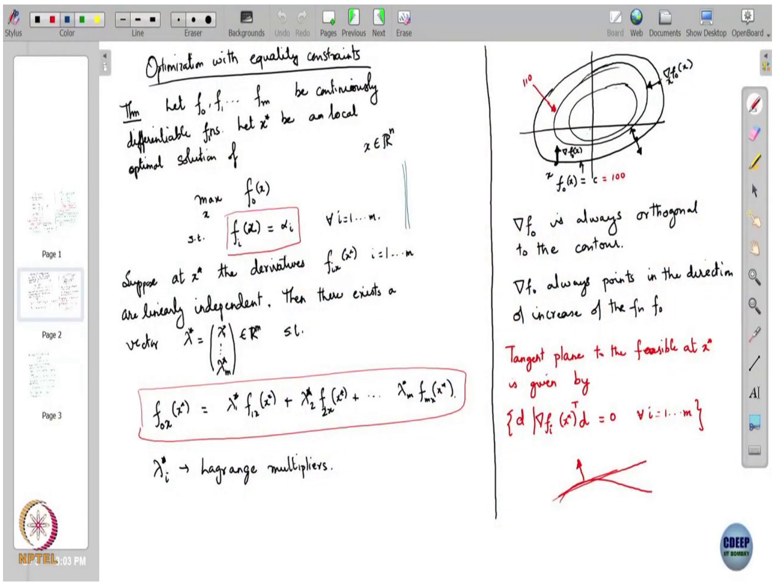
Optimization from Fundamentals
Prof. Ankur Kulkarni
Department of Systems and Control Engineering
Indian Institute of Technology, Bombay

Lecture – 6B
Optimization with equality constraints and introduction to Lagrange multipliers - II

(Refer Slide Time: 00:17)



So, let us look at what is happening here geometrically. Effectively what we are what is being claimed by this particular by this particular equation the box equation. Let us try to understand this ok. So, for to understand this we first need a basic idea of what the role of the derivative of a function derivative or the gradient of a function is, what does it tell you about the function.

So, any function if it is a function of say n variable I can this instead of plotting it in a way where I look at the n variables on the as independent variables and the value of the function in another axis instead of that I can look at the contours of the function right. So, I can plot a function like this is suppose these are the contours of f 0 ok. So, suppose these are the contours of f 0.

So, what does this mean? So, this means that your this is the locus of points. So, x so a contour like this is a locus of points where f 0 is a constant. So, it is all points x such that say f 0 of x is equal to c some constant c. And as I vary c I will keep getting different contours is contours were; obviously, not intersect because they correspond to different values of c ok.

And you may or may not get concentric contours like this that depends on the shape of the function I am just taking for simplicity this is how it that the contours look like this. Now, what is the role of the derivative or the gradient? So, if I take a point like this here and I say well the gradient is a vector as long as x itself, so I can draw the gradient with origin shifted here suppose.

So, this is the vector grad f 0 of x, but I have just instead of, what I have done is I have just shifted the origin of that vector to this point x itself to this particular point. Likewise at this point this would just for indication I am just drawing the gradient in this sort way here.

If this is my x then if this is the point x then this is the gradient; this is the gradient. Now, what does this, what does the direction to which the gradient is a vector its pointing in a certain direction if I look if I shifted shift its origin to x its pointing in a certain direction, what does that direction tell me about the function? So, the, so first thing is the gradient is always normal to the contour ok.

So, the gradient if the function is differentiable than is always normal to the contour. Why is that? Yeah right. So, contour comprises of points there where the function values are constant right. So, the gradient and were as. So, effectively what must be happening is that the, so the gradient is you can think of as if in the scalar case is basically the derivative effectively the derivative along these points has to be as you travel along this should start vanishing right.

So, you should if you take the rate of change of the function as you travel around the contour that should vanish. What that if means in R n is that the gradient itself must be orthogonal to this to the direction in which you are travelling while you are on the contour, alright.

So, the gradient is always orthogonal to the contour. So, gradient is orthogonal to the contour, but which direction does it point in? So, there are two orthogonal directions right; one is a direction here and the other. So, for example, if this is a point I could talk of this orthogonal direction I could also talk of this orthogonal direction yeah.

So, the this is the other thing about the gradient. So, this direct the gradient always points in the direction of increase of the function of the function f 0 ok. So, the way to see this in the in this contour plot is to see suppose here c let suppose the c is equal to say 100 and this c inner c here is suppose this one is say 110.

So, the gradient is pointing inwards means it is pointing towards the direction in which the contour values keep increasing right. So, if this was reversed if this was not 100 say it was 90 if the function was function values were decreasing as you went in into inner and inner contours then the gradient would have been pointing outwards in this direction ok. Inwards or outwards depends on what happens what is happening to the function value you cannot tell that from just looking at the shape of the contours ok.

So, the gradient is always pointing in the direction of increase of the functional 0 ok. So, now let us go back to this boxed condition, what is this condition saying? It saying that if you take the derivative of the function f 0 it is given as it is a linear combination of the derivatives of the linear combination of the derivatives of the constraints right ok.

So, what does this what would that mean? That means, the direction in which you are the direction in which f 0 was increasing at x star, if you look at the point x star and look at the direction in which f 0 was increasing that is the direction of the gradient, if you to look at that direction that direction is itself a linear combination of the directions of the gradients of the constraint. Now, what is that; what is that further mean? Satisfying the constraints also yes.

So, your you are on the surface correct, you are on the surface? What you are on the surface and then what does that say? So, let very good. So, let me write it in this sort of way. So, the linear. So, once we have linear independence of the derivatives what we can say is that the plane tangent plane to the surface or the common surface area of common area in all surfaces right. The tangent plane that is the tangent plane the feasible to the feasible region at x star is given by all directions d such that.

So, it is oh sorry. All directions d such that if I take the derivative at x star that times d or I equivalently; let me this in terms of the gradient, gradient at x star transpose d this is equal to 0 for all IN. So, if you take those directions d for which are such that the gradient is that such they are orthogonal to all the individual gradients. This means if you take the subspace band by these gradients b is orthogonal to that subspace right so because it is orthogonal to all the individual gradients.

So, that those these are your tangent directions ok. Now, if your f 0 of x if your f 0 of x ok. If it f 0 of x lied in one of in that tangent direction and what does that mean, if the gradient of the objective was in that tangent direction, then what would that mean? What it would mean is that you could increase your function while also gently going along the tangent to your surfaces. Means there was scope for you to be on the surface and at the same time increase the function further.

So, there is these surfaces that is, so you have they have all these surfaces that are intersecting the tangent to them is orthogonal to the space to the subspace formed by the gradients of the constraints. And your objective gradient also lies in that subspace because it is a linear combination of the gradients of the constraints right.

So, the directions that you need to go in order to be on the surface is orthogonal to the direction in which you need to go in to increase the function. So, this is the compromise that gets reached here essentially when the if you are at local maximum of this problem, then this has to be true, that you cannot both be on continue to be on the surface and also increase the objective value for them.

Because if you could increase the objective value and also be on the constraint and stay on the constraint. Then the gradient which is the direction in which objective value increases that gradient would have some positive component on the on the subspace in which you need to be back to remain on the surface right.

So, if you had to remain on this particular surface and then for that you need to be you need to travel along tangent to it, but your gradient is pushing you in this sort of direction orthogonal to that tangent right.

So, the gradient here is a linear combination of the of the gradient of the objective is a linear combination of the gradient of the constraints whereas, the tangent space is orthogonal to this to the subspace formed by the linear combinations of this of these constrained gradient is this clear.

So, this is basically the condition that that is being expressed here that these are two orthogonal directions, if there was even a slight component between of the gradient of the objective onto the subspace. Then there would have been scope for you to travel further in that direction and you would have potentially gotten a better object ok.

So, that is what this constraints condition effectively says, that its effect it is just expressing that the very fact that you are at an at a local minimum means that this is the very least local maximum means that this is the very least that must be true. Is this clear ok?

So, this is the; this is the geometric interpretation behind what behind what is happening here. Now, can someone suppose if I ask you a related question? Suppose what if I instead of instead of a maximization which is what I have here I have a maximization problem here written here I had a minimization problem, how would this change? So, I could just replace f 0 by a minus f 0 and that would correspond to minimizing the function f 0.

So, you are maximizing the negative of f 0, f 0 gets replaced by minus f 0 all the other conditions are not put up by this. What I would get is a negative sign sitting out here right; I would get a negative sign sitting out here behind next to the gradient here.

What, but then that negative sign I can observe in my definition of lambda star itself, but this condition simply says that exists some lambda star. So, I can observe that into the lambda stars and this the condition would still look the same just that the lambdas would have reversed the sign ok.

So, but there exist some other lambdas that is all put here ok. So, this condition would not change even if I had a if I had a minimization problem. Now, that gives you a clue about what is going on here; it tells you that this is actually a very weak condition it works for both maximum as well as minimum.

So, the very fact that you have found lambda star and x star such that these equations are satisfied does not mean that you have actually got to a local maximum, you could as well have got to a local minimum right.

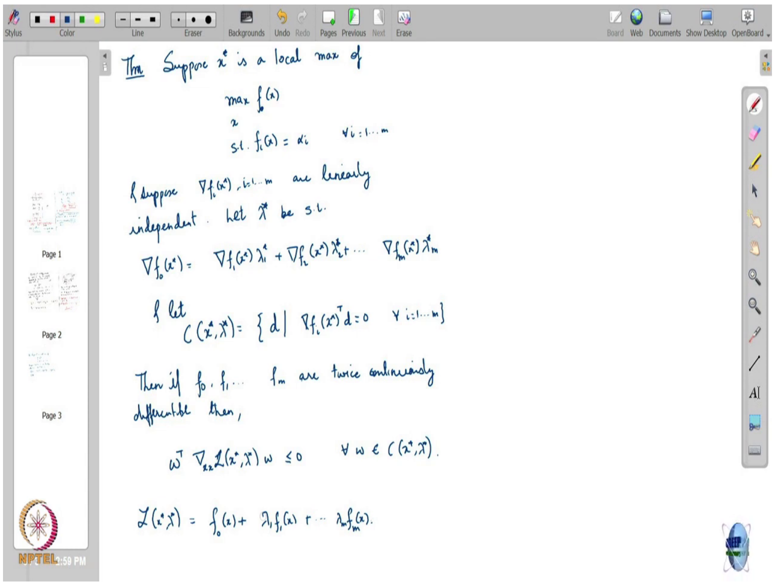
So, geometrically the same compromise that I just mentioned that you cannot go in the direction of increase of the function because the gradient should not have a component in that direction. Likewise if the gradient had a component along that tangent space then the negative gradient would also have a component and you could go in the direction of decrease also right.

So, the same argument can be worked for the minimization also. So, the point so this condition tells you that well they are that there exist such lambda star such that this equations are satisfied is both for local maximum as well as for local minimum ok. So, to decide further if you are at a local minimum or maximum you need additional information right.

So, far as you need it is not just enough that you cannot you do not have a component along that the gradient does not have a component along the tangent space. What also matters is whether you have approached the tangent space that point along an increasing curve or a decreasing curve right

So, effect that is what we need to check. So, something analogous to the second derivative condition of scalar optimization as to be invoked. So, the let me state that now.

(Refer Slide Time: 16:48)



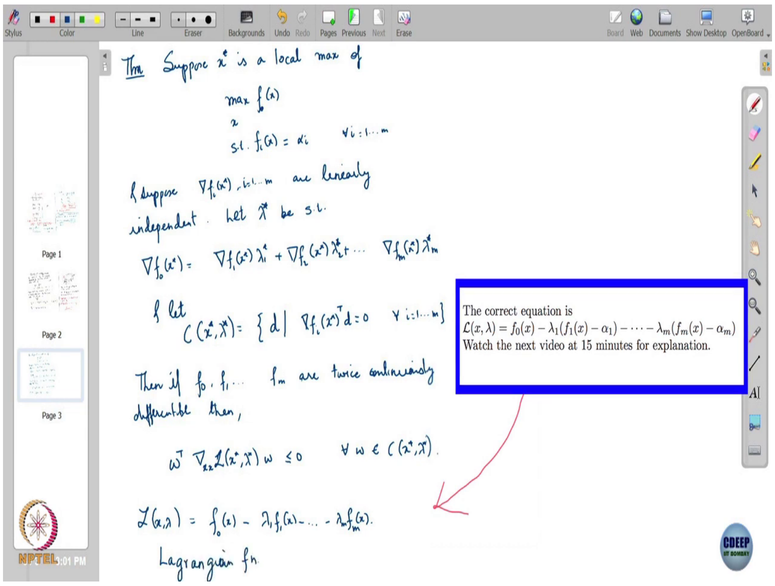
Suppose x star is a local maximum of your optimization problem which is maximizing f 0 of x subject to fi of x equal to alpha i for all i and suppose the gradients of the constraints are linearly independent.

Now, let lambda star be such that those equations are satisfied let lambda star be such that this condition is satisfied. And define c of x star lambda star as this it is all directions d such that this is that tangent space that I was referring its those directions d such that the they are orthogonal to all the all of these gradients right this is the tangent space ok. Then, if it is a local maximum, then if f 0, f 1 dot dot dot fm are twice continuously differentiable.

Then you can write the then the following must be true you can take w transpose. So, I will explain what this notation is this is another piece of notation. Now, here this L here L of this; this is the function defined as. So, this is your second order condition which ah; which tells you that if your if you have found some a condition which satisfies the previous condition found a point that satisfies the previous condition; that means, this boxed it satisfies the constraint and you it satisfies all this ok.

So, then now if it is a local maximum and you know that your functions are twice considered continuously differentiable then this must also hold. Then what is this? This is looking at the Hessian of this function L; L is formed by taking a linear combination of the objective with the constraints ok. It is formed by taking a linear combination of the objective with the constraints, so we just make there is small correction here.

(Refer Slide Time: 22:38)



So, this should be a should be a negative here and I should write this as ok. So, what you need to look at? What you do is look at this function L which is formed by a taking the objective function along with the a linear combination of the constraints. So, you take f 0 of x minus lambda 1 f 1 lambda 2 f 2 dot dot dot minus lambda m f m, put all of that that is called let called this function l it has a name it is called the Lagrangian function.

You take the Lagrangian function look at it only in the x space its a function of both x and lambda to take the Lagrangian function and look at it only in the x space calculate the Hessian of the Lagrangian only with respect to x ok. Assuming lambda fixed you just think of it as a function of x calculate the Hessian of the Lagrangian with respect to x evaluate that Hessian at x star lambda star alright.

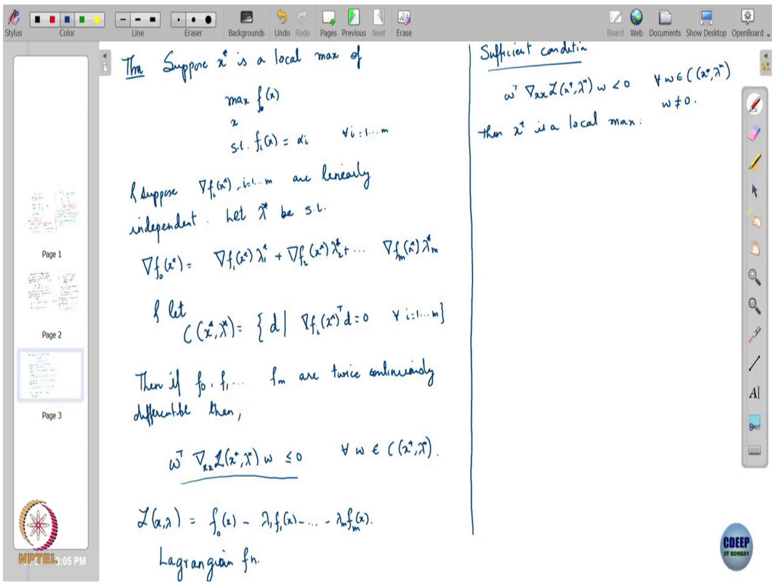
And then you look at this thing, this looks like, so what is this? This is w transpose Hessian of the x Hessian of the Lagrangian times w, that has to be less than equal to 0 for all w that lie in this tangent space. So, which means that the function should have a certain curvature along the tangent space defined by the constraints ok. The second order derivative of the Lagrangian is telling you, how the Lagrangian curves in the x space.

But, then we want it is necessary that this is less than equal to 0 along the directions defined by the tangent space of the constraints. Is this clear? So, this is the thing that is this is one way of further reducing the number of, ah further pruning from the set of points that satisfy these box constraints box equations.

So, you when you get an optimization problem you will you and you try to solve for this you will you can it is you are looking for an x star and lambda star. So, they will you will you will be able you need to satisfy the constraints and you need to satisfy this boxed equation.

But then this will could still get you to a local minimum or a local maximum. Now, if you are looking for a local maximum, then you will further check if your point satisfies this condition ok. But for you to check this you need that your functions are twice continuously differentiable ok.

(Refer Slide Time: 25:36)



Like, so this is a necessary condition the sufficient condition ah; sufficient condition meaning that if you want to be sure that it is a local min local maximum, then what you need to do is the same thing here gets replaced by a strict inequality.

For all w in this set c star x of c of x star lambda star w not equal to 0. So, what this means is that this, your what your referring to is that this Hessian should be negative semi definite along a certain subspace only not necessarily along the full space or we just need that it is negative semi definite along that particular sub space the sufficient this is a sufficient condition. So, if this is less than 0 for all w in this then is a local maximum.