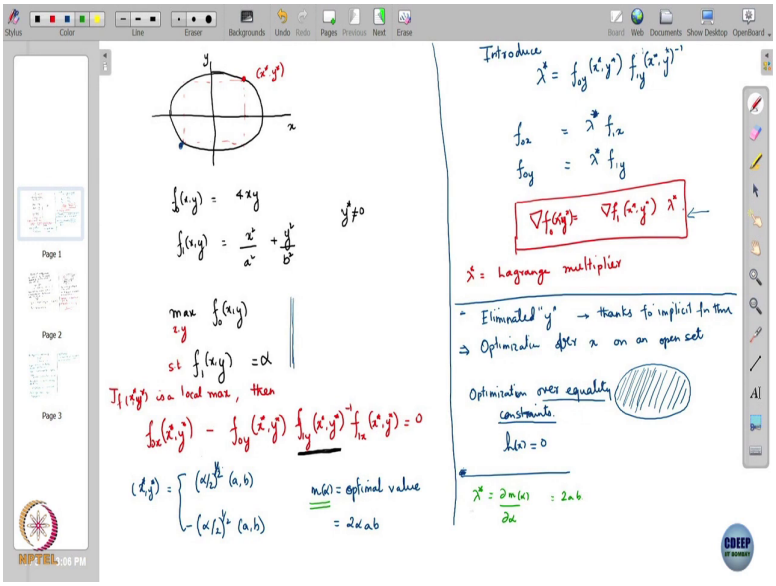
Optimization from Fundamentals
Prof. Ankur Kulkarni
Department of Systems and Control Engineering
Indian Institute of Technology, Bombay

Lecture – 6C
Least norm solution of underdetermined liner system

So, the like we can also talk a little bit further about what these Lagrange multipliers are actually saying and for that, let us go back to our example again.

(Refer Slide Time: 00:27)



And I will tell you what. So, you can actually calculate what the lambda star in this particular case was and you what you will notice is that in this case, actually if you take write it in a different color. So, if you take the lambda star actually satisfies this. It is the if I look at this

optimal value m of alpha and look at the partial derivative of m of alpha with respect to alpha. Then, that is actually lambda star.

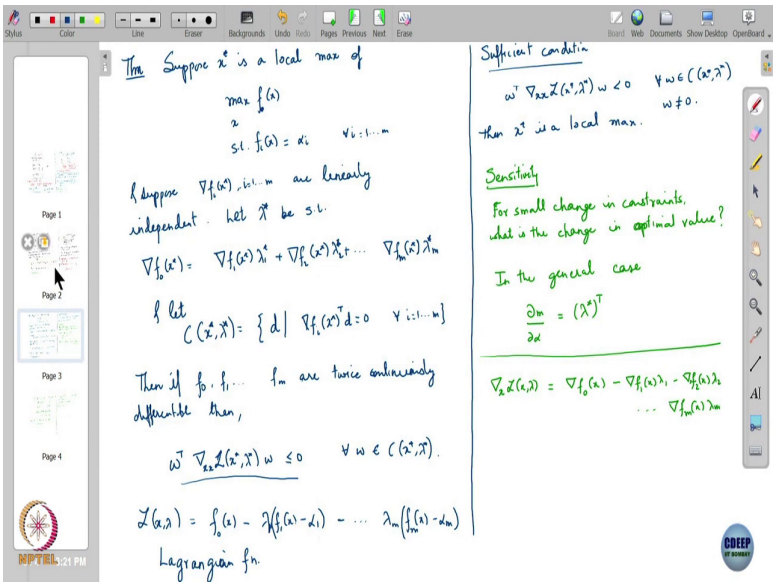
So, that is in short, it is equal to 2 a b. Now, what is the meaning of this? What this means is that the lambda star is telling me how much would the objective function change, if I changed my alpha? How much would be not the objective function, the optimal value change if I changed my alpha. So, m of alpha remember was the psi was the area of the largest rectangle. It is a function of alpha. Alpha was the right hand side here. So, alpha tells you how big is your ellipse right.

If I scale alpha, my ellipse magnifies or becomes smaller. So, if I change my alpha slightly, how much would the size of the largest rectangle, how much would the area of the largest rectangle change by? That is what my lambda star is telling you. My lambda star is actually equal is the derivative of the optimal value with respect to with respect to alpha.

So, the so here is the interpretation and the importance of Lagrange multipliers. Lagrange multipliers are telling us how sensitive is the optimal value to changes in the constraint. So, think of the constraint as a resource ok. Suppose, I tell you that alpha is my is the size of my the controls say the plot of the size of, the plot of land this which is of this elliptical shape, plot of land and alpha controls the size of that. So, alpha is the way by which I am going to measure the size of that plot of land.

I am if I wanted to change my alpha a little bit, means if I wanted to go for a slightly bigger plot of land, how much bigger of a rectangle could I accommodate in that in terms of area? Well, the answer is for a delta alpha change in alpha, it would be lambda star times delta alpha would be the; change would be the change in the area of the optimal of the largest rectangle ok.

(Refer Slide Time: 03:27)



So, Lagrange multipliers tell us something about the inter. So, this is what is called sensitivity. Sensitivity is for small change in constraints, what is the change in optimal value? So, small change is in the right in the constants of the constraint, what is the change in the optimal value?

So, you can do this one constraint at a time also, you look you do not need to look at all constraints together. Look at if you are changing only one constraint by a slight amount, you look at how much is you are basically just making a change in that particular component of alpha ok.

So, what together by together with these, you will be able to see what the; so, I will explain what this is. So, you can get in the general case, if I look at just; so, this is the derivative of the optimal value with respect to alpha, that is always equal to lambda star transpose. Now,

one thing to note here is because you are talking of equality constraint because you are on surfaces right.

A bigger value of alpha does not necessarily mean you have more resource. It just happened because we are talking of this ellipse that larger alpha would mean a bigger ellipse and smaller alpha would mean a smaller ellipse. But in general, as you change your alpha, your surface can change in many in strange ways.

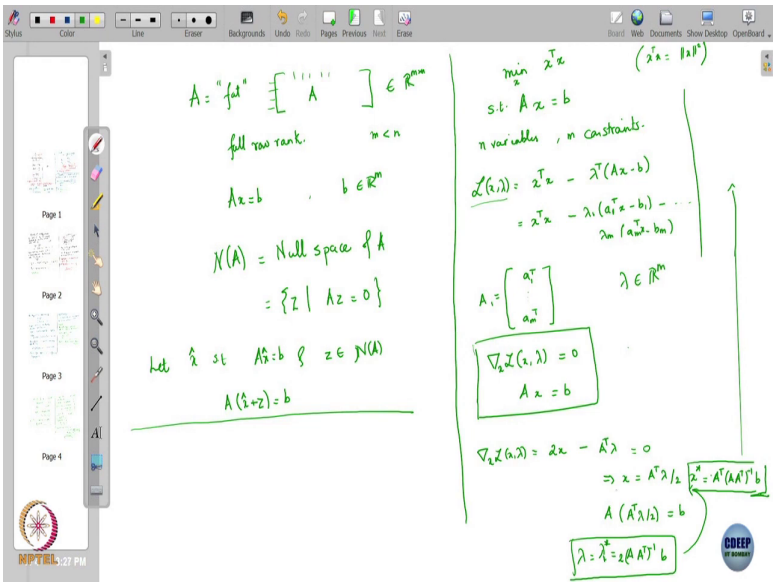
So, it does not necessarily mean you have your optimizing over a bigger region or that the earlier region is enclosed in the previous region or any of that, ok because we are talking of surfaces here. As alpha changes the shape of the surface or the contour on which you are operating will change ok.

So, it is possible that the objective can by increasing alpha, your objective could decrease or it is possible that by decreasing alpha, your objective could increase. All of that is encapsulated in the sign of lambda. Lambda is also then, the sign of lambda also tells you which constraints are sort of more binding than the others; which cons and in which direction should you be changing the constraint; whether you should be decreasing or increasing in order to get a better objective alright ok.

So, you are doing this problem of last time, we did this problem of least squares solutions of equations that were over determined right. So, these were; so, you could not satisfy all of the, all equations at once. We did this in the context of machine learning and also, in the context of maximum likelihood estimation. So, the. So, all equations could not be satisfied, together by a linear function by linear relation.

So, we were looking for the minimizing the sum of the squares of the residues. That was the problem; we were looking at that became a unconstrained optimization problem. So, today, I will look at a slightly different problem.

(Refer Slide Time: 07:25)



So, here suppose you have; suppose you have a matrix A and since this matrix is a “fat” matrix. What I mean by this is you should imagine it to be something like this. It has fewer rows and more columns. So, this is the nature of the of the matrix A ok. So, and let us assume that it is full row rank. Now, if I ask you for a solution of this, Ax equal to b ok where b is some other vector.

So, suppose my A is in R m cross n and b is in R n, R m sorry and I ask you for a solution of A x equal to b and so, I am in this region, where m is less than n; actually m is much less than in general ok. So, then, can you solve for x and how many solutions do we have? Yes, so this is there are fewer; so, number of unknowns here is n which is the number of columns of A and number of equations is m which is the number of rows of A. You have fewer rows than columns or fewer equations than variables.

So, you can easily of course, solve for this. In fact, you will get not one, but infinitely many solutions. Why infinitely many? Because you can this sort of matrix, a fat matrix like this will always have a null space right. So, the null space of A, the null space of A, this is z such that A z equals 0.

So, this is an entire subspace of R n right. So, if I have one solution like this x hat. Suppose, if I take x hat, let x hat be such that Ax hat equals b and I take any z in the null space of A. Then, what can I say? x hat plus z is also a solution of this right.

So, if I have one solution and of course, there is at least one I can always generate an infinitely man y, infinitely many more by just take picking points from the subspace alright. So, then in that in this sort of situation then the common problem that is posed is that you want to find a solution that has a certain structure.

Now, structure mean can mean many different things ok. Structure can mean sparsity, structure can mean close to something else, structure can mean having the least having the least norm ok. So, in this case, let us look at the least norm problem.

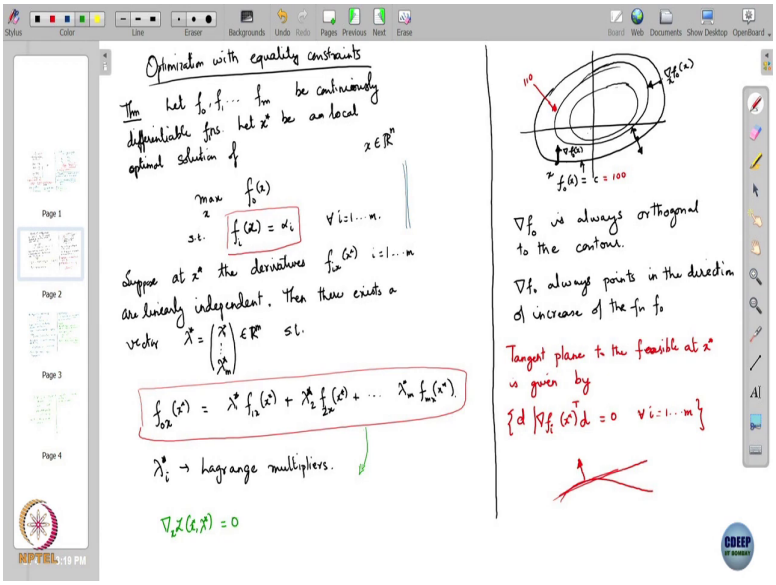
So, the problem there is then is to look at amongst all solutions x of the system of equations A x equal to b, you want to find the one which has the least norm. So, x transpose x or x transpose x, that is the same as norm of x whole square. So, we have this problem ok.

So, now, if A is in R m cross n, how many variables are we optimizing over? You have n variables here. x is in R n right, n variables; n scalar variable. How many constraints do we have? m of these right. All of them are put together, I have written it as a matrix equation.

But it is basically m, m individual scalar constraints right; n variables, m constraints. So, this is now an optimization problem of trying to find the solution of least norm that satisfies a linear system of equations alright. Let us try let us solve this. So, remember this function l that I introduced.

If I look if you look at this function l and I write here is the look at this quantity, the gradient with respect to x of l, what would that be? That would be f 0 of x minus gradient of f 1 of x times lambda 1 minus gradient of f 2 of x times lambda 2 dot dot dot gradient of f m of x and lambda m. Correct? And now, go back to the boxed equation.

(Refer Slide Time: 13:01)



Here, can you write this equation in terms of this function l. This is I mean if I just take the transpose of this; the what this is effectively would amount to is to simply say? It would amount to saying that the gradient of this Lagrangian equation with respect to x evaluated at x star lambda star should be equal to 0 right.

So, this boxed equation is basic all it is saying is that the gradient of the Lagrangian must be equal to 0. So, this is. So, we can this is one succinct way of writing this equation, the red the

red boxed equation. In addition, of course, you have you need to satisfy your the these boxed equations ok.

So, let us use that sort of notation here ok. So, let us write the Lagrangian. So, what would be the Lagrangian? I have my objective x transpose x and minus now I need to write. So, what we I can go back here, if you like. Lagrangian was linear you are taking linear combination of the constraints right. So, the constraints were; so, I ok. So, yeah so these constraints were written as f yeah the.

So, actually, I have I made a slight error here, let me just correct that. So, let me absorb all the alphas also in the definition of the functions. So, if f 1 of x minus alpha equal to 0 is my constraint ok. So, if this; so, its lambda 1 times f 1 of x minus alpha lambda m dot dot dot lambda m into f m of x minus alpha. This does not affect the way I.

After I take the partial derivative with respect to x, all the alphas will not mat will anyway go away. So, it does not affect this condition ok. So, let us write it in this sort of way. So, mine. So, I can write it as for my problem.

So, you have x transpose x minus now let lambda be your Lagrange multiplier vector and I will do lambda transpose Ax minus b. Now, can you verify that this is the same as doing? This is the same as doing actually lambda 1 into; so, where, a can be expressed as a 1 transpose dot dot dot a m transpose.

So, if my rows of a are a 1 transpose, a 2 transpose and a m transpose, those are my; those are my rows of A ok. Then, I can write this Lagrangian in this sort of way right. So, its x transpose x minus lambda transpose A x minus b; where, now lambda is just lambda is a vector in is just any vector in R n.

So, what I have to solve for is that is my previous boxed equation, which is the gradient of the Lagrangian should be equal to 0 and in addition to that, I need to solve I need to make sure I

am feasible which is this other boxed equation ok, which means I need to make sure that Ax equals b. These are my these are the equations I need to solve. Is this clear?

So, if I put the gradient of the Lagrangian with respect to x, let us solve let us calculate that. What would that be? What is the gradient of the Lagrangian with respect to x? It is 2 x minus A transpose lambda. It is 2 x minus A transpose lambda. So, this is my this is the Lagrangian. I am taking this, the it's gradient with respect to x. It gives me 2 x minus. You can check this, this is 2 x minus A transpose lambda. So, I need to put this equal to 0. So, that gives me that x is equal to A transpose lambda divided by 2 ok.

Now, I also need to satisfy Ax equals b. So, I can just substitute for this x out here and that would give me A into A transpose lambda by 2 equal's b. Now, A into A transpose, remember A was a "fat" matrix like this; A transpose would be a thin matrix. A and A, I have assumed is full row rank. So, A into A transpose is invertible right. So, consequently, I can take this on the other side and so, I have A A transpose whole thing inverse b and a 2 outside, that is my lambda.

With my lambda known, I can put it back here to get back my x. So, this is my lambda star; lambda equals lambda star as this, put this back here and that gives me my x star as equal to A transpose A A transpose inverse b; sorry, into yeah this (Refer Time:20:15). So, this is your least norm solution ok. So, the least norm solution of this optimization problem is this one here. So, everyone understood the what we did?

We have your we had our optimization problem; we wrote out the Lagrangian function and we took the derivative gradient with of the Lagrangian with respect to just the x variable put that equal to 0. And then, we had to also satisfy our constraints. These were the two equations, we need to satisfy; use, put that in, we get that we found actually that there is a unique solution. So, it has to be therefore, that this is the solution of your problem. Is this clear? Ok, Alright.

So, I will; so, we can end here, we will continue again next class.