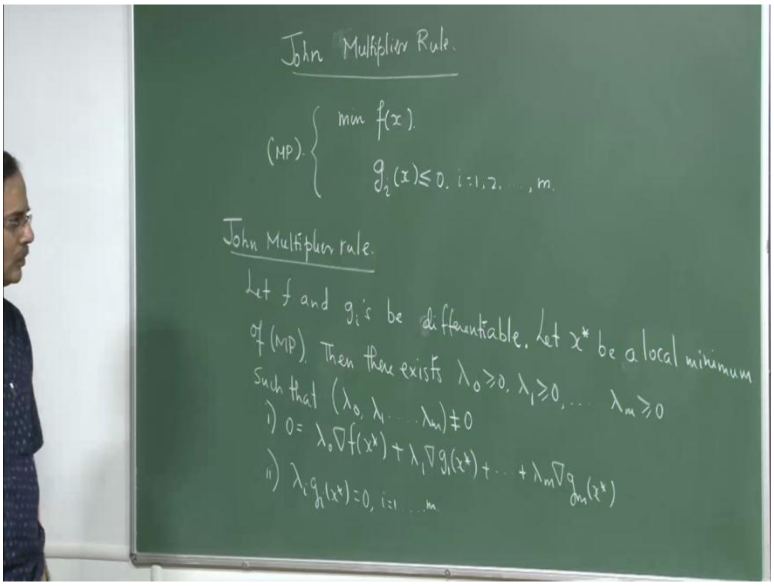
Foundation of Optimization

Prof. Dr. Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture – 16

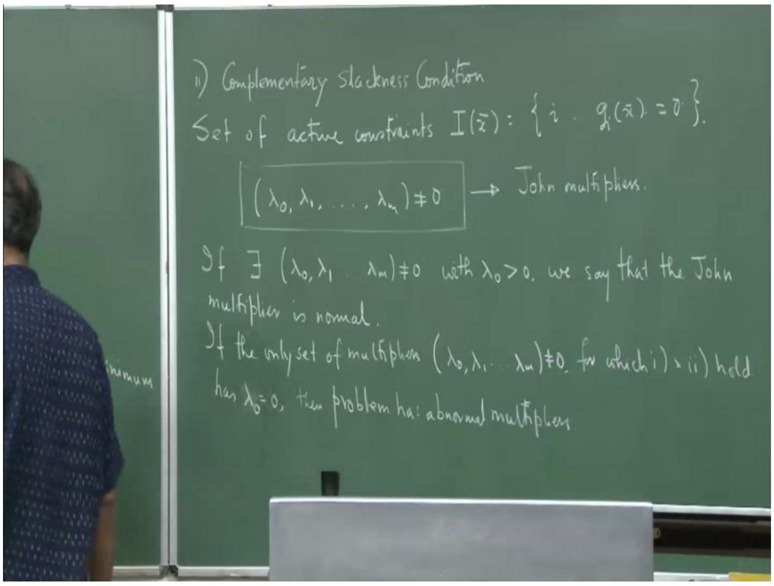
(Refer Slide Time: 00:26)



So, today we are going to speak about the Fritz John necessary conditions or the john multiplier rule and from where we will deduce the Kuhn tucker conditions, but as we will show gradually with examples that in most cases the john multiplier rule gives us what we really require. So the multiplier rule John establish in 1948 was for inequality constant, because as I told you that earlier the lagan multiplier rule dealt with equality constants though its you know validity was not established till late till in the sixties, but here Fritz John study the problem of this form. We will show that in most cases we will get what we want? So I will just write down the Fritz John condition and then try to explain to you the geometry associated with the constants, and what those things mean, so this is the theorem.

Now observe 1 thing, let us assume that all of these are differentiable function, just differentiable functions. If you have more equality constants we have to assume something more that is continuous differentiability, but if you just have this you will have differentiable function, then we will see why it is so? So, let us call this problem as MP at programming problem, let f and g I be differentiable maybe I should write on other side of the board to have space, let x bar be a minimum rather a local minimum that is a much better statement rather than just telling minimum, is a local minimum of MP, then there exists lambda naught greater than equal to 0, lambda 1 greater than equal to 0, lambda m greater than equal to 0; such that lambda naught, lambda 1, lambda m this is not a 0 vector, and lambda naught sorry 0 is equal to, so here you go.

(Refer Slide Time: 04:53)



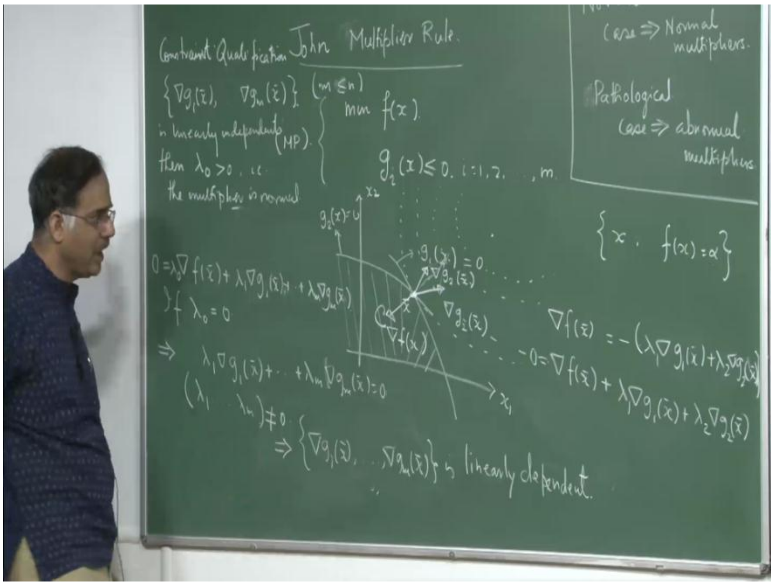
The second condition is called the complementary slackness condition. Condition number 2 is called the complementary slackness condition. In the sense that both these lambda i and g i cannot hold with strict in equality at the same time that is if g i x is strictly less than 0, and lambda strictly greater than 0, then the product would be strictly less than 0 that cannot happen if lambda is strictly greater than 0, then either g i x star is equal to 0. If lambda is strictly greater than 0 then g i x star must be equal to 0, if lambda is equal to 0 then g i x star could be strictly less than 0 could be equal to 0, but if g i x start strictly less than 0.

Then lambda be must be equal to 0, so that gives rise to what is called the set of active constants. Of course, we will start doing the proof of this, but let us go a bit to the geometry. So these are the points where the these are the indices for which the lambda I is need not be 0 may be 0 may not be see if all these lambda are 0, then lambda not cannot be 0. Now a very important fact here is this fact that this multiplier is not equal to 0, if we allow them all of them to be equal to 0, then every feasible point will satisfies the John conditions the fundamental and crucial facts of the John condition is this from modern terminology from the basis of from modern way of talking about things. If there exist lambda naught, lambda 1, lambda m not equal to 0 with lambda not greater than 0, we say that this thing this thing is called the Lagrange multiplier, we in case are calling it john multipliers.

So we say that the john multiplier is normal if no such lambda naught, if no such vector, no such vector of this form naught equal to 0 exist with lambda naught greater than 0, then we say that the the only possible multiplier is an abnormal multiplier; that is if the only set of multipliers lambda naught, lambda 1, lambda m not equal to 0 which for which 1, and 2 hold 1 and 2 hold has lambda naught equal to… Then we say then we say that the problem only has abnormal multipliers, then the problem has abnormal multipliers. What are the good case, and the bad case.

The good case is that that is enough just a 1 such multiplier with lambda naught strictly greater than 0, and the bad case is that you do not have any multipliers for which lambda naught which satisfies these two, and for which lambda naught is strictly greater than 0. And it looks that why are you bothered with lambda naught equal to 0, lambda naught strictly greater than 0 when we have lambda strictly greater than 0, then you can just write lambda naught equal to 1. So I am bothered, because if lambda naught is equal to 0, then grade f x would have no role, but the important part is that in most most examples which we will start learning may be from next lecture you would not have lambda naught 0 except for pathological cases.

(Refer Slide Time: 10:32)



This abnormality when abnormal normal case implies normal multipliers pathological case abnormal multipliers, now let us do the geoma. Let us look at it geometrically what is the normal case, and a good case and what is the bad case? So, here I will just rub this part which you have already seen or you can roll back the film and see it again. So g 1 suppose I have 2 constant g 1 and g 2, and so here is g 1 x, so this is the curve giving giving me g 1 x equal to 0. So, I am in two dimensions, so there is another curve giving me g 2 equal to 0, so this is g 2, and this is my common zone, so this shaded region is my feasible set g 1 x less than equal to 0 g 2 x less than equal to 0.

Now, suppose this is the point where I am interested in looking in to the nature of the behavior of the gradient of this, and this because for all this could be my solution point, I will let me assume that there is a this is, this is my solution point. Now, at this point whether I have the good case or the bad case, that is what we have to really decide, if you look at g 2, this is the surface of g 2, and then you really have to have the normal which has the perpendicular form the surface.

So, at this point x bar this is my grad g 2 x bar and at this point, if I look at g 1, this is my grad g 2 x bar, so you observe that; these 2 are linearly independent. And now assume that here is the un constant minima, and here is the level curves; level curves means the set of all x. If I take the minima f x, so that f x equal to some alpha, so alpha is 0 and then so this is a un constant minima. So, if so from un constant minima I am just trying to…

So, basically its some circulate so functional, and nice convex function basically parabola, it is coming this it is coming like here, and there is the level curve and there at this point the gradient of f on this level curve, where it is touching, and where the optimize achieved on that level curve this would be my grad of f x, because it is perpendicular to the surface of the level curve. Now, you see how can I represent grad f x in terms of this, I can always write that grad of f x bar see, if I take lambda say positive non negative lambda 1, and lambda lambda 1, and lambda 2 and add them let me see lambda 1 grad g 1 x bar plus lambda 2 grad g 2 x bar.

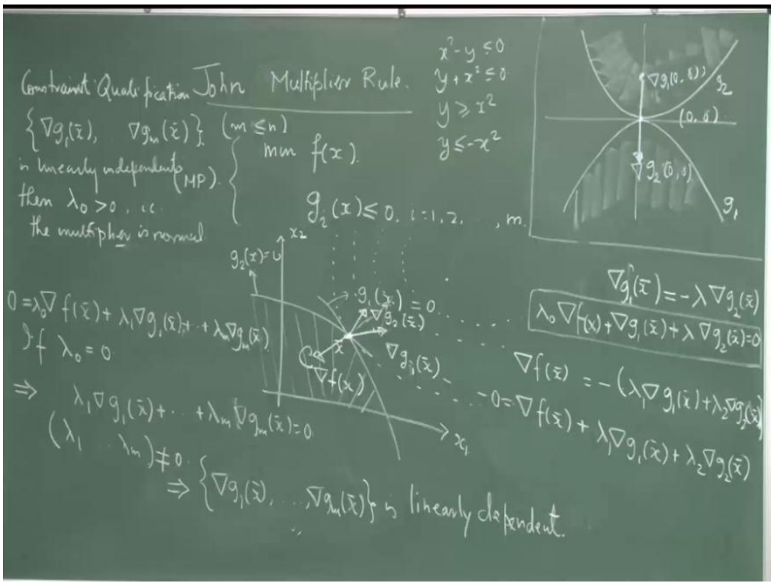
Now, if I because I am taken I have taken a positive multiplication it will remain on this side, and this will be my final linear combination. For example, but then this this this linear combination here is exactly opposite to grad f x, so minus of grade f x can be represented like this, so in so I can write instead like this grad f x is this. So finally, I will get and this is exactly the first equation of the john multiplier rule, so here you see they have a nice behavior here, you see I have got a regular I have got a nor normal multiplier I have I have shown at this lambda naught is greater than 0. So, this thing if you observe comes free naturally from the equation itself, because if you are having this scenario.

Now, you have this scenario now if lambda naught is equal to 0, you would imply that and because I have to because these are the multipliers, which has satisfying the John rule then and because lambda naught is 0, then lambda 1 to lambda m; these vector because lambda naught, and lambda 1 lambda m all are 0 cannot be 0. So among these one of them has to be non zero, these vector itself has to be non zero which is implying that grad g 1 to grad g m, this set of vectors is linearly independent which is implying that these set of vectors grad g 1 at x bar grad g m at x bar is linearly independent. Now, what does it mean? If lambda naught is equal to 0, these vectors are linearly independent, so these vectors are linearly independent, these vectors calculate at x bar the grad g 1, g 2, g m then lambda naught is not equal to 0, p implies q, negation of q implies negation of p, very simple fact of logic.

So, if this equal to 0 this are linearly independent, but these are linearly independent this cannot be 0, so which means a natural assumption that I can or natural assumption that one can impose on the constraints is that all these things cannot be all this these vectors are all linearly independent. Of course, when you take linearly independent, you have to take m less than equal to n, because the maximum number linearly independent vectors in r n is… Of course, just n so from the condition itself that what would an abnormal multiplier give me, that leads to way by which we can guarantee that the multiplier is normal.

So, once so we so this is, so this is as an assumption on the constant which is usually refer at as a constant qualification constraint qualification that if you say that this is linear independent of course this assumption would immediately imply that m is less than equal to n is linearly independent, whenever you may impose this has this is linearly independent, then lambda naught is strictly greater than 0, that is a multiplier is normal. This fact is imply shown here in the picture, because here you see in a two dimensional set up if the angle between two vectors either 0 or 1 80 degree, then the vectors are linearly independent here, they are not angle is a acute angle, so thus these two are linearly independent vectors guarantee in that what we have is a normal multiplier.

(Refer Slide Time: 21:42)



Now, let us see where what is the situation where we can have only abnormal multipliers, there is no way we can get a normal multiplier. So here I am just rubbing this part off may be here. Now let us look at this situation, we have 1 thing say y is greater than equal to x square; another thing say that y is less than equal to minus x square some sort thing like this at 2 right, and so this will this will give rise to x square minus y less than 0, and this will give rise to y plus x square less then equal to 0.

And only possible solution to these are this is just a point 0 0, this is a typical situation where a liner independency if it fails when you just have 1 element in the not, always one of the cases when liner independence have a high chance mini immediate independence have high chance of failing, because if this is my g 1, and this is my g 2 then this is what I have as, and this is what I have as, so these are the pathological cases where liner independence failing. And can lead to non regular multipliers, we will give a more detail example, after we prove the multiplier rule, because needs a lot of things to be done before multiplier rule can be proved even just for inequality constraint.

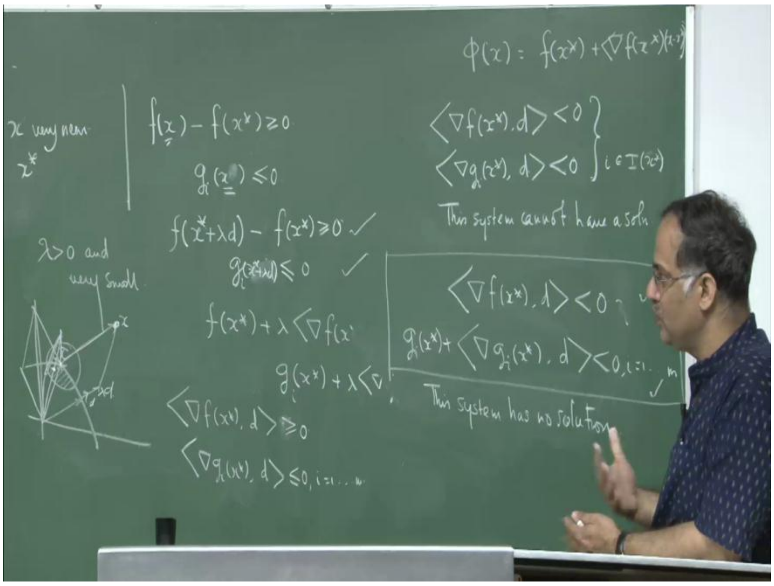
Then we will give a contrite example where showing that example is due to Karush Kuhn tucker in 1951 paper showing that only possible solution of the set of Lagan multipliers, only possible Lagan multipliers that could satisfy is the one with lambda naught equal to 0. So, that would be very important here you can see a grad g 1 and gad g 2 are linearly independent, and see what would happen in this situation, because you can always write lambda 1 has minus lambda g 2, sorry lambda g 1 x bar is equal to minus lambda or lambda g 1 x bar is equal to minus lambda g 2 x bar, I say lambda is greater than equal to 0.

So, you have g 1 x bar plus lambda grad g 2 x bar is equal to 0, so in that case it will really dependent on the objective function, you can always of course, you will you can always prove that there can be a multiplier with lambda if not equal to 0 particularly put lambda naught equal to 0. Then equate inequalities still holding does not matter, but it does not mean that lambda not only would be equal to 0, there, there is not see there cannot be a set of multipliers on g 1, g 2 where lambda naught actually is still non zero.

So, that is the thing that we are trying to figure out, we are trying to say that these are the cases where the high chance of generating multipliers, which are abnormal see what constant qualification does is that it tells you that the multiplier can never be abnormal or so on seem, but in the pathological situation the thing is much more interesting, I would say in some sense. Then in the pathological situation it is very difficult to say that the multipliers would be normal or abnormal, of course in some cases, one can show that only possible multipliers are the abnormal multipliers, they are very bad problems, actually and in some cases even if all this regularity as we are imposing on the constants, do not hold still the multiplier can be regular.

And that is a crux that this whole story of regularity on that is impose on the constant, which leads to what is called the Kuhn tucker conditions, on the constant qualifications one need not always worry about that, because in many, many situations even in the pathological cases which is coming out in latest last in current research also that in pathological cases, you have lambda naught strictly greater than 0. And in fact in many situation the problem the condition of the problem itself guides, you to show that the lambda naught would be anyway greater than equal to 0. So, if you can just find the set of regular multipliers, then it is alright. And the bad cases are where you cannot find it as a as a regular multiplier. Now, you cannot find and you cannot find it not regular find a normal multiplier. So, let us try to prove what we have mentioned, so what we are going to prove from the how how are you going to prove the John multiplier rule, what is meaning of a solution even locally.

(Refer Slide Time: 28:34)



So for x very near x bar for which this is true and this is true, this is your local minimum for x near x bar, you can understand x very near x means, we are talking about neighborhood and all those points now from here what can I say, so x is very near x bar. Then I can replace f x with its linear approximation that is instead of f x, I will look I will write down its linear approximation around x star, that linear approximation of a function around a given point x star is, sorry x minus x star. So, if instead of x point which is nearby I can write as lambda d. For some given direction d x plus lambda d x star lambda d, so x can be written like this for lambda very, very small lambda greater than 0, very

very small I think I should write sufficiently small, but if I am writing loosely, sorry in g x; that is equal to 0 this is the meaning of local optimality.

Now, I want to replace f x here by because x is very near x star, we can replace the function by it is linear approximation earlier, that is the first order approximation of f around x star, basically I write the tailor expansion ignore the remain dot term that is the model. So you are replacing a original function which could be non-linear by linear function; for example, sin x is replaced by x for example, those who are forgotten this basic fact, we will look at this six x graph. So, very near x when x is very near x is difference with the line y equal to x is very less, but as you go out if you go more towards this side away from 0, the difference start increasing, but as you come very near 0; that is why in you will you have leant in say in physics they always use sin x is almost x, when x is very, very small so the same idea is been used here.

So, what we get here is now g x star plus lambda d, sorry I made a mistake g x should be less than 0 x should be feasible, please just correct this mistake, it should be like g x bar plus lambda d should be less than equal to 0. So I am just finding 1 of the x's, so 1 x could be like this a a point which is very near x, and it is feasible feasible means this, and is also satisfying this. So lambda you take any x d if you are not still comfortable with the geometry suppose this is your x, so what I do is a local minima means that if I take a ball around this. And then to get the intersection of the ball with the set and see and add this function point is the minimum for all this points here. Now what I do if I take this this as point as x bar, and I take any any direction d, then I take x bar plus lambda d which should be like this x bar plus lambda d. This is some lambda d this is x bar lambda d something may be may be this is d, and this lambda d x bar plus lambda d on this x bar lambda d, outside the this particular set right.

So here you keep on shrinking keep on shrinking keep on shrinking you can come here, but you can never come come very near, you can come very near to x bar, but you are still inside here. So, that is sort of d I do not want I want the d; for example, if I take the d like this that is sort of a d also would not give you anything I need a d where I make this sum, so if I take a lambda, if if I take a proper lambda or proper link I can actually come move along this, and come to this point so that is why this condition has been written.

So, now I am replacing it by its first order approximation, which is f x star plus lambda minus f x star, and here I also write the first order, I am just taking 1 g. So let me just g i g i, fine. So, for each g i write g i x star plus lambda say it cancels, so basically I will have there is a lot of geometry involved here, but what I am trying to show here is the following that if optimality occurs. Then this sort of a must be a d which has to solve this, so which means if you look at this if there is d for which this is strictly less than 0, and this is less than equal to 0, which means there is feasible point from where I can actually make a better move, I can decrease the value of f. So, I have not reach the optimal. So, if you have reach the optimal, so I will get something like this, so which means that this is this strictly less than 0 cannot occur, and as we will observe.

We will show that this 2 equations cannot occur this system cannot have a solution, so if x star is the local minima of the origin Mac programming problem, and this is this 2, this system where, i is sorry, i is belonging to i x bar. This system cannot have a solution, if you are uncomfortable with I x bar, we can still make it much more easy looking. So I can say that this system actually this is much more, and this system and this is for i equal to 1 to m this system has no solution, because if I take i equal to i x bar that is g i that is is 0 i, ok.

So because I am giving x start, because x star was the symbol I have given in the main result, so here if I take i x star is equal to 0, whether i x start strictly less than 0. Then it is quite simple to do that, because if i g x star strictly less than 0, and this is strictly less than 0 and both of them are strictly less than 0, if there such I deal which satisfies this but, if g i x star equal to 0. Then I can just put here as put g x star equal to 0 and get get just this point, so this system does not have a solution, this is what we need to show and here we gradually get into the more deeper depths of convex convexity.

And we have to talk about separation theorems, and all those staff, so let me tell you one thing, tomorrow we will start by proving this after we prove this, you might ask me how do this is not a difficult thing, how can show the system in consistence it not, so easy to show the system is in consistence possibly, it is easier to solve a system rather than showing a system in inconsistent. So we will first show that if this system is inconsistent what is consistent is there anything consistent, where which you can actually compute to do that, we will need to talk about separation theorem.

So first we will talk about this and in order to say that if this is inconsistent, what is consistent? What is the corresponding system, which is consistent the inconsistency of this would should be equivalent to consistency of something else. So we will show that in order to do that we should learn about convex. Of course, you know about convex set of it about separation theorem, and the golden and the theorem for golden then we will reach to the final lead to the john multiplier rule. So, here we are making a step by step study, and we will start giving a lot of examples, and we will show that actually the most cases the multiplier will gives the normal normal things, and we will in fact do lot lot of examples in this course.

Thank you very much