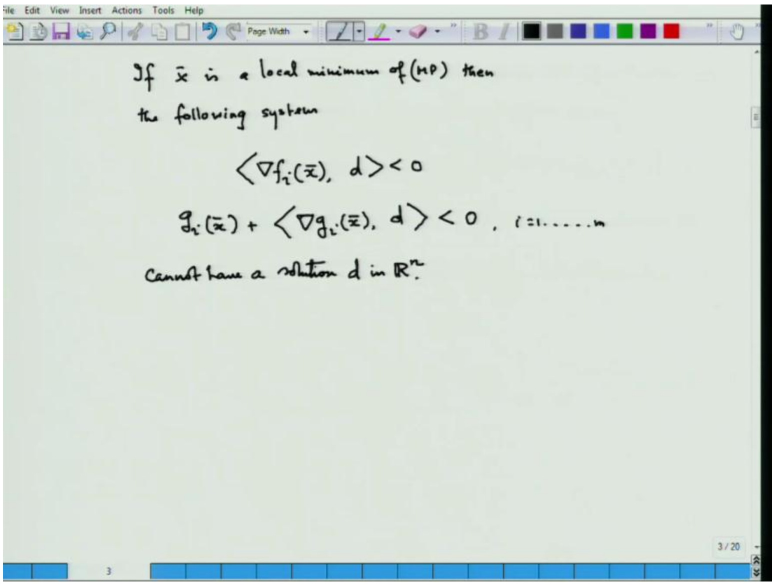
Foundation of Optimization
Prof. Dr. Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

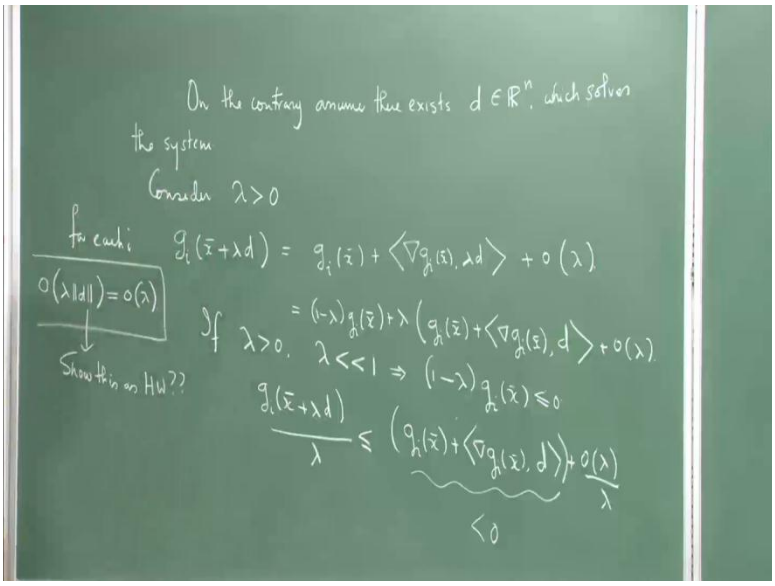
Lecture – 17

(Refer Slide Time: 00:24)



So, what we have told in the last lecture was the following that is what we want to prove today that if x bar is a local minimum of MP the math programming problem in inequality constraints only. Then the following system this system, so this m plus one inequalities this system cannot have a solution it, cannot have a solution d in R n. Now how do we prove this fact, so I will just go to the board proving this fact.

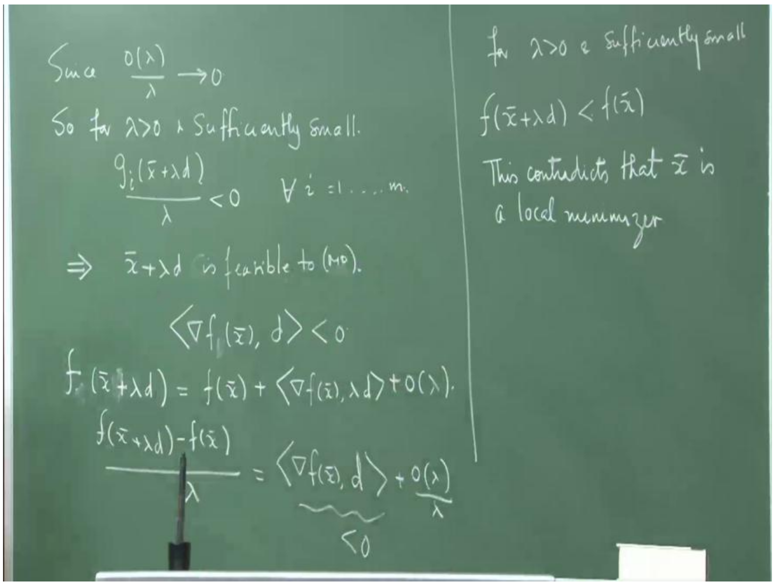
(Refer Slide Time: 02:03)



So, on the contrary, this is proved by contradiction a standard technique in mathematics. So, on the contrary, assume that d there exist d element of R n, which solves the system system means the system one this particular system which you which I have written down, so it solve this system. And now so what we have is a following. Now consider lambda strictly greater than 0, and take g i x bar plus lambda d for any d and then write down the Taylor's expansion, here expanding in a Taylor series around x bar, and then the error term actually it should be o of lambda norm d. And that can be replaced by o of lambda - that is o of lambda norm d is same is an o lambda quantity. So, just little bit of analysis you need to do, show that they are a small o o lambda norm d quantity is same as small o lambda quantity. So, show this fact as homework. So, as a little practices of basic analytical using analytical tools.

Now what I do is here, I break up as 1 minus lambda g i x bar plus lambda times g i x bar plus grad g i. Now, since x bar is a solution to the problem and g i x bar is less than equal to 0 for all i; this for each i, for each i, this is true. So, g i x bar is less than equal to 0. So, one minus lambda, if lambda is greater than 0 and lambda is sufficiently less than 1, when something is very less than one you make this sign. So, when lambda is very much less than 1, there is very near 0. Then it implies that 1 minus lambda g i x bar is less than equal to 0, because this becomes non negative - positive basically. So, what I have from here is the following; I have from this equation I am writing this. So, actually this whole thing now because this is less than equal to 0, this thing is now less than this part. So, I can now dived by lambda to write. Now because d is a solution to the system, this part is sticky less than 0; this whole part is sticky less than 0.

(Refer Slide Time: 06:44)



Now, I can chose lambda as small as I like. So, this part is now because o lambda by lambda, since o lambda by lambda is going to 0 which means that I can as I choose lambda smaller and smaller, this quantity is vanishing is coming to that 0. Of course, this quantity could be negative positive whatever in whichever way it is approaching could be on from the negative side could be from positive side it is approaching of 0; it is in the neighborhood of 0. So, what happens even this is positive say it becomes so small as I make lambda very small that this is a fix negative quantity that the overall sum becomes negative. For example, this is minus 1 and this finally, I made so small that it becomes say one-third or one-fourth something like that. Then minus one plus one fourth would became minus c 4. So, for lambda greater than 0 and sufficiently small, you have g i x bar plus lambda d naught how you will because for every i lambda choose would be different. So, choose, so there will be a threshold of lambda, there will be a lambda beyond whose after for any lambda below that value for a particular i this would be true.

So, for each i, there would be such a threshold lambda chose the one which is minimum and do it. This is a just as this is a standard sort of argument that you make in real analysis and so I expect the students to have some idea about real analysis; in the sense that they should be able to make these all arguments. Unfortunately I would again like to stress, because this is a sort of a public lecture and so I like to stress on one fact that this is a course on mathematical optimization and this is not just a course where you are talking about some modeling some problems and the software. So that course is better possibly delivered by someone who is in the forth front of applications specially from the departments of management and or for may be from the engineering departments.

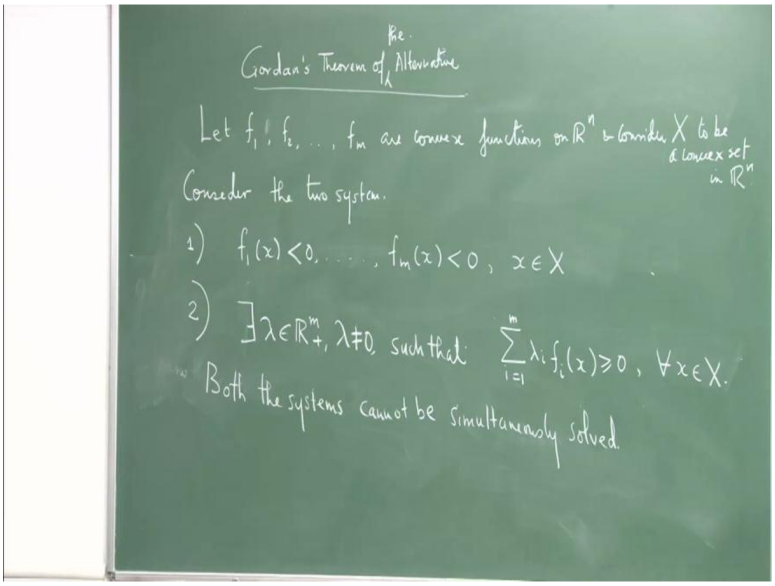
So, I would like to apologize and I would like to ask for the viewers pardon that I am using mathematics which possibly is not of your liking, but I cannot help optimization is a mathematical subject. I am not doing any involved more involved stuff like calculus of variations or any other thing like infinite dimensional optimization. I am just remaining in optimization with derivatives, but mathematics is a involved stuff and optimization is a mathematical involved stuff and you really have to go through this things. Those who are from the engineering background I am sorry that this is essentially a course in mathematical optimization, it is the foundations of optimization and optima the foundations of optimization is is based on heavy mathematics. So, it is a mathematical subject. So, but there are lot of insides which will start giving once we establish the so called Fritz John multiplier condition or Fritz John necessary condition. So, this is true for all i equal to 1 to m. This implies that x bar plus lambda d means g I x bar lambda d strictly less than 0 is feasible.

Now, once I know this then let me go to the first part of the thing; first part is so d because this is a solution also solves this. So, which means now I can write, because d is a element which is solving this system. Now again we will take the help of the Taylor’s expansion around x bar, sorry there is no f i is only f I have written f I sorry there is no f I made a mistake it is only f there is no, because this is only one object. So, now, f of x bar plus lambda d is equal to f of x bar plus grad f x bar lambda d plus o lambda - small o lambda. So, this means now that f of x bar plus lambda d minus f of x bar by lambda is equal to grad of f x bar d plus o lambda by lambda. Now this is strictly less than 0, again by a similar sort of argument as before, so we can conclude for lambda greater than 0 and sufficiently small f of x bar plus lambda d is strictly less than f x bar.

Now, for lambda sufficiently small x bar plus lambda d is feasible to the original problem. So, this would imply that this contrary x z x bar is a local minimum, because if I make lambda very very small I can bring x bar plus lambda d into the neighborhood because x bar plus lambda d is already I know it is feasible. And I can make lambda, so small that x bar plus lambda d is in the neighborhood of x bar where that local minima holds this contradicts that this x bar is a local minimum is a local minimizer. So, this contradiction leads to the fact that this system, please note those who have looked at f i might have question, but this is a way I have placed question. So, just have a look at this.

So, this system would not have a solution and that is what we proved. Now if this system does not have a solution what can we conclude. In order to conclude something, so if this system does not have a solution, something has a solution. So, if we we should get something which would have a solution, we can actually try to solve it and get our suspected point of suspected local minimzer. So, in order to do that we will now state something called the Gordon's theorem of alternative or theorem of alternative where basically there will be two systems of inequalities. And among those two systems, only one would have a solution at a given time; in the sense that if one of them has a solution the other would have would not have a solution vice versa. If the other has a solution - the if system one has a solution, system two has no solution; if system two has a solution, system one has no solution.

(Refer Slide Time: 15:18)



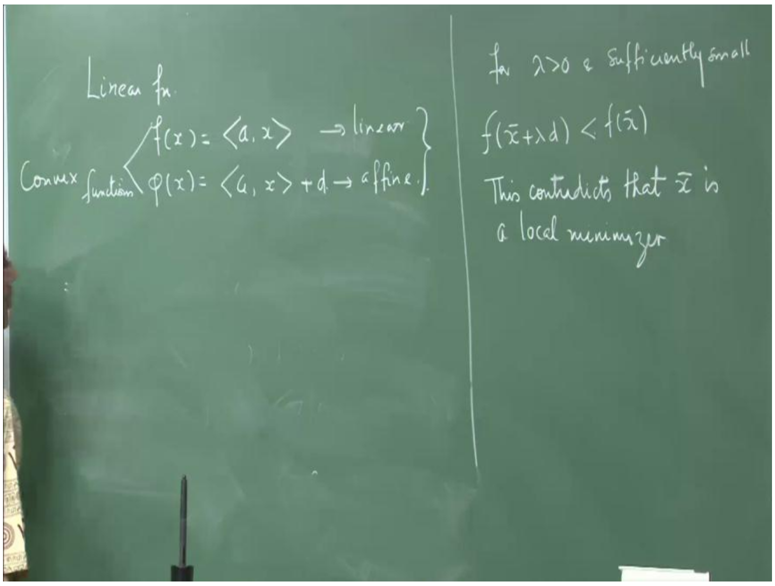
So, just let me write down. Now let us look at what are the two systems. So, let f 1, f 2, f m are convex functions on r m convex function of r m finite value of course. Consider the system we have studied a bit about convex sets and convex function very earlier. So, we will just try to recall that we will try to see what is there how much we have done and then we will try to recall that the notions of basic notions of convex sets and convex functions and so that would help you in understanding this thing. Consider the two systems and consider x to be a convex set in r n. So, this is one system where I am expecting that if this system is solvable means there is an x in this convex set x for which all of these gives me a values sticky less than 0. Or else if such a thing is not this is one system and the second system is basically if this system does not have a solution, this will have a solution. And there would exist this is a system of the there would exist lambda element of r m plus and lambda not a 0 vector such that this quantity is greater than equal to 0 for all x in capital x this is the Gordon's theorem of the alternative. It is called theorem of the alternative, because final statement is that both the systems cannot be simultaneously solved, so that is the meaning of what theorem of the alternative.

Theorem of alternative means I think people better use a theorem of the alternative. Theorem of the alternative actually means that there are two systems here and at the same time you cannot solve the system that is if you find an x for which the system one is solved, you cannot find a lambda for which this will be true. That is any lambda that you for which you say that this is true that is such a lambda must be equal to 0. This has a lot of interesting connotations in optimization and we will use this fact to get what will be called as the Lagrange multiplier rule, because let us observe one fact. In our systems, if what we proved here we have a chain of inequalities, which does not have a solution. It is like our system one which does not have a solution. This is a function in d and if the function of d this is this is a liner function and these are affine function. So, they are convex function. So, here I will just take few minutes to actually make a little bit of recall of your ideas of convex sets and functions because this is quite heavy stuff compare to what little stuff we had learnt in the beginning.

So, then we can see how to first use this to get the John multiplier rule or the John Fritz necessary condition and then try to give a brief sketch of how to prove this fact. So, here we go back once again to recall what you have here the definition of a convex set and a convex function. So, that is fine, because you see convex set is a set who with all its points contains all the line segments joining any pair of points. So, whenever there are two points x and y in c the line segment joining them should also lie in c. And this is the definition of convex function which for a nice looking convex function tells you that if you take two points on the graph and join it if your function is somewhat 2 R n you will have the line above the graph that chord. You see we have studied a bit more about convex series. List of examples about which have convex functions here, we have famous quadratic function, then what would happen, if it is differentiable and the fact that for a convex function every critical point is a global minimum and then we came to line such method that is enough for us to know.

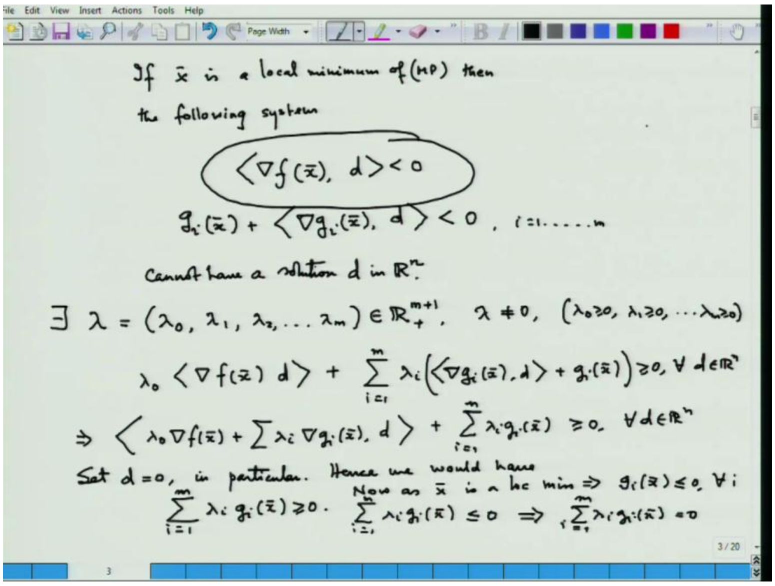
So, but the let us first try to apply what is there into our thing before knowing how to prove it, because to prove this you would require a thing called separation theorem. So, that proof or how the poof would go about for that particular result, I will just give you an outline. Keeping again in view that of the broad nature of the audience, but still it will be quite mathematically involved, but mathematically involved does not mean that if the word mathematically involved should not deter you because it is very important to know that if you love mathematics will love you back and help in your work. So, those who are doing other sciences, and they have to use optimization, I mean most engineering sciences they have to use optimization. And if you say that I want to shy away from mathematics it is very strange, because if you go and look make a search on the internet a good number of mathematical optimizers are in engineering departments, and many of them are had training in original training was in engineering and then of course, they did different things.

(Refer Slide Time: 23:28)



So, let us try to apply what is there. So, we know what is a convex function what is a convex set and all these things. So, what is very important what important example of a convex set is a liner function a liner function is a function f x which is given in terms of the inner product. These are simple liner algebra fact, which you can prove if you know what is the basis of a vector space. So, then this is a liner function if you add something to it, for example, if you say some other real number to it then this becomes an affine function. So, this is linear and this is affine linear plus some stuff. So, you can show that these two are convex functions. So, in that sense, go back to this one, this is a liner function and this is a affine function. So, basically what I am telling is that these two things does not have a solution. So, now, I want to apply the result there on the board to this thing then what I would get, so I would get a vector. So, if this system does not have a solution in d here d is in R n, so my x is in R n and x capital x here is R n.

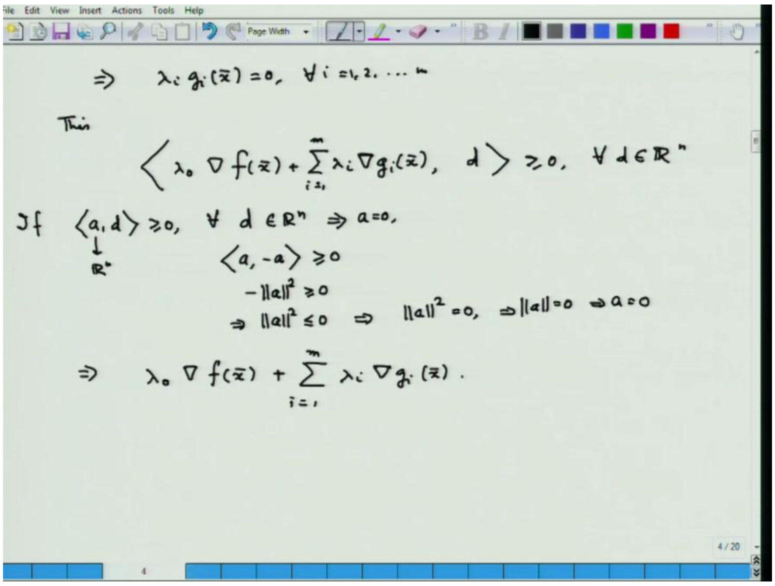
(Refer Slide Time: 24:50)



So, there exists a lambda vector which should contain we will put lambda naught for the thing corresponding to this lambda naught, lambda 1, lambda 2 dot dot dot lambda m. This is in r m plus 1 and lambda is not a 0 vector, you will observe that this is a very important conclusion that this vector is not equal to 0. Such that lambda naught into grad f x bar d plus summation i is equal to 1 to m lambda i grad g i x bar d plus g i x bar. This whole thing that has to be greater than equal to 0 for all d in R n. Now all this is true you simply have the following fact. So, you combine a lot of things and which you can do this little home work yourself, which is a simple manipulation; once you know this, set d is equal to 0 in particular, hence you would have so d is 0. So, this is left, so this is greater than equal to 0.

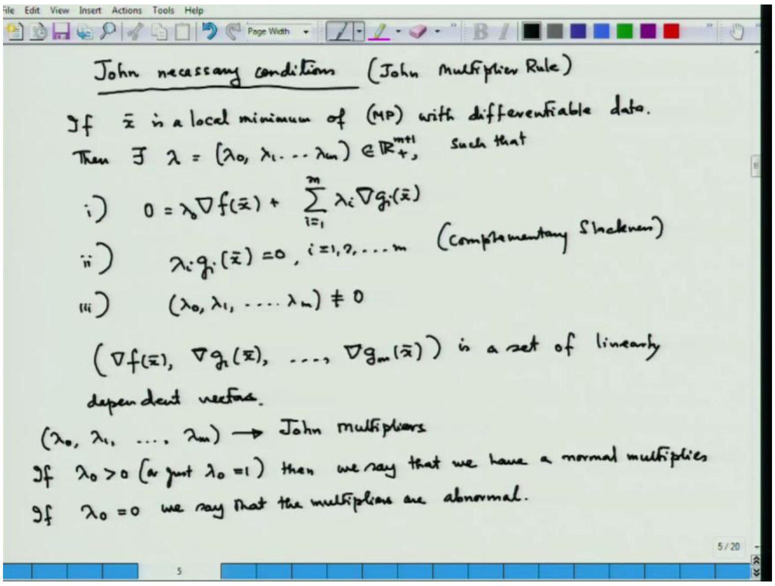
Now, as x bar is a solution is a is a local minimum, it would imply that for all i g i x bar must be less than equal to 0, for all i from 1 to m because it x bar has to be a feasible set feasible element. And because lambda is all greater than equal to 0 which it is an R m plus. So, here I forgot to put the plus. So, it will be R m plus one plus. So, all of these quantities that is lambda naught is greater than 0, lambda 1 is greater than 0 till lambda m is greater than equal to 0 that is the minimum this is an R m plus R m plus 1 plus. So, what I should have is that summation lambda I should be less than 0, but then this we have greater than 0 from here. So, implies that summation lambda i g i x bar is equal to 0 consists each of them lambda i g i x bar is a negative quantity.

(Refer Slide Time: 28:54)



Finally, I will we will get a following implication that lambda i g i x bar is equal to 0 for all i. So, this implies, now it says basically a liner function is greater than equal to 0 over all over the whole of R n, this fact this is not possible unless the function itself is 0, this is a simple fact. So, if a, d is greater than equal to 0 for all d element of R m of course, a is element in R n also; otherwise you cannot define the inner product. This would imply that a is equal to 0, it is simply done by the fact that if you put instead of d, if you put minus a that minus a is also in R n you will get this. So, what you will get is minus norm a square is greater than equal to 0 which will give you norm a square is less than equal to 0, but norm a square is actually greater than equal to 0. So, this would imply that norm a square is equal to 0 which will imply that norm a is equal to 0 which will imply that a is equal to 0. So, in the same logic that we would do we can show now that so this would imply that lambda naught.

(Refer Slide Time: 31:17)

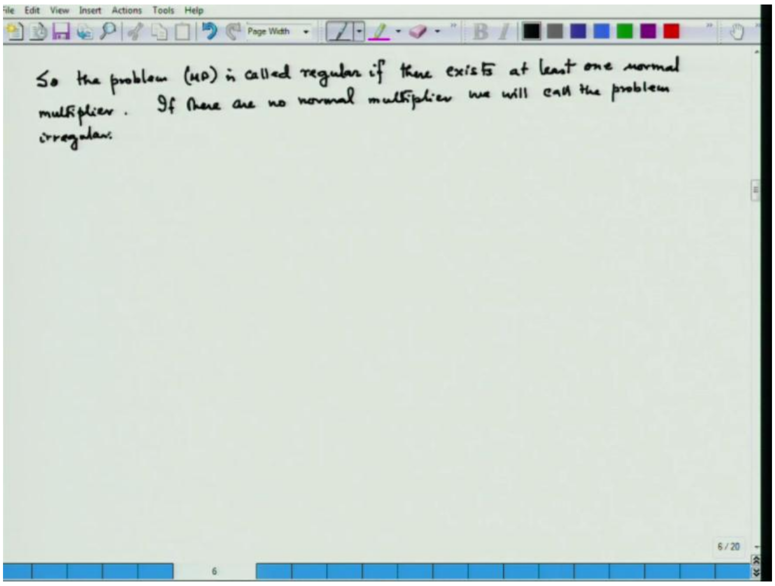


So, this would now allow us to summarize what is called the John necessary conditions or the John multiplier rule. If x bar is a local minimum for m p just this john multiplier rule is usually used by some people just to make it look similar with what we call as a Lagrange multiplier rule which was just for equality constraints to which we will come in the next classes. And then we will combine both the equality and inequality – M P. Now this math we will all we will consider all our functions to be differentiable at the least could be continuously differentiable could be twice continuously differentiable, but at least at the least differentiable you with differentiable data. So, it is a inequality constraint program then there exists lambda such that number one as I have told you in the explained you on the board in the last class that this condition, this condition is called the complementary slackness condition at both g i and lambda i cannot hold with strict inequalities at the same time. If one is strict, the other has to be 0.

Now come to the third and most crucial third and most crucial point the point is this, this cannot be 0. This is a most crucial point this central point of the result. So, what does this say, what does this fact say. This fact and this fact combines it says that if x bar is a local minima then you must be sure of one thing is a set of linearly dependent vectors. Now one might get worried about the fact that what would happen if lambda naught is equal to 0. Of course, if lambda naught is equal to 0 then f is not a part of the problem then such a problem are such a multiplier is called an abnormal multiplier. So, this set of so called multipliers with the objective and constraints is called the john multipliers. See the role of these multipliers is essentially to relate the coordinates, the components of the vector x. So, relate the components of the vector x, so that you can actually find the solution that is the whole idea of the multipliers at the end when it just look at from a look at it from a computational prospective. So, this is called as set of john multipliers and then if lambda not is strictly greater than 0 or one, we just if lambda naught is sticky greater than 0 or just say lambda naught is one.

Then may be if lambda naught is greater than 0, we can divide by lambda naught and basically the same thing the same story remains. Then we say that we have a normal multiplier that is a good case. If lambda naught is equal to 0, we say that the multipliers are abnormal that is important to know at a given x bar, there can be more than one multipliers, which will satisfy the condition. Once you know the x bar then essentially what you have to find what does the thing says that if x bar is a local minima, it has to satisfy this condition means you have to find a lambda which satisfies this and this. Now note this fact that even though I am getting there could be a situation there could be more than one multipliers which will satisfy this. So, I could get a set of multipliers where lambda naught is 0, and could as well as possibly figure out a set of multipliers where lambda naught is 1 or strictly greater than 0.

(Refer Slide Time: 37:40)



So, the problem MP is called regular, if at least one normal multiplier; and if there are no normal multiplier, we call the problem is irregular. Now, we shall show in the next class with examples that there will be a problem whose Fritz multiplier could be, there could be abnormal multiplier to that given solution could be a normal multiplier. There could be a situation where there are no normal multipliers and there could be situation where they are only normal multipliers, but how do you assure that there is a normal multiplier. For that you have impose certain conditions one of conditions which come out very simply from this is that ok, if this is 0 then in order to maintain three the lambda 1 to lambda m this vector cannot be 0. So, which means the gradient of g i x bar is linearly dependent. So, if lambda naught is 0, this gradients of the constraints are linearly dependent.

So, the statement is that if these gradients of the constraints are linearly independent then we will never have lambda naught equal to 0. So, that statement that the gradient of g i x bar are all linearly independent can be imposed as what is called as a constraint qualification to get that I will never get a normal multiplier. Whenever if there exists a normal multiplier to this fritz john inequality system, we say that the KKT condition holds or the Karush Kuhn Tucker condition holds, because we already spoken about the Karush Kuhn Tucker conditions and we have spoken about Kuhn and Tucker’s work of 1951. And here you see and then which is now can be view as corollary to the Fritz John system, or a John conditions. So, thank you very much. We will come with examples in the next class.