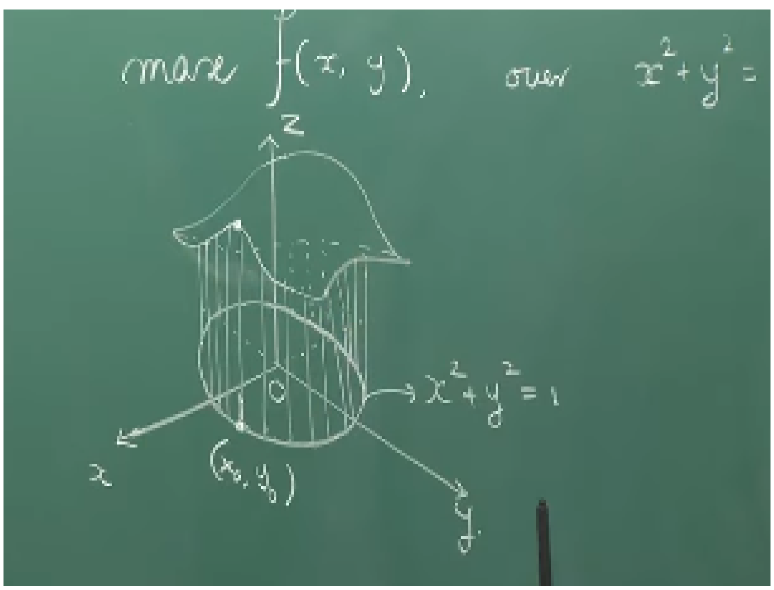
Calculus of Several Real Variables
Prof. Joydeep Dutta
Department of Economic Sciences
Indian Institute of Technology - Kanpur

Lecture – 17
Constrained Optimization and The Lagrange Multiplier Rule

Okay it gives me a great pleasure to talk on this subject optimization and that to constraint optimization I have been studying the subject for more than 20 years as a researcher and gone into quite a depth into various aspects of it. So I would now put forward to you my own private portion of this subject and I have already given you an example of a constraint minimization problem or maximization problem whatever you want to call.

When we spoke about the use of n dimensions more than 3 dimensions 4 dimensions we give an example of the diode problem and we formed what is called a linear programming problem and that we had done that was the important example of a constrained optimization problem. Here we are going to simply study first to start with.

(Refer Slide Time: 01:19)



A simple situation where we have to minimize a function of x,y such that some x,y is not free that it is not unconstrained like what we did in the last two lectures but has to satisfy a condition like this that g x,y has to be always. So x,y must be within in the level curve of this function g.

For example if you want to say maximize you want to maximize a function of x,y/x square + y square=1.

So this is an example given in the book but I will talk about Lagrangian multiplier rule in my own way and you must be asked me why there was one constraint in the linear programming problem that you have said there are a couple of constraints more than 1 but yes here we are just talking about the situation very preliminary situation which first arose in the natural sciences and Lagrangian actually use these technique not while studying optimization as per say but while studying mechanics.

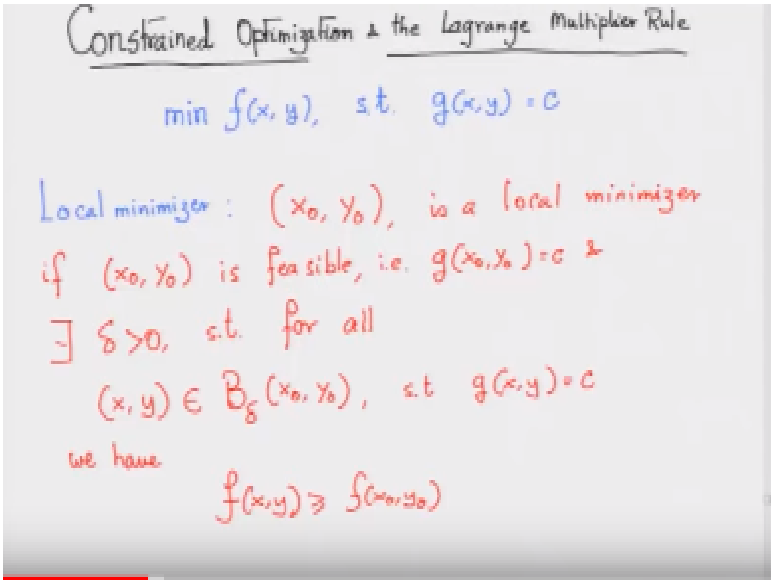
Mechanics is the key to many subjects that have come as a part of mathematics or rather applied mathematics. So suppose we want to do this and suppose a function of the graph so here is your so this is your x square + y square=1. So basically I am looking at this this is my constraint set so I want to bother about the maximum value of f x,y for all x,y among all the x,y which lie here I am not bothered what happens here or outside this circular circumference.

So if the graph is for example like this so what do you do for example graph is like this and inner graph is like this what do you do essentially you draw the cylinders basically you draw from every x,y you touch the function value z you keep on drawing cylinders we usually construct a kind of cylinder basically at every x,y on this circle you are evaluating the values the values of z right.

Also you have things inside and so forth you will see among these at this particular point here which drops down to this place this place this x0 y0 maximizes the function value over this. Now this maximum may not always be the maximum over rn the maximum over rn could be here at this point if you drop it is here but maximum over this space is this particular zone is different further you have to also come to the fact that we can also talk about local minimizers and local maximizers.

So I will define now local minimum you have to yourself sit down and define what is the local maximum but here you have to understand that we have constraints. So we will first define a constraint problem.

(Refer Slide Time: 06:02)



So what we do is we consider to minimize f of x,y where x and y are real variables such that g x,y= c. We consider the minimization of this now we have to talk about a local minimizer x0,y0 is a local minimizer if x0,y0 is feasible that is feasible means g of x0,y0 satisfies the constraints this is called a constraint. So in optimization language this is called the objective f x,y is called the objective and g x,y = c is called a constraint.

So if anything any point which satisfies the constraint are said to be feasible for that problem because those are the points we are interested in and we are not interested in other points. So optimization by the way is a generic name given for maximization and minimization it could mean either but mathematical optimizes usually speak in the language of minimization.

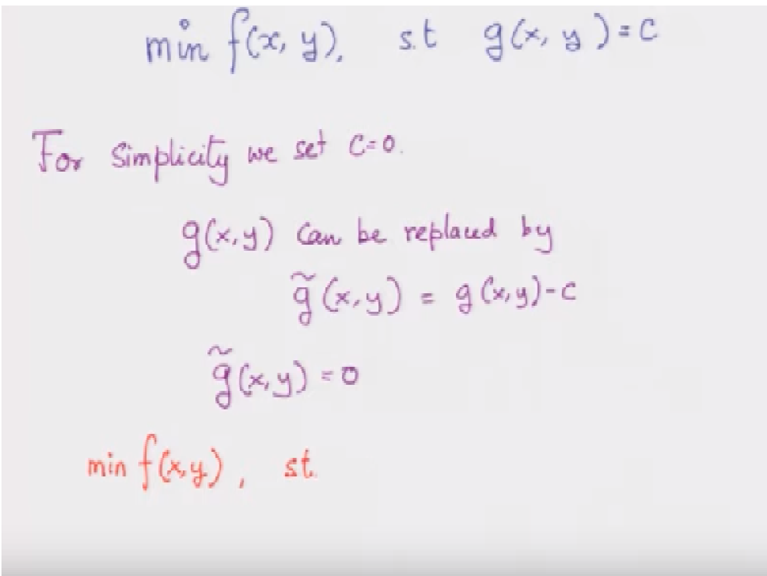
You can always speak in the language of maximization engineer speaking the language of minimization it is a it depends on what context you are speaking because economics you are talking about maximizing profit maximizing utility, engineers are talking about minimizing some kind of you know some kind of load or some kind of current or some issues.

So x0,y0 is the local minimizer if it is feasible and there exists delta greater than 0 such that for all x,y which is in the open ball around x0,y0 and you know what is open ball the norm of x,y - norm of x0,y0 must be strictly less than delta. These are these are such points for all x,y in this such that g x,y = c we have so you have to you understand we cannot just take the ball and take points on the ball.

We have to take only those points in the ball which satisfy the feasibility condition satisfy the constraint and that is the important thing that we have to remember. We have f of x,y = f of x0,y0 so writing the definition of a local maximizer is not a problem because you have to say stories same just change the inequality. So I am not going to talk about a local maximum as a global maximizer to understand where you just say that for all x,y you do not say this is there for all x,y which satisfies this I have this.

So but the story that we are going to talk about is for a local maxi the local minimizer the main theorem is actually written for a local minimizer.

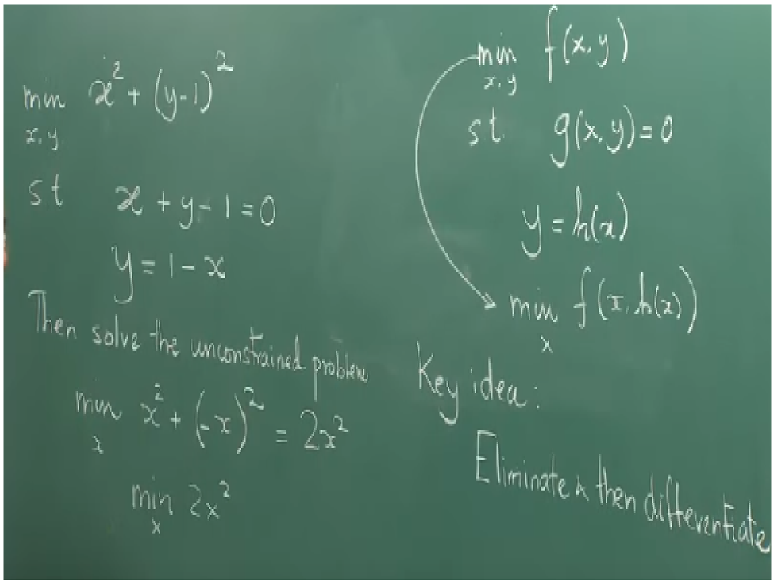
(Refer Slide Time: 09:49)



Now let us look at this problem minimize f x,y such that g x,y = c all simple purposes we said for simplicity in the x position we said c = 0 and in doing so we lose no generality we do not have a loss of generality which means that I can always instead of g x,y g x,y can be replaced by g tilde x,y g tilde x,y = g x,y – c. So our constraint can now be replaced written as g x,y g tilde x,y =0. So without loss of generality we can always take c = 0.

And hence we shall now focus on the problem. Now if I want to solve this problem what should be my first thinking let me just try to understand that let us try to understand that algebraically.

(Refer Slide Time: 11:44)



For example if I have say minimize x square + y -1 whole square such that so you minimize over x,y such that x - y = 1 or take x + y = 1 let so gx,y = c so or x + y - 1 = 0 basically if this is what I have written then I can see that I can from here I can always write I can convert the unconstrained problem in constraint problem in to an unconstraint one by just setting y to be 1 - x and then solve the unconstrained problem then solve the unconstraint problem minimize over x x square + y -1 = - x - x square.

So I am replace it here because I can write it like this I can do this. So what I have so this is 2 x square so you know the answer is 0 say x =0 if x = 0 is the answer y =1 right if you minimize this I have x = 0 if x = 0 then from the constraint I have y = 1 so if you see this is a non-negative function because this is x square y - 1 whole square. So if I put x = 0 and y = 1 this gives me the value 0 which is the least value of this function.

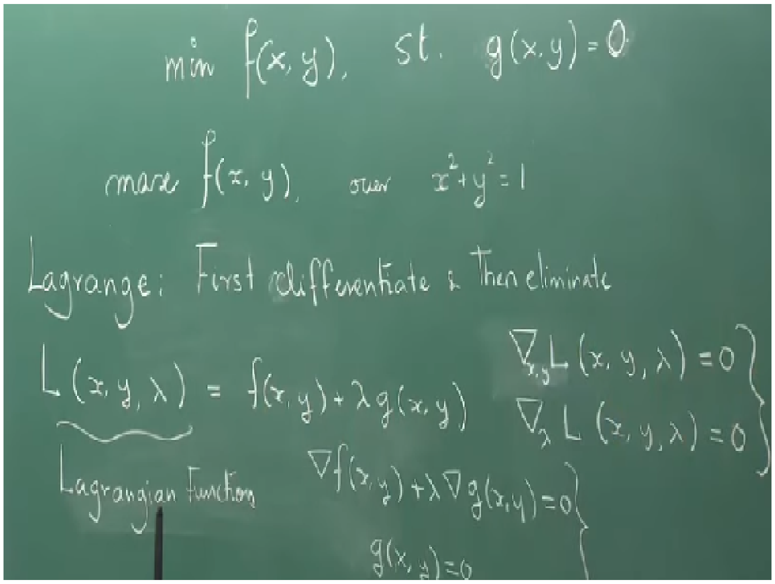
Because this is a non-negative function all the values are greater than = 0 because these are points lying on a circle actually of with center 0 and 1 centre 0,1. So you know so I have changed the problem from this to minimizing over x 2 x square and you know what is the minima and hence you can solve it. Now you also know simultaneous is all this idea what if I write it in a very general framework what is my idea even in a general framework I have the following idea.

So if I minimize fx, y such that subject or such that whatever you want to say gx, y = 0. Suppose y is expressed as h of x just like here y is expressed as 1 - x h of x is 1 - x then I can write this problem I can write this problem as minimize over x f of x, h of x and then go on differentiating it but that really need not happen every time because this function g x,y can be complex enough so that you would not be able to write it like this because for example if you have exists here some x square + y square = 1.

Then you will have y = + - root/1 - x so I can have 2 different values and then you are stuck you cannot put it there straight away. So here so what is the idea the idea is at first you eliminate y and then you differentiate eliminate y make this as a function of x and then differentiate. So the key idea is key idea behind in constraint optimization key idea is eliminate and then differentiate. So that is the key idea eliminate and then differentiate so that is what I am trying to tell you that Lagrange said that sweet swap these two swap these towards.

Lagrange says first differentiate and then eliminate so what does so let me write it down here and I am sure you have taken down the diagram.

(Refer Slide Time: 17:09)



So Lagrange says first different just reverses the whole policy first differentiate and then eliminate. This is the key approach key idea behind the Lagrange multiplier rule first eliminate and then differentiate. Sorry first differentiate and then eliminate what he does is he forms a function called the Lagrangian function where he writes a combined function of the objective and the constraint with a multiplier associated with a constraint function where this lambda is usually called the Lagrange multiplier.

So this function is often called the Lagrangian function it is good to hear some names Lagrange was a great mathematician by the way spend most of his life in French but a Italian mathematician from Turin and Lagrange if I tell you how he used to work you might be surprised people do not work like that. So he used to start to work in the evening and then go working till the next lunch.

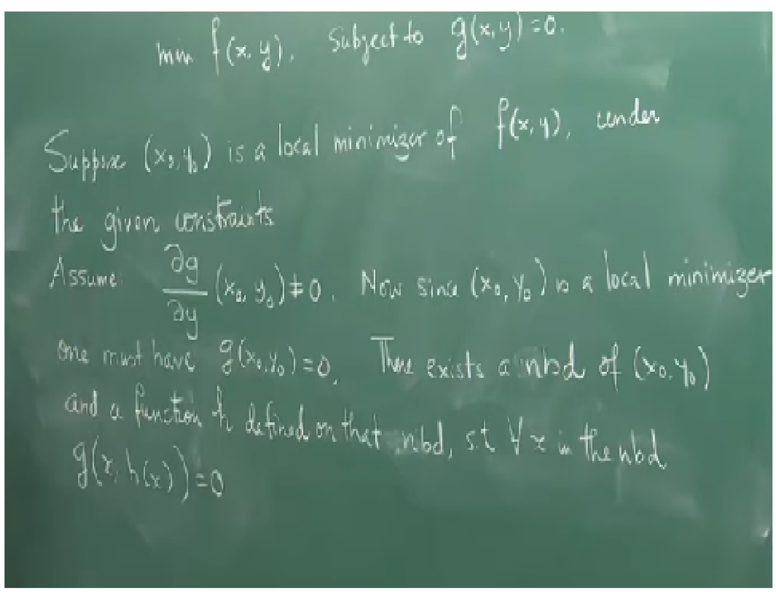
And then he would take rest and come back again in the evening that kind of dedication Lagrange had to mathematics and so what did Lagrange said Lagrange says in this first eliminate to differentiate first differentiate and then eliminate that you differentiate with this function with respect to x and then y take the partial derivatives. That is, you basically take the gradient and then you take the one with respect to lambda and equate them to 0 and solve that system of equation.

So what you do so first show you basically Lagrange says okay take the gradient with respect to x,y. So I am just taking the gradient I am just writing x,y because just to say that okay x,y has some value of course my name you put in something you solve that system. So this one also says you take the gradient just with respect to lambda so keep x and y fixed up. So Lagrange says that if you solve these two equations you can find an x,y which satisfies this system.

But if so this is what is called the optimality condition or necessary condition for solving a for if there is if the point is a local minimizer of this problem minimize f x,y subject to gx,y = c or g x,y = 0. For example then this is these two are the necessary optimality condition which simply says that grad of f x,y + lambda grad g x,y = 0 and if I do this differentiation I will simply have gx,y = 0. So this is the system that he wants to solve you eliminate the lambda compute the x and y but if you get such an x,y it does not tell me that it is really a maxima or minima.

Only looking at the structure of the problem, you can determine whether this is a maxima minima or whatever but how do you reach such a condition how do you know that such a lambda exists because every time you may not be able to write like this y = hx and here the implicit function theorem that you have learned which I want you to go and see again that would play a very key role and that I really want to explain in the board. So how the implicit function theorem is used.

(Refer Slide Time: 21:17)



So again we start by writing the problem now the issue that comes in here is important for every function if it is a very implicitly written function like x,y +. So for example if I write x,y + y square x + x cube = 0 if I write something like this you cannot write y = fx y the function of x cannot be written what you have to understand here comes in the role of implicit function theorem.

Suppose I have I know suppose x0,y0 is a local minimizer local minimizer of fx,y under the given constraints. So assume that the following condition holds assume that del g del y x0,y0 is not equal to 0. Now because it is a local minimizer now since x0,y0 is a local minimizer it must have a must satisfy this feasibility condition. One must have g x0, y0 = 0 and you also have this condition and these two conditions are right for application of the implicit function theorem.

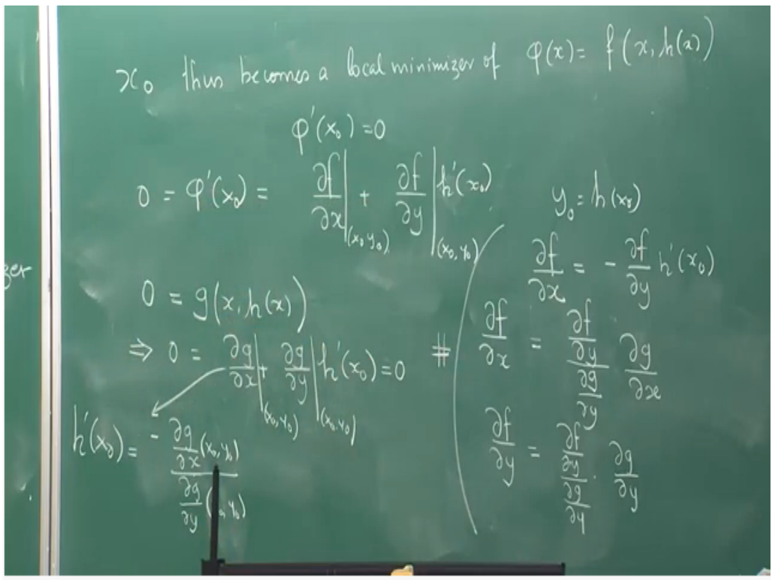
So if these two conditions are made. This is always made if x0, y0 see these necessary conditions assume this fact that we start we assume that we have a local minimizer and want to see what condition it satisfies. So if you know what condition is satisfied you know that you if you want to find a maximizer or minimizer in this particular case. We really have to figure out a points which satisfy these conditions.

We find the lambda x, y such that these things are satisfied. Once we can do that then we can further check whether maximization and minimization holds I will give you again I will write it down in your extra things which I will come within few days. I think by next Monday you will have it on your portal that if your f is a convex function and g is a convex function sorry g is a linear function kind of a nice kind of convex function linear function.

Then we will have the fact that whenever you have the any point which satisfies the Lagrange multiplier rule will give me a solution. But if not then we have to figure out on our own looking at the condition otherwise their second-order conditions which we do not want to get into that is pretty slightly advanced. Now suppose this is there so what so there exists a neighbourhood of x0, y0 such that for all there is a neighbourhood of x0, y0 and a function h defined on that neighbourhood.

So that is a definition function is defined from neighbourhood to r defined on that neighbourhood such that for all x in the neighbourhood I am not writing the technical things just telling you story. So we would have gx hx = 0. So basically in that neighbourhood ys are expressed as y = hx. Function as an explicit representation in that given neighbourhood so which means now in that small neighbourhood around x0, y0 why the function we have an explicit representation.

(Refer Slide Time: 27:42)



The function can be from here we can write y can be represented as h, x0 thus becomes a local minimizer of phi x = f of x hx. So we have replaced y with the hx. So in that local zone x0 is the minimizer because x0 is the local minimizer. Because h x0 is y0, h x0 has to be y0 right. And so if you take any other point x, y it is of the form where y is of the form hx in that neighbourhood. So you have a neighbourhood at where this feasibility satisfied plus because x0, y0 is already a local minima.

There is already a neighbourhood around x0, y0 such that fx, y is always bigger. So you consider that part the neighbourhood so small on which this is satisfied you can get a neighbourhood pretty big but even consider the neighbourhood to be contained inside the neighbourhood where the local minimizer stuff holds. So what happens that x0 becomes a local minimizer of this. Listen to this talk stop it repeatedly do back and forth and listen to what I have said.

So you have x0, y0 which is a local minimizer. So you have a neighbourhood a circular disc around it. So there and for all points on that f of x0, y0 >= fx, y. If those point satisfies for those points which satisfy gx, y = 0. Now here we have a class of such points. We have a neighbourhood around x0, y0 by the implicit function theorem such that this happens. Now if that neighbourhood is bigger than what the earlier neighbourhood but I can have a neighbourhood smaller than this.

This basically if this neighbourhood is bigger than the neighbourhood used for the definition of the local minimizer I can actually consider only a small part of it which is inside that local minimizing thing okay. So which means that x0 becomes a local minimizer of this. So which means phi dash of x0 = 0. But what is phi dash x, if you go by that if you go by your chain rule it is del f del x + del f del y into del h del x but del y del x which is y is hx so it is h dash x.

So basically it is x0 this things evaluated at x0, y0 I am not writing them. So evaluated at x0, y0 that is x0 here y0 is h of x0, y0 of course is h of x0 because it is x0, y0 is in that neighbourhood. So this is 0. So phi dash x0 is this and this is 0. So instead of writing del f del x at x0 we will just simply just write it. So again you know that 0 = g of x, hx. So again you take the derivative and use a chain rule.

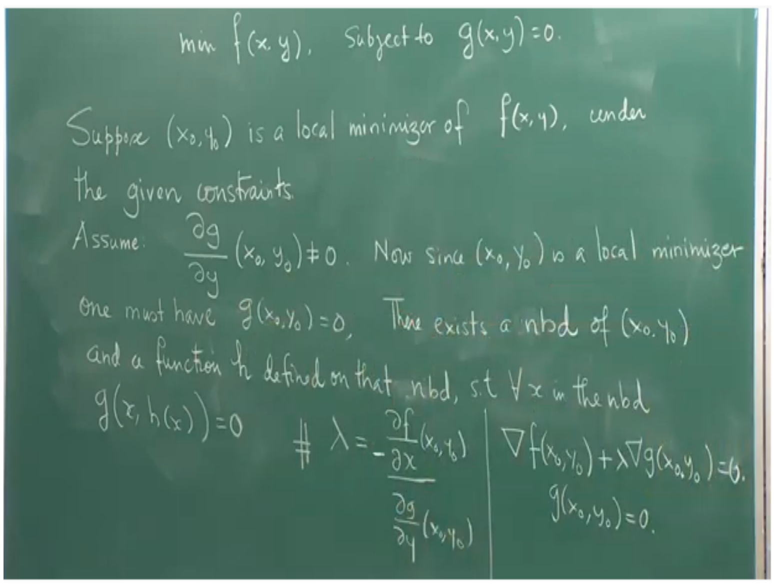
This will imply that 0 = del g del x + del g del y into we are taking the derivative with respect to x now only with respect to h dash x at x0 basically 0. So here we have 2 equations. So from here I will conclude from here the del f del x = -del f del y h dash x0. But what is h dash x0 I can write from here it is -del g del x/del g del y and del g del y is not = 0 at x0, y0. So these are all evaluated at x0, y0 x0, y0.

So if you want to be so clear put x0, y0. So this is non 0 so what do I have from here? From here I have h dash x0 = -del g del x/del g del y which is non 0. This is evaluated at x0, y0 I can also write on the sides x0, y0. So if you want the writing to be much more clearer you can write like this del g del x at x0, y0, del g del y at x0, y0 it is non 0 so it is very this is absolutely proper math.

This is minus of this. So that is what you get from here. So you put that back here. So you get del f del x = del f del y/del g del y all evaluated at x0, y0 del f del x also is evaluated at x0, y0. If I am not writing x0, y0 every time. So I will not write x0, y0 I will write x0, y0 at the end. So that I do not make the whole thing look very complex and this is the into del g del x okay. Now interestingly enough if you observe what is del f del y?

This is all evaluated at x0, y0 please. It is nothing but del f/del g del y into del g del x del g del y so this cancels and I have del f del y. So what is important here is the following. So I will now what I want from whatever have there on this part, from this part I will now take things from this part and come take an hashtag and come to this spot.

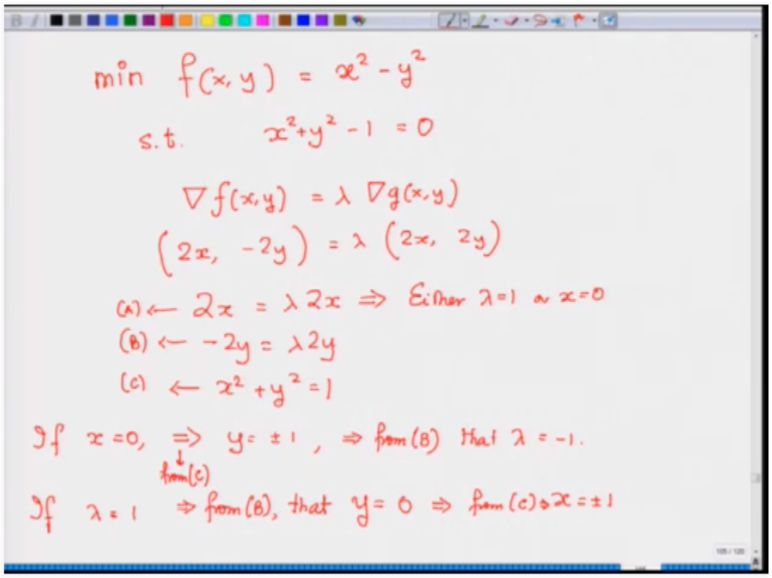
(Refer Slide Time: 35:02)



Now I will set lambda = del f or del f del x/del g del y evaluated at x0, y0. This is by Lagrange multiplier I am showing the existence of something. So lambda I can put here as minus of this. Then what I have? I have grad of f of so this is -lambda. This is -lambda, this is -lambda. So I have at the end grad of f0 x0 + lambda grad of g x0, y0 = 0 and of course you have g x0, y0 = 0. So that condition that we have written earlier actually satisfied. This is already known to you.

So this is the Lagrangian multiplier rule it shows the existence and existence of the Lagrange multiplier rule the Lagrange multiplier comes from the implicit function theorem and that is the key idea that one has to keep in mind. And so let us do an example here and through that example we solve this problem okay. So let us take one example and then try to see what we can get from the Lagrange multiplier rule. You can go to higher dimensions also but I am not getting into higher dimensions with you at this moment.

(Refer Slide Time: 36:40)



So take a simpler problem minimize f of x, y equal to I am taking this example from book so standard one which are using subject to x square + y square -1 = 0. If what is my f? So instead of if I call this just lambda then I can write this as is if for example instead of lambda I put lambda is -mu or I can replace lambda is - lambda dot lambda dash or lambda 0. Then I can write this as grad f x0 =lambda 0 grad g x0, y0.

So whichever way you can write. So I am writing in that simpler form where grad f x0 is lambda times grad g x0 it says that the 2 vectors are parallel but they are linearly dependent grad f is linearly dependent on grad g those who know linear algebra I am using for that terminology and those who do not know linear algebra just do not worry nothing is going to happen. You just apply it mechanically at least you know why such a lambda exists the implicit function theorem that you have learned is actually working.

So math works that is very important thing that is the very important reminder of this lecture. So what you do is you try to find out an x, y which satisfies this. So here are my x, y is f is x square - y square it is 2x and -2y times g is 2x, 2y. So what do we have? We have 2x = lambda 2x - 2y =

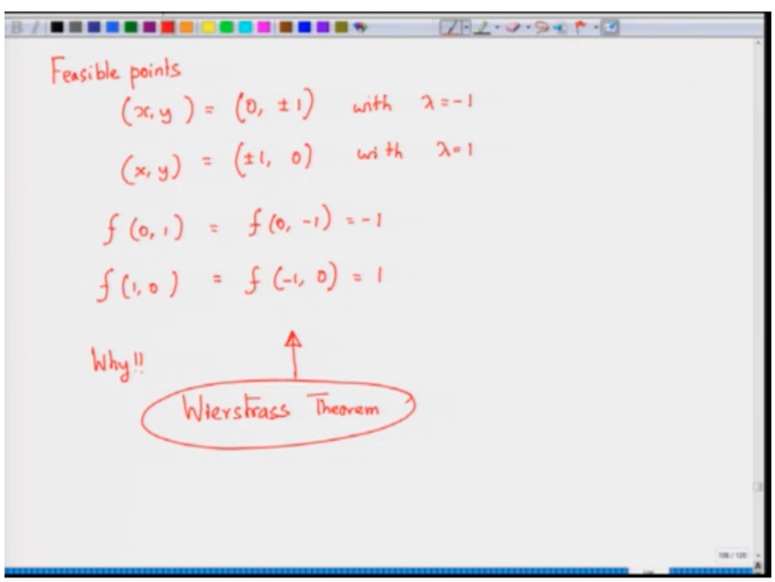
lambda 2y and of course we have x square + y square = 1. So if lambda = 1. So from this equation what does it imply either lambda = 1 or x = 0.

If x = 0 whatever lambda you want it will work. If lambda if you put x lambda = 1 then whatever x you choose works. So either lambda = 1 or x = 0. If x =0 then it implies that from this equation y + - 1 because y square =1. So if y + - 1 say if I put y = 1 here, so I will get lambda = -1. If I put y = -1 here still I will get lambda = -1. So then it implies so I will write the question as A, B and C.

So it implies from B here it implies from C, it implies from B that lambda = -1. So this is what I now if so from the first we also have another conclusion. If lambda = 1, it implies that if lambda = 1 if I put lambda is 1 okay. So lambda means once means I can take any x lambda means lambda = 1 simply means I have to put y to be 0s here. Because I cannot put y to be 1 because it will become - 2 = 2 so lambda is 1 then it implies from B that y = 0.

Because there is no other choice you if lambda = 1 for this equation - a = a is any - x = x hold only if x = 0. Basically if lambda = 1, you will have 4y = 0 so y = 0. So y = 0 so that implies from C x = + -1 okay. That is what we have till now. So now you have to see what are the eligible points.

(Refer Slide Time: 41:57)



What are the feasible points basically? The feasible points here are x, y is 0, +-1 with lambda = -1. So these are the possible chances of maximizing or minimizing the problem. So you can actually find maxima or minima let us so maybe I can make the problem more minimize or maximize whatever you want do both let us try out both and see if we can get. So there is another set of points x, y where x is +-1 and 0 with lambda = 1.

Because observed that when lambda = 1 we have y = 0 and x is this. When x = 0 we have y is this and for that lambda is equal to this. So these are the 2 situations. Now interestingly enough f 0, 1 = f 0, - 1 = - 1 if you look at the function value which is and f 1, 0 = f of –1, 0 which is 1. So you know which are the values which are maximizing which are minimizing. Now why these are maximizing and minimizing value why? That is the question we have not answered.

How do you know that these are not maximizing and minimizing well I can have something less than I can have something worse, I can this function value can be lesser than - 1 subject to this? But you have to observe subject to this it cannot have value lesser than - 1 or + 1. So how do we know that this problem has a maxima or minima this goes back to the age old question of Weierstrass which I have not mentioned here.

So if you take a set which is closed and bounded in the sense that it is a kind of boundary and it is in the sense that they all the xs the norm of all these xs in this set is less than some given number so it is bounded and closed means the set is something with a boundary for that kind of sets Weierstrass said that if you have a you want to have maximize or minimize continuous function over it.

You will always finds points inside the set which will have a maximizer and minimizer and that what these books do not explain that because of the Weierstrass theorem that a continuous function can be maximized or minimized over a closed and convex said it is bounded over a sorry bounded over a closed and bounded set. And there are points in that set where the maximum value and the minimum value would be achieved.

So because such maximizer and minimizer exists and that maximizes and minimizes must satisfy this Lagrangian multiplier rule. So whatever points we just satisfied the Lagrangian multiplier rule we have checked with them what are the values have been observed. So the maximizer and minimizer must lie among them and so the 2 gives the minimum, the -1 is a minimum value and +1 is the maximum value.

So it is a Weierstrass theorem which we know for a function of 1 variable that if I take a function of a real variable from a closed interval A, B to R then the function is bounded on the interval A, B and there are points x when a points x1 and x2 in the interval A, B such that f x1 is the maximum value and f x2 is the minimum value of the function. The same story can be repeated here where a function of 2 variables is continuous what the set that you have is closed and bounded I am not getting the geometry in too much detail.

But that is what it is because of that Weierstrass theorem you can you know that there are minimizers and maximizers are existing and you get the points which satisfy the Lagrangian multiplier rule and you have jotted down the function values and you know what is the maximum value and minimum value and you know that the maximum value and minimum value would exist. So and you know which are the maximizers and which are the minimizes.

And with this I end the talk and hope that you have enjoyed and hope that you enjoyed this proof which shows the power of the implicit function theorem. There are many ways to prove the Lagrangian multiplier rule. People are still publishing papers like a short proof of the Lagrange multiplier rule and elementary way to publish a Lagrangian multiplier rule they are still coming in actually. So the interest with Lagrangian multiplier rule has not waned out. It is still extremely important in science. Thank you very much see you in the next course.