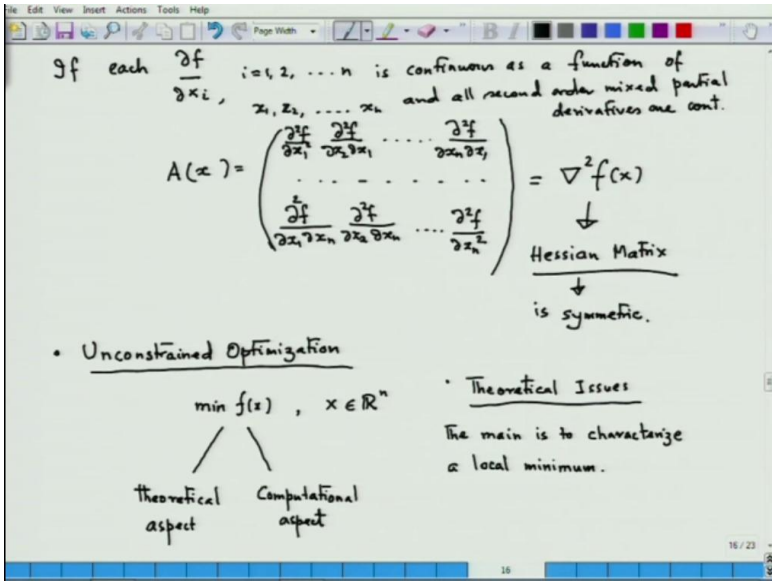
Foundation of Optimization
Prof. Dr. Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

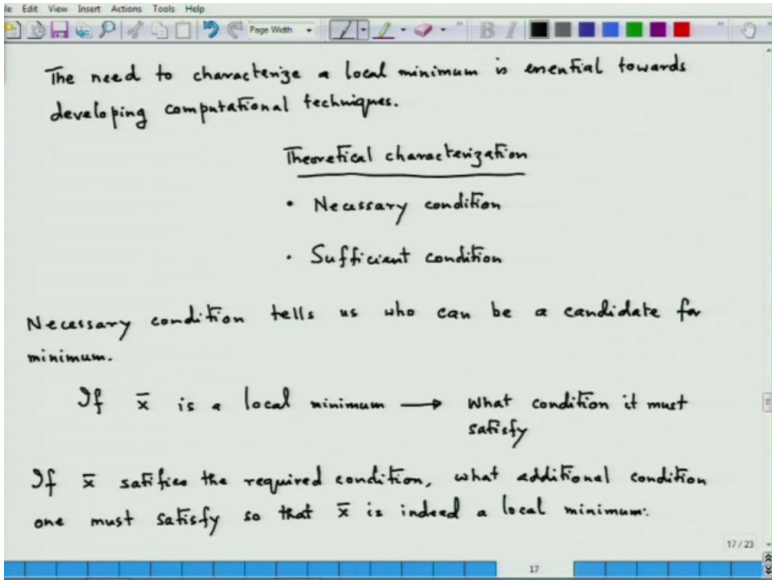
Lecture - 3

(Refer Slide Time: 00:25)



So, today we are going to concentrate on unconstrained optimization. That is here we are going to look at minimizing effects where x is in R n whole R n. Now, what are the two aspects of this problem? So, any optimization problem has two aspects them two aspects; first is the theoretical aspect and the second aspect is the computational aspect. Now, in the theoretical aspect, what are we suppose to do in the computational aspect, what are we suppose to do these are the two main themes of discussion we will carry out today. So, let us see what is the, what is the theoretical issues here. The main issue is to characterize a local minimum. Why you need to characterize a local minimum?

(Refer Slide Time: 02:05)



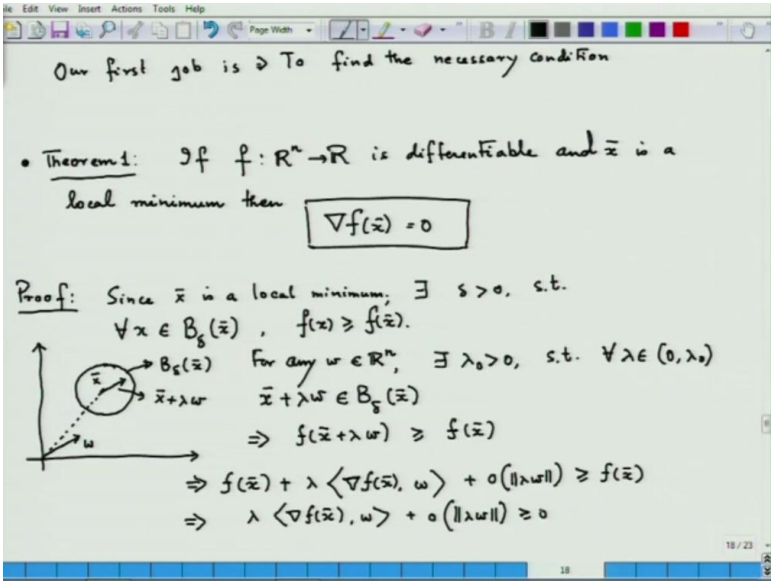
It is important to note the need to characterize local minima, minimum is essential to develop is essential towards developing computational techniques. So, in the theoretical characterization there are 2 things that we really have to bother about these. So, there are 2 aspects of it namely the necessary condition and other is the sufficient condition. Now, what do these two conditions do? A large amount of optimization literature is bothered about really geared towards designing these sorts of conditions for various types of optimization problems. Now, why you need a necessary condition, why you need a sufficient condition? Necessary condition tells us how to compute a candidate point for a minimum.

So, necessary condition tells us who can be a candidate for minimum, who can be a candidate for minimum. Now, if he tells you who can be a candidate for minimum. So, what it should do? So, necessary condition will do this. If x bar is a local minimum, what condition it must satisfy? Means a point, which does not satisfy that condition cannot be a local minimum. So, that is very important. So, in order that your point that you say that possibly be a local minimum has to satisfy this condition. First you have to find the point which satisfies this condition. Then you check whether it is a local minimum or global minimum or whatever.

If a point does not satisfy this condition, then it is ruled out it is it can never be a local minimum. So, that is why necessary conditions are very important or rather central to

the study of optimization. Now, if x bar satisfies the condition the required condition, what additional assumption would guarantee that it is really a local minimum that is where sufficient condition comes. It tells you if these conditions are satisfied along with the fact that x bar has come from the computation that you do with the necessary condition then that x bar is a minimum. What condition it must satisfy? So that, x bar is indeed a local minimum. Now once you know this. So, our first job is let us layout our job properly.

(Refer Slide Time: 07:38)



Our first job to find the necessary condition, this we state as follows, which we write as theorem one of our study. It says that if, if from R n to R is differentiable and x bar is a local minimum, then my next question is, then what would happen then gradient of f at x bar must vanish. Like any other part of mathematics, optimization in a mathematical subjects it demands proof of what statement you are making. You have heard this when n is equal to 1 that is when f is from to R you have studied this in school if you are trying if you are given a ordinary function say x square plus two x plus theta when I ask you to find the minimum. You will immediately go and find the derivative you got it to 0 find a point and try to do something. So, here so what we have to do to find the local minima is to first find a point which satisfies this.

So, let us see whether this condition, how this condition gets satisfied? So, what we are writing here is formally the first proof of this talk. How does we how does we do we

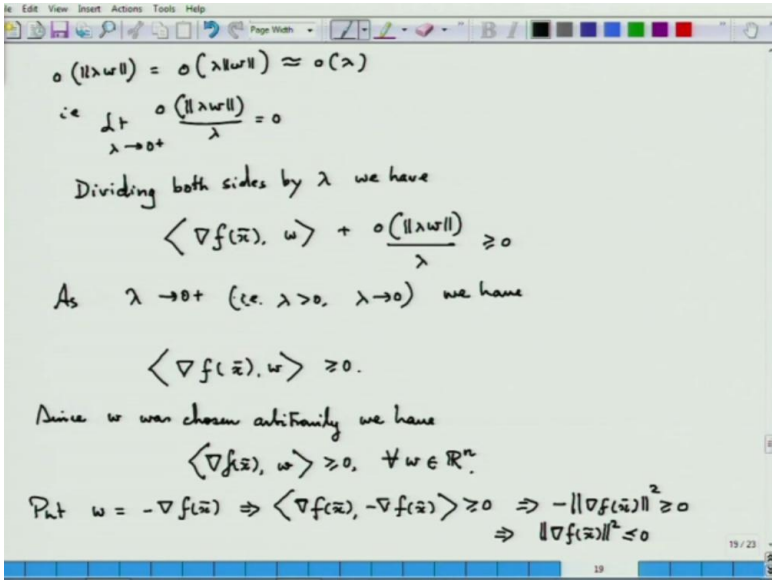
proceed to do the proof? No one has to understand, we are talking about local minima so around x bar we are looking at a neighborhood around x bar, we are not looking around, around the looking at the nature of the function very far from x bar. So, then we can employ tailors theorem or which is another way of talking about differentiability. So, we will use this idea of differentiability in a very, very sensible way. So, what does it mean by a local minimum. Since x bar is a local minimum. So, I will write it down completely. So, you can follow the proof step by step.

Now, once you know that x bar is a local minimum what you have there exists delta greater than 0, such that for all x in the delta neighborhood of x bar, the ball around x bar f of x must be bigger than f of x bar. Now, I will try to draw the picture of this scenario in order to which can, which will make proof much clear. So, here is my x bar and here is the ball here is my vector x bar. So, this is my B delta of x bar. What I can do is let me take any vector w in any direction. Then I can choose I can choose to move from x bar along the direction w, say this is my w vector and so this was my x bar vector.

So, I can choose to move from w from x bar long a direction parallel to w. So, let this point where I come and stop the x bar plus lambda w. Now, you see if I increases lambda there will be a threshold beyond which it will come out of the ball and then I cannot do anything. So, if I am within the ball I can relate the function values of this and with that of x bar that this function value at this point must be bigger than function value at x bar. So, for any w element of R n there exists lambda naught strictly greater than 0, such that for all lambda between s naught you see even if I lambda naught is the point where x bar plus lambda w this vector comes and touches the boundary, because once you are in the boundary you are out of the neighborhood.

So, that is why you have a opened here. For all of these, so that is for all lambda of these x bar plus lambda w remains in the ball this naturally implies f of x bar plus lambda w is bigger than f of x bar, but knowing the definition of the derivative. Because the function is differentiable at x bar I can write I would just like to recall where I had written the definition of the derivative, this is the definition of the derivative. So, I can write here f of x bar plus lambda times grade f x bar to w plus mole o of ohm of lambda w. And this is bigger than f this also means the following this means that lambda times grade f x bar, this is what you get, because I have cancelled f x bar from both sides.

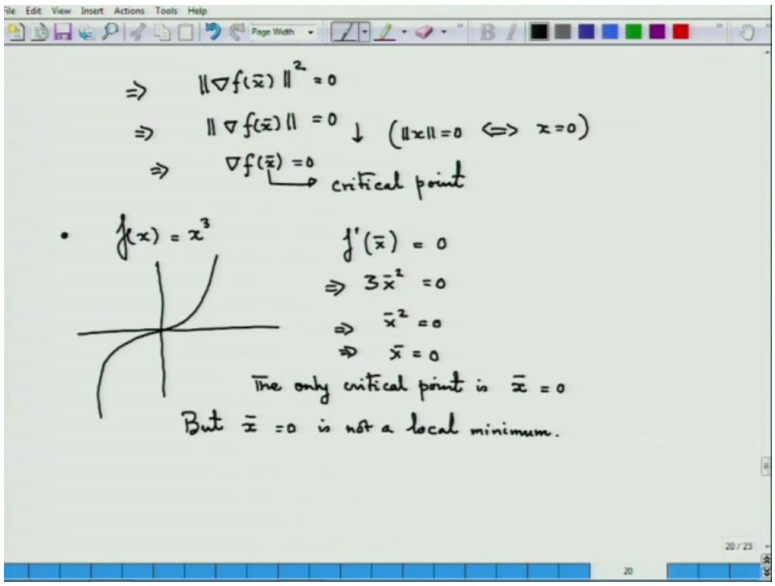
(Refer Slide Time: 15:08)



Now, let us look at this term o lambda w. So, whatever be the powers of ohm w, lambda would have the same powers. So, if you have, because you can bring out lambda outside here. So, you can write this as o of lambda ohm w. So, basically here if ohm w is fixed if I vary the lambda I can keep on changing the things. So, basically this is nothing but a o lambda quantity that is o of ohm lambda w by lambda as limit lambda tends to 0 plus, I have taken lambda as positive that should be equal to 0. So, here what I have observed now I have got if I divide by because lambda here is positive, because lambda is positive I can divide both sides by lambda.

So, let me write down more clearly, we have now as lambda tends to 0 plus that is lambda strictly greater than 0. And lambda is going to 0 so as lambda tends to 0 plus we have because this will go to 0. Now, w was chosen arbitrarily, w was not fixed thing fix vector. Since w was chosen arbitrarily, we have for all w element of R n for all w element of R n. So, once this is known, now put w is equal to the negative of grad f of x bar. So, this would imply from this condition that grad f of x bar minus grad f of x bar is greater than equal to 0 which would imply minus of because I take the minus outside and normal inner product of this a vector with itself is nothing but square of its own is greater than equal to 0 which would imply at the known, but it is this is the length. So, the length of a vector from the length of a vector cannot be less than equal to 0.

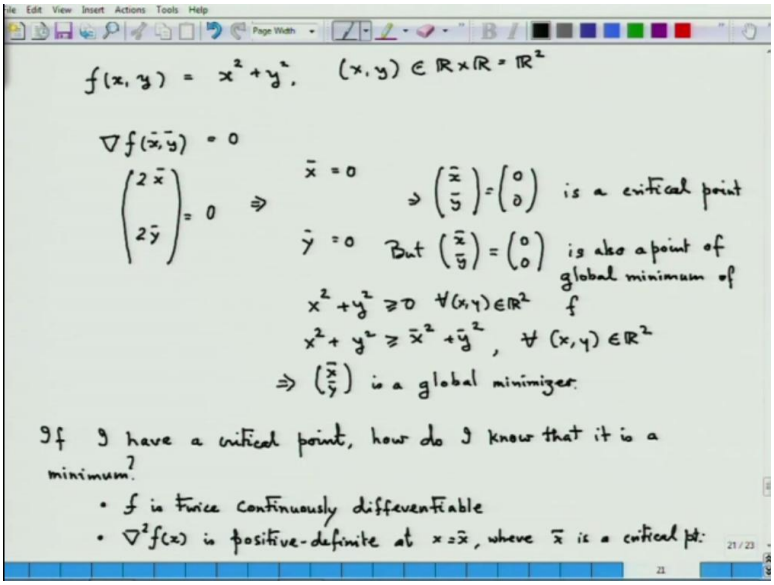
(Refer Slide Time: 15:08)



So, this would imply that and you know from the property of the norm this will imply that the gradient of f at x bar is 0. So, here to here we have applied the formula norm of x is equal to 0 if and only if x is equal to 0, this is the property of the know and here at this point from here to here we have employed the fact that x x, which have been a product of this is nothing but square of the norm. So, we have this condition. Now we need to test it up for example, you take first the function f x is equal to x cube. We want to show that it is truly a necessary condition at a point x bar which satisfies this need not be a point of minimum. Any point which satisfies, this is called a critical point.

Now, you have f x equal to x cube, let us look at the picture of f x equal to x cube a graph. Now, f dash x bar equal to 0 would imply 3 times x bar square is equal to 0 and that would imply x bar square is equal to 0 and that would imply x bar is equal to 0. So, the only critical point in this case is x bar is equal to 0. But x bar is not a point or not a local x bar equal to 0 is not a minimizer is x bar equal to 0 is not a local minimum. So, this is very important that this is just a necessary condition and not a sufficient condition.

(Refer Slide Time: 21:48)



Let us look at another example where this is possibly alright. So, x y here is in R cross R each individually from the real line which is nothing but R 2. Now, if I want to find the gradient of f x y, I want to find this means I would have two of x bar 2 of y bar. This vector would be equal to the 0 vector. This would imply x bar is equal to 0 vector and y bar is equal to 0 vector. So, it is x bar y bar is equal to 0 0 is the critical point. But x bar y bar equal to 0 0 is also a point of global minima of the function. How do I know this fact? Because you observe that x square plus y square is always greater than equal to 0. So, x square plus y square because this is all non negative is bigger than x bar greater than 0. this is true for all x y in R 2 where this is 0 this is 0.

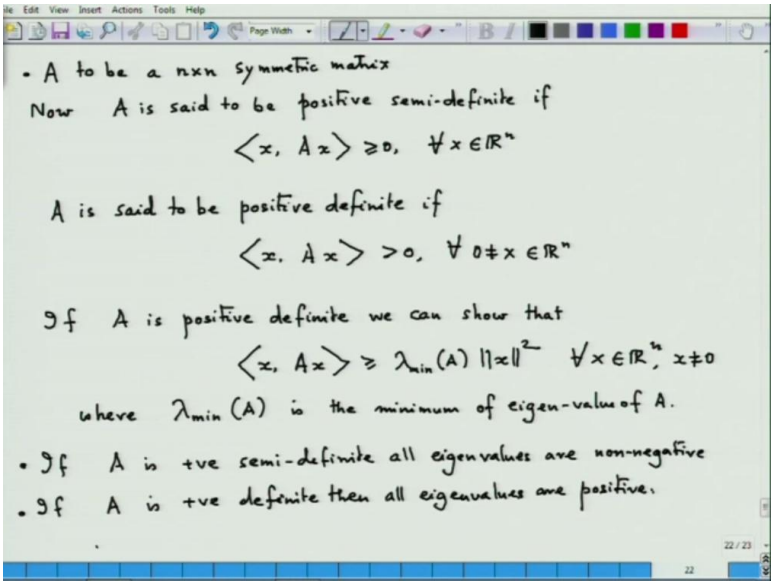
So, this would imply that x bar y bar is truly a global minimizer. So, here you see the two different notions of differentiable function in 1 case, I am having the critical point to be a global minimizer this forms an important class of functions called convex which will focus on little later for some part of for a little part of the course. But which is none the less important and there is a complete course on optimization problems with convex functions which are already been delivered and you will see it quite soon. But here we are dealing with all types of functions which are differentiable could be convex could be non convex whatever. So, here is so there are two examples; one example here for example, this function is really not convex.

Now if, now the question is what is the point here? Actually here you see the curve it changes the nature. So, if you have a function from R to R then a point is either a local or global minima a critical point or it is a like this where the curve is changing shape. In general, I could find the critical point which is neither a local minimum, local maxima nor nothing it is just a critical point for and it is not also a saddle point. Now, I am not going to get into the saddle point issue right now, because they might just confuse you. So, what I am going to get going at this movement is that if I have a critical point, how do I know that it is a minimum. To answer this question we have to additionally assume certain conditions on function f.

So, what are those conditions? so trying making an attempt to answer it just by having the knowledge that f is differentiable or even continuously differentiable does not help me at all it does not tell me anything, because that is where we get stuck. But now if I put some more additional conditions on the function can I device a condition which if satisfied by a critical point with guarantee that such a point is a local minimum. First condition is that if f is twice continuously differentiable.

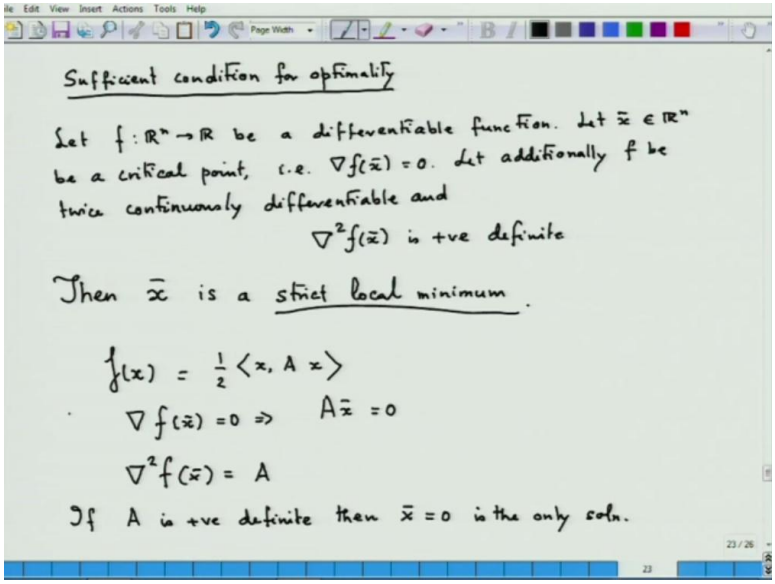
And second condition is the Hessian matrix is positive definite at x equal to x bar, where x bar is a critical point. Now, what do you mean by the term positive definite and what do you mean by the term positive semi definite these are terms which comes from matrix theory, but which are very important in optimization. And so I would just like to take a minute to explain what these terms are. So, our matrices by because we have taken the function do be twice continuously differentiable the hessian matrix is a symmetric matrix.

(Refer Slide Time: 28:23)



So, let us consider a to be A n cross n symmetric matrix. Now, now A is said to be positive definite positive, semi definite positive semi definite. If the inner product of x with A x is always bigger then equal to 0 for every x in R n. So, this is condition is always taken and now A is said to be positive definite if x A x is strictly bigger that 0 for all non zero x in R n. In fact if A is we can show that, this is where lambda min of A is the minimum Eigen value. It is important, however to note that if A is positive semi definite all Eigen values are non negative. If A these are also 2 bullet points, if A is positive definite, then all Eigen values are positive. Of course, these Eigen values are all real, because A is a symmetric matrix. Now, this is what you have now we go to the main question the second order result.

(Refer Slide Time: 32:25)

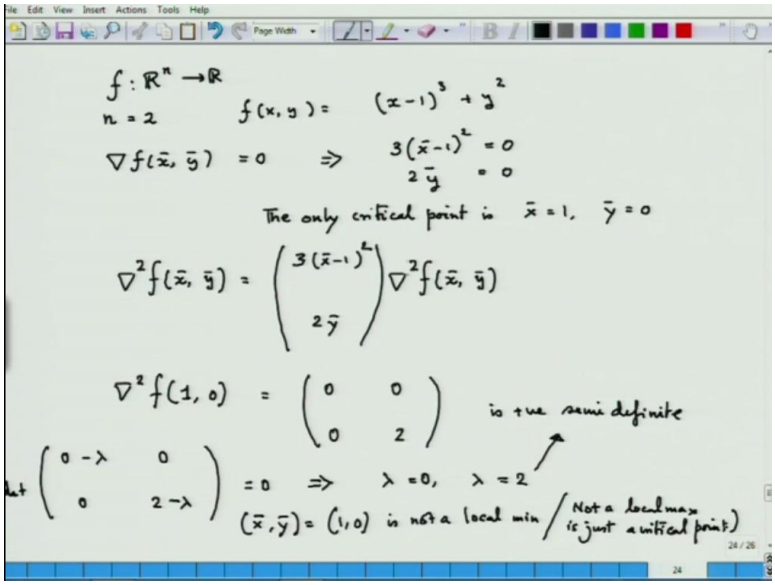


So, we are now going to state the sufficient condition for optimality and you would be amazed how strong a result we get. Now, here is the result. You might ask me for a proof of this, but the proof keeping in view that this optimization course is really for a very broad audience from very different disciplines in engineering in the sciences, so in mathematics of course. But would not like to bog you down with the proof of this which is quite technical and needs a slightly more deeper understanding on mathematics I would rather give you the result. And you would see for yourself how to apply it, then we would go more into the algorithmic aspect computational aspect which would be really of useful to you, and and you would be really able to use it in several things specially those who are in engineering sciences. So, let us just state this result.

So, let f from R n to R be a differentiable function. Let x bar elemental of R n be a critical point that is gradient of f x bar is equal to 0. Let additionally f be twice continuously differentiable which you had added here also I just did not do it. And the hessian matrix is positive definite, then x bar is a strict local minimum proof of this would be added to the f A Q of this subject. But here we just do not do the proof. So, you can see it later on in the website of this course main conclusion here is that it is not just a local minimum it is a strict local minimum. So, it forbids global maximum to take this position so global maximum can never be a strict local minimum. So, we are really getting a local minimum and not been fooled by flat type of functions where you have you can say a local minimum can be also global maximum.

So, those sorts of anomalies won't come because it will give you this. So, this is a very important result and this result has to be really appreciated. Now, if you look at for example, I take an example of a quadratic function. Now, then grad of f x bar equal to 0 would imply A of x bar would be equal to 0. But if A is positive definite, then you will observe that the hessian matrix is a here for any x actually the hessian matrix is A. So, if A is positive definite here the answer would be of course, 0 as we will soon see if a is positive definite. Then x bar equal to 0 is the only solution it is a strict local minimum there are many other cases where a your second order condition you have learnt in school actually gives you the strict local minima. So, this is very simply case actually which is to illustrate you that how to how do you find the hessian.

(Refer Slide Time: 38:13)



For example, if you have a say a more slightly strong slightly different looking scenario, put n equal to 2. Now, try to figure out say let me put f of x y. So, let me put out try to figure out this. So, this would imply 3 x power minus 1 whole square is equal to 0 and 2 y bar is equal to 0. So, the only critical point is, point is x bar is 1 and y bar is 0. Now if I want to compute the hessian of this. So, I compute the hessian of this. So, you I take the gradient of this right. So, I take the gradient of this in x bar, then in y bar and write it as a row vector. Then this will become 6 x bar minus 1 and the next term would be 0, because there is no y term. And here first I will differentiate with respect to x term which is 0 and then I have 2. Now, grad square f at 1 0 would give me 0 0 0 2.

So, let me try to find the Eigen values of this matrix, which Eigen values of this matrix means you have to find, so that determinant of this is equal to 0 which would imply lambda is 0 and lambda is 2. So, this matrix is positive definite. So, this is positive semi definite, but not positive definite. In fact, in fact x bar, y bar equal to 1 0 is not a local min. In fact, it is not a local max is just a critical point, see it is positive semi definite does not give you, what you want? What you require it is positive definiteness, which will give you what you require. So, this example illustrates that.

So, with this I end today's discussion. And tomorrow we would talk about decent direction as to how to really compute a point. But you know in optimization we will learn a very important lesson tomorrow that real optimization problems cannot be solved exactly. So, we will get in to that issue tomorrow and in detail try to develop computational methods to solve or minimize a differentiable function over R n.

Thank you very much.