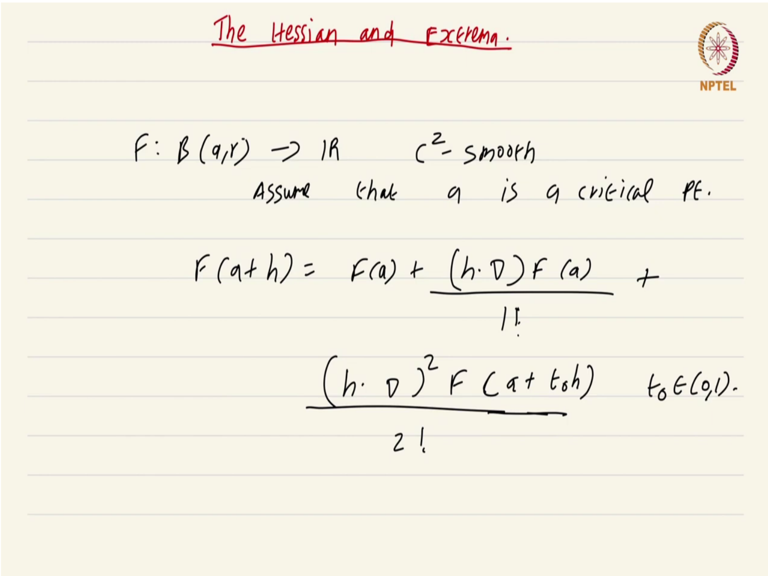
Real Analysis II
Prof. Jaikrishnan J
Department of Mathematics
Indian Institute of Technology, Palakkad

Lecture – 22.2
The Hessian and Extrema

(Refer Slide Time: 00:22)



In this video we are going to use the Hessian of a function at a critical point to characterize whether that critical point is a point of local maxima, minima or a saddle point. So, the setup is as follows. F is from B a, r to the real numbers, it is a C 2 smooth function. We are assuming C 2 smoothness so that we can take the Hessian that should be clear. Now, what I am going to do is I am going to also assume that a is a critical point.

Now, we will invoke this assumption a bit later. So, for the time being we are just going to use the C 2 smoothness part of F. Now, what we are going to do is we are going to Taylor expand this function F about the point a.

We already know that for points near a we can write F of a plus h is F of a plus h dot grad F at a divided by one factorial which is just 1 of course, plus h dot grad squared F at a by 2 factorial and here I cannot just put a this is actually going to be that error term I am using the Taylors theorem with the remainder.

So, I can only say a plus t naught h where t naught is in 0, 1. So, this is what Taylor's theorem would tell me. I can approximate the value of F at the point a plus h using the value of h dot grad of the point F at a plus this second order term h dot grad squared F of a plus t naught h where t naught is a point in between 0 and 1.

(Refer Slide Time: 02:21)

$$(h\cdot\nabla)^2 F(a+t_0 h)$$

$$= \sum_{i,j=1}^{n} D_i D_j F(a+t_0 h)\, h_i h_j.$$

$$\parallel$$

$$h \cdot \left(\mathcal{H} F(a+t_0 h)\, h \right)$$

$$DF(a) = 0.$$

Now, a straightforward an utterly easy computation will tell you that this h dot grad squared of F of a plus t naught h this mysterious quantity is none other than summation i equals i comma j running from 1 to n D i D j F of a plus t naught h multiplied by h i h j ok.

So, I leave it to you to do this straightforward computation and conclude this fact that h dot grad squared F of F at the point a plus t naught h is nothing but this big sum. Now, this is looking a bit complicated, but if you think about it for a second I can express this in terms of the Hessian. This whole thing is nothing but this is nothing but h dot product the standard scalar dot product the Hessian of the function F at the point a plus t naught h.

Now, this Hessian is going to be a n cross n matrix. Recall that the Hessian is nothing but the matrix of partial derivatives D i D j. It is going to be a symmetric matrix because we are assuming that this function F is c 2, but for the time being that is irrelevant this Hessian at the point you can treat it as a linear transformation because it is an n cross n matrix and it eats up vectors to produce another vector.

So, I act this matrix on the vector h that will produce for me another vector in R n and I can take the dot product of that vector back with h itself. So, again this is a straightforward computation why these two quantities are equal and I am going to leave it to you ok.

So, we have also assumed that gradient of F at the point a is 0 because we are assuming that it is a critical point. So, the net upshot of all this is I can express F near the point a just using the value at a and the Hessian.

(Refer Slide Time: 04:28)

$$F(a+h) = F(a) + \frac{1}{2}\left(h \cdot \left(\mathcal{H}F(a+t_0h)h\right)\right)$$

$$\left| h \cdot \left(\mathcal{H}F(a+t_0h)h\right) - h \cdot \left(\mathcal{H}F(a)h\right) \right|$$

Cauchy-Schwarz

$$\leq \|h\| \times \left(\| \mathcal{H}F(a+t_0h)h - \mathcal{H}F(a)h \| \right)$$

$$\leq \|h\|^2 \times \| \mathcal{H}F(a+t_0h) - \mathcal{H}F(a) \|_{op}.$$

F of a plus h is nothing but F of a plus half of this h dot product Hessian of F at a plus t naught h acting on h ok. I hope I have put the right number of parenthesis. Now, this quantity

Hessian of F at a plus t naught h is annoying me a bit. I want to determine the nature of the critical point a using only the data at the point a. I do not want this t naught h term.

So, what I am going to do is I am going to somehow try to get rid of this t naught h term. So, to do that what I am going to do is I am going to see how much of an error I would make if I just replace this a plus t naught h by just a naught a. I mean I just ignore this term t naught h and I want to see how much of an error I would be making ok.

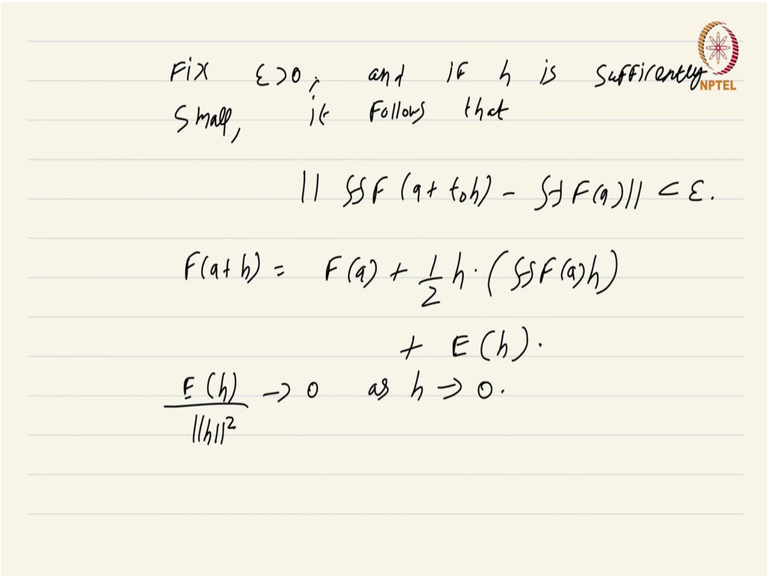
So, what I am going to do is I am going to consider norm of h dot the Hessian of F at a plus t naught h acting on h minus h dot product Hessian of F at the point a. So, this is solely to determine how much of an error I am making by replacing the term a plus t naught h inside the F Hessian by just a.

So, I want to analyze this term. Now, what I do is first I notice that I can take the dot product outside and then apply Cauchy Schwartz I can say this is less than or equal to norm h times this difference of Hessian of F at a plus t naught h acting on h minus minus; one second I made a minor error in the previous step I forgot a h there is a h here also ok yeah. So, what this will lead me to is I would get h F a acting on h ok.

So, this step is just this equality inequality is just Cauchy Schwartz is this Cauchy Schwartz ok. Now, what I am going to do is since I am treating this matrix the Hessian as just a linear transformation the natural thing to do would be to now use the properties of the operator norm and write this as less than or equal to norm h squared multiplied by Hessian of F at a plus t naught h minus Hessian of F at a and of course, this is the operator norm.

Whenever you just have a norm and the quantity inside is either a linear transformation or a matrix it is always the operator norm that I am taking unless otherwise specified. So, a bit soon I will intentionally start forgetting to put this op ok. Now, what does this tell us? Well, we know that the function F is C 2 smooth. So, the derivatives the second derivatives of F at the point a plus t naught h will get very close to the second derivatives of F at a.

(Refer Slide Time: 08:01)



So, the net upshot is if you fix epsilon greater than 0 if you fix epsilon greater than 0 and if h is sufficiently small if h is sufficiently small, it follows that I can make that quantity. It follows that I can make this Hessian of F at a plus t naught h minus Hessian of F at a, this I can make it less than epsilon ok. So, this just follows from the C 2 smoothness this is a rather easy to show.

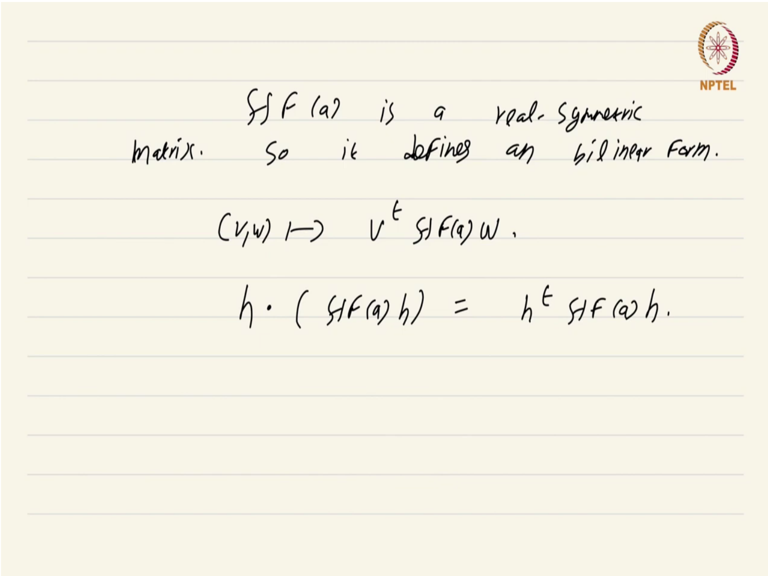
Now, the net upshot of this discussion is I can write F of a plus h to be equal to F of a plus half Hessian of F at a; I forgot that h term. So, half h dot product Hessian of F at a acting on the vector h plus an error term plus an error term ok.

So, this error term occurs because I have replaced h Hessian of F of a plus t naught h by just Hessian of F, but since we have chosen h sufficiently small such that this inequality is true you can easily check that E of h will be sub linear.

What I mean by that is he E of h by norm h squared goes to 0 as h goes to 0. This error term is quadratic because of this inequality because of this inequality ok. Now, because this error term is small we will be able to get some useful information about the behavior of F at the point a. I am going to come to that in a moment, but let us just focus on this particular key term this h dot H F a h.

So, this seems like an elaborate way of writing that one term and since I am assuming you have taken a course on linear algebra some of you might have already recognized what this quantity is going to be.

(Refer Slide Time: 10:27)



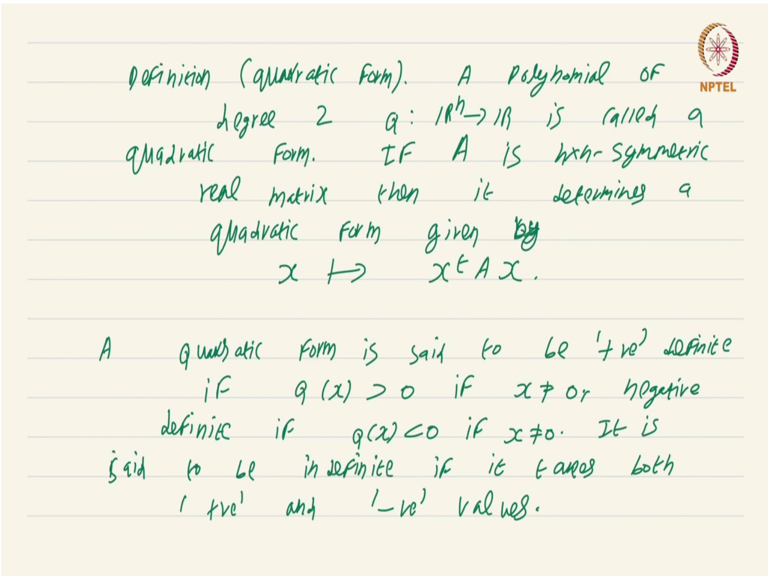
Well recall that h F a is a real symmetric matrix it is a real symmetric matrix. So, it defines an inner product rather it may not define an inner product it just defines a bilinear form. Inner products usually we assume positive definiteness if you recall. So, this Hessian for the time being since it is just a real symmetric matrix all I can say is its going to define a bilinear form. How is it going to define a bilinear form?

It is going to define it in this following way v comma w just goes to v transpose h F a w ok where I am treating the vectors v and w as column vectors. So, I am taking a transpose to make it a row vector and then I am acting the matrix F a on w. Of course, what I mean by this is I write down v and w in the standard basis. I write down the coordinates and do this computation. This will turn out to be a bilinear form.

So, the term that we are interested in h dot product H F a acting on h the silly term is actually nothing but h transpose H F a h ok. So, this is just taking the inner product of h with itself not taking the inner product taking the bilinear form h h where the bilinear form is the bilinear form determined by this real symmetric matrix the Hessian ok.

So, there is a name for this, it is called a quadratic form. I hope you are familiar with it. If not the main result that I need about quadratic forms I am going to do it anyway in the next video; so, no problem. So, let me just state the definition on a fresh slide.

(Refer Slide Time: 12:21)



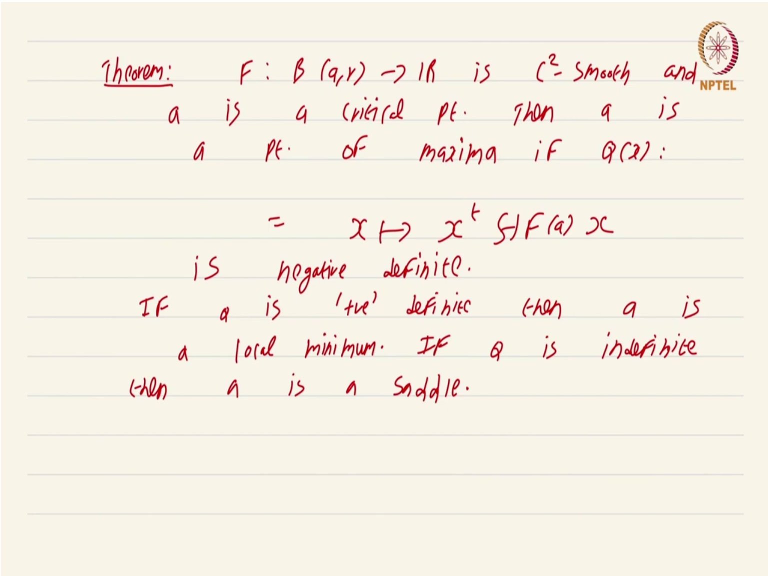
So, this is the definition of a quadratic form, definition of quadratic form ok. A polynomial of degree 2 let us just call it Q from R n to R is called a quadratic form is called a quadratic form ok. So, the definition is utterly straightforward. In fact, you could have guessed what it is just from the name. So, the key part is that symmetric matrices determine quadratic form.

If A is a n cross n symmetric real matrix real matrix then it determines a quadratic form given by x maps to again x transpose a acting on x that this is going to be a quadratic polynomial it just follows from the very definition of matrix multiplication.

Now, the key fact that I need is this quadratic form being positive definite or negative definite or indefinite. A quadratic form is said to be positive definite if Q of x greater than 0 if x not equal to 0, negative definite if Q of x less than 0 x not equal to 0 ok. It is said to be indefinite it is said to be indefinite if it takes both positive and negative values ok.

Now, I am going to prove a theorem that characterizes the nature of the critical point using the nature of the quadratic form determined by the Hessian. So, the theorem is as follows.

(Refer Slide Time: 15:13)



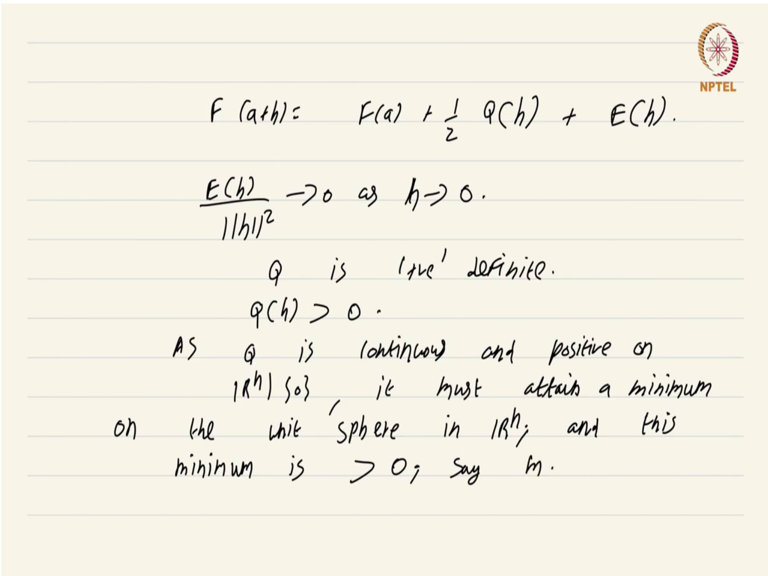
I am sure you would have used this theorem at least the special case when n equal to 2 or 3 in a basic course in multivariable calculus. So, the setup is as follows. F from B a, r to R is C 2 smooth is C 2 smooth and a is a critical point is a critical point. Then a is a point of maxima if this quadratic form Q of x which is just by definition equal to x mapping to x transpose Hessian of F at a x is negative definite.

Note this weird role reversal you might think that maxima means things should be positive, but no, at a point of maxima this quadratic form should be negative definite. And this should actually not be shocking to you because if you go back all the way to kindergarten when you studied maxima and minima of one variable functions using calculus you would notice that the second derivative has to be less than 0 at a point of maxima right.

So, or rather the other way around if the second derivative is less than 0 at a critical point then it is a point of maxima. So, this sort of generalizes that basic fact from elementary calculus in a similar way if Q is positive definite Q is positive definite then Q is not Q a is a local minimum.

If Q is indefinite then no price is here for you then a is a saddle ok. So, we are going to prove this. Most of the proof is already done in the discussion above. Let us just finish up the details. What we have to do is the following.

(Refer Slide Time: 17:44)



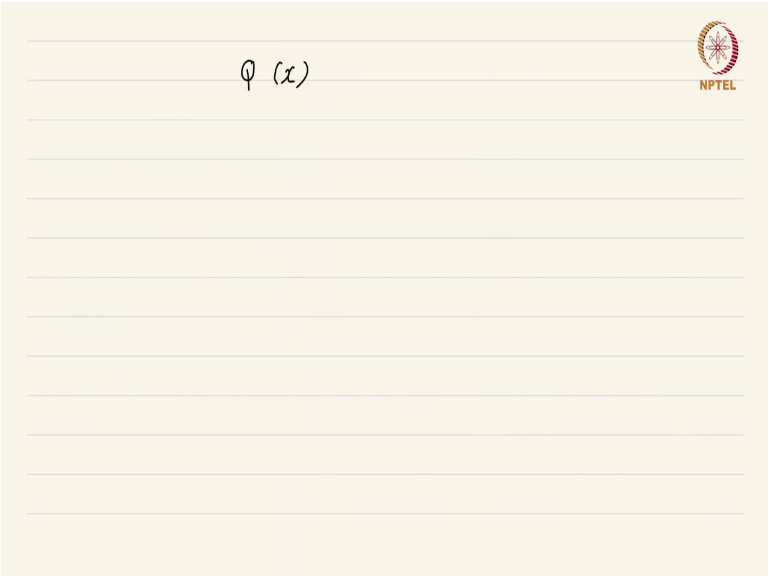
We know already know that F of a plus h is F of a plus half the Hessian of F at a. Again I keep forgetting to put now let me just put Q of h. Let me be lazy, I have earned that right plus the error term ok and we know that this error term is small its sub linear this quantity goes to 0 as h goes to 0 ok.

Now, we are going to just prove one case and leave the analogous case to you. I am going to assume Q is positive definite and end up proving that this is a point of local minimum. I am going to leave the negative definite case to you. Now, what this means is that Q of h is of course, greater than 0 ok. Now, Q is continuous as Q is continuous and positive on R n set minus 0, it must attain a minimum on the unit sphere on the unit sphere in R n right.

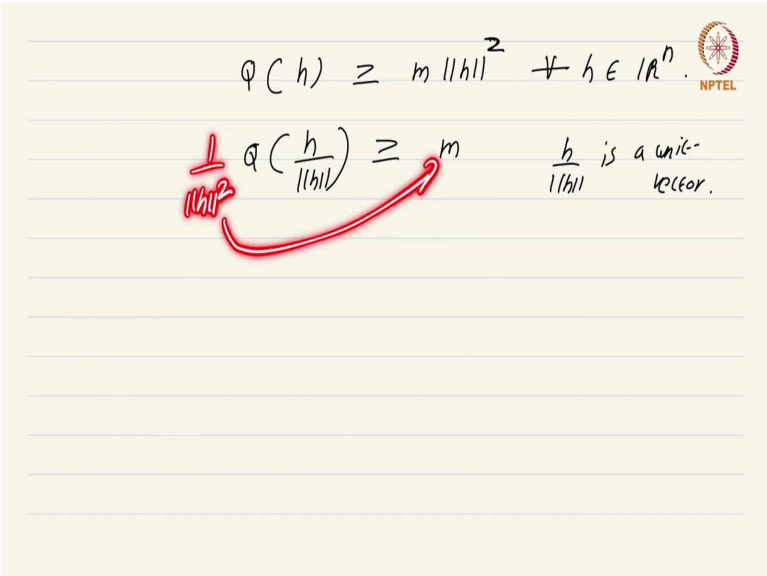
The unit sphere is a closed and bounded set. The unit sphere is a set of all points in R n that are precisely distance one away from the origin. This unit sphere is of course, a closed and bounded set. Therefore, by the Heine-Borel theorem for R n we know that it is going to be a compact set and we have also seen that continuous functions on compact sets must attain their maxima and minima.

So, it must attain a minimum value and this minimum value and this minimum value this is the key this minimum is strictly greater than 0. Because it does not take the value 0 at any point on the sphere, this minimum must be strictly greater than 0. So, I am just going to call it say m little m ok.

(Refer Slide Time: 20:03)



(Refer Slide Time: 20:10)

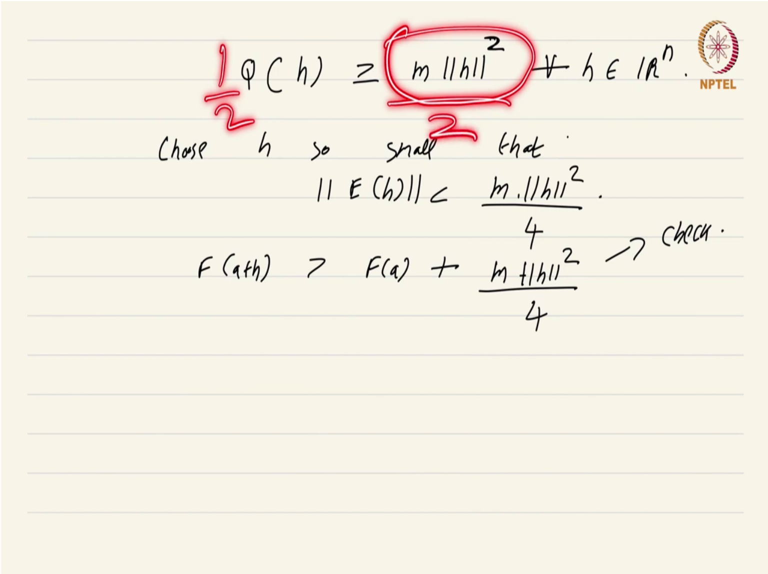


Now, what this essentially means is that Q of x is rather Q of h Q of any vector h this is going to be greater than or equal to m norm h for all h in R n ok. Why is this the case? Well, actually this is going to be norm squared m norm h squared. Why is this the case? Because what will happen is if you know that Q of h by norm h Q of h by norm h this will be greater than or equal to m simply because h by norm h is a unit vector is a unit vector.

Now, if Q were linear I can take this norm h 1 by norm h outside, but Q is in fact, quadratic, it is a quadratic form and just by the definition of Q you can immediately see that this norm h two of those norm h s will club together and produce a 1 by norm h squared outside which I have just taken to the other side ok.

So, this is a trivial proof. Just look at the definition of Q its just h transpose Hessian times h. When you replace h by h by norm h there will be a 1 by norm h squared sticking outside just shove it to the other side ok. So, we have that Q of h by norm h or rather let me just erase this. We have Q of h is greater than m times norm h squared and this is true for all h in R n ok.

(Refer Slide Time: 21:59)

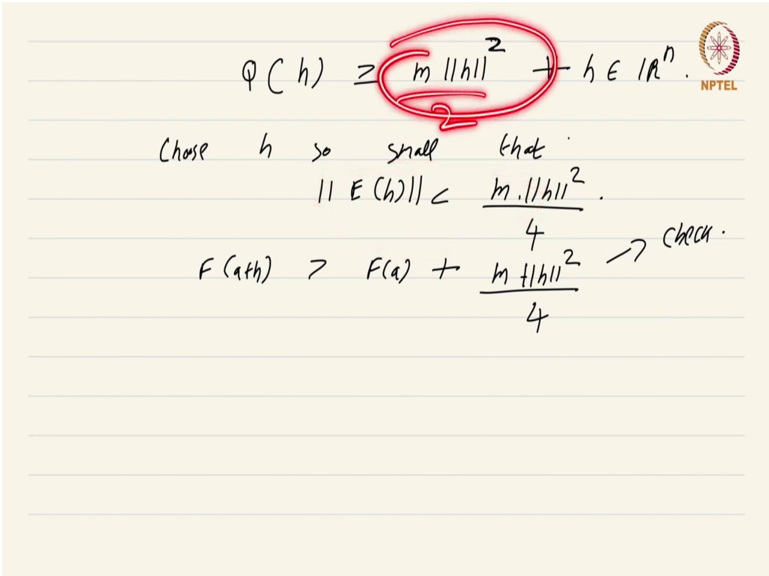


Now, what you do is choose h so small that norm of E h is less than m norm h squared by 4 ok. This can be done again because E of h by norm h squared goes to 0. Therefore, norm E of h by norm h squared also goes to 0 as h goes to 0, the error term is sub linear ok.

So, the net upshot is F of a plus h will be greater than F of a plus m by 4 by norm h squared ok. Check this check this. Wait a second, I made a slight error here. Somehow the norm h squared came to the denominator, this is m norm h squared by 4.

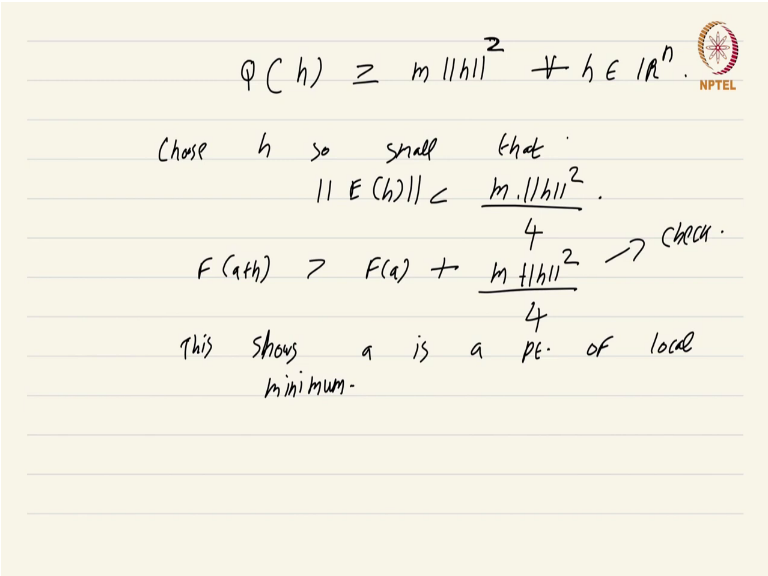
So, I want you to check this, check how we got this. It is rather easy. I have just analyzed the fact that this Q of h is going to be at least m norm h squared. There is a half Q of h. So, there will be a by 2 here and I am just going to combine this E of h term with this m norm h squared by 2 term.

(Refer Slide Time: 23:27)



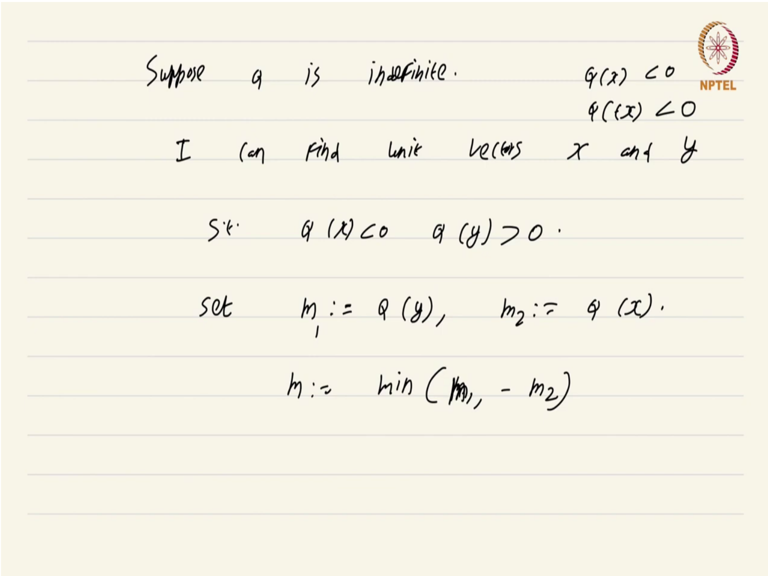
And see that it is got to be at least m norm h squared by 4 ok.

(Refer Slide Time: 23:35)



So, this shows a is a point of local minimum ok. So, I am going to leave the proof for you when the Hessian is negative definite. I conclude that a is a point of local maximum in the exact same way.

(Refer Slide Time: 24:02)

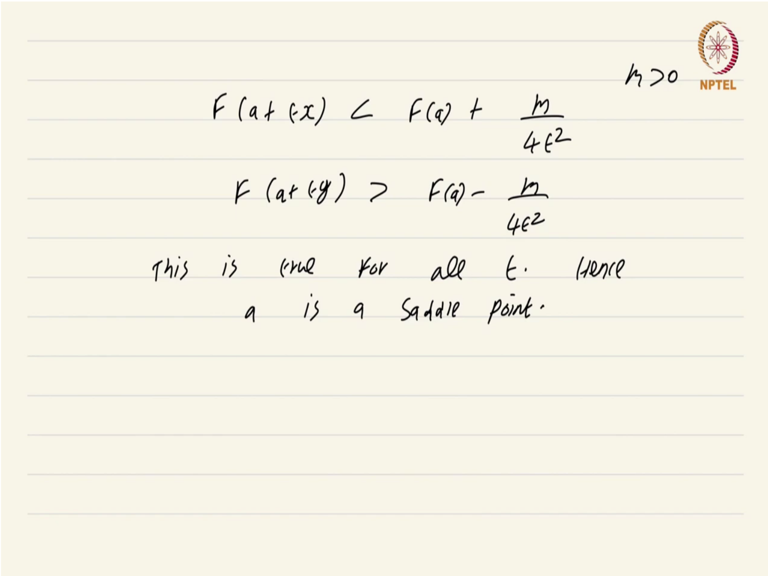


Now, one case remains, suppose Q is indefinite suppose Q is indefinite. What this means is that it takes positive and negative values on R n, but because Q is a quadratic form you can scale a vector x where Q takes a negative value. So, if Q of x is less than 0 you can easily see that Q of t x will also be less than 0 where t is any real number. So, this just means that there are vectors of any given length at which Q is positive and at which Q is negative as well.

The moment there are two vectors x and y where at one point Q of x is positive at the other point Q of y is negative, then you I can just scale these vectors around and continue to get vectors of any length that I desire at which Q takes positive values or Q takes negative values. So, what this means is I can find I can find unit vectors x and y x and y such that Q of x is less than 0, Q of y is greater than 0 ok.

Now, what you do is what you do is set m to be the yeah just set m 1 to be Q of y and m 2 to be Q of x ok and set m to be set m to be the minimum of these two values; not just the minimums m 1 comma minus m 2. Set it to be the minimum of m 1 and minus m 2 ok. Now, we use the observation that I had said before that when you scale the vectors by t when you scale x and when you scale y you can predict the value of Q of t x and Q of t y.

(Refer Slide Time: 26:28)



From that a very simple computation will tell you that F of a plus p x is greater than F of a plus m of m divided by 4 t square ok. And similarly you can conclude that F of a plus t y I think I got the inequality reversed, this is less than and I can conclude that F of a plus t y would be greater than F of a minus m by 4 t squared. And note m is positive by the way I have defined m ok. And this is true this is true for all t, this is true for all t. Hence a is a saddle point ok.

So, this concludes the theorem. The proof is not hard, it is just involves simple inequalities once you know the behavior of the Hessian. Now, you might be wondering what happens if instead of positive definite I have positive def in semi definite; that means, Q of x could be 0 at points other than x, but Q of x is always greater than or equal to 0.

What happens in that situation? What happens it is if it is positive negative semi definite, Q of x is less than or equal to 0, but at some point it is in fact, 0 and x is not 0, what happens to these cases? Well, you cannot really predict what happens and there is an exercise in the notes where I give you three examples three example functions and ask you to work out what happens to these functions at the critical point ok.

So, this is essentially determining the nature of a critical point using the Hessian. Still the question will arise if I am given this Hessian matrix how do I know whether it is positive definite or negative definite. Well, one criteria, if you have taken a sophisticated course in linear algebra not just a basic first course you might have heard the word spectral theorem.

If you have learned the spectral theorem, this will be rather easy to show; you can show that this quadratic form will be positive definite if and only if all the eigen values are greater than 0. And negative definite if and only if all the eigen values are less than 0 and indefinite if and only if some eigen value is greater than 0 and some eigen value is less than 0.

So, the spectral theorem will immediately tell you how and determining the eigen values is not that hard, it just involves some computation at least in the 2 by 2 or the 3 by 3 case which is mostly what we are interested in.

In those cases it is rather easy, but for sophisticated applications like the thing that Google uses or NASA uses you will typically not be dealing just with 2 by 2 or 3 by 3. So, there finding the eigen values is not so easy.

There is another elementary way of finding out whether a particular quadratic form is positive definite or negative definite, that is by what is known as completing the squares that is an elementary way of seeing whether a quadratic form is positive definite or negative definite.

We will see that in the next video. This is a course on Real Analysis and you have just watch the video on criteria for extrema using the Hessian.