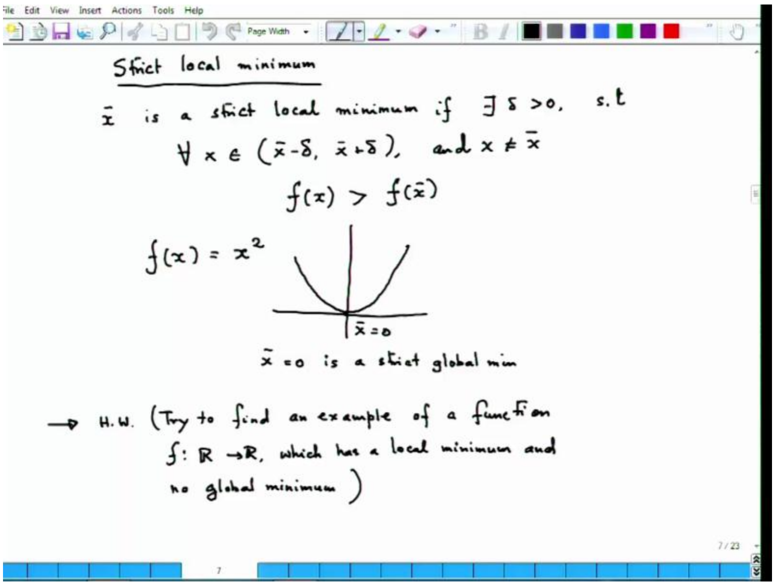
Foundation of Optimization
Prof. Dr. Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture - 2

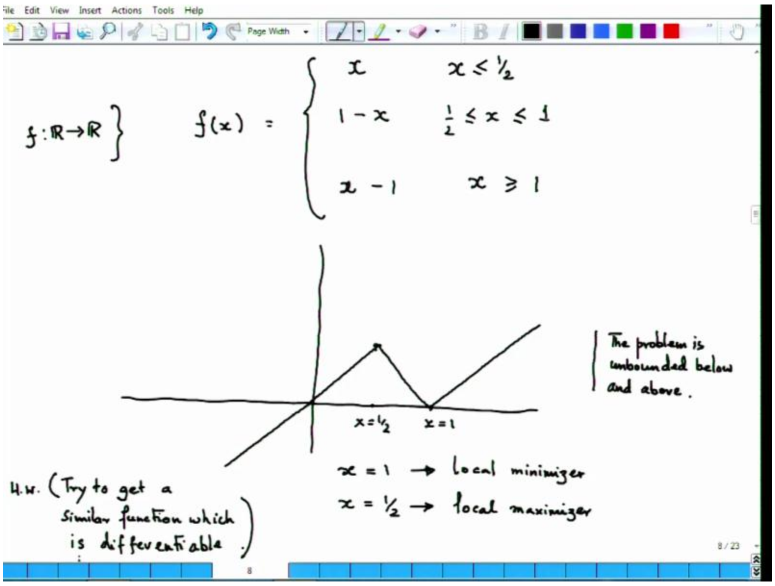
Good evening everybody; maybe I should say good morning or good afternoon because I really not know what time you are really seeing this video, but since I am shooting it at the evening hours, so good evening from my side.

(Refer Slide Time: 00:46)



So, we ended yesterday's discussions with a small homework which says try to find an example of a function from R to R which has a local minimum and no global minimum. In fact, I will just give you an example of a function.

(Refer Slide Time: 00:55)

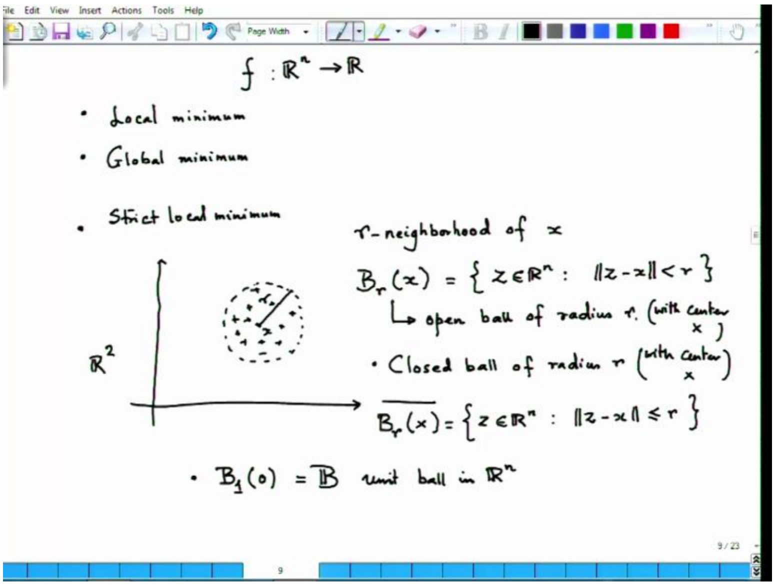


So, let me try to write down a function. So, this is the function that I have written down. So, f is a function from R to R and that is given as f x is x when x is less than half, f x is 1 minus x when x is between half and 1 and it is equal to x minus 1 when it is greater than equal to 1. So, if I draw the graph of this function it would look like this. So, here I draw the graph up to x equal to half.

So x equal to half, y is also equal to half because of both of these. So, this is x equal to half, then I come down to x equal to 1, then I go up at an angle of 45 degrees. So, this is my graph. Now you observe that it would be nice to write as x equal to 1. So x equal to 1, x equal to half, both have interesting characters; x equal to 1 is a local minimizer, x equal to half is a local maximizer.

But if you look at this problem this continues to go up and this continues to go down. So, the problem is unbounded both below and above. So, the problem in general is unbounded below and above. So, this problem is unbounded below and above. So, this is the example that of course you can say, oh, this is not differentiable at this point and this point; it is quite right but try to figure out such an example where the function is differentiable. So, I start to talk today with this homework for you. Try to get a similar example with a differentiable function. So, in this function for example only local optimization matters, global optimization does not matter which is differentiable. Now we move on to a bit more sophistication.

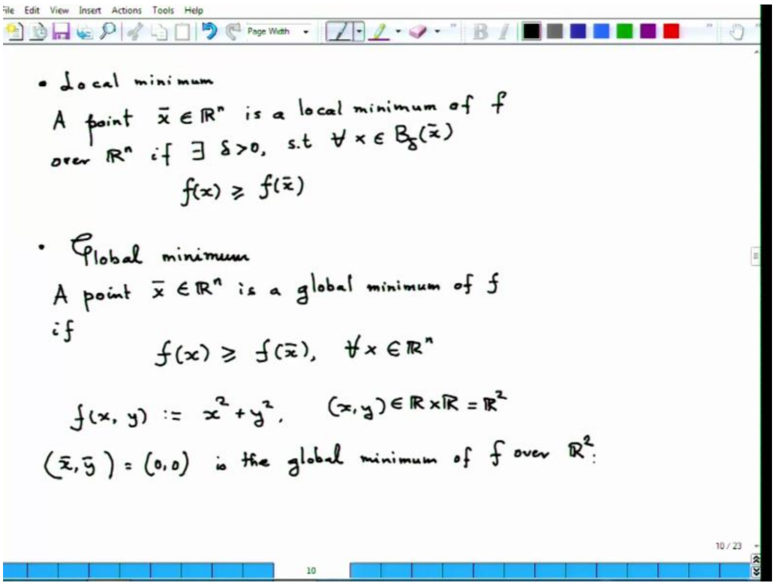
(Refer Slide Time: 04:57)



So, our aim would now be to consider functions from R n to R and discuss meaning of the three things, local minimum, global minimum and strict local minimum. The definitions are all similar but there is a crux here. It is that what we have to now do is to define what is the meaning of a neighborhood? So, if you say let us our prototype model for anything of higher dimension is a two dimensional plane; in the plane R 2 take any point. So, neighborhood is nothing but you draw a circular disk of radius r around it, but do not consider any point on the circle. So, I have kept it as a dotted line; only consider points inside, this is my x. So, R neighborhood of x is nothing but the open ball around x which is given as the set of all x such that norm of set of all z in this case x is fixed, set of all z in R n in our case. So, let me be more specific that I am in R n.

So, whatever I am writing it for writing for R 2 can be easily changed to R n by replacing two with n; norm of as the distance between z and x is strictly less then r. Of course, R neighborhood is nothing but open, this open ball of radius r. Of course, you can similarly draw on certaintion to the closed ball of radius r. So, this is usually symbolized as radius r with center x in this case. This is nothing but all z in R n such that norm of z minus x is less than or equal to r; you did not have the equality there. Now once I know about these things, I would go to the notion definitions but just this symbol B 1 0 when x is 0 is usually denoted as this unit ball in R n. So, this is an open disk or open ball of radius 1 and centre at 0. So, it is usually denoted as this and it is called the unit ball in R n and now we will go to our definition.

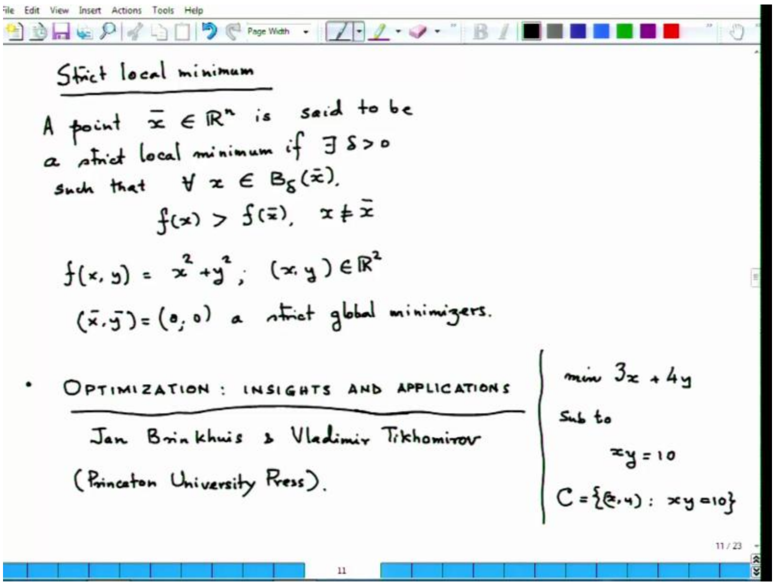
(Refer Slide Time: 09:31)



So, first we will define a local minimum. So, we are doing it for the unconstraint case. Now how do I do a local minimum; the same definition just like the constraint case same definition, but here your neighborhood gets changed with the neighborhood that we have just defined your delta neighborhood or that open interval in case of R which was this one gets changed to this one. So, a local minimum definition is as follows. A point x bar element of R n is a local minimum of f over R n, if there exist delta greater than 0 such that for all x in B delta x bar f of x is bigger than f of x bar; global minimum means of course it is for all x in R.

So, again I will give the definition of it. So, what do you mean by global minimum? A point x bar in R n is a global minimum of f if f of x is bigger than equal to f of x bar for all x in R n. So, consider this function f of x, y is equal to x square plus y square where x and y is element of R cross R, both are real numbers, is same as R 2. In this case if you see that this function value is always greater than equal to 0 and if I put x is equal to 0 and y equal to 0, that 0 value is achieved. So, x bar y bar is equal to 0, 0 is the global minimum of f over R 2. So, now we will talk about the strict local minimum.

(Refer Slide Time: 13:02)



Now, the strict local minimum has the same approach as yesterday. A point x bar element of R n is said to be a strict local minimum if there exist delta greater than 0 such that for all x element of B delta of x bar f of x is strictly bigger than f of x bar whenever x is in here but x is not equal to x bar; that is enough to write this. So, this is the crucial fact that x is not equal to x bar. So, if you look at this function again, this one, one we just took, then x, y is equal to x bar y bar is equal to 0, 0 is actually a strict global minimizer. So, we have some idea now in higher dimension how to write the things. So, as we said that this is a largely story telling. So, because this is story telling I really have to lighten up the atmosphere; one important book that one might read which will make him or convert into an optimizer, you would be attracted towards this field is the following book.

The name of the book is Optimization Insights and Applications by Jan Brinkhuis and Vladimir Tikhomirov published by the Princeton University Press in the Princeton citizen applied mathematics, but I do not know whether there is an Indian edition of this book. So, it is a good idea; however, to write this down in detail. Of course, I told you about another book of Vladimir Tikhomirov; I am a great fan of his. Vladimir Tikhomirov is an extremely deep person and a great mathematician. So, what I would like to look into is to write down the name of this book. So, it is Optimization Insights and Applications. This book by Jan Brinkhuis he is from Netherlands and Vladimir Tikhomirov is published by the Princeton University Press. Once this is done, may be let us look at and take a little inside into this book; may be I will do some reading from this book, you might like this idea.

What you really need in optimization is to able to compute the minimizer. So, if you want to compute the minimizer you must know how to compute the minimizer. So, there must be some equations or something from which you can compute a candidate for a minimizer. So, to do that you need to do you need to take into account the development of necessary conditions.

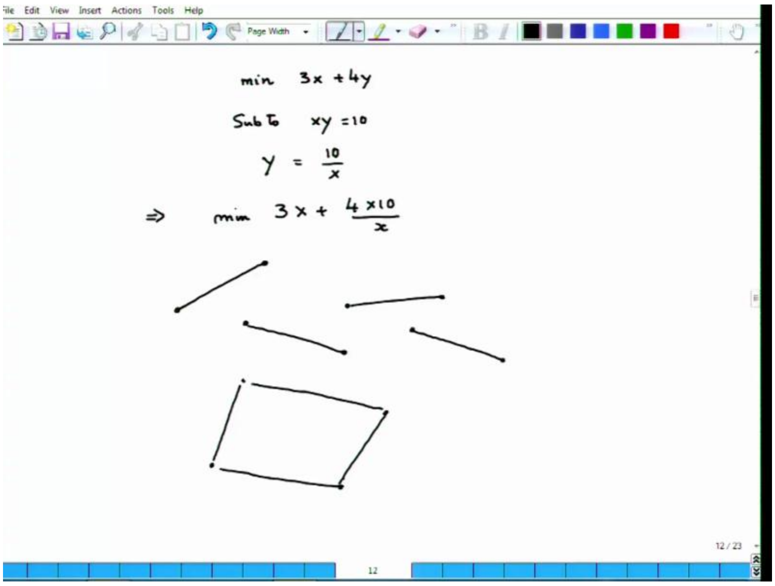
So, if you talk about necessary optimality conditions. So, using necessary optimality conditions you can compute out a point for which could be considered as a minimizing point and then really test whether it is a minimizer or not. So, I am reading out from the introduction of this book, just which I have mentioned, optimization insights and applications; thing is necessary conditions what is the point. So, let us hear from the experts as Lagrange as Laplace has once told, "Go to Euler, go to Euler, he is the master of us all." So, you really have to learn from the masters, that is what we are trying to do.

Six reasons to optimize, suppose you have just read for the first time about a new optimization methods say the Lagrange multiplier method or the method of putting the first derivative equal to 0. So, which you already know from school, at first sight it might not make a great impression on you. So, why not take an example, find positive numbers x, y with product 10 for which 3 x plus 4 y is as small as possible. So, here you see he is asking you to solve this problem; minimize 3 x plus 4 y subject to x, y is equal to 10. So here my set C, the set c is set of all x, y such that x into y is equal to 10. See this is consisting of all the points 10 1 5 2 2 5 at the product is 10 and many, many other points 1 by 10 into 1 by 10 100 100 1 by 10. So, there are many such points. You try your luck with multiplier rule and after a minor struggle and may be a after the unsuccessful attempts you succeed in finding solution. This is how the multiplier rule comes to life; you have conquered a new method.

You see a very simple looking a problem like this is actually a very hard problem in optimization because of these restrictions; these constrains are not very hard known liner constraints while this is a linear function, you are minimizing a liner objective over a hard non-linear constraints. So, this is a very interesting example of a constraint problem which is so simple, but it is very hard but which looks so simple when you look at it; but it is actually very hard to solve. After a few new numerical examples you start to lose interest in these exercises; of course, if you just keep on doing applying the Lagrange multiplier rule or whatever optimality conditions we proposed and you would really lose interest. You want more than applying the method to the problem that comes from nowhere and leading to nowhere.

Then it is time for a puzzle and here is one; that is if you are getting bored with problem we should tell you a puzzle. To save the world from destruction agent 0 0 7 has to reach a skiff 50 meters offshore from a point 100 meters further along a straight beach and then disarm a timing device. The agent can run along the shore at 5 meters per second, swim at 2 meters per second and disarm a timing device in 30 seconds. Can 0 0 7 save the world if the device is set to trigger destruction in 73 seconds. The satisfaction of solving such a puzzle by putting derivative equal to 0 in this case is already much greater. So, this is also an optimization problem that how what is the minimum time in which he can reach the device.

(Refer Slide Time: 23:04)

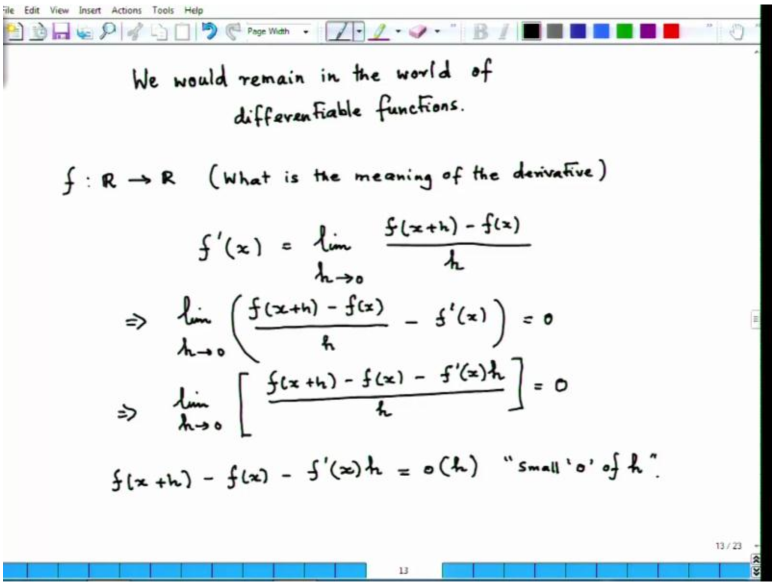


So, but you can say as he shows here in this book what Brinkhuis and Tikhomirov shows that even this problem is actually quite difficult. What I can do is why do not I write y from here as 10 by x and then I write it down as minimize 3 x plus 4 into 10 by x. So, it is just a problem in x, the problem in single variable and we take derivative equal to 0 and whatever and we try to solve it. And once you know that optimal x you will know the optimal y. The question is that do we need to really use the multiplier method for this sort of problems; it is so simple. So, does the multiplier rule have the right to exist; that is what he asks in this book. He says that it is the ancient problems of geometry which really had shown the strength of multiplier method, I think which we will also discuss; we will discuss a couple of examples from this book.

So, time for the test of strength of this method. How can we put down four sticks to form a quadrangle of maximum area; that is you again I told you we will come with a geometrical problem. We have four sticks given to you which are of fixed lengths. Now these four sticks from a quadrilateral and this quadrilateral should enclose the maximum area and the sticks need not have the same size. So, then the answer is square if the sticks of the same size. So, the sticks can have different sizes. So, what is in which way I should arrange the four sticks so that they will have maximal area enclosing; they would enclose the maximum area. As Vladimir Tikhomirov and Brinkhuis write that this is not just a run of the main problem.

In ancient times the Greek geometers knew this problem already. They tried very hard to solve it using geometrical methods; yesterday we actually used an algebraic and geometrical method, arithmetic method and geometrical method to solve such a maximization-minimization problem; we will do some more of them as we come along. They tried very hard to solve it but without success, but with the multiplier rule with the Lagrange multiplier rule which will be one of the most interesting topics in our discussion, we will solve it without any difficulty; as he writes here we do not know any where to solve it. So, this is very, very important step. So, here with this very basic introduction we start our further discussions. So, what we need to discuss is that we would be in the world of differentiable functions.

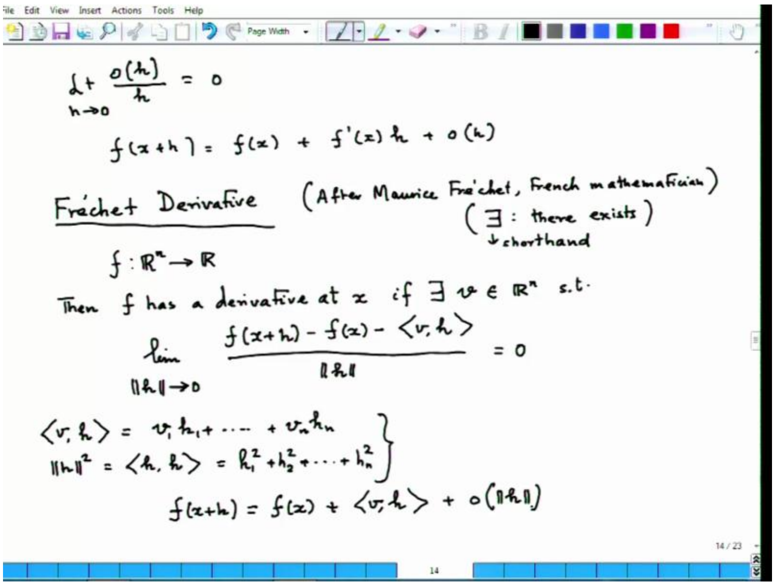
(Refer Slide Time: 26:22)



So, you can first ask me what are differentiable function; I assume all of you know it, but once you are on a higher dimensional from function from R n to R differentiability might not be such a straightforward concept as you see in function from R to R. But we will take a clue from the functions from R to R and then try to develop the notion of differentiability of a function from R n to R. So, if I have a function from R to R. So, what is the meaning of the derivative? So, a derivative is defined, as we have also known in high school, at a given point x take the limit as h tends to 0 means h is moving from left and right both ways; what more information can I have from here? The information that I can have from here is as follows.

See f dash x does not really depend on h, this is free of h. So, I can write the whole thing as. So, I hope you all agree with this writing. So, once you do that this would actually come down to doing some computations here; that is pretty interesting. So, what you have there is a quantity because x is fixed, this is the top quantity here is the function of h. So, I have a function of h when I am dividing that by h and as I am tending the limit as h and I am looking at the limit as h goes to 0, I am getting 0. Then this quantity on the top is a quantity which is called the small o of h of the small order of h. So, this quantity on the top is defined as follows, small o of h.

(Refer Slide Time: 30:25)

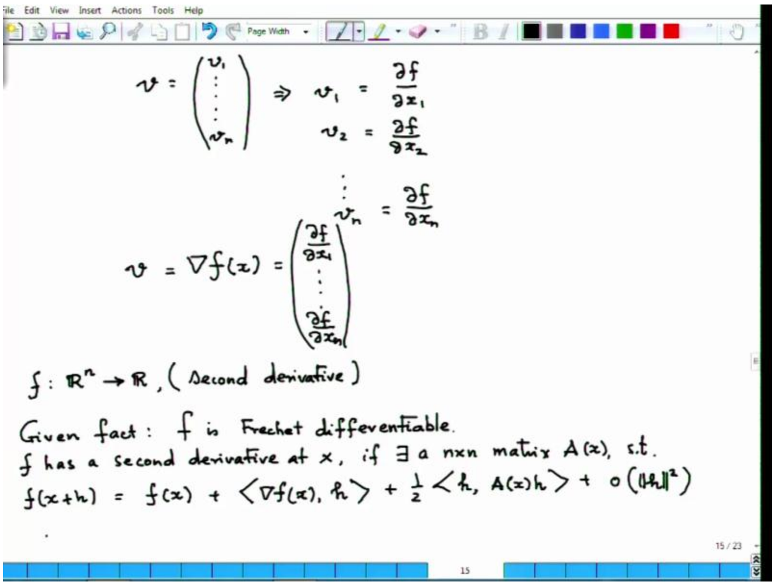


Now this actually means small o of any quantity is called small o of h if this is true. So, of course you can write the derivative; the derivative has also leads to the following expression which is called the first order Taylor's theorem. Basically if I take out this, then this is nothing but a linear approximation of the function value near x. But actually there is an error; if the function is linear there will be no error, otherwise the function is actually linear there is an error. So now, how to define it for higher dimensions that is the whole question. To do that definition, we can generalize the whole idea. Here you see I have multiplied the derivative into h the increment and then there is this error.

But there when I go to higher dimensions I am moving in the space R n my vector h the increment is a vector; it is no longer a scalar and then multiplication has got to be changed with inner product. So, this definition was originally given by a Frachet and one of the most useful definition of derivative, the French mathematician Frachet, which is called the Frachet derivative, named after Maurice Frachet, French mathematician. So, what he says that if you have a function from R n to R then f has a derivative at x if there exists this is a sign of their exist. So, if you those who are getting confused this is a shorthand of the word there exists; if there exist v element of R n such that limit, now I cannot divide by h it is a vector, but I can dived by norm of h which is the length of h.

Here I cannot write the same thing because now my norm the multiplication is changed in to the inner product divided by the norm of h that should be equal to 0. In fact those who are uncomfortable with idea of inner products, just to recall because this is a slightly advance level course inner product of two vectors in R n. So, there will have n components is nothing but this one and the norm of h square is nothing but the inner product of h with itself which is nothing but the Euclidean distance from 0, square of the Euclidean distance from 0, where h 1, h 2, h n are the components of h. So, this is what is the meaning of the derivative which also means that f of x plus h is equal to f of x plus v of h plus o of norm of h.

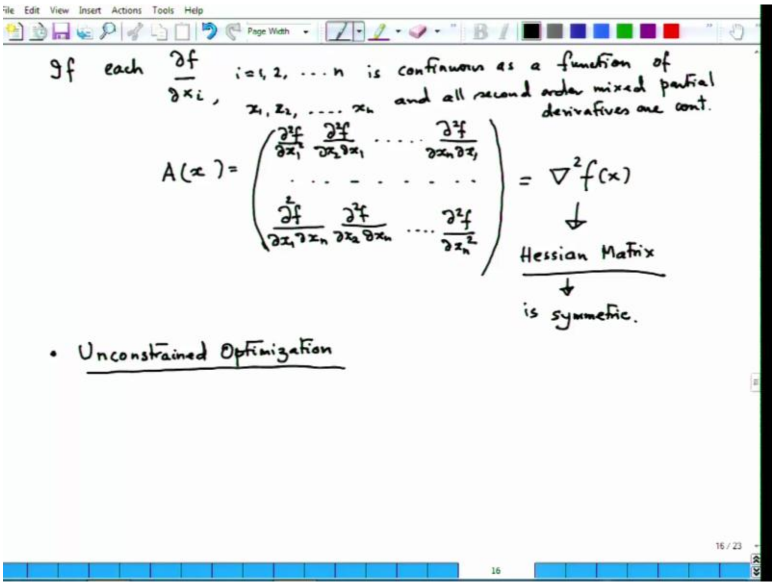
(Refer Slide Time: 35:26)



Now if you do some simple calculations, then you can show that if v is a vector, v 1, v 2, v n, then v 1 it would imply that v 1 is equal to del f del x 1, v 2 is equal to del f del x 2 dot dot dot v n is equal to is a partial derivative of f with respect to these variables. So, v is then symbolized as the gradient of f at x and which is nothing but del f x 1 dot dot dot del f del x n. Similarly you can define the second derivative. The second derivative would also be of big help when you do second order conditions. So, it is not only the necessary conditions which is of help; sometimes you have to go to what are called sufficient condition that if I figure out a point, and now it is a one of the candidates by which could form minimum.

I really have to show that it is a minimum and one of the simplest paths to do it is to check whether this point satisfies certain conditions which are called sufficient condition and which are usually involving the second derivative. So, it is important for us in the case from function from R n to R decide what is the second derivative. So, second derivative means f is already having the first. So, I am given fact is f is Frachet differentiable. So, once I know this, how do I define the second derivative? f has a second derivative at x if there exist n cross n matrix A x such that f of x plus h is equal to f of x plus grad f of x of h plus half h A x h plus small o of norm h square this time. Now, after certain calculations if you assume that all these derivatives are continuous.

(Refer Slide Time: 39:10)



So, if each del f del x i is continuous as a function of x 1, x 2, x n, then A of x this matrix can be written as. So, you are taking the derivative of the gradient matrix first with respect to x 1 then del 2 f del x 2 del x 1 del 2 f del x n del x 1, del 2 f del x n del x 1. Of course, I am taking this continuous so that you know this and this are equal, then the matrix is symmetric and nice and can have much better properties which is useful for optimization; del x n del x; sorry del x, I am making a mistake here; I will just write it clearly. Now I have last line as del f x and del f x n is now I am going to take its gradient of del f x n; del f x n is a function of x 1 to x n and I am taking its gradient. So, del 2 this symbolizes is del 2 f del x 1 del x n del 2 f del x 2 del x n.

So, this and this are equal; if all these are all the first del f x these are continuous of the function of x next to x n and all mixed partial derivatives are also continuous and all second order mixed partial derivatives are continuous, then you have this equal this Young's theorem. So, that is interesting actually; I am not writing what is in the middle, the trace is the Laplacian. So, this usually symbolizes the Hessian matrix of f at x and this is very, very important to determine whether point is in minimum or not. So, with all these assumptions here the Hessian matrix is symmetric.

So, now we have done enough material at hand to start our progress step by step. So now, what we will start with is unconstrained optimization; that will be our goal tomorrow. Thank you very much. We will start with this topic; you must state, oh, what about the c, what about this constraints. We will not come to constrained optimization immediately; we will talk more in detail about constraints when we start constrained optimization. The goal is to talk about the basic theory of unconstrained optimization, how do you find locate a minimize, how do you compute it, how do you know it is a minimizer use of this and then can you develop algorithms to really find out an unconstraint minimize; this is the goal. So, we will keep in mind these little things before we make further progress.

Thank you very much.