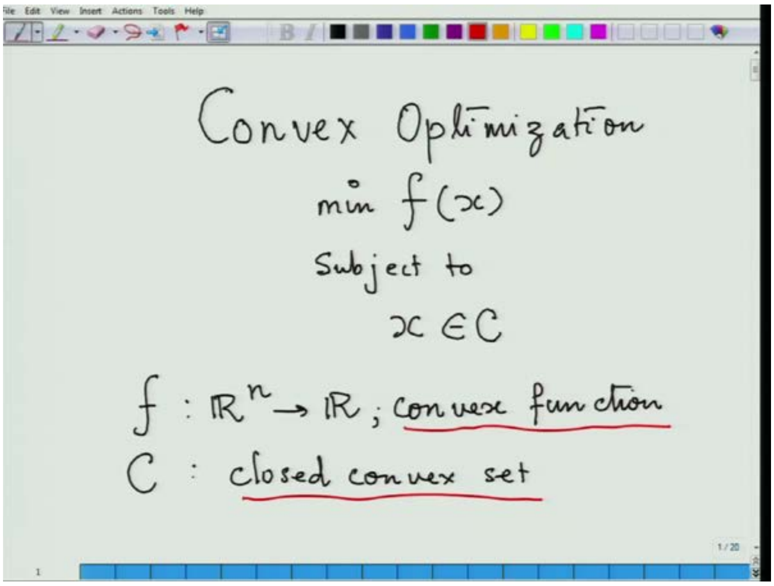
Convex Optimization
Prof. Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

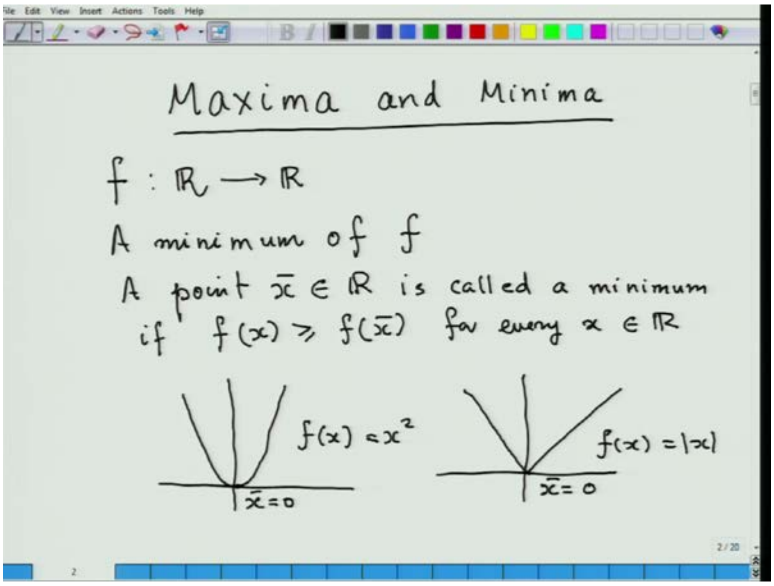
Module No. # 01
Lecture No. # 01

(Refer Slide Time: 00:22)



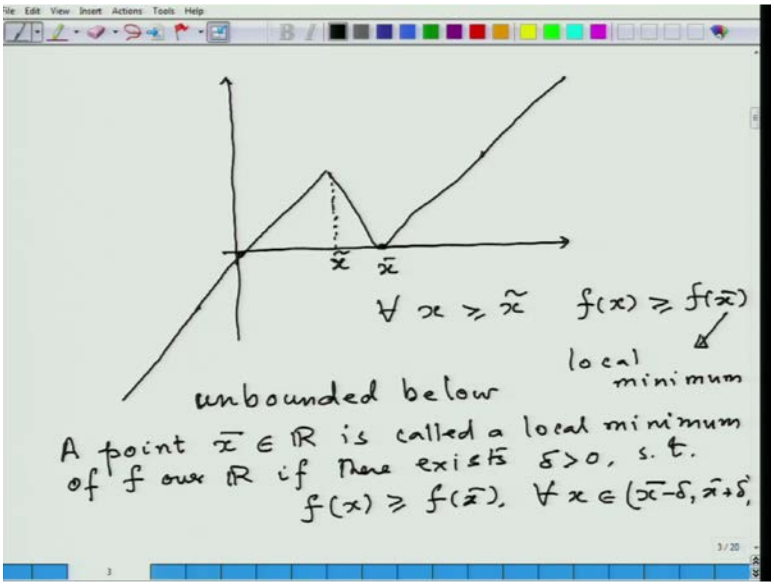
Today we are going to start this exciting journey on a subject call Convex Optimization. Convex Optimization involves, the Minimization of a convex function subject to a convex set C. Now, of course this is a very special class of optimization problems. This class is special and why it is important to applications, and why good algorithms can be constructed about it, that would be essentially the matter of this course. So, we would now like to refresh our knowledge about Maxima and Minima.

(Refer Slide Time: 01:04)



So, in doing so we will go back to high school once again, and we will concentrate on a function say from R to R, and let us see what we have learnt in high school and let us see what we mean by minimum (()). Of course, this minimum minimum is clearly understood, because A point x bar of R is called a minimum, if f of x is bigger than f of x bar for every x in R. So, you see here, let us draw a diagram to show you one such example. So, this function f x is a parabola, x square where you can see at x equal to 0, the function reaches a minimum, this function f x is the absolute value of x, and at this point x bar equal to 0, this also achieves the minimum. Both of these two functions, though we have not defined convex functions yet are also convex functions.

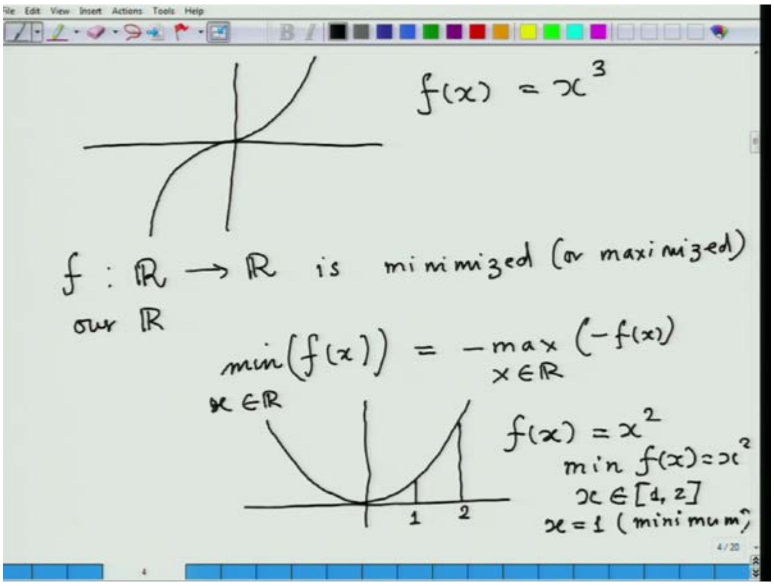
(Refer Slide Time: 03:37)



But, this is not the only part of the story, because one has to remember that I can give you a function, where this might not always be true. For example, let me look into this function, this function is nice continuous, but here I pull it down. So, this function becomes unbounded below. So, there is no lower bound. So, we cannot really speak about a global minimum or the minimum the as the way we had defined earlier. So, the question now arises, can anything be extracted from this picture, if you look at this picture carefully, look at this point x bar, if I want to call it x bar, look at this range.

So, for all x which is bigger than x tilde here, f of x is bigger than f of x bar, this x bar is not really the global minimum, but then it could be helpful in many other ways and we call such an x bar a local minimum, more clearly if I want to define a local minimum, then let us do it in a more eluded fashion, point x bar in R is called a local minimum. Local minimum of f over R, if there exists delta greater than 0, such that f of x is bigger in value then f of x bar, for every x, this is the mathematics symbol for all, the what for all is symbolize like this, for all x in the open interval x bar minus delta to x bar plus delta. For all x in this open interval excluding these 2 points, x bar minus delta and x bar plus delta.

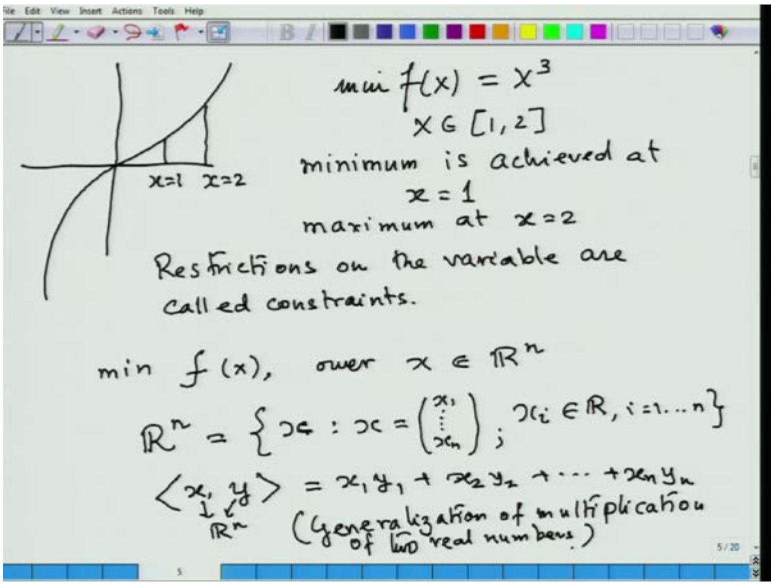
(Refer Slide Time: 07:03)



Now, this definition is there, and I would request you to look into this interesting thing, where I have a function which does not have either a local Minima or a local Maxima. It, is both unbounded below and above, this function f x is equal to x cube; this function does not have a local Minima, a local Maxima nothing, in fact nor global Minima nor global maxima. Of course you might be ask me I have been using the term local Maxima without defining it, but you can define local Maxima in the way I have defined local Minima or and global Maxima, in the way I have defined global maxima, but there are certain function which can be both unbounded below unbounded above and does not have any such things like local or global maxima. Once we have said this, let us observe on little fact that f from R to R that we have defined, is minimized or maximized over R.

Can these maximization and minimization be done if we restrict this function over a subset. Answer is very simple, but before I go into anything else, let me tell you in the whole course we would be essentially talking about minimization of a function, because minimizing a function f x over x belonging to R is same as, taking the maximum of the negative and then take a Maxima again, to take a Maxima over the negative and then take a negative again to get the minimum. So, it is enough to speak about the minimum, once you talk about the minimum you can also talk about maximum. Now, let me tell you something interesting in the sense that, let us look back again at this function, f x equal to x square. So, if I talk about this class of functions f x equal to x square, then 0 was the minimum over all over, but now let me say, let me pose you the problem of minimizing this function f x equal to x square over x restricted say to with the interval 1 2. So, basically now you can easily see the minimum value here is achieved at x equal to 1, actually is a global minimum.

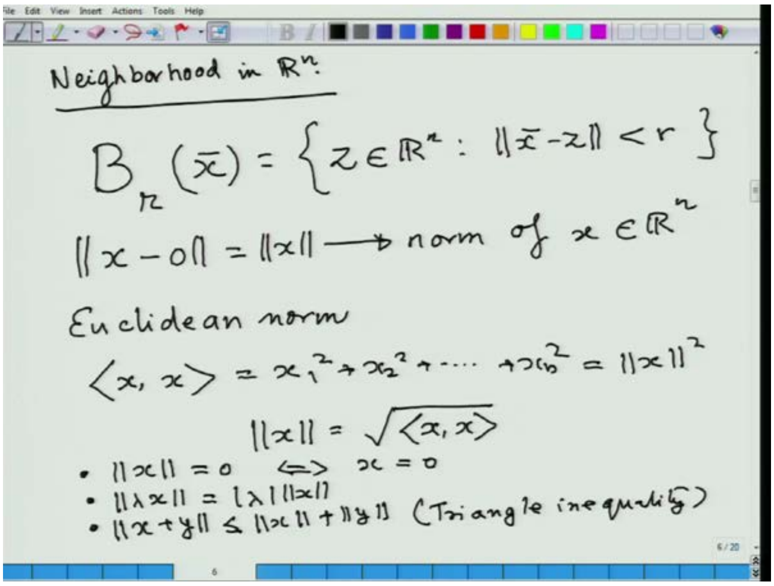
(Refer Slide Time: 10:42)



Now let us do a thing, let us go back look at again at this function f x equal to x cube, and if I restrict, if I call for minimizing f x equal to x cube, over the same interval 1 2. You will immediately see that the minimum is achieved. In fact, the global minimum is achieved at x equal to one. So, it clearly shows that the same problem which did not have a global Maxima, global Minima, local Maxima and local Minima has a global minimum at x equal to 1, and has a global maximum at x equal to 2. So, the whole problem paradigm changes once you add certain restrictions on the variables, such restrictions are called constrains, restrictions on the variable are called constrains. So, with this very basic idea, we will now expand our dimensions in the sense that, we will now look at the problem of minimizing a function f over x in R n. Of course, you know R n, the n dimensional Euclidean's phase is a collection of vectors x, where x is represented through a n duple of n numbers, or other coordinates of n numbers where each of this x i belongs to real numbers set, there is each of the x are real numbers.

R n as you know is an important phase and is equipped with inner product. So, it takes any x in R n, and any y in R n. This inner product is defined as follows, that is you take the first component of x multiply it with the first component of y, and then take the second component of y multiply them, add them together and keep on doing so. So, this inner product is basically a generalization of multiplication of two real numbers. We must now go and try to define what is the meaning of a local Minima, you know what how to define a global Minima, need not stress this issue of global Minima, we need to define a local Minima for a function from R n to R, but before doing so we need to speak about the notion of Neighborhood in R n.

(Refer Slide Time: 14:53)

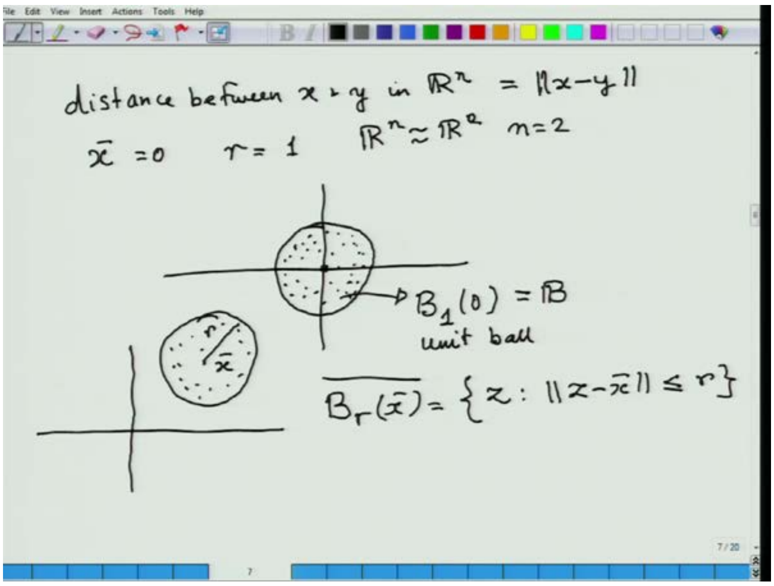


I understand that this course for this lecture are been followed by a wide category of people rather, it is not that it is just been observed by mathematicians or people interested in mathematical optimization. I believe that it is also seen my engineers and engineering students and people working in operation research, people working in industrial engineering. So, the idea here is not to be to mathematically correct, to say a neighborhood in R n we can say mathematically as follows that if I take a point x bar in R n, then any open set which contents x bar can is defined as neighborhood in R n.

But instead of doing so we will be slightly not so rigorous, but for us largely neighborhood would mean the following ball. A ball of radius r centered at x bar, it's consists of all z in R n, whose distance from x bar is strictly less than r. You might call suddenly I have given this signs and I am calling this as distance between x bar and z. Now, this distance, the distance of the point x from the origin of the phase R n is often written like this, and refer to as the norm of x, but this norm actually of point in R n actually means the distance of x from 0, so how the norm is defined. So, for our case what we defined would be called Euclidean norm or a two norm, in doing so how do we define Euclidean norm, and that is done again done through a inner product.

Suppose, I take the inner product of x with respect to x, so this would become by definition, this is simply a repeated application of the Pythagorean theorem and that would lead to the definition of norm x, because you see this is nothing, but the distance of x from 0 calculated by the Pythagorean theorem, or celebrated to result in Euclidean geometry which everybody seeing this course would know, because it is start in the school level. Now, the very definition of the norm, because it is a length we cannot put plus minus root over, we just have to put the positive sign. So, the norm as every criteria of a distance, because certain properties of norm I want to write down. So, norm of x is equal to 0, if and only if x is the 0 vector, if you scale up the vector x, whether you scaling it up through a negative number or positive number the answer is, where the length of the new vector is the absolute value of lambda times the norm of x. And then there is a famous rule which says that the triangle inequality, which is based on the fact that some of two sides of triangle are always greater than the third, this is called the triangle inequality, it would be quiet helpful. Now, once I have this idea of distance from 0, of course we can define the distance between 2 points x and z or x and y.

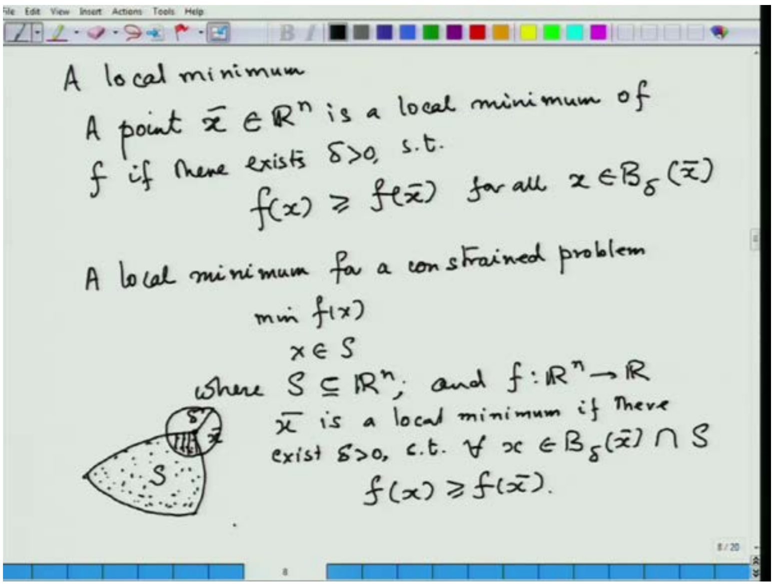
(Refer Slide Time: 19:36)



Distance, between x and y in R n is given as. So, once this is done, once you know this you can actually draw a ball. Now, let me take x bar equal to 0 and let me take R equal to 1, this is sometimes call as the unit ball. So, how would it look like, and of course take R n to be R 2, basically I am taking n equal to 2. So, if you look at this picture. So, you have 0 and anything which is within the distance one, but it should not be exactly one. So, it is look at this circular path the circle. So, points on the circle and not in the neighborhood, because the neighborhood has to be an open set. I am assuming that you have some idea about open and close sets. Any point here, this dotted part apart from the boundary, this dotted part is what is called, sometimes just denoted as this bold B.

So, this is called the unit ball. So, if you take any ball centered at any other x bar, it would look like this of any radius R. Of course, you can define the close ball which is define like this, which is the closure of the ball that we have defined, it will be set of all z such that z minus x bar, of course z is in R n less than or equal to what we means that we are now considering the boundary also, so then this whole thing would become this one. Now how do I go about defining a local minimum.

(Refer Slide Time: 22:19)



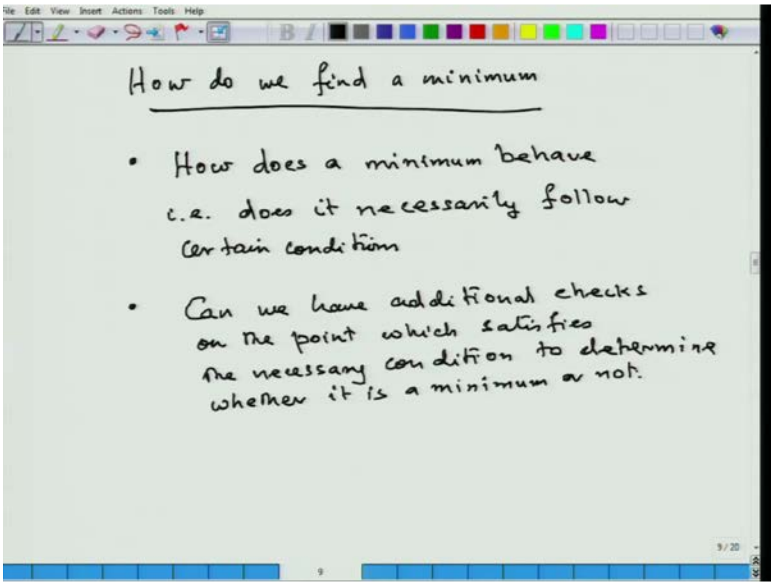
The definition of a local minimum go that is follows, a point, see we are now what is the neighborhood as such, neighborhood is nothing, but generalization of the notion of a interval, a open interval around a point, let like we have said x bar plus delta and x bar minus delta, is just the generalization of that in higher dimensional phase. A point x bar in R n is a local minimum of f, if there exists delta greater than 0, such that the functional value of x is bigger than the functional value at x bar for all x in the open ball centered at x of radius delta. Just remember when we write this, we always said this balls, this is a center of the ball. Now, once you know this, you could now come in write down the definition for a local Maxima. You can now try to write down the definition of a local minimum for a constrain problem that is now we put restrictions on x. So, you might ask me what is your convex optimization, whereas your convex functions etcetera all this things vanished.

See, convex optimization sub class of optimization of problems are very important sub class, but; however, it is imperative on us that we have a look at the basic facts about Maxima and Minima, because without understanding this very basic things it is not really possible to go get in to the depth of convex optimization. Good news is that when we do convex optimization we need not think about local Minima, we just have to think about global Minima. Because the importance of convexity or convex optimization from the fact that whenever we minimize the convex function over convex set every local Minima is a global Minima, there is nothing called a local Minima in convex optimization and that is why the subject is so powerful, so important, so important in applications. Now, we go for local minimum for a constrained problem. So, here we are looking at the problem of minimizing f x over x element of sum S, where S is a subset of R n which could be hollow R n, it could be a proper subset and f as before a function from R n to R.

So, you see these functions are always mapping an element in R n to R, because then we can actually do the comparison. Of course there is a wealth of literature, wealth of optimization which talks about minimization of functions when functions are defined from R n to say R m. So, that is the whole of a difference subject call vector optimization which we would not discuss. So, here what we do is, we have to observe that when we talk about local minimum, we need to restrict ourselves to the set S. Suppose, this is my set S, suppose this is not really the thing, because this is just an abstract discussion. So, interesting part of mathematics is that just by doing certain abstract discussions, you can do upon certain very important points immediately to the focus of the discussion. You see if this is my x bar and I want to say it is a local minimum; then obviously, you assured that for every x in the set S, f x bar need not be smaller than f x.

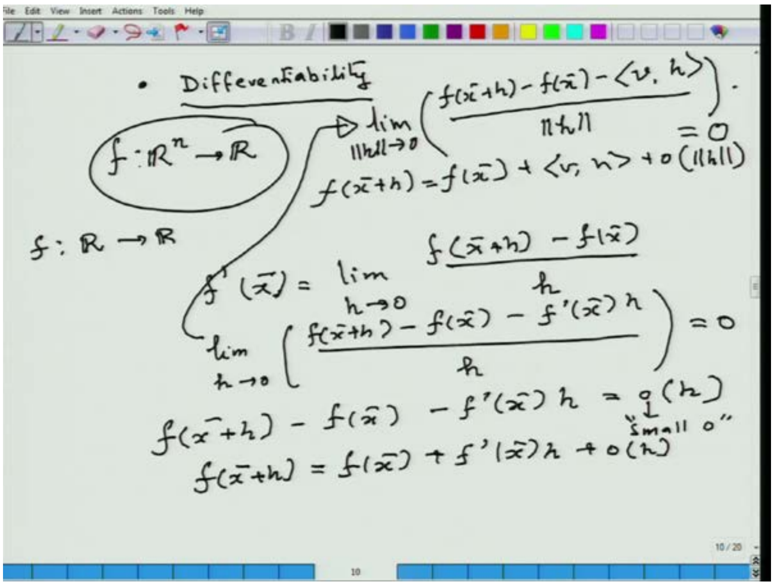
So, what we do is, to show that there would be a ball of radius delta; such that for this intersection of the ball with S that is on this part, for every x in this part function value at x here would be bigger than f x bar, that is x bar is a local minimum if there exist delta greater than 0; such that for all x in the ball surrounding x bar of radius delta, that is centered at x bar and of radius delta, S and also in c. So, all x which is in the in this particular set which is intersection of B delta x bar and c, f of x goes over x bar. So, you now take the responsibility of defining local Maxima and global Maxima, global Maxima is obviously, any of this definition over all x belonging to R n. So, once these basic facts are known, let us go ahead and also this question.

(Refer Slide Time: 28:41)



How do we find a minimum. So, in order to find a minimum we must first know How does a minimum behave, at if I have A local minimum, does it has have some property, if it has some property does it allow us to compute easily such a point which satisfies the property. Then the point that we get is it really a minimum. So, the few steps in finding a minimum the first thing that we have to know, how does a minimum behave; that is, does it necessarily follow certain conditions. When we compute such a point which satisfies such a condition, the necessary condition, we cannot just declare it to be a minimum unless we have certain additional checks. Can we have additional checks on the point which satisfies the necessary condition, to determine whether it is actually a minimum or not, these lead us to certain questions. The questions are just given enough it is very difficult to say what sort of a condition, this minimum, local minimum or even the global one follow. So, what we want to say is that, if you put certain additional conditions on f, something more you know about f, it might help, one such condition is Differentiability.

(Refer Slide Time: 31:38)



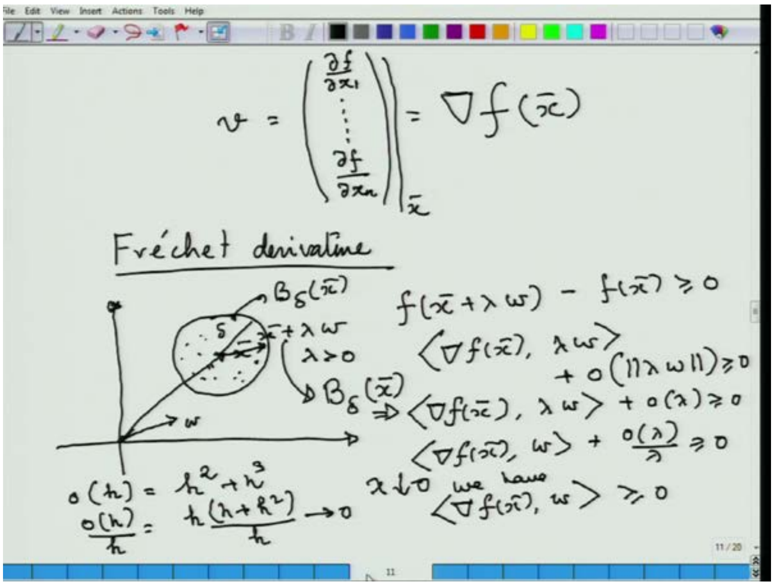
Now, when we are talking about Differentiability, it is very very important to note the following, that here we are going to talk about Differentiability of a function from R n to R, but most of us are usually habituated in talking about the Differentiability of a function from R to r. Now, how do we manage by some intuition to move from the definition of a derivative from R to R, to a definition of a derivative in R n to R. So, let us just go back and for the time being consider the standard case of R to R, and the derivative of this function if that exists at x bar, is defined as a limit as h tends to 0, where h is the increment of x bar. Now, I can write this thing in slightly more compact way. I can say, that look I can just bring in here, because now this is the constant function it, the limit will have no effect on it. So, I can write down the whole thing as. This is very clear, this manipulation is simple and now observe one thing, that if this quantity when divided by h, and as h tends to 0 in the limit gives me 0; such a quantity is called the small 0 of h, that is f of x bar plus h, minus f x bar, minus f dash x bar h is called the small 0 of h, this is the called the small 0.

So, in general the derivative for a function from R to R would satisfy this expression. Now, how do I generalize this idea to a function for from R n to R that is the question.

So, it is all about generalizing this idea, now let me tell you that the very beginning, I am not been extremely rigorous, 1 can be extremely rigorous and defining gatho derivative and Fre'chet derivative, and doing lot of all the things and showing relations between them, but in straight of that, I preferred to go by the intuition and that would really help you have some fun, and really get an hang off the answer, before you have logically figured things out. So, that a bit of fun we can always do. Now, here I will go back just go back for second, I would generalize this part for this case. Instead of going back to a different page, let me generalize it from here, how do I generalize. Now, instead of h going to 0, because h would be now vector not a real number, I can just take the norm of h is going to 0, and then I have f of x bar plus h minus f of x bar, minus some vector v which we call the derivative of f at x bar.

Now, here instead of multiplication here would be change to inner product, because multiplication is generalized as inner product in R n, a high dimensions. This divided by norm h. This limit should be equal to 0. This is the same thing where I am writing this is equal to 0. Now, let us note a certain thing, this v here is called the derivative of the function from R n to R. So, we have just generalized this fact, just put in here. So, you can also write it like this. You can also put f of x bar plus h, minus f of x bar are equal to f of x bar plus v h, plus small 0 of norm of h. So, you can say this is the definition that there must be a vector v which satisfies this condition, and that v is called the derivative of f at x bar. Now, how does f of x bar look like, so that is the very important question, f of x bar.

(Refer Slide Time: 37:26)



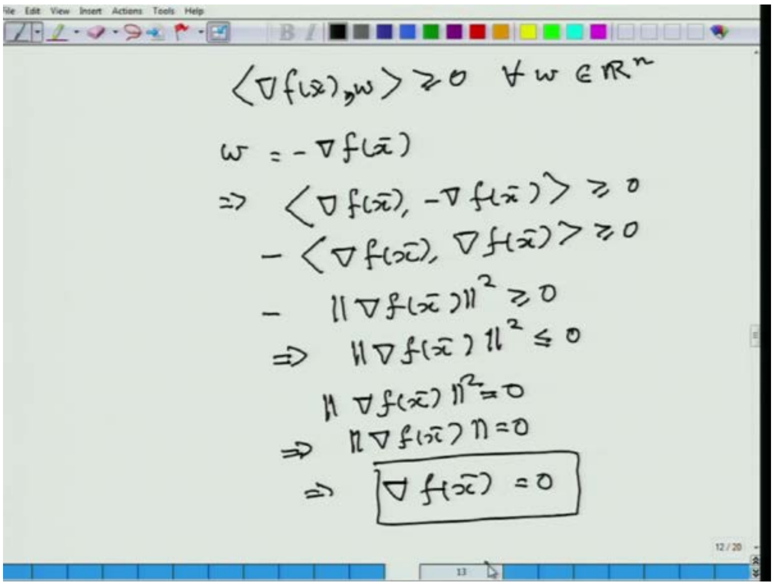
Now, the derivative of f at x bar which is v, how does it look like, it would be a small exercise for you to compute that this is a vector consisting of the partial derivatives, because there are n variables which we write, and that this partial derivative is evaluated at x bar, and this is given as the symbol gradient of f at x bar. These are standard thing known to most of us. So, this is what the derivative is, and more a precisely it is call the Frechet derivative. From more mathematical point of view Frechet derivative named after the French mathematician Morris Frechet. So, we will not get in to so much of details, but just for us this is what we would use as a derivative. So, here you can now replace v with grad of f of x bar. Now, let us show something; now if I have this definition of grad of f of x bar, can I say something about the local Minima if x bar is a local Minima.

Now, let us look at local Minima, see every drawing has to be R 2, drawings cannot be in R 3 or R n, R 3 sometimes, but it quiet difficult to really get a hang off, because we are ourselves in 3 dimension phases, we can only visualize things in 3 dimensional phases, but it is better to visualize thing the we can much easily visualize things in 2 dimension. So, that is why most of this drawings where even when people draw abstract sets, in say bench phases or topological vector phases, they thing are in pictures in R 2 . So, was x bar is a local minimum, now you have open neighborhood delta, that is open ball; p delta x, such that for every x inside this f of x is bigger than f of x bar. Now, what does this mean, so this is your vector x bar. Now, take any direction w, now move a certain distance along the direction w from x, so that is x bar plus lambda w, but I can make such a small movement, so that I can remain I can make take lambda greater than 0 to be small, so that I can remain within this set. You can take it less than 0 and come this side also does not matter, because of the symmetry of the ball. So, let me take lambda greater than 0, whatever w I take I can always move in the direction w, because you see moving in a direction parallel. So, I am parallel to w. So, I can move belong the direction w from x bar, but I can choose my lambda, so small there I can still remain in the ball B delta x bar. So, this is my.

So, what I can do is that, this thing can always be kept in B delta x bar, x bar plus lambda w. So, when I choose my lambda sufficient is small, I can always keep this x bar plus lambda w which is this point inside the ball. So, by definition of A local Minima f of x bar plus lambda w minus f x bar must be greater than equal to 0. But, then you must observe a fact that now I can apply the definition of differentiation, once I apply the definition of differentiation, I will have from here gradient of x bar, x bar plus lambda w minus x bar which will just leave me with lambda w, plus order small 0 of norm of lambda w, that is greater than equal to 0. So, this would imply gradient of f x bar, lambda w plus 0 of lambda. Now, you might ask me where did this w vanish. You see how does what are these 0 h functions, a small 0 h functions. So, small 0 h functions, suppose I have functions like this h square plus h. So, this is the small h function, because if I divide by h this becomes h into h plus, sorry I will just. So, take the function h square plus h cube. So, now, what I can do, I can take h and here keep h square. Now, if I divide by h and take limit as h tends to 0, this will go to 0, because here h will go to 0 and h square root go to 0. So, this is an 0 h quantity.

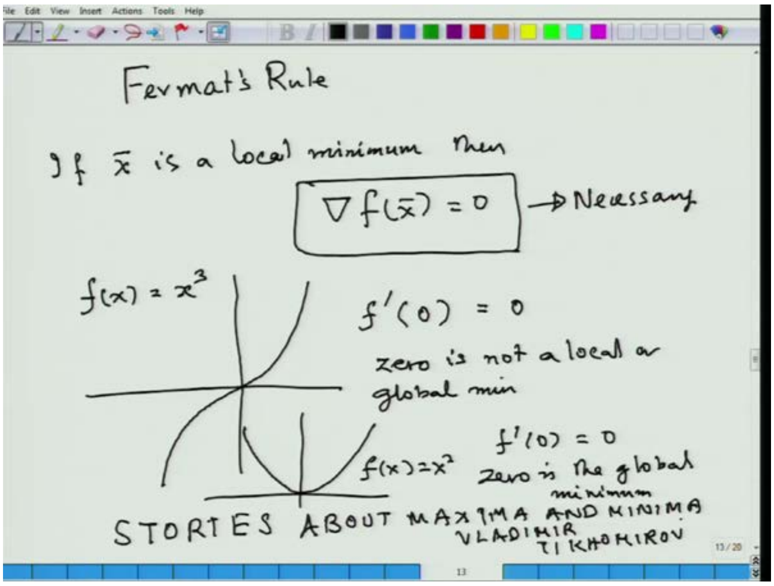
Now, if you see if I make multiply this by lambda here, 0 h, 0 lambda h here, I will have lambda square h square and lambda cube h cube. So, I can just take the lambda out in the similar fashion and divide by lambda and take lambda to 0, and get the same answer. So, instead of writing whenever you have lambda multiplied with a vector, it is same as writing the as if order quantity is same as the order quantity of lambda. So, you can always do this trick. Now, what I do because lambda is strictly bigger than 0, I can divide by lambda to get this condition. Now, I will take lambda going to 0 from the positive side, because lambda is positive. This is modern day symbolism lambda going down arrow 0. Now, this quantity by the very definition will vanish will go to 0 as we take the limit. So, as lambda tends to 0 we have. Now, you must note that, this thing that grad f x bar w is greater than equal to 0. But you should also note that this w was arbitrary, I could have taken any other w, I could have taken any other w, and I could have gone and have the same sort of argument and did the same calculations.

(Refer Slide Time: 45:03)



So, that would lead to the fact that, this is true every w you have in R n. So, now, I will put w equal to minus grad of f x bar. So, that would give me grad of f x bar minus grad of f x bar to be greater than equal to 0, but you see I can take the minus sign out just by the very definition of inner product, I can take the minus sign out. And then what you can do is that, you can now observe that this is, once you forget the minus here this is the inner product of x and x means not x and x 2 same vector. So, this by definition is the non-square of f x bar, this implies that grad of f x bar, but the distance is always non negative. Norm of x has to be non negative, because this the distance from 0. So, which means grad of f x bar square here, because it cannot be less than equal to 0, if it is less than equal to 0 the only option left is it has to be 0, and this would imply that grad of f x bar equal to 0, and this would imply grad of f x bar equal to 0, because norm of x equal to 0 if and only if x is equal to 0. So, grad of f x bar has to be the 0 vector.

(Refer Slide Time: 47:07)



And this is what is called the Necessary optimality condition or the Fermat's Necessary optimality condition or the Fermat's rule. So, if x bar is a local minimum, then. Now, one has to remember certain things here that this condition is a Necessary condition. I can show you a function like f x equal to x cube, whose derivative is 0 at a point, but that point is not A local minimum, because here if you observe if you calculate f dash 0, it is nothing, but 0 for this function. But, 0 is not a local or a global Minima. So, this condition is Necessary it is not always sufficient, but for some cases S if you take f x equal to x square. You see in this case when f x is x square, you have f dash 0 is equal to 0, and 0 is the global minimum. So, if f is from R to R, it is left to you to decide what is this condition. This is the standard thing that you have learnt in school f dash high school, rather f dash x equal to 0.

So, this condition is necessary this is very important, not sufficient. So, any point which satisfies this condition need not be a Minima. You have to put additional conditions to show that it is a minimum. Those conditions etcetera will not be as per say the discussion of our story, because for a convex case this condition becomes necessary and sufficient. So, tomorrow we are going to talk about convex functions and convex sets, and we shall show, we shall proof that f for a convex function every local minimum is global, and then show the certain additional property is which makes convexly important. So, with this very basic idea about Maxima Minima and the derivation of the necessary condition, and with the knowledge of the Differentiability of a function, we will stop our discussion for today, and tomorrow will start our discussion on convex optimization.

So, before I go, I would say that you might ask me about some books to have a general idea about Maxima, Minima. The book that I would really prefer is a book call stories about maxima and minima by Vladimir M. Tikhomirov. Vladimir M. Tikhomirov is a great optimization theorist of Russian optimization theory; stories about maxima and minima published by A M S, now a cheaper addition is also available, but mind you need to have some sort of little mathematical maturity at least at the high school level to understand this book.

So, thank you very much for your attention and we hope to continue tomorrow with convex optimization in proper thank you.