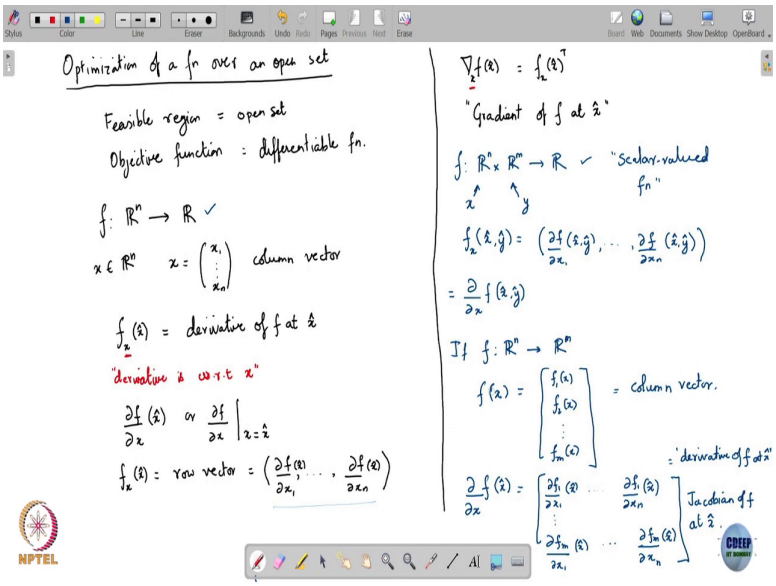
Optimization from Fundamentals
Prof. Ankur Kulkarni
Department of Systems and Control Engineering
Indian Institute of Technology, Bombay

Lecture – 4A
Taylor's theorem

(Refer Slide Time: 00:27)



Now, we will begin with our study of first class of Optimization problems, which is optimization of a function over an open set. So, what this means is that the feasible region is a open set and the function or objective function more precisely. Objective function, we will assume this to be a differentiable function.

Now, since we are going to talk about differentiable functions, I need to tell you a little bit about my notation for derivatives second derivatives and so on. So, now, for that let f be a

function from R n to R. So, any x in Rn I will write its I will denote it in this sort of way x will be denoted with its coordinates as x one till x n.

So, this is so every vector x in R n this way will be interpreted as a column vector. Now when I write something like f x evaluated at x hat, what this means is; this is the derivative of f at x hat ok. So, the small x here the small x in this notation just simply says that, derivative is with respect to x; with respect to x.

So, this will be important when we are talking about functions that are functions of more than 1 variable. So, we may want to take derivative only with respect to one of the variables ok. So, this is this just stands for derivative with respect to x.

Another notation for the same thing is this evaluated at x hat or this. Now the important thing to note is that this quantity f sub f x at x hat, this is actually a row vector. Row vector and it is defined in this way you [vocalised-noise] consider the partial derivatives of f with respect to each of the components of x and put them into a row vector evaluate all of these at x hat.

So, f sub x evaluated at x hat is this row vector of this kind. There is a related quantity which is denoted in this sort of way or to be more explicit about what we are differentiating with respect to that is it is denoted this way and that is simply the transpose of this derivative. This thing is called the gradient of f at x hat.

And once again the small x here simply denotes that the gradient is with respect to the variable x alright. So, suppose if I have a function f which is a function from R n cross R m to R and.

So, my variable my variable x the x variable lives in R n and the y variable lives in R m. Then if I write something like this f sub x, if I write something like this f sub x at x hat comma y hat this here is, can someone tell me is this a row vector or a column vector.

It is a row vector, with how many components? n components right; because that is the number of components. I am taking derivative with respect to x and x has n component. So, this come with this is a row vector this particular thing also denoted by ok.

Now, if f is so in both of these cases here in this case as well as in the earlier case in both of these cases f was a function that mapped to R right the. So, f itself did not have multiple components it has just one component right. So, f is a scalar valued function ok.

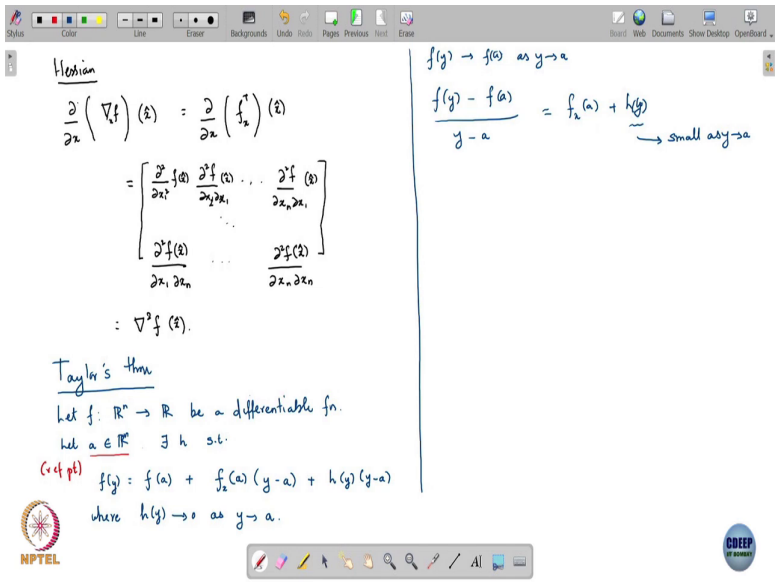
So, then now if f is vector valued let me just write this for your reference here. So, this is a case of a scalar valued function. If f is a vector valued function, then say we have suppose f is a function from R n to some R m right. And what we will do is, we will think of this every output of this vector every point in the image of this of f is itself a column vector.

So, f should be thought of like this, f at a point x is equal to this kind of column vector f 1 of x f 2 of x dot dot dot f m of x, it has m columns. Now if I want to, if I take the derivative of f with respect to x what I need to do is, take the derivative of each of these components with respect to x.

So, this evaluated at x hat now becomes a. So, every row every component of f will have a derivative and will its derivative would be a row vector like this and. So, what you are going to get now is row vectors of stacked one below the other. So, the whole thing could become a matrix.

So, this would now this would become at x hat ok. So, that is your it should be x n ok. So, that is your right this is also there are there are multiple names for this. You can call this is called the derivative of f at x hat that is one of the name, another name for it is that is called the Jacobian of f at x hat ok. Now we can also write high another higher order derivative. So, this is what is called the hessian.

(Refer Slide Time: 11:13)



So, the hessian is simply this. So, if I take the derivative of the derivative, but the derivative of the I need to. You know when I am taking the derivative of the derivative the inner thing has. So, let me sorry let me put it like this. So, I am taking the derivative of the gradient.

Evaluating that at x hat. So, this is simply another term for the same thing, another way of writing the same thing would be this is whole thing evaluated at itself. So, this will again because the gradient is a column vector, I can now take the derivative of that and that then will give me a again a matrix that looks like this.

All of these evaluated at x hat ok another notation for the same thing is simply del square f evaluated at x hat alright. So, along with derivatives we also need a an important theorem that pertains to derivatives, derivatives of derivatives and approximations of differentiable

functions. So, this theorem is it would be known to you in some or the other form probably before this is Taylors theorem.

Taylors theorem basically says that if you have a differentiable function ok. And if you look at the value of the function close to take a reference point and look at the value of the function very close to that reference point ok. What Taylors theorem is basically saying is that close to that reference point the value of the function is very well approximated by a linear function, that you can construct using the value of the function at the point and the derivative of the function at the bottom.

So, what does it mean by well approximated and what is the sense of the approximation that is the what is what is made precise by Taylors theorem, but the main idea is basically that, when you are close if your function is differentiable then and you want to look at how the function behaves near a point.

You have some reference point and you want to know, well near it how does the function behave? Well, it tells you that a linear approximation is can be obtained and it tells you what that approximation is ok, it tells you in a very precise sense.

So, the theorem is the following. So, let f from R n to R be a differentiable function let a be a point in Rn. So, this is my reference point this here this you can this is my reference point and what I want to know is how does the function behave at another point x ok. Then the theorem says, there exists and there exists this sort of function h, such that if you take f of x then that value of the function at x that other point x is given by this it is given as f of a plus.

Now, f of a plus I am just change my notation, I want to instead of using the point instead of denoting this by x let me denote this by a y ok. So, you so my reference point is a and my the other point at which I am want to evaluate the function, let us call that y. So, f of y is equal to f of a plus, what is this vector now? This is a row vector.

Row vector which is the derivative of f with respect to x evaluated at a. So, that transpose y minus a plus another function h of y times y minus a, where and this is the important part, where h of y tends to 0 as y tends to a ok.

So, now I want you to appreciate, what this sort what this theorem is actually saying and it will become evident as we go into the next main result also, but remember just for clarity I want you to see what this is saying.

See as y tends to a of course, it is true that f of y and f of a will come close to each other, as y tends to a f of y will approach f of a. So, that in so that is not saying anything here ok. What this theorem is saying is that as y tends to a if you look at f of y minus f of a, f of y minus f of a divide that whole thing by y minus a, then that starts behaving more and more like f x of a ok.

So, what this theorem is effectively saying is that, if you look at the it is of course, true that f as y tends to a as y tends to a f of y tends to f of a, but what this theorem is actually telling you is how fast does y tend does f of y tend to f of a as y tends to a.

So, what this is saying is its telling you telling you somehow a measure on this difference it is telling you that this difference f of y minus f of a divided by y minus a. This is something this is equal to f x of a plus h of y, where h of y is a quantity that will become smaller and smaller, this will become small as y tends to a. It will become close to 0 as y tends to a, is this clear?

So, this is what the theorem is actually effectively telling you. So, it is not only telling you that the function values near a are close to f of a that we already know, but it is also telling you how fast they approach f of a, they approach f of a at a rate that is linear in y minus a ok.

And the constant of linearity there is roughly equal to f x of a, its f x of a plus this h of y where h of y becomes smaller and smaller as you come close to it. So, all of optimization is about these about how fast different quantities converge ok. The relative rates at which different quantities converge is something that we keep exploiting all the time in

optimization. So, that is why an estimate like the like this one which comes from Taylors theorem is you can say a cornerstone of optimization ok.