Essential Mathematics for Machine Learning

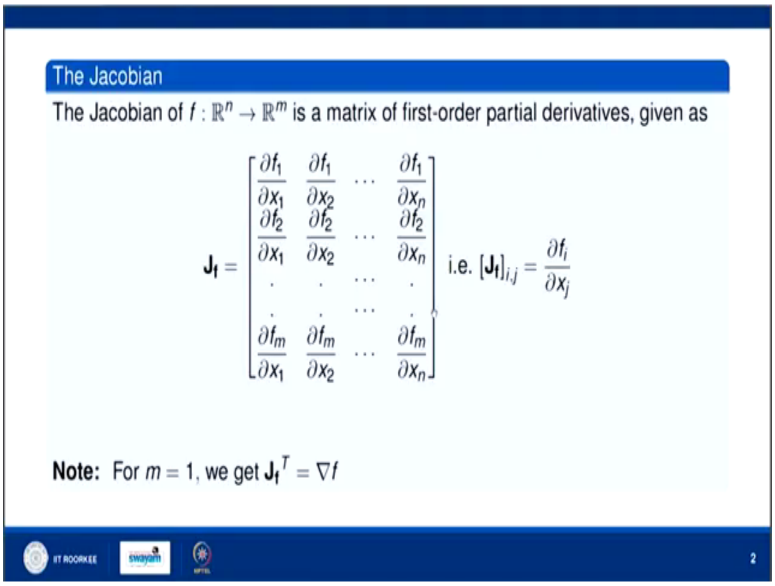
Prof. S. K. Gupta
Department of Mathematics
Indian Institute of Technology, Roorkee

Lecture - 22

Basic Concepts of Calculus - II

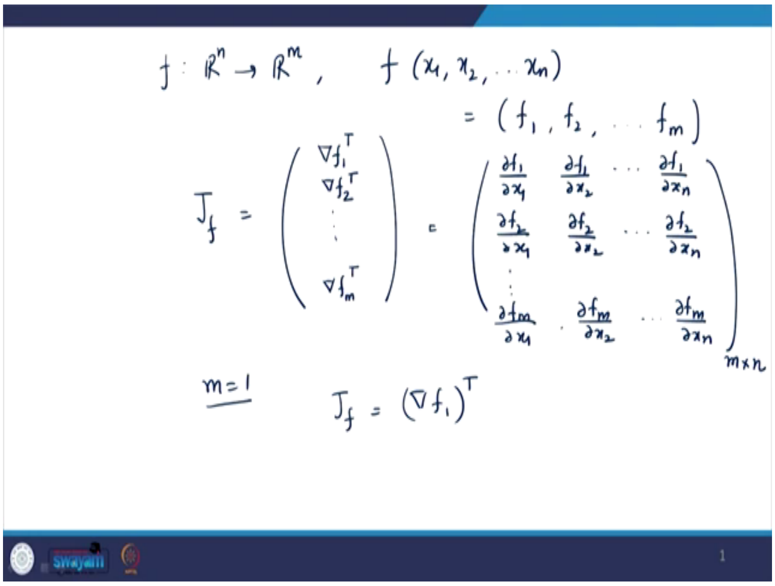
Hello friends. So, welcome to lecture series on Essential Mathematics for Machine Learning. In the last lecture, we have seen some of the basic concepts of calculus like gradient, directional derivative, partial derivatives etcetera. Now, in this lecture we will see some more concepts of calculus. So, first of all we will discuss Jacobian. What do you mean by Jacobian?

(Refer Slide Time: 00:48)



So, let function is from R n to R m ok.

(Refer Slide Time: 00:54)

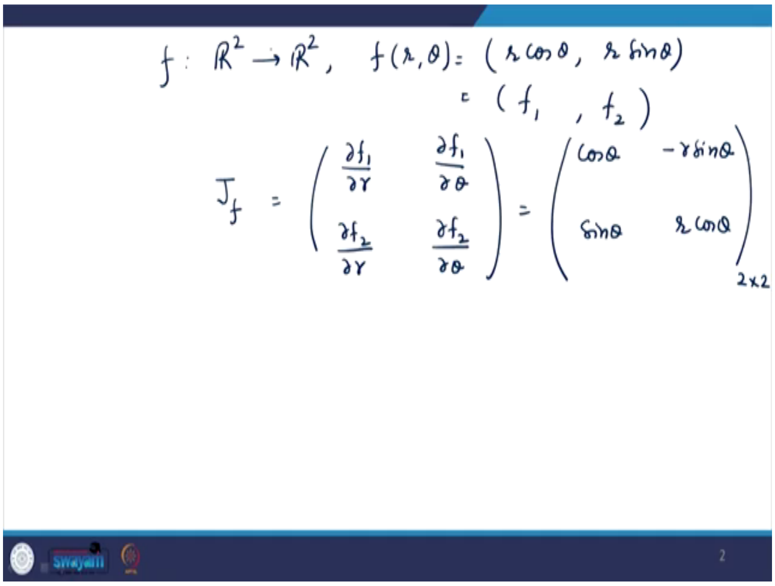


So, how we can define Jacobian? Now, here function is from R n to R m. So, that means, we are having an input variable x 1 x 2 up to x n and we are having m number of output variables f 1 f 2 up to f m. So, here the Jacobian of f; it is a matrix basically, first you take f 1. Differentiate partial with respect to x 1 x 2 up to x n; that means, it is nothing but gradient of f 1, gradient of f 2 and gradient of f m. And of course, transpose would be there.

So, that will be nothing, but del f 1 upon del x 1, del f 1 upon del x 2 and so on, del f 1 upon del x n. Similarly, del f 2 upon del x 1, del f 2 upon del x 2, del f 2 upon del x n. And last is del f m upon del x 1, del f m upon del x 2 and so on del fm upon del x n. So, this is basically of matrix of a m rows and n columns. So, that is how we can define Jacobian of f. It is nothing, but matrix of first order partial derivatives ok.

So, of course, if f if m equal to 1 if you take m equal to 1. So, this Jacobian of f will be nothing, but gradient of f 1 transpose the only the first row if m equal to 1.

(Refer Slide Time: 02:57)

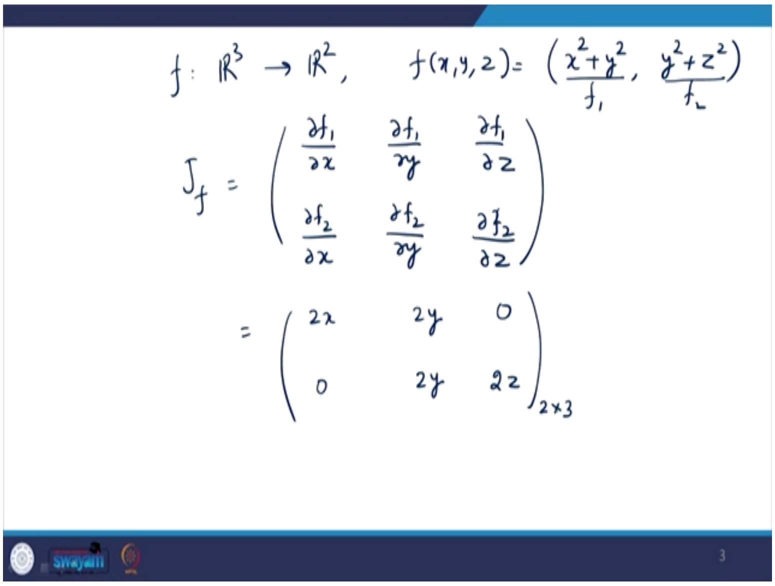


So, how we can compute it? So, suppose you have this example f is from suppose R 2 to R 2. And f so, f of r theta is given by r cos theta and r sin theta. So, here f 1 is basically if you deal with, so we are calling it as f 1 and this we are calling as f 2, f 1 f 2 again the functions of r and theta. So, if here if we want to compute Jacobian of f.

So, Jacobian of f will be, how many unknowns here? We are having two unknowns. Unknowns are independent parameters of r and theta. So, you take f 1 with respect to r you take f 1 with respect to theta you take f 2 with respect to r f 2 with respect to theta. So, that will be equal to; now this is f 1, f 1 is r cos theta. Differentiate partial with respect to r. So, this is cos theta. Now, f 1 this with respect to theta is minus r sin theta.

Now, f 2 with respect to r f 2 is this with respect to r it is sin theta and with respect to theta it is r cos theta. So, this 2 cross 2 matrix is basically Jacobian of this f ok. So, this is a function; this is a function from R 2 to R 2. Let us take another function from higher degree that maybe from R 3 to R 3.

(Refer Slide Time: 04:32)

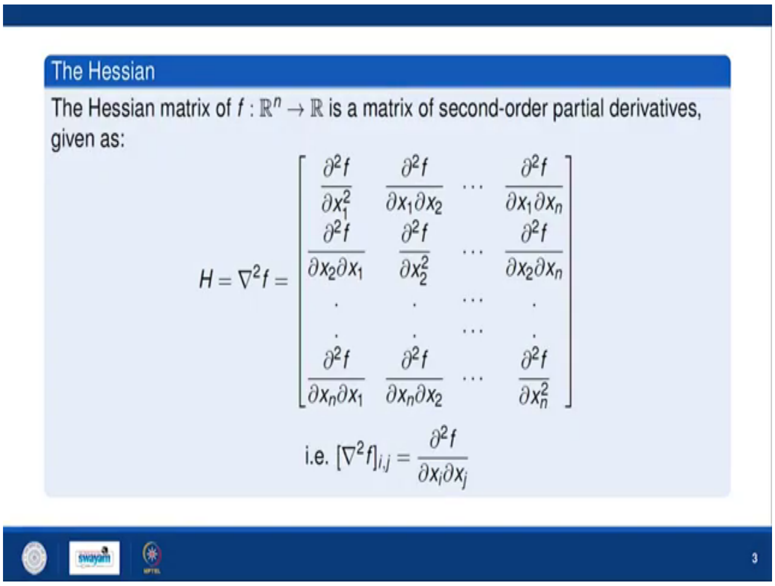


Let us take a function f from say R 3 to R 2; R 3 to R 2 and say f x y z is defined like this 3 input variable, say it is a x square plus y square and y square plus z square and suppose you want to compute Jacobian of f. So, again here you are having three 2 functions f 1 and f 2 ok. So, this will be del f 1 upon del x 3 input variables del f upon del y and it is del f 1 upon del z. Here it is del f 2 upon del x del f 2 upon del y del f 2 upon del z.

So now, what is del f 1 upon del x? See this is f 1 and this is f 2. So, the partial derivative of this f 1 with respect to x. So, what it is? It is 2 x with respect to y, it is 2 y with respect to z, this is 0. Now, for f 2 with respect to x it is 0, with respect to y it is 2 y, with respect to z it is 2 z. So, this matrix of 2 rows and 3 columns 2 cross 3 is basically Jacobian of this particular example.

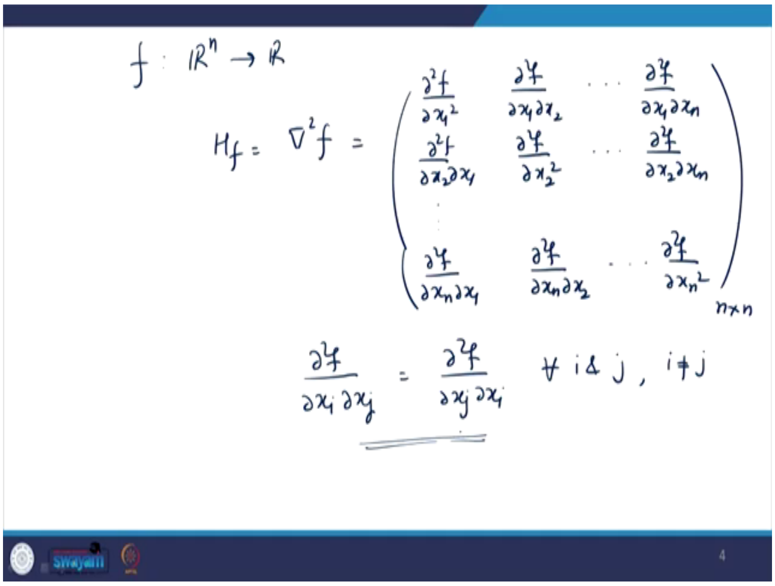
So, in this way we can compute Jacobian of a function from R m to R n ok. So, this is how we can define Jacobian suppose you want to compute a to 1 this is element a to one. So, this is nothing but del f 2 upon del x 1. So, if f if m equal to 1 then this is nothing but gradient of this we have already seen.

(Refer Slide Time: 06:16)



Now, next is Hessian matrix. So, what do you mean by Hessian matrix? A Hessian matrix is a matrix of second order partial derivatives. So, how we can constitute this? So, let us discuss this again by an example and.

(Refer Slide Time: 06:33)



So, let us suppose a function is from R n to R. So, Hessian matrix of f is nothing but second order partial derivatives of f. So, this is given by del square f upon del x 1 square del square f upon del x 1 del x 2 and this is del square f upon del x 1 del x n ok; that means, that means this is nothing but del by del x 1 of gradient of f.

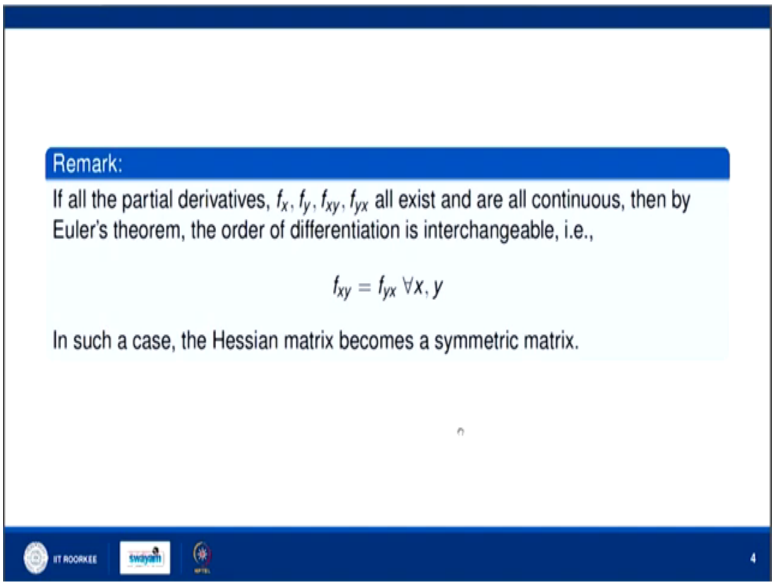
Similarly, del by del x 2 of gradient of f, similarly del by del x n of gradient of f. So, it is del by del x 2 x 1 of f del square f upon del x 2 square del square f upon del x 2 del x n and here it is del square f upon del x n del x 1, del square f upon del x n del x 2, del square f upon del x n square.

So, this n cross n matrix is basically called Hessian matrix of f ok. Now, if you carefully see that diagonal elements of this matrix are simply second order partial derivatives of f with respect to x 1 x 2 up to x n. And the and these diagonals, these elements are basically this is x 1 x 2, this is x 2 x 1, this is x 1 x n, this is x n x 1.

Now, if these 2 are equal, if these 2 are equal I mean I want to say that if del square f upon del x i del x j is equal to del square f upon del x j del x i. For all i and j i not equal to j then this matrix definitely will be a symmetric matrix. And when this will be equal?

We know from the Euler's theorem then this will be equal if the partial derivatives that is f x f y f x i x j f x j x i are continuous then they will be equal, this is by Euler's theorem.

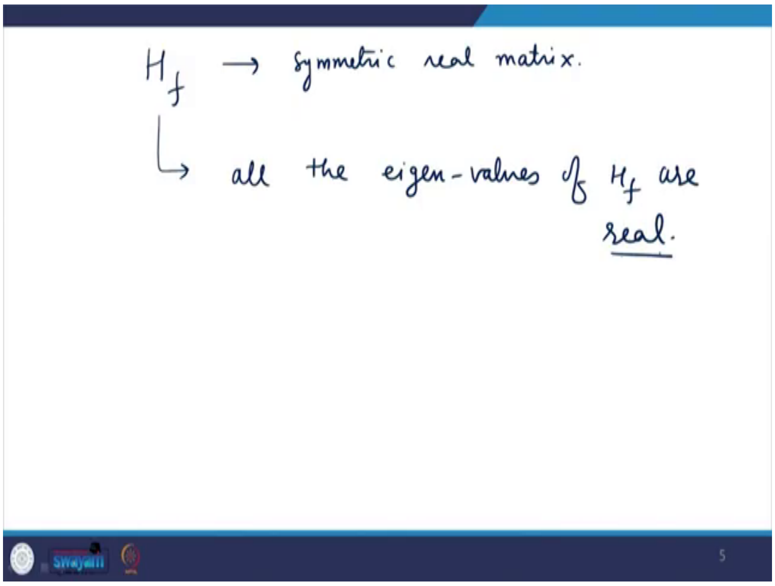
(Refer Slide Time: 08:57)



So, here it is mentioned that if all the partial derivatives f x f y f xy f yx all exist and are all continuous then by Euler's theorem the order of differentiation can be interchanged; that means, this is equal to this for all x and y if these are exist and all are continuous. So, if it is so, if it is hold for all xi x j then this matrix will be a symmetric matrix

So, if we assume that all partial derivatives are continuous throughout the open region where the function is defined throughout the open region, then Hessian matrix are symmetric.

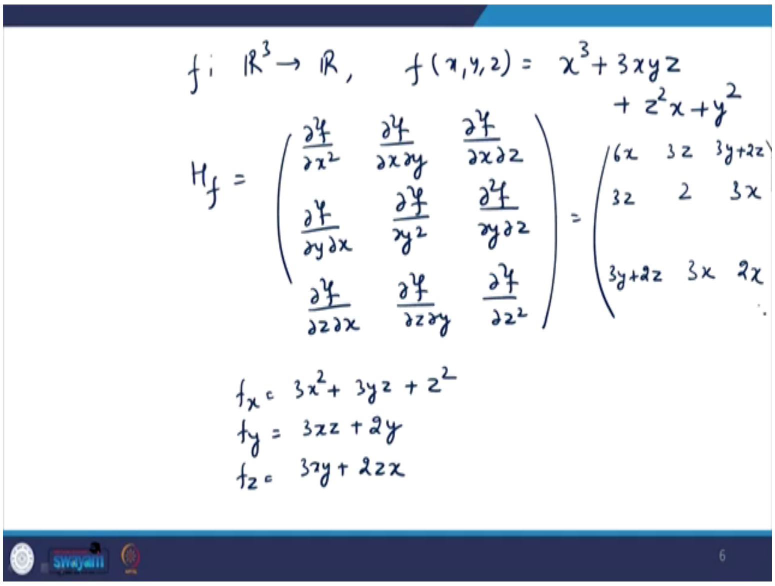
(Refer Slide Time: 09:41)



So, why we are; why we are dealing with symmetric matrix? See, so, suppose Hessian matrix we assume that Hessian matrix is symmetric; symmetric and real matrix. So, the first important property of this is that all it is eigen values are real because, if it is a symmetric matrix then we know that all the eigen value of symmetric matrix are real.

So, first of all we can say that all the eigenvalues of Hessian matrix are real. See this important property is used in finding other aspects of Hessian matrix. We will discuss later on that how why we are discussing eigen values of Hessian matrix. So, we will discuss it later on that it has some significant role in deciding the definiteness of a Hessian matrix ok.

(Refer Slide Time: 10:48)



So, the next is let us find the Hessian matrix of a simple function. Let us suppose I have function from say r 3 to r and the function f xy z is defined like this suppose. So, function is given by x cube plus 3 x y z plus z square x plus y square.

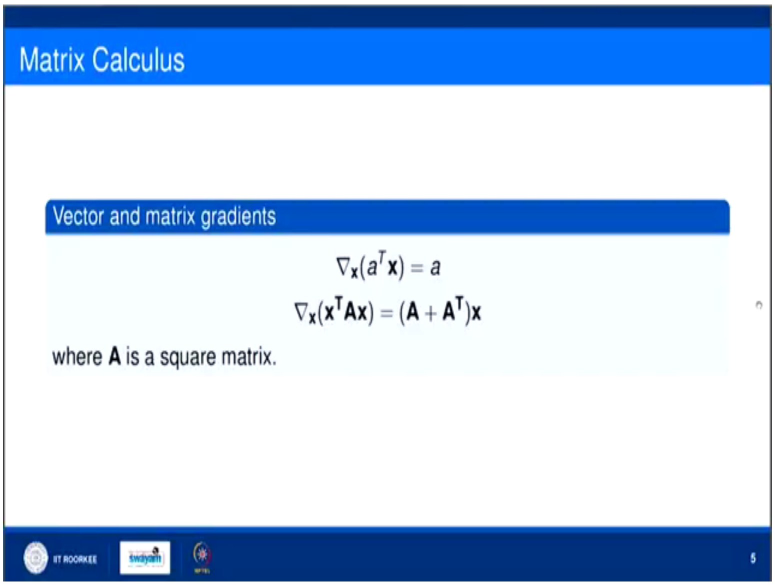
So, what would be the Hessian matrix of f here? It will be function of 3 variables; that means, del square f upon del x square, del square f upon del x del y, del square f upon del x del z. Again del square f upon del del y del x, del square f upon del y square del square f upon del y del z.

It is del square f upon del z del x del square f upon del z del y del square f upon del z square. So, let us compute this. So, del square f upon del x square is what? So, first you can compute f x f x is 3 x square plus 3 y z plus z square. What is f y? You will compute f y from here 3 it is 3 x z plus 2 y and f z is what? f z is 3 x y plus 2 z x.

So, what is second order partial derivative with respect of x? Is 6 x. What is derivative with respect to y or with respect to x? Sorry. So, it is 3 z ok, then with respect to z with respect to x of this with respect to x this is 3 y 3 y plus 2 z ok. Since it is symmetric. So, this will come here and this will come here ok.

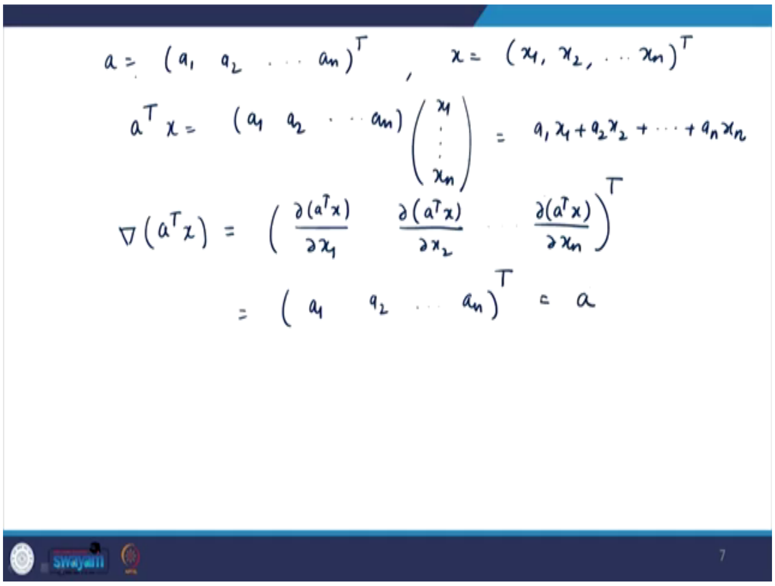
Partial derivative with respect to y is 2, the 2 will come here and del by del z y with respect to z or z with respect to y. So, this will be with respect to y with respect to why this is 3 x. If you take with respect to z then it is also 3 x. So, 3 x will come here because matrix is symmetric and as z is nothing but 2 x. So, this is the Hessian matrix of this f.

(Refer Slide Time: 13:21)



Now, we have some computation of matrix also in terms of gradient. So, vector and matrix gradient can be defined like this. If you want to compute gradient of this a transpose x, where a is a fixed vector there is not it is independent of free from x, then the gradient of this will be nothing, but a. And the gradient of x transpose ax will be given by this where a is a square matrix. So, let us discuss where it how we can obtain this.

(Refer Slide Time: 13:56)



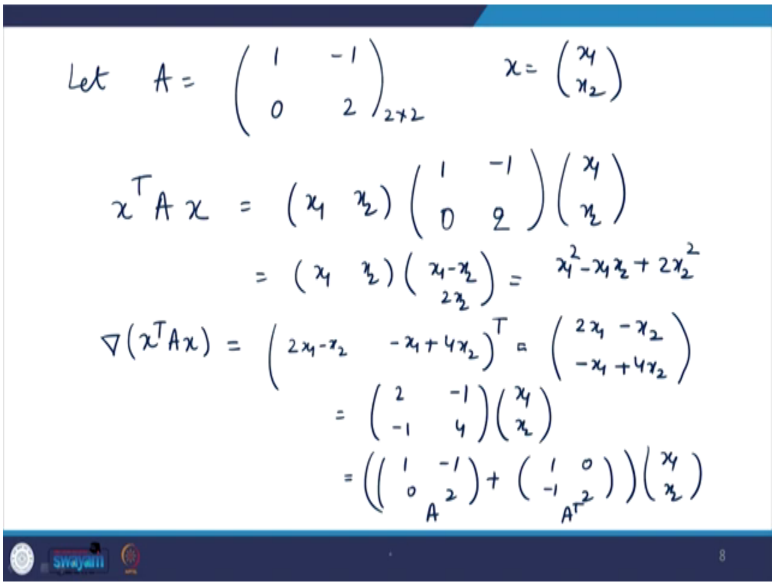
So, first is a first is a transpose x a is a fixed. So, a is a fixed vector which is a 1 a 2 up to a n. This is this transpose is a and x is basically x 1 x 2 up to x n x transpose.

So what is a transpose x? a transpose x is nothing but, it is a 1 a 2 up to a n and it is x 1 up to x n. So, this is simply this row this column which is a 1 x 1 plus a 2 x 2 plus and so on a n x n. Now, the gradient of f, gradient of this f f here is a transpose x will be will be del of a transpose x upon del x 1, del of a transpose x upon del x 2 and so on del of a transpose x upon del x n. This whole transpose this will be the gradient of f by the definition of gradient.

Now, if we take this f this is a transpose x. So, what is the; what is the derivative, what is the first order of what is the partial derivative of this a transpose x with respect to x 1? With respect because all are independent variables. So, with respect to x 1 it is nothing, but a 1 only.

So this will be nothing but a 1 a 2 and so on up to a n whole transpose. And this is nothing but a, as you have already seen here this is a. So, the first result. So, the first result we have shown that the gradient of a transpose x is nothing but a only ok. Now, let us go to the second result. Gradient of x transpose ax which is a plus a transpose x. Let us try to prove this.

(Refer Slide Time: 16:01)



So, before proving it, let us understand what a x transpose ax is. So, for this let us take a simple example. Let a is simply say you can take 1 minus 1 0 2. Let us take a simple example.

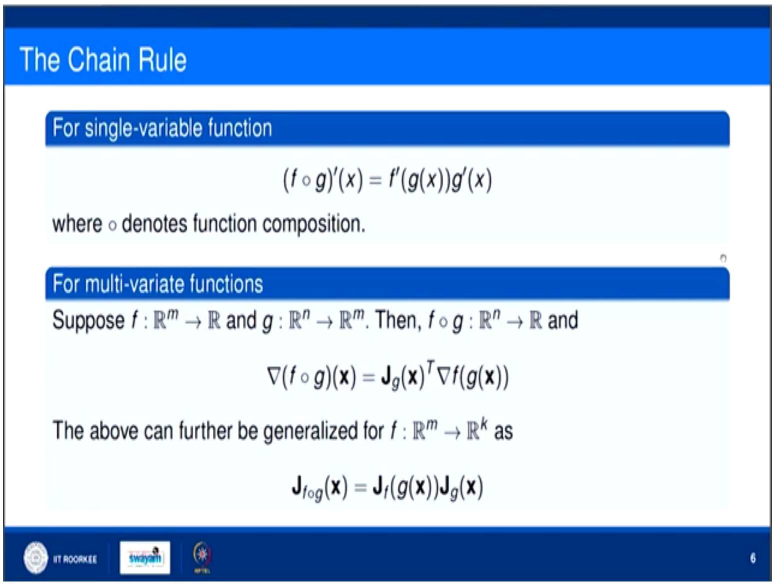
So, what is x transpose a x? So, x here will be which are 2 cross 2 matrix. So, x will be x 1 x 2 ok. So, this will be nothing but x 1 x 2 this is 1 minus 1 1 0 2 0 2 and this is oh sorry 0 2 and this is x 1 x 2. So, this will be x 1 x 2. This row this column is x 1 minus x 2 this row this column is 2 x 2. So, this is x 1 square minus x 1 x 2 plus 2 x 2 square. So, by simple calculation we can obtain x transpose ax is this.

Now, let us compute gradient of this. Gradient of this means; here there are only 2 variables. So, means, del del of this with respect to x 1 and del of this with respect to x 2 del of this with respect to x 2. So, this is nothing but, so you first differentiate this with respect to x 1 with respect to x 1 it is 2 x 1 minus x 2, with respect to x 2 it is minus x 1 plus 4 x 2 transpose. So, this is nothing, but if you see here it is 2 x 1 minus x 2 it is minus x 1 plus four x 2 ok.

So, this is further equal to 2 minus 1 minus 1 4 x 1 x 2. See this row this column gives first factor, this row this column gives second element. Now, if you see this matrix. So, this matrix is nothing but, 1 minus 1 0 2 plus 1 minus 1 0 2 with x 1 x 2. See this plus this is 2, this plus this minus 1, this plus this minus 1, this plus this 2, this is a and this is a transpose ok.

So, we can say we have verified this second point this is this verification is only for illustration that what do you mean by x transpose ax and how we have obtained this.

(Refer Slide Time: 18:40)

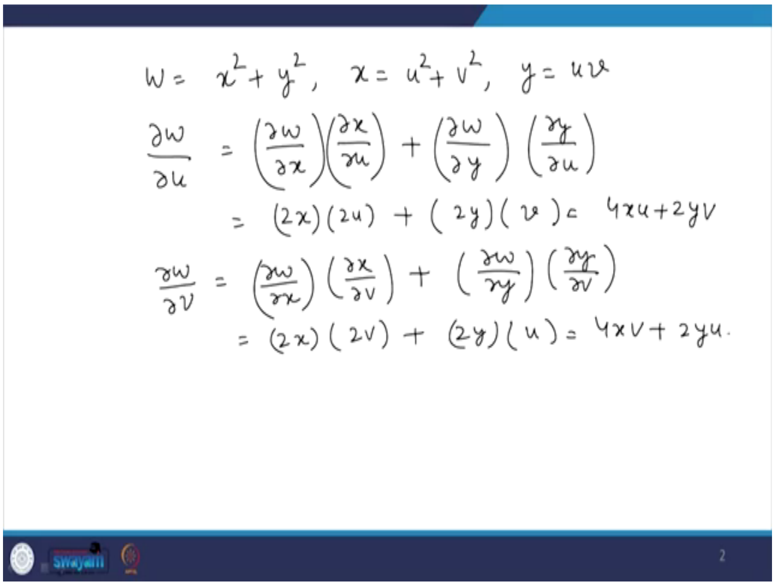


So, next we are having chain rule. So, for a single variable function composition of 2 functions f and g of dash f x is nothing but, f dash g x into g dash x; where this denotes function composition.

So, this is basically chain rule which we can which we can other extended to multi variable also. So, suppose f is from R m to R and g is from R m to R n, then the composition of this which is from R n to R, the gradient of this is nothing but, Jacobian of g transpose into gradient of f g x. So, this can be easily obtained using this calculus of single variable functions and definition of Jacobian in gradient.

And this can be further generalized for function from R m to R k as Jacobian of fog is nothing but Jacobin of g of f x into Jacobian of g of x ok. So, for chain rule let us discuss 1 example for this.

(Refer Slide Time: 19:47)



Suppose you are having w equal to x square plus y square, suppose x is u square plus v square and y is equals to uv ok. So, w is a function of x and y and x and y are again the function of u and v. And suppose you want to compute del w upon del u.

So, what it is? It is del of w upon del x into del x upon del u, because x and w is the function of x and y and x and y in terms of the function of u and v. So, it is del y into del y upon del u. So, what is del upon del x? Is 2 x. What is del x upon del u? It is 2 u plus del upon del y is 2 y and this is 2 u again del y.

So del y is del y is basically v. So, this is four x u plus 2 y v. So, now, if you want to compute del upon del w upon del v. So, this is nothing but del upon del x into del x upon del v plus del w upon del y into del y upon del v.

So, this is 2 x into this del x upon del y is 2 v plus this is 2 y into this is u. So, this is 4 x v plus 2 y v u. So, this is basically one simple example of chain rule. So, in this way we can also find out higher order derivatives using chain rules.

So, thank you very much for hearing me.