Course Name: Optimization Theory and Algorithms

Professor Name: Dr. Uday K. Khankhoje

Department Name: Electrical Engineering

Institute Name: Indian Institute of Technology Madras

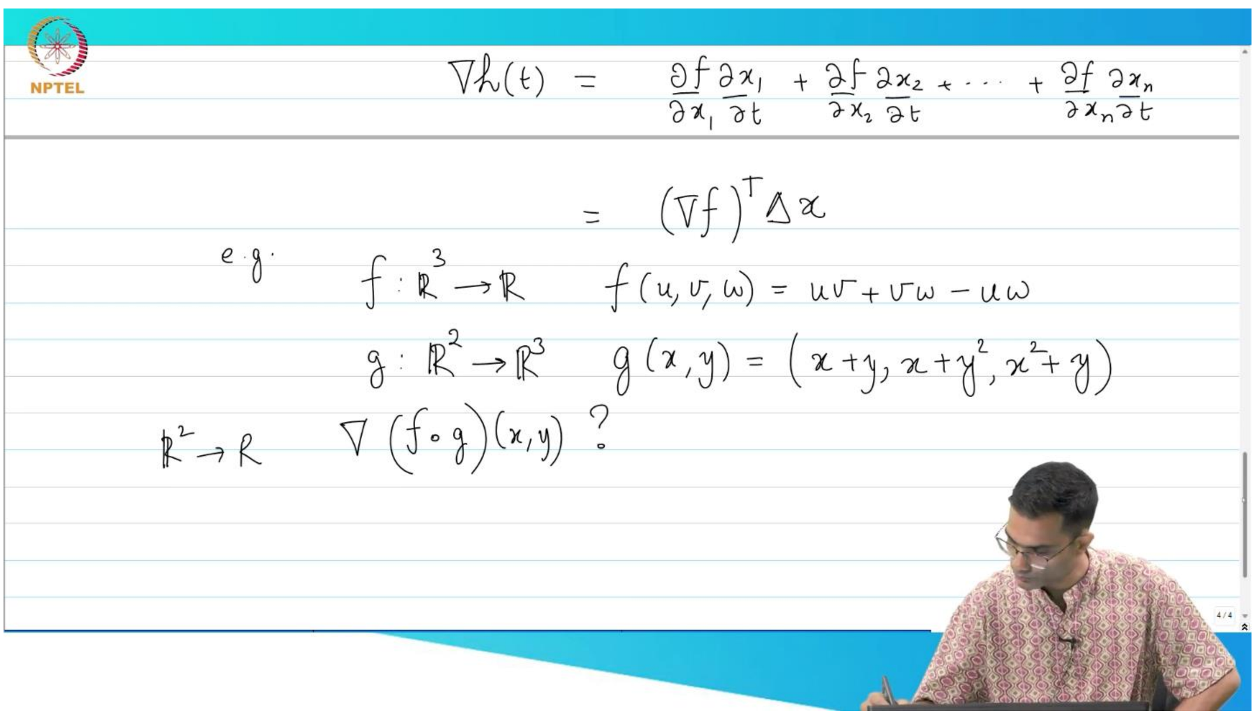
Week - 02

Lecture - 12

Example of Multivariate Differentiation

I am going to give you two functions; we will combine them and then try to find a gradient. Let us take our function f which is going from $\mathbb{R}^3$ to $\mathbb{R}$. This means inputting three variables and outputting one variable, so let us say $f(u, v, w)$. Here is another function which takes 2 inputs and gives 3 outputs, and I will call it g. Let $g(x, y)$ produce the outputs $x + y$ and $(x + y)^2$.

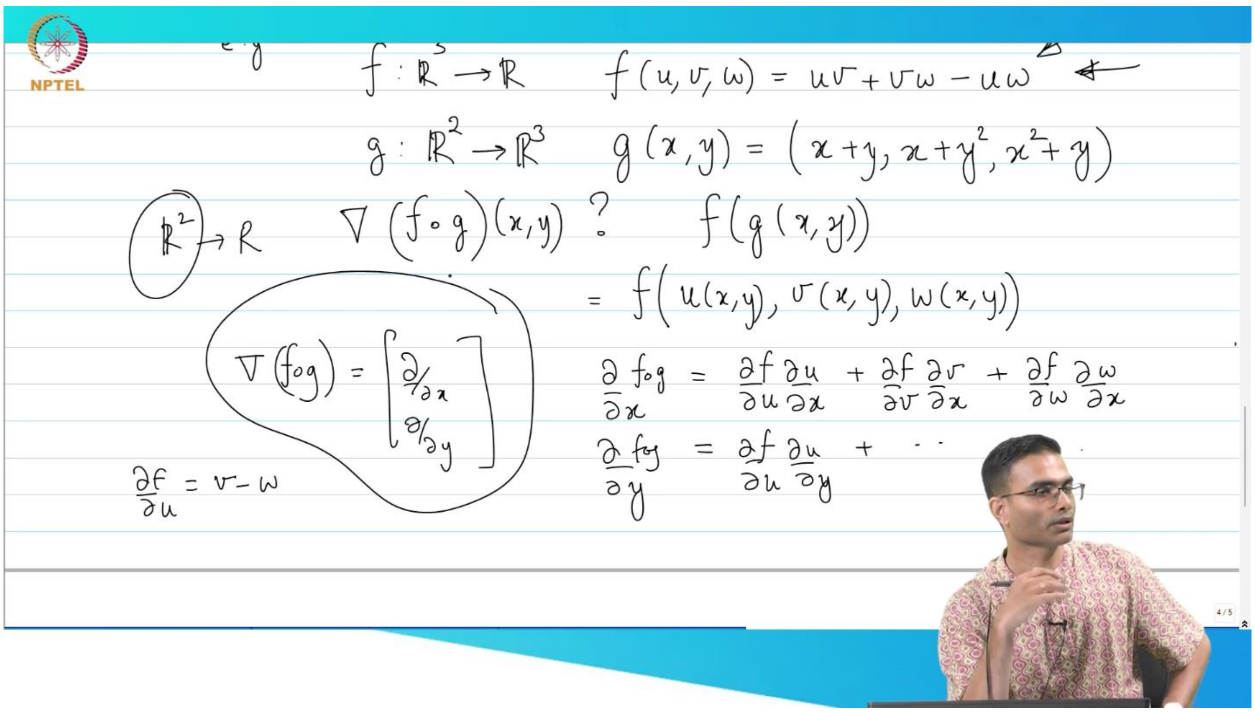
In general, we will use the following notation to describe a function with many outputs, separated by commas. This function takes two variables and outputs three variables. Now, the question is: if I compose f with g, that means g takes two variables and spits out three variables, while f takes three variables and outputs one variable. So, the composition $f \circ g$ is going to be a function from where to where? It maps $\mathbb{R}^2$ to $\mathbb{R}$.



Finally, we want the gradient of this function. So, we are looking at $f\big(g(x,y)\big)$. Remember, f takes in three inputs u, v, w; except now u, v, and w are actually functions of x and y. Thus, this can be represented as $f\big(u(x,y), v(x,y), w(x,y)\big)$.

To find the gradient, we note that we previously wrote the gradient as a column vector with n elements. Now, when we write the gradient of this function, how many elements will there be? There will be 2 elements: $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$, because those are the two variables that this function depends on. Hence, we express it as:

$$\nabla(f \circ g) = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \end{pmatrix}.$$



Now, the first thing we need to do is $\frac{\partial(f\circ g)}{\partial x}$. Here we apply the chain rule. The first component depends on $u, v,$ and w, each of which in turn depends on x and y. Therefore, we write:

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial x} + \frac{\partial f}{\partial w}\frac{\partial w}{\partial x}.$$

Similarly, for $\frac{\partial(f\circ g)}{\partial y}$:

$$\frac{\partial(f \circ g)}{\partial y} = \frac{\partial f}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial y} + \frac{\partial f}{\partial w}\frac{\partial w}{\partial y}.$$

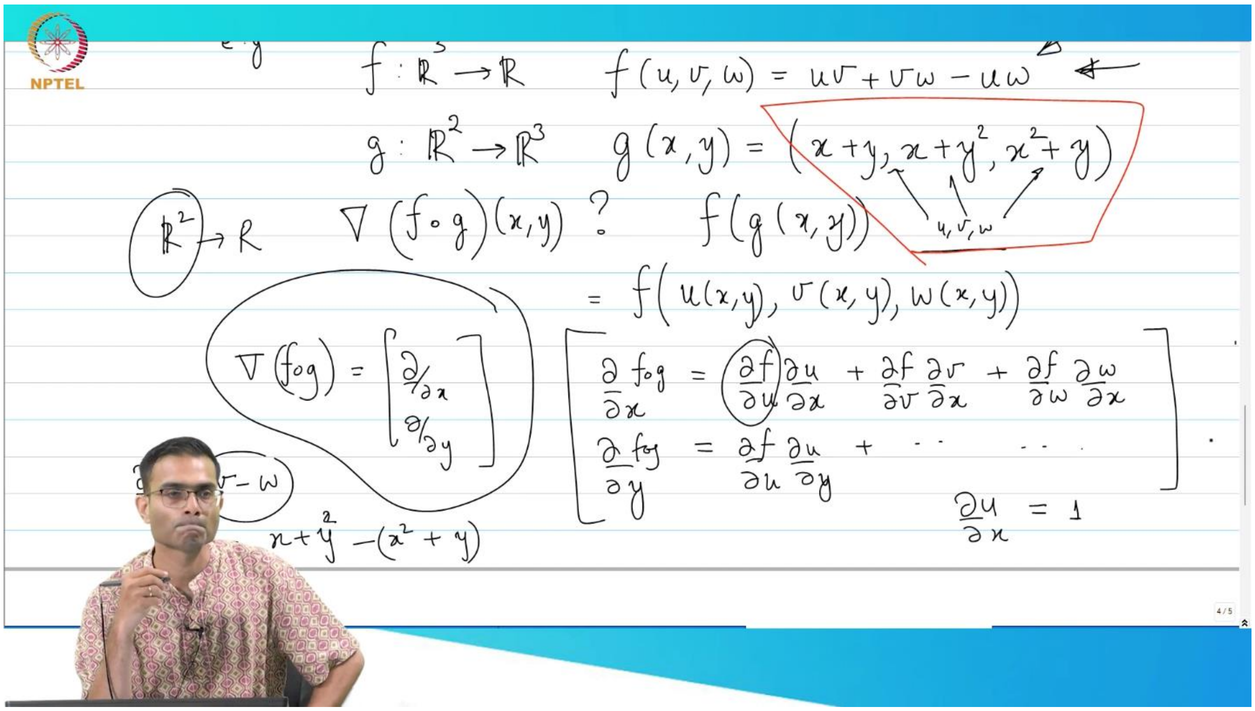
For example, let us compute $\frac{\partial f}{\partial u}$. Looking at the definition of f, we see that:

$$\frac{\partial f}{\partial u} = v - w.$$

To eliminate $u, v,$ and w from the final expression, we need to substitute in terms of x and y. Here, v is given by $x + y^2$ and w by $x^2 + y$. Thus, we can compute:

$$\frac{\partial f}{\partial u} = (x + y^2) - (x^2 + y) = x + y^2 - x^2 - y.$$

Next, we need $\frac{\partial u}{\partial x}$, which is simply 1 from our earlier expression.



Finally, to arrive at the final gradient expression, we can express it in a more compact notation. Notice that ∇f^T is a row vector being multiplied by the partial derivatives of $u, v,$ and w with respect to x and y. Hence, multivariate calculus is not difficult; we just have to keep track of the dimensions carefully and repeat the same logic for every dimension.

Any questions on this?

Now, as I mentioned previously, there is one type of function that we will spend a lot of time dealing with. Let us look at its calculus beforehand, so that when we work with it later, we are at peace. There is something called a quadratic form.