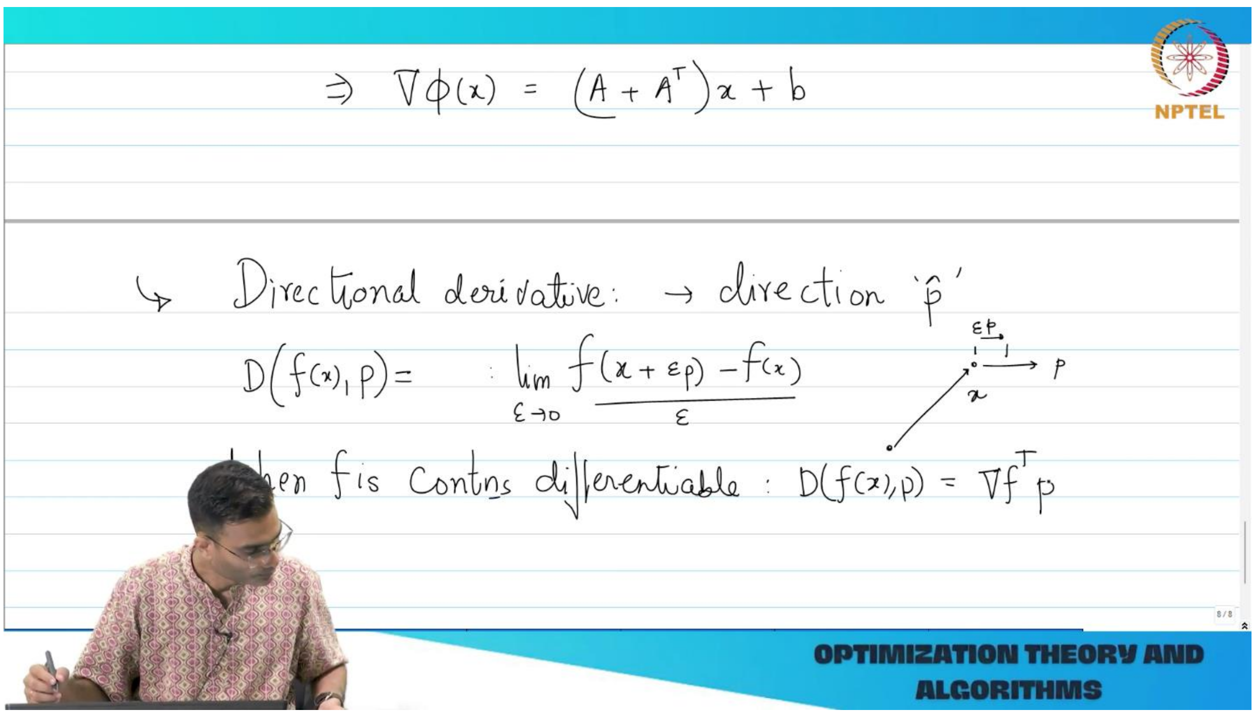
Course Name: Optimization Theory and Algorithms
Professor Name: Dr. Uday K. Khankhoje
Department Name: Electrical Engineering
Institute Name: Indian Institute of Technology Madras
Week - 02
Lecture - 14

Directional Derivative, Hessian, and Mean Value Theorem

The other thing that gets used quite a bit in optimization is something called a **directional derivative**. A direction is given by the symbol $\mathbf{p}$. Before I go into the definition, let us get a geometric picture of what we are saying: the rate of change of the function along the direction $\mathbf{p}$ is the meaning of the directional derivative.

The definition is very intuitive: take the function value along the direction $\mathbf{p}$ by making a small change ϵ and find the limit. Mathematically, if I am at $\mathbf{x}$, I need to go a small distance in the direction $\mathbf{p}$. So, I will denote a small walk along the direction $\mathbf{p}$ starting from $\mathbf{x}$ as:



$$\mathbf{x} + \epsilon \mathbf{p}.$$

Thus, the directional derivative is given by:

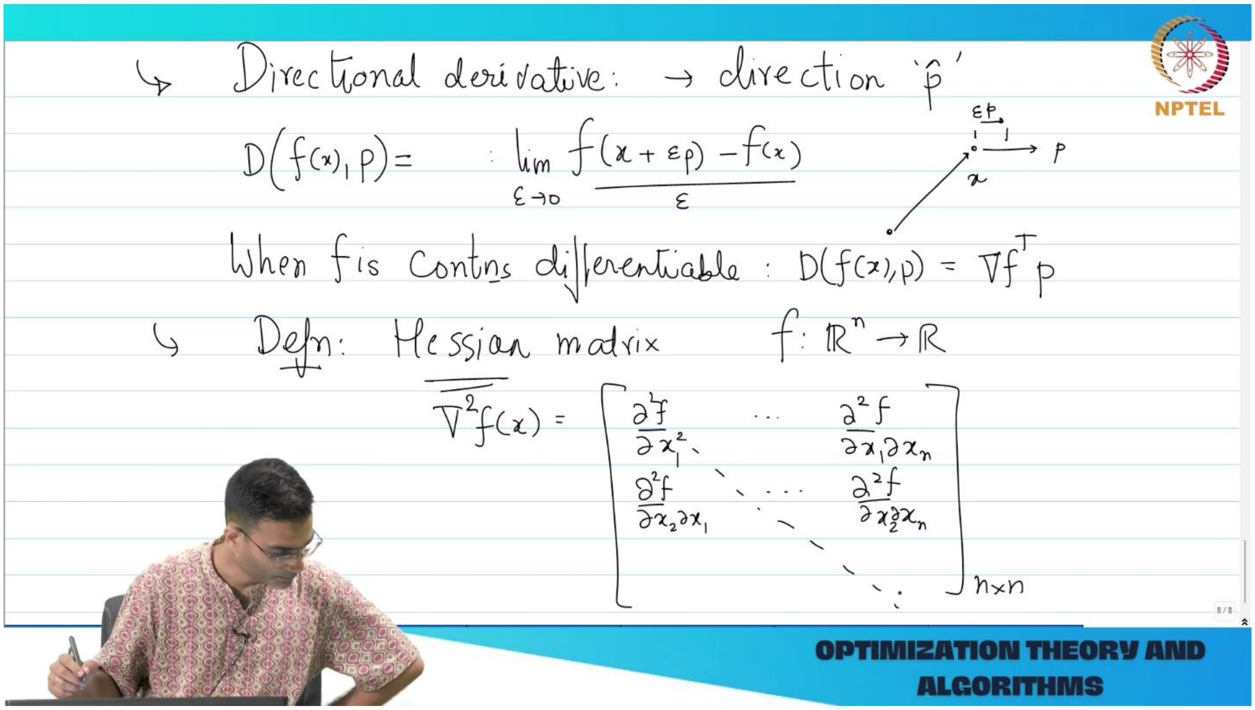
$$\lim_{\epsilon \to 0} \frac{f(\mathbf{x} + \epsilon \mathbf{p}) - f(\mathbf{x})}{\epsilon}.$$

Typically, $\mathbf{p}$ is specified as a unit vector, but we can define it as a small vector $\epsilon \mathbf{p}$. The directional derivative can be expressed as:

$$D_{\mathbf{p}} f(\mathbf{x}) = \nabla f(\mathbf{x})^T \mathbf{p},$$

where ∇f is the gradient of f. The gradient can be visualized as a vector with n components, and the dot product translates to the transpose. This means we are taking the inner product of the rate of change vector and the direction vector $\mathbf{p}$.

Now, the next topic is the **Hessian**, which is the generalization of the second derivative, represented as a matrix of second derivatives. For simplicity, consider $f: \mathbb{R}^n \rightarrow \mathbb{R}$. The Hessian matrix H is defined as:



$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \, \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \, \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \, \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \, \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \, \partial x_1} & \frac{\partial^2 f}{\partial x_n \, \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

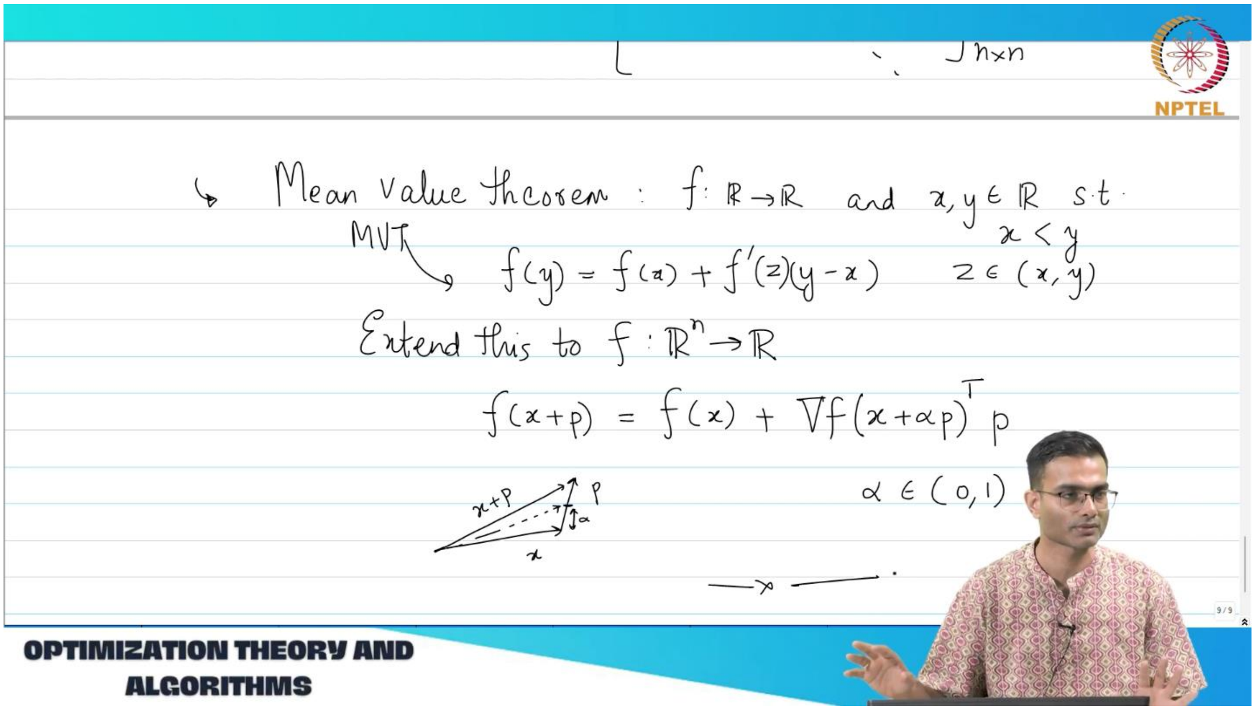
The diagonal terms contain the pure double derivatives, and this is a straightforward generalization of single-variable calculus. We will use this when we start discussing Newton's methods in optimization.

Lastly, we have the **Mean Value Theorem (MVT)**, which states that for a function $f: \mathbb{R} \rightarrow \mathbb{R}$ and $x, y \in \mathbb{R}$ such that $x < y$, we have:

$$f(y) = f(x) + f'(z)(y - x),$$

for some z between x and y. This shows that I can determine the value at y using the derivative at some intermediate point z.

When extending this to functions $f: \mathbb{R}^n \to \mathbb{R}$, we write:



$$f(\mathbf{x} + \mathbf{p}) = f(\mathbf{x}) + \nabla f(\mathbf{z})^T \mathbf{p},$$

where $\mathbf{z}$ lies on the line segment between $\mathbf{x}$ and $\mathbf{x} + \mathbf{p}$. Here, α is a scalar in $(0,1)$, such that:

$$\mathbf{z} = \mathbf{x} + \alpha \mathbf{p}.$$

In summary, we covered the definitions of the directional derivative, the Hessian matrix, and the Mean Value Theorem, which are essential concepts in optimization.

Summary:

- The directional derivative measures the rate of change of a function in the direction of a vector $\mathbf{p}$.
- The Hessian matrix is a matrix of second derivatives that generalizes the second derivative for functions of multiple variables.
- The Mean Value Theorem provides a way to relate the values of a function at two points to the value of its derivative at some intermediate point.