Gradient, Jacobian, Hessian, Laplacian and all that

In this article I will explain the different derivative operators used in calculus. Before we start looking into the operators let's first revise the different types of mathematical functions and the concept of derivatives.

In mathematics, a function is a mapping between a set of inputs and a set of permissible outputs with the property that each input is related to exactly one output.

If the mapping is from a scalar to another scalar then we call it a **scalar function**.

$$f : \mathbb{R} \to \mathbb{R}$$

For example, the function $f(x) = x^2$ is a mapping from the set of real numbers to the set of non-negative real numbers $f : \mathbb{R} \to \mathbb{R}$.

If the mapping is from a vector to a scalar then the function is known as a multivariable or **multivariate function**.

$$f : \mathbb{R}^n \to \mathbb{R},$$

where n is the dimension of the input vector. For example, the function $f(X) = \sqrt{x_1^2 + x_2^2}$ is a mapping from the set of 2-dimensional real vectors $X = [x_1, x_2]$ to the set of real numbers $f : \mathbb{R}^2 \to \mathbb{R}$.

If the mapping is from a vector (or a scalar) to another vector then the function is known as a vector function or a **vector-valued function**.

$$f : \mathbb{R}^n \to \mathbb{R}^m,$$

where n is the dimension of the input vector and m is the dimension of the output vector. For example, a function taking the 3-dimensional position as input and giving the velocity of the object in 3-dimensions is a vector valued function $f : \mathbb{R}^3 \to \mathbb{R}^3$.

Derivative

Now let's look at what the derivative of a function is. The derivative of a function is a measure of how much the value of a function $f(x)$ changes when we change the input x to the function, $\frac{\text{change in f(x)}}{\text{change in x}}$. In the case of scalar functions the concept of derivative is very simple as there is only one variable whose value need to be changed and there is only one output for which we need to measure the change. The derivative of a scalar function is denoted as

$$f'(x) = \frac{df(x)}{dx}$$

For example, the derivative of position with respect to time tells us how much the position of an object changes when we advance time and this is known as the velocity. We can also repeat the derivative operator and find the derivative of the derivative, for example, the derivative of the velocity with respect to time is known as the accelaration, which is another very useful physical quantity.

Gradient

To generalize the notion of derivative to the multivariate functions we use the gradient operator. The gradient of a multivariate function is a vector with each component proportional to the derivative of the function with respect to that component. For example, if we have a three dimensional multivariate function, $f(x_1, x_2, x_3)$, then gradient is given by

$$\nabla f = \left[\frac{\partial f(x_1, x_2, x_3)}{\partial x_1}, \quad \frac{\partial f(x_1, x_2, x_3)}{\partial x_2}, \quad \frac{\partial f(x_1, x_2, x_3)}{\partial x_3} \right]$$

Let's say the function $f(x_1, x_2, x_3)$ gives us the temperature of the point $[x_1, x_2, x_3]$ in a room. Now if we take a very small step in the x_1 direction, and note the change in temperature, then we get the derivative of f with respect to direction x_1, $\frac{\partial f(x_1, x_2, x_3)}{\partial x_1}$, let its value be 5. Now if we repeat it in the other two directions, we will find out how much the temperature changes in x_2 and x_3 directions, let their values be 15 and 0, respectively. Since the value of the function changes 3 times more in the x_2 direction than that of x_1 direction, if we take a step in the $[5, 15, 0]$ direction, it will give us maximum change in the temperature $f(x_1, x_2, x_3)$. Thus the gradient vector gives us the magnitude and direction of maximum change of a multivariate function.

Jacobian

The Jacobian operator is a generalization of the derivative operator to the vector-valued functions. As we have seen earlier, a vector-valued function is a mapping from $f : \mathbb{R}^n \to \mathbb{R}^m$, hence, now instead of having a scalar value of the function f, we will have a mapping $[x_1, x_2, \ldots, x_n] \to [f_1, f_2, \ldots, f_n]$. Thus, we now need the rate of change of each component of f with respect to each component of the input variable x, this is exactly what is captured by a matrix called Jacobian matrix J

$$J = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

Hessian

The gradient is the first order derivative of a multivariate function. To find the second order derivative of a multivariate function, we define a matrix called a Hessian matrix given by

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

Laplacian

The trace of the Hessian matrix is known as the Laplacian operator denoted by ∇^2,

$$\nabla^2 f = trace(H) = \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2}$$

I hope you enjoyed reading. Your feedback on this article will be highly appreciated.