

# MATHEMATICS

## Multivariate Calculus Cheat Sheet

### QUICK DERIVATION RULES

#### Derivative definition

$$f'(x) = \frac{df(x)}{dx} = \lim_{\Delta x \rightarrow 0} \left( \frac{f(x + \Delta x) - f(x)}{\Delta x} \right)$$

#### Sum Rule

$$\frac{d}{dx}(f(x) + g(x)) = \frac{d}{dx}(f(x)) + \frac{d}{dx}(g(x))$$

#### Power Rule

$$\text{Given } f(x) = ax^b, \\ \text{then } f'(x) = abx^{(b-1)}$$

#### Product Rule

$$\text{Given } A(x) = f(x)g(x), \\ \text{then } A'(x) = f'(x)g(x) + f(x)g'(x)$$

#### Chain Rule

$$\text{Given } h = h(p) \text{ and } p = p(m), \\ \text{then } \frac{dh}{dm} = \frac{dh}{dp} \times \frac{dp}{dm}$$

#### Total derivative

For the function  $f(x, y, z, \dots)$ , where each variable is a function of parameter  $t$ , the total derivative is:

$$\frac{df}{dt} = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} + \frac{\partial f}{\partial z} \frac{dz}{dt} + \dots$$

### DERIVATIVE STRUCTURES

Given  $f = f(x, y, z)$ , we have:

#### Jacobian

$$\mathbf{J}_f = \left[ \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right]$$

The Jacobian is a matrix representing the **partial derivatives** of a vector function with respect to its input variables. It is useful for understanding how the function changes as its independent variables change.

#### Hessian

$$\mathbf{H}_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial x \partial z} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial y \partial z} \\ \frac{\partial^2 f}{\partial z \partial x} & \frac{\partial^2 f}{\partial z \partial y} & \frac{\partial^2 f}{\partial z^2} \end{bmatrix}$$

The Hessian is a matrix of the **second partial derivatives** of a scalar function with respect to its input variables. This symmetric matrix, on the other hand, provides crucial information about the concavity and slope change of the function with respect to its variables.

### TAYLOR SERIES

#### Univariate

The Taylor polynomial is a way to approximate a complex function by a simpler polynomial. For a function  $f(x)$  that is differentiable infinitely many times around a point  $c$ , the Taylor polynomial centered in  $c$  is given by the following expression:

$$f(x) = f(c) + f'(c)(x - c) + \frac{1}{2}f''(c)(x - c)^2 + \dots \\ = \sum_{n=0}^{\infty} \frac{f^{(n)}(c)}{n!}(x - c)^n \quad (\text{Compact Formalization})$$

This polynomial approximates  $f(x)$  around point  $c$  using the derivatives of  $f$  evaluated at  $c$ , considering all successive derivatives.

💡 If we consider the univariate Taylor polynomial centered at  $c = 0$ , this becomes known as the **Maclaurin series**, which is a special form of the Taylor polynomial centered in the origin.

#### Multivariate

In the case of functions with several variables, the Taylor polynomial becomes an expansion around a point  $c$  in a multivariate space. For a function  $f(\mathbf{x})$  that can be differentiated infinitely many times around  $c$ , the Taylor polynomial centered in  $c$  is expressed as:

$$f(\mathbf{x}) = f(\mathbf{c}) + \mathbf{J}_f(\mathbf{c})(\mathbf{x} - \mathbf{c}) + \frac{1}{2}(\mathbf{x} - \mathbf{c})^t \mathbf{H}_f(\mathbf{c})(\mathbf{x} - \mathbf{c}) + \dots$$

Where  $\mathbf{J}_f(\mathbf{c})$  represents the vector of partial derivatives of  $f$  evaluated in  $c$  (the Jacobian) and  $\mathbf{H}_f(\mathbf{c})$  is the Hessian matrix, always evaluated in  $c$ . This expansion makes it possible to approximate  $f(\mathbf{x})$  around the point  $c$  in a multidimensional space.

### GRADIENTS

#### Gradient

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{bmatrix} \quad (\text{Vector of Partial Derivatives})$$

The gradient ( $\nabla f$ ) of a function is a vector that indicates the direction and maximum growth intensity of the function in a multidimensional space. It provides crucial information about the gradient of the function in each direction.

#### Gradient Descent

$$s_{n+1} = s_n - \gamma \nabla f$$

The gradient descent algorithm is an optimization method used to minimize a function iteratively. Using the negative gradient multiplied by a learning rate ( $\gamma$ ), the current position is updated to move toward the local or global minimum of the function