DIFFERENTIAL CALCULUS – HIGHER-ORDER DERIVATIVES

- Successive differentiation: taking the derivative of a derivative (of a derivative…)

- **SECOND DERIVATIVE:** 
  \[ f''(x) \equiv \frac{d^2 f(x)}{dx^2} = \frac{d}{dx} \left( \frac{df(x)}{dx} \right) \]

- In KINEMATICS, one encounters a DISTANCE function of TIME, whose first derivative corresponds to SPEED (itself may be considered VELOCITY of unspecified DIRECTION), and whose second derivative corresponds to ACCELERATION.

- A function is said to be **TWICE DIFFERENTIABLE** at a point \( x = a \) if \( \exists f''(x) \big|_{x=a} \).

- Moreover, a function’s CURVATURE may be stated in terms of the 2\(^{nd}\) derivative: a function is said to **CURVE UPWARD (DOWNWARD)** when the 2\(^{nd}\) derivative is **positive (negative)**.

- **INFLECTION POINT** \( \equiv \) a point where the curvature **changes sign**:
  \[
  \left( x^-, f(x^-) \right) \exists \forall x_{left}, x_{right} \in \delta(x^-) \exists x_{left} < x^- < x_{right}, \quad \text{sgn}(f''(x_{left})) = -\text{sgn}(f''(x_{left})) \]
  \( (66) \)

- **HIGHER-ORDER DERIVATIVE:**
  \[ f^{(n)}(x) \equiv \frac{d^n f(x)}{dx^n} = \frac{d}{dx} \left( \frac{\ldots}{dx} \right( f(x) ) \]

- A function is said to be \( N^{th} \)-TIME DIFFERENTIABLE at \( x = a \) if \( \exists f^{(n)}(x) \big|_{x=a} \).

- **TAYLOR’S THEOREM**, which allows for the **TAYLOR SERIES EXPANSION** of a value of differentiable function near a point of differentiation (for which derivatives exist):
  \[
  f(x) = \sum_{i=0}^{\infty} \frac{f^{(i)}(a)}{i!} (x-a)^i
  = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots \quad (67)
  \]

- Where the **LAGRANGE REMAINDER**, historically derived much later, is given by:
  \[
  R_n^{\text{Lagrange}} = \int_a^x \frac{f^{(n+1)}(t)}{(n+1)!} (x-t)^{n+1} \, dt \quad (68)
  \]

- Which, by the virtue of the **MEAN VALUE THEOREM**, equals to the **CAUCHY REMAINDER**:
  \[
  R_n^{\text{Cauchy}} = \frac{f^{(n+1)}(\xi)}{(n+1)!} (x-a)^{n+1}, \quad \text{for some } \xi \in [a, x] \quad (69)
  \]

- **MACLAURIN SERIES**: Taylor series expansion around \( x_0 = 0 \).
DIFFERENTIAL CALCULUS – UNIVARIATE EXTREMA

- **LOCAL/RELATIVE MAXIMUM & LOCAL/RELATIVE MINIMUM:**
  
  Local Maximum \( (x^*, f(x^*)) \) \( \forall \delta(x^*) \rightarrow f(x) \leq f(x^*) \) \( f(x) \leq f(x^*) \) \( \forall x \in \delta(x^*) \) \( \text{(70)} \)
  
  Local Minimum \( (x^*, f(x^*)) \) \( \forall \delta(x^*) \rightarrow f(x) \geq f(x^*) \) \( f(x) \geq f(x^*) \) \( \forall x \in \delta(x^*) \) \( \text{(70)} \)

  - EXTREMUM \( \equiv \) either a maximum or a minimum.

- **FIRST DERIVATIVE TEST:**
  
  \[
  f'(x) \begin{cases} 
  > (\langle) 0 & x < x^* \\
  < (\rangle) 0 & x > x^* 
  \end{cases} \Rightarrow (x^*, f(x^*)) = \text{Local Maximum (Minimum)} \quad \text{(71)}
  \]

  Provided the function is continuous at \( x = x^* \), this will always work, i.e. even if the derivative does not even exist at that point, as exemplified by \( f(x) = \pm |x - x^*| \).

- **SECOND DERIVATIVE TEST** – requires that the function is twice differentiable:

  - **FIRST-ORDER NECESSARY CONDITION:** \( f'(x^*) = 0 \), whence \( x^* \), \( f(x^*) \), and the coordinate pair \( (x^*, f(x^*)) \) are termed, respectively, **CRITICAL VALUE**, **STATIONARY VALUE**, and **STATIONARY POINT**.

  - **SADDLE POINT** \( \equiv \) a stationary point that is not also an extremum.

  - **SECOND-ORDER NECESSARY CONDITION** on a stationary point \( (x^*, f(x^*)) \):
    
    \[
    f''(x^*) \begin{cases} 
    \leq 0 & \Rightarrow (x^*, f(x^*)) \text{Local Maximum} \\
    \geq 0 & \Rightarrow (x^*, f(x^*)) \text{Local Minimum} 
    \end{cases} \quad \text{(72)}
    \]

  - **SECOND-ORDER SUFFICIENT CONDITION** on a stationary point \( (x^*, f(x^*)) \):
    
    \[
    f''(x^*) \begin{cases} 
    < 0 & \Rightarrow (x^*, f(x^*)) \text{Local Maximum} \\
    > 0 & \Rightarrow (x^*, f(x^*)) \text{Local Minimum} 
    \end{cases} \quad \text{(73)}
    \]

  Essentially, \( f''(x^*) < 0 \) \( (> 0) \) locates \( x^* \) within the concave (convex) part of the function.

  In other words, given a stationary point, this test is conclusive only when \( f''(x^*) \neq 0 \), as exemplified by \( f(x) = \pm (x - a)^2 \) at \( x^* = a \); the case where \( f''(x^*) = 0 \) remains inconclusive: it could be a local extremum, as exemplified by \( f(x) = \pm (x - a)^4 \) at \( x^* = a \), or a saddle point, as exemplified by \( f(x) = \pm (x - a)^3 \) at \( x^* = a \).

- **N\textsuperscript{th}-DERIVATIVE TEST** (aka general EXTREMUM TEST):
  
  \[
  \frac{df(x^*)}{dx} = \frac{d^2 f(x^*)}{dx^2} = \cdots = \frac{d^{k-1} f(x^*)}{dx^{k-1}} = 0 \neq \frac{d^k f(x^*)}{dx^k} \Rightarrow (x^*, f(x^*)) = \begin{cases} 
  \text{Local Maximum} & \frac{d^k f(x^*)}{dx^k} < 0 \\
  \text{Local Minimum} & \frac{d^k f(x^*)}{dx^k} > 0 \\
  \text{Saddle Point} & k \text{ even}
  \end{cases} \quad \text{(74)}
  \]
For instance:

\[ f(x) = (x - a)^{2n+1}, \quad n \geq 1 \]

\[ \frac{df(a)}{dx} = \cdots = \frac{d^{2n}f(a)}{dx^{2n}} = 0 \neq \frac{d^{2n+1}f(a)}{dx^{2n+1}} \Rightarrow (a, f(a)) = \text{Saddle Point} \quad (75) \]

**GLOBAL/ABSOLUTE MAXIMUM & GLOBAL/ABSOLUTE MINIMUM:**

- **Global Maximum** \( (x^*, f(x^*)) \) \( \forall x \in D_1, \quad f(x) \leq f(x^*) \)
- **Global Minimum** \( (x^*, f(x^*)) \) \( \forall x \in D_1, \quad f(x) \geq f(x^*) \) \( (76) \)

Here, \( f'(x^*) = 0 \) is not a necessary condition.

In fact, any **MONOTONICALLY INCREASING (DECREASING) FUNCTION** over a closed interval \( x \in [a, b] \) will have its minimum (maximum) occurring at the left endpoint \( x^* = a \), and hence its maximum (minimum) will be at the right endpoint \( x^* = b \).

**DIFFERENTIAL CALCULUS – MULTIVARIATE DIFFERENTIATION**

- For a multivariate function, define the **PARTIAL DERIVATIVES:**
  \[ \frac{\partial f(x_1, \ldots, x_n)}{\partial x_i} \equiv \lim_{\Delta x_i \to 0} \frac{f(x_1, \ldots, x_i + \Delta x_i, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)}{\Delta x_i}, \quad i = 1, \ldots, n \quad (77) \]

  **GRADIENT VECTOR** \( \equiv \) a vector of first partial derivatives:
  \[ f : \mathbb{R}^n \to \mathbb{R}, \quad \nabla f \equiv \left( \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_i}, \ldots, \frac{\partial f}{\partial x_n} \right), \quad i = 1, \ldots, n \quad (78) \]

- Note that **GRADIENT** itself, denoted \( \nabla \), is an **operator**.
- For a problem in **MINIMISATION/MAXIMISATION**, the gradient gives the **SOLUTION ALGORITHM** the **DIRECTION OF GREATEST DESCENT/ASCENT**.

- **TOTAL DERIVATIVE** – encapsulates the notion that, for \( \bar{\xi} \in \mathbb{R}^n \) “very close to” \( \bar{x} \in \mathbb{R}^n \), i.e. close w.r.t. some vector norm defined in \( \mathbb{R}^n \), it makes sense to obtain the following approximation of \( f(\bar{\xi}) \) if one knew precisely both \( f(\bar{x}) \) and \( \nabla f(\bar{x}) \), i.e. the gradient vector \( \nabla f \) evaluated at \( \bar{x} \):
  \[ f(\bar{\xi}) \approx f(\bar{x}) + \nabla f(\bar{x}) \cdot (\bar{\xi} - \bar{x}) \quad (79) \]
Given some *(REFERENCE) UNIT VECTOR* $\vec{v} \in \mathbb{R}^n$, one may be interested in further defining a **DIRECTIONAL DERIVATIVE** as the instantaneous rate of change of the value of a multivariate function $y = f(\vec{x})$, $\vec{x} \in \mathbb{R}^n$ in the direction of $\vec{v}$:

$$D_v f \equiv \lim_{h \to 0} \frac{f(\vec{x} + h\vec{v}) - f(\vec{x})}{h} = \lim_{h \to 0} \frac{f(x_1 + hv_1, \ldots, x_n + hv_n) - f(x_1, \ldots, x_n)}{h} \quad (80)$$

In fact, with defined gradient vector $\nabla f$, a directional derivative is identical to the following *dot product*:

$$D_v f = \nabla f \cdot \vec{v} \quad (81)$$

Suppose that **SECOND (MIXED) PARTIAL DERIVATIVES** exist and are continuous, then these are commutative in the sense of **CLAIRAUT’S THEOREM**:

$$f : \mathbb{R}^{n+1} \to \mathbb{R}$$

$$\exists \text{ continuous } \frac{\partial F}{\partial x_i \partial x_j}, \frac{\partial F}{\partial x_j \partial x_i}, i, j \in \{1, \ldots, n\} \Rightarrow \frac{\partial F}{\partial x_i \partial x_j} = \frac{\partial F}{\partial x_j \partial x_i} \quad (82)$$

A well-known **COUNTEREXAMPLE**, where the mixed partial derivatives are not continuous at $(x, y) = (0, 0)$, is given by:

$$f(x, y) = \frac{xy(x^2 - y^2)}{x^2 + y^2} \Rightarrow \frac{\partial}{\partial y} \left( \frac{\partial f(0, 0)}{\partial x} \right) \neq \frac{\partial}{\partial x} \left( \frac{\partial f(0, 0)}{\partial y} \right) \quad (83)$$

Consider $n$ multivariate functions in $n$ unknowns, i.e. where it is possible to define a *vector field*. If all $n$ partial derivatives exist for these $n$ multivariate functions, then it is possible to construct, i.e. by “stacking” gradient row vectors vertically (one top one another), the **JACOBIAN MATRIX**:

$$f_i : \mathbb{R}^n \to \mathbb{R}, \quad i = 1, \ldots, n,$$

$$J(f) = J(f_1(x), \ldots, f_n(x)) \equiv \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix} \quad (84)$$

**JACOBIAN DETERMINANT** $\equiv$ determinant of the Jacobian matrix (confusingly, the term “The Jacobian” may well refer to either the Jacobian matrix or its determinant):

$$D_{\text{Jacobian}}(f) \equiv \left| J(f) \right| = \left| J(f_1(x), \ldots, f_n(x)) \right| \quad (85)$$

**NONLINEAR (FUNCTIONAL) INDEPENDENCE**:

$$f_i : \mathbb{R}^n \to \mathbb{R}, \quad i = 1, \ldots, n, \quad \exists J(x_1, \ldots, x_n) \neq 0 \quad (86)$$
For example, the following functions are **NONLINEARLY DEPENDENT**:

\[
f_1(x_1, x_2) = x_1 + 3x_2
\]
\[
f_2(x_1, x_2) = e^{x_1+3x_2} = \exp(f_1(x_1, x_2))
\]

\[
\therefore J(f_1(\mathbf{x}), f_2(\mathbf{x})) = \begin{vmatrix}
\frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\
\frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2}
\end{vmatrix} = \begin{vmatrix}
1 & 3 \\
3e^{x_1+3x_2} & 3e^{x_1+3x_2}
\end{vmatrix} = 3e^{x_1+3x_2} - 3e^{x_1+3x_2} = 0
\] (87)

Now consider just one multivariate function, again in \(n\) unknowns. If all \(n\) partial derivatives exist for this one multivariate function, then let’s agree that these \(n\) partial derivatives themselves (i.e. elements of the gradient vector) may be said to comprise a collection of \(n\) multivariate functions. And if second partial derivatives also exist for all these multivariate functions (there will be \(n^2\)-square of these), then the resulting Jacobian matrix/determinant is called the **HESSIAN MATRIX/DETERMINANT**.

\[
f : \mathbb{R}^n \to \mathbb{R},
\]
\[
H(f) = \left( \begin{array}{ccc}
\frac{\partial^2 f}{\partial x_i \partial x_j} & \ldots & \frac{\partial^2 f}{\partial x_i \partial x_n} \\
\ldots & \ddots & \ldots \\
\frac{\partial^2 f}{\partial x_n \partial x_j} & \ldots & \frac{\partial^2 f}{\partial x_n \partial x_n}
\end{array} \right)
\] (88)

**HESSIAN DETERMINANT** \(H^{1/2}(\mathbf{x})\) = determinant of the Hessian matrix (confusingly, the term “The Hessian” may well refer to either the Hessian matrix or its determinant):

\[
D_{Hessian} (f) = \left| H(f) \right| 
\] (89)

Note that:

\[
H(f) = J(\nabla f) = J \left( \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right)
\]

\[
\Rightarrow D_{Hessian} (f) = D_{Jacobi} (\nabla f) = D_{Jacobi} \left( \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right)
\] (90)
Second Derivative Test for Multivariate Extremum:

1\textsuperscript{st}-order necessary condition (establish a \textit{multivariate} stationary point):

\[
(x^*, f(x^*)) \ni \nabla f(x^*) = \begin{bmatrix} \frac{\partial f}{\partial x_1} = 0 \\ \vdots \\ \frac{\partial f}{\partial x_n} = 0 \end{bmatrix} = 0
\]  
\hspace{1cm} (91)

For \( z = f(x, y) \), this reduces to establishing whether or not \( \frac{\partial f(x, y)}{\partial x} = \frac{\partial f(x, y)}{\partial y} = 0 \).

2\textsuperscript{nd}-order necessary condition:

\[
(x^*)^T H(f)(x^*) \begin{cases} \leq 0 & \text{Local Maximum} \\ \geq 0 & \text{Local Minimum} \end{cases}
\]  
\hspace{1cm} (92)

For \( z = f(x, y) \), this reduces to determining:

\[
\frac{\partial^2 f}{\partial x^2} (x^*)^2 + 2 \frac{\partial^2 f}{\partial x \partial y} (x^*) (y^*) + \frac{\partial^2 f}{\partial y^2} (y^*)^2 \begin{cases} \geq 0 & \text{Local Maximum} \\ \leq 0 & \text{Local Minimum} \end{cases}
\]  
\hspace{1cm} (93)

2\textsuperscript{nd} -order sufficient condition:

\[
(x^*)^T H(f)(x^*) \begin{cases} < 0 & \text{Local Maximum} \\ > 0 & \text{Local Minimum} \end{cases}
\]  
\hspace{1cm} (94)

For \( z = f(x, y) \), this reduces to determining:

\[
\frac{\partial^2 f}{\partial x^2} (x^*)^2 + 2 \frac{\partial^2 f}{\partial x \partial y} (x^*) (y^*) + \frac{\partial^2 f}{\partial y^2} (y^*)^2 \begin{cases} > 0 & \text{Local Minimum} \\ < 0 & \text{Local Maximum} \end{cases}
\]  
\hspace{1cm} (95)

If the Hessian has both positive and negative eigenvalues, then it’s a \textit{saddle point}; otherwise the test is inconclusive.
DIFFERENTIAL CALCULUS – IMPLICIT FUNCTION THEOREM

Let \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) be a multivariate, scalar-valued function:

- Create an **IMPLICIT FUNCTION** representation:
  \[
  y = f(x_1, \ldots, x_n) \iff F(y; x_1, \ldots, x_n) = 0
  \tag{96}
  \]

- Taking the total derivative yields:
  \[
  F(y; x_1, \ldots, x_n) = 0 \Rightarrow \frac{\partial F}{\partial y} dy + \sum_{j=1}^{n} \frac{\partial F}{\partial x_j} dx_j = 0
  \tag{97}
  \]

- Keep \( dy \neq 0 \) and let \( dx_j \neq 0 \), \( j \in \{1, \ldots, n\} \) one at a time:
  \[
  \forall j, \quad j = 1, \ldots, n, \\
  dx_k \left\{ \begin{array}{ll}
  \neq 0 & k = j \\
  = 0 & \text{otw.}
  \end{array} \right.
  \Rightarrow \frac{dy}{dx_j} \quad \text{other variables held constant}
  \equiv \frac{\partial y}{\partial x_j} = -\frac{\partial F/\partial x_j}{\partial F/\partial y}
  \tag{98}
  \]

- Note the proviso: \( \frac{\partial F}{\partial y} \neq 0 \).

- Hence the **IMPLICIT FUNCTION THEOREM**:
  \[
  F(y; x_1, \ldots, x_n) = 0 \Rightarrow \frac{\partial y}{\partial x_j} = -\frac{\partial F/\partial x_j}{\partial F/\partial y}, \quad j = 1, \ldots, n
  \tag{99}
  \]

- This could be extended to \( m > 1 \) **SIMULTANEOUS EQUATIONS** (each still in \( n > 1 \) variables), in other words, for a multivariate, vector-valued function \( f : \mathbb{R}^n \rightarrow \mathbb{R}^m \).