UNIVERSITÀ
DI SIENA
1240

Prof. Lonzi Marco

Lecture notes for the course

QUANTITATIVE METHODS FOR ECONOMIC APPLICATIONS

MATHEMATICS FOR ECONOMIC APPLICATIONS

Volume 2

Differential calculus for vector valued functions of a vector variable

AA. 2022/23

DIFFERENTIAL CALCULUS FOR VECTOR-VALUED FUNCTIONS OF A VECTOR VARIABLE

Let $\mathbb{R}^n$ be the n-dimension vector space, whose elements are n-tuples of real numbers, or vectors, resulting from the Cartesian product of $\mathbb{R}$ by itself n times.
Any function f has domain and range (codomain) contained in proper vector spaces, $\mathbb{R}^n$ and $\mathbb{R}^m$, with $n \geq 1$ and $m \geq 1$; both the independent and the dependent variable may then assume real or vector values. So we define:
$f : \mathbb{R} \to \mathbb{R}$, real function of a real variable;
$f : \mathbb{R} \to \mathbb{R}^n$, vector valued function of a real variable;
$f : \mathbb{R}^n \to \mathbb{R}$, real function of a vector variable;
$f : \mathbb{R}^n \to \mathbb{R}^m$, vector valued function of a vector variable.

TOPOLOGY IN $\mathbb{R}^n$
An element $\mathbb{X} = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$ is also called a point or a vector.
Given two points $\mathbb{X}, \mathbb{Y} \in \mathbb{R}^n$, $\mathbb{X} = (x_1, x_2, ..., x_n)$, $\mathbb{Y} = (y_1, y_2, ..., y_n)$ we have:
Definition 1 : The (Euclidean) distance between $\mathbb{X}$ and $\mathbb{Y}$ is given by the norm (or lenght) of their difference: $\mathrm{d}(\mathbb{X},\mathbb{Y}) = \|\mathbb{X} - \mathbb{Y}\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$.
Many are the distances that can be defined in $\mathbb{R}^n$, and here we will only use the Euclidean one.
Definition 2 : A neighbourhood of the point $\mathbb{X}_0 \in \mathbb{R}^n$ with radius ε is the set:
$\mathfrak{I}(\mathbb{X}_0, \varepsilon) = \{\mathbb{X} \in \mathbb{R}^n : \mathrm{d}(\mathbb{X}, \mathbb{X}_0) < \varepsilon\} = \{\mathbb{X} \in \mathbb{R}^n : \|\mathbb{X} - \mathbb{X}_0\| < \varepsilon\}$.
A neighbourhood $\mathfrak{I}(\mathbb{X}_0, \varepsilon)$ in $\mathbb{R}^2$ consists of the points inside a circle having $\mathbb{X}_0$ as center and ε as radius; a neighbourhood $\mathfrak{I}(\mathbb{X}_0, \varepsilon)$ in $\mathbb{R}^3$ consists of the points inside a sphere having $\mathbb{X}_0$ as center and ε as radius.
The topological definitions of the various types of points in $\mathbb{R}^n$ are similar to those given in $\mathbb{R}$; given a point $\mathbb{X}_0$ and a set $\mathbb{A} \subset \mathbb{R}^n$ we have the following:
Definition 3 : $\mathbb{X}_0$ is an accumulation (or limit) point of the set $\mathbb{A}$ if any neighborhood of $\mathbb{X}_0$ has a non-empty intersection with $\mathbb{A}$, different from the single point $\mathbb{X}_0$, i.e. if:
$\forall\, \varepsilon > 0 : \Big\{\mathfrak{I}(\mathbb{X}_0, \varepsilon)/\{\mathbb{X}_0\}\Big\} \cap \mathbb{A} \neq \emptyset$;
Definition 4 : $\mathbb{X}_0 \in \mathbb{A}$ is an isolated point of the set $\mathbb{A}$ if there exists at least a neighborhood of $\mathbb{X}_0$ which has no common points with $\mathbb{A}$, except for the point $\mathbb{X}_0$ itself, i.e. if: $\exists\, \varepsilon > 0 :\ \mathfrak{I}(\mathbb{X}_0, \varepsilon) \cap \mathbb{A} = \{\mathbb{X}_0\}$.
Definition 5 : $\mathbb{X}_0 \in \mathbb{A}$ is an interior point of the set $\mathbb{A}$ if there exists at least a neighborhood of $\mathbb{X}_0$ all contained in $\mathbb{A}$, i.e. if: $\exists\, \varepsilon > 0 :\ \mathfrak{I}(\mathbb{X}_0, \varepsilon) \subset \mathbb{A}$.
Definition 6 : $\mathbb{X}_0 \in \mathcal{C}(\mathbb{A})$ is an external point of the set $\mathbb{A}$ if $\mathbb{X}_0$ is an interior point of the set $\mathcal{C}(\mathbb{A})$, the complementary set of $\mathbb{A}$.
Definition 7 : $\mathbb{X}_0$ is a boundary point of the set $\mathbb{A}$ if every neighborhood of $\mathbb{X}_0$ has non-empty intersection with both $\mathbb{A}$ and $\mathcal{C}(\mathbb{A})$, i.e. if:
$\forall\, \varepsilon > 0 :\ \mathfrak{I}(\mathbb{X}_0, \varepsilon) \cap \mathbb{A} \neq \emptyset$ **and** $\mathfrak{I}(\mathbb{X}_0, \varepsilon) \cap \mathcal{C}(\mathbb{A}) \neq \emptyset$.
To be an isolated point or an interior point of $\mathbb{A}$, $\mathbb{X}_0$ must belong to $\mathbb{A}$; this is not required to be an accumulation point or a boundary point.
If $\mathbb{X}_0$ is an isolated point of $\mathbb{A}$ then $\mathbb{X}_0$ is also a boundary point of $\mathbb{A}$; if $\mathbb{X}_0$ is an interior point of $\mathbb{A}$ then $\mathbb{X}_0$ is also an accumulation point of $\mathbb{A}$.

From the topological definitions of point, the topological definitions for sets follow:

Definition 8 : A set $\mathbb{A} \subset \mathbb{R}^n$ is an open set if all its points are interior points.
So a set $\mathbb{A} \subset \mathbb{R}^n$ is open if none of its boundary points belongs to it.
Definition 9 : A set $\mathbb{A} \subset \mathbb{R}^n$ is a closed set if its complementary set is open.
So a set $\mathbb{A} \subset \mathbb{R}^n$ is closed if all its boundary points (or all its accumulation points) belong to it.
An **interval** in $\mathbb{R}^n$ is given by the Cartesian product of n intervals of $\mathbb{R}$.
If all intervals are closed : $[x_i^a, x_i^b]$, the interval $\prod_{i=1}^{n}[x_i^a, x_i^b]$ will be a closed interval, if all the intervals are open : $]x_i^a, x_i^b[$, the interval $\prod_{i=1}^{n}]x_i^a, x_i^b[$ will be an open interval.
Definition 10 : A set $\mathbb{A} \subset \mathbb{R}^n$ is said to be bounded if there exists a neighborhood $\mathfrak{J}(\mathbb{X}_0, \varepsilon)$, having center in an appropriate point $\mathbb{X}_0$ and with appropriate radius ε, such that $\mathbb{A} \subset \mathfrak{J}(\mathbb{X}_0, \varepsilon)$.
Definition 11 : A closed and bounded set $\mathbb{A} \subset \mathbb{R}^n$ is also said a compact set.

VECTOR VALUED FUNCTIONS OF A REAL VARIABLE $f : \mathbb{R} \to \mathbb{R}^n$

Consider a vector $\mathbb{X} \in \mathbb{R}^n$ each of whose components is a function $\mathbb{R} \to \mathbb{R}$ of the real variable t. We write $\mathbb{X}(t) = (x_1(t), x_2(t), ..., x_n(t))$ or $f(t) = (f_1(t), f_2(t), ..., f_n(t))$ to indicate such a function. Each function $x_i(t) = f_i(t)$ is a function $f_i : \mathbb{R} \to \mathbb{R}$, $1 \le i \le n$.
The function $f : t \to \mathbb{X}(t)$ is called a vector valued function of a real variable, as the image of the real variable t is a vector $\mathbb{X}(t) \in \mathbb{R}^n$. Such type of functions are also called curves in $\mathbb{R}^n$. For this type of functions graph means also codomain (range). The resulting line is also said the curve support. The field of existence of such a function is given by the intersection of the fields of existence of the n functions $x_i(t) = f_i(t)$.

Example 1 : The graph of the function $f_1 : \mathbb{R} \to \mathbb{R}^2, t \to (\cos t, \sin t)$ is the Cartesian circumference $x^2 + y^2 = 1$, having center in the origin and radius equal to 1. The circumference is traveled countless counterclockwise for $-\infty < t < +\infty$.
The graph of the function $f_2 : [0, 2\pi[\to \mathbb{R}^2, t \to (\cos t, \sin t)$ is the same circumference covered only once counterclockwise starting, for $t = 0$, from the point $(1, 0)$.
The graph of the function $f_3 : [0, 2\pi[\to \mathbb{R}^2, t \to (\sin t, \cos t)$ is the same circumference covered only once clockwise starting, for $t = 0$, from the point $(0, 1)$.
As can be seen from these examples, it is not sufficient to describe geometrically the points of a graph, but it must also be considered how, how many times and in which direction such a graph is covered.

Example 2 : The graph of the function $f_1 : \mathbb{R} \to \mathbb{R}^2, t \to (t, t^2)$ is the same of the cartesian parabola $y = x^2$.
The graph of the function $f_2 : \mathbb{R} \to \mathbb{R}^2, t \to (t^2, t^4)$ is only the right side of the same parabola, covered twice, from right to left for $t < 0$ and from left to right for $t \ge 0$.
The graph of the function $f_3 : \mathbb{R} \to \mathbb{R}^2, t \to (\sin t, \sin^2 t)$ is the part of the parabola $y = x^2$ included between the point $(-1, 1)$ and the point $(1, 1)$, covered innumerable times to and from between the extreme points.

Example 3 : The graph of the function $f : \mathbb{R} \to \mathbb{R}^3, t \to (\cos t, \sin t, t)$ is an helix that wraps up itself, going up if $t \ge 0$, going down if $t < 0$, along the axis of the third variable.

Suppose we have a function in cartesian form $f:\mathbb{R}\to\mathbb{R}, x\to y=f(x)$; if there exists a function $F:\mathbb{R}\to\mathbb{R}^2, t\to(x(t),y(t))$, such that $y(t)=f(x(t)), \forall\, t\in D_f$, $F(t)$ is called a parametric version for $y=f(x)$.

Example 4 : The function $f:\mathbb{R}\to\mathbb{R}^2, t\to(x_0+\rho\cos t, y_0+\rho\sin t)$ is a parametric version for the circumference with center (x_0,y_0) and radius $\rho>0$.

Example 5 : The function $f:\mathbb{R}\to\mathbb{R}^2, t\to(a\cos t, b\sin t)=(x(t),y(t)), a,b\in\mathbb{R}_+$ is a parametric version for the ellipse with center $(0,0)$, axes parallel to the coordinate axes, and semiaxes a and b. In fact : $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=\dfrac{(a\cos t)^2}{a^2}+\dfrac{(b\sin t)^2}{b^2}=1$.
$f:\mathbb{R}\to\mathbb{R}^2, t\to(x_0+a\cos t, y_0+b\sin t)$ is instead a parametric version for the ellipse with center (x_0,y_0), axes parallel to axes x and y, and semiaxes a and b.

LIMITS AND CONTINUITY FOR FUNCTIONS $f:\mathbb{R}\to\mathbb{R}^n$

Given $f:\mathbb{R}\to\mathbb{R}^n, t\to f(t)=(f_1(t),f_2(t),...,f_n(t))$, and t_0 an accumulation point of D_f.

Definition 12 : $\lim\limits_{t\to t_0}f(t)=\lim\limits_{t\to t_0}(f_1(t),f_2(t),...,f_n(t))$ is defined as:

$\lim\limits_{t\to t_0}f(t)=\left(\lim\limits_{t\to t_0}f_1(t),\lim\limits_{t\to t_0}f_2(t),...,\lim\limits_{t\to t_0}f_n(t)\right)$, that is the limit of a vector is defined as a vector having as components the limits, for $t\to t_0$, of each of the components of $f(t)$.

To calculate the limit of a function $f:\mathbb{R}\to\mathbb{R}^n$ consists therefore in calculating n limits for functions $\mathbb{R}\to\mathbb{R}$.

Definition 13 : We have $\lim\limits_{t\to t_0}f(t)=\mathfrak{L}\in\mathbb{R}^n$, with $t_0\in\mathbb{R}$, if:
$\forall\,\varepsilon>0\;\exists\,\delta(\varepsilon):0<|t-t_0|<\delta(\varepsilon)\Rightarrow\|f(t)-\mathfrak{L}\|<\varepsilon$.

Definition 14 : We have $\lim\limits_{t\to-\infty}f(t)=\mathfrak{L}\in\mathbb{R}^n$ if:
$\forall\,\varepsilon>0\;\exists\,\delta(\varepsilon):t<\delta(\varepsilon)\Rightarrow\|f(t)-\mathfrak{L}\|<\varepsilon$.

Definition 15 : We have $\lim\limits_{t\to+\infty}f(t)=\mathfrak{L}\in\mathbb{R}^n$ if:
$\forall\,\varepsilon>0\;\exists\,\delta(\varepsilon):t>\delta(\varepsilon)\Rightarrow\|f(t)-\mathfrak{L}\|<\varepsilon$.

We shall not deal here with the concept of infinite limit.

If t_0 is an accumulation point of D_f, and $t_0\in D_f$, we have the following

Definition 16 : The function $f:\mathbb{R}\to\mathbb{R}^n, t\to f(t)=(f_1(t),f_2(t),...,f_n(t))$ is continuous at the point t_0 if $\lim\limits_{t\to t_0}f(t)=f(t_0)$.

This definition corresponds to requiring that:
$\lim\limits_{t\to t_0}f_1(t)=f_1(t_0), \lim\limits_{t\to t_0}f_2(t)=f_2(t_0),..., \lim\limits_{t\to t_0}f_n(t)=f_n(t_0)$,
that is that each of the components $f_i(t)$ be continuous at point t_0.

In metrical form, $f(t)$ is continuous at t_0 if:
$\forall\,\varepsilon>0\;\exists\,\delta(\varepsilon):|t-t_0|<\delta(\varepsilon)\Rightarrow\|f(t)-f(t_0)\|<\varepsilon$.

If t_0 is an isolated point for D_f the function is defined as a continuous one at t_0.

DERIVATIVE FOR FUNCTIONS $f:\mathbb{R}\to\mathbb{R}^n$

Given $f:\mathbb{R}\to\mathbb{R}^n, t\to f(t)=(f_1(t),f_2(t),...,f_n(t))=\mathbb{X}(t)=(x_1(t),x_2(t),...,x_n(t))$, and considering point $t_0\in D_f$, we define the derivative of the function $f(t)$ at point t_0.

Definition 17 : The derivative of the function $f(t)$ at t_0 is given by the limit:

$\lim\limits_{t\to t_0}\dfrac{f(t)-f(t_0)}{t-t_0}=f'(t_0)=\mathbb{X}'(t_0)\in\mathbb{R}^n$, provided that this limit exists and is finite.

If a function has derivative at t_0 we say that the function is differentiable at t_0.

Since $\dfrac{1}{t-t_0} \in \mathbb{R}$ while $(f(t)-f(t_0)) \in \mathbb{R}^n$:

$$\lim_{t\to t_0}\frac{f(t)-f(t_0)}{t-t_0} = \lim_{t\to t_0}\frac{1}{t-t_0}\cdot(f(t)-f(t_0)) =$$
$$= \lim_{t\to t_0}\frac{1}{t-t_0}\cdot(f_1(t)-f_1(t_0), f_2(t)-f_2(t_0), ..., f_n(t)-f_n(t_0)) =$$
$$= \left(\lim_{t\to t_0}\frac{f_1(t)-f_1(t_0)}{t-t_0}, \lim_{t\to t_0}\frac{f_2(t)-f_2(t_0)}{t-t_0}, ..., \lim_{t\to t_0}\frac{f_n(t)-f_n(t_0)}{t-t_0}\right) =$$
$$= (f_1'(t_0), f_2'(t_0), ..., f_n'(t_0)).$$

Therefore the function $f : \mathbb{R} \to \mathbb{R}^n$ has the derivative at point t_0 if each of its components $f_i(t)$ has the derivative at point t_0.

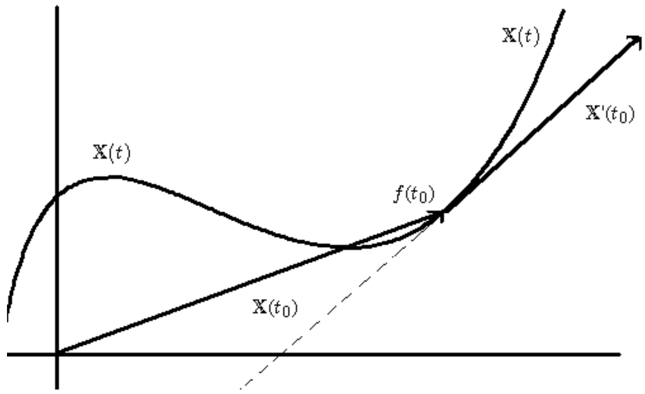
For practical calculus, the result we have found is that the derivative of a vector is still a vector, whose components are the derivatives of its components.

Example 6 : Given $f : \mathbb{R} \to \mathbb{R}^2, t \to (t, t^2)$, we have $f'(t) = (1, 2t)$.
Given $f : \mathbb{R} \to \mathbb{R}^2, t \to (t^2, t^4)$, we have $f'(t) = (2t, 4t^3)$.
Given $f : \mathbb{R} \to \mathbb{R}^3, t \to f(t) = (\cos t, \sin t, t)$, we have $f'(t) = (-\sin t, \cos t, 1)$.

Theorem 1 : If $f(t)$ is differentiable at t_0, then $f(t)$ is continuous at t_0.
Proof : We must verify that $\lim_{t\to t_0} f(t) = f(t_0)$ or that $\lim_{t\to t_0}(f(t)-f(t_0)) = \mathbb{O}$.

But $\lim_{t\to t_0}(f(t)-f(t_0)) = \lim_{t\to t_0}(t-t_0)\cdot\dfrac{f(t)-f(t_0)}{t-t_0} = 0\cdot f'(t_0) = \mathbb{O}$, i.e. the thesis.●

The definition of a differentiable function at t_0 can also be written as:
$\lim_{t\to t_0}\dfrac{\mathbb{X}(t)-\mathbb{X}(t_0)}{t-t_0} = \mathbb{X}'(t_0)$; the vector $\mathbb{X}'(t_0) \in \mathbb{R}^n$ is called the tangent vector to the curve $\mathbb{X}(t)$ at t_0.
We need $\mathbb{X}'(t_0) \neq \mathbb{O}$, null vector, to get the tangent vector $\mathbb{X}'(t_0)$.



The equation of the tangent line to the curve $\mathbb{X}(t)$ at the point $f(t_0) = \mathbb{X}(t_0)$ is the function:
$r : \mathbb{R} \to \mathbb{R}^n, t \to r(t) = \mathbb{X}(t_0) + t\cdot\mathbb{X}'(t_0)$ or
$t \to (f_1(t_0), f_2(t_0), ..., f_n(t_0)) + t\cdot(f_1'(t_0), f_2'(t_0), ..., f_n'(t_0))$.

Example 7 : Given $f : \mathbb{R} \to \mathbb{R}^3$, $f(t) = (t^2, t^3, e^t)$ and $t_0 = 1$, we get $f(t_0) = (1, 1, e)$, $f'(t) = (2t, 3t^2, e^t)$ and so $f'(t_0) = (2, 3, e)$. The equation of the tangent line to the curve at $t_0 = 1$ is then: $t \to (1, 1, e) + t\cdot(2, 3, e) = (1+2t, 1+3t, e+et)$.

Given $\mathcal{D}f(t)$ the derivative of $f(t)$, we have the following:

Theorem 2 : If $f(t)$ and $g(t)$ are differentiable functions at t, then:
$\mathcal{D}(f(t)\pm g(t)) = \mathcal{D}f(t)\pm\mathcal{D}g(t)$.
That is, the derivative of a sum (of a difference) of functions is equal to the sum (the difference) of the derivatives.
As for the product, since now images are vectors, we must consider two cases: the product of a scalar and a vector and the scalar (or dot) product of two vectors.
The following theorems are valid:

Theorem 3 : If $f : \mathbb{R} \to \mathbb{R}^n, t \to f(t) = (f_1(t), f_2(t), ..., f_n(t))$ and $g : \mathbb{R} \to \mathbb{R}, t \to g(t)$ are differentiable functions at t, then $\mathcal{D}(g(t) \cdot f(t)) = \mathcal{D}g(t) \cdot f(t) + g(t) \cdot \mathcal{D}f(t)$.

Example 8 : Given $f : \mathbb{R} \to \mathbb{R}^3, f(t) = \left(\cos t, \sin t, e^t\right)$ and $g : \mathbb{R} \to \mathbb{R}, g(t) = t^2$. Then:
$g(t) \cdot f(t) = t^2 \left(\cos t, \sin t, e^t\right) = \left(t^2 \cos t, t^2 \sin t, t^2 e^t\right)$. And so:
$\mathcal{D}(g(t) \cdot f(t)) = \left(2t \cos t - t^2 \sin t, 2t \sin t + t^2 \cos t, 2t\, e^t + t^2 e^t\right) =$
$= 2t \left(\cos t, \sin t, e^t\right) + t^2 \left(- \sin t, \cos t, e^t\right) = \mathcal{D}g(t) \cdot f(t) + g(t) \cdot \mathcal{D}f(t)$.

Theorem 4 : If f and g : $\mathbb{R} \to \mathbb{R}^n$ are differentiable functions at t, then:
$\mathcal{D}(f(t) \cdot g(t)) = \mathcal{D}f(t) \cdot g(t) + f(t) \cdot \mathcal{D}g(t)$.

Example 9 : Given $f : \mathbb{R} \to \mathbb{R}^3, f(t) = \left(t, \sin t, e^t\right)$ and $g : \mathbb{R} \to \mathbb{R}^3, g(t) = \left(t^2, e^t, t^3\right)$, then $f(t) \cdot g(t) = \left(t, \sin t, e^t\right) \cdot \left(t^2, e^t, t^3\right) = t^3 + e^t \sin t + t^3 e^t$. And so:
$\mathcal{D}(f(t) \cdot g(t)) = 3t^2 + e^t \sin t + e^t \cos t + 3t^2 e^t + t^3 e^t =$
$= \left(1, \cos t, e^t\right) \cdot \left(t^2, e^t, t^3\right) + \left(t, \sin t, e^t\right) \cdot \left(2t, e^t, 3t^2\right) = \mathcal{D}f(t) \cdot g(t) + f(t) \cdot \mathcal{D}g(t)$.

In both cases, therefore, independently from the product, the rule is that the derivative of a product is the sum of two terms, each of which is the product of the derivative of a factor and the non-derivative of the other.

As far as composite functions are concerned, at this stage we can only deal with this case:
Theorem 5 : If $g : \mathbb{R} \to \mathbb{R}, t \to g(t)$ is a differentiable function at t and $f : \mathbb{R} \to \mathbb{R}^n$, $t \to f(t) = (f_1(t), f_2(t), ..., f_n(t))$ is a differentiable function at $g(t)$, then the composite function: $\mathbb{R} \xrightarrow{g} \mathbb{R} \xrightarrow{f} \mathbb{R}^n, t \to f(g(t)) = (f_1(g(t)), f_2(g(t)), ..., f_n(g(t)))$ is differentiable at t and: $\mathcal{D}(f(g(t))) = f'(g(t)) \cdot g'(t)$.
Proof : As $\mathcal{D}f(g(t)) = (\mathcal{D}f_1(g(t)),\ \mathcal{D}f_2(g(t)), ..., \mathcal{D}f_n(g(t)))$, we get:
$\mathcal{D}f(g(t)) = (f_1'(g(t)) \cdot g'(t), f_2'(g(t)) \cdot g'(t), ..., f_n'(g(t)) \cdot g'(t)) = f'(g(t)) \cdot g'(t)$.•

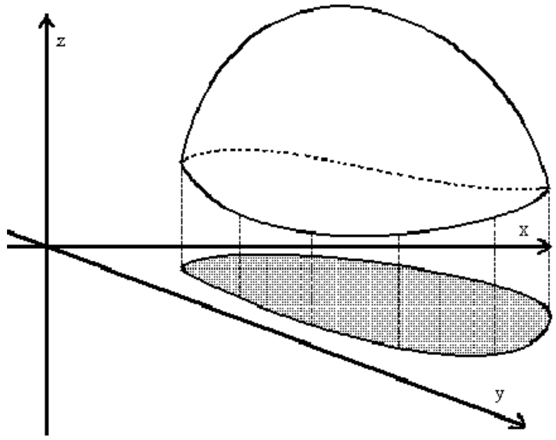
Example 10 : Given $g : \mathbb{R} \to \mathbb{R}, g(t) = t^2$, and $f : \mathbb{R} \to \mathbb{R}^3, f(t) = \left(\cos t, \sin t, e^t\right)$, then:
$f(g(t)) = \left(\cos t^2, \sin t^2, e^{t^2}\right)$. And so:
$\mathcal{D}(f(g(t))) = \left(- \sin t^2 \cdot 2t, \cos t^2 \cdot 2t, e^{t^2} \cdot 2t\right) =$
$= \left(- \sin t^2, \cos t^2, e^{t^2}\right) \cdot (2t) = f'(g(t)) \cdot g'(t)$.

For a vector valued function $\mathbb{R} \to \mathbb{R}^n$ it is meaningless to speak about reciprocal and quotient derivative.

REAL FUNCTIONS OF A VECTOR VARIABLE $f : \mathbb{R}^n \to \mathbb{R}$

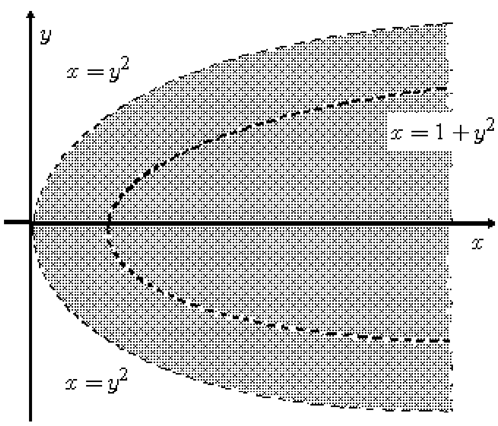
Since the domain is now given by $\mathbb{R}^n$ or by an appropriate subset of its, the independent variable is a vector and so functions $f : \mathbb{R}^n \to \mathbb{R}$ are also called functions of several variables.

These functions will be represented in the form $y = f(\mathbb{X}) = f(x_1, x_2, ..., x_n), x_i \in \mathbb{R}$.
We also use notations like $z = f(x, y)$ for $f : \mathbb{R}^2 \to \mathbb{R}$, while it is also commonly used $w = f(x, y, z)$ for $f : \mathbb{R}^3 \to \mathbb{R}$.



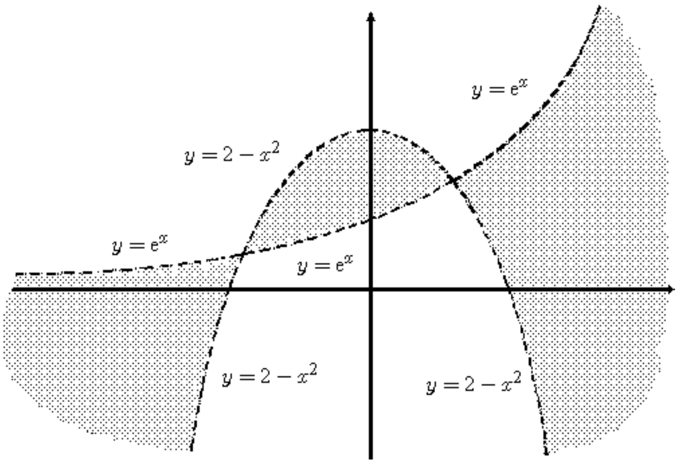
The graph of a function $f : \mathbb{R}^n \to \mathbb{R}, y = f(\mathbb{X}) = f(x_1, x_2, ..., x_n)$ is defined as the subset of $\mathbb{R}^n \times \mathbb{R} = \mathbb{R}^{n+1}$ consisting of points $\{(x_1, x_2, ..., x_n, y), y = f(x_1, x_2, ..., x_n)\}$.
If $n = 2$, being the domain a subset of the real plane, the graph of $z = f(x, y)$ is a two-dimensional surface lying in $\mathbb{R}^3$ that can be represented as in the previous figure, crushing, in perspective, the plane of the independent variables x and y.
If $n > 2$ we say that the graph is an n-dimensional hypersurface in $\mathbb{R}^{n+1}$.
The existence field of a function of two variables is a subset of $\mathbb{R}^2$, and can therefore be represented graphically.

Example 11 : Let us consider the function $f : \mathbb{R}^2 \to \mathbb{R}, f(x, y) = \dfrac{1}{\log(x - y^2)}$.
Let's determine and represent its existence field.
We should put $\begin{cases} x - y^2 > 0 \\ \log(x - y^2) \neq 0 \end{cases}$ and so $\begin{cases} x > y^2 \\ x - y^2 \neq 1 \Rightarrow x \neq 1 + y^2 \end{cases}$.
The existence field of the function is represented by the dark region of the next figure, formed by the points on the right of the parabola $x = y^2$, removed the points of the parabola $x = 1 + y^2$.



Example 12 : Let us consider the function $f : \mathbb{R}^2 \to \mathbb{R}, f(x, y) = \log\left(\dfrac{y - e^x}{2 - x^2 - y}\right)$.
Let's determine and represent its existence field.

We should put $\dfrac{y-e^x}{2-x^2-y}>0$, which is satisfied if:
$\begin{cases} y-e^x>0 \\ 2-x^2-y>0 \end{cases}$ or if $\begin{cases} y-e^x<0 \\ 2-x^2-y<0 \end{cases}$, respectively equivalent to $\begin{cases} y>e^x \\ y<2-x^2 \end{cases}$ and to $\begin{cases} y<e^x \\ y>2-x^2 \end{cases}$.
The existence field of the given function is represented by the union of the dark regions, the central one is the solution of the first system, the remaining two, the one on the left and the one on the right, represent the solution of the second. The edges of the areas are dashed as they do not belong to the field of existence, since the inequalities are tight.



Example 13 : Let us consider the function $f:\mathbb{R}^2 \to \mathbb{R}, f(x,y)=\dfrac{1}{\sin(x-y)}$.
Let's determine its existence field.
We should put $\sin(x-y)\neq 0$, or: $x-y\neq k\pi \Rightarrow y\neq x-k\pi, \forall\, k\in\mathbb{Z}$.
We must therefore remove from the plane $\mathbb{R}^2$ all the straight lines, parallel to the bisector $y=x$, with equation $y=x-k\pi,\, k\in\mathbb{Z}$.

Definition 18 : $f:\mathbb{R}^n\to\mathbb{R}, y=f(\mathbb{X})$ is a bounded function on $\mathbb{A}\subseteq\mathbb{R}^n$ if there are two values, $y_1,y_2\in\mathbb{R}$, such that : $y_1\le f(\mathbb{X})\le y_2, \forall\,\mathbb{X}\in\mathbb{A}$.
Definition 19 : A point $\mathbb{X}_0\in\mathbb{R}^n$ is called a relative maximum (minimum) point if there exists a neighborhood $\mathfrak{I}(\mathbb{X}_0,\varepsilon)$ for which $f(\mathbb{X})\le f(\mathbb{X}_0)$ $(f(\mathbb{X})\ge f(\mathbb{X}_0))\ \forall\,\mathbb{X}\in\mathfrak{I}(\mathbb{X}_0,\varepsilon)$.

LIMITS FOR FUNCTIONS $f:\mathbb{R}^n\to\mathbb{R}$
Let $\mathbb{X}_0\in\mathbb{R}^n$ be an accumulation point for the domain of the function $y=f(x_1,x_2,...,x_n)$.
$\mathbb{X}=(x_1,x_2,...,x_n)$, $\mathbb{X}_0=\left(x_1^0,x_2^0,...,x_n^0\right)$, and $l\in\mathbb{R}$.
Definition 20 : $\lim\limits_{\mathbb{X}\to\mathbb{X}_0} f(x_1,x_2,...,x_n)=l$ if :
$\forall\,\varepsilon>0\ \exists\,\delta(\varepsilon): 0<\|\mathbb{X}-\mathbb{X}_0\|<\delta\Rightarrow|f(\mathbb{X})-l|<\varepsilon$;
Definition 21 : $\lim\limits_{\mathbb{X}\to\mathbb{X}_0} f(x_1,x_2,...,x_n)=+\infty$ if :
$\forall\,\varepsilon\ \exists\,\delta(\varepsilon): 0<\|\mathbb{X}-\mathbb{X}_0\|<\delta\Rightarrow f(\mathbb{X})>\varepsilon$;
Definition 22 : $\lim\limits_{\mathbb{X}\to\mathbb{X}_0} f(x_1,x_2,...,x_n)=-\infty$ if :
$\forall\,\varepsilon\ \exists\,\delta(\varepsilon): 0<\|\mathbb{X}-\mathbb{X}_0\|<\delta\Rightarrow f(\mathbb{X})<\varepsilon$.

For the limits of functions $f:\mathbb{R}^n\to\mathbb{R}$ the so called "uniqueness of the limit theorem", "permanence of the sign theorem" and "comparison theorem" apply, as for functions $\mathbb{R}\to\mathbb{R}$.

When we write $\mathbb{X} \to \mathbb{X}_0$, that is that $\mathbb{X}$ takes values closer and closer to $\mathbb{X}_0$, it means that $\mathbb{X}$ belongs to a neighborhood with center $\mathbb{X}_0$ and radius δ; as $\mathbb{X}_0 \in \mathbb{R}^n$, we use, in the limit definition, the Euclidean norm $\|\mathbb{X} - \mathbb{X}_0\|$.
When we say that $(x, y) \to (x_0, y_0)$ we mean that this should be through all possible paths (or curves) continuous (i.e. without gaps or jumps) that lead from (x, y) to (x_0, y_0). Limit must always be the same for every path used to say that the limit exists. If multiple paths lead to different results, the conclusion will be that the limit does not exist.

Example 14 : Let us compute $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x\,y^2}{x^2 + y^2}$. The function is defined in $\mathbb{R}^2 \setminus \{(0,0)\}$.
Since the numerator is an infinitesimal polynomial of the third degree, while the denominator is of the second degree, we assume that the limit is 0 and we try to verify this result using the definition. It must result $\displaystyle\left|\frac{x\,y^2}{x^2 + y^2} - 0\right| < \varepsilon$ in a neighborhood of $(0,0)$.
But $\displaystyle\left|\frac{x\,y^2}{x^2 + y^2}\right| = |x| \cdot \frac{y^2}{x^2 + y^2} \le |x|$, as $\displaystyle\frac{y^2}{x^2 + y^2} \le 1, \forall\,(x, y) \in \mathbb{R}^2 \setminus \{(0,0)\}$.
Then it is sufficient to impose $|x| < \varepsilon$ to obtain $\displaystyle\left|\frac{x\,y^2}{x^2 + y^2}\right| < \varepsilon$.
But $|x| < \varepsilon \Leftrightarrow -\varepsilon < x < \varepsilon$, that is we get a vertical strip inside which it is always possible to find a neighborhood of $(0,0)$: just take $\delta < \varepsilon$.
So it is verified that $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x\,y^2}{x^2 + y^2} = 0$.

Example 15 : Given $f(x, y) = \begin{cases} x \sin \dfrac{1}{y} & : y \neq 0 \\ 0 & : y = 0 \end{cases}$, let us compute $\displaystyle\lim_{(x,y)\to(0,0)} f(x, y)$.
As x approaches 0, while $\sin \dfrac{1}{y}$ is bounded, also now let us assume that the limit is 0, consistent, moreover, with the behavior of the function along the x axis. Then let us check that: $\displaystyle\left|x \sin \frac{1}{y} - 0\right| < \varepsilon$ in a neighborhood of $(0,0)$. But $\displaystyle\left|x \sin \frac{1}{y}\right| = |x| \cdot \left|\sin \frac{1}{y}\right| \le |x|$, as $\displaystyle\left|\sin \frac{1}{y}\right| \le 1$. Imposing $|x| < \varepsilon$ we find the solution as in the previous example. For the x axis points we have instead: $|f(x, y) - 0| = |0 - 0| < \varepsilon$, which is always verified.
So $\displaystyle\lim_{(x,y)\to(0,0)} f(x, y) = 0$.

Example 16 : Let us compute $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2}{x^2 + y^2}$. As $(x, y) \to (0,0)$, let us use as approaching paths the straight lines passing through the origin, whose equation is $y = m\,x$.
Studying the limit along these paths with the substitution $y = m\,x$, we get:
$\displaystyle\lim_{x\to 0} \frac{x^2}{x^2 + m^2\,x^2} = \lim_{x\to 0} \frac{1}{1 + m^2} = \frac{1}{1 + m^2}$. The result depends on m, varying with the function line used. So $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2}{x^2 + y^2}$ does not exist.

Example 17 : Let us compute $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2}{x^2 + y^4}$. Operating as in the previous example, given $y = m\,x$, we compute $\displaystyle\lim_{x\to 0} \frac{x^2}{x^2 + m^4\,x^4} = \lim_{x\to 0} \frac{1}{1 + m^4\,x^2} = 1$. Then moving along any line passing through the origin the limit is 1. However is not strictly correct to deduct from this that $\lim\limits_{(x,y)\to(0,0)} \dfrac{x^2}{x^2+y^4} = 1$.

In fact, if we use different paths to get closer to $(0,0)$, such as parabolas like $x = k\,y^2$, studing the limit with these restrictions, we have:

$\lim\limits_{y\to 0} \dfrac{k^2\,y^4}{k^2\,y^4+y^4} = \lim\limits_{y\to 0} \dfrac{k^2}{k^2+1} = \dfrac{k^2}{k^2+1}$.

Varying the parabolas, the result is different, and then $\lim\limits_{(x,y)\to(0,0)} \dfrac{x^2}{x^2+y^4}$ does not exist.

ITERATED LIMITS

We should note that to calculate $\lim\limits_{(x,y)\to(x_0,y_0)} f(x,y)$, it is incorrect to calculate the two limits: $\lim\limits_{x\to x_0} \lim\limits_{y\to y_0} f(x,y)$ and $\lim\limits_{y\to y_0} \lim\limits_{x\to x_0} f(x,y)$. These two limits, called iterated limits, consist in the successive calculation of two limits of functions of a single variable, holding the other as a constant.

Should both exist and be equal, however, this would not allow us to conclude anything about $\lim\limits_{(x,y)\to(x_0,y_0)} f(x,y)$.

Even if $\lim\limits_{x\to x_0} \lim\limits_{y\to y_0} f(x,y) = \lim\limits_{y\to y_0} \lim\limits_{x\to x_0} f(x,y) = l$, it is incorrect to attribute the resulting value l to $\lim\limits_{(x,y)\to(x_0,y_0)} f(x,y)$.

Example 18 : We have already seen that $\lim\limits_{(x,y)\to(0,0)} \dfrac{x^2}{x^2+y^2}$ does not exist. Moreover:

$\lim\limits_{x\to 0} \lim\limits_{y\to 0} \dfrac{x^2}{x^2+y^2} = \lim\limits_{x\to 0} \dfrac{x^2}{x^2} = \lim\limits_{x\to 0} 1 = 1$ and

$\lim\limits_{y\to 0} \lim\limits_{x\to 0} \dfrac{x^2}{x^2+y^2} = \lim\limits_{y\to 0} \dfrac{0}{y^2} = \lim\limits_{y\to 0} 0 = 0$,

consistently with the non-existence of the limit.

Example 19 : We have already seen that for the function $f(x,y) = \begin{cases} x \sin \dfrac{1}{y} & : y \neq 0 \\ 0 & : y = 0 \end{cases}$ it results $\lim\limits_{(x,y)\to(0,0)} f(x,y) = 0$.

Yet we have that: $\lim\limits_{y\to 0} \lim\limits_{x\to 0} x \sin \dfrac{1}{y} = \lim\limits_{y\to 0} 0 = 0$, whereas $\lim\limits_{x\to 0} \lim\limits_{y\to 0} x \sin \dfrac{1}{y}$ does not exist. In fact the existence of $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ does not imply the existence of the limit along all possible paths from (x,y) to $(0,0)$.

Example 20 : Using the iterated limits for $\lim\limits_{(x,y)\to(0,0)} \dfrac{x\,y}{x^2+y^2}$, we get:

$\lim\limits_{x\to 0} \lim\limits_{y\to 0} \dfrac{x\,y}{x^2+y^2} = \lim\limits_{x\to 0} \dfrac{0}{x^2} = \lim\limits_{x\to 0} 0 = 0$, and

$\lim\limits_{y\to 0} \lim\limits_{x\to 0} \dfrac{x\,y}{x^2+y^2} = \lim\limits_{y\to 0} \dfrac{0}{y^2} = \lim\limits_{y\to 0} 0 = 0$.

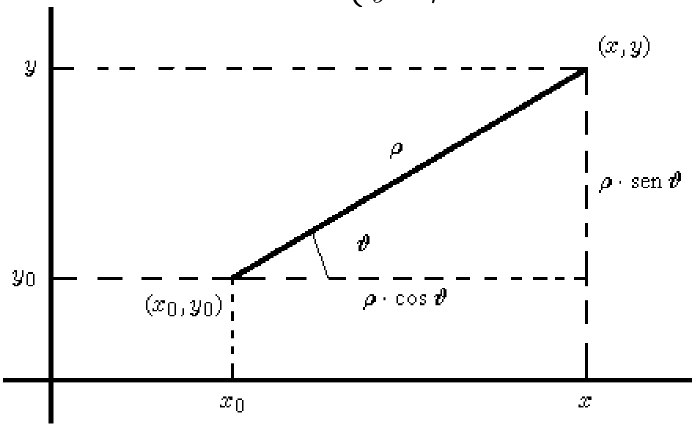
If we use the straight lines passing through the origin $y = m\,x$, we get instead $\lim\limits_{x\to 0} \dfrac{m\,x^2}{x^2 + m^2\,x^2} = \dfrac{m}{1+m^2}$, and therefore the given limit does not exist, even though the two iterated limits exist and have the same value.

LIMITS FOR FUNCTIONS $f:\mathbb{R}^2 \to \mathbb{R}$ USING POLAR COORDINATES

It may be useful, but only for functions of two variables, to change from Cartesian to polar coordinates to calculate a limit. Each point $(x, y) \in \mathbb{R}^2$ can be expressed, referred to a given point (x_0, y_0), as: $\begin{cases} x = x_0 + \rho\cos\vartheta \\ y = y_0 + \rho\sin\vartheta \end{cases}$.

These are the polar coordinates of the point (x, y) compared to the point (x_0, y_0).

If $(x_0, y_0) = (0,0)$, we get, as a special case, $\begin{cases} x = \rho\cos\vartheta \\ y = \rho\sin\vartheta \end{cases}$.



Operating the substitution, we have:

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y) \Rightarrow \lim_{\rho\to 0} f(\rho\cos\vartheta, \rho\sin\vartheta) = \lim_{\rho\to 0} F(\rho,\vartheta).$$

The limit as $(x, y) \to (x_0, y_0)$ becomes a limit in the single variable ρ, as $\rho \to 0$, since $\rho = \|(x,y) - (x_0,y_0)\|$.

But the value of the limit should not depend on the particular path used, and this requires that $\lim\limits_{\rho\to 0} F(\rho,\vartheta)$, if it exists, should not depend on the particular direction, i.e. from ϑ; we say that convergence (or divergence) of $\lim\limits_{\rho\to 0} F(\rho,\vartheta)$ must be uniform with respect to ϑ, i.e. that δ depends only on ε, not on ϑ.

We have, using polar coordinates, the following limit definitions:

Definition 23 : It is $\lim\limits_{\rho\to 0} F(\rho,\vartheta) = l \in \mathbb{R}$ if

$\forall\, \varepsilon > 0\; \exists\, \delta(\varepsilon) : \forall\, \vartheta, 0 < \rho < \delta \Rightarrow |F(\rho,\vartheta) - l| < \varepsilon$;

Definition 24 : It is $\lim\limits_{\rho\to 0} F(\rho,\vartheta) = +\infty$ if

$\forall\, \varepsilon\; \exists\, \delta(\varepsilon) : \forall\, \vartheta, 0 < \rho < \delta \Rightarrow F(\rho,\vartheta) > \varepsilon$;

Definition 25 : It is $\lim\limits_{\rho\to 0} F(\rho,\vartheta) = -\infty$ if

$\forall\, \varepsilon\; \exists\, \delta(\varepsilon) : \forall\, \vartheta, 0 < \rho < \delta \Rightarrow F(\rho,\vartheta) < \varepsilon$.

Uniform convergence is expressed by: $\forall\, \varepsilon > 0\; \exists\, \delta(\varepsilon) : \forall\, \vartheta$, which expresses precisely the independence from ϑ in the choice of $\delta(\varepsilon)$.

So if $\lim\limits_{\rho\to 0} F(\rho,\vartheta) = l$ (or $+\infty$ or $-\infty$) uniformly with respect to ϑ, we have that:

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y) = \lim_{\rho\to 0} F(\rho,\vartheta).$$

Example 21 : Let us compute $\lim\limits_{(x,y)\to(0,0)} \dfrac{x^4}{x^2+y^2}$ using polar coordinates.

From $\begin{cases} x = \rho\cos\vartheta \\ y = \rho\sin\vartheta \end{cases}$ we get $\lim\limits_{\rho\to 0} \dfrac{(\rho\cos\vartheta)^4}{(\rho\cos\vartheta)^2 + (\rho\sin\vartheta)^2} = \lim\limits_{\rho\to 0} \rho^2\cos^4\vartheta = 0$.

Let us see whether the convergence to 0 is uniform with respect to ϑ.

From the limit definition, it results that: $\left|\rho^2\cos^4\vartheta\right| < \varepsilon$ in $\mathfrak{J}(0,\ \delta(\varepsilon))$.

But $\left|\rho^2\cos^4\vartheta\right| \le \rho^2$ as $\cos^4\vartheta \le 1$, and so, if $\rho^2 < \varepsilon$, i.e. $-\sqrt{\varepsilon} < \rho < \sqrt{\varepsilon}$, (or $0 < \rho < \sqrt{\varepsilon}$, as ρ is always positive), if $\delta(\varepsilon) = \sqrt{\varepsilon}$ we get:
$0 < \rho < \delta(\varepsilon) = \sqrt{\varepsilon} \Rightarrow \left|\rho^2\cos^4\vartheta\right| < \varepsilon$, and so $\lim\limits_{\rho\to 0} F(\rho,\vartheta) = 0$ and the convergence of the limit is uniform. Then $\lim\limits_{(x,y)\to(0,0)} \dfrac{x^4}{x^2+y^2} = 0$.

Example 22 : Let us compute $\lim\limits_{(x,y)\to(0,0)} \dfrac{x^2\,y}{x^4+y^2}$ using polar coordinates.

Using the lines $y = m\,x$ we get: $\lim\limits_{(x,y)\to(0,0)} \dfrac{x^2\,y}{x^4+y^2} = \lim\limits_{x\to 0} \dfrac{m\,x^3}{x^4+m^2\,x^2} = \lim\limits_{x\to 0} \dfrac{m\,x}{x^2+m^2} = 0$,
but we know that this is not enough to guarantee the existence of the limit.

From $\begin{cases} x = \rho\cos\vartheta \\ y = \rho\sin\vartheta \end{cases}$ we get:

$\lim\limits_{\rho\to 0} \dfrac{\rho^2\cos^2\vartheta\cdot\rho\sin\vartheta}{\rho^4\cos^4\vartheta + \rho^2\sin^2\vartheta} = \lim\limits_{\rho\to 0} \rho\cdot\dfrac{\cos^2\vartheta\cdot\sin\vartheta}{\rho^2\cos^4\vartheta + \sin^2\vartheta} = 0$, as the first factor ρ approaches to 0 while the second factor approaches to $\dfrac{\cos^2\vartheta}{\sin\vartheta}$.

So $\lim\limits_{\rho\to 0} F(\rho,\vartheta) = 0$, but it is not true that $\lim\limits_{(x,y)\to(0,0)} \dfrac{x^2\,y}{x^4+y^2} = 0$, because the convergence is not uniform. In fact, from the limit definition it must result that:
$\left|\rho\cdot\dfrac{\cos^2\vartheta\cdot\sin\vartheta}{\rho^2\cos^4\vartheta + \sin^2\vartheta} - 0\right| = \rho\cdot\dfrac{\cos^2\vartheta\cdot|\sin\vartheta|}{\rho^2\cos^4\vartheta + \sin^2\vartheta} < \varepsilon$. The fact that the convergence is not uniform can be explained noticing that the quantity $\dfrac{\cos^2\vartheta\cdot|\sin\vartheta|}{\rho^2\cos^4\vartheta + \sin^2\vartheta}$ can take arbitrarily large values, for values of ϑ near to 0 or to π, when $\rho \to 0$; so we cannot find a greater term for $\dfrac{\cos^2\vartheta\cdot|\sin\vartheta|}{\rho^2\cos^4\vartheta + \sin^2\vartheta}$ being indipendent from ϑ.

So $\lim\limits_{(x,y)\to(0,0)} \dfrac{x^2\,y}{x^4+y^2}$ does not exist.

CONTINUOUS FUNCTIONS

Similarly to $f:\mathbb{R}\to\mathbb{R}$, also for functions $f:\mathbb{R}^n\to\mathbb{R}$ we have the following:

Definition 26 : Given $\mathbb{X}_0 \in \mathbb{R}^n$ an accumulation point belonging to the domain of $f(\mathbb{X})$, the function $f(\mathbb{X})$ is continuous at $\mathbb{X}_0$ if $\lim\limits_{\mathbb{X}\to\mathbb{X}_0} f(\mathbb{X}) = f(\mathbb{X}_0)$.

Example 23 : Let us verify that $\lim\limits_{(x,y)\to(0,0)} \dfrac{x^3}{x^2+y^2} = 0$. In fact, since:

$\left|\dfrac{x^3}{x^2+y^2}\right| = |x| \cdot \left|\dfrac{x^2}{x^2+y^2}\right| \le |x|$, if $|x| < \delta = \varepsilon$, we get $\left|\dfrac{x^3}{x^2+y^2} - 0\right| < \varepsilon$, and then the limit is 0. We can then define the function $f(x,y) = \begin{cases} \dfrac{x^3}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$.

Since $\lim\limits_{(x,y)\to(0,0)} f(x,y) = f(0,0) = 0$, the function $f(x,y)$ is continuous at $(0,0)$.

For continuous functions $f : \mathbb{R}^n \to \mathbb{R}$ theorems similar to those established for continuous functions of only one variable apply, namely:

- Adding, subtracting and multiplying continuous functions we obtain continuous functions;
- The reciprocal and the quotient of continuous functions (with no infinitesimal denominator) are continuous functions;
- Composing continuous functions we obtain continuous functions.

And also:

Theorem 6 : (Weierstrass) If $f : \mathbb{R}^n \to \mathbb{R}$ is continuous in a compact set, then it admits absolute maximum and minimum values.

Important Note: The terminology used in Italy is different from that used in the Anglo-Saxon literature. In Italy derivable function means a function that has the derivative at a point; differentiable function means a function that can be linearly approximated. In Anglo-Saxon literature the term differentiable function is used in both cases. For real functions of a real variable being derivable implies being differentiable and vice versa, but this does not apply to real functions of a vector variable.

PARTIAL DERIVATIVES

For a function $f : \mathbb{R} \to \mathbb{R}$ we try to determine the instantaneous rate of change (i.e. the derivative) at a point x_0 giving to the independent variable values on the left and on the right of x_0, since the domain of f is contained in $\mathbb{R}$, which is a one dimension space. The derivative, denoted by $f'(x_0)$, is defined as the limit:

$\lim\limits_{x\to x_0} \dfrac{f(x)-f(x_0)}{x-x_0} = \lim\limits_{h\to 0} \dfrac{f(x_0+h)-f(x_0)}{h} = f'(x_0)$, provided that this limit exists and is finite, and it gives us the slope of the tangent at the point $(x_0, f(x_0))$ to the graph of the function.

For functions $f : \mathbb{R}^n \to \mathbb{R}$, being the domain a subset of an n-dimensional space, taken $\mathbb{X}_0$, we must choose one of the infinite directions (straight lines) passing through the point $\mathbb{X}_0$ to determine the instantaneous rate of change of the function f at $\mathbb{X}_0$ relative to the chosen direction.

If point $\mathbb{X}_0$ is interior to D_f, it is possible to develope this analysis in any direction, while, if point $\mathbb{X}_0$ is on the boundary of D_f, this is possible only in some directions.

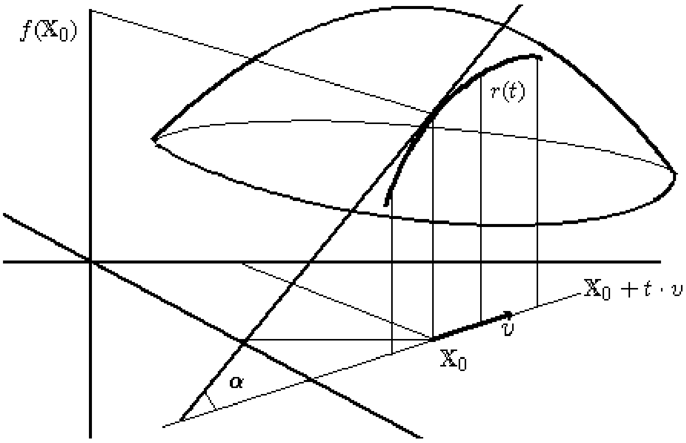
The above leads us to the definition of directional derivative.

If $\mathbb{X}_0$ is an interior point of D_f, we have the following:

Definition 27 : Choosen a unit vector $v \in \mathbb{R}^n$ ($\|v\| = 1$), the directional derivative of f in the direction of v at the point $\mathbb{X}_0$ is the limit $\lim\limits_{t\to 0} \dfrac{f(\mathbb{X}_0 + t\,v) - f(\mathbb{X}_0)}{t} = \mathcal{D}_v f(\mathbb{X}_0)$, provided that this limit exists and is finite. We note that in this definition $t \in \mathbb{R}$.

Since the graph of f is a (hyper)surface, $\mathbb{X}_0 + t\,v$ being a segment that, varying t, starting from $\mathbb{X}_0$ leads in the direction of v, the projection through f of this segment generates a curve $r(t)$, lying on the (hyper)surface rapresenting the graph; to this curve we can draw the tangent line at $(\mathbb{X}_0, f(\mathbb{X}_0))$, which will form an angle α with the line passing through $\mathbb{X}_0$ in the direction of v. As a result, is $\mathcal{D}_v f(\mathbb{X}_0) = \text{tg}\,\alpha$, which represents the slope of the tangent line.

If $f : \mathbb{R}^2 \to \mathbb{R}$, $z = f(x, y)$, every unit vector $v \in \mathbb{R}^2$ can be expressed in the form $v = (\cos\alpha, \sin\alpha)$, with $0 \le \alpha < 2\pi$. The directional derivative $\mathcal{D}_v f(\mathbb{X}_0)$ can then be expressed as: $\mathcal{D}_v f(\mathbb{X}_0) = \lim\limits_{t\to 0} \dfrac{f(x_0 + t\cos\alpha, y_0 + t\sin\alpha) - f(x_0, y_0)}{t}$.
The figure shows the example for a function $f : \mathbb{R}^2 \to \mathbb{R}$.



If we compute the directional derivative using an unit vector of the canonical (also called natural or standard) basis $E = \{e_1, e_2, ..., e_n\}$ we obtain the partial derivative with respect to the variable x_i:
Definition 28 : Chosen $e_i \in E$, the partial derivative of a function f in the direction e_i (or respect to the variable x_i) at $\mathbb{X}_0$ is defined as $\lim\limits_{t\to 0} \dfrac{f(\mathbb{X}_0 + t\, e_i) - f(\mathbb{X}_0)}{t} = \dfrac{\partial f(\mathbb{X}_0)}{\partial x_i}$, provided that this limit exists and is finite.
If $\mathbb{X}_0 = (x_1, x_2, ..., x_n)$ this limit can also be written:

$$\lim_{h\to 0} \frac{f(x_1, x_2, ..., x_i + h, ..., x_n) - f(x_1, x_2, ..., x_i, ..., x_n)}{h} = \frac{\partial f(\mathbb{X}_0)}{\partial x_i}.$$

We use also other notations such as: $\dfrac{\partial f(\mathbb{X}_0)}{\partial x_i} = f'_{x_i}(\mathbb{X}_0) = f'_i(\mathbb{X}_0) = \mathcal{D}_{x_i} f(\mathbb{X}_0) = \mathcal{D}_i f(\mathbb{X}_0)$.
To move in a direction parallel to an axis means to increase only one variable and to keep all the other constant.
If we consider a function of two variables $f(x, y)$, we have two possible partial derivatives at point (x_0, y_0):
-the partial derivative with respect to x, defined as :

$$\lim_{h\to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h} = \frac{\partial f(x_0, y_0)}{\partial\, x} = f'_x(x_0, y_0)$$

-the partial derivative with respect to y, defined as :

$$\lim_{h\to 0} \frac{f(x_0, y_0 + h) - f(x_0, y_0)}{h} = \frac{\partial f(x_0, y_0)}{\partial\, y} = f'_y(x_0, y_0)$$

if these limits exist and are finite.
If we consider a function of three variables $f(x, y, z)$ we have three possible partial derivatives, defined as :

$$\lim_{h\to 0} \frac{f(x_0 + h, y_0, z_0) - f(x_0, y_0, z_0)}{h} = \frac{\partial f(x_0, y_0, z_0)}{\partial\, x} = f'_x(x_0, y_0, z_0)$$

$$\lim_{h\to 0} \frac{f(x_0, y_0 + h, z_0) - f(x_0, y_0, z_0)}{h} = \frac{\partial f(x_0, y_0, z_0)}{\partial\, y} = f'_y(x_0, y_0, z_0)$$

$$\lim_{h\to 0}\frac{f(x_0,y_0,z_0+h)-f(x_0,y_0,z_0)}{h}=\frac{\partial f(x_0,y_0,z_0)}{\partial\, z}=f_z'(x_0,y_0,z_0)$$
if these limits exist and are finite.
For a function $f:\mathbb{R}^n\to\mathbb{R}$ at $\mathbb{X}_0$, interior point of the domain, we can have then, if they exist, exactly n partial derivatives.

If a function is differentiable at $\mathbb{X}_0$ with respect to all its variables, then there exists a vector, called the gradient of the function, denoted by $\nabla f(\mathbb{X}_0)$, whose components are the partial derivatives of the function at the point $\mathbb{X}_0$:
$\nabla f(\mathbb{X}_0)=\left(\frac{\partial\, f(\mathbb{X}_0)}{\partial\, x_1},\frac{\partial\, f(\mathbb{X}_0)}{\partial\, x_2},...,\frac{\partial\, f(\mathbb{X}_0)}{\partial\, x_n}\right)$. The symbol ∇f is read "del f ".

As the partial derivatives are defined by the limit of a difference quotient in which only one variable increases while all others remain constant, we have an important consequence for the practical calculus of partial derivatives: it is sufficient to apply the usual rules to find the derivative for functions of only one variable, the one with respect to which we derive, treating all other variables as constant.

Example 24 : Given $f:\mathbb{R}^2\to\mathbb{R}$, $f(x,y)=x^y$ we have:
$\frac{\partial f}{\partial x}=y\,x^{y-1}$ (derivative of a power, as x is the variable while y is a constant)
$\frac{\partial f}{\partial y}=x^y\log x$ (derivative of an exponential, as y is the variable while x is a constant).

Example 25 : Given $f:\mathbb{R}^2\to\mathbb{R}$, $f(x,y)=x\,\text{arctg}\,\frac{x}{y}$ we have:
$$\frac{\partial f}{\partial x}=1\cdot\text{arctg}\,\frac{x}{y}+x\cdot\frac{1}{1+\left(\frac{x}{y}\right)^2}\cdot\frac{1}{y}\ ;$$
$$\frac{\partial f}{\partial y}=x\cdot\frac{1}{1+\left(\frac{x}{y}\right)^2}\cdot x\cdot\left(-\frac{1}{y^2}\right)\ .$$

Example 26 : Given $f:\mathbb{R}^2\to\mathbb{R}$, $f(x,y)=y^{x-y}$ we have:
$\frac{\partial f}{\partial x}=y^{x-y}\cdot\log y$ (derivative of an exponential, as x is the variable while y is a constant)
$\frac{\partial f}{\partial y}=\mathcal{D}_y\,e^{(x-y)\log y}=y^{x-y}\left(-1\cdot\log y+\frac{x-y}{y}\right)$ (derivative of a $f(y)^{g(y)}$).

Example 27 : Given $f:\mathbb{R}^3\to\mathbb{R}$, $f(x,y,z)=z-\sin\left(\frac{x-z}{x^2}\right)$ we have:
$\frac{\partial f}{\partial x}=0-\cos\left(\frac{x-z}{x^2}\right)\cdot\frac{x^2-2x\,(x-z)}{x^4}$;
$\frac{\partial f}{\partial y}=0$ as the given function is constant with respect to y ;
$\frac{\partial f}{\partial z}=1-\cos\left(\frac{x-z}{x^2}\right)\cdot\frac{1}{x^2}\cdot(-1)$.

Example 28 : Given $f:\mathbb{R}^2\to\mathbb{R}$, $f(x,y)=e^{x+y}$, we calculate the directional derivative at point (x,y) in the direction of the vector $w=(1,1)$.

As $\|w\| = \sqrt{2}$, we have $v = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ and so:

$$\mathcal{D}_v f(x,y) = \lim_{t\to 0}\frac{f((x,y)+t\,v)-f(x,y)}{t} = \lim_{t\to 0}\frac{f\left(x+\frac{1}{\sqrt{2}}t, y+\frac{1}{\sqrt{2}}t\right)-f(x,y)}{t} =$$

$$= \lim_{t\to 0}\frac{e^{x+\frac{1}{\sqrt{2}}t+y+\frac{1}{\sqrt{2}}t}-e^{x+y}}{t} = \lim_{t\to 0}\frac{e^{x+y}\left(e^{\sqrt{2}\,t}-1\right)}{t} =$$

$$= e^{x+y}\cdot\sqrt{2}\cdot\lim_{t\to 0}\frac{e^{\sqrt{2}\,t}-1}{\sqrt{2}\,t} = e^{x+y}\cdot\sqrt{2}\cdot\lim_{w\to 0}\frac{e^w-1}{w} = \sqrt{2}\,e^{x+y}\,.$$

DIFFERENTIABILITY AND CONTINUITY

For functions of only one variable we know that continuity is a necessary condition for differentiability, and therefore also that differentiability is a sufficient condition for continuity. This does not apply, however, to functions $\mathbb{R}^n \to \mathbb{R}$.

Example 29 : Given the function $f(x,y) = \begin{cases} 1: & x\,y = 0 \\ 0: & x\,y \neq 0 \end{cases}$, we compute its partial derivatives at $(0,0)$. We get:

$$\frac{\partial f(0,0)}{\partial x} = \lim_{h\to 0}\frac{f(0+h,0)-f(0,0)}{h} = \lim_{h\to 0}\frac{1-1}{h} = 0\,,$$

$$\frac{\partial f(0,0)}{\partial y} = \lim_{h\to 0}\frac{f(0,0+h)-f(0,0)}{h} = \lim_{h\to 0}\frac{1-1}{h} = 0\,.$$

So the function has partial derivatives at $(0,0)$ although it is clearly discontinuous, as in every neighborhood of $(0,0)$ there are points where $f(x,y)=1$ and points where $f(x,y)=0$.

Example 30 : For $f(x,y) = \begin{cases} \dfrac{x^2\,y}{x^4+y^2} & (x,y)\neq(0,0) \\ 0 & (x,y)=(0,0) \end{cases}$, $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ does non exist, see Example 22, so the function is not continuous at $(0,0)$. Let us check, however, if the function at $(0,0)$ has derivatives in some direction $v = (\cos\alpha, \sin\alpha)$.

We must then calculate:

$$\lim_{t\to 0}\frac{f((0,0)+t\,v)-f(0,0)}{t} = \lim_{t\to 0}\frac{f(t\cos\alpha, t\sin\alpha)-0}{t} =$$

$$= \lim_{t\to 0}\frac{1}{t}\cdot\frac{t^2\cos^2\alpha\cdot t\sin\alpha}{t^4\cos^4\alpha+t^2\sin^2\alpha} = \lim_{t\to 0}\frac{t^3\cos^2\alpha\cdot\sin\alpha}{t^3\left(t^2\cos^4\alpha+\sin^2\alpha\right)} = \frac{\cos^2\alpha\cdot\sin\alpha}{\sin^2\alpha} = \frac{\cos^2\alpha}{\sin\alpha}\,,$$

provided that $\sin\alpha\neq 0$, i.e. $\alpha\neq 0$ e $\alpha\neq\pi$.

If $\alpha = 0$ or $\alpha=\pi$, the direction is the one of the x axis, so the directional derivative is the partial derivative with respect to x, for which:

$$\lim_{t\to 0}\frac{f(0+t,0)-f(0,0)}{t} = \lim_{t\to 0}\frac{1}{t}\cdot\frac{t^2\cdot 0}{t^4+0} = 0\,.$$

So this function has derivatives in every direction without being continuous at $(0,0)$. It is no longer so necessary to be a continuous function to have derivatives.

DIFFERENTIABLE FUNCTIONS

Let us recall the concept of differentiable functions for $f:\mathbb{R}\to\mathbb{R}$, to extend it to functions $f:\mathbb{R}^n\to\mathbb{R}$.

Given $f:\mathbb{R}\to\mathbb{R}$ and x_0 interior point of D_f, the function $f(x)$ is differentiable at x_0 if there exists a constant $\alpha\in\mathbb{R}$ for which the following relation is valid:

$f(x) = f(x_0) + \alpha\,(x - x_0) + o(x - x_0)\,.$

Some important theorems apply:

- a differentiable function at x_0 is continuous at x_0;
- a differentiable function at x_0 has derivative at x_0 and $\alpha = f'(x_0)\,.$

So we have: $f(x) = f(x_0) + f'(x_0)\,(x - x_0) + o(x - x_0)\,.$

To be differentiable means, therefore, that the function can be linearly approximated by the tangent line to the graph of $f(x)$ in x_0, with an error, $o(x - x_0)$, which is negligible compared with $(x - x_0)$, i.e. such that $\lim\limits_{x \to x_0} \dfrac{o(x - x_0)}{x - x_0} = 0\,.$

For functions $f : \mathbb{R}^n \to \mathbb{R}$ the following applies:

Definition 29 : Given $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathbb{X}_0$ interior point of D_f, $f(\mathbb{X})$ is differentiable at $\mathbb{X}_0$ if there is one constant terms vector $\mathbb{K} \in \mathbb{R}^n$ for which:

$f(\mathbb{X}) = f(\mathbb{X}_0) + \mathbb{K} \cdot (\mathbb{X} - \mathbb{X}_0) + o(\|\mathbb{X} - \mathbb{X}_0\|)\,,$

where $\mathbb{K} \cdot (\mathbb{X} - \mathbb{X}_0)$ is the scalar (or dot) product of two vectors of $\mathbb{R}^n$.

The error that arises from this approximation should be negligible compared to $\|\mathbb{X} - \mathbb{X}_0\|$, the norm (or lenght) of $\mathbb{X} - \mathbb{X}_0\,.$

The definition of differentiable function can also be written so:

$$\lim_{\mathbb{X} \to \mathbb{X}_0} \frac{f(\mathbb{X}) - f(\mathbb{X}_0) - \mathbb{K} \cdot (\mathbb{X} - \mathbb{X}_0)}{\|\mathbb{X} - \mathbb{X}_0\|} = 0\,.$$

We note that $\mathbb{K} \cdot (\mathbb{X} - \mathbb{X}_0)$ is a linear application $\mathbb{R}^n \to \mathbb{R}\,.$

Let us consider the relationship between differentiability and continuity, differentiability and the existence of partial derivatives, differentiability and the existence of directional derivatives. The following applies:

Theorem 7 : If f is differentiable at $\mathbb{X}_0$ then f is continuous at $\mathbb{X}_0$.

Proof : Since by hypothesis f is differentiable at $\mathbb{X}_0$, then:

$f(\mathbb{X}) - f(\mathbb{X}_0) = \mathbb{K} \cdot (\mathbb{X} - \mathbb{X}_0) + o(\|\mathbb{X} - \mathbb{X}_0\|)\,.$

We need to show that $\lim\limits_{\mathbb{X} \to \mathbb{X}_0} f(\mathbb{X}) = f(\mathbb{X}_0)\,$, or that $\lim\limits_{\mathbb{X} \to \mathbb{X}_0} f(\mathbb{X}) - f(\mathbb{X}_0) = 0\,.$ But

$\lim\limits_{\mathbb{X} \to \mathbb{X}_0} f(\mathbb{X}) - f(\mathbb{X}_0) = \lim\limits_{\mathbb{X} \to \mathbb{X}_0} (\mathbb{K} \cdot (\mathbb{X} - \mathbb{X}_0) + o(\|\mathbb{X} - \mathbb{X}_0\|)) = 0\,$, since:

$\lim\limits_{\mathbb{X} \to \mathbb{X}_0} \mathbb{K} \cdot (\mathbb{X} - \mathbb{X}_0) = 0\,$, as $\mathbb{K}$ is a constant vector while $\mathbb{X} - \mathbb{X}_0 \to \mathbb{O}\,$, while by hypothesis $o(\|\mathbb{X} - \mathbb{X}_0\|) \to 0\,$, and then the theorem is proved.●

Being a continuous function is therefore a necessary condition to be a differentiable function.

Also the following applies:

Theorem 8 : If f is differentiable at $\mathbb{X}_0$, interior point of D_f, then f has all its partial derivatives at $\mathbb{X}_0$, so exists $\nabla f(\mathbb{X}_0)$, and $\mathbb{K} = \nabla f(\mathbb{X}_0)\,.$

Furthermore f has directional derivatives $\forall\ v$ at $\mathbb{X}_0$ and also: $\mathcal{D}_v f(\mathbb{X}_0) = \nabla f(\mathbb{X}_0) \cdot v\,.$

Proof : First of all, we prove that f has all its partial derivatives at $\mathbb{X}_0$.

If $\mathbb{K} = (k_1, k_2, ..., k_n)$: $\dfrac{\partial f(\mathbb{X}_0)}{\partial x_i} = \lim\limits_{t \to 0} \dfrac{f(\mathbb{X}_0 + t\, e_i) - f(\mathbb{X}_0)}{t} = \lim\limits_{t \to 0} \dfrac{\mathbb{K} \cdot t\, e_i + o(\|t\, e_i\|)}{t} =$

$$= \lim_{t \to 0} \frac{t \cdot [(k_1, k_2, ..., k_i, ..., k_n) \cdot (0, 0, ..., 1_i, ..., 0)] + o(|t| \cdot \|e_i\|)}{t} = \lim_{t \to 0} \frac{t\, k_i + o(|t|)}{t} =$$

$$= \lim_{t \to 0} \left(k_i + \frac{o(|t|)}{t} \right) = k_i\,,$$ as $\lim\limits_{t \to 0} \dfrac{o(|t|)}{t} = 0$ by definition.

Being $k_i \in \mathbb{R}\,$, the limit exists finite, so $\dfrac{\partial f(\mathbb{X}_0)}{\partial x_i} = k_i$ and then $\mathbb{K} = \nabla f(\mathbb{X}_0)\,.$

Therefore, differentiability for f at $\mathbb{X}_0$ can be expressed as:
$f(\mathbb{X}) = f(\mathbb{X}_0) + \nabla f(\mathbb{X}_0) \cdot (\mathbb{X} - \mathbb{X}_0) + o(\|\mathbb{X} - \mathbb{X}_0\|)$.
The gradient of a function $f : \mathbb{R}^n \to \mathbb{R}$ at a point $\mathbb{X}_0$ characterizes the best linear approximation for f at $\mathbb{X}_0$.

$y = f(\mathbb{X}_0) + \nabla f(\mathbb{X}_0) \cdot (\mathbb{X} - \mathbb{X}_0)$ gives the equation of the tangent plane (if $f : \mathbb{R}^2 \to \mathbb{R}$) or of the hyperplane (if $n > 2$) tangent to the (hyper)surface at $(\mathbb{X}_0, f(\mathbb{X}_0))$.
To be differentiable means, therefore, to be approximated by tangent plane (or hyperplane) with an error that is negligible compared to $\|\mathbb{X} - \mathbb{X}_0\|$.

Finally let us calculate the derivative of f at $\mathbb{X}_0$ in any direction v. We have:

$$\mathcal{D}_v f(\mathbb{X}_0) = \lim_{t\to 0}\frac{f(\mathbb{X}_0 + t\,v) - f(\mathbb{X}_0)}{t} = \lim_{t\to 0}\frac{\mathbb{K}\cdot(\mathbb{X}_0 + t\,v - \mathbb{X}_0) + o(\|\mathbb{X}_0 + t\,v - \mathbb{X}_0\|)}{t} =$$
$$= \lim_{t\to 0}\frac{t\,\nabla f(\mathbb{X}_0)\cdot v + o(|t|)}{t} = \lim_{t\to 0}\left(\nabla f(\mathbb{X}_0)\cdot v + \frac{o(|t|)}{t}\right) = \nabla f(\mathbb{X}_0)\cdot v = \mathcal{D}_v f(\mathbb{X}_0)\,.$$

Then the function is differentiable in every direction v, and we also see that to calculate directional derivatives we need not to compute the limit set by the definition, but rather just to compute the scalar (or dot) product of the gradient of f at point $\mathbb{X}_0$ and the unit vector v, so it is sufficient to know the n partial derivatives of f at $\mathbb{X}_0$. This of course if the function is differentiable at $\mathbb{X}_0$.

THE MEANING OF THE GRADIENT

From Schwarz's formula we know that: $\mathbb{X}\cdot\mathbb{Y} = \|\mathbb{X}\|\cdot\|\mathbb{Y}\|\cdot\cos\alpha$, where $\alpha : 0 \le \alpha < \pi$ is the angle between the vectors $\mathbb{X}$ and $\mathbb{Y}$. If f is differentiable at $\mathbb{X}_0$, as v is a unit vector we get: $\mathcal{D}_v f(\mathbb{X}_0) = \nabla f(\mathbb{X}_0)\cdot v = \|\nabla f(\mathbb{X}_0)\|\cdot\|v\|\cdot\cos\alpha = \|\nabla f(\mathbb{X}_0)\|\cdot\cos\alpha$.
If $\alpha = 0$, then $\cos\alpha = 1$ and so $\mathcal{D}_v f(\mathbb{X}_0) = \|\nabla f(\mathbb{X}_0)\|$, and this is the maximum possible value $\mathcal{D}_v f(\mathbb{X}_0)$ can take; but $\alpha = 0$ means that $\nabla f(\mathbb{X}_0)$ and v are on the same straight line and are oriented towards the same side; as v is the direction of the derivative, and as $\nabla f(\mathbb{X}_0)$ expresses the same direction, we can deduce that $\nabla f(\mathbb{X}_0)$ expresses the direction of the maximum growth (or maximum change) of f at $\mathbb{X}_0$. The gradient of a real function is a vector that points towards the direction of the greatest rate of increase of the function, and whose norm is the greatest rate of change.
Similarly, if $\alpha = \pi$, $\nabla f(\mathbb{X}_0)$ and v are on the same straight line but are oriented towards opposite directions, and as $\cos\alpha = -1$ it follows that $\mathcal{D}_v f(\mathbb{X}_0) = -\|\nabla f(\mathbb{X}_0)\|$, and this is the minimum value that $\mathcal{D}_v f(\mathbb{X}_0)$ can take.

CONDITIONS FOR DIFFERENTIABILITY

The existence of the gradient at a certain point is a necessary but not sufficient condition to ensure the differentiability of the function at the same point.
Function $f(x,y) = \begin{cases} 1: & x\,y = 0 \\ 0: & x\,y \ne 0 \end{cases}$ (see Example 29) provides us an example of a function that has all its partial derivatives at $(0,0)$, and therefore it has the gradient, but it is not continuous, and so, following Theorem 7, it is not differentiable.

Example 31 : Let us check if $f(x,y) = \sqrt{|x\,y|}$ is differentiable at $(0,0)$. First we calculate $\nabla f(0,0)$:

$$\frac{\partial f(0,0)}{\partial x} = \lim_{h\to 0}\frac{f(0+h,0) - f(0,0)}{h} = \lim_{h\to 0}\frac{\sqrt{|(0+h)\cdot 0|} - 0}{h} = 0\,,$$
$$\frac{\partial f(0,0)}{\partial y} = \lim_{h\to 0}\frac{f(0,0+h) - f(0,0)}{h} = \lim_{h\to 0}\frac{\sqrt{|0\cdot(0+h)|} - 0}{h} = 0\,.$$

So $\nabla f(0,0) = (0,0)$. If the function were differentiable at $(0,0)$ it should be:

$$f(x,y) - f(0,0) = \frac{\partial f(0,0)}{\partial x} \cdot (x-0) + \frac{\partial f(0,0)}{\partial y} \cdot (y-0) + o\left(\sqrt{(x-0)^2 + (y-0)^2}\right)$$

$$f(x,y) - 0 = 0 + 0 + o\left(\sqrt{(x-0)^2 + (y-0)^2}\right)$$

i.e. it should be $\lim\limits_{(x,y)\to(0,0)} \dfrac{\sqrt{|x\,y|}}{\sqrt{x^2+y^2}} = 0$; using polar coordinates we have:

$\lim\limits_{\rho\to 0} \dfrac{\sqrt{|\rho^2 \cos\vartheta \cdot \sin\vartheta|}}{\sqrt{\rho^2}} = \lim\limits_{\rho\to 0} \sqrt{|\cos\vartheta \cdot \sin\vartheta|} = \sqrt{|\cos\vartheta \cdot \sin\vartheta|}$, whose value is 0 only when $\vartheta = k\,\dfrac{\pi}{2}$. Even if $\nabla f(0,0)$ exists, the function is therefore not differentiable at $(0,0)$.

However, if the partial derivatives are continuous functions at $\mathbb{X}_0$, this condition is sufficient to ensure differentiability at $\mathbb{X}_0$. In fact, the following applies:

Theorem 9 (Total differential) : If the function has all its partial derivatives at $\mathbb{X}_0$, that is if $\nabla f(\mathbb{X}_0)$ exists, and if partial derivatives are continuous functions at $\mathbb{X}_0$, then f is differentiable at $\mathbb{X}_0$.

We do not give the proof of this theorem.

We observe, however, that to have continuous partial derivatives at $\mathbb{X}_0$ is only sufficient but not necessary condition to be a differentiable function at $\mathbb{X}_0$.

Example 32 : For $f(x,y) = e^{x-y}$ we compute $\mathcal{D}_v f(0,0)$ when $v = \left(\dfrac{1}{\sqrt{2}}, -\dfrac{1}{\sqrt{2}}\right)$.

The function is the result of the composition of an exponential with the polynomial $x-y$, therefore it is continuous and differentiable $\forall\,(x,y) \in \mathbb{R}^2$.

As $\dfrac{\partial f(x,y)}{\partial x} = e^{x-y}$ and $\dfrac{\partial f(x,y)}{\partial y} = -\,e^{x-y}$, also the two partial derivatives are continuous functions $\forall\,(x,y) \in \mathbb{R}^2$, and then the given function is differentiable $\forall\,(x,y) \in \mathbb{R}^2$.

As $\dfrac{\partial f(0,0)}{\partial x} = 1$ and $\dfrac{\partial f(0,0)}{\partial y} = -\,1$, we get:

$$\mathcal{D}_v f(0,0) = \nabla f(0,0) \cdot \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) = (1,-1) \cdot \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) = \sqrt{2}.$$

PARTIAL AND DIRECTIONAL DERIVATIVES OF HIGHER ORDER

If $f : \mathbb{R}^n \to \mathbb{R}$ has its n partial derivatives, each of these is still a function $\dfrac{\partial f(\mathbb{X})}{\partial x_i} : \mathbb{R}^n \to \mathbb{R}$, which, if differentiable, can be derived with respect to each of its n variables: $\dfrac{\partial}{\partial x_j}\left(\dfrac{\partial f(\mathbb{X})}{\partial x_i}\right)$.

So we get $n^2 = n \cdot n$ second order partial derivatives, which, if differentiable, can be derived with respect to its n variables, giving course to $n^3 = n^2 \cdot n$ third order partial derivatives and so on.

Therefore we can draw the following scheme, valid for $f : \mathbb{R}^2 \to \mathbb{R}$, $(x,y) \to f(x,y)$:

$$f(x,y) \Rightarrow \begin{cases} \dfrac{\partial f(x,y)}{\partial x} \Rightarrow \begin{cases} \dfrac{\partial^2 f(x,y)}{\partial x^2} = f''_{xx} \Rightarrow \begin{cases} \dfrac{\partial^3 f(x,y)}{\partial x\,\partial x\,\partial x} = \dfrac{\partial^3 f(x,y)}{\partial x^3} = f'''_{xxx} \\ \dfrac{\partial^3 f(x,y)}{\partial x\,\partial x\,\partial y} = \dfrac{\partial^3 f(x,y)}{\partial x^2\ \partial y} = f'''_{xxy} \end{cases} \\ \dfrac{\partial^2 f(x,y)}{\partial x\,\partial y} = f''_{xy} \Rightarrow \begin{cases} \dfrac{\partial^3 f(x,y)}{\partial x\,\partial y\,\partial x} = f'''_{xyx} \\ \dfrac{\partial^3 f(x,y)}{\partial x\,\partial y\,\partial y} = \dfrac{\partial^3 f(x,y)}{\partial x\,\partial y^2} = f'''_{xyy} \end{cases} \end{cases} \\ \dfrac{\partial f(x,y)}{\partial y} \Rightarrow \begin{cases} \dfrac{\partial^2 f(x,y)}{\partial y\,\partial x} = f''_{yx} \Rightarrow \begin{cases} \dfrac{\partial^3 f(x,y)}{\partial y\,\partial x\,\partial x} = \dfrac{\partial^3 f(x,y)}{\partial y\,\partial x^2} = f'''_{yxx} \\ \dfrac{\partial^3 f(x,y)}{\partial y\,\partial x\,\partial y} = f'''_{yxy} \end{cases} \\ \dfrac{\partial^2 f(x,y)}{\partial y^2} = f''_{yy} \Rightarrow \begin{cases} \dfrac{\partial^3 f(x,y)}{\partial y\,\partial y\,\partial x} = \dfrac{\partial^3 f(x,y)}{\partial y^2\,\partial x} = f'''_{yyx} \\ \dfrac{\partial^3 f(x,y)}{\partial y\,\partial y\,\partial y} = \dfrac{\partial^3 f(x,y)}{\partial y^3} = f'''_{yyy} \end{cases} \end{cases} \end{cases}$$

Note the different position of the pseudo-exponent 2: in $\dfrac{\partial^2 f(x,y)}{\partial x^2}$ it means to derive 2 times $(\partial^2 f)$ the function f with respect to the variable x both times (∂x^2), thus explaining the different position.

The derivatives made with respect to the same variable, $\dfrac{\partial^2 f(x,y)}{\partial x^2}$ and $\dfrac{\partial^2 f(x,y)}{\partial y^2}$, are called pure derivatives, while $\dfrac{\partial^2 f(x,y)}{\partial x\,\partial y}$ and $\dfrac{\partial^2 f(x,y)}{\partial y\,\partial x}$ are called mixed derivatives.

For $f : \mathbb{R}^n \to \mathbb{R}$, similarly to the partial derivatives of the second and subsequent orders, we can define the directional derivatives of the second and subsequent orders.

We formally define a second order partial derivative as:
$\dfrac{\partial^2 f(\mathbb{X}_0)}{\partial x_i\,\partial x_j} = \lim\limits_{t\to 0} \dfrac{1}{t} \cdot \left(\dfrac{\partial f(\mathbb{X}_0 + t\,e_j)}{\partial x_i} - \dfrac{\partial f(\mathbb{X}_0)}{\partial x_i} \right)$, provided that this limit exists and is finite.
We formally define a second order directional derivative as:
$\mathcal{D}_w(\mathcal{D}_v f(\mathbb{X}_0)) = \lim\limits_{t\to 0} \dfrac{\mathcal{D}_v f(\mathbb{X}_0 + t\,w) - \mathcal{D}_v f(\mathbb{X}_0)}{t} = \mathcal{D}^2_{v\,w} f(\mathbb{X}_0)$, provided that this limit exists and is finite.
In the same way we define partial and directional derivatives of higher orders.
It is not generally true that $\dfrac{\partial^2 f(\mathbb{X}_0)}{\partial x_i\,\partial x_j} = \dfrac{\partial^2 f(\mathbb{X}_0)}{\partial x_j\,\partial x_i}$, nor that $\mathcal{D}^2_{v\,w} f(\mathbb{X}_0) = \mathcal{D}^2_{w\,v} f(\mathbb{X}_0)$.
However, the following applies:
Theorem 10 (Schwarz) : If $f : \mathbb{R}^n \to \mathbb{R}$ has second order mixed derivatives $\dfrac{\partial^2 f(\mathbb{X})}{\partial x_i\,\partial x_j}$ and $\dfrac{\partial^2 f(\mathbb{X})}{\partial x_j\,\partial x_i}$ in a neighborhood of $\mathbb{X}_0$, and if they are continuous functions at $\mathbb{X}_0$, then they are equal: $\dfrac{\partial^2 f(\mathbb{X}_0)}{\partial x_i\,\partial x_j} = \dfrac{\partial^2 f(\mathbb{X}_0)}{\partial x_j\,\partial x_i}$.
We omit the proof of this theorem.

There is also a more general form of this theorem, whose assumptions include existence and continuity of only one of the two second order mixed derivatives, then it proves that the other second order mixed derivative exists, is continuous and is equal to the first.

Example 33 : Let $f(x,y) = x\,e^y - y\,e^x$, then we have:

$$f(x,y) = x\,e^y - y\,e^x \Rightarrow \begin{cases} f'_x = e^y - y\,e^x \Rightarrow \begin{cases} f''_{xx} = -\,y\,e^x \\ f''_{xy} = e^y - e^x \end{cases} \\ f'_y = x\,e^y - e^x \Rightarrow \begin{cases} f''_{yx} = e^y - e^x \\ f''_{yy} = x\,e^y \end{cases} \end{cases} \text{, and so } f''_{xy} = f''_{yx} \ .$$

Schwarz's theorem expresses a sufficient and not necessary condition, however, for the equality of the second order mixed derivatives.

Remark 1 : Schwarz's theorem applies not only to the second order partial derivatives, but also to mixed derivatives of any order, since a derivative of order n is still the second order derivative of a derivative of order $n-2$: $\partial^{\,n} f = \partial^{\,2}\big(\partial^{\,n-2} f\big)$.

Extending the hypotheses of Schwarz's theorem to the continuity of mixed derivatives of the proper order, we can write, for a function of two variables, $\dfrac{\partial^m f(\mathbb{X})}{\partial x_i^p\,\partial x_j^q}$, with $p+q=m$, to denote the mixed partial derivative of order m, obtained by differentiating p times with respect to x_i and q times with respect to x_j, without specifying the order with respect to which we have derived with respect to x_i and x_j, since this is irrelevant using Schwarz's theorem.

Example 34 : Given $f:\mathbb{R}^n \to \mathbb{R}$, if its third order derivatives are continuous, it results:

$$\frac{\partial^3 f(\mathbb{X})}{\partial x_i^2\,\partial x_j} = \frac{\partial^3 f(\mathbb{X})}{\partial x_i\,\partial x_i\,\partial x_j} = \frac{\partial^3 f(\mathbb{X})}{\partial x_i\,\partial x_j\,\partial x_i} = \frac{\partial^3 f(\mathbb{X})}{\partial x_j\,\partial x_i\,\partial x_i} = \frac{\partial^3 f(\mathbb{X})}{\partial x_j\,\partial x_i^2} \, .$$

Remark 2 : Schwarz's theorem applies not only to partial derivatives but also to all directional derivatives at least of the second order. If $\mathcal{D}^2_{v\,w} f(\mathbb{X})$ exists and is continuous at $\mathbb{X}_0$, then $\mathcal{D}^2_{w\,v} f(\mathbb{X}_0)$ exists, is continuous and also $\mathcal{D}^2_{v\,w} f(\mathbb{X}_0) = \mathcal{D}^2_{w\,v} f(\mathbb{X}_0)$.
Similar conclusions are valid for the higher orders directional derivatives.

DIFFERENTIABILITY OF THE SECOND AND HIGHER ORDERS

Let us consider $f:\mathbb{R}^n \to \mathbb{R}$ and suppose that it is a differentiable function $\forall\,\mathbb{X} \in \mathbb{A}$, $\mathbb{A} \subseteq D_f$. There exist at $\mathbb{X}_0 \in \mathbb{A}$ the n partial derivative functions $\dfrac{\partial f(\mathbb{X})}{\partial x_i} : \mathbb{R}^n \to \mathbb{R}$. Let's take the following:

Definition 30 : $f(\mathbb{X})$ is twice differentiable at $\mathbb{X}_0$ if each of its first order derivative functions $\dfrac{\partial f(\mathbb{X})}{\partial x_i}$ is differentiable at $\mathbb{X}_0$.

So to be a twice differentiable function means to be a function having differentiable first order derivatives.

Therefore it is easy to extend this definition to that of k times differentiable functions.

Definition 31 : $f(\mathbb{X})$ is k-times differentiable at $\mathbb{X}_0$ if each of its $k-1$ order derivative functions is differentiable at $\mathbb{X}_0$.

For twice differentiable functions the following theorem applies:

Theorem 11 : If a function is twice differentiable at $\mathbb{X}_0$, interior point of D_f, then all the second order (partial and directional) derivatives exist at $\mathbb{X}_0$. In addition, the mixed (partial and directional) derivatives are equal: $\dfrac{\partial^2 f(\mathbb{X}_0)}{\partial x_i\,\partial x_j} = \dfrac{\partial^2 f(\mathbb{X}_0)}{\partial x_j\,\partial x_i}$ and $\mathcal{D}^2_{v\,w} f(\mathbb{X}_0) = \mathcal{D}^2_{w\,v} f(\mathbb{X}_0)$.

The latter theorem provides a further sufficient, and not necessary, condition for the equality of the mixed derivatives in addition to the one extablished by Schwarz's theorem.
Similarly to what we have seen for the differentiability of the first order, it is not necessary but only sufficient that the second order derivatives are continuous to ensure that the function is twice differentiable.

TOTAL DIFFERENTIALS OF THE FIRST AND HIGHER ORDERS

For f differentiable at $\mathbb{X}_0$ we can write:
$f(\mathbb{X}) = f(\mathbb{X}_0) + \nabla f(\mathbb{X}_0)\cdot(\mathbb{X}-\mathbb{X}_0) + o(\|\mathbb{X}-\mathbb{X}_0\|)$.
$\nabla f(\mathbb{X}_0)\cdot(\mathbb{X}-\mathbb{X}_0)$ is called the total differential of the first order, and we also write $\mathrm{d}f(\mathbb{X}_0) = \nabla f(\mathbb{X}_0)\cdot(\mathbb{X}-\mathbb{X}_0)$.
If $\nabla f(\mathbb{X}_0) = \left(\dfrac{\partial f(\mathbb{X}_0)}{\partial x_1}, \dfrac{\partial f(\mathbb{X}_0)}{\partial x_2}, ..., \dfrac{\partial f(\mathbb{X}_0)}{\partial x_n}\right) = (f_1'(\mathbb{X}_0), f_2'(\mathbb{X}_0), \dots, f_n'(\mathbb{X}_0))$ and if $(\mathbb{X}-\mathbb{X}_0) = \left(x_1 - x_1^0, x_2 - x_2^0, ..., x_n - x_n^0\right) = (\mathrm{d}x_1, \mathrm{d}x_2, \dots, \mathrm{d}x_n)$, with these symbols we can also write:

$$\mathrm{d}f(\mathbb{X}_0) = (f_1'(\mathbb{X}_0), f_2'(\mathbb{X}_0), \dots, f_n'(\mathbb{X}_0))\cdot(\mathrm{d}x_1, \mathrm{d}x_2, \dots, \mathrm{d}x_n) = \sum_{i=1}^n f_i'(\mathbb{X}_0)\cdot \mathrm{d}x_i\,.$$

If it does not matter to specify the point $\mathbb{X}_0$, we can also write:

$$\mathrm{d}f = (f_1', f_2', .., f_n')\cdot(\mathrm{d}x_1, \mathrm{d}x_2, \dots, \mathrm{d}x_n) = \sum_{i=1}^n f_i'\mathrm{d}x_i = f_1'\,\mathrm{d}x_1 + f_2'\,\mathrm{d}x_2 + \ \dots\ + f_n'\,\mathrm{d}x_n\,.$$

This expression defines the first order total differential for a function of n variables.
If $f:\mathbb{R}^2\to\mathbb{R}$, we get $\mathrm{d}f = f_1'\,\mathrm{d}x_1 + f_2'\,\mathrm{d}x_2$, if the variables are x_1 and x_2, or $\mathrm{d}f = f_x'\,\mathrm{d}x + f_y'\,\mathrm{d}y$, if the variables are x and y.
For a function of three variables $f(x,y,z)$ we have $\mathrm{d}f = f_x'\,\mathrm{d}x + f_y'\,\mathrm{d}y + f_z'\,\mathrm{d}z$.

Let us now define the second order total differential for $f:\mathbb{R}^2\to\mathbb{R}$. Starting from:
$\mathrm{d}^2 f = \mathrm{d}(\mathrm{d}f) = \dfrac{\partial(\mathrm{d}f)}{\partial x}\,\mathrm{d}x + \dfrac{\partial(\mathrm{d}f)}{\partial y}\,\mathrm{d}y$, as f_x' and f_y' are functions of x and y, while $\mathrm{d}x$ and $\mathrm{d}y$ are constant with respect to x and y, we get:

$$\mathrm{d}^2 f = \frac{\partial\left(f_x'\,\mathrm{d}x + f_y'\,\mathrm{d}y\right)}{\partial\,x}\,\mathrm{d}x + \frac{\partial\left(f_x'\,\mathrm{d}x + f_y'\,\mathrm{d}y\right)}{\partial\,y}\,\mathrm{d}y =$$
$$= \left(f_{xx}''\,\mathrm{d}x + f_{yx}''\,\mathrm{d}y\right)\mathrm{d}x + \left(f_{xy}''\,\mathrm{d}x + f_{yy}''\,\mathrm{d}y\right)\mathrm{d}y = f_{xx}''\,(\mathrm{d}x)^2 + 2\,f_{xy}''\,\mathrm{d}x\,\mathrm{d}y + f_{yy}''\,(\mathrm{d}y)^2\,,$$

if we suppose that the function is twice differentiable, wherefore $f_{xy}'' = f_{yx}''$.
For $f:\mathbb{R}^3\to\mathbb{R}$ similarly we get:

$$\mathrm{d}^2 f = f_{xx}''\,(\mathrm{d}x)^2 + f_{yy}''\,(\mathrm{d}y)^2 + f_{zz}''\,(\mathrm{d}z)^2 + 2f_{xy}''\,\mathrm{d}x\,\mathrm{d}y + 2f_{xz}''\,\mathrm{d}x\,\mathrm{d}z + 2f_{yz}''\,\mathrm{d}y\,\mathrm{d}z\,.$$

For $f:\mathbb{R}^n\to\mathbb{R}$, $y = f(x_1, x_2, ..., x_n)$, we shall write in a compact form:

$$\mathrm{d}^2 f = \mathrm{d}^2 y = \sum_{i=1}^n\sum_{j=1}^n \frac{\partial^2 f(\mathbb{X})}{\partial x_i\,\partial x_j}\,\mathrm{d}x_i\,\mathrm{d}x_j = \sum_{i,j=1}^n \frac{\partial^2 f(\mathbb{X})}{\partial x_i\,\partial x_j}\,\mathrm{d}x_i\,\mathrm{d}x_j\,.$$

Note the similarities (not the identity) between the second order differential of a function of two variables and the square of a binomial, between the second order differential of a function of three variables and the square of a trinomial, and so on for second order differential of a function of n variables, which is analogous to the square of an n-omial. These similarities are also reflected in the total differentials of third, fourth order an so on, analogous to a third, fourth power of a binomial, a trinomial and so on.

In fact, the third order total differential of $f : \mathbb{R}^2 \to \mathbb{R}$ is:
$\mathrm{d}^3 f = f'''_{xxx}\,(\mathrm{d}x)^3 + 3\,f'''_{xxy}\,(\mathrm{d}x)^2\,\mathrm{d}y + 3\,f'''_{xyy}\,\mathrm{d}x\,(\mathrm{d}y)^2 + f'''_{yyy}\,(\mathrm{d}y)^3$.
The fourth order total differential of $f : \mathbb{R}^2 \to \mathbb{R}$ is:
$\mathrm{d}^4 f = f^{(4)}_{xxxx}\,(\mathrm{d}x)^4 + 4f^{(4)}_{xxxy}\,(\mathrm{d}x)^3\mathrm{d}y + 6f^{(4)}_{xxyy}\,(\mathrm{d}x)^2(\mathrm{d}y)^2 + 4f^{(4)}_{xyyy}\,\mathrm{d}x(\mathrm{d}y)^3 + f^{(4)}_{yyyy}\,(\mathrm{d}y)^4$.

VECTOR-MATRIX FORM OF SECOND ORDER TOTAL DIFFERENTIALS

The second order total differential of a function of any number of variables can be expressed also in matrix-vector form, using the so-called Hessian matrix $\mathbb{H}$.
The Hessian matrix is the matrix formed with the second order partial derivatives, ordered by row with respect to the first derivation variable and by column with respect to the second, and in case of a function of n variables it takes the form: $\mathbb{H} = \left\| \begin{matrix} f''_{11} & f''_{12} & \dots & \dots & f''_{1n} \\ f''_{21} & f''_{22} & \dots & \dots & f''_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ f''_{n1} & f''_{n2} & \dots & \dots & f''_{nn} \end{matrix} \right\|$.

If the hypothesis of Schwarz's theorem are valid, or if the function is twice differentiable, the Hessian matrix is a symmetric one.

If $\mathrm{d}\mathbb{X} = \|\mathrm{d}x_1, \mathrm{d}x_2, \dots, \mathrm{d}x_n\|$, the following equality holds:

$$\mathrm{d}^2 f = \mathrm{d}\mathbb{X} \cdot \mathbb{H} \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}} = \|\mathrm{d}x_1, \mathrm{d}x_2, \dots, \mathrm{d}x_n\| \cdot \left\| \begin{matrix} f''_{11} & f''_{12} & \dots & \dots & f''_{1n} \\ f''_{21} & f''_{22} & \dots & \dots & f''_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ f''_{n1} & f''_{n2} & \dots & \dots & f''_{nn} \end{matrix} \right\| \cdot \left\| \begin{matrix} \mathrm{d}x_1 \\ \mathrm{d}x_2 \\ \dots \\ \dots \\ \mathrm{d}x_n \end{matrix} \right\|.$$

Verifying it only in the simplest case, the one of a function of two variables, $f : \mathbb{R}^2 \to \mathbb{R}$, we have:

$$\mathrm{d}^2 f = \|\mathrm{d}x, \mathrm{d}y\| \cdot \left\| \begin{matrix} f''_{xx} & f''_{xy} \\ f''_{yx} & f''_{yy} \end{matrix} \right\| \cdot \left\| \begin{matrix} \mathrm{d}x \\ \mathrm{d}y \end{matrix} \right\| = \left\| \begin{matrix} f''_{xx}\,\mathrm{d}x + f''_{yx}\,\mathrm{d}y \\ f''_{xy}\,\mathrm{d}x + f''_{yy}\,\mathrm{d}y \end{matrix} \right\| \cdot \left\| \begin{matrix} \mathrm{d}x \\ \mathrm{d}y \end{matrix} \right\| =$$
$$= f''_{xx}\,(\mathrm{d}x)^2 + 2\,f''_{xy}\,\mathrm{d}x\,\mathrm{d}y + f''_{yy}\,(\mathrm{d}y)^2.$$

Higher order than the second differentials cannot be expressed in matrix-vector form.

As far as the practical calculus of a second-order directional derivative is concerned, there is a result similar to that found for the first-order directional derivatives $\mathcal{D}_v f(\mathbb{X}_0) = \nabla f(\mathbb{X}_0) \cdot v$.
If $v = (v_1, v_2, \dots, v_n)$ and $w = (w_1, w_2, \dots, w_n)$ are unit vectors of $\mathbb{R}^n$, the following holds:
Theorem 12 : Let $f(\mathbb{X})$ be twice differentiable at $\mathbb{X}_0$. Then:
$\mathcal{D}^2_{vw} f(\mathbb{X}_0) = v \cdot \mathbb{H}(\mathbb{X}_0) \cdot w^{\mathrm{T}} = w \cdot \mathbb{H}(\mathbb{X}_0) \cdot v^{\mathrm{T}} = \mathcal{D}^2_{wv} f(\mathbb{X}_0)$.

Example 35 : For $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = e^{x-y}$, we compute $\mathcal{D}^2_{vw} f(x, y)$, using two different procedures, leaving general v, w and (x, y).
If $v = (\cos\alpha, \sin\alpha)$ and $w = (\cos\beta, \sin\beta)$, the function $f(x, y) = e^{x-y}$, being the composition of an exponential and a polynomial, which are continuous and derivable functions with continuous derivatives of any order, is twice differentiable throughout $\mathbb{R}^2$.
So $\mathcal{D}^2_{vw} f(x, y) = v \cdot \mathbb{H}(x, y) \cdot w^{\mathrm{T}}$. Then:

$\dfrac{\partial f}{\partial x} = e^{x-y}$; $\dfrac{\partial f}{\partial y} = -e^{x-y}$; $\dfrac{\partial^2 f}{\partial x^2} = e^{x-y}$; $\dfrac{\partial^2 f}{\partial x\,\partial y} = \dfrac{\partial^2 f}{\partial y\,\partial x} = -e^{x-y}$; $\dfrac{\partial^2 f}{\partial y^2} = e^{x-y}$, and so:

$$\mathcal{D}^2_{vw} f(x,y) = \|\cos\alpha\ \sin\alpha\| \cdot \left\| \begin{matrix} e^{x-y} & -e^{x-y} \\ -e^{x-y} & e^{x-y} \end{matrix} \right\| \cdot \left\| \begin{matrix} \cos\beta \\ \sin\beta \end{matrix} \right\| =$$
$$= \|\cos\alpha\ \sin\alpha\| \cdot \left\| \begin{matrix} e^{x-y}\cos\beta - e^{x-y}\sin\beta \\ -e^{x-y}\cos\beta + e^{x-y}\sin\beta \end{matrix} \right\| =$$
$$= e^{x-y}\cos\beta\cdot\cos\alpha - e^{x-y}\sin\beta\cdot\cos\alpha - e^{x-y}\cos\beta\cdot\sin\alpha + e^{x-y}\sin\beta\cdot\sin\alpha =$$
$$= e^{x-y}\cos(\alpha-\beta) - e^{x-y}\sin(\alpha+\beta) = e^{x-y}\left(\cos(\alpha-\beta) - \sin(\alpha+\beta)\right).$$

Alternatively, we can compute the second order directional derivative as the directional derivative of the first order directional derivative, and then we get:

$$\mathcal{D}_v f(x,y) = \nabla f(x,y)\cdot v = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)\cdot v = (e^{x-y},\ -e^{x-y})\cdot(\cos\alpha, \sin\alpha) =$$
$$= e^{x-y}\cos\alpha - e^{x-y}\sin\alpha.$$

We now calculate $\nabla(\mathcal{D}_v f(x,y))$ and then we shall calculate $\nabla(\mathcal{D}_v f(x,y))\cdot w$. We get:

$$\frac{\partial(\mathcal{D}_v f(x,y))}{\partial x} = e^{x-y}\cos\alpha - e^{x-y}\sin\alpha \ \text{ and } \ \frac{\partial(\mathcal{D}_v f(x,y))}{\partial y} = -e^{x-y}\cos\alpha + e^{x-y}\sin\alpha.$$

And so:

$$\mathcal{D}^2_{vw} f(x,y) = \nabla(\mathcal{D}_v f(x,y))\cdot w = \left(\frac{\partial(\mathcal{D}_v f(x,y))}{\partial x}, \frac{\partial(\mathcal{D}_v f(x,y))}{\partial y}\right)\cdot(\cos\beta, \sin\beta) =$$
$$= (e^{x-y}\cos\alpha - e^{x-y}\sin\alpha;\ -e^{x-y}\cos\alpha + e^{x-y}\sin\alpha)\cdot(\cos\beta, \sin\beta) =$$
$$= e^{x-y}\cos\alpha\cdot\cos\beta - e^{x-y}\sin\alpha\cdot\cos\beta - e^{x-y}\cos\alpha\cdot\sin\beta + e^{x-y}\sin\alpha\cdot\sin\beta =$$
$$= e^{x-y}\left(\cos(\alpha-\beta) - \sin(\alpha+\beta)\right),$$

i.e. the same result as with the other process.

TAYLOR AND MACLAURIN POLYNOMIAL

Also for functions $f:\mathbb{R}^n \to \mathbb{R}$ we can provide a better approximation than that obtained with the differentiability formula, creating a polynomial (in n variables) of degree m suitably chosen, for which the followig equality holds:

$f(\mathbb{X}) - \mathbb{P}_m(\mathbb{X}, \mathbb{X}_0) = o(\|\mathbb{X} - \mathbb{X}_0\|^m)$.

For a function $f:\mathbb{R}\to\mathbb{R}$ Taylor's polynomial at x_0 is:

$$\mathbb{P}_m(x, x_0) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \frac{f'''(x_0)}{3!}(x-x_0)^3 + ...$$

There is a unique polynomial of degree m if the function is m times differentiable in a neighborhood of x_0.

For a function $f:\mathbb{R}^n\to\mathbb{R}$ the following applies instead:

Theorem 13 : (Taylor) If the function $f(\mathbb{X})$, $\mathbb{R}^n\to\mathbb{R}$, is differentiable up to order m in a neighborhood of the point $\mathbb{X}_0$, then there exists a unique Taylor's polynomial of degree m such that: $f(\mathbb{X}) - \mathbb{P}_m(\mathbb{X}, \mathbb{X}_0) = o(\|\mathbb{X} - \mathbb{X}_0\|^m)$.

This polynomial has the following expression:

$$\mathbb{P}_m(\mathbb{X}, \mathbb{X}_0) = f(\mathbb{X}_0) + \mathrm{d}f(\mathbb{X}_0) + \frac{\mathrm{d}^2 f(\mathbb{X}_0)}{2!} + \frac{\mathrm{d}^3 f(\mathbb{X}_0)}{3!} + ... + \frac{\mathrm{d}^m f(\mathbb{X}_0)}{m!}.$$

As can be seen, the polynomial is built using total differentials, from the first order up to the order m.

If $\mathbb{X}_0 = \mathbb{O}$, null vector, we call it, instead Taylor's, MacLaurin's polynomial.

The second-degree polynomial can be expressed, as it has been already seen, using a vector-matrix form, and we have the following expression:

$\mathbb{P}_2(\mathbb{X}, \mathbb{X}_0) = f(\mathbb{X}_0) + \nabla f(\mathbb{X}_0)\cdot \mathrm{d}\mathbb{X} + \dfrac{1}{2}\,\mathrm{d}\mathbb{X}\cdot\mathbb{H}(\mathbb{X}_0)\cdot(\mathrm{d}\mathbb{X})^{\mathrm{T}}$; since $\mathrm{d}\mathbb{X} = \mathbb{X} - \mathbb{X}_0$ we get:

$$f(\mathbb{X}) = f(\mathbb{X}_0) + \nabla f(\mathbb{X}_0)\cdot(\mathbb{X}-\mathbb{X}_0) + \frac{1}{2}(\mathbb{X}-\mathbb{X}_0)\cdot\mathbb{H}(\mathbb{X}_0)\cdot(\mathbb{X}-\mathbb{X}_0)^{\mathrm{T}} + o\left(\|\mathbb{X}-\mathbb{X}_0\|^2\right).$$

Example 36 : Given $f(x,y,z) = x^2 \sin(y-z)$, let us determine the expression of second-degree Taylor's polynomial at the point $(1,1,1)$.
First of all, we get $f(\mathbb{X}_0) = f(1,1,1) = 0$. Then:
$f'_x = 2x\sin(y-z) \Rightarrow f'_x(1,1,1) = 0$;
$f'_y = x^2\cos(y-z) \Rightarrow f'_y(1,1,1) = 1$;
$f'_z = -x^2\cos(y-z) \Rightarrow f'_z(1,1,1) = -1$;
$f''_{xx} = 2\sin(y-z) \Rightarrow f''_{xx}(1,1,1) = 0$;
$f''_{yy} = -x^2\sin(y-z) \Rightarrow f''_{yy}(1,1,1) = 0$;
$f''_{zz} = -x^2\sin(y-z) \Rightarrow f''_{zz}(1,1,1) = 0$;
$f''_{xy} = 2x\cos(y-z) \Rightarrow f''_{xy}(1,1,1) = 2$;
$f''_{xz} = -2x\cos(y-z) \Rightarrow f''_{xz}(1,1,1) = -2$;
$f''_{yz} = x^2\sin(y-z) \Rightarrow f''_{yz}(1,1,1) = 0$.
We get then, in analytical form:

$$\begin{aligned}
&\mathbb{P}_2(\mathbb{X},(1,1,1)) = 0 + 0\cdot(x-1) + 1\cdot(y-1) - 1\cdot(z-1) + \\
&+ \frac{1}{2}\left(0\cdot(x-1)^2 + 0\cdot(y-1)^2 + 0\cdot(z-1)^2\right) + \\
&+ \frac{1}{2}\left(2\cdot 2\,(x-1)\,(y-1) + 2\,(-2)\,(x-1)\,(z-1) + 2\cdot 0\cdot(y-1)\,(z-1)\right) = \\
&= z - y + 2x\,y - 2x\,z\,.
\end{aligned}$$

In vector-matrix form, putting $\mathrm{d}x = x-1, \mathrm{d}y = y-1, \mathrm{d}z = z-1$, we get:

$$\mathbb{P}_2(\mathbb{X},(1,1,1)) = 0 + (0,1,-1)\cdot(\mathrm{d}x,\mathrm{d}y,\mathrm{d}z) + \frac{1}{2}\,\|\mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z\| \cdot \begin{Vmatrix} 0 & 2 & -2 \\ 2 & 0 & 0 \\ -2 & 0 & 0 \end{Vmatrix} \cdot \begin{Vmatrix} \mathrm{d}x \\ \mathrm{d}y \\ \mathrm{d}z \end{Vmatrix}.$$

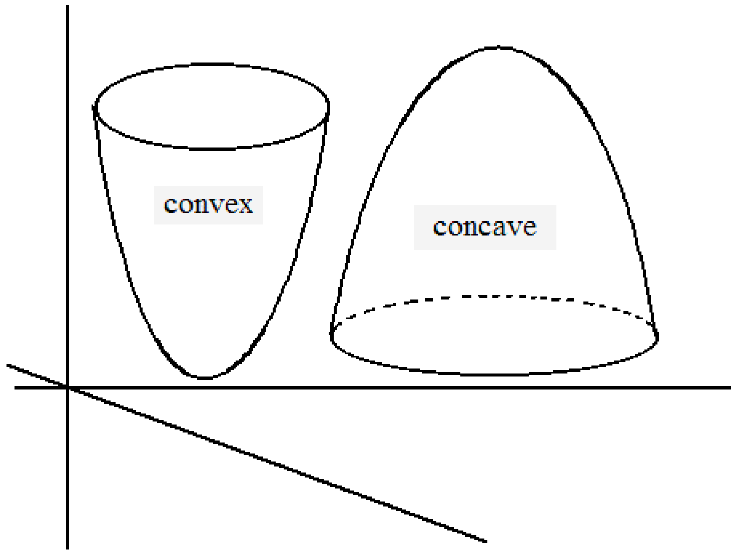
CONVEX AND CONCAVE FUNCTIONS

First of all, it is important to know that for functions $f:\mathbb{R}^n \to \mathbb{R}$ the concept of increasing or decreasing function is not definable. This concept could be recovered if we spoke about increasing or decreasing in a certain direction, i.e. bringing it back to a one-dimensional type analysis, which is not generally useful to draw global conclusions. There are not, therefore, criteria like that of the study of the sign of $f'(x)$, valid for functions $f:\mathbb{R}\to\mathbb{R}$. It is however valid and useful for applications, primarily for the study of maxima and minima, the definition of convex and concave function.
First of all we give the following:
Definition 32 : A set $\mathbb{A} \subseteq \mathbb{R}^n$ is said a convex set if $\forall \mathbb{X}_1, \mathbb{X}_2 \in \mathbb{A}$, the segment that joins them is all contained in $\mathbb{A}$.
This condition is equivalent to:
$\forall\, \mathbb{X}_1, \mathbb{X}_2 \in \mathbb{A} : \forall\, \alpha \in [0,1] \;\Rightarrow\; \alpha\cdot\mathbb{X}_1 + (1-\alpha)\cdot\mathbb{X}_2 \in \mathbb{A}$,
where $\alpha\cdot\mathbb{X}_1 + (1-\alpha)\cdot\mathbb{X}_2,\ \alpha\in\mathbb{R}$ is the line passing through the points $\mathbb{X}_1$ and $\mathbb{X}_2$.
If a set $\mathbb{A} \subseteq \mathbb{R}^n$ is not convex, it is concave.
Definition 33 : For $f:\mathbb{R}^n\to\mathbb{R}$, the epigraph of the function on $\mathbb{A}\subseteq\mathbb{R}^n$, is the set:
$\mathcal{E}(f) = \left\{(\mathbb{X},y)\in\mathbb{R}^{n+1}, \mathbb{X}\in\mathbb{A} :\ y \geq f(\mathbb{X})\right\}$.
The epigraph is therefore the region above the graph, including the graph.
From this it follows the
Definition 34 : Given $f:\mathbb{R}^n\to\mathbb{R}$, and $\mathbb{A}\subseteq\mathbb{R}^n$ a convex set, the function f is said to be convex on $\mathbb{A}$ if its epigraph on $\mathbb{A}$ is a convex set.
So a function is convex if the region above its graph is a convex set.
The definition of convex function is not given in any set, but only in a convex domain.

As for functions $\mathbb{R}\to\mathbb{R}$, unlike sets, a function that is not convex is not called concave, but:

Definition 35 : Given $f:\mathbb{R}^n \to \mathbb{R}$, and $\mathbb{A} \subseteq \mathbb{R}^n$ a convex set, the function f is said to be concave on $\mathbb{A}$ if the function $-f$ is convex on $\mathbb{A}$.

Concave functions are therefore the symmetrical ones, with respect to the (hyper)plane of independent variables, of convex functions and, in a given set, there are functions which are neither concave nor convex, while sets are always concave or convex.
Considering for obvious reasons only the case of functions $f:\mathbb{R}^2 \to \mathbb{R}$, we see the figure of an example of a convex and one of a concave function.



Finally we state two theorems that link the convexity of a function to the differentiability of the first and second order.
Theorem 14 : Let $f:\mathbb{R}^n \to \mathbb{R}$ be differentiable on $\mathbb{A} \subseteq \mathbb{R}^n$ convex set. f is convex on $\mathbb{A}$ if and only if $\forall\, \mathbb{X}, \mathbb{X}_0 \in \mathbb{A} : f(\mathbb{X}) \geq f(\mathbb{X}_0) + \mathrm{d}f(\mathbb{X}_0)$, or, equivalently:
$f(\mathbb{X}) \geq f(\mathbb{X}_0) + \nabla f(\mathbb{X}_0) \cdot (\mathbb{X} - \mathbb{X}_0)$.
This formula allows us to express the convexity of a differentiable function by stating that the graph of the function does not lie below the tangent (hyper)plane at any point $\mathbb{X}_0 \in \mathbb{A}$.
Using Taylor's polynomial we can then prove that is also true that:
Theorem 15 : Given $f:\mathbb{R}^n \to \mathbb{R}$ twice differentiable on $\mathbb{A} \subseteq \mathbb{R}^n$ convex set. Then f is convex on $\mathbb{A}$ if and only if $\mathrm{d}^2 f(\mathbb{X}) \geq 0, \forall\, \mathbb{X} \in \mathbb{A}$.
In fact, being f twice differentiable, we get:
$f(\mathbb{X}) = f(\mathbb{X}_0) + \mathrm{d}f(\mathbb{X}_0) + \dfrac{1}{2}\mathrm{d}^2 f(\mathbb{X}_0) + o\big(\|\mathbb{X} - \mathbb{X}_0\|^2\big)$, and so:
$f(\mathbb{X}) - [f(\mathbb{X}_0) + \mathrm{d}f(\mathbb{X}_0)] = \dfrac{1}{2}\mathrm{d}^2 f(\mathbb{X}_0) + o\big(\|\mathbb{X} - \mathbb{X}_0\|^2\big)$, and if $\mathrm{d}^2 f(\mathbb{X}_0) \geq 0$ we get:
$f(\mathbb{X}) \geq f(\mathbb{X}_0) + \mathrm{d}f(\mathbb{X}_0)$.
Using instead the vector-matrix form, the previous theorem leads to:
$f(\mathbb{X}) - [f(\mathbb{X}_0) + \nabla f(\mathbb{X}_0) \cdot (\mathbb{X} - \mathbb{X}_0)] = \dfrac{1}{2}(\mathbb{X} - \mathbb{X}_0) \cdot \mathbb{H}(\mathbb{X}_0) \cdot (\mathbb{X} - \mathbb{X}_0)^{\mathrm{T}} + o\big(\|\mathbb{X} - \mathbb{X}_0\|^2\big)$
from which we get: $\mathrm{d}^2 f(\mathbb{X}_0) \geq 0 \Leftrightarrow \mathrm{d}\mathbb{X} \cdot \mathbb{H}(\mathbb{X}_0) \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}} \geq 0$.

The term $\mathrm{d}^2 f(\mathbb{X}_0) = \mathrm{d}\mathbb{X} \cdot \mathbb{H}(\mathbb{X}_0) \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}} = (\mathbb{X} - \mathbb{X}_0) \cdot \mathbb{H}(\mathbb{X}_0) \cdot (\mathbb{X} - \mathbb{X}_0)^{\mathrm{T}}$ represents a quadratic form in n variables $\mathrm{d}x_1, \mathrm{d}x_2, \ldots, \mathrm{d}x_n$; if $\mathrm{d}^2 f(\mathbb{X}_0) > 0$ (≥ 0) it is called a positive definite (semi-definite) quadratic form.

The study of convexity and concavity of a function is related to the study of the sign of the quadratic form $\mathrm{d}^2 f(\mathbb{X}_0)$, and this will be discussed below, in the section dealing with the search for the maximum and minimum relative points.

If $f(\mathbb{X}) > f(\mathbb{X}_0) + \mathrm{d}f(\mathbb{X}_0)$ or if $\mathrm{d}^2 f(\mathbb{X}_0) > 0$, we have a strictly convex function.

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

Let us extend the concepts up here worked out to vector valued functions of a vector variable, i.e. to functions $\mathbb{Y} = f(\mathbb{X})$, with $\mathbb{X} = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$, $\mathbb{Y} = (y_1, y_2, ..., y_m) \in \mathbb{R}^m$.
A first example of such a function $f : \mathbb{R}^n \to \mathbb{R}^m$ are linear applications, that is functions that can be expressed as $\mathbb{Y} = f(\mathbb{X}) = \mathbb{A} \cdot \mathbb{X}$, where $\mathbb{A}$ is a $m \cdot n$ matrix whose elements are real or complex numbers.

In case of need, the notation $\mathbb{Y} = f(\mathbb{X}), f : \mathbb{R}^n \to \mathbb{R}^m$ can be written, in an expanded form, as: $(y_1, y_2, ..., y_m) = (f_1(x_1, x_2, ..., x_n), f_2(x_1, x_2, ..., x_n), ..., f_m(x_1, x_2, ..., x_n))$.
Therefore, a function $f : \mathbb{R}^n \to \mathbb{R}^m$ can be seen as an m-dimension vector whose components are functions $f_i : \mathbb{R}^n \to \mathbb{R}$.
It will be sufficient to repeat what we did for functions $f : \mathbb{R} \to \mathbb{R}^m$ adapting the theory to what we saw for functions $f : \mathbb{R}^n \to \mathbb{R}$ to get the extension of main definitions and properties relevant to functions $f : \mathbb{R}^n \to \mathbb{R}^m$.

LIMITS, CONTINUITY, DERIVABILITY, DIFFERENTIABILITY
Definition 36 : The function $f : \mathbb{R}^n \to \mathbb{R}^m, \mathbb{Y} = f(\mathbb{X})$ has limit when $\mathbb{X} \to \mathbb{X}_0$, being $\mathbb{X}_0$ an accumulation point of D_f, if each component $f_i : \mathbb{R}^n \to \mathbb{R}, 1 \le i \le m$ has limit when $\mathbb{X} \to \mathbb{X}_0$.
Definition 37 : Given $\mathbb{X}_0$ an accumulation point belonging to D_f, $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous at $\mathbb{X}_0$ if $\lim\limits_{\mathbb{X} \to \mathbb{X}_0} f(\mathbb{X}) = f(\mathbb{X}_0)$, or $\lim\limits_{\mathbb{X} \to \mathbb{X}_0} f_i(\mathbb{X}) = f_i(\mathbb{X}_0), 1 \le i \le m$, i.e. if each component f_i is continuous at $\mathbb{X}_0$.

As far as the differentiability is concerned, since each component is a function $f_i : \mathbb{R}^n \to \mathbb{R}$, we can define partial and directional derivatives.
Given $\mathbb{X}_0$ interior point of D_f and $v \in \mathbb{R}^n$ unit vector, we have
Definition 38 : The directional derivative of $f : \mathbb{R}^n \to \mathbb{R}^m, \mathbb{Y} = f(\mathbb{X})$ in the direction of v at the point $\mathbb{X}_0$ is the limit $\lim\limits_{t \to 0} \dfrac{f(\mathbb{X}_0 + t\,v) - f(\mathbb{X}_0)}{t} = \mathcal{D}_v f(\mathbb{X}_0)$, provided that this limit exists and is finite.
So each function $f_i : \mathbb{R}^n \to \mathbb{R}, 1 \le i \le m$ must have the directional derivative at $\mathbb{X}_0$ in the direction of v and $\lim\limits_{t \to 0} \dfrac{f_i(\mathbb{X}_0 + t\,v) - f_i(\mathbb{X}_0)}{t} = \mathcal{D}_v f_i(\mathbb{X}_0), 1 \le i \le m$.
Partial derivatives for $f : \mathbb{R}^n \to \mathbb{R}^m$ are partial derivatives of its components $f_i : \mathbb{R}^n \to \mathbb{R}$.
$f : \mathbb{R}^n \to \mathbb{R}^m$ has partial derivative with respect to the variable x_j if each component f_i has partial derivative with respect to x_j.

So we compute $m \cdot n$ partial derivatives, differentiating $f_i,\ 1 \le i \le m$, with respect to the variables $x_j,\ 1 \le j \le n$. All these first order partial derivatives form the so-called Jacobian matrix $m \cdot n$:

$$J_f(\mathbb{X}) = \frac{\partial(f_1, f_2, ..., f_m)}{\partial(x_1, x_2, ..., x_n)} = \left\| \begin{array}{cccc} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & .... & \dfrac{\partial f_1}{\partial x_n} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & .... & \dfrac{\partial f_2}{\partial x_n} \\ .... & .... & .... & .... \\ \dfrac{\partial f_m}{\partial x_1} & \dfrac{\partial f_m}{\partial x_2} & .... & \dfrac{\partial f_m}{\partial x_n} \end{array} \right\|,$$

in which the element (i, j) is the partial derivative of f_i with respect to x_j.
Each row of the Jacobian matrix is a gradient: the i-th row is in fact the gradient $\nabla f_i(\mathbb{X})$.

For a function $f : \mathbb{R}^n \to \mathbb{R}^m, f = (f_1, f_2, ..., f_m)$, we follow the previous definitions, i.e.:
Definition 39 : $f : \mathbb{R}^n \to \mathbb{R}^m, \mathbb{Y} = f(\mathbb{X})$ is differentiable at $\mathbb{X}_0$ if each of its components, $y_i = f_i(\mathbb{X})$, is differentiable at $\mathbb{X}_0$ i.e.:
$f_i(\mathbb{X}) = f_i(\mathbb{X}_0) + \mathbb{K}_i \cdot (\mathbb{X} - \mathbb{X}_0) + o(\|\mathbb{X} - \mathbb{X}_0\|)$, with $\mathbb{K}_i \in \mathbb{R}^n, \forall\, i : 1 \leq i \leq m$.

Equivalently, we may request that there exists a constant terms matrix $\mathbb{M}_{m,n}$, for which:
$f(\mathbb{X}) = f(\mathbb{X}_0) + \mathbb{M} \cdot (\mathbb{X} - \mathbb{X}_0) + o(\|\mathbb{X} - \mathbb{X}_0\|)$ is valid.
Vectors $\mathbb{K}_i$ are the rows of the matrix $\mathbb{M}$.

As a consequence of what we have seen for the functions $f : \mathbb{R}^n \to \mathbb{R}$, if f_i is differentiable at $\mathbb{X}_0$ we have $\mathbb{K}_i = \nabla f_i(\mathbb{X}_0)$, and so the differentiability of $f : \mathbb{R}^n \to \mathbb{R}^m$ can be expressed as: $f_i(\mathbb{X}) = f_i(\mathbb{X}_0) + \nabla f_i(\mathbb{X}_0) \cdot (\mathbb{X} - \mathbb{X}_0) + o(\|\mathbb{X} - \mathbb{X}_0\|), \forall\, i : 1 \leq i \leq m$.
These equalities can be written with a unique formula, as $\mathbb{M} = J_f(\mathbb{X}_0)$:
$f(\mathbb{X}) = f(\mathbb{X}_0) + J_f(\mathbb{X}_0) \cdot (\mathbb{X} - \mathbb{X}_0) + o(\|\mathbb{X} - \mathbb{X}_0\|)$,
where $f(\mathbb{X}), f(\mathbb{X}_0) \in \mathbb{R}^m, o(\|\mathbb{X} - \mathbb{X}_0\|) \in \mathbb{R}^m$ while $(\mathbb{X} - \mathbb{X}_0) \in \mathbb{R}^n$ and $J_f(\mathbb{X}_0)$ is a $(m \cdot n)$ matrix.
In order to compute $\mathcal{D}_v f(\mathbb{X}_0)$, if f is differentiable at $\mathbb{X}_0$, it will be sufficient to compute $J_f(\mathbb{X}_0) \cdot v$, to get the vector $(\mathcal{D}_v f_1(\mathbb{X}_0), \mathcal{D}_v f_2(\mathbb{X}_0), ... , \mathcal{D}_v f_m(\mathbb{X}_0)) \in \mathbb{R}^m$.

The Jacobian matrix expresses and summarizes the concept of derivative in the more general case, i.e. that of $f : \mathbb{R}^n \to \mathbb{R}^m$, including as special cases all types of derivatives encountered so far.
For $f : \mathbb{R} \to \mathbb{R}$, the derivative $f'(x)$ is a $(1 \cdot 1)$ Jacobian matrix, i.e. a real number;
for $f : \mathbb{R}^n \to \mathbb{R}$ we have the gradient, which is a $(1 \cdot n)$ Jacobian matrix, that is a Jacobian having only a row;
for $f : \mathbb{R} \to \mathbb{R}^n$, we have the tangent vector, which is a $(n \cdot 1)$ Jacobian matrix, a Jacobian having only a column.

Example 37 : If $f : \mathbb{R}^n \to \mathbb{R}^m$ is a linear application, $\mathbb{Y} = f(\mathbb{X}) = \mathbb{A} \cdot \mathbb{X}$, it is easy to see that $J_f(\mathbb{X}_0) = \mathbb{A}$, $\forall\, \mathbb{X}_0 \in \mathbb{R}^n$.

Example 38 : If $f : \mathbb{R}^2 \to \mathbb{R}^2$ is a change of coordinates from Cartesian (x, y) to polar ones (ρ, ϑ): $f(\rho, \vartheta) = (x_0 + \rho \cos \vartheta, y_0 + \rho \sin \vartheta)$, we get:

$$J_f(\rho, \vartheta) = \frac{\partial(x, y)}{\partial(\rho, \vartheta)} = \left\| \begin{array}{cc} \dfrac{\partial x}{\partial \rho} & \dfrac{\partial x}{\partial \vartheta} \\ \dfrac{\partial y}{\partial \rho} & \dfrac{\partial y}{\partial \vartheta} \end{array} \right\| = \left\| \begin{array}{cc} \cos \vartheta & -\rho \sin \vartheta \\ \sin \vartheta & \rho \cos \vartheta \end{array} \right\|.$$

DERIVATIVES FOR COMPOSITE FUNCTIONS (CHAIN RULE)

Purpose of this section is to explain the rule of the derivative of a composite function (chain rule) in the more general case: $\mathbb{R}^m \xrightarrow{g} \mathbb{R}^n \xrightarrow{f} \mathbb{R}^p, \mathbb{T} \xrightarrow{g} \mathbb{X} \xrightarrow{f} \mathbb{Y}$, where:
$\mathbb{T} = (t_1, t_2, ..., t_m); \mathbb{X} = (x_1, x_2, ..., x_n); \mathbb{Y} = (y_1, y_2, ..., y_p)$;
$\mathbb{X} = g(\mathbb{T}) = (g_1(\mathbb{T}), g_2(\mathbb{T}), .., g_n(\mathbb{T})) = (g_1(t_1, .., t_m), g_2(t_1, .., t_m), .., g_n(t_1, .., t_m))$;
$\mathbb{Y} = f(\mathbb{X}) = (f_1(\mathbb{X}), f_2(\mathbb{X}),.., f_p(\mathbb{X})) = (f_1(x_1, ..., x_n), f_2(x_1, ..., x_n), .., f_p(x_1, ..., x_n))$
so as to get $\mathbb{Y} = f(\mathbb{X}) = f(g(\mathbb{T}))$.

If $m = n = p = 1$ we get $\mathcal{D}(f(g(x_0))) = f'(g(x_0)) \cdot g'(x_0)$, i.e. the derivative of composite function is given by the product of two numbers: $f'(g(x))$ and $g'(x)$.

If $\mathbb{R} \xrightarrow{g} \mathbb{R} \xrightarrow{f} \mathbb{R}^p, t \xrightarrow{g} x \xrightarrow{f} \mathbb{Y}, \mathbb{Y} = f(g(t))$, see Theorem 5, we get:
$\mathcal{D}(f(g(t_0))) = \dfrac{\mathrm{d}\mathbb{Y}(x(t_0))}{\mathrm{d}t} = \mathbb{Y}'(x(t_0)) \cdot x'(t_0)$,
i.e. the derivative of composite function is given by the product of the tangent vector $\mathbb{Y}'(x(t_0))$ and the scalar $x'(t_0)$.

Let us determine now the rule if $\mathbb{R} \xrightarrow{g} \mathbb{R}^n \xrightarrow{f} \mathbb{R}, t \xrightarrow{g} \mathbb{X} \xrightarrow{f} y$, $n > 1$.
If $y = f(\mathbb{X}) = f(x_1, x_2, ..., x_n) = f(g(t)) = f(x_1(t), x_2(t), ..., x_n(t))$, we have the following
Theorem 16 : If $\mathbb{X}(t) = g(t) = (x_1(t), x_2(t), ..., x_n(t)), \mathbb{R} \to \mathbb{R}^n$ is differentiable at $t = t_0$ and $y = f(\mathbb{X}), \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbb{X}(t_0) = g(t_0)$, then $y = f(g(t))$ is differentiable at $t = t_0$ and: $\mathcal{D}(f(g(t_0))) = \dfrac{\mathrm{d}y(t_0)}{\mathrm{d}t} = \nabla f(\mathbb{X}(t_0)) \cdot \mathbb{X}'(t_0)$,
i.e. the derivative of the composite function is given by the scalar (or dot) product of the gradient vector $\nabla f(\mathbb{X}(t_0))$ and the tangent vector $\mathbb{X}'(t_0) = (x_1'(t_0), x_2'(t_0), ..., x_n'(t_0))$.
Proof : By definition, we have:
$$\frac{\mathrm{d}y(t_0)}{\mathrm{d}t} = \lim_{t \to t_0} \frac{f(g(t)) - f(g(t_0))}{t - t_0} = \lim_{t \to t_0} \frac{f(\mathbb{X}(t)) - f(\mathbb{X}(t_0))}{t - t_0}.$$
As f is differentiable at $\mathbb{X}_0$ we get:
$f(\mathbb{X}(t)) - f(\mathbb{X}(t_0)) = \nabla f(\mathbb{X}(t_0)) \cdot (\mathbb{X}(t) - \mathbb{X}(t_0)) + o(\|\mathbb{X}(t) - \mathbb{X}(t_0)\|)$,
from which, by substitution, we get:
$$\lim_{t \to t_0} \frac{f(\mathbb{X}(t)) - f(\mathbb{X}(t_0))}{t - t_0} = \lim_{t \to t_0} \frac{\nabla f(\mathbb{X}(t_0)) \cdot (\mathbb{X}(t) - \mathbb{X}(t_0)) + o(\|\mathbb{X}(t) - \mathbb{X}(t_0)\|)}{t - t_0} =$$
$$= \lim_{t \to t_0} \nabla f(\mathbb{X}(t_0)) \cdot \frac{\mathbb{X}(t) - \mathbb{X}(t_0)}{t - t_0} + \lim_{t \to t_0} \frac{o(\|\mathbb{X}(t) - \mathbb{X}(t_0)\|)}{\|\mathbb{X}(t) - \mathbb{X}(t_0)\|} \cdot \left(\pm \left\| \frac{\mathbb{X}(t) - \mathbb{X}(t_0)}{t - t_0} \right\| \right).$$
But $\nabla f(\mathbb{X}(t_0))$ is a constant vector, while
$$\lim_{t \to t_0} \frac{\mathbb{X}(t) - \mathbb{X}(t_0)}{t - t_0} = \left(\lim_{t \to t_0} \frac{x_1(t) - x_1(t_0)}{t - t_0}, \dots, \lim_{t \to t_0} \frac{x_n(t) - x_n(t_0)}{t - t_0} \right) =$$
$$= (x_1'(t_0), x_2'(t_0), ..., x_n'(t_0)) = \mathbb{X}'(t_0)$$
as $\mathbb{X}(t)$ is differentiable at t_0. Finally
$\lim\limits_{t \to t_0} \dfrac{o(\|\mathbb{X}(t) - \mathbb{X}(t_0)\|)}{\|\mathbb{X}(t) - \mathbb{X}(t_0)\|} = 0$ by definition and
$\lim\limits_{t \to t_0} \left\| \dfrac{\mathbb{X}(t) - \mathbb{X}(t_0)}{t - t_0} \right\| = \|\mathbb{X}'(t_0)\|$, finite number as $\mathbb{X}(t)$ is differentiable at t_0. So
$\mathcal{D}(f(g(t_0))) = \dfrac{\mathrm{d}y(t_0)}{\mathrm{d}t} = \nabla f(\mathbb{X}(t_0)) \cdot \mathbb{X}'(t_0)$.●

This result can also be expressed in the form:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n}\right) \cdot \left(\frac{\mathrm{d}x_1}{\mathrm{d}t}, \frac{\mathrm{d}x_2}{\mathrm{d}t}, ..., \frac{\mathrm{d}x_n}{\mathrm{d}t}\right) \text{ and so:}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{\partial f}{\partial x_1} \cdot \frac{\mathrm{d}x_1}{\mathrm{d}t} + \frac{\partial f}{\partial x_2} \cdot \frac{\mathrm{d}x_2}{\mathrm{d}t} + ... + \frac{\partial f}{\partial x_n} \cdot \frac{\mathrm{d}x_n}{\mathrm{d}t} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \cdot \frac{\mathrm{d}x_i}{\mathrm{d}t}.$$

Using this result, we treat the case $\mathbb{R} \xrightarrow{g} \mathbb{R}^n \xrightarrow{f} \mathbb{R}^p$, $t \xrightarrow{g} \mathbb{X} \xrightarrow{f} \mathbb{Y}$.
Now $\mathbb{Y} = (y_1, y_2, ..., y_p)$, with $y_i = f_i(x_1, x_2, ..., x_n), 1 \le i \le p$, and $x_j = g_j(t)$. If each $f_i(x_1, x_2, .., x_n) = f_i(\mathbb{X})$ is differentiable at $\mathbb{X}(t_0)$ and if $\mathbb{X}(t) = (x_1(t), x_2(t), ., x_n(t))$ is differentiable at t_0, from the previous theorem we get:

$$\frac{\mathrm{d}y_i(t_0)}{\mathrm{d}t} = \nabla f_i(\mathbb{X}(t_0)) \cdot \mathbb{X}'(t_0) \text{ or:}$$

$$\frac{\mathrm{d}y_i}{\mathrm{d}t} = \frac{\partial f_i}{\partial x_1} \cdot \frac{\mathrm{d}x_1}{\mathrm{d}t} + \frac{\partial f_i}{\partial x_2} \cdot \frac{\mathrm{d}x_2}{\mathrm{d}t} + ... + \frac{\partial f_i}{\partial x_n} \cdot \frac{\mathrm{d}x_n}{\mathrm{d}t},\ 1 \le i \le p.$$

These p equalities can be written in matrix form as:

$$\left\|\begin{matrix} \frac{\mathrm{d}y_1}{\mathrm{d}t} \\ \frac{\mathrm{d}y_2}{\mathrm{d}t} \\ .... \\ \frac{\mathrm{d}y_p}{\mathrm{d}t} \end{matrix}\right\| = \left\|\begin{matrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & .... & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & .... & \frac{\partial f_2}{\partial x_n} \\ .... & .... & .... & .... \\ \frac{\partial f_p}{\partial x_1} & \frac{\partial f_p}{\partial x_2} & .... & \frac{\partial f_p}{\partial x_n} \end{matrix}\right\| \cdot \left\|\begin{matrix} \frac{\mathrm{d}x_1}{\mathrm{d}t} \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} \\ .... \\ \frac{\mathrm{d}x_n}{\mathrm{d}t} \end{matrix}\right\|, \text{ or:}$$

$$\mathcal{D}(f(g(t_0))) = \frac{\mathrm{d}\mathbb{Y}(\mathbb{X}(t_0))}{\mathrm{d}t} = \frac{\partial(f_1, f_2, ..., f_p)(\mathbb{X}(t_0))}{\partial(x_1, x_2, ..., x_n)} \cdot \mathbb{X}'(t_0) \text{ and also:}$$

$\mathbb{Y}'(\mathbb{X}(t_0)) = J_f(\mathbb{X}(t_0)) \cdot \mathbb{X}'(t_0)$.
The tangent vector $\mathbb{Y}'(\mathbb{X}(t_0))$ is then given by the product of the Jacobian matrix $J_f(\mathbb{X}(t_0))$ and the tangent vector $\mathbb{X}'(t_0)$.

Finally we treat the general case $\mathbb{R}^m \xrightarrow{g} \mathbb{R}^n \xrightarrow{f} \mathbb{R}^p$, $\mathbb{T} \xrightarrow{g} \mathbb{X} \xrightarrow{f} \mathbb{Y}$. The following applies:
Theorem 17 : Given $\mathbb{T} = (t_1, t_2, ..., t_m)$, $\mathbb{X} = (x_1, x_2, ..., x_n)$, $\mathbb{Y} = (y_1, y_2, ..., y_p)$ with:
$x_i = g_i(t_1, t_2, ..., t_m)$; $y_j = f_j(x_1, x_2, ..., x_n)$.
If $\mathbb{Y}_0 = f(\mathbb{X}(\mathbb{T}_0)) = f(g(\mathbb{T}_0))$, and $\mathbb{X} = g(\mathbb{T})$ is differentiable at $\mathbb{T}_0$, and $\mathbb{Y} = f(\mathbb{X})$ is differentiable at $\mathbb{X}(\mathbb{T}_0)$, we have the following:

$$\mathcal{D}(f(g(\mathbb{T}_0))) = \frac{\partial(y_1, y_2, ..., y_p)(\mathbb{X}(\mathbb{T}_0))}{\partial(t_1, t_2, ..., t_m)} = \frac{\partial(f_1, f_2, ..., f_p)(\mathbb{X}(t_0))}{\partial(x_1, x_2, ..., x_n)} \cdot \frac{\partial(x_1, x_2, ..., x_n)(\mathbb{T}_0)}{\partial(t_1, t_2, ..., t_m)}.$$

So: $J_{f(g)}(\mathbb{T}_0) = J_f(\mathbb{X}(\mathbb{T}_0)) \cdot J_g(\mathbb{T}_0)$, which in general form can also be expressed as:

$$\frac{\partial(y_1, y_2, ..., y_p)}{\partial(t_1, t_2, ..., t_m)} = \frac{\partial(y_1, y_2, ..., y_p)}{\partial(x_1, x_2, ..., x_n)} \cdot \frac{\partial(x_1, x_2, ..., x_n)}{\partial(t_1, t_2, ..., t_m)}.$$

In matrix form, we have:

$$\left\|\begin{matrix} \frac{\partial y_1}{\partial t_1} & \frac{\partial y_1}{\partial t_2} & .... & \frac{\partial y_1}{\partial t_m} \\ \frac{\partial y_2}{\partial t_1} & \frac{\partial y_2}{\partial t_2} & .... & \frac{\partial y_2}{\partial t_m} \\ .... & .... & .... & .... \\ \frac{\partial y_p}{\partial t_1} & \frac{\partial y_p}{\partial t_2} & .... & \frac{\partial y_p}{\partial t_m} \end{matrix}\right\| = \left\|\begin{matrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & .... & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & .... & \frac{\partial y_2}{\partial x_n} \\ .... & .... & .... & .... \\ \frac{\partial y_p}{\partial x_1} & \frac{\partial y_p}{\partial x_2} & .... & \frac{\partial y_p}{\partial x_n} \end{matrix}\right\| \cdot \left\|\begin{matrix} \frac{\partial x_1}{\partial t_1} & \frac{\partial x_1}{\partial t_2} & .... & \frac{\partial x_1}{\partial t_m} \\ \frac{\partial x_2}{\partial t_1} & \frac{\partial x_2}{\partial t_2} & .... & \frac{\partial x_2}{\partial t_m} \\ .... & .... & .... & .... \\ \frac{\partial x_n}{\partial t_1} & \frac{\partial x_n}{\partial t_2} & .... & \frac{\partial x_n}{\partial t_m} \end{matrix}\right\|$$

It is easy now to see how all cases of derivatives of composite function encountered before are included in $J_{f(g)}(\mathbb{T}_0) = J_f(g(\mathbb{T}_0)) \cdot J_g(\mathbb{T}_0)$, where $J_{f(g)}(\mathbb{T}_0)$ is a $p \cdot m$ matrix, $J_f(g(\mathbb{T}_0))$ is a $p \cdot n$ matrix, and $J_g(\mathbb{T}_0)$ a $n \cdot m$ matrix.

If we want to highlight the expression of a single derivative of $f(g(\mathbb{T}))$ we write:
$\dfrac{\partial y_i}{\partial t_j} = \dfrac{\partial f_i}{\partial x_1} \cdot \dfrac{\partial x_1}{\partial t_j} + \dfrac{\partial f_i}{\partial x_2} \cdot \dfrac{\partial x_2}{\partial t_j} + ... + \dfrac{\partial f_i}{\partial x_n} \cdot \dfrac{\partial x_n}{\partial t_j}, 1 \le i \le p, 1 \le j \le m$, and also:
$\dfrac{\partial y_i}{\partial t_j} = \dfrac{\partial y_i}{\partial x_1} \cdot \dfrac{\partial x_1}{\partial t_j} + \dfrac{\partial y_i}{\partial x_2} \cdot \dfrac{\partial x_2}{\partial t_j} + ... + \dfrac{\partial y_i}{\partial x_n} \cdot \dfrac{\partial x_n}{\partial t_j}, 1 \le i \le p, 1 \le j \le m$.

For a three functions composition: $\mathbb{R}^k \xrightarrow{h} \mathbb{R}^m \xrightarrow{g} \mathbb{R}^n \xrightarrow{f} \mathbb{R}^p, \mathbb{W} \xrightarrow{h} \mathbb{T} \xrightarrow{g} \mathbb{X} \xrightarrow{f} \mathbb{Y}$, to get $\mathbb{Y} = f(g(h(\mathbb{W})))$ derivatives, we must compute:
$J_{f(g(h))}(\mathbb{W}_0) = J_f(\mathbb{X}_0) \cdot J_g(\mathbb{T}_0) \cdot J_h(\mathbb{W}_0) = J_f(g(h(\mathbb{W}_0))) \cdot J_g(h(\mathbb{W}_0)) \cdot J_h(\mathbb{W}_0)$, or:
$\dfrac{\partial(y_1, y_2, ..., y_p)}{\partial(w_1, w_2, ..., w_k)} = \dfrac{\partial(y_1, y_2, ..., y_p)}{\partial(x_1, x_2, ..., x_n)} \cdot \dfrac{\partial(x_1, x_2, ..., x_n)}{\partial(t_1, t_2, ..., t_m)} \cdot \dfrac{\partial(t_1, t_2, ..., t_m)}{\partial(w_1, w_2, ..., w_k)}$.
Since matrices product is not commutative, it is important to stress the proper order of this product.

Example 39 : Given $z = f(x, y)$, let us change from Cartesian coordinates (x, y) to polar coordinates (ρ, ϑ) with $g(\rho, \vartheta) = (x_0 + \rho\cos\vartheta, y_0 + \rho\sin\vartheta)$.
We get a composite function $z = f(g(\rho, \vartheta))$. If f is differentiable we have:
$J_{f(g)}(\rho, \vartheta) = J_f(x(\rho, \vartheta), y(\rho, \vartheta)) \cdot J_g(\rho, \vartheta)$, or: $\dfrac{\partial z}{\partial(\rho, \vartheta)} = \dfrac{\partial z}{\partial(x, y)} \cdot \dfrac{\partial(x, y)}{\partial(\rho, \vartheta)}$ from which:

$$\left\| \begin{matrix} \dfrac{\partial z}{\partial \rho} & \dfrac{\partial z}{\partial \vartheta} \end{matrix} \right\| = \left\| \begin{matrix} \dfrac{\partial z}{\partial x} & \dfrac{\partial z}{\partial y} \end{matrix} \right\| \cdot \left\| \begin{matrix} \dfrac{\partial x}{\partial \rho} & \dfrac{\partial x}{\partial \vartheta} \\ \dfrac{\partial y}{\partial \rho} & \dfrac{\partial y}{\partial \vartheta} \end{matrix} \right\| = \left\| \begin{matrix} z'_x & z'_y \end{matrix} \right\| \cdot \left\| \begin{matrix} \cos\vartheta & -\rho\sin\vartheta \\ \sin\vartheta & \rho\cos\vartheta \end{matrix} \right\| =$$
$$= \left\| \begin{matrix} z'_x\cos\vartheta + z'_y\sin\vartheta & -z'_x\,\rho\sin\vartheta + z'_y\,\rho\cos\vartheta \end{matrix} \right\|.$$

Example 40 : Given $g : \mathbb{R}^3 \to \mathbb{R}, x = g(t_1, t_2, t_3)$ and $f : \mathbb{R} \to \mathbb{R}^2, (y_1, y_2) = (f_1(x), f_2(x))$ differentiable functions. So: $J_{f(g)}(\mathbb{T}) = J_f(g(\mathbb{T})) \cdot J_g(\mathbb{T})$, i.e.:
$\dfrac{\partial(y_1, y_2)}{\partial(t_1, t_2, t_3)} = \dfrac{\partial(y_1, y_2)}{\partial x} \cdot \dfrac{\partial x}{\partial(t_1, t_2, t_3)}$ and so:

$$\left\| \begin{matrix} \dfrac{\partial y_1}{\partial t_1} & \dfrac{\partial y_1}{\partial t_2} & \dfrac{\partial y_1}{\partial t_3} \\ \dfrac{\partial y_2}{\partial t_1} & \dfrac{\partial y_2}{\partial t_2} & \dfrac{\partial y_2}{\partial t_3} \end{matrix} \right\| = \left\| \begin{matrix} \dfrac{\partial y_1}{\partial x} \\ \dfrac{\partial y_2}{\partial x} \end{matrix} \right\| \cdot \left\| \begin{matrix} \dfrac{\partial x}{\partial t_1} & \dfrac{\partial x}{\partial t_2} & \dfrac{\partial x}{\partial t_3} \end{matrix} \right\| \text{ from which:}$$
$$\left\| \begin{matrix} \dfrac{\partial y_1}{\partial t_1} & \dfrac{\partial y_1}{\partial t_2} & \dfrac{\partial y_1}{\partial t_3} \\ \dfrac{\partial y_2}{\partial t_1} & \dfrac{\partial y_2}{\partial t_2} & \dfrac{\partial y_2}{\partial t_3} \end{matrix} \right\| = \left\| \begin{matrix} \dfrac{\partial y_1}{\partial x} \cdot \dfrac{\partial x}{\partial t_1} & \dfrac{\partial y_1}{\partial x} \cdot \dfrac{\partial x}{\partial t_2} & \dfrac{\partial y_1}{\partial x} \cdot \dfrac{\partial x}{\partial t_3} \\ \dfrac{\partial y_2}{\partial x} \cdot \dfrac{\partial x}{\partial t_1} & \dfrac{\partial y_2}{\partial x} \cdot \dfrac{\partial x}{\partial t_2} & \dfrac{\partial y_2}{\partial x} \cdot \dfrac{\partial x}{\partial t_3} \end{matrix} \right\|.$$

Example 41 : Given $g : \mathbb{R}^2 \to \mathbb{R}^3, (t_1, t_2) \to (x_1, x_2, x_3) = \left(2t_1 - t_2^2, t_1 t_2, t_1^2\right)$ and:
$f : \mathbb{R}^3 \to \mathbb{R}^2, (x_1, x_2, x_3) \to (y_1, y_2) = (x_1 + x_2, x_1 x_3)$. We have:
$\dfrac{\partial(y_1, y_2)}{\partial(t_1, t_2)} = \dfrac{\partial(y_1, y_2)}{\partial(x_1, x_2, x_3)} \cdot \dfrac{\partial(x_1, x_2, x_3)}{\partial(t_1, t_2)}$ and so:

$$\left\|\begin{matrix} \dfrac{\partial y_1}{\partial t_1} & \dfrac{\partial y_1}{\partial t_2} \\ \dfrac{\partial y_2}{\partial t_1} & \dfrac{\partial y_2}{\partial t_2} \end{matrix}\right\| = \left\|\begin{matrix} \dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_1}{\partial x_2} & \dfrac{\partial y_1}{\partial x_3} \\ \dfrac{\partial y_2}{\partial x_1} & \dfrac{\partial y_2}{\partial x_2} & \dfrac{\partial y_2}{\partial x_3} \end{matrix}\right\| \cdot \left\|\begin{matrix} \dfrac{\partial x_1}{\partial t_1} & \dfrac{\partial x_1}{\partial t_2} \\ \dfrac{\partial x_2}{\partial t_1} & \dfrac{\partial x_2}{\partial t_2} \\ \dfrac{\partial x_3}{\partial t_1} & \dfrac{\partial x_3}{\partial t_2} \end{matrix}\right\|$$, and substituting:

$$\left\|\begin{matrix} \dfrac{\partial y_1}{\partial t_1} & \dfrac{\partial y_1}{\partial t_2} \\ \dfrac{\partial y_2}{\partial t_1} & \dfrac{\partial y_2}{\partial t_2} \end{matrix}\right\| = \left\|\begin{matrix} 1 & 1 & 0 \\ x_3 & 0 & x_1 \end{matrix}\right\| \cdot \left\|\begin{matrix} 2 & -2t_2 \\ t_2 & t_1 \\ 2t_1 & 0 \end{matrix}\right\| = \left\|\begin{matrix} 2+t_2 & t_1-2t_2 \\ 2x_3+2x_1t_1 & -2x_3t_2 \end{matrix}\right\| =$$

and finally substituting x_1 and x_3 with their expressions:

$$= \left\|\begin{matrix} 2+t_2 & t_1-2t_2 \\ 6t_1^2-2t_1t_2^2 & -2t_1^2t_2 \end{matrix}\right\|.$$

SECOND ORDER DERIVATIVES FOR COMPOSITE FUNCTIONS

We begin treating the following: $g:\mathbb{R}\to\mathbb{R}^2, t\to(x_1,x_2)$ and $f:\mathbb{R}^2\to\mathbb{R},(x_1,x_2)\to y$ or $y=f(x_1,x_2)=f(x_1(t),x_2(t))$.

If g and f are twice differentiable functions, $\mathbb{X}=(x_1,x_2)$ and $\mathbb{X}'=\dfrac{\mathrm{d}\mathbb{X}}{\mathrm{d}t}=(x_1',x_2')$, using the chain rule we get:

$$\frac{\mathrm{d}y}{\mathrm{d}t}=\nabla f(\mathbb{X}(t))\cdot\mathbb{X}'(t)=\frac{\partial f}{\partial x_1}\cdot\frac{\mathrm{d}x_1}{\mathrm{d}t}+\frac{\partial f}{\partial x_2}\cdot\frac{\mathrm{d}x_2}{\mathrm{d}t}.$$

Now we must compute:

$$\frac{\mathrm{d}^2y}{\mathrm{d}t^2}=\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)=\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f}{\partial x_1}\cdot\frac{\mathrm{d}x_1}{\mathrm{d}t}+\frac{\partial f}{\partial x_2}\cdot\frac{\mathrm{d}x_2}{\mathrm{d}t}\right).$$

But $\dfrac{\partial f}{\partial x_1}$ and $\dfrac{\partial f}{\partial x_2}$ are functions of the variables x_1 and x_2 and so, using again the chain rule and the sum and product derivative, we get:

$$\frac{\mathrm{d}^2y}{\mathrm{d}t^2}=\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f}{\partial x_1}\right)\cdot\frac{\mathrm{d}x_1}{\mathrm{d}t}+\frac{\partial f}{\partial x_1}\cdot\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}x_1}{\mathrm{d}t}\right)+\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f}{\partial x_2}\right)\cdot\frac{\mathrm{d}x_2}{\mathrm{d}t}+\frac{\partial f}{\partial x_2}\cdot\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}x_2}{\mathrm{d}t}\right).$$

From composite function derivative it results:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f}{\partial x_1}\right)=\frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_1}\right)\cdot\frac{\mathrm{d}x_1}{\mathrm{d}t}+\frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_1}\right)\cdot\frac{\mathrm{d}x_2}{\mathrm{d}t}=\frac{\partial^2 f}{\partial x_1^2}\cdot\frac{\mathrm{d}x_1}{\mathrm{d}t}+\frac{\partial^2 f}{\partial x_1\partial x_2}\cdot\frac{\mathrm{d}x_2}{\mathrm{d}t}$$, and

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f}{\partial x_2}\right)=\frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_2}\right)\cdot\frac{\mathrm{d}x_1}{\mathrm{d}t}+\frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_2}\right)\cdot\frac{\mathrm{d}x_2}{\mathrm{d}t}=\frac{\partial^2 f}{\partial x_2\partial x_1}\cdot\frac{\mathrm{d}x_1}{\mathrm{d}t}+\frac{\partial^2 f}{\partial x_2^2}\cdot\frac{\mathrm{d}x_2}{\mathrm{d}t},$$

from which we get:

$$\frac{\mathrm{d}^2y}{\mathrm{d}t^2}=\left(\frac{\partial^2 f}{\partial x_1^2}\cdot\frac{\mathrm{d}x_1}{\mathrm{d}t}+\frac{\partial^2 f}{\partial x_1\partial x_2}\cdot\frac{\mathrm{d}x_2}{\mathrm{d}t}\right)\cdot\frac{\mathrm{d}x_1}{\mathrm{d}t}+\frac{\partial f}{\partial x_1}\cdot\frac{\mathrm{d}^2x_1}{\mathrm{d}t^2}+$$

$$+\left(\frac{\partial^2 f}{\partial x_2\partial x_1}\cdot\frac{\mathrm{d}x_1}{\mathrm{d}t}+\frac{\partial^2 f}{\partial x_2^2}\cdot\frac{\mathrm{d}x_2}{\mathrm{d}t}\right)\cdot\frac{\mathrm{d}x_2}{\mathrm{d}t}+\frac{\partial f}{\partial x_2}\cdot\frac{\mathrm{d}^2x_2}{\mathrm{d}t^2}=$$

$$=\frac{\partial^2 f}{\partial x_1^2}\left(\frac{\mathrm{d}x_1}{\mathrm{d}t}\right)^2+2\,\frac{\partial^2 f}{\partial x_1\partial x_2}\,\frac{\mathrm{d}x_1}{\mathrm{d}t}\,\frac{\mathrm{d}x_2}{\mathrm{d}t}+\frac{\partial^2 f}{\partial x_2^2}\left(\frac{\mathrm{d}x_2}{\mathrm{d}t}\right)^2+\frac{\partial f}{\partial x_1}\,\frac{\mathrm{d}^2x_1}{\mathrm{d}t^2}+\frac{\partial f}{\partial x_2}\,\frac{\mathrm{d}^2x_2}{\mathrm{d}t^2},$$

as $\dfrac{\partial^2 f}{\partial x_1\partial x_2}=\dfrac{\partial^2 f}{\partial x_2\partial x_1}$.

In shorthand notation the above equality can be written as:

$y''=f''_{11}\,(x_1')^2+2\,f''_{12}\,x_1'\,x_2'+f''_{22}\,(x_2')^2+f_1'\,x_1''+f_2'\,x_2''$ and also:

$y''=y''_{11}\,(x_1')^2+2\,y''_{12}\,x_1'\,x_2'+y''_{22}\,(x_2')^2+y_1'\,x_1''+y_2'\,x_2''$.

From $\mathbb{X} = (x_1, x_2)$ and $\mathbb{X}' = \dfrac{\mathrm{d}\mathbb{X}}{\mathrm{d}t} = (x_1', x_2')$ we get $\mathbb{X}'' = \dfrac{\mathrm{d}^2\mathbb{X}}{\mathrm{d}t^2} = (x_1'', x_2'')$, and since then $\nabla f(x_1, x_2) = (f_1', f_2') = (y_1', y_2')$ and $\mathbb{H}(f(x_1, x_2)) = \left\|\begin{matrix} y_{11}'' & y_{12}'' \\ y_{12}'' & y_{22}'' \end{matrix}\right\| = \left\|\begin{matrix} f_{11}'' & f_{12}'' \\ f_{12}'' & f_{22}'' \end{matrix}\right\|$, the previous equality can also be expressed as:
$y'' = \mathbb{X}'(t) \cdot \mathbb{H}(f(\mathbb{X}(t))) \cdot (\mathbb{X}'(t))^{\mathrm{T}} + \nabla f(\mathbb{X}(t)) \cdot \mathbb{X}''(t)$,
where $\mathbb{X}'(t), \mathbb{X}''(t), \nabla f(\mathbb{X}(t)) \in \mathbb{R}^2$ and $\mathbb{H}(f(\mathbb{X}(t)))$ is a 2×2 matrix.

This equality can be generalized. Let us see the various cases.

If $\mathbb{X} = (x_1, x_2, ..., x_n)$, let be:
$g : \mathbb{R} \to \mathbb{R}^n, t \to (x_1, x_2, ..., x_n)$ and $f : \mathbb{R}^n \to \mathbb{R}, (x_1, x_2, ..., x_n) \to y$, i.e.:
$\mathbb{R} \to \mathbb{R}^n \to \mathbb{R}, t \to \mathbb{X} \to y, y = f(\mathbb{X}(t))$.
Similarly we obtain:
$\dfrac{\mathrm{d}^2 y}{\mathrm{d}t^2} = \mathbb{X}'(t) \cdot \mathbb{H}(f(\mathbb{X}(t))) \cdot (\mathbb{X}'(t))^{\mathrm{T}} + \nabla f(\mathbb{X}(t)) \cdot \mathbb{X}''(t)$,
where $\mathbb{X}'(t), \mathbb{X}''(t), \nabla f(\mathbb{X}(t)) \in \mathbb{R}^n$ and $\mathbb{H}(f(\mathbb{X}(t)))$ is a $n \times n$ matrix.

If $\mathbb{X} = (x_1, x_2, ..., x_n)$, $\mathbb{Y} = (y_1, y_2, ..., y_p)$ and $y_i = f_i(\mathbb{X})$, let be:
$g : \mathbb{R} \to \mathbb{R}^n, t \to (x_1, x_2, ..., x_n)$ and $f : \mathbb{R}^n \to \mathbb{R}^p, (x_1, x_2, ..., x_n) \to (y_1, y_2, ..., y_p)$, or:
$\mathbb{R} \to \mathbb{R}^n \to \mathbb{R}^p, t \to \mathbb{X} \to \mathbb{Y}, \mathbb{Y} = f(\mathbb{X}(t)) = (f_1(\mathbb{X}(t)), ..., f_p(\mathbb{X}(t)))$.
Similarly we obtain p equalities:
$\dfrac{\mathrm{d}^2 y_i}{\mathrm{d}t^2} = \mathbb{X}'(t) \cdot \mathbb{H}(f_i(\mathbb{X}(t))) \cdot (\mathbb{X}'(t))^{\mathrm{T}} + \nabla f_i(\mathbb{X}(t)) \cdot \mathbb{X}''(t)$, $1 \le i \le p$
where $\mathbb{X}'(t), \mathbb{X}''(t), \nabla f(\mathbb{X}(t)) \in \mathbb{R}^n$ and $\mathbb{H}(f_i(\mathbb{X}(t)))$ is a $n \times n$ matrix.

Finally, if $\mathbb{T} = (t_1, t_2, ..., t_m)$, $\mathbb{X} = (x_1, x_2, ..., x_n)$ and $\mathbb{Y} = (y_1, y_2, ..., y_p)$, the following applies:
Theorem 18 : Given $g : \mathbb{R}^m \to \mathbb{R}^n, (t_1, t_2, ..., t_m) \to (x_1, x_2, ..., x_n)$,
$f : \mathbb{R}^n \to \mathbb{R}^p, (x_1, x_2, ..., x_n) \to (y_1, y_2, ..., y_p)$, or:
$\mathbb{R}^m \to \mathbb{R}^n \to \mathbb{R}^p, \mathbb{T} \to \mathbb{X} \to \mathbb{Y}, \mathbb{Y} = f(\mathbb{X}(\mathbb{T})) = (f_1(\mathbb{X}(\mathbb{T})), ..., f_p(\mathbb{X}(\mathbb{T}))) = f(g(\mathbb{T}))$
both twice differentiable functions. We get the following general formula:
$\dfrac{\partial^2 y_i}{\partial t_j \partial t_k} = \dfrac{\partial \mathbb{X}(\mathbb{T})}{\partial t_j} \cdot \mathbb{H}(f_i(\mathbb{X}(\mathbb{T}))) \cdot \left(\dfrac{\partial \mathbb{X}(\mathbb{T})}{\partial t_k}\right)^{\mathrm{T}} + \nabla f_i(\mathbb{X}(\mathbb{T})) \cdot \dfrac{\partial^2 \mathbb{X}(\mathbb{T})}{\partial t_j \partial t_k}$,
$1 \le i \le p$, $1 \le j, k \le m$, formed by $p \cdot \dfrac{m^2 + m}{2}$ equalities, where:
$\dfrac{\partial \mathbb{X}(\mathbb{T})}{\partial t_j}, \dfrac{\partial^2 \mathbb{X}(\mathbb{T})}{\partial t_j \partial t_k}, \nabla f_i(\mathbb{X}(\mathbb{T})) \in \mathbb{R}^n$ and $\mathbb{H}(f_i(\mathbb{X}(\mathbb{T})))$ is a $n \times n$ matrix.

Example 42 : Let $g : \mathbb{R}^3 \to \mathbb{R}, x = g(t_1, t_2, t_3)$, $f : \mathbb{R} \to \mathbb{R}^2, (y_1, y_2) = (f_1(x), f_2(x))$ be twice differentiable functions.
As: $\dfrac{\partial x(\mathbb{T})}{\partial t_i} = \dfrac{\partial x}{\partial t_i}$, $\mathbb{H}(f_1(\mathbb{X}(\mathbb{T}))) = \dfrac{\mathrm{d}^2 y_1}{\mathrm{d}x^2}$, $\nabla f_1(\mathbb{X}(\mathbb{T})) = \dfrac{\mathrm{d}y_1}{\mathrm{d}x}$, $\dfrac{\partial^2 x(\mathbb{T})}{\partial t_1 \partial t_3} = \dfrac{\partial^2 x}{\partial t_1 \partial t_3}$,
we get:
$\dfrac{\partial^2 y_1}{\partial t_1 \partial t_3} = \dfrac{\partial \mathbb{X}(\mathbb{T})}{\partial t_1} \cdot \mathbb{H}(f_1(\mathbb{X}(\mathbb{T}))) \cdot \left(\dfrac{\partial \mathbb{X}(\mathbb{T})}{\partial t_3}\right)^{\mathrm{T}} + \nabla f_1(\mathbb{X}(\mathbb{T})) \cdot \dfrac{\partial^2 \mathbb{X}(\mathbb{T})}{\partial t_1 \partial t_3}$ and so:
$\dfrac{\partial^2 y_1}{\partial t_1 \partial t_3} = \dfrac{\partial x}{\partial t_1} \cdot \dfrac{\mathrm{d}^2 y_1}{\mathrm{d}x^2} \cdot \dfrac{\partial x}{\partial t_3} + \dfrac{\mathrm{d}y_1}{\mathrm{d}x} \cdot \dfrac{\partial^2 x}{\partial t_1 \partial t_3}$.

Example 43 : Let $g:\mathbb{R}^2\to\mathbb{R}^2, (x_1,x_2)=g(t_1,t_2)$, $f:\mathbb{R}^2\to\mathbb{R}, y=f(x_1,x_2)$ be twice differentiable functions. As:

$$\frac{\partial\mathbb{X}(\mathbb{T})}{\partial t_i}=\left(\frac{\partial x_1}{\partial t_i},\frac{\partial x_2}{\partial t_i}\right),\ \mathbb{H}(f(\mathbb{X}(\mathbb{T})))=\left\|\begin{matrix}\dfrac{\partial^2 y}{\partial x_1^2} & \dfrac{\partial^2 y}{\partial x_1\partial x_2}\\ \dfrac{\partial^2 y}{\partial x_1\partial x_2} & \dfrac{\partial^2 y}{\partial x_2^2}\end{matrix}\right\|=\left\|\begin{matrix}y''_{11} & y''_{12}\\ y''_{12} & y''_{22}\end{matrix}\right\|,$$

$$\nabla f(\mathbb{X}(\mathbb{T}))=\left(\frac{\partial y}{\partial x_1},\frac{\partial y}{\partial x_2}\right)=(y'_1,y'_2),\ \frac{\partial^2\mathbb{X}(\mathbb{T})}{\partial t_1^2}=\left(\frac{\partial^2 x_1}{\partial t_1^2},\frac{\partial^2 x_2}{\partial t_1^2}\right)$$, we get:

$$\frac{\partial^2 y}{\partial t_1^2}=\left\|\frac{\partial x_1}{\partial t_1},\frac{\partial x_2}{\partial t_1}\right\|\cdot\left\|\begin{matrix}y''_{11} & y''_{12}\\ y''_{12} & y''_{22}\end{matrix}\right\|\cdot\left\|\begin{matrix}\dfrac{\partial x_1}{\partial t_1}\\ \dfrac{\partial x_2}{\partial t_1}\end{matrix}\right\|+\left(\frac{\partial y}{\partial x_1},\frac{\partial y}{\partial x_2}\right)\cdot\left(\frac{\partial^2 x_1}{\partial t_1^2},\frac{\partial^2 x_2}{\partial t_1^2}\right)=$$

$$\frac{\partial^2 y}{\partial t_1^2}=\left(\frac{\partial x_1}{\partial t_1}\right)^2 y''_{11}+2\,\frac{\partial x_1}{\partial t_1}\,\frac{\partial x_2}{\partial t_1}\,y''_{12}+\left(\frac{\partial x_2}{\partial t_1}\right)^2 y''_{22}+y'_1\,\frac{\partial^2 x_1}{\partial t_1^2}+y'_2\,\frac{\partial^2 x_2}{\partial t_1^2}\,;$$

$$\frac{\partial^2 y}{\partial t_1\partial t_2}=\left\|\frac{\partial x_1}{\partial t_1},\frac{\partial x_2}{\partial t_1}\right\|\cdot\left\|\begin{matrix}y''_{11} & y''_{12}\\ y''_{12} & y''_{22}\end{matrix}\right\|\cdot\left\|\begin{matrix}\dfrac{\partial x_1}{\partial t_2}\\ \dfrac{\partial x_2}{\partial t_2}\end{matrix}\right\|+\left(\frac{\partial y}{\partial x_1},\frac{\partial y}{\partial x_2}\right)\cdot\left(\frac{\partial^2 x_1}{\partial t_1\partial t_2},\frac{\partial^2 x_2}{\partial t_1\partial t_2}\right)=$$

$$\frac{\partial^2 y}{\partial t_1\partial t_2}=\frac{\partial x_1}{\partial t_1}\,\frac{\partial x_1}{\partial t_2}\,y''_{11}+\left(\frac{\partial x_1}{\partial t_2}\,\frac{\partial x_2}{\partial t_1}+\frac{\partial x_1}{\partial t_1}\,\frac{\partial x_2}{\partial t_2}\right)y''_{12}+\frac{\partial x_2}{\partial t_1}\,\frac{\partial x_2}{\partial t_2}\,y''_{22}+$$
$$+\,y'_1\,\frac{\partial^2 x_1}{\partial t_1\partial t_2}+y'_2\,\frac{\partial^2 x_2}{\partial t_1\partial t_2}\,;$$

$$\frac{\partial^2 y}{\partial t_2^2}=\left\|\frac{\partial x_1}{\partial t_2},\frac{\partial x_2}{\partial t_2}\right\|\cdot\left\|\begin{matrix}y''_{11} & y''_{12}\\ y''_{12} & y''_{22}\end{matrix}\right\|\cdot\left\|\begin{matrix}\dfrac{\partial x_1}{\partial t_2}\\ \dfrac{\partial x_2}{\partial t_2}\end{matrix}\right\|+\left(\frac{\partial y}{\partial x_1},\frac{\partial y}{\partial x_2}\right)\cdot\left(\frac{\partial^2 x_1}{\partial t_2^2},\frac{\partial^2 x_2}{\partial t_2^2}\right)=$$

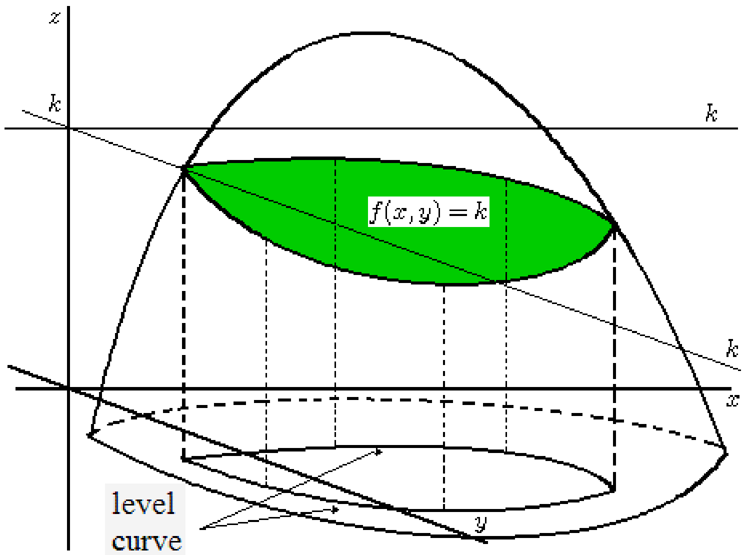
$$\frac{\partial^2 y}{\partial t_2^2}=\left(\frac{\partial x_1}{\partial t_2}\right)^2 y''_{11}+2\,\frac{\partial x_1}{\partial t_2}\,\frac{\partial x_2}{\partial t_2}\,y''_{12}+\left(\frac{\partial x_2}{\partial t_2}\right)^2 y''_{22}+y'_1\,\frac{\partial^2 x_1}{\partial t_2^2}+y'_2\,\frac{\partial^2 x_2}{\partial t_2^2}\,.$$

IMPLICIT FUNCTIONS

Let $f:\mathbb{R}^n\to\mathbb{R}^m$, $n\geq 1, m\geq 1$. The function is called "in explicit form" when we know the law f that allows us to associate to each element of the domain its unique image.
All functions used up to now are functions in explicit form.
Now let us consider the equation $x^2+y^2=1$. Points $(x,y)\in\mathbb{R}^2$ that satisfy this equation form the circumference with center in the origin and radius equal to 1, i.e. the trigonometric circle.
However this circle is not the graph of a function $y=f(x)$, because to each value x correspond two values y (excluding $x=\pm 1$) as well as to each value y correspond two values x (excluding $y=\pm 1$). Solving the equation algebrically we get $y=\pm\sqrt{1-x^2}$, that is, the explicit expression of two possible functions $y=y(x)$, having respectively as graph the upper and the lower semicircle, or we can get $x=\pm\sqrt{1-y^2}$, that is, two possible explicit expressions of functions $x=x(y)$, where the roles of the independent and dependent variables have been exchanged, these latter functions have as graph, respectively, the semicircle on the right and the one on the left.
It is easily seen that $x^2+(y(x))^2=(x(y))^2+y^2=1$.

We could still break the circle into a suitable number of not overlapping traits, to form other functions which, unlike the four previous ones, would not be continuous.
Now let us consider instead the equation $f(x,y) = x\,e^y + x^2\,y^2 = 1$.
The point $P_0 = (1,0)$ satisfies it; this equation is not solvable with respect to y while it can be solved with respect to x, since it is a second degree polynomial $y^2\,x^2 + e^y\,x - 1 = 0$ having as solutions $x = \dfrac{-\,e^y \pm \sqrt{e^{2y} + 4y^2}}{2y^2}$, which are the two possible forms of explicit function $x = x(y)$ obtainable from this equation. We easily verify that $f(x(y), y) = 1$.
Finally, let us consider the equation $f(x,y) = x\,e^y + x^3\,y^2 = 1$. The point $P_0 = (1,0)$ satisfies it, but this equation cannot be solved with respect neither to x nor to y. We wonder wheter by the latter equation it is possible to guarantee the existence of functions $y = y(x)$ or $x = x(y)$, which we, being unable to obtain the explicit expression, will call functions in implicit form.
We will examine below some introductory examples, each time increasing first the number of variables and then the number of the equations, to arrive finally to establish a theorem valid for the general case.
We start with the simplest case, namely that of an equation with two variables.

FUNCTIONS IMPLICITLY DEFINED BY AN EQUATION
I case: Equation $f(x,y) = k$: implicit function $\mathbb{R} \to \mathbb{R}$
Let us suppose we have a general equation with two variables $f(x,y) = k, k \in \mathbb{R}$.
Let us first find a point $\mathrm{P}_0 = (x_0, y_0)$ that satisfies such equation $f(x_0, y_0) = k$.
Once we have found P_0 we want to see if, in a neighborhood of x_0 (or y_0), the set of points (x,y) such that $f(x,y) = k$ is the graph of a function $y = y(x)$ or $x = x(y)$, even if we cannot find its explicit expression, but however we try to verify if it is a continuous and differentiable function.
If so, i.e. if $y_0 = y(x_0)$ and $f(x, y(x)) = k$, $\forall\, x \in \mathfrak{I}(x_0)$ we shall say that the function $y = y(x)$ is implicitly defined by the equation $f(x,y) = k$.



This problem can also be seen geometrically using the so-called level curves. Confining ourselves to the case of $f : \mathbb{R}^2 \to \mathbb{R}$ to graphically represent the situation, if $w = f(x,y)$, the equation $f(x,y) = k$ leads to determine the pairs (x,y) for which $w = k$, and this is similar to cut the surface $w = f(x,y)$ with a plane, parallel to the $(x;y)$ plane, at an height equal to k.

The intersection between the surface and the plane generates a curve, which is the projection by means of f of a curve lying on the $(x;y)$ plane: the level curve.
We can then see this level curve as a curve in the true sense of the term, that is a function $g:\mathbb{R}\to\mathbb{R}^2, t\to(x(t),y(t))$.
We then have the following composite function:
$\mathbb{R}\xrightarrow{g}\mathbb{R}^2\xrightarrow{f}\mathbb{R}, t\xrightarrow{g}(x(t),y(t))\xrightarrow{f}w$, where $w=f(x,y)=k$ is a constant.
We get so, differentiating:
$$\frac{\mathrm{d}w}{\mathrm{d}t}=\nabla f(x,y)\cdot\frac{\mathrm{d}(x,y)}{\mathrm{d}t}=\frac{\partial f}{\partial x}\cdot\frac{\mathrm{d}x}{\mathrm{d}t}+\frac{\partial f}{\partial y}\cdot\frac{\mathrm{d}y}{\mathrm{d}t}=f_x'\cdot x'(t)+f_y'\cdot y'(t),$$
where $(x'(t),y'(t))$ is the tangent vector to the given curve.
But $\frac{\mathrm{d}w}{\mathrm{d}t}=0$, since w is constant, and so $\nabla f(x,y)\cdot(x'(t),y'(t))=0$, i.e. the gradient of f and the tangent vector are perpendicular, at any point of the level curve. We will use this property below.

Let us come back to the problem of the existence of the implicit function $y=y(x)$ defined by the equation $f(x,y)=k$.
The existence and the properties of such an implicit function are established as follows:
Teorema 19 (U. Dini) : Let us suppose that $f:\mathbb{R}^2\to\mathbb{R}$ is a continuous function having continuous derivative f_y' in $\mathbb{A}\subseteq\mathbb{R}^2$; let $f(x_0,y_0)=k$ and $f_y'(x_0,y_0)\neq 0$. Then there exists a neighborhood $\mathfrak{J}(x_0)$ and a single continuous function $y=y(x)$, such that $y_0=y(x_0)$ and $f(x,y(x))=k$, $\forall\, x\in\mathfrak{J}(x_0)$.
Moreover, if also f_x' is continuous in $\mathbb{A}$, then $y(x)$ is differentiable in $\mathfrak{J}(x_0)$ and so:
$$y'(x)=\frac{\mathrm{d}y}{\mathrm{d}x}=-\frac{f_x'}{f_y'},\ \forall\, x\in\mathfrak{J}(x_0).$$
Finally, the function $y'(x)$ is also continuous $\forall\, x\in\mathfrak{J}(x_0)$.

We don't give the proof of this theorem, but we shall verify, using the derivative of composite function, the result found with regard to the derivative $y'(x)$.
Note how the hypotheses of the theorem give a sufficient and not necessary condition for the existence of an implicit function.
In addition, if all assumptions were met, they would imply the differentiability of $f(x,y)$ in $\mathbb{A}$.
With appropriate changes in assumptions we can deduce existence, continuity and differentiability of an implicit function $x=x(y)$ in $\mathfrak{J}(y_0)$.

Example 44 : Let us consider the equation $f(x,y)=x^2-y=0$. At $(x_0,y_0)=(0,0)$ is $f(0,0)=0$; $f(x,y)$ is continuous with continuous derivatives throughout the whole $\mathbb{R}^2$ and moreover $f_y'=-1\neq 0$ while $f_x'=2x$, which vanishes for $x=0$.
The function defined by $f(x,y)=x^2-y=0$ can be made explicit as $y=y(x)=x^2$, is continuous and differentiable throughout the whole $\mathbb{R}$. On the contrary, in a neighborhood of $y=0$ it is not possible to define a function $x=x(y)$, as each $y>0$ has two corresponding $x=\pm\sqrt{y}$. Searching for the inverse function leads in fact to $y=\sqrt{x}$ or $y=-\sqrt{x}$, depending on whether we restrict the domain of $g(x)=x^2$ to $\mathbb{R}_+$ or $\mathbb{R}_-$.

Example 45 : Given the equation $f(x,y)=x^3-y=0$. At $(x_0,y_0)=(0,0)$ is $f(0,0)=0$; moreover $f(x,y)$ is continuous with continuous derivatives throughout $\mathbb{R}^2$ and $f_y'=-1\neq 0$ while $f_x'=3x^2$, which vanishes for $x=0$.

The function defined by $f(x,y) = x^3 - y = 0$ is made explicit as $y = y(x) = x^3$, is continuous and differentiable throughout $\mathbb{R}$, where it is also invertible, with inverse $x = \sqrt[3]{y}$.
Even if $f'_x(0,0) = 0$ so, in a neighborhood of $y = 0$ the function (explicit) exists, confirming the fact that condition $f'_y \neq 0$ (or $f'_x \neq 0$) is only sufficient and not necessary.
Geometrically, the function $y = x^3$ has at $x = 0$ an horizontal tangent line, but this happens at an inflection point, and not at a minimum point, as in the previous example, and this allows the existence of the function $x = x(y)$.

Implicit Function $\mathbb{R} \to \mathbb{R}$: First order derivative
If $f(x,y) = k$ and if we define implicitly y as $y = y(x)$, we get this functions composition:
$\mathbb{R} \to \mathbb{R}^2 \xrightarrow{f} \mathbb{R}, x \to (x, y(x)) \xrightarrow{f} w = f(x,y) = k$. We can then calculate:

$$\frac{\mathrm{d}w}{\mathrm{d}x} = \frac{\partial(f)}{\partial(x,y)} \cdot \frac{\partial(x,y)}{\partial(x)} = \nabla f(x,y) \cdot \left(\frac{\mathrm{d}x}{\mathrm{d}x}, \frac{\mathrm{d}y}{\mathrm{d}x}\right) = f'_x \cdot 1 + f'_y \cdot y'(x) = 0,$$

as f is a constant, and then we get:

$$y'(x) = \frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{f'_x}{f'_y}.$$

If we had implicitly defined $x = x(y)$ we would instead have:
$\mathbb{R} \to \mathbb{R}^2 \xrightarrow{f} \mathbb{R}, y \to (x(y), y) \xrightarrow{f} w = f(x,y) = k$, and from this:

$$\frac{\mathrm{d}w}{\mathrm{d}y} = \frac{\partial(f)}{\partial(x,y)} \cdot \frac{\partial(x,y)}{\partial(y)} = \nabla f(x,y) \cdot \left(\frac{\mathrm{d}x}{\mathrm{d}y}, \frac{\mathrm{d}y}{\mathrm{d}y}\right) = f'_x \cdot x'(y) + f'_y \cdot 1 = 0,$$

and so: $x'(y) = \dfrac{\mathrm{d}x}{\mathrm{d}y} = -\dfrac{f'_y}{f'_x}$.

As $\dfrac{f'_y}{f'_x}$ is the reciprocal of $\dfrac{f'_x}{f'_y}$, we find again the inverse function derivative rule:

$y'(x) = \dfrac{1}{x'(y)}$, if $f'_x \neq 0$ and $f'_y \neq 0$.

Implicit Function $\mathbb{R} \to \mathbb{R}$: Tangent line equation
From $f(x,y) = k$, assuming it is defined $y = y(x)$, we can write the equation of the tangent line to the curve $y = y(x)$ at x_0, which will be: $y - y(x_0) = y'(x_0)(x - x_0)$, and so:

$$y - y_0 = -\frac{f'_x(x_0,y_0)}{f'_y(x_0,y_0)} \cdot (x - x_0),$$

which can also be written as: $f'_x(x_0,y_0)(x - x_0) + f'_y(x_0,y_0)(y - y_0) = 0$ or:
$\nabla f(x_0,y_0) \cdot (x - x_0, y - y_0) = 0$, that is, we find again the orthogonality between the gradient of f and the tangent vector to the level curve.

Example 46 : Let us consider the equation $f(x,y) = x\,e^y + x^2 y^2 = 1$.
It is $f(1,0) = 1$; and also $f'_x = e^y + 2x\,y^2$ while $f'_y = x\,e^y + 2x^2 y$. So, as $f'_x(1,0) = 1$ and $f'_y(1,0) = 1 \neq 0$, it exists, in a neighborhood of the point $x = 1$, the implicit function $y = y(x)$ and the result is: $y'(1) = -\dfrac{f'_x(1,0)}{f'_y(1,0)} = -1$.

We get also, in the same neighborhood of $x = 1$: $y'(x) = -\dfrac{e^y + 2x\,y^2}{x\,e^y + 2x^2 y}$.

For the equation of the tangent line to $y = y(x)$ at $x = 1$, as $y(1) = y_0 = 0$, we get: $y = -1 \cdot (x - 1) = 1 - x$.

Implicit Function $\mathbb{R} \to \mathbb{R}$: Second order derivative; Taylor's polynomial

If $w = f(x,y) = k$ and $f(x_0,y_0) = k$, assuming that f is twice differentiable with continuous derivatives, we calculate the second order derivative of the implicit function $y = y(x)$, if $f'_y(x_0,y_0) \neq 0$. We can follow two different ways.

From $\dfrac{\mathrm{d}w}{\mathrm{d}x} = f'_x + f'_y \cdot y' = 0$, we derive again with respect to x; for composite functions derivative (chain) rule we obtain, if $\dfrac{\mathrm{d}y}{\mathrm{d}x} = y'$:

$$\frac{\mathrm{d}^2w}{\mathrm{d}x^2} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}w}{\mathrm{d}x}\right) = \frac{\mathrm{d}}{\mathrm{d}x}\left(f'_x + f'_y\,y'\right) = \frac{\mathrm{d}}{\mathrm{d}x}(f'_x) + \frac{\mathrm{d}}{\mathrm{d}x}\left(f'_y\right)\cdot y' + f'_y \cdot \frac{\mathrm{d}}{\mathrm{d}x}(y') =$$
$$= \frac{\partial}{\partial x}(f'_x)\cdot\frac{\mathrm{d}x}{\mathrm{d}x} + \frac{\partial}{\partial y}(f'_x)\cdot\frac{\mathrm{d}y}{\mathrm{d}x} + \left(\frac{\partial}{\partial x}\left(f'_y\right)\cdot\frac{\mathrm{d}x}{\mathrm{d}x} + \frac{\partial}{\partial y}\left(f'_y\right)\cdot\frac{\mathrm{d}y}{\mathrm{d}x}\right)\cdot y' + f'_y\cdot\frac{\mathrm{d}}{\mathrm{d}x}(y') =$$
$$= f''_{xx} + f''_{xy}\cdot y' + \left(f''_{yx} + f''_{yy}\cdot y'\right)\cdot y' + f'_y\cdot y'' =$$
$$= f''_{xx} + 2f''_{xy}\cdot y' + f''_{yy}\cdot (y')^2 + f'_y\cdot y'' = 0\,,$$

as, w being constant, its second order derivative is zero, and $f''_{xy} = f''_{yx}$. Solving for y'' we get:

$$y'' = -\,\frac{f''_{xx} + 2f''_{xy}\,y' + f''_{yy}\,(y')^2}{f'_y}$$ from which, replacing $y' = -\,\dfrac{f'_x}{f'_y}$, finally we get:

$$y'' = -\,\frac{f''_{xx}\left(f'_y\right)^2 - 2f''_{xy}\,f'_x\,f'_y + f''_{yy}\,(f'_x)^2}{\left(f'_y\right)^3}\,.$$

We can also apply Theorem 18, for $x \to (x,y) \to w = f(x,y) = k$ with $\mathbb{X} = \mathbb{X}(x) = (x,y)$, from which $\mathbb{X}'(x) = (1,y')$ and $\mathbb{X}''(x) = (0,y'')$ and so:

$w'' = \mathbb{X}'(x)\cdot\mathbb{H}(f(\mathbb{X}(x)))\cdot(\mathbb{X}'(x))^{\mathrm{T}} + \nabla f(\mathbb{X}(x))\cdot\mathbb{X}''(x) = 0$ and:

$\begin{Vmatrix} 1 & y' \end{Vmatrix} \cdot \begin{Vmatrix} f''_{xx} & f''_{xy} \\ f''_{xy} & f''_{yy} \end{Vmatrix} \cdot \begin{Vmatrix} 1 \\ y' \end{Vmatrix} + \left(f'_x, f'_y\right)\cdot(0,y'') = 0$ so finally we obtain:

$$y'' = -\,\frac{f''_{xx} + 2f''_{xy}\,y' + f''_{yy}\,(y')^2}{f'_y}$$, and then we substitute $y' = -\,\dfrac{f'_x}{f'_y}$.

Having the second derivative $y''(x_0)$ we can then determine, for the implicit function $y = y(x)$, the expression of Taylor's polynomial of second degree at $x = x_0$, which will be:

$$\mathrm{P}_2(x,x_0) = y_0 - \frac{f'_x(x_0,y_0)}{f'_y(x_0,y_0)}\,(x - x_0) + \frac{1}{2}\,y''(x_0)\,(x - x_0)^2\,,$$

where $y''(x_0)$ must be computed using the previous formula.

Example 47 : Let us compute $y''(1)$ from $f(x,y) = x\,e^y + x^2\,y^2 = 1$.

As $f'_x = e^y + 2x\,y^2$ and $f'_y = x\,e^y + 2x^2\,y$, we get:

$f''_{xx} = 2y^2,\ f''_{xy} = f''_{yx} = e^y + 4x\,y,\ f''_{yy} = x\,e^y + 2x^2$, from which we obtain:

$f''_{xx}(1,0) = 0,\ f''_{xy}(1,0) = 1,\ f''_{yy}(1,0) = 3$, and so, as $f'_x(1,0) = 1$ and $f'_y(1,0) = 1$, we get: $y''(1) = -\,\dfrac{0\cdot 1 - 2\cdot 1\cdot 1\cdot 1 + 3\cdot 1}{1^3} = -\,1$.

Then we can write the expression of Taylor's polynomial of second degree at $x = 1$, which will be:

$$\mathrm{P}_2(x,1) = y(1) + y'(1)\,(x - 1) + \frac{1}{2}\,y''(1)\,(x - 1)^2 = 0 - (x - 1) - \frac{1}{2}\,(x - 1)^2\,.$$

II case: Equation $f(x_1,x_2,y) = k$: Implicit Function $\mathbb{R}^2 \to \mathbb{R}$

Now let us suppose that we have an equation in three variables $f(x_1,x_2,y) = k$, $k \in \mathbb{R}$, and that the point (x_1^0,x_2^0,y_0) satisfies it: $f\left(x_1^0,x_2^0,y_0\right) = k$; f is differentiable with continuous derivatives. A single equation allows us to define implicitly (or explicitly) a variable, say y, as a function of the other two: $y = y(x_1, x_2)$, thus obtaining an implicit function $\mathbb{R}^2 \to \mathbb{R}$. Dini's theorem here applied tells us that we will obtain an implicit continuous and differentiable function if $f'_y\big(x_1^0, x_2^0, y_0\big) \neq 0$.

Implicit Function $\mathbb{R}^2 \to \mathbb{R}$: First order derivatives

If $w = f(x_1, x_2, y) = k$ and $(\mathbb{X}|\, y) = (x_1, x_2, y)$, we have the following functions composition:

$\mathbb{R}^2 \to \mathbb{R}^3 \xrightarrow{f} \mathbb{R},\ \mathbb{X} \to (\mathbb{X}|\, y) \xrightarrow{f} w = f(\mathbb{X}|\, y) = k$.

If $y'_{x_i} = y'_i,\ f'_{x_i} = f'_i$, we can then compute:

$\dfrac{\partial(w)}{\partial(x_1, x_2)} = \dfrac{\partial(w)}{\partial(\mathbb{X}|\, y)} \cdot \dfrac{\partial(\mathbb{X}|\, y)}{\partial(x_1, x_2)} = 0$ as f is constant, and as $\dfrac{\partial x_i}{\partial x_i} = 1$, while $\dfrac{\partial x_i}{\partial x_j} = 0$, as variables x_1 and x_2 are mutually independent, we get:

$$\left\| \begin{matrix} \dfrac{\partial w}{\partial x_1} & \dfrac{\partial w}{\partial x_2} \end{matrix} \right\| = \nabla f(\mathbb{X}|\, y) \cdot \left\| \begin{matrix} 1 & 0 \\ 0 & 1 \\ y'_1 & y'_2 \end{matrix} \right\| = \mathbb{O} \text{ or:}$$

$$\begin{cases} \dfrac{\partial w}{\partial x_1} = \big(f'_1, f'_2, f'_y\big) \cdot (1, 0, y'_1) = f'_1 + f'_y \cdot y'_1 = 0 \\ \dfrac{\partial w}{\partial x_2} = \big(f'_1, f'_2, f'_y\big) \cdot (0, 1, y'_2) = f'_2 + f'_y \cdot y'_2 = 0 \end{cases},$$

from which we obtain: $(y'_1, y'_2) = \left(-\dfrac{f'_1}{f'_y}, -\dfrac{f'_2}{f'_y} \right)$ and so:

$$\nabla y\big(x_1^0, x_2^0\big) = \left(-\frac{f'_1(x_1^0, x_2^0, y_0)}{f'_y(x_1^0, x_2^0, y_0)}, -\frac{f'_2(x_1^0, x_2^0, y_0)}{f'_y(x_1^0, x_2^0, y_0)} \right).$$

Implicit Function $\mathbb{R}^2 \to \mathbb{R}$: Tangent plane equation

We can write the equation of the tangent plane to the surface $y = y(x_1, x_2)$ at (x_1^0, x_2^0), with $y\big(x_1^0, x_2^0\big) = y_0$:

$$y - y_0 = -\frac{f'_1(x_1^0, x_2^0, y_0)}{f'_y(x_1^0, x_2^0, y_0)} \cdot \big(x_1 - x_1^0\big) - \frac{f'_2(x_1^0, x_2^0, y_0)}{f'_y(x_1^0, x_2^0, y_0)} \cdot \big(x_2 - x_2^0\big),$$

which can also be written, as $f'_y\big(x_1^0, x_2^0, y_0\big) \neq 0$, as:

$f'_1\big(x_1^0, x_2^0, y_0\big)\,\big(x_1 - x_1^0\big) + f'_2\big(x_1^0, x_2^0, y_0\big)\,\big(x_2 - x_2^0\big) + f'_y\big(x_1^0, x_2^0, y_0\big)\,(y - y_0) = 0$,

or $\nabla f\big(x_1^0, x_2^0, y_0\big) \cdot \big(x_1 - x_1^0, x_2 - x_2^0, y - y_0\big) = 0$.

Enlarging the dimension of the problem, now the gradient vector is orthogonal to the tangent plane to the level surface.

Implicit Function $\mathbb{R}^2 \to \mathbb{R}$: Second order derivatives

From: $\mathbb{R}^2 \to \mathbb{R}^3 \xrightarrow{f} \mathbb{R}$: $\mathbb{X} \to (\mathbb{X}|\, y) \xrightarrow{f} w = f(\mathbb{X}|\, y) = k$, applying Theorem 18, if: $y''_{x_i x_j} = y''_{ij}$, $f''_{x_i x_j} = f''_{ij}$ as $(\mathbb{X}|\, y) = (x_1, x_2, y)$, from which:

$\dfrac{\partial(\mathbb{X}|\, y)}{\partial x_1} = (1, 0, y'_1)$ and $\dfrac{\partial(\mathbb{X}|\, y)}{\partial x_2} = (0, 1, y'_2)$, we get also:

$\dfrac{\partial^2(\mathbb{X}|\, y)}{\partial x_1^2} = (0, 0, y''_{11})$, $\dfrac{\partial^2(\mathbb{X}|\, y)}{\partial x_1\, \partial x_2} = \dfrac{\partial^2(\mathbb{X}|\, y)}{\partial x_2\, \partial x_1} = (0, 0, y''_{12})$ and $\dfrac{\partial^2(\mathbb{X}|\, y)}{\partial x_2^2} = (0, 0, y''_{22})$,

and so: $\dfrac{\partial^2 w}{\partial x_1^2} = \| 1 \quad 0 \quad y'_1 \| \cdot \left\| \begin{matrix} f''_{11} & f''_{12} & f''_{1y} \\ f''_{12} & f''_{22} & f''_{2y} \\ f''_{1y} & f''_{2y} & f''_{yy} \end{matrix} \right\| \cdot \left\| \begin{matrix} 1 \\ 0 \\ y'_1 \end{matrix} \right\| + \big(f'_1, f'_2, f'_y\big) \cdot (0, 0, y''_{11}) = 0$, from which:

$$y''_{11} = -\frac{f''_{11} + 2f''_{1y}\,y'_1 + f''_{yy}\,(y'_1)^2}{f'_y}\,;$$

$$\frac{\partial^2 w}{\partial x_1\,\partial x_2} = \|\,1 \quad 0 \quad y'_1\,\| \cdot \left\| \begin{matrix} f''_{11} & f''_{12} & f''_{1y} \\ f''_{12} & f''_{22} & f''_{2y} \\ f''_{1y} & f''_{2y} & f''_{yy} \end{matrix} \right\| \cdot \left\| \begin{matrix} 0 \\ 1 \\ y'_2 \end{matrix} \right\| + \left(f'_1, f'_2, f'_y\right) \cdot (0, 0, y''_{12}) = 0\,, \text{ or:}$$

$$y''_{12} = -\frac{f''_{12} + f''_{1y}\,y'_2 + f''_{2y}\,y'_1 + f''_{yy}\,y'_1\,y'_2}{f'_y}\,;$$

$$\frac{\partial^2 w}{\partial x_2^2} = \|\,0 \quad 1 \quad y'_2\,\| \cdot \left\| \begin{matrix} f''_{11} & f''_{12} & f''_{1y} \\ f''_{12} & f''_{22} & f''_{2y} \\ f''_{1y} & f''_{2y} & f''_{yy} \end{matrix} \right\| \cdot \left\| \begin{matrix} 0 \\ 1 \\ y'_2 \end{matrix} \right\| + \left(f'_1, f'_2, f'_y\right) \cdot (0, 0, y''_{22}) = 0\,, \text{ or:}$$

$$y''_{22} = -\frac{f''_{22} + 2f''_{2y}\,y'_2 + f''_{yy}\,(y'_2)^2}{f'_y}\,.$$

Implicit Function $\mathbb{R}^2 \to \mathbb{R}$: I and II order total differentials; Taylor's polynomial

Since we have now a function of several variables, to search its maximum and minimum points, as we shall see below, first and second order total differentials are more useful. Let us see how to compute, starting from the equation $f(x_1, x_2, y) = k$, the first and second order total differentials of an implicit function $y = y(x_1, x_2)$.

As $y'_1 = -\dfrac{f'_1}{f'_y}$ and $y'_2 = -\dfrac{f'_2}{f'_y}$, replacing we have:

$\mathrm{d}y = y'_1\,\mathrm{d}x_1 + y'_2\,\mathrm{d}x_2 = -\dfrac{f'_1}{f'_y}\,\mathrm{d}x_1 - \dfrac{f'_2}{f'_y}\,\mathrm{d}x_2$, which could also be computed differentiating:

$\mathrm{d}w = f'_1\,\mathrm{d}x_1 + f'_2\,\mathrm{d}x_2 + f'_y\,\mathrm{d}y = 0$, and from this we deduce $\mathrm{d}y = -\dfrac{f'_1}{f'_y}\,\mathrm{d}x_1 - \dfrac{f'_2}{f'_y}\,\mathrm{d}x_2$, if $f'_y \neq 0$.

To get $\mathrm{d}^2 y$ we differentiate again $\mathrm{d}w = f'_1\,\mathrm{d}x_1 + f'_2\,\mathrm{d}x_2 + f'_y\,\mathrm{d}y = 0$, bearing in mind that f'_1, f'_2 and f'_y depend on x_1, x_2 and y, but now also $\mathrm{d}y$ depends on x_1 and x_2; so we get:

$$\mathrm{d}(\mathrm{d}w) = 0 = \mathrm{d}\big(f'_1\,\mathrm{d}x_1 + f'_2\,\mathrm{d}x_2 + f'_y\,\mathrm{d}y\big) = \frac{\partial}{\partial x_1}\big(f'_1\,\mathrm{d}x_1 + f'_2\,\mathrm{d}x_2 + f'_y\,\mathrm{d}y\big)\,\mathrm{d}x_1 +$$
$$+ \frac{\partial}{\partial x_2}\big(f'_1\,\mathrm{d}x_1 + f'_2\,\mathrm{d}x_2 + f'_y\,\mathrm{d}y\big)\,\mathrm{d}x_2 + \frac{\partial}{\partial y}\big(f'_1\,\mathrm{d}x_1 + f'_2\,\mathrm{d}x_2 + f'_y\,\mathrm{d}y\big)\,\mathrm{d}y =$$
$$= \left(f''_{11}\,\mathrm{d}x_1 + f''_{21}\,\mathrm{d}x_2 + f''_{y1}\,\mathrm{d}y + f'_y\,\frac{\partial}{\partial x_1}(\mathrm{d}y)\right)\mathrm{d}x_1 +$$
$$+ \left(f''_{12}\,\mathrm{d}x_1 + f''_{22}\,\mathrm{d}x_2 + f''_{y2}\,\mathrm{d}y + f'_y\,\frac{\partial}{\partial x_2}(\mathrm{d}y)\right)\mathrm{d}x_2 +$$
$$+ \left(f''_{1y}\,\mathrm{d}x_1 + f''_{2y}\,\mathrm{d}x_2 + f''_{yy}\,\mathrm{d}y + f'_y\,\frac{\partial}{\partial y}(\mathrm{d}y)\right)\mathrm{d}y = 0\,.$$

But

$$\big(f''_{11}\,\mathrm{d}x_1 + f''_{21}\,\mathrm{d}x_2 + f''_{y1}\,\mathrm{d}y\big)\,\mathrm{d}x_1 + \big(f''_{12}\,\mathrm{d}x_1 + f''_{22}\,\mathrm{d}x_2 + f''_{y2}\,\mathrm{d}y\big)\,\mathrm{d}x_2 +$$
$$+ \big(f''_{1y}\,\mathrm{d}x_1 + f''_{2y}\,\mathrm{d}x_2 + f''_{yy}\,\mathrm{d}y\big)\,\mathrm{d}y = \mathrm{d}^2 f(\mathbb{X}\,|\,y)$$

while

$f'_y \cdot \dfrac{\partial}{\partial x_1}(\mathrm{d}y)\,\mathrm{d}x_1 + f'_y \cdot \dfrac{\partial}{\partial x_2}(\mathrm{d}y)\,\mathrm{d}x_2 + f'_y \cdot \dfrac{\partial}{\partial y}(\mathrm{d}y)\,\mathrm{d}y = f'_y\,\mathrm{d}(\mathrm{d}y) = f'_y\,\mathrm{d}^2 y$, and so we have: $\mathrm{d}^2 f(\mathbb{X}\,|\,y) + f'_y\,\mathrm{d}^2 y = 0$, from which we finally get:

$$\mathrm{d}^2 y = -\frac{\mathrm{d}^2 f(x_1, x_2, y)}{f'_y} = -\frac{\mathrm{d}^2 f(\mathbb{X}\,|\,y)}{f'_y}\,.$$

We use the expressions just found for $\mathrm{d}y$ and d^2y to write the expression of second degree Taylor's polynomial of the implicit function $y = y(x_1, x_2)$ which will be:

$$\mathrm{P}_2(x_1, x_2) = y_0 + \mathrm{d}y + \frac{1}{2}\,\mathrm{d}^2y = y_0 - \frac{f_1'}{f_y'}\,\mathrm{d}x_1 - \frac{f_2'}{f_y'}\,\mathrm{d}x_2 - \frac{1}{2}\,\frac{\mathrm{d}^2 f(\mathbb{X}\,|\,y)}{f_y'}\,.$$

Example 48 : Given $f(x,y,z) = x\,e^{x\,(y-z)} + y - z = 0$, it is $f(0,0,0) = 0$. As:
$f_x' = e^{x\,(y-z)} + x\,e^{x\,(y-z)}\,(y-z) \Rightarrow f_x'(0,0,0) = 1 \neq 0$;
$f_y' = x^2\,e^{x\,(y-z)} + 1 \Rightarrow f_y'(0,0,0) = 1 \neq 0$,
$f_z' = -\,x^2\,e^{x\,(y-z)} - 1 \Rightarrow f_z'(0,0,0) = -\,1 \neq 0$,
we have three possible choices to define an implicit function: $x = x(y,z)$, $y = y(x,z)$ or $z = z(x,y)$. We choose the third option and we get:

$$\frac{\partial z(0,0)}{\partial x} = -\,\frac{f_x'(0,0,0)}{f_z'(0,0,0)} = -\,\frac{1}{-\,1} = 1\,,\quad \frac{\partial z(0,0)}{\partial y} = -\,\frac{f_y'(0,0,0)}{f_z'(0,0,0)} = -\,\frac{1}{-\,1} = 1$$

from which:
$\mathrm{d}z(0,0) = 1\cdot\mathrm{d}x + 1\cdot\mathrm{d}y = \mathrm{d}x + \mathrm{d}y$. Since then:
$f_{xx}'' = 2\,(y-z)\,e^{x\,(y-z)} + x\,e^{x\,(y-z)}\,(y-z)^2 \Rightarrow f_{xx}''(0,0,0) = 0$,
$f_{yy}'' = x^3\,e^{x\,(y-z)} \Rightarrow f_{yy}''(0,0,0) = 0$,
$f_{zz}'' = x^3\,e^{x\,(y-z)} \Rightarrow f_{zz}''(0,0,0) = 0$,
$f_{xy}'' = f_{yx}'' = 2x\,e^{x\,(y-z)} + x^2\,e^{x\,(y-z)}\,(y-z) \Rightarrow f_{xy}''(0,0,0) = 0$,
$f_{xz}'' = f_{zx}'' = -\,2x\,e^{x\,(y-z)} - x^2\,e^{x\,(y-z)}\,(y-z) \Rightarrow f_{xz}''(0,0,0) = 0$,
$f_{yz}'' = f_{zy}'' = -\,x^3\,e^{x\,(y-z)} \Rightarrow f_{yz}''(0,0,0) = 0$,
from which we get: $\mathrm{d}^2z = -\,\dfrac{\mathrm{d}^2 f}{f_z'} = -\,\dfrac{0}{-\,1} = 0$, and so:
$\mathrm{P}_2(x,y,0,0) = z_0 + \mathrm{d}z + \dfrac{1}{2}\,\mathrm{d}^2z = 0 + \mathrm{d}x + \mathrm{d}y + 0 = \mathrm{d}x + \mathrm{d}y$; as:
$\mathrm{d}x = x - 0 = x$ and $\mathrm{d}y = y - 0 = y$, finally, we have $\mathrm{P}_2(x,y,0,0) = x + y$.

Now we use the equation $f(x,y,z) = x\,e^{x\,(y-z)} + y - z = 1$. It is $f(1,0,0) = 0$.
Moreover, since first and second derivatives of the function do not change, we have: $f_x'(1,0,0) = 1 \neq 0$; $f_y'(1,0,0) = 2 \neq 0$ and $f_z'(1,0,0) = -\,2 \neq 0$, the same three possibilities for implicit function remain valid; we choose again $z = z(x,y)$, and so:

$$\mathrm{d}z = -\,\frac{f_x'(1,0,0)}{f_z'(1,0,0)}\,\mathrm{d}x - \frac{f_y'(1,0,0)}{f_z'(1,0,0)}\,\mathrm{d}y = \frac{1}{2}\,\mathrm{d}x + \mathrm{d}y\,.$$ Then:

$f_{xx}''(1,0,0) = 0$, $f_{yy}''(1,0,0) = 1$, $f_{zz}''(1,0,0) = 1$, $f_{xy}''(1,0,0) = 2$, $f_{xz}''(1,0,0) = -\,2$, $f_{yz}''(1,0,0) = -\,1$, and so:
$\mathrm{d}^2 f(1,0,0) = (\mathrm{d}y)^2 + (\mathrm{d}z)^2 + 4\mathrm{d}x\,\mathrm{d}y - 4\mathrm{d}x\,\mathrm{d}z - 2\mathrm{d}y\,\mathrm{d}z$, from which we get:

$$\mathrm{d}^2 z(1,0,0) = -\,\frac{\mathrm{d}^2 f(1,0,0)}{f_z'(1,0,0)} = -\,\frac{(\mathrm{d}y)^2 + (\mathrm{d}z)^2 + 4\mathrm{d}x\,\mathrm{d}y - 4\mathrm{d}x\,\mathrm{d}z - 2\mathrm{d}y\,\mathrm{d}z}{-\,2}\,;$$

as $\mathrm{d}z = \dfrac{1}{2}\,\mathrm{d}x + \mathrm{d}y$, and replacing, we obtain:

$$\mathrm{d}^2 z = \frac{1}{2}\left[(\mathrm{d}y)^2 + \left(\frac{1}{2}\,\mathrm{d}x + \mathrm{d}y\right)^2 + 4\mathrm{d}x\,\mathrm{d}y - 4\mathrm{d}x\left(\frac{1}{2}\,\mathrm{d}x + \mathrm{d}y\right) - 2\mathrm{d}y\left(\frac{1}{2}\,\mathrm{d}x + \mathrm{d}y\right)\right]$$

and so $\mathrm{d}^2 z(1,0,0) = -\,\dfrac{7}{8}\,(\mathrm{d}x)^2$, wherefore Taylor's second degree polynomial will be:
$\mathrm{P}_2(x,y,0,0) = z_0 + \mathrm{d}z + \dfrac{1}{2}\,\mathrm{d}^2 z = 0 + \dfrac{1}{2}\,\mathrm{d}x + \mathrm{d}y - \dfrac{7}{16}\,(\mathrm{d}x)^2$; as:
$\mathrm{d}x = x - 1$ and $\mathrm{d}y = y - 0 = y$, we finally get:

$$\mathrm{P}_2(x,y,0,0)=\frac{1}{2}(x-1)+y-\frac{7}{16}(x-1)^2 .$$

III Case: Equation $f(x_1,x_2,...,x_n,y)=k$: Implicit Function $\mathbb{R}^n\to\mathbb{R}$

Generalizing the two cases above, let us suppose that we have a single equation in $n+1$ variables : $f(x_1,x_2,...,x_n,y)=k$, $k\in\mathbb{R}$, and $f:\mathbb{R}^{n+1}\to\mathbb{R}$ differentiable function with continuous derivatives. If $\mathbb{X}=(x_1,x_2,...,x_n)$ and $(\mathbb{X}|\,y)=(x_1,x_2,...,x_n,y)$, let $f(\mathbb{X}_0|\,y_0)=k$ and $f'_y(\mathbb{X}_0|\,y_0)\neq 0$. Then the implicit function $y=y(x_1,x_2,...,x_n)$ exists in a neighborhood $\mathfrak{I}(\mathbb{X}_0)$, with $y_0=y(\mathbb{X}_0)$, and we have the following functions composition:

$\mathbb{R}^n\to\mathbb{R}^{n+1}\xrightarrow{f}\mathbb{R}$, $\mathbb{X}\to(\mathbb{X}\,|\,y)\xrightarrow{f}w=f(\mathbb{X}\,|\,y)=k$.

Implicit Function $\mathbb{R}^n\to\mathbb{R}$: First order derivatives

From $\mathbb{X}\to(\mathbb{X}\,|\,y)\xrightarrow{f}w=f(\mathbb{X}\,|\,y)=k$, deriving with respect to the variable x_i we get:

$$\frac{\partial w}{\partial x_i}=\frac{\partial(w)}{\partial(\mathbb{X}\,|\,y)}\cdot\frac{\partial(\mathbb{X}\,|\,y)}{\partial x_i}=\nabla f(\mathbb{X}\,|\,y)\cdot\frac{\partial(\mathbb{X}\,|\,y)}{\partial x_i}=\frac{\partial f}{\partial x_i}+\frac{\partial f}{\partial y}\cdot\frac{\partial y}{\partial x_i}=f'_i+f'_y\cdot\frac{\partial y}{\partial x_i}=0$$

as $\dfrac{\partial x_i}{\partial x_i}=1$ and $\dfrac{\partial x_j}{\partial x_i}=0$ if $i\neq j$, as the variables x_i are mutually independent, for which we obtain: $\dfrac{\partial y}{\partial x_i}=-\dfrac{f'_i}{f'_y}$, $1\le i\le n$.

Implicit Function $\mathbb{R}^n\to\mathbb{R}$: Tangent hyperplane equation

It is $\nabla y=\left(-\dfrac{f'_1}{f'_y},-\dfrac{f'_2}{f'_y},...,-\dfrac{f'_n}{f'_y}\right)$ and the equation of the tangent hyperplane to the hypersurface $y=y(x_1,x_2,...,x_n)$ at $(x_1^0,x_2^0,...,x_n^0)$ is:

$y-y_0=\displaystyle\sum_{i:1}^{n}\left(-\frac{f'_i}{f'_y}\right)\cdot\left(x_i-x_i^0\right)$, i.e. $\displaystyle\sum_{i:1}^{n}f'_i\cdot\left(x_i-x_i^0\right)+f'_y\cdot(y-y_0)=0$, or:

$\nabla f\left(x_1^0,x_2^0,...,x_n^0,y_0\right)\cdot\left(x_1-x_1^0,x_2-x_2^0,...,x_n-x_n^0,y-y_0\right)=0$, that expresses the usual relation of orthogonality between the gradient of the function f and, now, the tangent hyperplane.

Implicit Function $\mathbb{R}^n\to\mathbb{R}$: Second order derivatives

From:

$\mathbb{R}^n\to\mathbb{R}^{n+1}\xrightarrow{f}\mathbb{R}\ :\ \mathbb{X}\to(\mathbb{X}|\,y)\xrightarrow{f}w=f(\mathbb{X}|\,y)=k$, deriving we get:

$\dfrac{\partial(\mathbb{X}|\,y)}{\partial x_i}=(0,..,1_i,..,0,y'_i)$ and also:

$\dfrac{\partial^2(\mathbb{X}|\,y)}{\partial x_i\,\partial x_j}=\left(0,..,0,y''_{ij}\right)$, and so, applying Teorema 18, we get:

$$\frac{\partial^2 w}{\partial x_i\,\partial x_j}=\frac{\partial(\mathbb{X}|\,y)}{\partial x_i}\cdot\mathbb{H}(f(\mathbb{X}|\,y))\cdot\left(\frac{\partial(\mathbb{X}|\,y)}{\partial x_j}\right)^{\mathrm{T}}+\left(f'_1,..,f'_n,f'_y\right)\cdot\left(0,..,0,y''_{ij}\right)=0$$

from which we get the second order partial derivatives:

$$y''_{ij}=\frac{\partial^2 y}{\partial x_i\,\partial x_j}=-\frac{1}{f'_y}\cdot\frac{\partial(\mathbb{X}|\,y)}{\partial x_i}\cdot\mathbb{H}(f(\mathbb{X}|\,y))\cdot\left(\frac{\partial(\mathbb{X}|\,y)}{\partial x_j}\right)^{\mathrm{T}},\ 1\le i,j\le n\ .$$

Implicit Function $\mathbb{R}^n\to\mathbb{R}$: I and II order total differentials; Taylor's polynomial

Operating in a similar manner to the case of implicit function $\mathbb{R}^2\to\mathbb{R}$, from the composition $\mathbb{X}\to(\mathbb{X}|\,y)\xrightarrow{f}w=f(\mathbb{X}|\,y)=k$ we get:

$\mathrm{d}w = f_1' \,\mathrm{d}x_1 + f_2' \,\mathrm{d}x_2 + .. + f_{x_n}' \,\mathrm{d}x_n + f_y' \,\mathrm{d}y = 0$, from which we get d$y$, and differentiating again, we get $\mathrm{d}^2 f(\mathbb{X}|\, y) + f_y' \,\mathrm{d}^2 y = 0$ and so $\mathrm{d}^2 y = -\dfrac{\mathrm{d}^2 f(\mathbb{X}|\, y)}{f_y'}$.

With these differentials we can write Taylor's polynomial.

We can summarize these cases, all with a single equation in two or more variables, saying that it is possible to define an implicit function, with a single dependent variable, while all others remain independent, if the gradient of the function (of the equation) at the considered point is different from the null vector, i.e. if it has at least one component different from zero. This derivation variable can then be taken as the dependent variable. Since the gradient is still a Jacobian matrix, even if formed by a single line, we can say that this Jacobian should have rank equal to 1, i.e. maximum, and this condition will be valid in the general case.

FUNCTIONS IMPLICITLY DEFINED BY A SYSTEM OF EQUATIONS

I Case: System $\begin{cases} f(x, y_1, y_2) = k_1 \\ g(x, y_1, y_2) = k_2 \end{cases}$ **: Implicit Function** $\mathbb{R} \to \mathbb{R}^2$

We don't increase now the number of variables but the number of equations, having, as a minimum case, a system of two equations in three variables: $\begin{cases} f(x, y_1, y_2) = k_1 \\ g(x, y_1, y_2) = k_2 \end{cases}$, $k_1, k_2 \in \mathbb{R}$.

Two equations can allow us to explain two variables, say y_1 and y_2, as a function of the remaining x: $\mathbb{R} \to \mathbb{R}^2, x \to (y_1(x), y_2(x))$, if the appropriate assumptions are met.

The equations number tells us how many the dependent variables can be, the remaining variables will then have the role of independent ones. Let us see what we get as far as the derivatives of a function $\mathbb{R} \to \mathbb{R}^2, x \to (y_1(x), y_2(x))$ defined by such a system of equations.

Implicit Function $\mathbb{R} \to \mathbb{R}^2$ **: First order derivatives**

From $\begin{cases} w_1 = f(x, y_1, y_2) = k_1 \\ w_2 = g(x, y_1, y_2) = k_2 \end{cases}$ with $(x|\, \mathbb{Y}) = (x, y_1, y_2)$, we have these two function compositions:

$\mathbb{R} \to \mathbb{R}^3 \xrightarrow{f} \mathbb{R},\ x \to (x|\, \mathbb{Y}) \xrightarrow{f} w_1 = f(x|\, \mathbb{Y}) = k_1$, and

$\mathbb{R} \to \mathbb{R}^3 \xrightarrow{g} \mathbb{R},\ x \to (x|\, \mathbb{Y}) \xrightarrow{g} w_2 = g(x|\, \mathbb{Y}) = k_2$,

from which, differentiating with respect to x, as w_1 and w_2 are constant, we get

$\dfrac{\partial(w_1, w_2)}{\partial(x)} = \dfrac{\partial(f, g)}{\partial(x)} = \dfrac{\partial(f, g)}{\partial(x|\, \mathbb{Y})} \cdot \dfrac{\partial(x|\, \mathbb{Y})}{\partial(x)} = \mathbb{O}$, i.e. the system:

$$\begin{cases} \dfrac{\mathrm{d}w_1}{\mathrm{d}x} = \nabla f(x|\, \mathbb{Y}) \cdot \dfrac{\partial(x|\, \mathbb{Y})}{\partial x} = 0 \\ \dfrac{\mathrm{d}w_2}{\mathrm{d}x} = \nabla g(x|\, \mathbb{Y}) \cdot \dfrac{\partial(x|\, \mathbb{Y})}{\partial x} = 0 \end{cases} \text{ equivalent to:}$$

$$\begin{cases} \dfrac{\partial f}{\partial x} \cdot \dfrac{\mathrm{d}x}{\mathrm{d}x} + \dfrac{\partial f}{\partial y_1} \cdot \dfrac{\mathrm{d}y_1}{\mathrm{d}x} + \dfrac{\partial f}{\partial y_2} \cdot \dfrac{\mathrm{d}y_2}{\mathrm{d}x} = 0 \\ \dfrac{\partial g}{\partial x} \cdot \dfrac{\mathrm{d}x}{\mathrm{d}x} + \dfrac{\partial g}{\partial y_1} \cdot \dfrac{\mathrm{d}y_1}{\mathrm{d}x} + \dfrac{\partial g}{\partial y_2} \cdot \dfrac{\mathrm{d}y_2}{\mathrm{d}x} = 0 \end{cases} \Rightarrow \begin{cases} f_{y_1}' \cdot y_1' + f_{y_2}' \cdot y_2' = -f_x' \\ g_{y_1}' \cdot y_1' + g_{y_2}' \cdot y_2' = -g_x' \end{cases}$$

, which is a linear system of two equations in two variables y_1' and y_2', and can be written in matrix form as: $\begin{Vmatrix} f_{y_1}' & f_{y_2}' \\ g_{y_1}' & g_{y_2}' \end{Vmatrix} \cdot \begin{Vmatrix} y_1' \\ y_2' \end{Vmatrix} = -\begin{Vmatrix} f_x' \\ g_x' \end{Vmatrix}$, and also, using Jacobian matrices, as:

$\dfrac{\partial(f, g)}{\partial(y_1, y_2)} \cdot \dfrac{\partial(y_1, y_2)}{\partial(x)} = -\dfrac{\partial(f, g)}{\partial(x)}$.

From Cramer's theorem, if $\left\|\begin{matrix} f'_{y_1} & f'_{y_2} \\ g'_{y_1} & g'_{y_2} \end{matrix}\right\| = \left|\dfrac{\partial(f,g)}{\partial(y_1,y_2)}\right| \neq 0$ the linear system has a unique solution $\left(\dfrac{\partial y_1}{\partial x}, \dfrac{\partial y_2}{\partial x}\right)$.

Since the Jacobian matrix of functions f and g with respect to the dependent variables y_1 and y_2 is not singular, (i.e. has maximum rank, in this case equal to 2), this is the condition that allows us to state the Dini's theorem for this case. In fact the following holds:

Theorem 20 : Given the system $\begin{cases} f(x,y_1,y_2) = k_1 \\ g(x,y_1,y_2) = k_2 \end{cases}$, with f and g, $\mathbb{R}^3 \to \mathbb{R}$, differentiable functions with continuous derivatives, being (x_0, y_1^0, y_2^0) a point that satisfies the system, and then is $\left|\dfrac{\partial(f,g)(x_0,y_1^0,y_2^0)}{\partial(y_1,y_2)}\right| \neq 0$. Then there exists a neighborhood $\mathfrak{J}(x_0)$ in which an implicit function $\mathbb{R} \to \mathbb{R}^2, x \to (y_1(x), y_2(x))$ is defined, which is continuous and differentiable $\forall\, x \in \mathfrak{J}(x_0)$, whose derivatives are:

$$\left\|\begin{matrix} y'_1 \\ y'_2 \end{matrix}\right\| = \frac{\partial(y_1,y_2)}{\partial(x)} = -\left\|\frac{\partial(f,g)}{\partial(y_1,y_2)}\right\|^{-1} \cdot \frac{\partial(f,g)}{\partial(x)} = -\left\|\begin{matrix} f'_{y_1} & f'_{y_2} \\ g'_{y_1} & g'_{y_2} \end{matrix}\right\|^{-1} \cdot \left\|\begin{matrix} f'_x \\ g'_x \end{matrix}\right\|.$$

In addition to the global solution expressed using the inverse of the Jacobian matrix, there is another process, practical consequence of Cramer's theorem on linear systems, which allows us to calculate individually each unknown. Each of them is in fact given by a quotient, the denominator of which is the determinant of the coefficient matrix of the unknowns (the Jacobian) and the numerator is the determinant of the matrix obtained replacing in the Jacobian matrix the column of the known terms to the column of the coefficients of the sought unknown. Then we will get, in the case we are dealing with:

$$\frac{\mathrm{d}y_1}{\mathrm{d}x} = -\frac{\begin{vmatrix} f'_x & f'_{y_2} \\ g'_x & g'_{y_2} \end{vmatrix}}{\begin{vmatrix} f'_{y_1} & f'_{y_2} \\ g'_{y_1} & g'_{y_2} \end{vmatrix}}\ ;\ \frac{\mathrm{d}y_2}{\mathrm{d}x} = -\frac{\begin{vmatrix} f'_{y_1} & f'_x \\ g'_{y_1} & g'_x \end{vmatrix}}{\begin{vmatrix} f'_{y_1} & f'_{y_2} \\ g'_{y_1} & g'_{y_2} \end{vmatrix}}.$$

Example 49 : Given the system $\begin{cases} f(x,y_1,y_2) = x - y_1 + \sin(y_1 - y_2) + \cos(x - y_2) = 1 \\ g(x,y_1,y_2) = x + y_1 - \sin(x - y_2) - \cos(y_2 - y_1) = 1 \end{cases}$, it is satisfied by the point $\mathrm{P}_0 = (x, y_1, y_2) = (1,1,1)$, and functions f and g are differentiable with continuous derivatives throughout $\mathbb{R}^2$. Let us calculate the Jacobian matrix:

$$\frac{\partial(f,g)}{\partial(x,y_1,y_2)} = \left\|\begin{matrix} 1 - \sin(x - y_2) & \cos(y_1 - y_2) - 1 & \sin(x - y_2) - \cos(y_1 - y_2) \\ 1 - \cos(x - y_2) & 1 - \sin(y_2 - y_1) & \cos(x - y_2) + \sin(y_2 - y_1) \end{matrix}\right\|.$$

Calculating the Jacobian at P_0 we get $\dfrac{\partial(f,g)(1,1,1)}{\partial(x,y_1,y_2)} = \left\|\begin{matrix} 1 & 0 & -1 \\ 0 & 1 & 1 \end{matrix}\right\|$, whose rank is maximum and equal to 2 as $\begin{vmatrix} 0 & -1 \\ 1 & 1 \end{vmatrix} \neq 0$, and then in a neighborhood of $x = 1$ an implicit function is defined $\mathbb{R} \to \mathbb{R}^2, x \to (y_1(x), y_2(x))$, continuous and differentiable function, whose derivatives are given by:

$$\frac{\mathrm{d}y_1}{\mathrm{d}x} = -\frac{\begin{vmatrix} 1 - \sin(x - y_2) & \sin(x - y_2) - \cos(y_1 - y_2) \\ 1 - \cos(x - y_2) & \cos(x - y_2) + \sin(y_2 - y_1) \end{vmatrix}}{\begin{vmatrix} \cos(y_1 - y_2) - 1 & \sin(x - y_2) - \cos(y_1 - y_2) \\ 1 - \sin(y_2 - y_1) & \cos(x - y_2) + \sin(y_2 - y_1) \end{vmatrix}}\ ;$$

$$\frac{\mathrm{d}y_2}{\mathrm{d}x} = -\frac{\begin{vmatrix} \cos(y_1 - y_2) - 1 & 1 - \sin(x - y_2) \\ 1 - \sin(y_2 - y_1) & 1 - \cos(x - y_2) \end{vmatrix}}{\begin{vmatrix} \cos(y_1 - y_2) - 1 & \sin(x - y_2) - \cos(y_1 - y_2) \\ 1 - \sin(y_2 - y_1) & \cos(x - y_2) + \sin(y_2 - y_1) \end{vmatrix}}.$$

At $x = 1$ we have: $\dfrac{\mathrm{d}y_1(1)}{\mathrm{d}x} = -\dfrac{\begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix}}{\begin{vmatrix} 0 & -1 \\ 1 & 1 \end{vmatrix}} = -1$ and $\dfrac{\mathrm{d}y_2(1)}{\mathrm{d}x} = -\dfrac{\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}}{\begin{vmatrix} 0 & -1 \\ 1 & 1 \end{vmatrix}} = 1$.

Example 50 : Given the system: $\begin{cases} z = f(x,y) \\ g(x,y) = 0 \end{cases}$, let's see it as $\begin{cases} F(x,y,z) = f(x,y) - z \\ g(x,y) = 0 \end{cases}$.
Let P_0 be a point satisfying the system, with f and g everywhere differentiable with continuous derivatives.
It will be $\dfrac{\partial(F,g)}{\partial(x,y,z)} = \left\| \begin{matrix} f'_x & f'_y & -1 \\ g'_x & g'_y & 0 \end{matrix} \right\|$. If $\left| \dfrac{\partial(F,g)(P_0)}{\partial(y,z)} \right| = \begin{vmatrix} f'_y & -1 \\ g'_y & 0 \end{vmatrix} = g'_y \neq 0$, the hypotheses of Dini's theorem are satisfied and so the system defines an implicit function $\mathbb{R} \to \mathbb{R}^2, x \to (y(x), z(x))$. To calculate the derivatives $y'(x)$ and $z'(x)$ we can proceed in two ways:

1) from $g(x,y) = 0$, differentiating with respect to x, we get: $g'_x \cdot 1 + g'_y \dfrac{\mathrm{d}y}{\mathrm{d}x} = 0$, from which $\dfrac{\mathrm{d}y}{\mathrm{d}x} = -\dfrac{g'_x}{g'_y}$. Differentiating the first equation $f(x,y) - z = 0$ with respect to x we get: $f'_x \cdot 1 + f'_y \dfrac{\mathrm{d}y}{\mathrm{d}x} + (-1)\dfrac{\mathrm{d}z}{\mathrm{d}x} = 0$, from which: $f'_x + f'_y\left(-\dfrac{g'_x}{g'_y}\right) - \dfrac{\mathrm{d}z}{\mathrm{d}x} = 0$ and so:

$$\frac{\mathrm{d}z}{\mathrm{d}x} = f'_x - f'_y \frac{g'_x}{g'_y} = \frac{f'_x g'_y - f'_y g'_x}{g'_y}.$$

2) using the Jacobian matrix, we have: $\left\| \begin{matrix} f'_y & -1 \\ g'_y & 0 \end{matrix} \right\| \cdot \left\| \begin{matrix} \dfrac{\mathrm{d}y}{\mathrm{d}x} \\ \dfrac{\mathrm{d}z}{\mathrm{d}x} \end{matrix} \right\| = -\left\| \begin{matrix} f'_x \\ g'_x \end{matrix} \right\|$, from which:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{\begin{vmatrix} f'_x & -1 \\ g'_x & 0 \end{vmatrix}}{\begin{vmatrix} f'_y & -1 \\ g'_y & 0 \end{vmatrix}} = -\frac{g'_x}{g'_y} \quad \text{and} \quad \frac{\mathrm{d}z}{\mathrm{d}x} = -\frac{\begin{vmatrix} f'_y & f'_x \\ g'_y & g'_x \end{vmatrix}}{\begin{vmatrix} f'_y & -1 \\ g'_y & 0 \end{vmatrix}} = \frac{f'_x g'_y - f'_y g'_x}{g'_y}.$$

Example 51 : The system: $\begin{cases} z = f(x,y) \\ g(x,y,z) = 0 \end{cases}$, is the same as $\begin{cases} F(x,y,z) = f(x,y) - z \\ g(x,y,z) = 0 \end{cases}$, the appropriate assumptions are met as in the previous example. Even now a variable, z, is already in explicit form, while the second equation allows us to get, even if only implicitly, another variable, say y.
As: $\dfrac{\partial(F,g)}{\partial(x,y,z)} = \left\| \begin{matrix} f'_x & f'_y & -1 \\ g'_x & g'_y & g'_z \end{matrix} \right\|$, if at P_0 is $\begin{vmatrix} f'_y & -1 \\ g'_y & g'_z \end{vmatrix} \neq 0$, we have a function $\mathbb{R} \to \mathbb{R}^2, x \to (y(x), z(x))$, whose derivatives are given by:

$\left\| \begin{matrix} f'_y & -1 \\ g'_y & g'_z \end{matrix} \right\| \cdot \left\| \begin{matrix} \dfrac{\mathrm{d}y}{\mathrm{d}x} \\ \dfrac{\mathrm{d}z}{\mathrm{d}x} \end{matrix} \right\| = -\left\| \begin{matrix} f'_x \\ g'_x \end{matrix} \right\|$, from which we obtain:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{\begin{vmatrix} f'_x & -1 \\ g'_x & g'_z \end{vmatrix}}{\begin{vmatrix} f'_y & -1 \\ g'_y & g'_z \end{vmatrix}} = -\frac{f'_x\, g'_z + g'_x}{f'_y\, g'_z + g'_y}\,, \quad \frac{\mathrm{d}z}{\mathrm{d}x} = -\frac{\begin{vmatrix} f'_y & f'_x \\ g'_y & g'_x \end{vmatrix}}{\begin{vmatrix} f'_y & -1 \\ g'_y & g'_z \end{vmatrix}} = -\frac{f'_y\, g'_x - f'_x\, g'_y}{f'_y\, g'_z + g'_y}\,.$$

The results, as we see, are different from those of the previous example, as the second equation is not constant with respect to z, but it depends upon all the variables.

II Case: System $\begin{cases} f(x_1, x_2, y_1, y_2) = k_1 \\ g(x_1, x_2, y_1, y_2) = k_2 \end{cases}$ **: Implicit Function** $\mathbb{R}^2 \to \mathbb{R}^2$

Let us consider now a system of two equations in four variables: $\begin{cases} f(x_1, x_2, y_1, y_2) = k_1 \\ g(x_1, x_2, y_1, y_2) = k_2 \end{cases}$, $k_1, k_2 \in \mathbb{R}$.

The system allows us, under suitable assumptions, to define a function:
$\mathbb{R}^2 \to \mathbb{R}^2, (x_1, x_2) \to (y_1(x_1, x_2), y_2(x_1, x_2))$.

Implicit Function $\mathbb{R}^2 \to \mathbb{R}^2$ **: First order derivatives**

If $\begin{cases} w_1 = f(x_1, x_2, y_1, y_2) = k_1 \\ w_2 = g(x_1, x_2, y_1, y_2) = k_2 \end{cases}$, and $(\mathbb{X}|\,\mathbb{Y}) = (x_1, x_2, y_1, y_2)$, we have these two function compositions:

$\mathbb{R}^2 \to \mathbb{R}^4 \xrightarrow{f} \mathbb{R},\ \mathbb{X} \to (\mathbb{X}|\,\mathbb{Y}) \xrightarrow{f} w_1 = f(\mathbb{X}|\,\mathbb{Y}) = k_1$, and
$\mathbb{R}^2 \to \mathbb{R}^4 \xrightarrow{g} \mathbb{R},\ \mathbb{X} \to (\mathbb{X}|\,\mathbb{Y}) \xrightarrow{g} w_2 = g(\mathbb{X}|\,\mathbb{Y}) = k_2$

from which, differentiating with respect to x_1 and x_2, w_1 and w_2 being constant, we obtain the system:

$$\begin{cases} \dfrac{\partial w_1}{\partial x_1} = \nabla f(\mathbb{X}|\,\mathbb{Y}) \cdot \dfrac{\partial(\mathbb{X}|\,\mathbb{Y})}{\partial x_1} = 0 \\ \dfrac{\partial w_2}{\partial x_1} = \nabla g(\mathbb{X}|\,\mathbb{Y}) \cdot \dfrac{\partial(\mathbb{X}|\,\mathbb{Y})}{\partial x_1} = 0 \\ \dfrac{\partial w_1}{\partial x_2} = \nabla f(\mathbb{X}|\,\mathbb{Y}) \cdot \dfrac{\partial(\mathbb{X}|\,\mathbb{Y})}{\partial x_2} = 0 \\ \dfrac{\partial w_2}{\partial x_2} = \nabla g(\mathbb{X}|\,\mathbb{Y}) \cdot \dfrac{\partial(\mathbb{X}|\,\mathbb{Y})}{\partial x_2} = 0 \end{cases}.$$

As $\dfrac{\partial(\mathbb{X}|\,\mathbb{Y})}{\partial x_1} = \left(1, 0, \dfrac{\partial y_1}{\partial x_1}, \dfrac{\partial y_2}{\partial x_1}\right)$ and $\dfrac{\partial(\mathbb{X}|\,\mathbb{Y})}{\partial x_2} = \left(0, 1, \dfrac{\partial y_1}{\partial x_2}, \dfrac{\partial y_2}{\partial x_2}\right)$ we get the system in the unknowns $\dfrac{\partial y_1}{\partial x_1}$, $\dfrac{\partial y_2}{\partial x_1}$, $\dfrac{\partial y_1}{\partial x_2}$ and $\dfrac{\partial y_2}{\partial x_2}$:

$$\begin{cases} \dfrac{\partial f}{\partial y_1} \cdot \dfrac{\partial y_1}{\partial x_1} + \dfrac{\partial f}{\partial y_2} \cdot \dfrac{\partial y_2}{\partial x_1} = -\dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial g}{\partial y_1} \cdot \dfrac{\partial y_1}{\partial x_1} + \dfrac{\partial g}{\partial y_2} \cdot \dfrac{\partial y_2}{\partial x_1} = -\dfrac{\partial g}{\partial x_1} \\ \dfrac{\partial f}{\partial y_1} \cdot \dfrac{\partial y_1}{\partial x_2} + \dfrac{\partial f}{\partial y_2} \cdot \dfrac{\partial y_2}{\partial x_2} = -\dfrac{\partial f}{\partial x_2} \\ \dfrac{\partial g}{\partial y_1} \cdot \dfrac{\partial y_1}{\partial x_2} + \dfrac{\partial g}{\partial y_2} \cdot \dfrac{\partial y_2}{\partial x_2} = -\dfrac{\partial g}{\partial x_2} \end{cases}$$, which in matrix form becomes:

$$\left\|\begin{matrix} \frac{\partial f}{\partial y_1} & \frac{\partial f}{\partial y_2} & 0 & 0 \\ \frac{\partial g}{\partial y_1} & \frac{\partial g}{\partial y_2} & 0 & 0 \\ 0 & 0 & \frac{\partial f}{\partial y_1} & \frac{\partial f}{\partial y_2} \\ 0 & 0 & \frac{\partial g}{\partial y_1} & \frac{\partial g}{\partial y_2} \end{matrix}\right\| \cdot \left\|\begin{matrix} \frac{\partial y_1}{\partial x_1} \\ \frac{\partial y_2}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_2} \end{matrix}\right\| = - \left\|\begin{matrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial g}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial g}{\partial x_2} \end{matrix}\right\|$$, which can be written as:

$$\left\|\begin{matrix} \frac{\partial f}{\partial y_1} & \frac{\partial f}{\partial y_2} \\ \frac{\partial g}{\partial y_1} & \frac{\partial g}{\partial y_2} \end{matrix}\right\| \cdot \left\|\begin{matrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{matrix}\right\| = - \left\|\begin{matrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \\ \frac{\partial g}{\partial x_1} & \frac{\partial g}{\partial x_2} \end{matrix}\right\|$$ and, using Jacobian matrices, as:

$$\frac{\partial(f,g)}{\partial(y_1,y_2)} \cdot \frac{\partial(y_1,y_2)}{\partial(x_1,x_2)} = -\frac{\partial(f,g)}{\partial(x_1,x_2)} .$$

If the point $(x_1^0, x_2^0, y_1^0, y_2^0)$ satisfies the system, if f and g are differentiable with continuous derivatives, and if $\left|\frac{\partial(f,g)(x_1^0,x_2^0,y_1^0,y_2^0)}{\partial(y_1,y_2)}\right| \neq 0$, i.e. $\frac{\partial(f,g)}{\partial(y_1,y_2)}$ has maximum rank, Dini's Theorem, in this case, assures the existence, in a neighborhood of the point (x_1^0, x_2^0) of the implicit function $\mathbb{R}^2 \to \mathbb{R}^2, (x_1,x_2) \to (y_1(x_1,x_2), y_2(x_1,x_2))$, whose derivatives are obtained as $\frac{\partial(y_1,y_2)}{\partial(x_1,x_2)} = -\left\|\frac{\partial(f,g)}{\partial(y_1,y_2)}\right\|^{-1} \cdot \frac{\partial(f,g)}{\partial(x_1,x_2)}$, which is the general formula from which we obtain the derivatives of y_1 and y_2 with respect to x_1 and x_2.

There is also another method of calculus, arising from Cramer's theorem on linear systems, from which we obtain:

$$\frac{\partial y_1}{\partial x_1} = -\frac{\begin{vmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial y_2} \\ \frac{\partial g}{\partial x_1} & \frac{\partial g}{\partial y_2} \end{vmatrix}}{\begin{vmatrix} \frac{\partial f}{\partial y_1} & \frac{\partial f}{\partial y_2} \\ \frac{\partial g}{\partial y_1} & \frac{\partial g}{\partial y_2} \end{vmatrix}} = -\frac{\left|\frac{\partial(f,g)}{\partial(x_1,y_2)}\right|}{\left|\frac{\partial(f,g)}{\partial(y_1,y_2)}\right|} ; \quad \frac{\partial y_1}{\partial x_2} = -\frac{\begin{vmatrix} \frac{\partial f}{\partial x_2} & \frac{\partial f}{\partial y_2} \\ \frac{\partial g}{\partial x_2} & \frac{\partial g}{\partial y_2} \end{vmatrix}}{\begin{vmatrix} \frac{\partial f}{\partial y_1} & \frac{\partial f}{\partial y_2} \\ \frac{\partial g}{\partial y_1} & \frac{\partial g}{\partial y_2} \end{vmatrix}} = -\frac{\left|\frac{\partial(f,g)}{\partial(x_2,y_2)}\right|}{\left|\frac{\partial(f,g)}{\partial(y_1,y_2)}\right|}$$

$$\frac{\partial y_2}{\partial x_1} = -\frac{\begin{vmatrix} \frac{\partial f}{\partial y_1} & \frac{\partial f}{\partial x_1} \\ \frac{\partial g}{\partial y_1} & \frac{\partial g}{\partial x_1} \end{vmatrix}}{\begin{vmatrix} \frac{\partial f}{\partial y_1} & \frac{\partial f}{\partial y_2} \\ \frac{\partial g}{\partial y_1} & \frac{\partial g}{\partial y_2} \end{vmatrix}} = -\frac{\left|\frac{\partial(f,g)}{\partial(y_1,x_1)}\right|}{\left|\frac{\partial(f,g)}{\partial(y_1,y_2)}\right|} ; \quad \frac{\partial y_2}{\partial x_2} = -\frac{\begin{vmatrix} \frac{\partial f}{\partial y_1} & \frac{\partial f}{\partial x_2} \\ \frac{\partial g}{\partial y_1} & \frac{\partial g}{\partial x_2} \end{vmatrix}}{\begin{vmatrix} \frac{\partial f}{\partial y_1} & \frac{\partial f}{\partial y_2} \\ \frac{\partial g}{\partial y_1} & \frac{\partial g}{\partial y_2} \end{vmatrix}} = -\frac{\left|\frac{\partial(f,g)}{\partial(y_1,x_2)}\right|}{\left|\frac{\partial(f,g)}{\partial(y_1,y_2)}\right|}$$

Example 52 : $P_0 = (1, 1, -1, -1)$ satisfies the system:

$$\begin{cases} f(x_1,x_2,y_1,y_2) = (x_1 + x_2)\, e^{y_2 - y_1} + (y_2 + y_1)\, e^{x_1 - x_2} = 0 \\ g(x_1,x_2,y_1,y_2) = (2x_2 - y_1)\, e^{x_1 + y_2} - (x_1 - y_2)\, e^{x_2 + y_1} = 1 \end{cases} .$$

Functions f and g are everywhere differentiable with continuous derivatives. Then:

$$\frac{\partial f}{\partial x_1} = e^{y_2 - y_1} + (y_2 + y_1)\, e^{x_1 - x_2} ; \qquad \frac{\partial g}{\partial x_1} = (2x_2 - y_1)\, e^{x_1 + y_2} - e^{x_2 + y_1} ;$$

$$\frac{\partial f}{\partial x_2} = e^{y_2 - y_1} - (y_2 + y_1)\, e^{x_1 - x_2}\,; \qquad \frac{\partial g}{\partial x_2} = 2\, e^{x_1 + y_2} - (x_1 - y_2)\, e^{x_2 + y_1}\,;$$
$$\frac{\partial f}{\partial y_1} = -(x_1 + x_2)\, e^{y_2 - y_1} + e^{x_1 - x_2}\,; \qquad \frac{\partial g}{\partial y_1} = -e^{x_1 + y_2} - (x_1 - y_2)\, e^{x_2 + y_1}\,;$$
$$\frac{\partial f}{\partial y_2} = (x_1 + x_2)\, e^{y_2 - y_1} + e^{x_1 - x_2}\,; \qquad \frac{\partial g}{\partial y_2} = (2x_2 - y_1)\, e^{x_1 + y_2} + e^{x_2 + y_1}\,.$$

We have $\dfrac{\partial(f,g)(1,1,-1,-1)}{\partial(x_1,x_2,y_1,y_2)} = \begin{Vmatrix} -1 & 3 & -1 & 3 \\ 2 & 0 & -3 & 4 \end{Vmatrix}$, and since $\begin{vmatrix} -1 & 3 \\ -3 & 4 \end{vmatrix} = 5 \neq 0$, with such system an implicit function can be defined:
$\mathbb{R}^2 \to \mathbb{R}^2, (x_1,x_2) \to (y_1(x_1,x_2), y_2(x_1,x_2))$, in a neighborhood of $(x_1,x_2) = (1,1)$.
For derivatives of this function, calculated in $(1,1)$, from:
$\dfrac{\partial(y_1,y_2)}{\partial(x_1,x_2)} = -\left\Vert \dfrac{\partial(f,g)}{\partial(y_1,y_2)} \right\Vert^{-1} \cdot \dfrac{\partial(f,g)}{\partial(x_1,x_2)}$ we get:

$$\begin{Vmatrix} \dfrac{\partial y_1(1,1)}{\partial x_1} & \dfrac{\partial y_1(1,1)}{\partial x_2} \\ \dfrac{\partial y_2(1,1)}{\partial x_1} & \dfrac{\partial y_2(1,1)}{\partial x_2} \end{Vmatrix} = -\begin{Vmatrix} -1 & 3 \\ -3 & 4 \end{Vmatrix}^{-1} \cdot \begin{Vmatrix} -1 & 3 \\ 2 & 0 \end{Vmatrix} =$$

$$= -\begin{Vmatrix} \frac{4}{5} & -\frac{3}{5} \\ \frac{3}{5} & -\frac{1}{5} \end{Vmatrix} \cdot \begin{Vmatrix} -1 & 3 \\ 2 & 0 \end{Vmatrix}, \text{ and so: } \begin{Vmatrix} \dfrac{\partial y_1(1,1)}{\partial x_1} & \dfrac{\partial y_1(1,1)}{\partial x_2} \\ \dfrac{\partial y_2(1,1)}{\partial x_1} & \dfrac{\partial y_2(1,1)}{\partial x_2} \end{Vmatrix} = \begin{Vmatrix} 2 & -\frac{12}{5} \\ 1 & -\frac{9}{5} \end{Vmatrix}.$$

DINI'S THEOREM IN THE GENERAL CASE

Systems of m equations in $n+m$ variables: Implicit Function $\mathbb{R}^n \to \mathbb{R}^m$

We conclude treating the general case, namely that of a system of m equations in $m+n$ variables. Stating Dini's theorem in this general case, we will find again, as particular cases, all cases previously treated. The system is given:

$$\begin{cases} f_1(x_1,x_2,...,x_n,y_1,y_2,...,y_m) = k_1 \\ f_2(x_1,x_2,...,x_n,y_1,y_2,...,y_m) = k_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ f_m(x_1,x_2,...,x_n,y_1,y_2,...,y_m) = k_m \end{cases}.$$

Such a system allows us, with the appropriate assumptions, to define an implicit function:
$\mathbb{R}^n \to \mathbb{R}^m, (x_1,x_2,...,x_n) \to (y_1,y_2,...,y_m)$, with $y_i = y_i(x_1,x_2,...,x_n)$.

We see that the number of equations corresponds to the number of dependent variables, and therefore, simply by difference, we get the number of the independent variables: m equations imply m dependent variables, and so $m+n-m=n$ is the number of the independent variables.

Theorem 21 : Let $\mathrm{P}_0 = \left(x_1^0, x_2^0, ..., x_n^0, y_1^0, y_2^0, ..., y_m^0\right)$ be a point satisfying the system:

$$\begin{cases} f_1(x_1,x_2,...,x_n,y_1,y_2,...,y_m) = k_1 \\ f_2(x_1,x_2,...,x_n,y_1,y_2,...,y_m) = k_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ f_m(x_1,x_2,...,x_n,y_1,y_2,...,y_m) = k_m \end{cases}.$$

Let f_i be differentiable functions with continuous derivatives in a neighborhood of P_0, and let also be $\left|\dfrac{\partial(f_1,f_2,..,f_m)(\mathrm{P}_0)}{\partial(y_1,y_2,...,y_m)}\right| \neq 0$.

Then the system defines, in a neighborhood of $(x_1^0, x_2^0, ..., x_n^0)$, an implicit function, continuous with continuous derivatives:

$\mathbb{R}^n \to \mathbb{R}^m, (x_1, x_2, ..., x_n) \to (y_1, y_2, ..., y_m)$, with $y_i = y_i(x_1, x_2, ..., x_n)$, whose Jacobian matrix is given by $\dfrac{\partial(y_1, y_2, ..., y_m)}{\partial(x_1, x_2, ..., x_n)} = -\left\|\dfrac{\partial(f_1, f_2, .., f_m)}{\partial(y_1, y_2, ..., y_m)}\right\|^{-1} \cdot \dfrac{\partial(f_1, f_2, .., f_m)}{\partial(x_1, x_2, ..., x_n)}$.

Each derivative $\dfrac{\partial y_i}{\partial x_k}$ can also be expressed, using Cramer's theorem, as a quotient:

$$\frac{\partial y_i}{\partial x_k} = -\frac{\begin{vmatrix} \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_{i-1}} & \frac{\partial f_1}{\partial x_k} & \frac{\partial f_1}{\partial y_{i+1}} & \cdots & \frac{\partial f_1}{\partial y_m} \\ \frac{\partial f_2}{\partial y_1} & \cdots & \frac{\partial f_2}{\partial y_{i-1}} & \frac{\partial f_2}{\partial x_k} & \frac{\partial f_2}{\partial y_{i+1}} & \cdots & \frac{\partial f_2}{\partial y_m} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_m}{\partial y_1} & \cdots & \frac{\partial f_m}{\partial y_{i-1}} & \frac{\partial f_m}{\partial x_k} & \frac{\partial f_m}{\partial y_{i+1}} & \cdots & \frac{\partial f_m}{\partial y_m} \end{vmatrix}}{\begin{vmatrix} \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_{i-1}} & \frac{\partial f_1}{\partial y_i} & \frac{\partial f_1}{\partial y_{i+1}} & \cdots & \frac{\partial f_1}{\partial y_m} \\ \frac{\partial f_2}{\partial y_1} & \cdots & \frac{\partial f_2}{\partial y_{i-1}} & \frac{\partial f_2}{\partial y_i} & \frac{\partial f_2}{\partial y_{i+1}} & \cdots & \frac{\partial f_2}{\partial y_m} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_m}{\partial y_1} & \cdots & \frac{\partial f_m}{\partial y_{i-1}} & \frac{\partial f_m}{\partial y_i} & \frac{\partial f_m}{\partial y_{i+1}} & \cdots & \frac{\partial f_m}{\partial y_m} \end{vmatrix}} =$$

$$\frac{\partial y_i}{\partial x_k} = -\frac{\left|\dfrac{\partial(f_1, f_2, .., f_m)}{\partial(y_1, ..., y_{i-1}, x_k, y_{i+1}, ..., y_m)}\right|}{\left|\dfrac{\partial(f_1, f_2, .., f_m)}{\partial(y_1, ..., y_{i-1}, y_i, y_{i+1}, ..., y_m)}\right|}.$$

That is, the derivative of y_i with respect to x_k is given by the opposite of a quotient between two Jacobian determinants: the denominator is the Jacobian of the equations with respect to the dependent variables, the numerator is the Jacobian obtained replacing, in the previous one, the i-th column, that of the variable we want to derive, with the derivatives of the equations made with respect to x_k, the variable respect to which we want to derive y_i.

Implicit Function $\mathbb{R}^n \to \mathbb{R}^m$: First order derivatives

Let's see how we can justify this result concerning the derivatives of the implicit function. From the composition of functions:

$(x_1, x_2, ..., x_n) \to (x_1, x_2, ..., x_n, y_1, y_2, ..., y_m) \xrightarrow{f_i} w_i = k_i = \text{cost.}$, $1 \le i \le m$ we get:

$$\frac{\partial(w_1, w_2, ..., w_m)}{\partial(x_1, x_2, ..., x_n)} = \frac{\partial(w_1, w_2, ..., w_m)}{\partial(x_1, x_2, ..., x_n, y_1, y_2, ..., y_m)} \cdot \frac{\partial(x_1, x_2, ..., x_n, y_1, y_2, ..., y_m)}{\partial(x_1, x_2, ..., x_n)} = \mathbb{O}.$$

The second term can be written as:

$$\left\| \begin{array}{c|c} \dfrac{\partial(f_1, f_2, .., f_m)}{\partial(x_1, x_2, ..., x_n)} & \dfrac{\partial(f_1, f_2, .., f_m)}{\partial(y_1, y_2, ..., y_m)} \end{array} \right\| \cdot \left\| \begin{array}{c} \dfrac{\partial(x_1, x_2, ..., x_n)}{\partial(x_1, x_2, ..., x_n)} \\ -\ -\ -\ -\ - \\ \dfrac{\partial(y_1, y_2, ..., y_m)}{\partial(x_1, x_2, ..., x_n)} \end{array} \right\| = \mathbb{O}.$$

The matrix on the left is a $(m, n+m)$ matrix, divided into two blocks, the first (m, n) and the second (m, m); the right matrix is a $(n+m, n)$ matrix, divided into two blocks, the upper (n, n) and the lower (m, n). Working by block, from the equality we obtain:

$\dfrac{\partial(f_1, f_2, .., f_m)}{\partial(x_1, x_2, ..., x_n)} \cdot \mathbb{I}_n + \dfrac{\partial(f_1, f_2, .., f_m)}{\partial(y_1, y_2, ..., y_m)} \cdot \dfrac{\partial(y_1, y_2, ..., y_m)}{\partial(x_1, x_2, ..., x_n)} = \mathbb{O}$ from which:

$$\frac{\partial(f_1,f_2,..,f_m)}{\partial(y_1,y_2,...,y_m)}\cdot\frac{\partial(y_1,y_2,...,y_m)}{\partial(x_1,x_2,...,x_n)} = -\frac{\partial(f_1,f_2,..,f_m)}{\partial(x_1,x_2,...,x_n)}$$ and so:

$$\frac{\partial(y_1,y_2,...,y_m)}{\partial(x_1,x_2,...,x_n)} = -\left\|\frac{\partial(f_1,f_2,..,f_m)}{\partial(y_1,y_2,...,y_m)}\right\|^{-1}\cdot\frac{\partial(f_1,f_2,..,f_m)}{\partial(x_1,x_2,...,x_n)}.$$

In compact form the above equalities can be expressed as:

$$\frac{\partial(\mathbb{W})}{\partial(\mathbb{X})} = \frac{\partial(\mathbb{W})}{\partial(\mathbb{X}\,|\,\mathbb{Y})}\cdot\frac{\partial(\mathbb{X}\,|\,\mathbb{Y})}{\partial(\mathbb{X})} = \frac{\partial(\mathbb{W})}{\partial(\mathbb{X})}\cdot\mathbb{I}_n + \frac{\partial(\mathbb{W})}{\partial(\mathbb{Y})}\cdot\frac{\partial(\mathbb{Y})}{\partial(\mathbb{X})} = \mathbb{O}$$, or:

$$\frac{\partial(\mathbb{Y})}{\partial(\mathbb{X})} = -\left(\frac{\partial(\mathbb{W})}{\partial(\mathbb{Y})}\right)^{-1}\cdot\frac{\partial(\mathbb{W})}{\partial(\mathbb{X})}.$$

MAXIMA AND MINIMA FOR FUNCTIONS $f:\mathbb{R}^n\to\mathbb{R}$

Let us now study the problem of finding maximum and minimum points (also known as extreme points) for functions of several variables.
The definition of relative (local) maximum or minimum point for functions $f:\mathbb{R}^n\to\mathbb{R}$ is similar to that given for functions of one variable, and it has been stated in Definition 19.
As regards the existence of maximum and minimum absolute points Weierstrass's theorem 6 is valid.

FIRST ORDER CONDITIONS

For functions of one variable there is a theorem, known as Fermat's Theorem, which states that if a function is differentiable at x_0 and x_0 is a relative maximum or minimum point, then $f'(x_0)=0$.
For functions $f:\mathbb{R}^n\to\mathbb{R}$, similarly, we state the following:

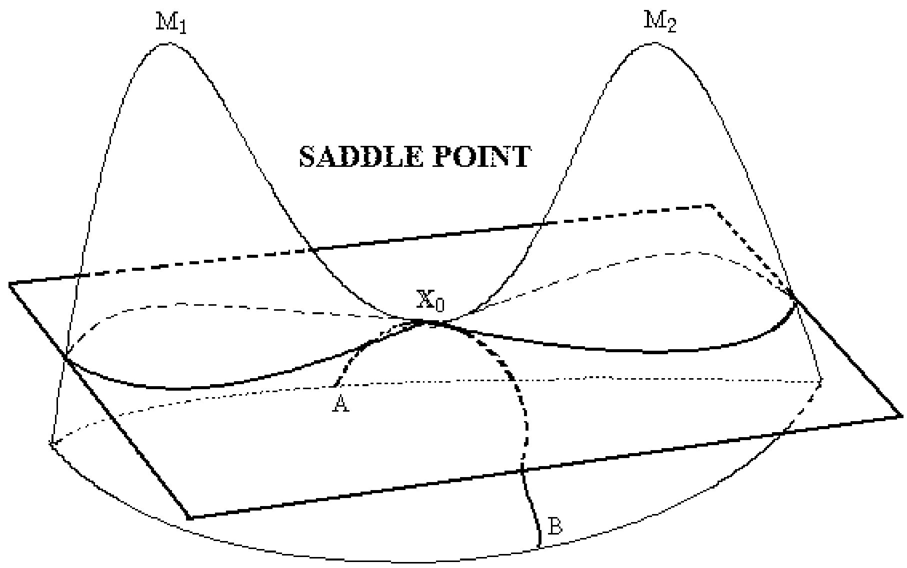
Theorem 22 : If $f(\mathbb{X})$ is differentiable at $\mathbb{X}_0$, interior point of D_f, and $\mathbb{X}_0$ is a relative maximum or minimum point for f, then $\nabla f(\mathbb{X}_0)=\mathbb{O}$, where $\mathbb{O}$ is the null vector.
Proof : If $f(\mathbb{X})=f(\mathbb{X}_0+t\,v)$, v unit vector, we have a composite function:
$\mathbb{R}\to\mathbb{R}^n\xrightarrow{f}\mathbb{R},\ t\to(\mathbb{X}_0+t\,v)\xrightarrow{f}f(\mathbb{X}_0+t\,v)$,
that we see as $g(t)=f(\mathbb{X}_0+t\,v)$, with $g(0)=f(\mathbb{X}_0)$.
If $\mathbb{X}_0$ is a relative maximum (or minimum) point, it will also be:
$f(\mathbb{X}_0+t\,v)\le f(\mathbb{X}_0)\ \ (f(\mathbb{X}_0+t\,v)\ge f(\mathbb{X}_0))$, $\forall\,\mathbb{X}\in\mathfrak{J}(\mathbb{X}_0)$, or:
$g(t)\le g(0)\ (g(t)\ge g(0))$, $\forall\,t\in\mathfrak{J}(0)$.
So if $\mathbb{X}_0$ is a relative maximum (or minimum) point for f, $t=0$ will be the same for g.
Since f is differentiable, and since $\mathbb{X}_0+t\,v$ is differentiable with respect to t, it follows that $g(t)$ is differentiable, as a composition of differentiable functions, and then from Fermat's Theorem it must be $g'(0)=0$. But

$$g'(0)=\lim_{t\to0}\frac{g(t)-g(0)}{t}=\lim_{t\to0}\frac{f(\mathbb{X}_0+t\,v)-f(\mathbb{X}_0)}{t}=\mathcal{D}_vf(\mathbb{X}_0)=0.$$

But $\mathcal{D}_vf(\mathbb{X}_0)=\nabla f(\mathbb{X}_0)\cdot v$, since f is differentiable, and then:
$\forall\,v:\ \nabla f(\mathbb{X}_0)\cdot v=0\Leftrightarrow\nabla f(\mathbb{X}_0)=\mathbb{O}$.•

The above condition $\nabla f(\mathbb{X}_0)=\mathbb{O}$ is necessary but not sufficient for a relative maximum or minimum point, and is a necessary condition only if the function is differentiable. These are called the first-order conditions in the maximum and minimum points research.
Points where $\nabla f(\mathbb{X}_0)=\mathbb{O}$ are called stationary points, and are points at which the tangent plane (or hyperplane) to the graph of the function is horizontal (i.e. parallel to the plane (or hyperplane) of the independent variables).

To search for relative maximum and minimum points we must therefore satisfy the first-order conditions, i.e. we must impose $\nabla f(\mathbb{X}) = \mathbb{O}$, and solve the system of n equations in n unknowns that arises from this. After all the solutions have been found, i.e. all points $\mathbb{X}_0$ for which $\nabla f(\mathbb{X}_0) = \mathbb{O}$, we must check whether those points are really relative maximum or minimum points, or if they are saddle points.



Saddle points are points where $\nabla f(\mathbb{X}_0) = \mathbb{O}$, but neither the definition of maximum point nor the definition of minimum point are satisfied , as in every neighborhood of $\mathbb{X}_0$ there are points where $f(\mathbb{X}) > f(\mathbb{X}_0)$ and points where $f(\mathbb{X}) < f(\mathbb{X}_0)$. Geometrically speaking, in a stationary point of maximum (or minimum) the tangent plan (or hyperplane) is, in a neighborhood of $\mathbb{X}_0$, all above (all below) the graph of the function. If $\mathbb{X}_0$ is, instead, a saddle point, the tangent plane crosses the graph of the function, and then, in every neighborhood of $\mathbb{X}_0$, there are points of the graph above and points of the graph below the tangent plane.
Therefore we need criteria and methodologies, called second-order conditions, whose purpose is to establish the true nature of a stationary point. Meanwhile, let us see now with an example how the analysis along particular directions, as seen for the limit operation, is not generally valid to draw positive conclusions about the true nature of a maximum or minimum point.

Example 53 : Given $f(x,y) = \left(y - x^2\right)\left(y - x^4\right)$, first of all we search for its stationary points; we must put:

$$\begin{cases} f'_x = -2x\left(y - x^4\right) - 4x^3\left(y - x^2\right) = 6x^5 - 2xy - 4x^3y = 0 \\ f'_y = y - x^4 + y - x^2 = 2y - x^4 - x^2 = 0 \quad \text{or } 2y = x^4 + x^2 \end{cases};$$

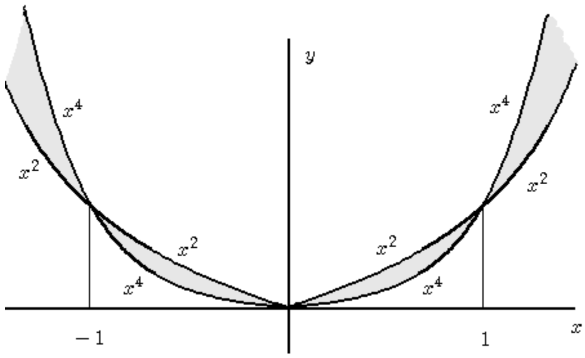
substituting in the first equation we have:

$2x^7 - 3x^5 + x^3 = x^3\left(2x^4 - 3x^2 + 1\right) = 0$, from which we obtain $x = 0$ and:

$x^2 = \dfrac{3 \pm \sqrt{9-8}}{4}$ and so: $\begin{cases} x^2 = 1 & \text{or } x = \pm 1 \\ x^2 = \frac{1}{2} & \text{or } x = \pm \frac{1}{\sqrt{2}} \end{cases}$. We have found five stationary points:

$$\begin{cases} x = 0 \\ y = 0 \end{cases};\ \begin{cases} x = 1 \\ y = 1 \end{cases};\ \begin{cases} x = -1 \\ y = 1 \end{cases};\ \begin{cases} x = \frac{1}{\sqrt{2}} \\ y = \frac{3}{8} \end{cases};\ \begin{cases} x = -\frac{1}{\sqrt{2}} \\ y = \frac{3}{8} \end{cases}.$$

In this example let us study only the point $(0,0)$. Is $f(0,0)=0$. Meanwhile, let us study the behavior of $f(x,y)$ along any line passing through the origin. Given $y=mx$, we get:
$f(x,mx)=\left(mx-x^2\right)\left(mx-x^4\right)=x^2\left(x^4-mx^3-mx+m^2\right)$.
Studying the sign of $f(x,mx)$, as $x^2>0, \forall\, x\neq 0$, for the "permanence of the sign theorem", since $x^4-mx^3-mx+m^2$ is $m^2>0$ at $x=0$, and since the function is continuous, $f(x,mx)$ is strictly positive in a neighborhood of the point $x=0$.
So the point $(0,0)$ is a minimum along each line passing through the origin; on the line $x=0$ we have $f(0,y)=y^2$, and then also on this line the point $(0,0)$ is a minimum point. Although the analysis of all lines passing through the origin give the same answer, $(0,0)$ is a saddle point. In fact, if we study the sign of $f(x,y)$ we have:
$f(x,y)>0$ if $\begin{cases} y>x^2 \\ y>x^4 \end{cases}$ or if $\begin{cases} y<x^2 \\ y<x^4 \end{cases}$.



In the figure the black area represents the points where $f(x,y)<0$. In every neighborhood of $(0,0)$ there are points where $f(x,y)>0$ and points where $f(x,y)<0$. As $f(0,0)=0$, it follows that $(0,0)$ is a saddle point, contrary to what could be deduced analyzing the function along all the lines passing through the origin.
We will resume and complete the analysis of stationary points of this function when we have the right tools.

SECOND ORDER CONDITIONS
Since for functions $f:\mathbb{R}^n\to\mathbb{R}$ we can not define an increasing or decreasing function, it therefore is of no use studying the sign of the first-order partial derivatives. To distinguish between maximum or minimum points and saddle points, if any, we must instead use the second-order conditions, which are sufficient conditions, and are connected to concavity and convexity of f at $\mathbb{X}_0$. Indeed the following is valid:

Theorem 23 : If $\mathbb{X}_0$ is a stationary point for f and if the function is differentiable and concave in a neighborhood of $\mathbb{X}_0$, then $\mathbb{X}_0$ is a relative (local) maximum point.
Proof : From Theorem 14, since f is concave:
$f(\mathbb{X})\leq f(\mathbb{X}_0)+\nabla f(\mathbb{X}_0)\,(\mathbb{X}-\mathbb{X}_0)$, as the graph of the function lies below the tangent (hyper) plane at $\mathbb{X}_0$. But $\nabla f(\mathbb{X}_0)=\mathbb{O}$, from which $f(\mathbb{X})\leq f(\mathbb{X}_0)$ in $\mathfrak{I}(\mathbb{X}_0)$, and so $\mathbb{X}_0$ is a relative (local) maximum point.●

Similarly, if $\mathbb{X}_0$ is a stationary point for f and if the function is differentiable and convex in a neighborhood of $\mathbb{X}_0$, then $\mathbb{X}_0$ is a relative (local) minimum point.

Assuming now that the function is twice differentiable at $\mathbb{X}_0$, we have, using Taylor's polynomial, if $\mathrm{d}\mathbb{X}=\mathbb{X}-\mathbb{X}_0$:

$f(\mathbb{X}) = f(\mathbb{X}_0) + \nabla f(\mathbb{X}_0) \cdot \mathrm{d}\mathbb{X} + \frac{1}{2}\,\mathrm{d}\mathbb{X} \cdot \mathbb{H}(\mathbb{X}_0) \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}} + o\big(\|\mathrm{d}\mathbb{X}\|^2\big)$,
and being $\mathbb{X}_0$ a stationary point ($\nabla f(\mathbb{X}_0) = \mathbb{O}$), we get:
$f(\mathbb{X}) - f(\mathbb{X}_0) = \frac{1}{2}\,\mathrm{d}\mathbb{X} \cdot \mathbb{H}(\mathbb{X}_0) \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}} + o\big(\|\mathrm{d}\mathbb{X}\|^2\big) = \frac{1}{2}\,\mathrm{d}^2 f(\mathbb{X}_0) + o\big(\|\mathrm{d}\mathbb{X}\|^2\big)$.
So the sign of $f(\mathbb{X}) - f(\mathbb{X}_0)$ is the same as that of $\mathrm{d}\mathbb{X} \cdot \mathbb{H}(\mathbb{X}_0) \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}} = \mathrm{d}^2 f(\mathbb{X}_0)$.
If $\mathrm{d}^2 f(\mathbb{X}_0) < 0$ f is concave at $\mathbb{X}_0$, whereas if $\mathrm{d}^2 f(\mathbb{X}_0) > 0$ f is convex at $\mathbb{X}_0$.

To check if the definition of maximum or minimum point is satisfied, we study, in a neighborhood of $\mathbb{X}_0$, the sign of the difference $f(\mathbb{X}) - f(\mathbb{X}_0)$, as:
.) $f(\mathbb{X}) < f(\mathbb{X}_0) \Leftrightarrow f(\mathbb{X}) - f(\mathbb{X}_0) < 0 \Leftrightarrow \mathrm{d}^2 f(\mathbb{X}_0) < 0$ and so $\mathbb{X}_0$ is a maximum point;
.) $f(\mathbb{X}) > f(\mathbb{X}_0) \Leftrightarrow f(\mathbb{X}) - f(\mathbb{X}_0) > 0 \Leftrightarrow \mathrm{d}^2 f(\mathbb{X}_0) > 0$ and so $\mathbb{X}_0$ is a minimum point.

If the sign of the difference $f(\mathbb{X}) - f(\mathbb{X}_0)$, i.e. the sign of $\mathrm{d}^2 f(\mathbb{X}_0)$, is not constant in a neighborhood of $\mathbb{X}_0$, then $\mathbb{X}_0$ is certainly a saddle point.
The second-order total differential $\mathrm{d}^2 f(\mathbb{X}_0) = \sum_{i,j=1}^{n} \frac{\partial^2 f(\mathbb{X}_0)}{\partial x_i\, \partial x_j}\,\mathrm{d}x_i\,\mathrm{d}x_j$ is a quadratic form, i.e. a polynomial in the n variables $\mathrm{d}x_i$, with all terms of the second degree, so now we need sufficient criteria to establish the sign of a quadratic form..

QUADRATIC FORMS
Quadratic forms are homogeneous quadratic polynomials in n variables like:
$$Q(\mathbb{X}) = Q(x_1, x_2, ..., x_n) = \sum_{i,j=1}^{n} a_{i\,j}\, x_i\, x_j\,,$$
i.e. polynomials in n variables having solely second-degree terms.

Let's see how every quadratic form can be written in the form $Q(\mathbb{X}) = \mathbb{X} \cdot \mathbb{A} \cdot \mathbb{X}^{\mathrm{T}}$, where $\mathbb{X} \in \mathbb{R}^n$ and $\mathbb{A}$ is a square matrix of order n.

Example 54 : $Q(x_1, x_2) = \|x_1\ x_2\| \cdot \begin{Vmatrix} 1 & 2 \\ 4 & 5 \end{Vmatrix} \cdot \begin{Vmatrix} x_1 \\ x_2 \end{Vmatrix} = 1\,x_1^2 + 2x_1\,x_2 + 4x_1\,x_2 + 5x_2^2 =$
$= x_1^2 + 6x_1\,x_2 + 5x_2^2$.

Example 55 : $Q(x_1, x_2, x_3) = \|x_1\ x_2\ x_3\| \cdot \begin{Vmatrix} 1 & 2 & 3 \\ 0 & 3 & 1 \\ 1 & 3 & 4 \end{Vmatrix} \cdot \begin{Vmatrix} x_1 \\ x_2 \\ x_3 \end{Vmatrix} =$
$= x_1^2 + 3x_2^2 + 4x_3^2 + 2x_1\,x_2 + 4x_1\,x_3 + 4x_2\,x_3$.

But at once we see that every quadratic form can be generated by a symmetric matrix: starting from matrix $\mathbb{A}$, the symmetric matrix $\mathbb{B}$ is built placing $b_{i\,j} = \dfrac{a_{i\,j} + a_{j\,i}}{2}$.

Example 56 : Using the examples above, we can easily verify that:
$Q(x_1, x_2) = \|x_1\ x_2\| \cdot \begin{Vmatrix} 1 & 2 \\ 4 & 5 \end{Vmatrix} \cdot \begin{Vmatrix} x_1 \\ x_2 \end{Vmatrix} = \|x_1\ x_2\| \cdot \begin{Vmatrix} 1 & 3 \\ 3 & 5 \end{Vmatrix} \cdot \begin{Vmatrix} x_1 \\ x_2 \end{Vmatrix}$ and that:
$Q(x_1, x_2, x_3) = \|x_1\ x_2\ x_3\| \cdot \begin{Vmatrix} 1 & 2 & 3 \\ 0 & 3 & 1 \\ 1 & 3 & 4 \end{Vmatrix} \cdot \begin{Vmatrix} x_1 \\ x_2 \\ x_3 \end{Vmatrix} = \|x_1\ x_2\ x_3\| \cdot \begin{Vmatrix} 1 & 1 & 2 \\ 1 & 3 & 2 \\ 2 & 2 & 4 \end{Vmatrix} \cdot \begin{Vmatrix} x_1 \\ x_2 \\ x_3 \end{Vmatrix}$.

Then we will consider every quadratic form as generated by a symmetric matrix, and this symmetry is guaranteed when we have the quadratic form generated by the second-order total differential of a twice differentiable function, as the matrix of this quadratic form is the Hessian matrix $\mathbb{H}(\mathbb{X}_0)$ of f. For our purposes, we will write a quadratic form always in the form $Q(\mathrm{d}\mathbb{X}) = \mathrm{d}\mathbb{X} \cdot \mathbb{H} \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}}$, with $\mathrm{d}\mathbb{X} = (\mathrm{d}x_1, \mathrm{d}x_2, .., \mathrm{d}x_n)$, so as to represent it in the form of a second-order total differential.

The study of the sign of quadratic forms is based on the following definitions:
Definition 40 : The quadratic form $Q(\mathrm{d}\mathbb{X}) = \mathrm{d}\mathbb{X} \cdot \mathbb{H} \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}}$ is called:
positive definite if $Q(\mathrm{d}\mathbb{X}) > 0, \forall\, \mathrm{d}\mathbb{X} \neq \mathbb{O}$;
negative definite if $Q(\mathrm{d}\mathbb{X}) < 0, \forall\, \mathrm{d}\mathbb{X} \neq \mathbb{O}$.
Definition 41 : The quadratic form $Q(\mathrm{d}\mathbb{X}) = \mathrm{d}\mathbb{X} \cdot \mathbb{H} \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}}$ is called:
positive semidefinite if $(Q(\mathrm{d}\mathbb{X}) \geq 0, \forall\, \mathrm{d}\mathbb{X} \neq \mathbb{O})$ **and** $(\exists\, \mathrm{d}\mathbb{X} \neq \mathbb{O} : Q(\mathrm{d}\mathbb{X}) = 0)$;
negative semidefinite if $(Q(\mathrm{d}\mathbb{X}) \leq 0, \forall\, \mathrm{d}\mathbb{X} \neq \mathbb{O})$ **and** $(\exists\, \mathrm{d}\mathbb{X} \neq \mathbb{O} : Q(\mathrm{d}\mathbb{X}) = 0)$.
Definition 42 : The quadratic form $Q(\mathrm{d}\mathbb{X}) = \mathrm{d}\mathbb{X} \cdot \mathbb{H} \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}}$ is called:
indefinite if $(\exists\, \mathrm{d}\mathbb{X}_1 : Q(\mathrm{d}\mathbb{X}_1) > 0)$ **and** $(\exists\, \mathrm{d}\mathbb{X}_2 : Q(\mathrm{d}\mathbb{X}_2) < 0)$.

Example 57 : $Q(x_1, x_2) = \|x_1\ x_2\| \cdot \left\|\begin{matrix} 1 & 0 \\ 0 & 3 \end{matrix}\right\| \cdot \left\|\begin{matrix} x_1 \\ x_2 \end{matrix}\right\| = x_1^2 + 3x_2^2$ is a positive definite form as $x_1^2 + 3x_2^2 \geq 0\, \forall\, (x_1, x_2)$ while $x_1^2 + 3x_2^2 = 0$ if and only if $x_1 = x_2 = 0$.

Example 58 : $Q(x_1, x_2, x_3) = \|x_1\ x_2\ x_3\| \cdot \left\|\begin{matrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{matrix}\right\| \cdot \left\|\begin{matrix} x_1 \\ x_2 \\ x_3 \end{matrix}\right\| = x_1^2 + 3x_2^2$ is a positive semidefinite form as $x_1^2 + 3x_2^2 \geq 0\, \forall\, (x_1, x_2, x_3)$ while $x_1^2 + 3x_2^2 = 0$ if $x_1 = x_2 = 0$, $\forall\, x_3 \neq 0$.

Example 59 : $Q(x_1, x_2) = \|x_1\ x_2\| \cdot \left\|\begin{matrix} 1 & 1 \\ 1 & 1 \end{matrix}\right\| \cdot \left\|\begin{matrix} x_1 \\ x_2 \end{matrix}\right\| = x_1^2 + 2x_1 x_2 + x_2^2 = (x_1 + x_2)^2$ is a positive semidefinite form as $(x_1 + x_2)^2 \geq 0, \forall\, (x_1, x_2)$ while $(x_1 + x_2)^2 = 0$ whenever $x_1 = -x_2$.

Example 60 : $Q(x_1, x_2, x_3) = \|x_1\ x_2\ x_3\| \cdot \left\|\begin{matrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{matrix}\right\| \cdot \left\|\begin{matrix} x_1 \\ x_2 \\ x_3 \end{matrix}\right\| = x_1^2 - x_2^2 + x_3^2$ is an indefinite form as $x_1^2 - x_2^2 + x_3^2 > 0$ if, for example, $x_1 = x_2 = 0, \forall\, x_3 \neq 0$ but it is instead $x_1^2 - x_2^2 + x_3^2 < 0$ if, for example, $x_1 = x_3 = 0, \forall\, x_2 \neq 0$.

Let us now expose sufficient criteria to ensure a definite or semidefinite quadratic form.
Let us study the sign of the second-order total differential for $f : \mathbb{R}^2 \to \mathbb{R}$, and then generalize the process to $f : \mathbb{R}^n \to \mathbb{R}$. Then the result is:
$$f(x, y) - f(x_0, y_0) = \frac{1}{2}\, (\mathrm{d}x, \mathrm{d}y) \cdot \mathbb{H}(x_0, y_0) \cdot (\mathrm{d}x, \mathrm{d}y)^{\mathrm{T}} + o\big(\|(x, y) - (x_0, y_0)\|^2\big),$$
and we must study the sign of:
$$\mathrm{d}^2 f(x_0, y_0) = f''_{xx}(x_0, y_0)\, (\mathrm{d}x)^2 + 2\, f''_{xy}(x_0, y_0)\, \mathrm{d}x\, \mathrm{d}y + f''_{yy}(x_0, y_0)\, (\mathrm{d}y)^2.$$
Writing in short form $\mathrm{d}^2 f = f''_{xx}\, (\mathrm{d}x)^2 + 2\, f''_{xy}\, \mathrm{d}x\, \mathrm{d}y + f''_{yy}\, (\mathrm{d}y)^2$, we get:
$$\mathrm{d}^2 f = f''_{xx}\, (\mathrm{d}x)^2 + 2\, f''_{xy}\, \mathrm{d}x\, \mathrm{d}y + \frac{\left(f''_{xy}\right)^2}{f''_{xx}}\, (\mathrm{d}y)^2 - \frac{\left(f''_{xy}\right)^2}{f''_{xx}}\, (\mathrm{d}y)^2 + f''_{yy}\, (\mathrm{d}y)^2 =$$

$$\mathrm{d}^2 f = f''_{xx}\left((\mathrm{d}x)^2 + 2\,\frac{f''_{xy}}{f''_{xx}}\,\mathrm{d}x\,\mathrm{d}y + \frac{\left(f''_{xy}\right)^2}{\left(f''_{xx}\right)^2}\,(\mathrm{d}y)^2\right) - \frac{\left(f''_{xy}\right)^2}{f''_{xx}}\,(\mathrm{d}y)^2 + f''_{yy}\,(\mathrm{d}y)^2 =$$

$$\mathrm{d}^2 f = f''_{xx}\left(\mathrm{d}x + \frac{f''_{xy}}{f''_{xx}}\,\mathrm{d}y\right)^2 + \frac{f''_{xx}\,f''_{yy} - \left(f''_{xy}\right)^2}{f''_{xx}}\,(\mathrm{d}y)^2 .$$

Similarly we can obtain also a second equality:

$$\mathrm{d}^2 f = f''_{yy}\left(\mathrm{d}y + \frac{f''_{xy}}{f''_{yy}}\,\mathrm{d}x\right)^2 + \frac{f''_{xx}\,f''_{yy} - \left(f''_{xy}\right)^2}{f''_{yy}}\,(\mathrm{d}x)^2 .$$

In any case we have the sum of two terms, each of which is the product of a square (so always positive) for another term, whose sign instead is variable:

f''_{xx} and $\dfrac{f''_{xx}\,f''_{yy} - \left(f''_{xy}\right)^2}{f''_{xx}}$ or f''_{yy} and $\dfrac{f''_{xx}\,f''_{yy} - \left(f''_{xy}\right)^2}{f''_{yy}}$.

If we consider the Hessian matrix $\left\|\begin{matrix} f''_{xx} & f''_{xy} \\ f''_{yx} & f''_{yy} \end{matrix}\right\|$, we will see that f''_{xx} and f''_{yy} are the so-called first-order leading principal minors of the matrix, while $f''_{xx}\,f''_{yy} - \left(f''_{xy}\right)^2$ is the determinant of the Hessian matrix, also called second-order leading principal minor.

We have then the following:

Definition 43 : the principal minors of a matrix are minors having as elements of its main diagonal only elements belonging to the main diagonal of the given matrix.

Definition 44 : the leading principal minors of a symmetric matrix are the n principal minors, whose order gradually increases from 1 to n, starting from any element of the main diagonal.

Example 61 : If $\mathbb{H} = \left\|\begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{matrix}\right\|$, we have two first-order leading principal minors, which are $|a_{11}|$ and $|a_{22}|$, only one second-order leading principal minor:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}\,a_{22} - a_{12}\,a_{21} .$$

So there are only two possible sequences of leading principal minors:

$|\mathbb{H}_1| = |a_{11}|$ and $|\mathbb{H}_2| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$ or $|\mathbb{H}_1| = |a_{22}|$ and $|\mathbb{H}_2| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$.

For a matrix of order 2, principal minors and leading principal minors are the same minors.

Example 62 : If $\mathbb{H} = \left\|\begin{matrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{matrix}\right\|$ we have three first-order leading principal minors: $|a_{11}|$, $|a_{22}|$ e $|a_{33}|$; we have two second-order leading principal minors: $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$ and $\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$, only one third-order leading principal minor: $\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$.

Minor $\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}$ is a principal minor but not a leading principal minor.

So there are four possible sequences of leading principal minors:

1) $|\mathbb{H}_1| = |a_{11}|$, $|\mathbb{H}_2| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$ and $|\mathbb{H}_3| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$; or

2) $|\mathbb{H}_1| = |a_{22}|$, $|\mathbb{H}_2| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$ and $|\mathbb{H}_3| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$; or

3) $|\mathbb{H}_1| = |a_{22}|$, $|\mathbb{H}_2| = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$ and $|\mathbb{H}_3| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$; finally

4) $|\mathbb{H}_1| = |a_{33}|$, $|\mathbb{H}_2| = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$ and $|\mathbb{H}_3| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$.

The first is called also the North-West leading principal minors sequence, the last is called the South-East leading principal minors sequence.

The previously obtained second-order total differential for $f : \mathbb{R}^2 \to \mathbb{R}$ can then be written as:
$$\mathrm{d}^2 f = |\mathbb{H}_1|\, Q_2^2 + \frac{|\mathbb{H}_2|}{|\mathbb{H}_1|}\, Q_1^2 \,,$$
where $|\mathbb{H}_1| = |a_{11}|$ or $|\mathbb{H}_1| = |a_{22}|$ but always $|\mathbb{H}_2| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$.
The terms Q_1^2 and Q_2^2 represent, respectively, the square of a monomial and the square of a binomial.

If $\mathrm{d}^2 f < 0$ or $\mathrm{d}^2 f > 0$ in a neighborhood of (x_0, y_0) means that the two previous expressions are negative or positive $\forall\, \mathrm{d}x$ and $\forall\, \mathrm{d}y$, i.e. their sign is independent of the choice of $\mathrm{d}x$ and $\mathrm{d}y$. We can achieve this independence in only two cases:

M) $\begin{cases} f''_{xx} < 0 \\ f''_{xx}\, f''_{yy} - \left(f''_{xy}\right)^2 > 0 \end{cases}$ or $\begin{cases} f''_{yy} < 0 \\ f''_{xx}\, f''_{yy} - \left(f''_{xy}\right)^2 > 0 \end{cases}$ $\Leftrightarrow \mathrm{d}^2 f < 0$, $\forall\, \mathrm{d}x$ and $\forall\, \mathrm{d}y$
and then (x_0, y_0) is a maximum point;
m) $\begin{cases} f''_{xx} > 0 \\ f''_{xx}\, f''_{yy} - \left(f''_{xy}\right)^2 > 0 \end{cases}$ or $\begin{cases} f''_{yy} > 0 \\ f''_{xx}\, f''_{yy} - \left(f''_{xy}\right)^2 > 0 \end{cases}$ $\Leftrightarrow \mathrm{d}^2 f > 0$, $\forall\, \mathrm{d}x$ and $\forall\, \mathrm{d}y$
and then (x_0, y_0) is a minimum point.

If $f''_{xx}\, f''_{yy} - \left(f''_{xy}\right)^2 < 0$, $\mathrm{d}^2 f$ is the sum of two terms of opposite sign, and therefore its sign varies with $\mathrm{d}x$ and $\mathrm{d}y$; so $f''_{xx}\, f''_{yy} - \left(f''_{xy}\right)^2 < 0$ is a sufficient condition to ensure that (x_0, y_0) is a saddle point. We note that this always happens when f''_{xx} and f''_{yy} have different sign.
So far, nothing can be concluded when $f''_{xx}\, f''_{yy} - \left(f''_{xy}\right)^2 = 0$.
This case will be treated later with semi-definite forms.

If now we study the $\mathrm{d}^2 f$ for $f : \mathbb{R}^3 \to \mathbb{R}$, with a similar procedure but with much more consistent calculations, we obtain an expression like:
$\mathrm{d}^2 f = |\mathbb{H}_1|\, Q_3^2 + \dfrac{|\mathbb{H}_2|}{|\mathbb{H}_1|}\, Q_2^2 + \dfrac{|\mathbb{H}_3|}{|\mathbb{H}_2|}\, Q_1^2$, where $|\mathbb{H}_1|$, $|\mathbb{H}_2|$ and $|\mathbb{H}_3|$ is any of the four possible sequences of leading principal minors seen in Example 62, and the terms Q_1^2, Q_2^2 e Q_3^2 represent, respectively, the square of a monomial, a binomial, a trinomial.

Examining the general case of $\mathrm{d}^2 f$ for $f : \mathbb{R}^n \to \mathbb{R}$, we have the expression:

$\mathrm{d}^2 f = |\mathbb{H}_1|\, Q_n^2 + \dfrac{|\mathbb{H}_2|}{|\mathbb{H}_1|}\, Q_{n-1}^2 + ... + \dfrac{|\mathbb{H}_{n-1}|}{|\mathbb{H}_{n-2}|}\, Q_2^2 + \dfrac{|\mathbb{H}_n|}{|\mathbb{H}_{n-1}|}\, Q_1^2$, where $|\mathbb{H}_1|$, $|\mathbb{H}_2|$,...,$|\mathbb{H}_n|$ is any possible sequence of leading principal minors and Q_1^2, Q_2^2,...,Q_n^2 are the square of a monomial, a binomial, ..., an n-omial.

We can then formulate, similarly to what we have seen in the case of a function of two variables, the following criteria to determine whether a quadratic form in n variables is positive or negative definite. The following is valid:
Theorem 24 : The quadratic form $Q(\mathrm{d}\mathbb{X}) = \mathrm{d}\mathbb{X} \cdot \mathbb{H} \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}}$ is:
- positive definite if and only if $|\mathbb{H}_i| > 0, \forall\, i : 1 \leq i \leq n$;
- negative definite if and only if $(-1)^i \cdot |\mathbb{H}_i| > 0, \forall\, i : 1 \leq i \leq n$.

If the form $Q(\mathrm{d}\mathbb{X}) = \mathrm{d}\mathbb{X} \cdot \mathbb{H} \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}}$ is positive definite at $\mathbb{X}_0$, as all its leading principal minors have positive sign, this is a sufficient (not necessary) condition to ensure that $\mathbb{X}_0$ is a minimum point. If the form $Q(\mathrm{d}\mathbb{X}) = \mathrm{d}\mathbb{X} \cdot \mathbb{H} \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}}$ is negative definite at $\mathbb{X}_0$, as all its leading principal minors have alternating signs, those with odd index are negative and those with even index are positive, this is a sufficient (not necessary) condition to ensure that $\mathbb{X}_0$ is a maximum point.
Since the theorem is expressed in the form of a necessary and sufficient condition, we deduce that any sequence of leading principal minors, whatever the starting element on the main diagonal, always leads to the same conclusion: therefore there is not a better choice to determine the first-order leading principal minor with which to start the sequence.

Any sequence that is not the $(+\ +\ ...\ +\ +)$ or the $(-\ +\ -\ +\ .....)$ says that the point $\mathbb{X}_0$ is a saddle point. When even only one leading principal minor is zero the above considerations are no longer valid, and we are in the field of semi-definite quadratic forms.
For the study of semi-definite quadratic forms there is a criterion similar to that given for the definite forms. But now it is not enough to analyze any sequence of leading principal minors, but we need to examine all the principal minors of the matrix.
Principal minors of a matrix are minors having as elements of its main diagonal only elements belonging to the main diagonal of the given matrix.

Example 63 : If $\mathbb{A} = \begin{Vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{Vmatrix}$ we have three first-order principal minors: $|a_{11}|$, $|a_{22}|$ and $|a_{33}|$, and a third-order principal minor: $\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$; these are also leading principal minors; there are instead three second-order principal minors: $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$, $\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$ and $\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}$. The last of these is not a leading principal minor.

To see if a quadratic form is semidefinite the following criteria apply, for which we denote by $|MP_i|$ any principal minor of order i:
Theorem 25 : The quadratic form $Q(\mathrm{d}\mathbb{X}) = \mathrm{d}\mathbb{X} \cdot \mathbb{H} \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}}$ is:
- positive semidefinite if and only if $|MP_i| \geq 0, \forall\, i : 1 \leq i \leq n$;
- negative semidefinite if and only if $(-1)^i \cdot |MP_i| \geq 0, \forall\, i : 1 \leq i \leq n$.

Signs sequences are the same as those relating to the definite forms, but now including the possibility of the presence of zeros. Each sequence of signs that is not one of the two former ones leads to the conclusion that the quadratic form is indefinite.

Pay attention to the fact that when we have a definite quadratic form at $\mathbb{X}_0$ we can immediately deduce the nature of the stationary point; on the contrary, if we have a semidefinite quadratic form at $\mathbb{X}_0$, there is no conclusion allowed, we can only exclude a possibility:
-if at $\mathbb{X}_0$ the form $Q(\mathrm{d}\mathbb{X}) = \mathrm{d}\mathbb{X} \cdot \mathbb{H} \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}}$ is positive semidefinite, then $\mathbb{X}_0$ cannot be a maximum point, and then $\mathbb{X}_0$ may be either a minimum or a saddle point;
-if at $\mathbb{X}_0$ the form $Q(\mathrm{d}\mathbb{X}) = \mathrm{d}\mathbb{X} \cdot \mathbb{H} \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}}$ is negative semidefinite, then $\mathbb{X}_0$ cannot be a minimum point, and then $\mathbb{X}_0$ may be either a maximum or a saddle point.

How to decide between the remaining two possibilities depends on the function we are studying, depending on which we can use different methodologies. The most common is to study, with various tricks, the sign of the difference $f(\mathbb{X}) - f(\mathbb{X}_0)$.
Here are some examples.

Example 64 : Let $\mathbb{H} = \left\|\begin{matrix} 4 & 2 & 1 \\ 2 & 1 & 2 \\ 1 & 2 & 1 \end{matrix}\right\|$ be the matrix of the quadratic form $\mathrm{d}^2 f$. First-order principal minors are $4 > 0, 1 > 0$ and $1 > 0$; second-order principal minors are $\begin{vmatrix} 4 & 2 \\ 2 & 1 \end{vmatrix} = 0$, $\begin{vmatrix} 4 & 1 \\ 1 & 1 \end{vmatrix} = 3 > 0$; $\begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix} = -3 < 0$. The presence of an even order negative minor immediately leads to the conclusion that the quadratic form is indefinite.

Example 65 : Let $\mathbb{H} = \left\|\begin{matrix} -4 & 2 & 3 \\ 2 & -2 & -1 \\ 3 & -1 & -3 \end{matrix}\right\|$ be the matrix of the quadratic form $\mathrm{d}^2 f$. First-order principal minors are $-4 < 0$, $-2 < 0$ and $-3 < 0$; second-order principal minors are $\begin{vmatrix} -4 & 2 \\ 2 & -2 \end{vmatrix} = 4 > 0$, $\begin{vmatrix} -4 & 3 \\ 3 & -3 \end{vmatrix} = 3 > 0$; $\begin{vmatrix} -2 & -1 \\ -1 & -3 \end{vmatrix} = 5 > 0$. Third-order principal minor is $\begin{vmatrix} -4 & 2 & 3 \\ 2 & -2 & -1 \\ 3 & -1 & -3 \end{vmatrix} = -2 < 0$. The quadratic form is then negative definite. To reach this conclusion, however, it was enough to examine only the sequence of leading principal minors:

$$|\mathbb{H}_1| = -4 < 0;\ |\mathbb{H}_2| = \begin{vmatrix} -4 & 2 \\ 2 & -2 \end{vmatrix} = 4 > 0,\ |\mathbb{H}_3| = \begin{vmatrix} -4 & 2 & 3 \\ 2 & -2 & -1 \\ 3 & -1 & -3 \end{vmatrix} = -2 < 0.$$

Example 66 : Let $\mathbb{H} = \left\|\begin{matrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{matrix}\right\|$ be the matrix of the quadratic form $\mathrm{d}^2 f$. First-order principal minors are $1 > 0, 4 > 0$ and $9 > 0$; second-order principal minors are $\begin{vmatrix} 1 & 2 \\ 2 & 4 \end{vmatrix} = 0$, $\begin{vmatrix} 4 & 6 \\ 6 & 9 \end{vmatrix} = 0$; $\begin{vmatrix} 1 & 3 \\ 3 & 9 \end{vmatrix} = 0$. Third-order principal minor is the determinant of $\mathbb{H}$, obviously equal to zero. Then the quadratic form is positive semidefinite.

Let us see now with some examples how to apply these methods for the study of stationary points of a function of several variables.

Example 67 : Let us consider the function $f(x,y) = x^2 + y^4$. Since $\nabla f(\mathbb{X}) = \left(2x, 4y^3\right)$, the only stationary point is $(0,0)$. Is $\mathbb{H} = \left\| \begin{matrix} 2 & 0 \\ 0 & 12y^2 \end{matrix} \right\|$, from which, substituting, we get $\mathbb{H}(0,0) = \left\| \begin{matrix} 2 & 0 \\ 0 & 0 \end{matrix} \right\|$. The quadratic form $\mathrm{d}^2 f(0,0)$ is positive semidefinite, so $(0,0)$ it is not a maximum point. As $f(0,0) = 0$ and being $x^2 + y^4 > 0$, $\forall\,(x,y) \neq (0,0)$, we immediately see that $(0,0)$ is a minimum point, moreover, an absolute minimum point.

Example 68 : Let us consider the function $f(x,y) = x^2 - y^4$, as $\nabla f(\mathbb{X}) = \left(2x, -4y^3\right)$, the only stationary point is $(0,0)$. It results $\mathbb{H} = \left\| \begin{matrix} 2 & 0 \\ 0 & -12y^2 \end{matrix} \right\|$, from which, substituting, we get $\mathbb{H}(0,0) = \left\| \begin{matrix} 2 & 0 \\ 0 & 0 \end{matrix} \right\|$. The quadratic form $\mathrm{d}^2 f(0,0)$ is positive semidefinite, so $(0,0)$ it is not a maximum point. As $f(0,0) = 0$, $f(x,0) = x^2 > 0$ and $f(0,y) = -y^4 < 0$, we see that $(0,0)$ is a saddle point.

We note therefore that, for the same Hessian matrix $\mathbb{H}(0,0) = \left\| \begin{matrix} 2 & 0 \\ 0 & 0 \end{matrix} \right\|$, the conclusion changes depending on the function we are examining.

Example 69 : Given $f(x,y) = 3x^3 - y^3 - 3x^2y + 3xy^2 - 3x^2 - 3x + 3y$, let us determine its possible maximum and minimum points. Imposing the first-order condition, $\nabla f(\mathbb{X}) = 0$, we have the system: $\begin{cases} f'_x = 9x^2 - 6xy + 3y^2 - 6x - 3 = 0 \\ f'_y = -3y^2 - 3x^2 + 6xy + 3 = 0 \end{cases}$.

Adding the two equations we have the system:

$\begin{cases} 6x^2 - 6x = 0 \\ 2xy - y^2 - x^2 + 1 = 0 \end{cases}$, or: $\begin{cases} 6x(x-1) = 0 \\ 2xy - y^2 - x^2 + 1 = 0 \end{cases}$ from which:

$\begin{cases} x = 0 \\ 1 - y^2 = 0 \end{cases} \Rightarrow \begin{cases} x = 0 \\ y = 1 \end{cases}$ and $\begin{cases} x = 0 \\ y = -1 \end{cases}$, or:

$\begin{cases} x = 1 \\ 2y - y^2 = 0 \end{cases} \Rightarrow \begin{cases} x = 1 \\ y(2-y) = 0 \end{cases} \Rightarrow \begin{cases} x = 1 \\ y = 0 \end{cases}$ and $\begin{cases} x = 1 \\ y = 2 \end{cases}$.

There are four stationary points: $\mathrm{P}_1 = (0{,}1)$, $\mathrm{P}_2 = (0,-1)$, $\mathrm{P}_3 = (1{,}0)$, $\mathrm{P}_4 = (1{,}2)$.

Using second-order conditions, we have first: $\mathbb{H} = \left\| \begin{matrix} 18x - 6y - 6 & -6x + 6y \\ -6x + 6y & -6y + 6x \end{matrix} \right\|$.

Studying the Hessian in each of the four points we have:

$\mathbb{H}(\mathrm{P}_1) = \left\| \begin{matrix} -12 & 6 \\ 6 & -6 \end{matrix} \right\| \Rightarrow \begin{cases} f''_{xx} = -12 < 0 \\ f''_{xx}\, f''_{yy} - \left(f''_{xy}\right)^2 = 72 - 36 = 36 > 0 \end{cases}$,

and so $(0{,}1)$ is a maximum point;

$\mathbb{H}(\mathrm{P}_2) = \left\| \begin{matrix} 0 & -6 \\ -6 & 6 \end{matrix} \right\| \Rightarrow f''_{xx}\, f''_{yy} - \left(f''_{xy}\right)^2 = 0 - 36 = -36 < 0$,

and so $(0,-1)$ is a saddle point;

$\mathbb{H}(\mathrm{P}_3) = \left\| \begin{matrix} 12 & -6 \\ -6 & 6 \end{matrix} \right\| \Rightarrow \begin{cases} f''_{xx} = 12 > 0 \\ f''_{xx}\, f''_{yy} - \left(f''_{xy}\right)^2 = 72 - 36 = 36 > 0 \end{cases}$,

and so $(1{,}0)$ is a minimum point;

$\mathbb{H}(\mathrm{P}_4) = \left\| \begin{matrix} 0 & 6 \\ 6 & -6 \end{matrix} \right\| \Rightarrow f''_{xx}\, f''_{yy} - \left(f''_{xy}\right)^2 = 0 - 36 = -36 < 0$,

and so $(1{,}2)$ is a saddle point.

Example 70 : Let us conclude the study of the function $f(x,y) = \left(y - x^2\right)\left(y - x^4\right)$, whose stationary points are $\begin{cases} x = 0 \\ y = 0 \end{cases}$; $\begin{cases} x = 1 \\ y = 1 \end{cases}$; $\begin{cases} x = -1 \\ y = 1 \end{cases}$; $\begin{cases} x = \frac{1}{\sqrt{2}} \\ y = \frac{3}{8} \end{cases}$; $\begin{cases} x = -\frac{1}{\sqrt{2}} \\ y = \frac{3}{8} \end{cases}$. Using second-order conditions, we have $\mathbb{H}(x,y) = \begin{Vmatrix} 30x^4 - 2y - 12x^2y & -2x - 4x^3 \\ -2x - 4x^3 & 2 \end{Vmatrix}$, and so:

$\mathbb{H}(0,0) = \begin{Vmatrix} 0 & 0 \\ 0 & 2 \end{Vmatrix}$ so the quadratic form is positive semidefinite, but we have already seen that $(0,0)$ is a saddle point;

$\mathbb{H}(1,1) = \begin{Vmatrix} 16 & -6 \\ -6 & 2 \end{Vmatrix}$ and so $\begin{cases} |\mathbb{H}_1| = 16 > 0 \\ |\mathbb{H}_2| = 32 - 36 = -4 < 0 \end{cases}$,

and therefore $(1,1)$ is a saddle point;

$\mathbb{H}(-1,1) = \begin{Vmatrix} 16 & 6 \\ 6 & 2 \end{Vmatrix}$ and so $\begin{cases} |\mathbb{H}_1| = 16 > 0 \\ |\mathbb{H}_2| = 32 - 36 = -4 < 0 \end{cases}$,

and therefore $(-1,1)$ is a saddle point;

$\mathbb{H}\left(\frac{1}{\sqrt{2}}, \frac{3}{8}\right) = \begin{Vmatrix} \frac{9}{2} & -\frac{4}{\sqrt{2}} \\ -\frac{4}{\sqrt{2}} & 2 \end{Vmatrix}$ and so $\begin{cases} |\mathbb{H}_1| = \frac{9}{2} > 0 \\ |\mathbb{H}_2| = 9 - 8 = 1 > 0 \end{cases}$,

and then $\left(\frac{1}{\sqrt{2}}, \frac{3}{8}\right)$ is a minimum point;

$\mathbb{H}\left(-\frac{1}{\sqrt{2}}, \frac{3}{8}\right) = \begin{Vmatrix} \frac{9}{2} & \frac{4}{\sqrt{2}} \\ \frac{4}{\sqrt{2}} & 2 \end{Vmatrix}$ and so $\begin{cases} |\mathbb{H}_1| = \frac{9}{2} > 0 \\ |\mathbb{H}_2| = 9 - 8 = 1 > 0 \end{cases}$,

and then $\left(-\frac{1}{\sqrt{2}}, \frac{3}{8}\right)$ also is a minimum point.

Example 71 : Given $f(x,y,z) = x^2 + y^2 + z^2 - x\,y^2$, let us determine its possible maximum and minimum points. First-order conditions give rise to the system:

$\begin{cases} f'_x = 2x - y^2 = 0 \\ f'_y = 2y - 2x\,y = 2y\,(1-x) = 0 \\ f'_z = 2z = 0 \end{cases}$ from which we get the solutions:

$\begin{cases} x = 0 \\ y = 0 \\ z = 0 \end{cases}$ or $\begin{cases} x = 1 \\ y^2 = 2 \\ z = 0 \end{cases}$ from which $\begin{cases} x = 1 \\ y = \sqrt{2} \\ z = 0 \end{cases}$ and $\begin{cases} x = 1 \\ y = -\sqrt{2} \\ z = 0 \end{cases}$.

It follows that $\mathbb{H}(x,y,z) = \begin{Vmatrix} 2 & -2y & 0 \\ -2y & 2-2x & 0 \\ 0 & 0 & 2 \end{Vmatrix}$ from which we obtain:

$\mathbb{H}(0,0,0) = \begin{Vmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{Vmatrix}$ and so: $\begin{cases} |\mathbb{H}_1| = 2 > 0 \\ |\mathbb{H}_2| = 4 > 0 \\ |\mathbb{H}_3| = 8 > 0 \end{cases}$

and then $(0,0,0)$ is a minimum point;

$\mathbb{H}\left(1, \sqrt{2}, 0\right) = \begin{Vmatrix} 2 & -2\sqrt{2} & 0 \\ -2\sqrt{2} & 0 & 0 \\ 0 & 0 & 2 \end{Vmatrix}$ and so: $\begin{cases} |\mathbb{H}_1| = 2 > 0 \\ |\mathbb{H}_2| = -8 < 0 \\ |\mathbb{H}_3| = -16 < 0 \end{cases}$

and then $\left(1, \sqrt{2}, 0\right)$ is a saddle point;

$$\mathbb{H}\Big(1,\,-\sqrt{2},0\Big) = \left\|\begin{matrix} 2 & 2\sqrt{2} & 0 \\ 2\sqrt{2} & 0 & 0 \\ 0 & 0 & 2 \end{matrix}\right\| \text{ and so: } \begin{cases} |\mathbb{H}_1| = 2 > 0 \\ |\mathbb{H}_2| = -8 < 0 \\ |\mathbb{H}_3| = -16 < 0 \end{cases}$$

and then $\Big(1,\,-\sqrt{2},0\Big)$ is a saddle point.

Example 72 : Given $f(x,y) = (x-1)^4 - y\,(x-1)^2 + y^2$, let us determine its possible relative maximum and minimum points. First-order conditions give rise to the system:

$$\begin{cases} f'_x = 4(x-1)^3 - 2y(x-1) = 0 \\ f'_y = -(x-1)^2 + 2y = 0 \end{cases} \Rightarrow \begin{cases} 4(x-1)^3 - (x-1)^3 = 0 \\ 2y = (x-1)^2 \end{cases} \Rightarrow \begin{cases} 3(x-1)^3 = 0 \\ 2y = (x-1)^2 \end{cases},$$

so $\begin{cases} x = 1 \\ y = 0 \end{cases}$ is the only stationary point. Moreover:

$$\mathbb{H}(x,y) = \left\|\begin{matrix} 12(x-1)^2 - 2y & -2(x-1) \\ -2(x-1) & 2 \end{matrix}\right\| \text{ from which } \mathbb{H}(1,0) = \left\|\begin{matrix} 0 & 0 \\ 0 & 2 \end{matrix}\right\|.$$

So $\mathrm{d}^2 f(1,0)$ is a positive semidefinite quadratic form, so that $(1,0)$ cannot be a maximum point. As $\mathrm{d}^2 f(1,0) = 2\,(\mathrm{d}y)^2$, let us investigate in the direction $\mathrm{d}y = 0$, or $y = 0$. We get $f(x,0) = (x-1)^4$, which would indicate $x = 1$ (and $y = 0$) as a minimum point.
Such a study, a one-dimensional study, does not allow us, however, to conclude affirmatively that $(1,0)$ is a minimum point; it could be used, although this is not the case, to exclude that $(1,0)$ is a minimum point. But $(1,0)$ is indeed a minimum point, just write:

$$f(x,y) = (x-1)^4 - y\,(x-1)^2 + y^2 = \frac{1}{4}(x-1)^4 - 2y\,\frac{(x-1)^2}{2} + y^2 + \frac{3}{4}(x-1)^4 =$$

$$f(x,y) = \left(\frac{1}{2}(x-1)^2 - y\right)^2 + \frac{3}{4}(x-1)^4 > 0 = f(1,0)\,,\ \forall\,(x,y) \neq (1,0)\,.$$

Example 73 : Given $f(x,y) = 3y^2 + 6xy - x^3 - 9x - 6y$, let us determine its possible relative maximum and minimum points. First-order conditions give rise to the system:

$$\begin{cases} f'_x = 6y - 3x^2 - 9 = 0 \\ f'_y = 6y + 6x - 6 = 0 \end{cases} \Rightarrow \begin{cases} 2 - 2x - x^2 - 3 = 0 \\ y = 1 - x \end{cases} \Rightarrow \begin{cases} -(x+1)^2 = 0 \\ y = 1 - x \end{cases} \Rightarrow \begin{cases} x = -1 \\ y = 2 \end{cases}$$

which is the only solution.

Moreover $\mathbb{H}(x,y) = \left\|\begin{matrix} -6x & 6 \\ 6 & 6 \end{matrix}\right\|$ from which $\mathbb{H}(-1,2) = \left\|\begin{matrix} 6 & 6 \\ 6 & 6 \end{matrix}\right\|$, and so $\mathrm{d}^2 f(-1,2)$ is a positive semidefinite form.
So $(-1,2)$ cannot be a maximum point.
As $\mathrm{d}^2 f(-1,2) = 6\,(\mathrm{d}x + \mathrm{d}y)^2$, we get $\mathrm{d}^2 f(-1,2) = 0$ when $\mathrm{d}x = -\mathrm{d}y$.
Looking at the function on the line $y = -x + 1$, passing through $(-1,2)$ and parallel to $y = -x$, we have: $f(x, 1-x) = -x^3 - 3x^2 - 3x - 3 = -3\,(x+1)^3$ and also: $f'_x(x, 1-x) = -9\,(x+1)^2$, which is negative $\forall\, x \neq -1$.
Then along this line the function is always decreasing, then the point $(-1,2)$ is not even a minimum point, and hence is a saddle point. This time analyzing in a particular direction leads to a negative conclusion, that is, to exclude the minimum point and then it gives us the certainty of the saddle point.

THE HESSIAN MATRIX EIGENVALUES METHOD

There is also another method to study quadratic forms, based on the eigenvalues of the symmetric matrix that generates such quadratic form. We know that a symmetric matrix has only real eigenvalues, and that it can always be diagonalized by an orthogonal matrix. That is, if $\mathbb{H}$ is a symmetric matrix, there exists an orthogonal matrix $\mathbb{P}$ such that: $\mathbb{H} \cdot \mathbb{P} = \mathbb{P} \cdot \mathbb{D}$, from which $\mathbb{P}^{-1} \cdot \mathbb{H} \cdot \mathbb{P} = \mathbb{P}^{\mathrm{T}} \cdot \mathbb{H} \cdot \mathbb{P} = \mathbb{D}$, where $\mathbb{D}$ is the diagonal matrix having as elements of its

main diagonal the eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ of $\mathbb{H}$. The matrix $\mathbb{P}$ is the modal matrix of $\mathbb{H}$, i.e. the matrix having as columns the normalized eigenvectors of $\mathbb{H}$, which, for the properties of symmetric matrices, are orthogonal to each other.
In order for the quadratic form $\mathrm{d}^2 f = \mathrm{d}\mathbb{X} \cdot \mathbb{H} \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}}$ to be definite it must be $\mathrm{d}^2 f > 0$ or $\mathrm{d}^2 f < 0$, $\forall\, \mathrm{d}\mathbb{X} \neq \mathbb{O}$. Let us consider $\mathrm{d}\mathbb{X} = \mathrm{d}\mathbb{Y} \cdot \mathbb{P}^{\mathrm{T}}$, where $\mathbb{P}$ is the modal matrix of $\mathbb{H}$. As $|\mathbb{P}| \neq 0$, the linear application $\mathrm{d}\mathbb{Y} \to \mathrm{d}\mathbb{Y} \cdot \mathbb{P}^{\mathrm{T}} = \mathrm{d}\mathbb{X}$ is a bijective correspondence $\mathbb{R}^n \to \mathbb{R}^n$, so we get:
$\mathrm{d}\mathbb{X} \cdot \mathbb{H} \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}} = \mathrm{d}\mathbb{Y} \cdot \mathbb{P}^{\mathrm{T}} \cdot \mathbb{H} \cdot \left(\mathrm{d}\mathbb{Y} \cdot \mathbb{P}^{\mathrm{T}}\right)^{\mathrm{T}} = \mathrm{d}\mathbb{Y} \cdot \mathbb{P}^{\mathrm{T}} \cdot \mathbb{H} \cdot \mathbb{P} \cdot (\mathrm{d}\mathbb{Y})^{\mathrm{T}} = \mathrm{d}\mathbb{Y} \cdot \mathbb{D} \cdot (\mathrm{d}\mathbb{Y})^{\mathrm{T}}$.
If $\mathrm{d}\mathbb{Y} \cdot \mathbb{D} \cdot (\mathrm{d}\mathbb{Y})^{\mathrm{T}} > 0$ or if $\mathrm{d}\mathbb{Y} \cdot \mathbb{D} \cdot (\mathrm{d}\mathbb{Y})^{\mathrm{T}} < 0$, $\forall\, \mathrm{d}\mathbb{Y} \neq \mathbb{O}$, it is also $\mathrm{d}\mathbb{X} \cdot \mathbb{H} \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}} > 0$ or $\mathrm{d}\mathbb{X} \cdot \mathbb{H} \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}} < 0$, $\forall\, \mathrm{d}\mathbb{X} \neq \mathbb{O}$. But
$\mathrm{d}\mathbb{Y} \cdot \mathbb{D} \cdot (\mathrm{d}\mathbb{Y})^{\mathrm{T}} = \lambda_1 (\mathrm{d}y_1)^2 + \lambda_2 (\mathrm{d}y_2)^2 + ... + \lambda_n (\mathrm{d}y_n)^2$,
from which it immediately follows
Theorem 26 : The quadratic form $\mathrm{d}\mathbb{X} \cdot \mathbb{H} \cdot (\mathrm{d}\mathbb{X})^{\mathrm{T}}$ is:
- positive definite if and only if $\lambda_i > 0, \forall\, i : 1 \le i \le n$;
- negative definite if and only if $\lambda_i < 0, \forall\, i : 1 \le i \le n$;
- positive semidefinite if and only if $\lambda_i \ge 0, \forall\, i : 1 \le i \le n$ **and** $\exists\, \lambda_k = 0$;
- negative semidefinite if and only if $\lambda_i \le 0, \forall\, i : 1 \le i \le n$ **and** $\exists\, \lambda_k = 0$;
- indefinite if $\exists\, \lambda_i > 0$ **and** $\exists\, \lambda_j < 0$.

Theorem 24 and Theorem 26 express necessary and sufficient conditions to get a definite quadratic form, then the two methods (leading principal minors and eigenvalues) are not alternative but always lead to the same conclusion.

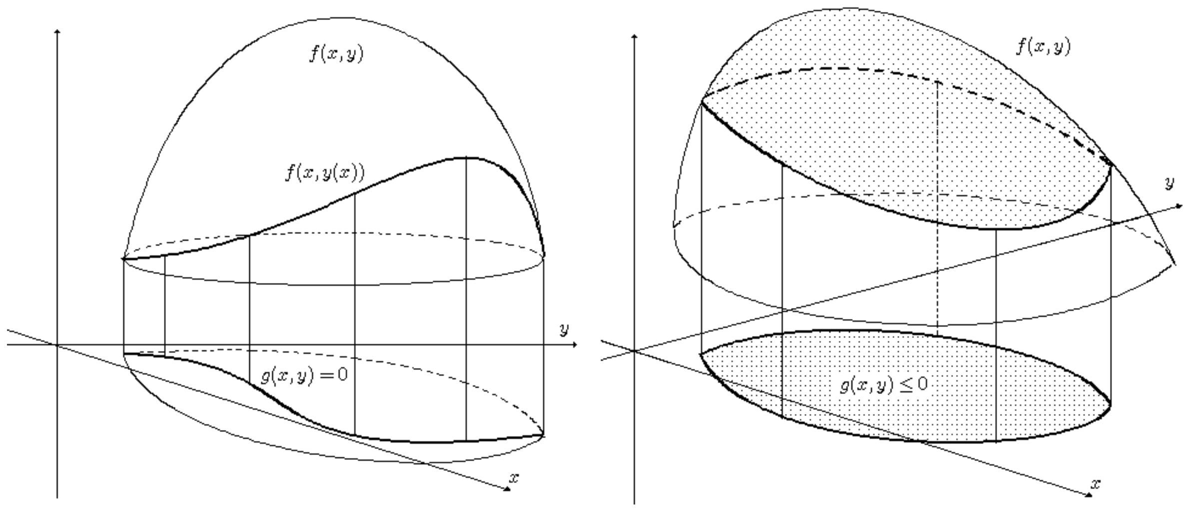
What we have seen so far about the second-order conditions for stationary points analysis is based on the analysis at $\mathbb{X}_0$, which verifies if $\mathrm{d}^2 f(\mathbb{X}_0)$ is a definite positive or negative form, or an indefinite one; if the form $\mathrm{d}^2 f(\mathbb{X}_0)$ is semidefinite we can only exclude one of the three possibilities, in order to decide, with a further analysis, between the two remaining, one of which will always be that of the saddle point.
The situation is different if we can lead a global analysis. If $\mathrm{d}^2 f(\mathbb{X})$ is positive or negative semi-definite, both at $\mathbb{X}_0$ than in the whole domain, then this is sufficient to guarantee that $\mathbb{X}_0$ is a, respectively, minimum or maximum point.

Example 74 : Let us use again the function $f(x, y) = x^2 + y^4$, with the stationary point $(0, 0)$ and the Hessian matrix $\mathbb{H}(x, y) = \left\|\begin{matrix} 2 & 0 \\ 0 & 12y^2 \end{matrix}\right\|$.
The result is $\mathrm{d}^2 f(x, y) = 2\,(\mathrm{d}x)^2 + 12\,y^2 (\mathrm{d}y)^2 \ge 0$, $\forall\,(\mathrm{d}x, \mathrm{d}y) \in \mathbb{R}^2$. So the form $\mathrm{d}^2 f(x, y)$ is positive semidefinite not only at $(0, 0)$ but throughout $\mathbb{R}^2$, and then, as already seen, $(0, 0)$ it is a minimum point.

CONSTRAINED MAXIMA AND MINIMA

The search for maximum and minimum points, both relative and absolute, for what we have seen so far, can be decomposed into three different problems. The first covers the search and analysis of the stationary points of a function in all its existence field, whose solutions are normally interior points of the domain. This type of research requires differentiable functions. The second problem, we are not dealing with, concerns the analysis of the points where a function is defined but is not differentiable. These too can be maximum or minimum points. The third problem concerns the search for maximum and minimum points relative to some appropriate subset of the existence field. Weierstrass's theorem gives a sufficient condition to

guarantee the existence of maximum and minimum for continuous functions in a compact set. There are two ways in which we can present this third type of problem.
The first is the so-called maxima and minima with equality constraints, which in the simplest case is in the form $\begin{cases} \text{Max/min } f(x,y) \\ \text{u.c. : } g(x,y)=0 \end{cases}$, where u.c. means "under constraints". We do not look for extremes for $f(x,y)$ throughout the domain (these maximum or minimum points from now on will be called "free"), but only between points that satisfy the equation $g(x,y)=0$.
As $g(x,y)=0$ is called the constraint, $f(x,y)$ is called the objective function.
If the function $g(x,y)$ satisfies suitable assumptions, related to those of Dini's theorem on implicit functions, we can consider one of the two variables as a function (explicit or implicit) of the other, for example $y=y(x)$, and then we can draw a curve in the plane: $(x,y(x))$; then the problem is reduced to finding maxima and minima of the function $f(x,y(x))$, i.e. of the curve, projection on the surface $f(x,y)$ of the points of the curve defined by $g(x,y)=0$.
Increasing the number of variables and equations we have geometric representations of the problem with larger dimensions than the one just described.



Let us consider instead, as a second type of problem, with a constraint expressed in the form of inequality: $g(x,y)\le 0$. Usually (but not necessarily) the points satisfying $g(x,y)\le 0$ are interior and boundary points of a region contained in the domain, and the problem then is to find maximum and minimum points of the function $f(x,y)$ inside or on the boundary of the selected region: $\begin{cases} \text{Max/min } f(x,y) \\ \text{u.c. : } g(x,y)\le 0 \end{cases}$. The search for extremes at interior points of the defined region is similar to that of free maxima and minima, the search in boundary points is similar to that of maxima and minima with equality constraints.
These two problems are named maximization or minimization with equality or inequality constraints.

EXTREMES WITH EQUALITY CONSTRAINTS - FIRST ORDER CONDITIONS

We begin dealing with the first case described: $\begin{cases} \text{Max/min } z=f(x,y) \\ \text{u.c. : } g(x,y)=0 \end{cases}$. Let $\mathcal{E}$ be the set of points satisfying the constraint $g(x,y)=0$; let us suppose that $f(x,y)$ and $g(x,y)$ are differentiable functions and that $\nabla g(x,y)\ne(0,0)$, $\forall\,(x,y)\in\mathcal{E}$. This last condition allows us to apply Dini's theorem at each point of $\mathcal{E}$, and then we can determine, explicitly or implicitly, a variable as a function of the other.

This condition can also be formulated as $\operatorname{Rank}\left(\dfrac{\partial g}{\partial(x,y)}\right)=1=\operatorname{Max}$.

If the rank of the Jacobian, i.e. the gradient of g, is maximum, or equal to 1, the two derivatives g'_x and g'_y cannot be simultaneously equal to zero. Let us suppose it is defined $y=y(x)$, so we have two functions compositions:

$\mathbb{R}\to\mathbb{R}^2\xrightarrow{f}\mathbb{R},\, x\to(x,y(x))\to f(x,y)=z$, and

$\mathbb{R}\to\mathbb{R}^2\xrightarrow{g}\mathbb{R},\, x\to(x,y(x))\to g(x,y)=0$.

Deriving in the first the variable z with respect to x we get:

$$\frac{\mathrm{d}z}{\mathrm{d}x}=\frac{\partial f}{\partial x}\cdot\frac{\mathrm{d}x}{\mathrm{d}x}+\frac{\partial f}{\partial y}\cdot\frac{\mathrm{d}y}{\mathrm{d}x}=f'_x+f'_y\cdot y'(x).$$

From the constraint equation we get, using the derivative of the implicit function, $y'(x)$ and so:

$y'(x)=-\dfrac{g'_x}{g'_y}$ with which, substituting, we obtain: $\dfrac{\mathrm{d}z}{\mathrm{d}x}=f'_x-f'_y\cdot\dfrac{g'_x}{g'_y}$.

Let us suppose, by assumption, that (x_0,y_0) is a solution, a maximum or a minimum point, to the problem. Since the function is a composition of differentiable functions, it must be, for Fermat's Theorem: $\dfrac{\mathrm{d}z}{\mathrm{d}x}=f'_x-f'_y\cdot\dfrac{g'_x}{g'_y}=0$, or $f'_x=f'_y\cdot\dfrac{g'_x}{g'_y}$ and so $f'_x\cdot g'_y=f'_y\cdot g'_x$ and finally:

$$\begin{vmatrix} f'_x & f'_y \\ g'_x & g'_y \end{vmatrix}=\left|\frac{\partial(f,g)}{\partial(x,y)}\right|=0.$$

As $\left|\dfrac{\partial(f,g)}{\partial(x,y)}\right|=\begin{vmatrix}\nabla f\\ \nabla g\end{vmatrix}=0$, ∇f and ∇g are linearly dependent vectors, and so $\nabla f=\lambda\,\nabla g$, $\lambda\in\mathbb{R}$. But then $\nabla f-\lambda\,\nabla g=\nabla(f-\lambda\,g)=\mathbb{O}$.

So we have obtained a necessary condition for the point (x_0,y_0) to be a solution of the problem $\begin{cases}\text{Max/min } f(x,y)\\ \text{u.c.}: g(x,y)=0\end{cases}$.

Even now a gradient must be cancelled, not that of the objective function, as in the case of free maxima and minima, but that of the function $f(x,y)-\lambda\,g(x,y)$.

The function $\Lambda(x,y,\lambda)=f(x,y)-\lambda\,g(x,y)$ is called "Lagrangian function" while λ is called "Lagrange's multiplier". So we have:

Theorem 27 : If f and g are differentiable functions $\mathbb{R}^2\to\mathbb{R}$ and $\nabla g(x,y)\neq(0,0)$; if (x_0,y_0) is a solution of the problem $\begin{cases}\text{Max/min } f(x,y)\\ \text{u.c.}: g(x,y)=0\end{cases}$, given the Lagrangian function $\Lambda(x,y,\lambda)=f(x,y)-\lambda\,g(x,y)$, then there is a value λ_0 such that: $\nabla\Lambda(x_0,y_0,\lambda_0)=\mathbb{O}$.

First order conditions become $\nabla\Lambda(x,y,\lambda)=\mathbb{O}$, bearing in mind, however, that any solution must satisfy the constraint, so we have to solve the system: $\begin{cases}\Lambda'_x=f'_x-\lambda\,g'_x=0\\ \Lambda'_y=f'_y-\lambda\,g'_y=0\\ g(x,y)=0\end{cases}$.

But $\Lambda'_\lambda=-\,g(x,y)$, so the latter system can be indeed be seen as $\nabla\Lambda(x,y,\lambda)=\mathbb{O}$, meaning, however, $\Lambda(x,y,\lambda)$ as a function of variables x, y and λ, and so we solve the system:

$$\begin{cases}\Lambda'_x=f'_x-\lambda\,g'_x=0\\ \Lambda'_y=f'_y-\lambda\,g'_y=0\\ \Lambda'_\lambda=-\,g(x,y)=0\end{cases}.$$

Let us study the problem $\begin{cases} \text{Max/min } w = f(x,y,z) \\ \text{u.c. : } g(x,y,z) = 0 \end{cases}$. We still have only one constraint, but independent variables now are three. If $\dfrac{\partial(g)}{\partial(x,y,z)}$ has rank equal to 1 throughout $\mathcal{E}$, from Dini's theorem we obtain (explicitly or implicitly) $z = z(x,y)$, obtaining the following functions composition: $\mathbb{R}^2 \to \mathbb{R}^3 \xrightarrow{f} \mathbb{R}$, $(x,y) \to (x,y,z(x,y)) \to f(x,y,z) = w$, and then we must find maxima and minima of a function not of one but of two variables.
If f and g are differentiable functions, we now must put:

$$\begin{cases} \dfrac{\partial w}{\partial x} = \dfrac{\partial f}{\partial x} \cdot \dfrac{\partial x}{\partial x} + \dfrac{\partial f}{\partial y} \cdot \dfrac{\partial y}{\partial x} + \dfrac{\partial f}{\partial z} \cdot \dfrac{\partial z}{\partial x} = f'_x + f'_y \cdot 0 + f'_z \cdot z'_x \\ \dfrac{\partial w}{\partial y} = \dfrac{\partial f}{\partial x} \cdot \dfrac{\partial x}{\partial y} + \dfrac{\partial f}{\partial y} \cdot \dfrac{\partial y}{\partial y} + \dfrac{\partial f}{\partial z} \cdot \dfrac{\partial z}{\partial y} = f'_x \cdot 0 + f'_y + f'_z \cdot z'_y \end{cases}.$$

But, from Dini's theorem, we obtain, from the constraint equation, as derivatives of the implicit function, z'_x and z'_y, i.e. $z'_x = -\dfrac{g'_x}{g'_z}$ and $z'_y = -\dfrac{g'_y}{g'_z}$, from which, substituting, we have: $\begin{cases} \dfrac{\partial w}{\partial x} = f'_x - f'_z \cdot \dfrac{g'_x}{g'_z} \\ \dfrac{\partial w}{\partial y} = f'_y - f'_z \cdot \dfrac{g'_y}{g'_z} \end{cases}$.

Suppose, by assumption, that (x_0, y_0, z_0) is a solution, a maximum or a minimum point, to the problem. It must be $\dfrac{\partial w}{\partial x} = \dfrac{\partial w}{\partial y} = 0$, or $\begin{cases} \dfrac{\partial w}{\partial x} = f'_x \cdot g'_z - f'_z \cdot g'_x = 0 \\ \dfrac{\partial w}{\partial y} = f'_y \cdot g'_z - f'_z \cdot g'_y = 0 \end{cases}$, from which it follows: $\begin{vmatrix} f'_x & f'_z \\ g'_x & g'_z \end{vmatrix} = \begin{vmatrix} f'_y & f'_z \\ g'_y & g'_z \end{vmatrix} = 0$. But then $\left\| \begin{matrix} f'_x & f'_y & f'_z \\ g'_x & g'_y & g'_z \end{matrix} \right\| = \dfrac{\partial(f,g)}{\partial(x,y,z)} = \left\| \begin{matrix} \nabla f \\ \nabla g \end{matrix} \right\|$ has rank equal to 1, and so $\nabla f = \lambda \nabla g$, from which we get again $\nabla(f - \lambda g) = \nabla \Lambda = \mathbb{O}$.
Increasing the number of independent variables does not lead to changes from an operational point of view: at a constrained maximum or minimum point it is still necessary to cancel the gradient of the Lagrangian function, but this time we get the system:
$\begin{cases} \Lambda'_x = f'_x - \lambda g'_x = 0 \\ \Lambda'_y = f'_y - \lambda g'_y = 0 \\ \Lambda'_z = f'_z - \lambda g'_z = 0 \\ \Lambda'_\lambda = -g(x,y,z) = 0 \end{cases}$, of four equations in four variables x, y, z and λ.
With the due changes, it can be stated for this case a theorem similar to Theorem 27.

The next generalization of the problem will involve constraints.
We study in fact the problem : $\begin{cases} \text{Max/min } w = f(x,y,z) \\ \text{u.c. : } \begin{cases} g(x,y,z) = 0 \\ h(x,y,z) = 0 \end{cases} \end{cases}$, that has not one but two constraints, in three variables x, y and z.
Let us take the hypothesis that $\dfrac{\partial(g,h)}{\partial(x,y,z)}$ has rank equal to 2, i.e. maximum, at every point of the set $\mathcal{E}$ in which the two constraints are simultaneously satisfied. From Dini's theorem this guarantees the existence of a function $\mathbb{R} \to \mathbb{R}^2$, for example $x \to (y(x), z(x))$, that generates the following functions composition:

$\mathbb{R} \to \mathbb{R}^3 \xrightarrow{f} \mathbb{R}$, $x \to (x, y(x), z(x)) \to f(x,y,z) = w$, and then we are led to find the extremes of a function of one variable. Suppose, by assumption, that (x_0, y_0, z_0) is a solution, a maximum or a minimum point, to the problem. It must be:

$\frac{\mathrm{d}w}{\mathrm{d}x} = \frac{\partial f}{\partial x} \cdot \frac{\mathrm{d}x}{\mathrm{d}x} + \frac{\partial f}{\partial y} \cdot \frac{\mathrm{d}y}{\mathrm{d}x} + \frac{\partial f}{\partial z} \cdot \frac{\mathrm{d}z}{\mathrm{d}x} = 0$, or $f'_x + f'_y \cdot y'(x) + f'_z \cdot z'(x) = 0$.

From the system $\begin{cases} g(x,y,z) = 0 \\ h(x,y,z) = 0 \end{cases}$, from Dini's theorem we obtain, since by hypothesis $\begin{vmatrix} g'_y & g'_z \\ h'_y & h'_z \end{vmatrix} \neq 0$:

$y'(x) = -\dfrac{\begin{vmatrix} g'_x & g'_z \\ h'_x & h'_z \end{vmatrix}}{\begin{vmatrix} g'_y & g'_z \\ h'_y & h'_z \end{vmatrix}}$ and $z'(x) = -\dfrac{\begin{vmatrix} g'_y & g'_x \\ h'_y & h'_x \end{vmatrix}}{\begin{vmatrix} g'_y & g'_z \\ h'_y & h'_z \end{vmatrix}}$, from which, replacing, and changing the order of the columns of the numerator of $z'(x)$, we get, computing common denominator:

$f'_x \cdot \begin{vmatrix} g'_y & g'_z \\ h'_y & h'_z \end{vmatrix} - f'_y \cdot \begin{vmatrix} g'_x & g'_z \\ h'_x & h'_z \end{vmatrix} + f'_z \cdot \begin{vmatrix} g'_x & g'_y \\ h'_x & h'_y \end{vmatrix} = 0$, or

$\begin{vmatrix} f'_x & f'_y & f'_z \\ g'_x & g'_y & g'_z \\ h'_x & h'_y & h'_z \end{vmatrix} = \begin{vmatrix} \nabla f \\ \nabla g \\ \nabla h \end{vmatrix} = 0$.

The three rows, or gradients, are then linearly dependent, and then we have:

$\nabla f = \lambda_1 \nabla g + \lambda_2 \nabla h$, or $\nabla(f - \lambda_1 g - \lambda_2 h) = 0$.

From a practical point of view increasing the constraints implies that the Lagrangian function is expressed as a difference between the objective function and a linear combination of the constraints, each one with its multiplier.

Increasing the number of the constraints does not lead to changes from a practical point of view: at a constrained maximum or minimum point it is still necessary to cancel the gradient of the Lagrangian function, but now we have the system: $\begin{cases} \Lambda'_x = f'_x - \lambda_1 g'_x - \lambda_2 h'_x = 0 \\ \Lambda'_y = f'_y - \lambda_1 g'_y - \lambda_2 h'_y = 0 \\ \Lambda'_z = f'_z - \lambda_1 g'_z - \lambda_2 h'_z = 0 \\ \Lambda'_{\lambda_1} = -g(x,y,z) = 0 \\ \Lambda'_{\lambda_2} = -h(x,y,z) = 0 \end{cases}$, of five equations in five variables x, y, z, λ_1 and λ_2.

Using these three introductory problems, let us formulate first order conditions in the general case of a maxima or minima problem with equality constraints.

Let $\mathcal{E}$ be the set of points that satisfy the system $\begin{cases} g_1(x_1, x_2, ..., x_n) = 0 \\ g_2(x_1, x_2, ..., x_n) = 0 \\ \dots\dots\dots\dots\dots\dots \\ g_m(x_1, x_2, ..., x_n) = 0 \end{cases}$.

We have the following:

Theorem 28 : Let f and g_i, $1 \leq i \leq m < n$, $\mathbb{R}^n \to \mathbb{R}$, be differentiable functions throughout $\mathcal{E} \subset \mathbb{R}^n$, with $\text{Rank}\left(\dfrac{\partial(g_1, g_2, ..., g_m)}{\partial(x_1, x_2, ..., x_n)}\right) = m$, and $(x_1^0, x_2^0, ..., x_n^0)$ is a solution of the problem $\begin{cases} \text{Max/min } f(x_1, x_2, ..., x_n) \\ \text{u.c.} \begin{cases} g_1(x_1, x_2, ..., x_n) = 0 \\ g_2(x_1, x_2, ..., x_n) = 0 \\ \dots\dots\dots\dots\dots \\ g_m(x_1, x_2, ..., x_n) = 0 \end{cases} \end{cases}$.

Then, if $\Lambda(x_1, x_2, ..., x_n, \lambda_1, \lambda_2, ..., \lambda_m) = f(x_1, x_2, ..., x_n) - \sum_{i=1}^{m} \lambda_i \cdot g_i(x_1, x_2, ..., x_n)$, there exists a vector $(\lambda_1^0, \lambda_2^0, ..., \lambda_m^0)$ for which $\nabla\Lambda(x_1^0, x_2^0, ..., x_n^0, \lambda_1^0, \lambda_2^0, ..., \lambda_m^0) = \mathbb{O}$.
So the point $(x_1^0, x_2^0, ..., x_n^0, \lambda_1^0, \lambda_2^0, ..., \lambda_m^0) = (\mathbb{X}_0, \lambda_0)$ must be a solution of the system of $n + m$ equations in $n + m$ unknowns: $\begin{cases} \dfrac{\partial\Lambda}{\partial x_i} = \dfrac{\partial f}{\partial x_i} - \sum_{j=1}^{m} \lambda_j \cdot \dfrac{\partial g_j}{\partial x_i} = 0 \quad 1 \leq i \leq n \\ \dfrac{\partial\Lambda}{\partial\lambda_i} = -g_i(x_1, x_2, ..., x_n) = 0 \quad 1 \leq i \leq m \end{cases}$.

It's worth repeating that these first order conditions, like $\nabla f = \mathbb{O}$ for free maxima and minima, are necessary, and not sufficient to ensure the nature of extremes. Among stationary points for the Lagrangian function there are, besides possible maxima and minima, also inflection or saddle points. We have inflection points when $m = n - 1$, since the constraints allow us to express $n - 1$ variables as functions of the only remaining one, and then the problem consists in finding the extremes for a function of one variable; we have instead saddle points if $n - m > 1$, as the variables that remain independent are more than 1.

EXTREMES WITH EQUALITY CONSTRAINTS - SECOND ORDER CONDITIONS

Now we treat the simplest example for second order conditions for maxima and minima subject to equality constraints; these will be sufficient and not necessary conditions. We shall go again over the simplest cases to be able to justify (not to prove) the formulation of the conditions in the general case.

Let us recall the problem $\begin{cases} \text{Max/min } z = f(x, y) \\ \text{u.c. : } g(x, y) = 0 \end{cases}$. Let's see how to study the sign of the second order total differential with the presence of a constraint.
As $\mathrm{d}z = f'_x\,\mathrm{d}x + f'_y\,\mathrm{d}y$, and if $y = y(x)$ from $g(x, y) = 0$, the differential $\mathrm{d}y$ depends on x and y, so:

$$\mathrm{d}^2z = \left(f''_{xx}\,\mathrm{d}x + f''_{yx}\,\mathrm{d}y + \frac{\partial(\mathrm{d}y)}{\partial x} \cdot f'_y\right) \cdot \mathrm{d}x + \left(f''_{yx}\,\mathrm{d}x + f''_{yy}\,\mathrm{d}y + \frac{\partial(\mathrm{d}y)}{\partial y} \cdot f'_y\right) \cdot \mathrm{d}y =$$

$$\mathrm{d}^2z = f''_{xx}\,(\mathrm{d}x)^2 + 2f''_{xy}\,\mathrm{d}x\,\mathrm{d}y + f''_{yy}\,(\mathrm{d}y)^2 + f'_y \cdot \left(\frac{\partial(\mathrm{d}y)}{\partial x} \cdot \mathrm{d}x + \frac{\partial(\mathrm{d}y)}{\partial y} \cdot \mathrm{d}y\right) =$$

$$\mathrm{d}^2z = \mathrm{d}^2f + f'_y \cdot \mathrm{d}(\mathrm{d}y) = \mathrm{d}^2f + f'_y \cdot \mathrm{d}^2y.$$

From $z = g(x, y) = 0$, if g is twice differentiable, similarly we get:
$\mathrm{d}^2z = \mathrm{d}^2g + g'_y \cdot \mathrm{d}^2y = 0$, as z is constant at the constraint points, hence we get:

$$\mathrm{d}^2y = -\frac{1}{g'_y} \cdot \mathrm{d}^2g.$$

As $\nabla f = \lambda\nabla g$, from which $f'_y = \lambda\,g'_y$ and so $\dfrac{f'_y}{g'_y} = \lambda$, substituting we obtain:

$$\mathrm{d}^2 z = \mathrm{d}^2 f + f'_y \cdot \mathrm{d}^2 y = \mathrm{d}^2 f - \frac{f'_y}{g'_y} \cdot \mathrm{d}^2 g = \mathrm{d}^2 f - \lambda \cdot \mathrm{d}^2 g = \mathrm{d}^2 (f - \lambda \cdot g) = \mathrm{d}^2 \Lambda\,.$$

Just as first order conditions can be expressed by cancelling a gradient (of Λ, not of f), also second order conditions can be expressed referring to the sign of a second order total differential, (once again of Lagrangian Λ, not of f).
However, not everything is equal to methods already described for free maxima and minima. In fact:
$\mathrm{d}^2\Lambda(x,y) = \Lambda''_{xx}\,(\mathrm{d}x)^2 + 2\,\Lambda''_{xy}\,\mathrm{d}x\,\mathrm{d}y + \Lambda''_{yy}\,(\mathrm{d}y)^2\,.$
From $g(x,y) = 0$ we get $\mathrm{d}y = -\dfrac{g'_x}{g'_y} \cdot \mathrm{d}x$, from which, substituting, we have:

$$\mathrm{d}^2\Lambda(x,y) = \Lambda''_{xx}\,(\mathrm{d}x)^2 - 2\,\Lambda''_{xy}\,\frac{g'_x}{g'_y} \cdot (\mathrm{d}x)^2 + \Lambda''_{yy}\left(\frac{g'_x}{g'_y}\right)^2 \cdot (\mathrm{d}x)^2 =$$

$$\mathrm{d}^2\Lambda(x,y) = \left[\Lambda''_{xx}\,(g'_y)^2 - 2\,\Lambda''_{xy}\,g'_x\,g'_y + \Lambda''_{yy}\,(g'_x)^2\right] \cdot \frac{(\mathrm{d}x)^2}{(g'_y)^2}\,.$$

But it is easily seen that:

$$\Lambda''_{xx}\,(g'_y)^2 - 2\,\Lambda''_{xy}\,g'_x\,g'_y + \Lambda''_{yy}\,(g'_x)^2 = -\begin{vmatrix} 0 & g'_x & g'_y \\ g'_x & \Lambda''_{xx} & \Lambda''_{xy} \\ g'_y & \Lambda''_{xy} & \Lambda''_{yy} \end{vmatrix} \text{ and also:}$$

$$\begin{vmatrix} 0 & g'_x & g'_y \\ g'_x & \Lambda''_{xx} & \Lambda''_{xy} \\ g'_y & \Lambda''_{xy} & \Lambda''_{yy} \end{vmatrix} = \begin{vmatrix} \Lambda''_{\lambda\lambda} & \Lambda''_{\lambda x} & \Lambda''_{\lambda y} \\ \Lambda''_{\lambda x} & \Lambda''_{xx} & \Lambda''_{xy} \\ \Lambda''_{\lambda y} & \Lambda''_{xy} & \Lambda''_{yy} \end{vmatrix} = \left|\overline{\mathbb{H}}(\Lambda(x,y,\lambda))\right|.$$

Matrix $\overline{\mathbb{H}}$ is called "Bordered Hessian matrix"; there is one border row at the top and one border column at the left, which, excluding the initial zero, are the gradient of the constraint $g(x,y)$.
The zero in the north-west corner corresponds to $\Lambda''_{\lambda\lambda}$, while remaining elements of the border are the opposite of the second order derivatives of the Lagrangian made with respect to λ and then with respect to x or y; for the determinant properties, changing the sign of two lines, the determinant remains unchanged.
The sign of the differential $\mathrm{d}^2\Lambda$ is not to be studied on varying two independent increments $\mathrm{d}x$ and $\mathrm{d}y$; replacing $\mathrm{d}y = -\dfrac{g'_x}{g'_y} \cdot \mathrm{d}x$ only $\mathrm{d}x$ remains independent, consistently with the fact that the presence of a constraint leaves only one independent variable among x and y.
This fact can be geometrically interpreted saying that we should investigate $\mathrm{d}^2\Lambda$ using increments lying only on directions tangential to the constraint.

As $\mathrm{d}^2\Lambda = -\left|\overline{\mathbb{H}}(\Lambda(x,y,\lambda))\right|$, we have the following:
Theorem 29 : (x_0, y_0, λ_0) is a solution of the system $\nabla\Lambda(x,y,\lambda) = 0$. Then
- $\left(\left|\overline{\mathbb{H}}(\Lambda(x_0,y_0,\lambda_0))\right| < 0 \Leftrightarrow \mathrm{d}^2\Lambda > 0\right) \Rightarrow (x_0,y_0)$ is a constrained minimum point;
- $\left(\left|\overline{\mathbb{H}}(\Lambda(x_0,y_0,\lambda_0))\right| > 0 \Leftrightarrow \mathrm{d}^2\Lambda < 0\right) \Rightarrow (x_0,y_0)$ is a constrained maximum point.

Nothing can be concluded if it is $\left|\overline{\mathbb{H}}(\Lambda(x_0,y_0,\lambda_0))\right| = 0$, as (x_0,y_0) may be a maximum, minimum or inflection point.
We need in this case a different type of analysis to determine the nature of the point.

Observation 1) Although the bordered Hessian is a third order matrix, we must take into account only the leading minor of third order, i.e. the determinant of the matrix itself; it is not necessary to examine the leading minor of first order, which is always zero, and is not necessary to examine the leading minor of the second order, as $\begin{vmatrix} 0 & g'_x \\ g'_x & \Lambda''_{xx} \end{vmatrix} = -\left(g'_x\right)^2 < 0$, therefore always negative.

Observation 2) Only North-West leading minors are to be examined ; there are no conditions that can be expressed through other sequences of leading minors and there are no conditions based on the eigenvalues of the Hessian bordered matrix.

If we study now the second-order conditions for the problem $\begin{cases} \text{Max/min } f(x,y,z) \\ \text{u.c. : } g(x,y,z) = 0 \end{cases}$ we will still be brought to the study of $\mathrm{d}^2\Lambda$. From the constraint $g(x,y,z) = 0$ we can, by hypothesis, get $z = z(x,y)$; so we are looking for the extremes of a two variables function, and the study of the sign of $\mathrm{d}^2\Lambda$ will depend on $\mathrm{d}x$ and $\mathrm{d}y$.
The developments of the calculus, which are omitted for brevity, leads to study the sign of a quantity whose opposite can be related to the bordered Hessian matrix:

$$\left\| \begin{matrix} 0 & g'_x & g'_y & g'_z \\ g'_x & \Lambda''_{xx} & \Lambda''_{xy} & \Lambda''_{xz} \\ g'_y & \Lambda''_{xy} & \Lambda''_{yy} & \Lambda''_{yz} \\ g'_z & \Lambda''_{xz} & \Lambda''_{yz} & \Lambda''_{zz} \end{matrix} \right\| = \overline{\mathbb{H}}(\Lambda(x,y,z,\lambda)) .$$

Similarly to what we saw in the case of free extremes for a two variables function, we must study the sign of two North-West leading minors.
As before, first order and second order leading minors are useless. We must study only the sign of the leading minors $\left|\overline{\mathbb{H}}_3\right|$ and $\left|\overline{\mathbb{H}}_4\right|$, and we have the following:
Theorem 30 : $(x_0, y_0, z_0, \lambda_0)$ is a solution of the system $\nabla\Lambda(x,y,z,\lambda) = 0$. Then
- $\left(\left(\left|\overline{\mathbb{H}}_3\right| < 0 \text{ e } \left|\overline{\mathbb{H}}_4\right| < 0\right) \Leftrightarrow \mathrm{d}^2\Lambda > 0\right) \Rightarrow (x_0, y_0, z_0)$ is a constrained minimum point;
- $\left(\left(\left|\overline{\mathbb{H}}_3\right| > 0 \text{ e } \left|\overline{\mathbb{H}}_4\right| < 0\right) \Leftrightarrow \mathrm{d}^2\Lambda < 0\right) \Rightarrow (x_0, y_0, z_0)$ is a constrained maximum point.

If $\left(\left|\overline{\mathbb{H}}_3\right| < 0 \text{ and } \left|\overline{\mathbb{H}}_4\right| > 0\right)$ or if $\left(\left|\overline{\mathbb{H}}_3\right| > 0 \text{ and } \left|\overline{\mathbb{H}}_4\right| > 0\right)$ surely the point is a saddle point.
Nothing can be concluded if it is $\left|\overline{\mathbb{H}}_3\right| = 0$ or $\left|\overline{\mathbb{H}}_4\right| = 0$, still respecting however the previous sequences of signs, as (x_0, y_0, z_0) may be a maximum, a minimum or a saddle point.
We need in this case a different type of analysis to determine the nature of the point.

Finally, for the problem: $\begin{cases} \text{Max/min } f(x,y,z) \\ \text{u.c. : } \begin{cases} g(x,y,z) = 0 \\ h(x,y,z) = 0 \end{cases} \end{cases}$ we will still be brought to the study of $\mathrm{d}^2\Lambda$, but from the constraints $\begin{cases} g(x,y,z) = 0 \\ h(x,y,z) = 0 \end{cases}$ we can, by hypothesis, obtain $\begin{cases} y = y(x) \\ z = z(x) \end{cases}$; so let us look for the extremes of a one variable function, and the study of the sign of $\mathrm{d}^2\Lambda$ will be based only on $\mathrm{d}x$.
With two (even number) constraints, there are no sign changes, so the sign of $\mathrm{d}^2\Lambda$ corresponds to that of the determinant of the bordered Hessian matrix:

$$\left\| \begin{matrix} 0 & 0 & g'_x & g'_y & g'_z \\ 0 & 0 & h'_x & h'_y & h'_z \\ g'_x & h'_x & \Lambda''_{xx} & \Lambda''_{xy} & \Lambda''_{xz} \\ g'_y & h'_y & \Lambda''_{xy} & \Lambda''_{yy} & \Lambda''_{yz} \\ g'_z & h'_z & \Lambda''_{xz} & \Lambda''_{yz} & \Lambda''_{zz} \end{matrix} \right\| = \overline{\mathbb{H}}(\Lambda(x,y,z,\lambda_1,\lambda_2)) .$$

The $(2,2)$ matrix in the North-West corner is null because $\Lambda = f(x,y,z) - \lambda_1\, g(x,y,z) - \lambda_2\, h(x,y,z)$, and so the four second order derivatives of the Lagrangian function with respect to the multipliers λ_1 and λ_2 are also null. We can now see that in the leading minors sequence (always and only North-West se-

quence) the first four (four is twice two, the number of constraints) are zero $\left(\left|\overline{\mathbb{H}}_1\right|=0, \left|\overline{\mathbb{H}}_2\right|=0, \left|\overline{\mathbb{H}}_3\right|=0\right)$ or have constant sign $\left(\left|\overline{\mathbb{H}}_4\right|<0\right)$, wherefore $\mathrm{d}^2\Lambda$ sign coincides with that of $\left|\overline{\mathbb{H}}_5\right|$, determinant of the bordered Hessian matrix.
We have the following:
Theorem 31 : $(x_0, y_0, z_0, \lambda_1^0, \lambda_2^0)$ is a solution of the system $\nabla\Lambda(x,y,z,\lambda_1,\lambda_2)=0$. Then
- $\left(\left|\overline{\mathbb{H}}_5\left(\Lambda\left(x_0,y_0,z_0,\lambda_1^0,\lambda_2^0\right)\right)\right|>0 \Leftrightarrow \mathrm{d}^2\Lambda>0\right) \Rightarrow (x_0,y_0,z_0)$ is a constrained minimum point;
- $\left(\left|\overline{\mathbb{H}}_5\left(\Lambda\left(x_0,y_0,z_0,\lambda_1^0,\lambda_2^0\right)\right)\right|<0 \Leftrightarrow \mathrm{d}^2\Lambda<0\right) \Rightarrow (x_0,y_0,z_0)$ is a constrained maximum point.

Nothing can be concluded if it is $\left|\overline{\mathbb{H}}_5\left(\Lambda\left(x_0,y_0,z_0,\lambda_1^0,\lambda_2^0\right)\right)\right|=0$, as (x_0,y_0,z_0) may be maximum, minimum or inflection point.
In this case too we need a different type of analysis to determine the nature of the point.

From what we saw in the three treated examples, the following considerations arise:
- we need to build the bordered Hessian matrix, consisting of the second order derivatives of the Lagrangian function made with respect both to variables and multipliers;
- we only need to study the signs of the North-West leading minors, subtracting an initial number of such minors equal to twice the number of the constraints; the number of leading minors whose sign is relevant is equal to the number of variables which remain independent;
- for the remaining leading minors, in order to find a minimum or a maximum point, two sequences of signs are valid, wich are the same as for free extremes if the number of constraints is even; have opposite signs than the previous sequences for an odd number of constraints;
- each sequence different from the two described, even if for only one sign, leads to the conclusion that the point is an inflection or saddle point;
- if between the relevant leading minors there is at least one equal to zero, nothing can be concluded about the nature of the point.

So let's look at the general case, i.e. the problem $\begin{cases} \text{Max/min } f(x_1,x_2,\dots,x_n) \\ \text{u.c.} \begin{cases} g_1(x_1,x_2,\dots,x_n)=0 \\ g_2(x_1,x_2,\dots,x_n)=0 \\ \dots\dots\dots\dots\dots\dots \\ g_m(x_1,x_2,\dots,x_n)=0 \end{cases} \end{cases}$.

The bordered Hessian matrix is a square matrix of order $m+n$, and is equal to:

$$\overline{\mathbb{H}}(\Lambda(x_1,x_2,\dots,x_n,\lambda_1,\lambda_2,\dots,\lambda_m)) = \left\|\begin{matrix} 0 & \dots & 0 & g'_{11} & \dots & g'_{1n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & g'_{m1} & \dots & g'_{mn} \\ g'_{11} & \dots & g'_{m1} & \Lambda''_{11} & \dots & \Lambda''_{1n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ g'_{1n} & \dots & g'_{mn} & \Lambda''_{n1} & \dots & \Lambda''_{nn} \end{matrix}\right\|,$$

where, for brevity, we settle $g'_{ij}=\dfrac{\partial g_i}{\partial x_j}$ and $\Lambda''_{ij}=\dfrac{\partial^2\Lambda}{\partial x_i\,\partial x_j}$, and can be so represented:

$$\overline{\mathbb{H}}(\Lambda(x_1,x_2,\dots,x_n,\lambda_1,\lambda_2,\dots,\lambda_m)) = \left\|\begin{matrix} \mathbb{O} & \dfrac{\partial(g_1,g_2,\dots,g_m)}{\partial(x_1,x_2,\dots,x_n)} \\ \left(\dfrac{\partial(g_1,g_2,\dots,g_m)}{\partial(x_1,x_2,\dots,x_n)}\right)^{\mathrm{T}} & \mathbb{H}(\Lambda(x_1,x_2,\dots,x_n)) \end{matrix}\right\|.$$

As we see, bordered Hessian matrix can be split into four blocks.
The block in the upper left is an $m\times m$ null matrix, consisting of the second order derivatives of the Lagrangian function done both times with respect to the multipliers, and hence equal to zero.

In the upper right and lower left block there are the derivatives of the Lagrangian obtained deriving once with respect to a multiplier and once with respect to a variable; these also form the border, which is completed so by the Jacobian of the constraints, an $m \times n$ matrix, at the top as rows and on the left, transposed, as columns.
The remaining $n \times n$ block, in the lower right, is the Hessian of the Lagrangian done with respect to the variables x_i.
Then the following applies:
Theorem 32 : $(x_1^0, x_2^0, ..., x_n^0, \lambda_1^0, \lambda_2^0, ..., \lambda_m^0)$ is a solution of the system:
$\nabla\Lambda(x_1, x_2, ..., x_n, \lambda_1, \lambda_2, ..., \lambda_m) = 0$. Then
- $\left(\begin{cases} (-1)^m \cdot \left|\overline{\mathbb{H}}_i(\Lambda)\right| > 0 \\ 2m+1 \le i \le m+n \end{cases} \Leftrightarrow \mathrm{d}^2\Lambda > 0\right) \Rightarrow (x_1^0, x_2^0, ..., x_n^0)$ is a constrained minimum point;
- $\left(\begin{cases} (-1)^{m+i} \cdot \left|\overline{\mathbb{H}}_i(\Lambda)\right| > 0 \\ 2m+1 \le i \le m+n \end{cases} \Leftrightarrow \mathrm{d}^2\Lambda < 0\right) \Rightarrow (x_1^0, x_2^0, ..., x_n^0)$ is a constrained maximum point.
Nothing can be concluded if some relevant leading minor is equal to zero.

If the constraints are in an even number (and also zero), an all positive signs sequence indicates a minimum point, while an alternating signs sequence, starting from the negative, indicates a maximum point.
If the constraints are in an odd number, an all negative signs sequence indicates a minimum point, while an alternating signs sequence, starting from positive, indicates a maximum point.
Each sequence that does not meet one of those described indicates an inflection or a saddle point.
The requirement $2m+1 \le i \le m+n$ means that only North-West relevant leading minors should be considered, obtained discarding the first $2m$ leading minors (a number equal to twice the number of constraints) so starting from the $2m+1$ order leading minor.
To change or not to change the signs of the sequence of the relevant leading minors is the role of the factor $(-1)^m$.

As $m < n$, we have $(m+n) - 2m = n - m$ relevant leading minors; if $m = 1$, or if there is only one constraint, we must consider the sign of $n-1$ relevant leading minors; if instead $m = n-1$, which is the maximum possible number of constraints, we will consider the sign of only one relevant leading minor, which is the determinant of the bordered Hessian matrix.
Since each constraint, explicitly or implicitly, makes a variable a dependent one, the number of relevant leading minors the sign of which must be studied coincides always with the number of variables that remain independent.

It should be finally pointed out that there are no second order conditions based on eigenvalues of the bordered Hessian matrix.

Example 75 : We study the problem $\begin{cases} \text{Max/min } f(x,y) = 2x^3 + 3y \\ \text{u.c.} : g(x,y) = x^2 + y^2 = 1 \end{cases}$.
$f(x,y)$ and $g(x,y)$ are polynomials, and then infinitely differentiable functions. Then $\nabla g(x,y) = (2x, 2y) = \mathbb{O} \Leftrightarrow (x,y) = (0,0)$; but $(0,0)$ does not satisfy the constraint equation, and therefore hypotheses of Theorem 27 are satisfied. We form the Lagrangian and then we have: $\Lambda(x,y,\lambda) = 2x^3 + 3y - \lambda\left(x^2 + y^2 - 1\right)$; imposing $\nabla\Lambda = 0$ we get:

$$\begin{cases} \Lambda'_x = 0 \Rightarrow 6x^2 - 2\lambda x = 2x(3x - \lambda) = 0 \\ \Lambda'_y = 0 \Rightarrow 3 - 2\lambda y = 0 \\ \Lambda'_\lambda = 0 \Rightarrow x^2 + y^2 = 1 \end{cases} \Rightarrow \begin{cases} x = 0 \\ \lambda = \dfrac{3}{2y} \\ y = \pm 1 \end{cases} \Rightarrow \begin{cases} x = 0 \\ \lambda = \frac{3}{2} \\ y = 1 \end{cases} \text{ and } \begin{cases} x = 0 \\ \lambda = -\frac{3}{2} \\ y = -1 \end{cases}$$

or $\begin{cases} x = \dfrac{\lambda}{3} \\ y = \dfrac{3}{2\lambda} \\ \dfrac{\lambda^2}{9} + \dfrac{9}{4\lambda^2} = 1 \end{cases} \Rightarrow \begin{cases} x = \dfrac{\lambda}{3} \\ y = \dfrac{3}{2\lambda} \\ \dfrac{4\lambda^4 + 81 - 36\lambda^2}{36\lambda^2} = 0 \end{cases} \Rightarrow \begin{cases} x = \dfrac{\lambda}{3} \\ y = \dfrac{3}{2\lambda} \\ (2\lambda^2 - 9)^2 = 0 \end{cases}$ so we have:

$\begin{cases} \lambda = \frac{3}{\sqrt{2}} \\ x = \frac{1}{\sqrt{2}} \\ y = \frac{1}{\sqrt{2}} \end{cases}$ e $\begin{cases} \lambda = -\frac{3}{\sqrt{2}} \\ x = -\frac{1}{\sqrt{2}} \\ y = -\frac{1}{\sqrt{2}} \end{cases}$. Solving the system, we have four points: $\mathrm{P}_1 = \left(0, 1, \dfrac{3}{2}\right)$, $\mathrm{P}_2 = \left(0, -1, -\dfrac{3}{2}\right)$, $\mathrm{P}_3 = \left(\dfrac{1}{\sqrt{2}}, \dfrac{1}{\sqrt{2}}, \dfrac{3}{\sqrt{2}}\right)$ and $\mathrm{P}_4 = \left(-\dfrac{1}{\sqrt{2}}, -\dfrac{1}{\sqrt{2}}, -\dfrac{3}{\sqrt{2}}\right)$.

We apply the second-order conditions. We form the bordered Hessian to get:

$\overline{\mathbb{H}}(\Lambda(x,y,\lambda)) = \left\|\begin{matrix} 0 & 2x & 2y \\ 2x & 12x - 2\lambda & 0 \\ 2y & 0 & -2\lambda \end{matrix}\right\|$ so we have:

$\left|\overline{\mathbb{H}}(\Lambda(\mathrm{P}_1))\right| = \begin{vmatrix} 0 & 0 & 2 \\ 0 & -3 & 0 \\ 2 & 0 & -3 \end{vmatrix} = 2 \cdot 6 = 12 > 0$, so P_1 is a maximum point;

$\left|\overline{\mathbb{H}}(\Lambda(\mathrm{P}_2))\right| = \begin{vmatrix} 0 & 0 & -2 \\ 0 & 3 & 0 \\ -2 & 0 & 3 \end{vmatrix} = -2 \cdot 6 = -12 < 0$, so P_2 is a minimum point;

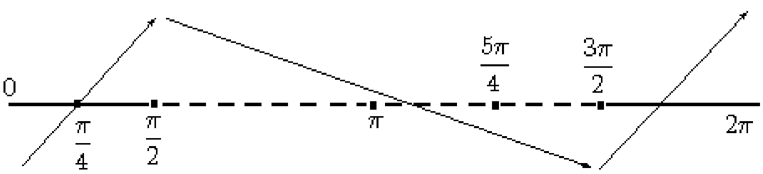
$\left|\overline{\mathbb{H}}(\Lambda(\mathrm{P}_3))\right| = \begin{vmatrix} 0 & \sqrt{2} & \sqrt{2} \\ \sqrt{2} & 3\sqrt{2} & 0 \\ \sqrt{2} & 0 & -3\sqrt{2} \end{vmatrix} = \begin{vmatrix} 0 & \sqrt{2} & \sqrt{2} \\ 0 & 3\sqrt{2} & 3\sqrt{2} \\ \sqrt{2} & 0 & -3\sqrt{2} \end{vmatrix} = 0$, so we cannot decide anything;

$\left|\overline{\mathbb{H}}(\Lambda(\mathrm{P}_4))\right| = \begin{vmatrix} 0 & -\sqrt{2} & -\sqrt{2} \\ -\sqrt{2} & -3\sqrt{2} & 0 \\ -\sqrt{2} & 0 & 3\sqrt{2} \end{vmatrix} = \begin{vmatrix} 0 & -\sqrt{2} & -\sqrt{2} \\ 0 & -3\sqrt{2} & -3\sqrt{2} \\ -\sqrt{2} & 0 & 3\sqrt{2} \end{vmatrix} = 0$, so we cannot decide anything.

Since the constraint is the trigonometric circle, to solve the two remaining cases we try replacing $\begin{cases} x = \cos t \\ y = \sin t \end{cases}$ to get: $f(t) = 2\cos^3 t + 3\sin t$. Is:

$f'(t) = 6\cos^2 t \cdot (-\sin t) + 3\cos t = 3\cos t \cdot (1 - 2\sin t\cos t) = 3\cos t \cdot (1 - \sin 2t) \geq 0$ if $\cos t \geq 0$ as $\sin 2t \leq 1$, $\forall t$.

So $f'(t) \geq 0$ if $0 \leq t \leq \dfrac{\pi}{2}$ and if $\dfrac{3\pi}{2} \leq t \leq 2\pi$.



Point $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ corresponds to $t = \frac{\pi}{4}$, while point $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$ corresponds to $t = \frac{5\pi}{4}$ and, as we see from the monotonicity analysis, they are both points of inflection and no maximum or minimum points.

Since the circle points form a bounded and closed set, and since the function $f(x, y)$ is continuous, points P_1 and P_2 are not only relative (local) but also absolute maximum and minimum points.

Example 76 : We study the problem $\begin{cases} \text{Max/min } f(x,y,z) = x + y - 2z \\ \text{u.c. : } g(x,y,z) = x^2 + y^2 - z = 0 \end{cases}$.

$f(x,y,z)$ and $g(x,y,z)$ are polynomials, and therefore infinitely differentiable functions. Then $\nabla g(x,y,z) = (2x, 2y, -1) \neq \mathbb{O} \ \ \forall (x,y,z)$; and therefore hypotheses of Theorem 27 are satisfied. We form the Lagrangian and then we have:

$\Lambda(x,y,z,\lambda) = x + y - 2z - \lambda\left(x^2 + y^2 - z\right)$; imposing $\nabla\Lambda = 0$ we get:

$\begin{cases} \Lambda'_x = 0 \Rightarrow 1 - 2\lambda x = 0 \\ \Lambda'_y = 0 \Rightarrow 1 - 2\lambda y = 0 \\ \Lambda'_z = 0 \Rightarrow -2 + \lambda = 0 \\ \Lambda'_\lambda = 0 \Rightarrow z - x^2 - y^2 = 0 \end{cases} \Rightarrow \begin{cases} x = \frac{1}{2\lambda} = \frac{1}{4} \\ y = \frac{1}{2\lambda} = \frac{1}{4} \\ \lambda = 2 \\ z = x^2 + y^2 = \frac{1}{8} \end{cases}$. So there is only one stationary point for the Lagrangian: $P_0 = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{8}, 2\right)$. Then:

$$\overline{\mathbb{H}}(\Lambda) = \left\|\begin{matrix} 0 & 2x & 2y & -1 \\ 2x & -2\lambda & 0 & 0 \\ 2y & 0 & -2\lambda & 0 \\ -1 & 0 & 0 & 0 \end{matrix}\right\| \text{ and } \overline{\mathbb{H}}(\Lambda(P_0)) = \left\|\begin{matrix} 0 & \frac{1}{2} & \frac{1}{2} & -1 \\ \frac{1}{2} & -4 & 0 & 0 \\ \frac{1}{2} & 0 & -4 & 0 \\ -1 & 0 & 0 & 0 \end{matrix}\right\|.$$

We need to compute two leading minors: $|\overline{\mathbb{H}}_3|$ and $|\overline{\mathbb{H}}_4|$, and so we get:

$$\left|\overline{\mathbb{H}}_3(\Lambda(P_0))\right| = \begin{vmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -4 & 0 \\ \frac{1}{2} & 0 & -4 \end{vmatrix} = \begin{vmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & -4 & 4 \\ \frac{1}{2} & 0 & -4 \end{vmatrix} = \frac{1}{2} \cdot (2 + 2) = 2 > 0, \text{ and}$$

$$\left|\overline{\mathbb{H}}_4(\Lambda(P_0))\right| = \begin{vmatrix} 0 & \frac{1}{2} & \frac{1}{2} & -1 \\ \frac{1}{2} & -4 & 0 & 0 \\ \frac{1}{2} & 0 & -4 & 0 \\ -1 & 0 & 0 & 0 \end{vmatrix} = 1 \cdot \begin{vmatrix} \frac{1}{2} & \frac{1}{2} & -1 \\ -4 & 0 & 0 \\ 0 & -4 & 0 \end{vmatrix} = -16 < 0.$$

As $\left|\overline{\mathbb{H}}_3(\Lambda(P_0))\right| > 0$ and $\left|\overline{\mathbb{H}}_4(\Lambda(P_0))\right| < 0$ it follows that P_0 is a maximum point.

We could also get the same result in a faster way, as the constraint allows us, using the equation $x^2 + y^2 - z = 0$, to solve z as $z = x^2 + y^2$, so, substituting, we obtain: $f\left(x, y, x^2 + y^2\right) = x + y - 2x^2 - 2y^2$. We can then look for free maxima and minima of this two variables function. The constraint is satisfied as contained in the replacement.

We have so: $\begin{cases} f'_x = 1 - 4x = 0 \\ f'_y = 1 - 4y = 0 \end{cases} \Rightarrow \begin{cases} x = \frac{1}{4} \\ y = \frac{1}{4} \end{cases} \Rightarrow z = x^2 + y^2 = \frac{1}{8}$. Then:

$\mathbb{H}(f) = \begin{Vmatrix} -4 & 0 \\ 0 & -4 \end{Vmatrix} = \mathbb{H}(f(\mathrm{P}))$ from which $|\mathbb{H}_1| = -4 < 0$ e $|\mathbb{H}_2| = 16 > 0$, so $\mathrm{P} = \left(\frac{1}{4}, \frac{1}{4}\right)$ is a maximum point for $f(x, y, x^2 + y^2)$ and then $\mathrm{P}_0 = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{8}\right)$ is a constrained maximum point for $f(x, y, z)$.

Example 77 : We study the problem $\begin{cases} \text{Max/min } f(x, y, z) = xy - z \\ \text{u.c. : } \begin{cases} g(x, y, z) = x^2 + y^2 = 1 \\ h(x, y, z) = 2x - 2y - z = 0 \end{cases} \end{cases}$.

We solve this problem in three different ways, achieving, of course, the same results. We begin using Lagrange's multipliers method, constructing the Lagrangian function after checking that all assumptions are satisfied.

Objective function and constraints are differentiable functions.

Then $\dfrac{\partial(g, h)}{\partial(x, y, z)} = \begin{Vmatrix} 2x & 2y & 0 \\ 2 & -2 & -1 \end{Vmatrix}$. The Jacobian rank is equal to 1 only if $x = y = 0$, but this point does not satisfy the first constraint, and then all the hypotheses are satisfied.

It is $\Lambda(x, y, z, \lambda_1, \lambda_2) = xy - z - \lambda_1 \left(x^2 + y^2 - 1\right) - \lambda_2 (2x - 2y - z)$.

Imposing $\nabla\Lambda = 0$ we get: $\begin{cases} \Lambda'_x = 0 \Rightarrow y - 2\lambda_1 x - 2\lambda_2 = 0 \\ \Lambda'_y = 0 \Rightarrow x - 2\lambda_1 y + 2\lambda_2 = 0 \\ \Lambda'_z = 0 \Rightarrow -1 + \lambda_2 = 0 \\ \Lambda'_{\lambda_1} = 0 \Rightarrow x^2 + y^2 = 1 \\ \Lambda'_{\lambda_2} = 0 \Rightarrow 2x - 2y - z = 0 \end{cases}$.

Summing first and second equation we get the system: $\begin{cases} (y + x)(1 - 2\lambda_1) = 0 \\ x - 2\lambda_1 y + 2\lambda_2 = 0 \\ \lambda_2 = 1 \\ x^2 + y^2 = 1 \\ 2x - 2y - z = 0 \end{cases}$ and then the two systems: $\begin{cases} (y + x) = 0 \\ x - 2\lambda_1 y + 2\lambda_2 = 0 \\ \lambda_2 = 1 \\ x^2 + y^2 = 1 \\ 2x - 2y - z = 0 \end{cases}$ and $\begin{cases} 1 - 2\lambda_1 = 0 \\ x - 2\lambda_1 y + 2\lambda_2 = 0 \\ \lambda_2 = 1 \\ x^2 + y^2 = 1 \\ 2x - 2y - z = 0 \end{cases}$.

For the first we have:

$\begin{cases} x = -y \\ \lambda_1 = \dfrac{2 - y}{2y} \\ \lambda_2 = 1 \\ y^2 = \frac{1}{2} \\ z = -4y \end{cases} \Rightarrow \begin{cases} x = -\frac{1}{\sqrt{2}} \\ \lambda_1 = \sqrt{2} - \frac{1}{2} \\ \lambda_2 = 1 \\ y = \frac{1}{\sqrt{2}} \\ z = -2\sqrt{2} \end{cases}$ and $\begin{cases} x = \frac{1}{\sqrt{2}} \\ \lambda_1 = -\sqrt{2} - \frac{1}{2} \\ \lambda_2 = 1 \\ y = -\frac{1}{\sqrt{2}} \\ z = 2\sqrt{2} \end{cases}$. For the second:

$\begin{cases} \lambda_1 = \frac{1}{2} \\ x - y + 2 = 0 \\ \lambda_2 = 1 \\ x^2 + y^2 = 1 \\ z = 2x - 2y \end{cases} \Rightarrow \begin{cases} \lambda_1 = \frac{1}{2} \\ y = x + 2 \\ \lambda_2 = 1 \\ x^2 + (x + 2)^2 = 1 \\ z = 2x - 2y \end{cases}$. The fourth equation, however, has no real solutions. So there are only two stationary points for the Lagrangian:

$\mathrm{P}_1 = \left(-\dfrac{1}{\sqrt{2}}, \dfrac{1}{\sqrt{2}}, -2\sqrt{2}, \sqrt{2} - \dfrac{1}{2}, 1\right)$ and

$$\mathrm{P}_2 = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 2\sqrt{2}, -\sqrt{2}-\frac{1}{2}, 1\right).$$

We construct the bordered Hessian: $\overline{\mathbb{H}}(\Lambda) = \left\|\begin{array}{ccccc} 0 & 0 & 2x & 2y & 0 \\ 0 & 0 & 2 & -2 & -1 \\ 2x & 2 & -2\lambda_1 & 1 & 0 \\ 2y & -2 & 1 & -2\lambda_1 & 0 \\ 0 & -1 & 0 & 0 & 0 \end{array}\right\|$.

Having two constraints, the first four North-West leading minors are useless, and so it is enough to calculate the determinant of the matrix: $\left|\overline{\mathbb{H}}(\Lambda)\right|$. We have:

$$\left|\overline{\mathbb{H}}(\Lambda(\mathrm{P}_1))\right| = \left|\begin{array}{ccccc} 0 & 0 & -\sqrt{2} & \sqrt{2} & 0 \\ 0 & 0 & 2 & -2 & -1 \\ -\sqrt{2} & 2 & 1-2\sqrt{2} & 1 & 0 \\ \sqrt{2} & -2 & 1 & 1-2\sqrt{2} & 0 \\ 0 & -1 & 0 & 0 & 0 \end{array}\right| =$$

$$= 1\cdot(-1)\cdot\left|\begin{array}{ccc} 0 & -\sqrt{2} & \sqrt{2} \\ -\sqrt{2} & 1-2\sqrt{2} & 1 \\ \sqrt{2} & 1 & 1-2\sqrt{2} \end{array}\right| = -\left|\begin{array}{ccc} 0 & -\sqrt{2} & \sqrt{2} \\ 0 & 2-2\sqrt{2} & 2-2\sqrt{2} \\ \sqrt{2} & 1 & 1-2\sqrt{2} \end{array}\right| =$$

$= -\sqrt{2}\cdot\left[-2\sqrt{2}\left(2-2\sqrt{2}\right)\right] = 8\left(1-\sqrt{2}\right) < 0$, so P_1 is a maximum point;

$$\left|\overline{\mathbb{H}}(\Lambda(\mathrm{P}_2))\right| = \left|\begin{array}{ccccc} 0 & 0 & \sqrt{2} & -\sqrt{2} & 0 \\ 0 & 0 & 2 & -2 & -1 \\ \sqrt{2} & 2 & 1+2\sqrt{2} & 1 & 0 \\ -\sqrt{2} & -2 & 1 & 1+2\sqrt{2} & 0 \\ 0 & -1 & 0 & 0 & 0 \end{array}\right| =$$

$$= 1\cdot(-1)\cdot\left|\begin{array}{ccc} 0 & \sqrt{2} & -\sqrt{2} \\ \sqrt{2} & 1+2\sqrt{2} & 1 \\ -\sqrt{2} & 1 & 1+2\sqrt{2} \end{array}\right| = -\left|\begin{array}{ccc} 0 & \sqrt{2} & -\sqrt{2} \\ 0 & 2+2\sqrt{2} & 2+2\sqrt{2} \\ -\sqrt{2} & 1 & 1+2\sqrt{2} \end{array}\right| =$$

$= \sqrt{2}\cdot\left[2\sqrt{2}\left(2+2\sqrt{2}\right)\right] = 8\left(1+\sqrt{2}\right) > 0$, so P_2 is a minimum point.

This problem can be solved in a second way. From the constraint $2x-2y-z=0$ we solve the variable $z : z = 2x-2y$ and then substituting in the expression of the objective function we get:

$f(x,y,z) = f(x,y,2x-2y) = xy-2x+2y$.

We can now solve the problem $\begin{cases} \text{Max/min } f(x,y) = xy-2x+2y \\ \text{s.v.} : g(x,y) = x^2+y^2 = 1 \end{cases}$, which is still the search of the extremes for a function of a single variable.

We form the Lagrangian and we have: $\Lambda(x,y,\lambda) = xy-2x+2y-\lambda\left(x^2+y^2-1\right)$; imposing $\nabla\Lambda = 0$, summing the first with the second equation, we get:

$\begin{cases} \Lambda'_x = 0 \Rightarrow y-2-2\lambda x = 0 \\ \Lambda'_y = 0 \Rightarrow x+2-2\lambda y = 0 \\ \Lambda'_\lambda = 0 \Rightarrow x^2+y^2 = 1 \end{cases} \Rightarrow \begin{cases} (y+x)(1-2\lambda) = 0 \\ x+2-2\lambda y = 0 \\ x^2+y^2 = 1 \end{cases}$ which gives two systems:

$\begin{cases} x = -y \\ \lambda = \dfrac{2-y}{2y} \\ 2y^2 = 1 \end{cases} \Rightarrow \begin{cases} x = -\frac{1}{\sqrt{2}} \\ \lambda = \sqrt{2}-\frac{1}{2} \\ y = \frac{1}{\sqrt{2}} \end{cases}$ and $\begin{cases} x = \frac{1}{\sqrt{2}} \\ \lambda = -\sqrt{2}-\frac{1}{2} \\ y = -\frac{1}{\sqrt{2}} \end{cases}$ or

$\begin{cases} \lambda = \frac{1}{2} \\ x + 2 - y = 0 \\ x^2 + y^2 = 1 \end{cases} \Rightarrow \begin{cases} \lambda = \frac{1}{2} \\ x = y - 2 \\ y^2 + 4 - 4y + y^2 = 1 \end{cases}$ which has no solutions.

So we have two stationary points:

$\mathrm{P}_1 = \left(-\dfrac{1}{\sqrt{2}}, \dfrac{1}{\sqrt{2}}, \sqrt{2} - \dfrac{1}{2} \right)$ and $\mathrm{P}_2 = \left(\dfrac{1}{\sqrt{2}}, -\dfrac{1}{\sqrt{2}}, -\sqrt{2} - \dfrac{1}{2} \right)$.

We compute the bordered Hessian and we have: $\overline{\mathbb{H}}(\Lambda(x, y, \lambda)) = \left\| \begin{matrix} 0 & 2x & 2y \\ 2x & -2\lambda & 1 \\ 2y & 1 & -2\lambda \end{matrix} \right\|$.

By substituting the points found, we have:

$$\left|\overline{\mathbb{H}}(\Lambda(\mathrm{P}_1))\right| = \begin{vmatrix} 0 & -\sqrt{2} & \sqrt{2} \\ -\sqrt{2} & 1 - 2\sqrt{2} & 1 \\ \sqrt{2} & 1 & 1 - 2\sqrt{2} \end{vmatrix} = \begin{vmatrix} 0 & -\sqrt{2} & \sqrt{2} \\ 0 & 2 - 2\sqrt{2} & 2 - 2\sqrt{2} \\ \sqrt{2} & 1 & 1 - 2\sqrt{2} \end{vmatrix} =$$

$= \sqrt{2} \cdot \left(8 - 4\sqrt{2}\right) > 0$, so P_1 is a maximum point; then we get $z = -2\sqrt{2}$;

$$\left|\overline{\mathbb{H}}(\Lambda(\mathrm{P}_2))\right| = \begin{vmatrix} 0 & \sqrt{2} & -\sqrt{2} \\ \sqrt{2} & 1 + 2\sqrt{2} & 1 \\ -\sqrt{2} & 1 & 1 + 2\sqrt{2} \end{vmatrix} = \begin{vmatrix} 0 & \sqrt{2} & -\sqrt{2} \\ 0 & 2 + 2\sqrt{2} & 2 + 2\sqrt{2} \\ -\sqrt{2} & 1 & 1 + 2\sqrt{2} \end{vmatrix} =$$

$= -\sqrt{2} \cdot \left(8 + 4\sqrt{2}\right) < 0$, so P_2 is a minimum point; then we get $z = 2\sqrt{2}$.

Obviously, results are the same as those previously found.

Finally we solve the problem in a third way.

After having explicited z, we resume the problem $\begin{cases} \text{Max/min } f(x, y) = xy - 2x + 2y \\ \text{u.c.} : g(x, y) = x^2 + y^2 = 1 \end{cases}$, and, since the constraint is the trigonometric circle, we put $\begin{cases} x = \cos t \\ y = \sin t \end{cases}$, from which we obtain:

$f(x(t), y(t), z(x(t), y(t))) = F(t) = \sin t \cos t - 2\cos t + 2\sin t$. So:

$F'(t) = \cos^2 t - \sin^2 t + 2\sin t + 2\cos t = (\sin t + \cos t)(\cos t - \sin t + 2)$.

As $\cos t - \sin t + 2 > 0, \forall t$, it is $F'(t) \geq 0$ if $\cos t \geq -\sin t$, which is verified if $0 \leq t \leq \dfrac{3}{4}\pi$ and if $\dfrac{7}{4}\pi \leq t \leq 2\pi$. So $t = \dfrac{3}{4}\pi$ is a maximum point, while $t = \dfrac{7}{4}\pi$ is a minimum point. If $t = \dfrac{3}{4}\pi$ we get $\begin{cases} x = \cos\dfrac{3}{4}\pi = -\dfrac{1}{\sqrt{2}} \\ y = \sin\dfrac{3}{4}\pi = \dfrac{1}{\sqrt{2}} \end{cases}$, while if $t = \dfrac{7}{4}\pi$ we get $\begin{cases} x = \cos\dfrac{7}{4}\pi = \dfrac{1}{\sqrt{2}} \\ y = \sin\dfrac{7}{4}\pi = -\dfrac{1}{\sqrt{2}} \end{cases}$. Then z is obtained as before.

EXTREMES WITH INEQUALITY CONSTRAINTS

Let us finally treat problems such as $\begin{cases} \text{Max/min } f(x_1, x_2, ..., x_n) \\ \text{u.c.} : g_i(x_1, x_2, ..., x_n) \leq 0, 1 \leq i \leq m \end{cases}$.

If $\mathbb{X} = (x_1, x_2, ..., x_n)$, $\mathcal{E} \subset \mathbb{R}^n$ is the set in which the m inequalities $g_i(\mathbb{X}) \leq 0$ are simultaneously satisfied; $\mathcal{E}$ is called the feasible region and is a closed (limited or unlimited) subset of $\mathbb{R}^n$. To search maxima and minima for $f(\mathbb{X})$ in $\mathcal{E}$ means then to search for extremes that are at interior points of $\mathcal{E}$ $(\forall\, i : g_i(\mathbb{X}) < 0)$ and those who are at boundary points of $\mathcal{E}$ $(\exists\, i : g_i(\mathbb{X}) = 0)$.

Searching points in the interior corresponds to the problem of free extremes, searching in boundary points corresponds to the problem of extremes with equality constraints, namely the two problems that have already been treated.

To describe the $\mathcal{E}$ region with inequality constraints we may have more constraints than the variables of the objective function, thus the rule $m < n$ does no more apply.

Definition 45 : At the point $\mathbb{X}_0 \in \mathcal{E}$ the constraint $g_i(\mathbb{X})$ is said:

- satisfied, if $g_i(\mathbb{X}_0) \leq 0$;
- active, if $g_i(\mathbb{X}_0) = 0$.

At a point $\mathbb{X}_0 \in \mathcal{E}$ all constraints must be satisfied, someone may be active.

Let us then begin to build what are commonly called the Kuhn-Tucker's conditions, which represent the most general form of first order conditions, and are therefore necessary conditions, to search the extremes. To get the general statement we need to make two choices: first we choose to search for maximum points. Minimum points are related to maximum ones, once it has been noted that $\min(f(x)) = -\max(-f(x))$.

Then we choose to represent the constraints, as before, in the form $g(\mathbb{X}) \leq 0$; writing them in the form $h(\mathbb{X}) = -g(\mathbb{X}) \geq 0$ leads to determine the same set $\mathcal{E}$, but with a different formulation of Kuhn-Tucker's conditions.

As seen above, to determine extremes with equality constraints leads to impose first order conditions not on the objective function but on the Lagrangian function:

$$\Lambda(\mathbb{X}, \lambda_1, ..., \lambda_m) = f(\mathbb{X}) - \sum_{i=1}^{m} \lambda_i\, g_i(\mathbb{X})\,.$$

We use the same function, and we observe that, if the point $\mathbb{X}_0 \in \mathcal{E}$, an hypothetical solution for the problem, is an interior point with respect to the constraint $g_k(\mathbb{X})$, i.e. if $g_k(\mathbb{X}_0) < 0$, with respect to this constraint it is as if we were searching for free extremes, and then it is enough to take $\lambda_k = 0$.

If, on the contrary, the constraint is active at $\mathbb{X}_0$, i.e. if $g_k(\mathbb{X}_0) = 0$, then it is like dealing with a problem of extremes with equality constraints.

Thus the expression of the Lagrangian function allows us to search the extremes both in the interior and on the boundary of $\mathcal{E}$, just vanishing (or not) the appropriate multipliers, simply seeing which constraints are active at $\mathbb{X}_0$.

So the first condition to be imposed is $\nabla\Lambda = \mathbb{O}$, if $\mathbb{X}_0$ is a maximum point, under the assumption that the functions $f(\mathbb{X})$ and $g_i(\mathbb{X})$, $1 \leq i \leq m$, are differentiable throughout $\mathcal{E}$.

In the first order conditions also the constraints must be respected, which can be obtained as derivatives of the Lagrangian with respect to the multipliers.

As $\dfrac{\partial\Lambda}{\partial\lambda_i} = -g_i(\mathbb{X})$, we need $\dfrac{\partial\Lambda}{\partial\lambda_i} = -g_i(\mathbb{X}) \geq 0$ to get the constraint satisfied.

The request for differentiability of $f(\mathbb{X})$ and $g_i(\mathbb{X})$ does not exhaust the assumptions, which will be completed later. We state for the moment the following:

Theorem 33 (Kuhn-Tucker's conditions) : $\mathbb{X}_0$ is a solution of the problem :

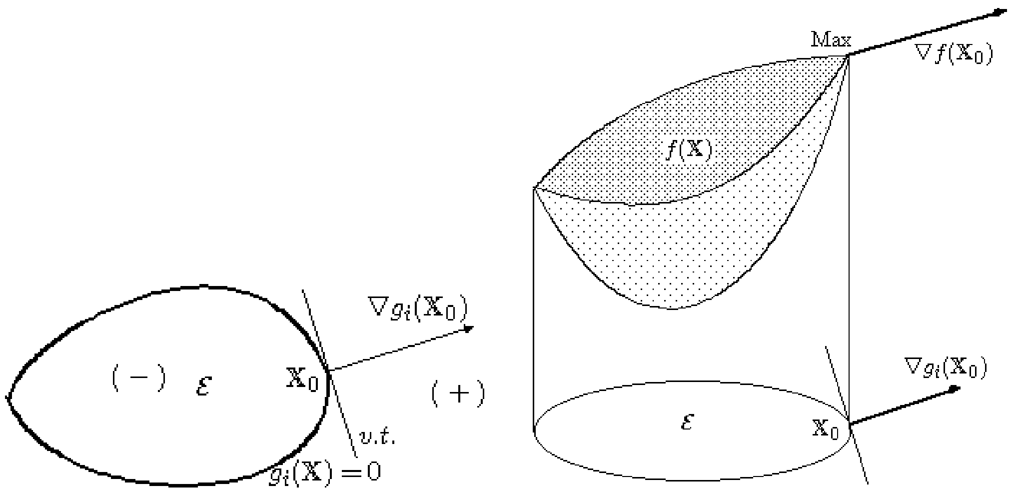
$$\begin{cases} \text{Max } f(x_1, x_2, ..., x_n) \\ \text{u.c. : } g_i(x_1, x_2, ..., x_n) \leq 0 \end{cases}$$

i.e is a maximum point for $f(\mathbb{X})$ subject to constraints $g_i(\mathbb{X}) \leq 0$, $1 \leq i \leq m$. $f(\mathbb{X})$ and $g_i(\mathbb{X})$ are differentiable throughout $\mathcal{E}$, and at $\mathbb{X}_0$ constraints are qualified.

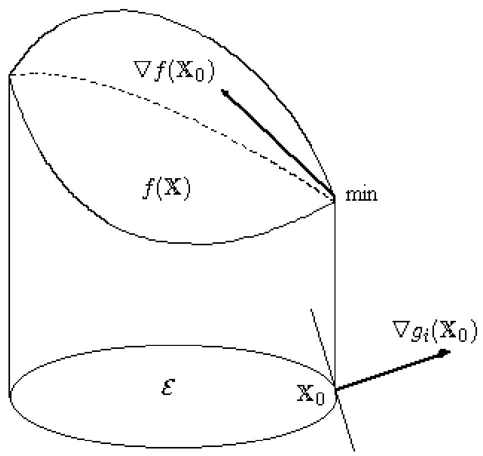
Then there exists a vector $\lambda = (\lambda_1, \lambda_2, ..., \lambda_m)$ such that:

$$\begin{cases} \dfrac{\partial \Lambda(\mathbb{X}_0)}{\partial x_i} = \dfrac{\partial f(\mathbb{X}_0)}{\partial x_i} - \displaystyle\sum_{j=1}^{m} \lambda_j \, \dfrac{\partial g_j(\mathbb{X}_0)}{\partial x_i} = 0 & 1 \le i \le n : \text{Stationarity} \\ \dfrac{\partial \Lambda}{\partial \lambda_i} = -\, g_i(\mathbb{X}_0) \ge 0 & 1 \le i \le m : \text{Primal feasibility} \\ \lambda_i \cdot g_i(\mathbb{X}_0) = 0 & 1 \le i \le m : \text{Complementary slackness} \\ \lambda_i \ge 0 & 1 \le i \le m : \text{Dual feasibility} \end{cases} .$$

The meaning of "qualified constraints at $\mathbb{X}_0$" will be explained later, when we complete the hypotheses of Theorem 33.
The first conditions: $\dfrac{\partial \Lambda(\mathbb{X}_0)}{\partial x_i} = 0$ follow the case of equality constraints, while the second ones: $\dfrac{\partial \Lambda}{\partial \lambda_i} = -\, g_i(\mathbb{X}_0) \ge 0$ reaffirm the need for constraints satisfaction.
Let's examine the third condition: $\lambda_i \cdot g_i(\mathbb{X}_0) = 0$.
In order to vanish a product, the first or the second factor must be zero. If it were $\lambda_i \ne 0$, then it must be $g_i(\mathbb{X}_0) = 0$, i.e. the constraint g_i is active at $\mathbb{X}_0$ and it is like searching extremes with equality constraints.
If it were $g_i(\mathbb{X}_0) \ne 0$, then it must be $\lambda_i = 0$, i.e. the point $\mathbb{X}_0$ is an interior point with respect to the constraint g_i, and with respect to that constraint it is like searching free extremes, then this constraint does not appear in the Lagrangian, and then $\lambda_i = 0$.
At $\mathbb{X}_0$ it could also be $\lambda_i = g_i(\mathbb{X}_0) = 0$, and this is because a constrained extreme may coincide with a free one, which happens when the coordinates of the free extreme satisfy also the constraint.
The third condition therefore means that the research is done both at interior points and at boundary ones, then throughout the whole $\mathcal{E}$.
From a practical point of view, the third condition shows how to set up calculations, i.e. we must impose equations $\dfrac{\partial \Lambda(\mathbb{X})}{\partial x_i} = 0$ and $\dfrac{\partial \Lambda(\mathbb{X})}{\partial \lambda_j} = 0$ for each of the 2^m possible combinations obtained by requiring each of the multipliers λ_i equal or different from zero.
If, for example, we had two constraints and then two multipliers, λ_1 and λ_2, we need four systems, those corresponding to the four cases:
($\lambda_1 = 0$ and $\lambda_2 = 0$), ($\lambda_1 \ne 0$ and $\lambda_2 = 0$), ($\lambda_1 = 0$ and $\lambda_2 \ne 0$) and ($\lambda_1 \ne 0$ and $\lambda_2 \ne 0$).
With three constraints the cases become eight and so on ...



Let us examine the fourth and final condition, wich depends on the sign of the constraints (we chose the negative one) and on the type of extreme we are looking for (we have chosen to look for maximum points).
Starting from $g_i(\mathbb{X}) = 0$ we have seen that $\nabla g_i(\mathbb{X}_0)$ is orthogonal to the vector tangent to the constraint at $\mathbb{X}_0$.
Since the gradient expresses the direction of maximum growth, and being $\mathcal{E}$ characterized by negative values for the constraint $g_i(\mathbb{X})$, outside $\mathcal{E}$ we have positive values for $g_i(\mathbb{X})$, from which it follows that the gradient $\nabla g_i(\mathbb{X}_0)$ must be oriented outside $\mathcal{E}$, as shown in the figure, which is based, as a particular case, on a single constraint.
Suppose that $\mathbb{X}_0$ is a maximum point and that the constraint $g_i(\mathbb{X})$ is active at $\mathbb{X}_0$: $g_i(\mathbb{X}_0) = 0$.
We have seen that $\nabla g_i(\mathbb{X}_0)$ is directed outwards $\mathcal{E}$.
Also the gradient $\nabla f(\mathbb{X}_0)$ of the function at the point $\mathbb{X}_0$, that in the figure, for easy reading, has been shifted on the surface, must be directed outwards $\mathcal{E}$, as it indicates the direction of maximum growth; if it goes to the inside, starting from the maximum point, it would indicate a decrease in the values of the function.
From the theory of the extremes with equality constraints we know that, at a point which is the solution of a problem of extremes, the gradient of the objective function must be a linear combination of the gradients of the constraints: $\nabla f(\mathbb{X}_0) = \lambda_i \nabla g_i(\mathbb{X}_0)$.
We have also seen that $\nabla f(\mathbb{X}_0)$ and $\nabla g_i(\mathbb{X}_0)$ must go towards the same direction: these two requests are both satisfied if $\lambda_i > 0$.
There is a theorem (Farkas) which extends this property to the case where more than one constraint is active at $\mathbb{X}_0$: $\nabla f(\mathbb{X}_0) = \sum_{i:1}^{m} \lambda_i \nabla g_i(\mathbb{X}_0)$.
Each λ_i must be non-negative so that $\nabla f(\mathbb{X}_0)$ is directed outwards, like the gradients $\nabla g_i(\mathbb{X}_0)$.



If, while maintaining the constraints in the form $g_i(\mathbb{X}) \leq 0$, we had looked for the minimum point at which the constraint is active: $g_i(\mathbb{X}_0) = 0$, as shown in the figure, the gradient $\nabla f(\mathbb{X}_0)$ must now move towards the interior, where the function takes values greater than the minimum. But then it must be $\lambda_i \leq 0$, as $\nabla f(\mathbb{X}_0)$ and $\nabla g_i(\mathbb{X}_0)$ are not oriented towards the same direction, since the first goes to the inside while the second goes to the outside. The search for minimum points requires non-positive multipliers, leaving unchanged the first three conditions of Theorem 33.
If the constraints were expressed in the form $g_i(\mathbb{X}) \geq 0$, the previous requests on multipliers' sign for a maximum point and a minimum point would have been reversed. The sign (not negative) of multipliers for a maximum at points with constraints $g_i(\mathbb{X}) \leq 0$ is therefore the same as that for a minimum at points with constraints $g_i(\mathbb{X}) \geq 0$.

CONSTRAINTS QUALIFICATION
Finally, let us complete the last part concerning the assumptions for the validity of Kuhn-Tucker's conditions: that is qualified constraints at $\mathbb{X}_0$.
Studying the sign of the multipliers λ_i in order to clarify the nature of the point we are examining, we need to establish clearly the $\nabla f(\mathbb{X}_0)$ direction with respect to the directions of the constraints gradients.
This problem is called "constraint qualification" at $\mathbb{X}_0$.
To describe this concept, we need some definitions, the first of which is "feasible direction".
Definition 46 : Vector $v \in \mathbb{R}^n$ is a feasible direction at $\mathbb{X}_0$ for $\mathcal{E}$ if it exists a curve, $t \to r(t) \in \mathbb{R}^n$, such that:
1) $r(0) = \mathbb{X}_0$;
2) $r'(0) = v$;
3) $\exists\, \varepsilon : r(t) \in \mathcal{E}, \forall\, t \in [0, \varepsilon[$.
So there exists at least one arc, however small, of a continuous curve starting from $\mathbb{X}_0$ entering in $\mathcal{E}$: the feasible direction v is the tangent line to such curve at $\mathbb{X}_0$.
This definition strictly concerns boundary points of $\mathcal{E}$, i.e. points where at least one constraint is active. If $\mathbb{X}_0$ is an interior point of $\mathcal{E}$, every direction is feasible.
Feasible directions at $\mathbb{X}_0$ for $\mathcal{E}$ are those which, starting from $\mathbb{X}_0$, go inside the feasible region $\mathcal{E}$, or, as a limiting case, are tangential to $\mathcal{E}$.

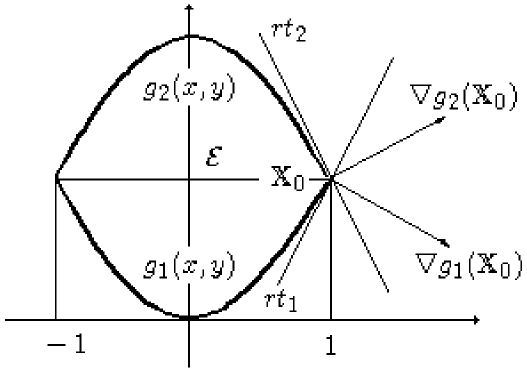
Definition 47 : A cone is a set $\mathbb{A} \subseteq \mathbb{R}^n$ such that: $\mathbb{X} \in \mathbb{A} \Rightarrow k \cdot \mathbb{X} \in \mathbb{A}, \forall\, k \in \mathbb{R}$.
A positive cone is a set $\mathbb{A} \subseteq \mathbb{R}^n$ such that: $\mathbb{X} \in \mathbb{A} \Rightarrow k \cdot \mathbb{X} \in \mathbb{A}, \forall\, k \in \mathbb{R}_+$.
Feasible directions at $\mathbb{X}_0$ for $\mathcal{E}$ form a positive cone: $\Gamma(\mathbb{X}_0)$.

To establish the nature of point $\mathbb{X}_0$ we saw that it is necessary, in the chosen formulation, that $\nabla f(\mathbb{X}_0)$ heads to the outside of $\mathcal{E}$, on the same side as $\nabla g_i(\mathbb{X}_0)$.
Then we need to study the directions v that we call "retroverted directions" with respect to $\nabla g_i(\mathbb{X}_0)$, i.e. directions for which $\nabla g_i(\mathbb{X}_0) \cdot v \leq 0$.
As $\nabla g_i(\mathbb{X}_0) \cdot v = \|\nabla g_i(\mathbb{X}_0)\| \cdot \|v\| \cdot \cos\alpha$, we get $\nabla g_i(\mathbb{X}_0) \cdot v \leq 0$ if $\dfrac{\pi}{2} \leq \alpha \leq \pi$.
The feasible directions at $\mathbb{X}_0$ for $\mathcal{E}$ and the retroverted directions with respect to $\nabla g_i(\mathbb{X}_0)$ at $\mathbb{X}_0$ must be the same directions in order that the analysis of $\nabla f(\mathbb{X}_0)$ is well made.
Let us consider all the active constraints at $\mathbb{X}_0$: $g_i(\mathbb{X}_0) = 0$.
$\widetilde{\Gamma}\,(\mathbb{X}_0)$ is the cone of the retroverted directions with respect to the gradients of the active constraints $\nabla g_i(\mathbb{X}_0)$ at $\mathbb{X}_0$, i.e.:
$\widetilde{\Gamma}\,(\mathbb{X}_0) = \{v \in \mathbb{R}^n : \nabla g_i(\mathbb{X}_0) \cdot v \leq 0; g_i(\mathbb{X}_0) = 0\}$.
In general, as we shall see in the examples, is $\Gamma(\mathbb{X}_0) \subseteq \widetilde{\Gamma}\,(\mathbb{X}_0)$.
We have the following:
Definition 48 : Constraints are qualified at $\mathbb{X}_0$ if it is $\Gamma(\mathbb{X}_0) = \widetilde{\Gamma}\,(\mathbb{X}_0)$, i.e. feasible directions and retroverted directions form the same cone.

This condition completes, with differentiability, Kuhn-Tucker's conditions for a maximum or minimum point.

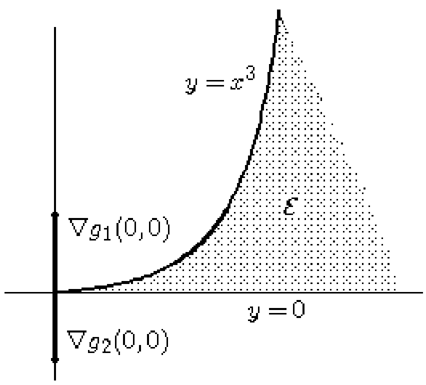
Example 78 : Let us consider as feasible region $\mathcal{E}$ the one given by the constraints:
$\begin{cases} g_1(x,y) = x^2 - y \leq 0 \\ g_2(x,y) = y - 2 + x^2 \leq 0 \end{cases}$.
The two constraints are simultaneously active at $(1,1)$ and $(-1,1)$. Let us compose $\Gamma(1,1)$.
To this effect, let us find equation of the tangent lines to the curves $g_1 : y = x^2$ and $g_2 : y = 2 - x^2$.

For the first, rt_1, we get $y = 2x - 1$ while for the second, rt_2, $y = -2x + 3$.



Let $v = (v_1, v_2)$. To go from $(1,1)$ to $\mathcal{E}$ we need $v_1 \leq 0$, to move to the left; v_2 must be in the range between the two tangent lines, i.e. $2v_1 \leq v_2 \leq -2v_1$ ($v_1 \leq 0$!!) and so:
$\Gamma(1,1) = \{(v_1, v_2) \in \mathbb{R}^2 : v_1 \leq 0, 2v_1 \leq v_2 \leq -2v_1\}$.
Then let us determine $\widetilde{\Gamma}\ (1,1)$.
From $g_1(x,y) = x^2 - y$ we get $\nabla g_1(x,y) = (2x, -1)$ and so $\nabla g_1(1,1) = (2, -1)$.
From $g_2(x,y) = y - 2 + x^2$ we get $\nabla g_2(x,y) = (2x, 1)$ and so $\nabla g_2(1,1) = (2,1)$.
Now let us look for retroverted directions to both gradients. It will be:
$\nabla g_1(1,1) \cdot (v_1, v_2) = (2, -1) \cdot (v_1, v_2) = 2v_1 - v_2 \leq 0$ for $v_2 \geq 2v_1$, while
$\nabla g_2(1,1) \cdot (v_1, v_2) = (2,1) \cdot (v_1, v_2) = 2v_1 + v_2 \leq 0$ for $v_2 \leq -2v_1$, and so:
$2v_1 \leq v_2 \leq -2v_1$, and this implies also $v_1 \leq 0$.
So $\widetilde{\Gamma}\ (1,1) = \{(v_1, v_2) \in \mathbb{R}^2 : v_1 \leq 0, 2v_1 \leq v_2 \leq -2v_1\} = \Gamma(1,1)$.
The constraints are then qualified at $(1,1)$. We can see in a similar way that this happens also at point $(-1,1)$, where:
$\widetilde{\Gamma}\ (-1,1) = \{(v_1, v_2) \in \mathbb{R}^2 : v_1 \geq 0, -2v_1 \leq v_2 \leq 2v_1\} = \Gamma(-1,1)$.

Example 79 : Let the feasible region $\mathcal{E}$ be that given by the two constraints:
$$\begin{cases} g_1(x,y) = y - x^3 \leq 0 \\ g_2(x,y) = -y \leq 0 \end{cases}.$$
At (0.0) both constraints are active. As $g_1 : y = x^3$ has at $(0,0)$ as tangent line the x-axis, the only feasible direction from $(0,0)$ to $\mathcal{E}$ is the positive semi-axis of x, so:
$\Gamma(0,0) = \{(v_1, v_2) \in \mathbb{R}^2 : v_1 \geq 0, v_2 = 0\}$. Then let us determine $\widetilde{\Gamma}\ (0,0)$.
From $g_1(x,y) = y - x^3$ we get $\nabla g_1(x,y) = (-3x^2, 1)$ and so $\nabla g_1(0,0) = (0,1)$.
From $g_2(x,y) = -y$ we get $\nabla g_2(x,y) = (0, -1)$ and so $\nabla g_2(0,0) = (0, -1)$.
Now let us find the retroverted directions to both gradients. It will be:
$\nabla g_1(0,0) \cdot (v_1, v_2) = (0,1) \cdot (v_1, v_2) = v_2 \leq 0$, while
$\nabla g_2(0,0) \cdot (v_1, v_2) = (0, -1) \cdot (v_1, v_2) = -v_2 \leq 0$, or $v_2 \geq 0$ and so: $v_2 = 0$.



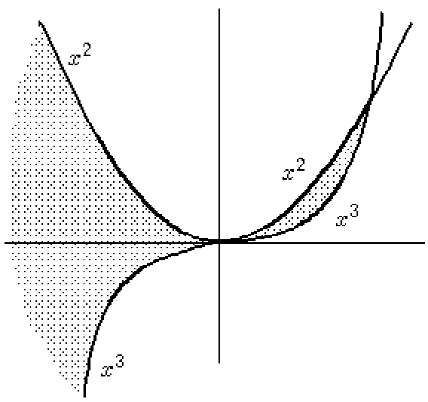
So $\widetilde{\Gamma}(0,0)=\{(v_1,v_2)\in\mathbb{R}^2 : v_2=0,\forall\, v_1\}$, i.e. the whole x-axis. In this example therefore it is $\Gamma(0,0)\subset\widetilde{\Gamma}(0,0)$ and not $\Gamma(0,0)=\widetilde{\Gamma}(0,0)$, so at $(0,0)$ the constraints are not qualified.

Example 80 : Let the feasible region $\mathcal{E}$ be that given now by the three constraints:
$$\begin{cases} g_1(x,y)=y-x^3\le 0\\ g_2(x,y)=-y\le 0\\ g_3(x,y)=-x\le 0\end{cases}.$$
The feasible region $\mathcal{E}$ is exactly the same as the previous example, even if characterized by a further constraint: $g_3(x,y)=-x$. Also this is active at $(0,0)$, so $\Gamma(0,0)$ remains unchanged but we must redefine $\widetilde{\Gamma}(0,0)$, since we have now an additional constraint.
From $g_3(x,y)=-x$ we get $\nabla g_3(x,y)=(-1,0)$ and so $\nabla g_3(0,0)=(-1,0)$. So:
$\nabla g_3(0,0)\cdot(v_1,v_2)=(-1,0)\cdot(v_1,v_2)=-v_1\le 0$ for $v_1\ge 0$,
and this, together with the two already found, gives:
$\Gamma(0,0)=\{(v_1,v_2)\in\mathbb{R}^2 : v_1\ge 0, v_2=0\}=\widetilde{\Gamma}(0,0)$.
With the third constraint, the constraints are now qualified at $(0,0)$.
Constraints qualification at a point depends not only on the shape of the region $\mathcal{E}$ but also on the constraints describing it. The same region described by different constraints may have qualified constraints at a point while they were not so if described by other constraints.

There are some sufficient conditions to ensure constraints qualification at a given point.
The most important is the following:
Theorem 34 : If $\nabla g_i(\mathbb{X}_0)$, gradients of active constraints at $\mathbb{X}_0$, are linearly independent vectors, then the constraints are qualified at $\mathbb{X}_0$.
This condition is however sufficient, and not necessary, for constraints qualification at $\mathbb{X}_0$.

Example 81 : Let us resume the feasible region $\mathcal{E}$ that is given by the two constraints:
$$\begin{cases} g_1(x,y)=x^2-y\le 0\\ g_2(x,y)=y-2+x^2\le 0\end{cases}.$$
It is $\nabla g_1(1,1)=(2,-1)$ and $\nabla g_2(1,1)=(2,1)$, as $\begin{vmatrix}2 & -1\\ 2 & 1\end{vmatrix}=4\neq 0$, so gradients are independent vectors and constraints are qualified.

Example 82 : Let us consider the feasible region $\mathcal{E}$ that is given by the two constraints:
$$\begin{cases} g_1(x,y)=x^3-y\le 0\\ g_2(x,y)=y-x^2\le 0\end{cases}.$$
Region $\mathcal{E}$ is the darkened region represented in Figure:



The feasible directions at $(0,0)$ for $\mathcal{E}$, both on the left and on the right, are the x-axis, which is the tangent line at $(0,0)$ for $g_1 : y = x^3$ and for $g_2 : y = x^2$. So we have:
$\mathit{\Gamma}(0,0) = \{(v_1, v_2) \in \mathbb{R}^2 : v_2 = 0, \forall\, v_1\}$.
From $g_1(x,y) = x^3 - y$ we get $\nabla g_1(x,y) = (3x^2, \, -1)$ and so $\nabla g_1(0,0) = (0, \, -1)$.
From $g_2(x,y) = y - x^2$ we get $\nabla g_2(x,y) = (\, -2x, 1)$ and so $\nabla g_2(0,0) = (0,1)$.
Let us search the retroverted directions to both gradients. It is
$\nabla g_1(0,0) \cdot (v_1, v_2) = (0, \, -1) \cdot (v_1, v_2) = \, -v_2 \leq 0$ for $v_2 \geq 0$, while
$\nabla g_2(0,0) \cdot (v_1, v_2) = (0,1) \cdot (v_1, v_2) = v_2 \leq 0$, and so: $0 \leq v_2 \leq 0$, and this double inequation implies $v_2 = 0$.
So $\widetilde{\mathit{\Gamma}}\,(0,0) = \{(v_1, v_2) \in \mathbb{R}^2 : v_2 = 0, \forall\, v_1\} = \mathit{\Gamma}(0,0)$. Constraints are qualified at $(0,0)$ even if $\nabla g_1(0,0) = (0, \, -1)$ and $\nabla g_2(0,0) = (0,1)$ are linearly dependent vectors, confirming that the independence condition is sufficient and not necessary for constraints qualification.

Example 83 : Let us study the problem $\begin{cases} \text{Max/min } f(x,y) = 1 - x^2 - y^2 \\ \text{u.c. : } \begin{cases} g_1(x,y) = x^3 - y \leq 0 \\ g_2(x,y) = y - x^2 \leq 0 \end{cases} \end{cases}$.

The feasible region $\mathcal{E}$ and the constraints qualification have been studied in Example 82 at $(0,0)$. At $(1,1)$ we have:
$g_1(x,y) = x^3 - y$ from which $\nabla g_1(x,y) = (3x^2, \, -1)$ and so $\nabla g_1(1,1) = (3, \, -1)$.
$g_2(x,y) = y - x^2$ from which $\nabla g_2(x,y) = (\, -2x, 1)$ and so $\nabla g_2(1,1) = (\, -2, 1)$.
Such vectors are indipendent: $\begin{vmatrix} 3 & -1 \\ -2 & 1 \end{vmatrix} = 1 \neq 0$ so at $(1,1)$ constraints are qualified.
Finally, we must check the points where only one constraint is active; for linear independence, the constraint gradient must not be the null vector. But the second components of these gradients are constant and nonzero, so the constraints are always qualified.
Finally, we note that the feasible region is not a bounded set, so Weierstrass's theorem does not apply.
Let us construct the Lagrangian function:
$\Lambda(x, y, \lambda_1, \lambda_2) = 1 - x^2 - y^2 - \lambda_1 \left(x^3 - y\right) - \lambda_2 \left(y - x^2\right)$ and study four cases.

I case: $\lambda_1 = 0, \lambda_2 = 0$: let us solve the system:

$\begin{cases} \Lambda'_x = 0 \Rightarrow \, -2x = 0 \\ \Lambda'_y = 0 \Rightarrow \, -2y = 0 \\ \Lambda'_{\lambda_1} \geq 0 \Rightarrow x^3 - y \leq 0 \\ \Lambda'_{\lambda_2} \geq 0 \Rightarrow y - x^2 \leq 0 \end{cases} \Rightarrow \begin{cases} x = 0 \\ y = 0 \\ 0 - 0 \leq 0 : \text{ satisfied} \\ 0 - 0 \leq 0 : \text{ satisfied} \end{cases}$. As $\lambda_1 = \lambda_2 = 0$, we study the free extremes of the function.

As $\mathbb{H}(x,y) = \begin{vmatrix} -2 & 0 \\ 0 & -2 \end{vmatrix} = \mathbb{H}(0,0)$, we get $\begin{cases} f''_{xx} = \, -2 < 0 \\ f''_{xx}\, f''_{yy} - \left(f''_{xy}\right)^2 = 4 > 0 \end{cases}$ and so $(0,0)$ is a maximum point, with $f(0,0) = 1$.

II case: $\lambda_1 \neq 0, \lambda_2 = 0$: let us solve the system:

$\begin{cases} \Lambda'_x = 0 \Rightarrow \, -2x - 3\lambda_1 x^2 = 0 \\ \Lambda'_y = 0 \Rightarrow \, -2y + \lambda_1 = 0 \\ \Lambda'_{\lambda_1} = 0 \Rightarrow y = x^3 \\ \Lambda'_{\lambda_2} \geq 0 \Rightarrow y - x^2 \leq 0 \end{cases} \Rightarrow \begin{cases} -x(2 + 3\lambda_1 x) = 0 \\ \lambda_1 = 2y \\ y = x^3 \\ y \leq x^2 \end{cases}$ which gives two systems:

$$\begin{cases} x = 0 \\ \lambda_1 = 0 \\ y = 0 \\ 0 \leq 0: \text{ satisfied} \end{cases}$$, but $(0,0)$ has already been studied, and

$$\begin{cases} x = -\dfrac{2}{3\lambda_1} \\ y = \dfrac{\lambda_1}{2} \\ \dfrac{\lambda_1}{2} = -\dfrac{8}{27\lambda_1^3} \Rightarrow \lambda_1^4 = -\dfrac{16}{27} \\ y \leq x^2 \end{cases}$$ which therefore has no solutions.

III case: $\lambda_1 = 0, \lambda_2 \neq 0$: let us solve the system:

$$\begin{cases} \Lambda'_x = 0 \Rightarrow -2x + 2\lambda_2 x = 0 \\ \Lambda'_y = 0 \Rightarrow -2y - \lambda_2 = 0 \\ \Lambda'_{\lambda_1} \geq 0 \Rightarrow x^3 \leq y \\ \Lambda'_{\lambda_2} = 0 \Rightarrow y = x^2 \end{cases} \Rightarrow \begin{cases} 2x(\lambda_2 - 1) = 0 \\ \lambda_2 = -2y \\ x^3 \leq y \\ y = x^2 \end{cases}$$ which gives two systems:

$$\begin{cases} x = 0 \\ \lambda_2 = 0 \\ 0 \leq 0: \text{ satisfied} \\ y = 0 \end{cases}$$, but $(0,0)$ has already been studied, and

$$\begin{cases} \lambda_2 = 1 \\ y = -\dfrac{1}{2} \\ x^3 \leq y \\ x^2 = -\dfrac{1}{2} \end{cases}$$ which therefore has no solutions.

IV case: $\lambda_1 \neq 0, \lambda_2 \neq 0$: let us solve the system:

$$\begin{cases} \Lambda'_x = 0 \Rightarrow -2x - 3\lambda_1 x^2 + 2\lambda_2 x = 0 \\ \Lambda'_y = 0 \Rightarrow -2y + \lambda_1 - \lambda_2 = 0 \\ \Lambda'_{\lambda_1} = 0 \Rightarrow x^3 = y \\ \Lambda'_{\lambda_2} = 0 \Rightarrow y = x^2 \end{cases} \Rightarrow \begin{cases} 0 = 0 \\ \lambda_1 = \lambda_2 \\ x = 0 \\ y = 0 \end{cases} \text{ and } \begin{cases} -2 - 3\lambda_1 + 2\lambda_2 = 0 \\ -2 + \lambda_1 - \lambda_2 = 0 \\ x = 1 \\ y = 1 \end{cases} \Rightarrow$$

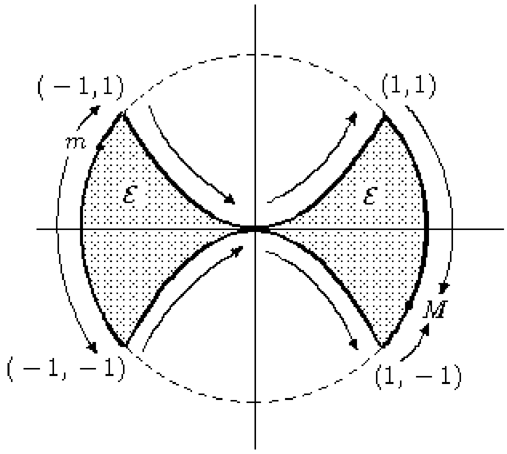
$$\begin{cases} -2 - 3\lambda_1 + 2\lambda_1 - 4 = 0 \\ \lambda_2 = \lambda_1 - 2 \\ x = 1 \\ y = 1 \end{cases} \Rightarrow \begin{cases} \lambda_1 = -6 \\ \lambda_2 = -8 \\ x = 1 \\ y = 1 \end{cases}.$$

Since both multipliers are negative, point $(1,1)$ may be a minimum point. To solve the problem, we note that function $f(x,y) = 1 - x^2 - y^2$, being a polynomial, is continuous throughout $\mathbb{R}^2$. The right side of $\mathcal{E}$ is a bounded and closed set, then by Weierstrass's theorem there are absolute minimum and maximum, which can be only at $(0,0)$ and at $(1,1,)$, as seen above.

Maximum $f(0,0) = 1$ is clearly an absolute one, while minimum $f(0,0)$ is only local, as, analyzing for example the function on the negative x-semi-axis, we get $\lim\limits_{x \to -\infty} f(x,0) = -\infty$, so the function can take indefinitely negative large values in $\mathcal{E}$, and therefore it can not have an absolute minimum.

Example 84 : Let us study the problem $\begin{cases} \text{Max/min } f(x,y) = 2x - y \\ \text{u.c. : } \begin{cases} g_1(x,y) = y^2 - x^4 \le 0 \\ g_2(x,y) = x^2 + y^2 - 2 \le 0 \end{cases} \end{cases}$.

Region $\mathcal{E}$ is the darkened region represented in Figure:



From $y^2 \le x^4$ it follows $-x^2 \le y \le x^2$, i.e. the part of the plan between the two parabolas $y = -x^2$ and $y = x^2$, inside the circle $x^2 + y^2 = 2$, with center $(0,0)$ and radius $r = \sqrt{2}$.

From $\begin{cases} y^2 - x^4 = 0 \\ y^2 = 2 - x^2 \end{cases}$ we get $2 - x^2 - x^4 = 0$, and $x^2 = \dfrac{-1 \pm \sqrt{1+8}}{2} \Rightarrow \begin{cases} x^2 = -2 \\ x^2 = 1 \end{cases}$ so $x = \pm 1$, so we have four intersections: $(1,1)$, $(1,-1)$, $(-1,1)$ and $(-1,-1)$.

From $g_1(x,y) = y^2 - x^4$ we get $\nabla g_1(x,y) = \left(-4x^3, 2y\right)$; from $g_2(x,y) = x^2 + y^2 - 2$ we get $\nabla g_2(x,y) = (2x, 2y)$ and so:

$\nabla g_1(1,1) = (-4,2)$ and $\nabla g_2(1,1) = (2,2)$: indipendent vectors;

$\nabla g_1(1,-1) = (-4,-2)$ and $\nabla g_2(1,-1) = (2,-2)$: indipendent vectors;

$\nabla g_1(-1,1) = (4,2)$ and $\nabla g_2(-1,1) = (-2,2)$: indipendent vectors;

$\nabla g_1(-1,-1) = (4,-2)$ and $\nabla g_2(-1,-1) = (-2,-2)$: indipendent vectors.

At $(0,0)$ it is $\nabla g_1(0,0) = (0,0)$, that does not allow constraints qualification.

The part of $\mathcal{E}$: $-x^2 \le y \le x^2$, described by two constraints $\begin{cases} h_1(x,y) = -y - x^2 \le 0 \\ h_2(x,y) = y - x^2 \le 0 \end{cases}$, gives $\nabla h_1(0,0) = (0,-1)$ and $\nabla h_2(0,0) = (0,1)$, so we can determine:

$\widetilde{\Gamma}(0,0) = \left\{(v_1,v_2) \in \mathbb{R}^2 : v_2 = 0, \forall v_1\right\} = \Gamma(0,0)$.

So at $(0,0)$ the constraints are qualified.

Where only one constraint is active, the constraint is always qualified. So let us construct the Lagrangian: $\Lambda(x,y,\lambda_1,\lambda_2) = 2x - y - \lambda_1\left(y^2 - x^4\right) - \lambda_2\left(x^2 + y^2 - 2\right)$ and study four cases.

I case: $\lambda_1 = 0, \lambda_2 = 0$: let us solve the system:

$\begin{cases} \Lambda'_x = 0 \Rightarrow 2 = 0 \\ \Lambda'_y = 0 \Rightarrow -1 = 0 \\ \Lambda'_{\lambda_1} \ge 0 \Rightarrow y^2 - x^4 \le 0 \\ \Lambda'_{\lambda_2} \ge 0 \Rightarrow x^2 + y^2 - 2 \le 0 \end{cases}$. The system has no solutions.

<u>II case: $\lambda_1 \neq 0, \lambda_2 = 0$</u> : let us solve the system:

$$\begin{cases} \Lambda'_x = 0 \Rightarrow 2 + 4\lambda_1 x^3 = 0 \\ \Lambda'_y = 0 \Rightarrow -1 - 2\lambda_1 y = 0 \\ \Lambda'_{\lambda_1} = 0 \Rightarrow y^2 = x^4 \\ \Lambda'_{\lambda_2} \geq 0 \Rightarrow x^2 + y^2 \leq 2 \end{cases} \Rightarrow \begin{cases} x = -\sqrt[3]{\dfrac{1}{2\lambda_1}} \\ y = -\dfrac{1}{2\lambda_1} \\ \dfrac{1}{4\lambda_1^2} = \sqrt[3]{\dfrac{1}{16\lambda_1^4}} \\ x^2 + y^2 \leq 2 \end{cases} \Rightarrow \begin{cases} x = -\sqrt[3]{\dfrac{1}{2\lambda_1}} \\ y = -\dfrac{1}{2\lambda_1} \\ 16\lambda_1^4 (4\lambda_1^2 - 1) = 0 \\ x^2 + y^2 \leq 2 \end{cases}.$$

The solution $\lambda_1 = 0$ is not acceptable, so we have:

$$\begin{cases} x = -\sqrt[3]{\dfrac{1}{2\lambda_1}} \\ y = -\dfrac{1}{2\lambda_1} \\ 4\lambda_1^2 - 1 = 0 \\ x^2 + y^2 \leq 2 \end{cases} \Rightarrow \begin{cases} x = -1 \\ y = -1 \\ \lambda_1 = \frac{1}{2} \\ 1 + 1 \leq 2 : \text{ satisfied} \end{cases} \text{ and } \begin{cases} x = 1 \\ y = 1 \\ \lambda_1 = -\frac{1}{2} \\ 1 + 1 \leq 2 : \text{ satisfied} \end{cases}.$$

So $(-1, -1)$ may be a maximum point $(\lambda_1 > 0)$, while $(1, 1)$ may be a minimum point $(\lambda_1 < 0)$.

<u>III case: $\lambda_1 = 0, \lambda_2 \neq 0$</u> : let us solve the system:

$$\begin{cases} \Lambda'_x = 0 \Rightarrow 2 - 2\lambda_2 x = 0 \\ \Lambda'_y = 0 \Rightarrow -1 - 2\lambda_2 y = 0 \\ \Lambda'_{\lambda_1} \geq 0 \Rightarrow y^2 \leq x^4 \\ \Lambda'_{\lambda_2} = 0 \Rightarrow x^2 + y^2 = 2 \end{cases} \Rightarrow \begin{cases} x = \dfrac{1}{\lambda_2} \\ y = -\dfrac{1}{2\lambda_2} \\ y^2 \leq x^4 \\ \dfrac{1}{\lambda_2^2} + \dfrac{1}{4\lambda_2^2} = 2 \end{cases} \Rightarrow \begin{cases} x = \dfrac{1}{\lambda_2} \\ y = -\dfrac{1}{2\lambda_2} \\ y^2 \leq x^4 \\ \lambda_2^2 = \dfrac{5}{8} \end{cases}$$

which gives the two solutions:

$$\begin{cases} x = 2\sqrt{\frac{2}{5}} \\ y = -\sqrt{\frac{2}{5}} \\ \frac{2}{5} \leq \frac{64}{25} : \text{ satisfied} \\ \lambda_2 = \frac{1}{2}\sqrt{\frac{5}{2}} \end{cases} \text{ and } \begin{cases} x = -2\sqrt{\frac{2}{5}} \\ y = \sqrt{\frac{2}{5}} \\ \frac{2}{5} \leq \frac{64}{25} : \text{ satisfied} \\ \lambda_2 = -\frac{1}{2}\sqrt{\frac{5}{2}} \end{cases}.$$

So $\left(2\sqrt{\frac{2}{5}}, -\sqrt{\frac{2}{5}}\right)$ may be a maximum point $(\lambda_2 > 0)$, while $\left(-2\sqrt{\frac{2}{5}}, \sqrt{\frac{2}{5}}\right)$ may be a minimum point $(\lambda_2 < 0)$.

<u>IV case: $\lambda_1 \neq 0, \lambda_2 \neq 0$</u> : let us solve the system:

$$\begin{cases} \Lambda'_x = 0 \Rightarrow 2 + 4\lambda_1 x^3 - 2\lambda_2 x = 0 \\ \Lambda'_y = 0 \Rightarrow -1 - 2\lambda_1 y - 2\lambda_2 y = 0 \\ \Lambda'_{\lambda_1} = 0 \Rightarrow y^2 = x^4 \\ \Lambda'_{\lambda_2} = 0 \Rightarrow x^2 + y^2 = 2 \end{cases} \text{ from which we obtain four systems:}$$

$$\begin{cases} 2+4\lambda_1-2\lambda_2=0 \\ -1-2\lambda_1-2\lambda_2=0 \\ x=1 \\ y=1 \end{cases} \Rightarrow \begin{cases} \lambda_1=-\frac{1}{2} \\ \lambda_2=0 \\ x=1 \\ y=1 \end{cases}$$

$(1,1)$ has already been studied: it may be a minimum point $(\lambda_1<0, \lambda_2\leq 0)$;

$$\begin{cases} 2+4\lambda_1-2\lambda_2=0 \\ -1+2\lambda_1+2\lambda_2=0 \\ x=1 \\ y=-1 \end{cases} \Rightarrow \begin{cases} \lambda_1=-\frac{1}{6} \\ \lambda_2=\frac{2}{3} \\ x=1 \\ y=-1 \end{cases}$$

this point cannot be neither a maximum nor a minimum point since the multipliers have a different sign;

$$\begin{cases} 2-4\lambda_1+2\lambda_2=0 \\ -1-2\lambda_1-2\lambda_2=0 \\ x=-1 \\ y=1 \end{cases} \Rightarrow \begin{cases} \lambda_1=\frac{1}{6} \\ \lambda_2=-\frac{2}{3} \\ x=-1 \\ y=1 \end{cases}$$

this point cannot be neither a maximum nor a minimum point since the multipliers have a different sign;

$$\begin{cases} 2-4\lambda_1+2\lambda_2=0 \\ -1+2\lambda_1+2\lambda_2=0 \\ x=-1 \\ y=-1 \end{cases} \Rightarrow \begin{cases} \lambda_1=\frac{1}{2} \\ \lambda_2=0 \\ x=-1 \\ y=-1 \end{cases}$$

$(-1,-1)$ has already been studied: it may be a maximum point $(\lambda_1>0, \lambda_2\geq 0)$.

As $\mathcal{E}$ is a bounded and closed set and the function $f(x,y)=2x-y$ is continuous throughout $\mathbb{R}^2$, there must be an absolute maximum and an absolute minimum, and also, perhaps, there might be some relative ones. Then let us examine the behavior of $f(x,y)$ on the boundary of $\mathcal{E}$.
On constraint $g_1: y^2=x^4 \Rightarrow y=\pm x^2$:
$f\big(x,x^2\big)=2x-x^2 \Rightarrow f'(x)=2-2x\geq 0$ if $x\leq 1$;
$f\big(x,-x^2\big)=2x+x^2 \Rightarrow f'(x)=2+2x\geq 0$ if $x\geq -1$.

If with an arrow we indicate the direction in which the function increases, we have that on the two parabolas the objective function increases from left to right, so it grows both in the path from $(-1,1)$ to $(1,1)$, and in the path from $(-1,-1)$ to $(1,-1)$.

Finally, let us analyze the behavior on the circumference.
If $\begin{cases} x=\sqrt{2}\cos t \\ y=\sqrt{2}\sin t \end{cases}$ we get $f(x(t),y(t))=2\sqrt{2}\cos t-\sqrt{2}\sin t$ and so:
$f'(t)=\sqrt{2}(-2\sin t-\cos t)\geq 0$ if $\cos t\leq -2\sin t$. This inequality is verified:
if $\alpha\leq t\leq\beta$, with $\cos\alpha=\dfrac{x}{\sqrt{2}}=-\dfrac{2}{\sqrt{5}}$ and $\sin\alpha=\dfrac{y}{\sqrt{2}}=\dfrac{1}{\sqrt{5}}$;
and if $\cos\beta=\dfrac{x}{\sqrt{2}}=\dfrac{2}{\sqrt{5}}$ and $\sin\beta=\dfrac{y}{\sqrt{2}}=-\dfrac{1}{\sqrt{5}}$.
So, starting from $m=\left(-2\sqrt{\dfrac{2}{5}},\sqrt{\dfrac{2}{5}}\right)$ the function increases in both directions, and so this is a minimum point; on the contrary, we arrive at $M=\left(2\sqrt{\dfrac{2}{5}},-\sqrt{\dfrac{2}{5}}\right)$ always increasing, both from above and from below, and so this is a maximum point. At $(1,1)$ we

arrive increasing on the parabola and we continue increasing on the circumference, so $(1,1)$ is not a minimum point. Similarly at $(-1,-1)$, which therefore is not a maximum point.

So $f\left(2\sqrt{\frac{2}{5}}, -\sqrt{\frac{2}{5}}\right) = 5\sqrt{\frac{2}{5}}$ and $f\left(-2\sqrt{\frac{2}{5}}, \sqrt{\frac{2}{5}}\right) = -5\sqrt{\frac{2}{5}}$ are, respectively, maximum and minimum for $f(x,y)$ in $\mathcal{E}$.

KUHN-TUCKER'S CONDITIONS WITH NON NEGATIVE VARIABLES

In many problems, especially economic ones, the search for the extremes of a function is accompained by the request of non-negative values for the independent variables; in this context, we can have a slightly different formulation for Kuhn-Tucker's conditions.

$$\begin{cases} \text{Max } f(x_1, x_2, ..., x_n) \\ \text{u.c. : } \begin{cases} g_i(x_1, x_2, ..., x_n) \le 0, 1 \le i \le m \\ x_i \ge 0 \Rightarrow -x_i \le 0, 1 \le i \le n \end{cases} \end{cases}$$. If $\mathbb{X} = (x_1, x_2, ..., x_n)$ and:

$$\widetilde{\Lambda}\,(\mathbb{X}, \lambda_1, .., \lambda_m, \mu_1, .., \mu_n) = f(\mathbb{X}) - \sum_{i:1}^{m} \lambda_i\, g_i(\mathbb{X}) - \sum_{i:1}^{n} \mu_i\,(-x_i) = \Lambda + \sum_{i:1}^{n} \mu_i\, x_i$$

imposing Kuhn-Tucker's conditions for a maximum point to multipliers λ_i and μ_i, we get:

$$\begin{cases} \dfrac{\partial \widetilde{\Lambda}}{\partial x_i} = \dfrac{\partial \Lambda}{\partial x_i} + \mu_i = 0 \Rightarrow \mu_i = -\dfrac{\partial \Lambda}{\partial x_i} \\ \dfrac{\partial \widetilde{\Lambda}}{\partial \lambda_i} = \dfrac{\partial \Lambda}{\partial \lambda_i} = -g_i(\mathbb{X}) \ge 0 \\ \dfrac{\partial \widetilde{\Lambda}}{\partial \mu_i} = x_i \ge 0 \end{cases}.$$

Having then to impose $\mu_i \ge 0$, we rewrite the first condition as $\dfrac{\partial \Lambda}{\partial x_i} \le 0$.

We also impose other conditions: $\begin{cases} \lambda_i \cdot g_i(\mathbb{X}) = 0 \\ \mu_i \cdot x_i = 0 \\ \lambda_i \ge 0 \\ \mu_i \ge 0 \end{cases}$. As $\mu_i = -\dfrac{\partial \Lambda}{\partial x_i}$, the second condition is also expressed in the form: $\dfrac{\partial \Lambda}{\partial x_i} \cdot x_i = 0$, like the fourth, which is equivalent to $\dfrac{\partial \Lambda}{\partial x_i} \le 0$, already seen above. So Kuhn-Tucker's conditions for a maximum, subject to constraints $g_i(\mathbb{X}) \le 0$ and under the nonnegativity condition of the independent variables, can be expressed as:

$$\begin{cases} \dfrac{\partial \Lambda}{\partial x_i} \le 0 \\ g_i(\mathbb{X}) \le 0 \\ x_i \ge 0 \\ \lambda_i \cdot g_i(\mathbb{X}) = 0 \\ \dfrac{\partial \Lambda}{\partial x_i} \cdot x_i = 0 \\ \lambda_i \ge 0 \end{cases}.$$