MATHEMATICS, english version

A. Brini

October 1, 2010

Contents

Part I

The euclidean space $\mathbf{R}^n$

1 The euclidean space $\mathbf{R}^n$ as a metric space

1.1 The euclidean metric (distance) in $\mathbf{R}^n$

Let us consider the set of of all ordered $n-$tuples

$$\mathbf{R}^n = \{(x_1, x_2, \dots, x_n); x_i \in \mathbf{R}, i = 1, \dots, n\}.$$

The set $\mathbf{R}^n$ is endowed with a structure of finite dimensional (real) vector space of dimension n in the usual way.

From now on, we will regard $\mathbf{R}^n$ as an *euclidean metric space*, that is a set endowed with a function d, called the *euclidean metric (distance)*, defined as follows.

given two vectors (points) $\underline{x} = (x_1, x_2, \dots, x_n)$, $\underline{x'} = (x'_1, x_2, \dots, x_n) \in \mathbf{R}^n$, we set:

$$d(\underline{x}, \underline{x'}) = \sqrt{(x_1 - x'_1)^2 + \cdots + (x_n - x'_n)^2}.$$

The euclidean metric satisfies the axioms of "abstract" metric, that is:

1. $d(\underline{x}, \underline{x'}) \geq 0$; $d(\underline{x}, \underline{x'}) = 0$ *se e solo se* $\underline{x} = \underline{x'}$.
2. $d(\underline{x}, \underline{x'}) = d(\underline{x'}, \underline{x})$.
3. $d(\underline{x}, \underline{x'}) \leq d(\underline{x}, \underline{x''}) + d(\underline{x''}, \underline{x'})$ (THE TRIANGULAR INEQUALITY).

1.2 Basic topological concepts in $\mathbf{R}^n$

Given $x_0 \in \mathbf{R}^n$ and $r \in \mathbf{R}^+$, the *spherical open neighbourhood with center* x_0 *and ray* r is the subset

$$I(x_0, r) = \{x \in \mathbf{R}^n; d(x, x_0) < r\} \subseteq \mathbf{R}^n.$$

A subset$A \subseteq \mathbf{R}^n$ is said to be *limited* if and only if its *diameter*

$$d(A) = sup\ \{d(x,x'); x, x' \in A\}$$

is a FINITE real number. A more "geometric" and intuitive equivalent definition is the following: A is a limited subset if and only if

$$\forall x \in \mathbf{R}^n\ \exists\ r \in \mathbf{R}^+\ such\ that\ A \subseteq I(x,r).$$

A subset $A \subseteq \mathbf{R}^n$ is said to be *open* if and only if

$$\forall\ x \in A\ \exists\ r \in \mathbf{R}^+\ such\ that\ I(x,r) \subseteq A.$$

Given $B \subseteq \mathbf{R}^n$ and a point $x_0 \in \mathbf{R}^n$, we say that x_0 is an *accumulation point (or, limit point) for* B if and only if

$$\forall\ r \in \mathbf{R}^+\ we\ have\ (I(x_0,r) - \{x_0\}) \cap B \neq \emptyset.$$

A subset $D \subseteq \mathbf{R}^n$ is said to be *closed* if and only if its complementary subset $D^c = \mathbf{R}^n - D$ is an open set in $\mathbf{R}^n$.

Proposition 1. *A subset $D \subseteq \mathbf{R}^n$ is closed if and only if it contains its accumulation points.*

The proof is left as an (useful) exercise.

REMARK 1.

- *In general, given a subset $C \subseteq \mathbf{R}^n$ it may be neither an open set nor a closed set. For example, the interval $[0,1[$ is neither an open subset nor a closed subset of the euclidean line $\mathbf{R}$. On the one hand, the point 1 is an accumulation point for $[0,1[$ but it doesn't belong to it; On the other hand, it isn't possible to find a neighbourhood with center 0 and contained in $[0,1[$.*
- *The subsets $\emptyset$ and $\mathbf{R}^n$ are simultaneously open and closed subsets.*

In the *euclidean* spaces $\mathbf{R}^n$, the following deep result holds:

Therem 1. *In the euclidean spaces $\mathbf{R}^n$, if a subset A is both an open subset and a closed subset, then $A =$ or $A = \mathbf{R}^n$.*

Corollary 1. *Let $C \subseteq \mathbf{R}^n$, $C \neq \emptyset$, $C \neq \mathbf{R}^n$.*

- *if C is open, then C is not closed.*
- *If C is closed, then C is not open.*

Proposition 2.

- *O_1) The union of any family of open sets is an open set.*
- *O_2) The intersection of a FINITE family of open sets is an open set.*
- *C_1) The union of a FINITE family of closed sets is a closed set.*

- C_2) *The intersection of any family of closed sets is a closed set.*

REMARK 2. *By the "De Morgan laws", assertion $O1$ is equivalent to assertion $C2$, and assertion $O2$ is equivalent to assertion $C1$.*

COUNTEREXAMPLES IN R

- $\cap_{n\in\mathbf{Z}^+}]-\frac{1}{n},\frac{1}{n}[= \{0\}$ is an intersection of open sets, but it is closed, and, therefore, *NOT* open.
- $\cup_{n\in\mathbf{Z}^+} [-1+\frac{1}{n},1-\frac{1}{n}] =]-1,1[$ is a union of closed sets, but it is open, therefore, *NOT* closed.

1.3 Continuous functions $f : A \subseteq \mathbf{R}^n \to \mathbf{R}$

Let $f : A \subseteq \mathbf{R}^n \to \mathbf{R}$ - $A \neq \emptyset$ - be a real valued function. In the following, we will assume that the *domain* $A \subseteq \mathbf{R}^n$ is a metric space with respect to the metric induced - by restriction - by the euclidean metric of $\mathbf{R}^n$.

Given a point $x_0 \in A$, we say that f is *continuous at the point* x_0 if and only if

$$\forall\varepsilon \in \mathbf{R}^+,\ \exists\ \delta \in \mathbf{R}^+ \textit{ such that } |f(x) - f(x_0)| < \varepsilon,\ \forall x \in I(x_0, \delta) \cap A.$$

This is a "local" definition; "globally", we say that f is *continuous on* A if and only if it is continuous at *every* point of A.

We have a quite useful characterization of the global continuity of a function

$$f : A \subseteq \mathbf{R}^n \to \mathbf{R}.$$

For the convenience of the reader, we recall the notion of *preimage* (or, *fiber*) of a subset $C \subseteq \mathbf{R}$ with respect to the function f.

Given $C \subseteq \mathbf{R}$, its preimage with respect to the function f is the set:

$$f^{-1}[C] = \{x \in A; f(x) \in C\} \subseteq A \subseteq \mathbf{R}^n.$$

Therem 2. *Consider a function $f : A \subseteq \mathbf{R}^n \to \mathbf{R}$, $A \neq \emptyset$.*

The following assertions are equivalent:

- *f is continuous on A.*
- *$\forall\ B \subseteq \mathbf{R}$, B open in $\mathbf{R}$, $\exists\ B_1 \subseteq \mathbf{R}^n$, B_1 open in $\mathbf{R}^n$, such that $f^{-1}[B] = A \cap B_1$.*
- *$\forall\ D \subseteq \mathbf{R}$, D closed in $\mathbf{R}$, $\exists\ D_1 \subseteq \mathbf{R}^n$, D_1 closed in $\mathbf{R}^n$, such that $f^{-1}[D] = A \cap D_1$.*

Corollary 2. *Consider a function $f : \mathbf{R}^n \to \mathbf{R}$.*

The following assertions are equivalent:

- *f is continuous on $\mathbf{R}^n$.*
- *$\forall\ B \subseteq \mathbf{R}$, B open in $\mathbf{R}$, $f^{-1}[B]$ is open in $\mathbf{R}^n$.*
- *$\forall\ D \subseteq \mathbf{R}$, D closed in $\mathbf{R}$, $f^{-1}[D]$ is closed in $\mathbf{R}^n$.*

1.4 The Theorem of Weierstrass

Una fondamentale conseguenza della continuita' globale e' data dal seguente risultato.

Therem 3. *(Weierstrass)*

Consider a function $f : A \subseteq \mathbf{R}^n \to \mathbf{R}$, $A \neq \emptyset$.

if A *is a closed and limited set (i.e.,* A *is a compact set), then the image* $f[A] = \{f(x); x \in A\} \subseteq \mathbf{R}$ *has a maximum and a minimum.*

1.5 Sequences in $\mathbf{R}^n$; convergent sequences and Cauchy sequences

Let $(x_n)_{n\in\mathbf{N}}$ be a *sequence* with elements in $\mathbf{R}^n$, that is $x_n \in \mathbf{R}^n$, for every $n \in \mathbf{N}$.

We say that the sequence $(x_n)_{n\in\mathbf{N}}$ *converges* to the point $x_0 \in \mathbf{R}^n$ if and only if

$$\forall\, \varepsilon \in \mathbf{R}^+,\ \exists\, \nu \in \mathbf{N} \textit{ such that } d(x_n, x_0) < \varepsilon,\ \forall n > \nu.$$

In this case, we also write, in short notation,

$$\lim_{n\to\infty} x_n = x_0.$$

In this case, we also say that the sequence $(x_n)_{n\in\mathbf{N}}$ is *convergent.*

A sequence $(x_n)_{n\in\mathbf{N}}$ with elements in $\mathbf{R}^n$ is said to be a *a Cauchy sequence* if and only if

$$\forall\, \varepsilon \in \mathbf{R}^+,\ \exists\, \nu \in \mathbf{N} \textit{ such that } d(x_n, x_m) < \varepsilon,\ \forall n, m > \nu.$$

From the "triangular inequality", it follows (useful and easy exercise) that *any convergent sequence is a Cauchy sequence.*

The converse assertion is NOT in general true, in abstract (general) metric spaces.

However, in the euclidean space $\mathbf{R}^n$, the following fundamental result is valid:

Therem 4. *(The Completeness Theorem for the euclidean space* $\mathbf{R}^n$*)*

A sequence $(x_n)_{n\in\mathbf{N}}$ *with elements in* $\mathbf{R}^n$ *is convergent if and only if it is a Cauchy sequence.*

In general, a metric space is said to be *complete* if and only se any Cauchy sequence is convergent (in general, this implication is FALSE; for example, in the space $\mathbf{Q}$ of rational numbers, endowed with the restriction of the euclidean metric of $\mathbf{R}$, this assertion is false.

We have an important characterization of the continuity of a function f at a point x_0 in terms of convergent sequences.

Proposition 3. *Consider a function* $f : A \subseteq \mathbf{R}^n \to \mathbf{R}$, $A \neq \emptyset$*, and a point* $x_0 \in A$.

The function f *is continuous at the point* $x_0 \in A$ *if and only if for every sequence* $(x_n)_{n\in\mathbf{N}}$ *in* A *the condition*

$$\lim_{x_n\to x_0} x_n = x_0$$

implies

$$\lim_{x_n\to x_0} f(x_n) = f(x_0)$$

2 The euclidean space $\mathbf{R}^n$ as a normed space

From now on, we will regard $\mathbf{R}^n$ also as a *normed space*, that is, we will endow $\mathbf{R}^n$ with a function, called*euclidean norm*:

$$|| \ || : \mathbf{R}^n \to \mathbf{R}$$

defined as follows. Given $\underline{x} = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n$, we set:

$$||\underline{x}|| = \sqrt{x_1^2 + \cdots + x_n^2}.$$

The euclidean norm satisfies the axioms of "abstract" norm, that is:

1. $||\underline{x}|| \geq 0$; $||\underline{x}|| = 0$ *se e solo se* $\underline{x} = \underline{0}$.
2. $||\lambda \cdot \underline{x}|| = |\lambda| \cdot ||\underline{x}||, \ \lambda \in \mathbf{R}$.
3. $||\underline{x} + \underline{x}'|| \leq ||\underline{x}|| + ||\underline{x}'||$.

2.1 Norms and metrics

In general, given a norm function $|| \ ||$, we can canonically associate to it two-variable function $d : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ by setting

$$d(x, x') = ||x - y||;$$

it is not difficult to prove that the function d is a metric.

In plain words, the normed space $(\mathbf{R}^n, || \ ||)$ may be regarded as a metric space. This is a general fact about normed spaces. In the case "euclidean" $\mathbf{R}^n$, the situation is extremely simple and intuitive.

Given $\underline{x} = (x_1, x_2, \ldots, x_n)$, $\underline{x'} = (x_1', x_2, \ldots, x_n) \in \mathbf{R}^n$, and denoted by $|| \ ||$ the euclidean norm, we have

$$||\underline{x} - \underline{x'}|| = \sqrt{(x_1 - x_1')^2 + \cdots + (x_n - x_n')^2} = d(\underline{x}, \underline{x'}),$$

that is, the euclidean metric.

3 The euclidean space $\mathbf{R}^n$ as a space with inner product

From now on, we will regard $\mathbf{R}^n$ also as a *space with inner product*, that is, we will endow $\mathbf{R}^n$ with a function, called*euclidean inner product*

$$< \ >: \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$$

defined as follows. Given $\underline{x} = (x_1, x_2, \ldots, x_n)$, $\underline{x'} = (x'_1, x_2, \ldots, x_n) \in \mathbf{R}^n$, we set:

$$< \underline{x}, \underline{x'} >= x_1 x'_1 + \cdots + x_n x'_n = \sum_{i=1}^{n} x_i x'_i.$$

The euclidean inner product satisfies the axioms of "abstract" inner product, that is:

1. $$< \underline{x} + \underline{x'}, \underline{y} > \; = \; < \underline{x}, \underline{y} > + < \underline{x'}, \underline{y} >,$$

   e

   $$< \underline{x}, \underline{y} + \underline{y'} > \; = \; < \underline{x}, \underline{y} > + < \underline{x}, \underline{y'} >,$$

2. $$< \lambda \cdot x, y >= < x, \lambda \cdot y > \; = \lambda \cdot < \underline{x}, \underline{y} >, \quad \lambda \in \mathbf{R}$$

3. $$< \underline{x}, \underline{y} > \; = \; < \underline{y}, \underline{x} > .$$

4. $$< \underline{x}, \underline{x} > \; \geq 0; \; < \underline{x}, \underline{x} > \neq 0 \textit{ per ogni } x \neq \underline{0}.$$

The properties 1) e 2) are also also expressed by saying that the inner product is *bilinear*, the property 3) by saying that it is *symmetric*, the property 4) by saying that it is *positively defined*.

3.1 Geometric interpretation: inner products and angles

The euclidean inner product in $\mathbf{R}^n$ allows us to provide a transparent and rigorous definition of the notion of *angle* between a pair of non-zero vectors in $\mathbf{R}^n$.

The starting point is the following result.

Therem 5. *(The Cauchy-Schwarz inequality)*
Let x, y be vectors in $\mathbf{R}^n$. Then:

$$| < x, y > | \; \leq \sqrt{< x, x >} \sqrt{< y, y >} \; = \; ||x|| \cdot ||y||.$$

In geometric language, we may interpret the Cauchy-Schwarz inequality in the following way.

Let x, y non-zero vectors in $\mathbf{R}^n$. The preceding inequality may be rewritten in the form

$$-1 \; \leq \; \frac{< x, y >}{||x|| \cdot ||y||} \; \leq 1.$$

Thus, there exists a*unique* ureal number $\theta \in [0, \pi]$ such that

$$\cos(\theta) \; = \; \frac{< x, y >}{||x|| \cdot ||y||}.$$

The value θ is called the *angle* between the non-zero vectors $x, y \in \mathbf{R}^n$.

Example 1. *let* $x = (x_1, 0), y = (0, y_2)$ *be vectors in* $\mathbf{R}^2$*,* $x_1, y_2 \neq 0$*. Then*

$$< x, y > = 0 = \cos(\frac{\pi}{2}),$$

that is the vectors x *e* y *are orthogonal.*

3.2 Inner products, norms and metrics

In general, given an inner product function $< >$, the one variable function defined by setting $||x|| = \sqrt{< x, x >}$ turns out to be a norm function.

In plain words, the inner product space $(\mathbf{R}^n, < >)$ may be¡ regarded as a normed space. This is a general fact about inner product spaces. In the case "euclidean" $\mathbf{R}^n$, the situation is extremely simple and intuitive.

Given $\underline{x} = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n$, and denoted by $< , >$ the euclidean inner product, we have:

$$\sqrt{< \underline{x}, \underline{x} >} = \sqrt{x_1^2 + \cdots + x_n^2} = ||\underline{x}||,$$

that is the euclidean norm.

In plain words, we say that the norm (canonically) associated to the euclidean inner product is the euclidean norm. Furthermore, we recall that the metric associated to the euclidean norm is the euclidean norm.

Thus, all the concepts we introduced for $\mathbf{R}^n$ as an euclidean metric space (for example, limited set, open sets, closed sets, continuous functions, convergent and Cauchy sequences, completeness) preserve their meaning when we regard $\mathbf{R}^n$ as a space endowed with the euclidean inner product.

Part II

Real-valued differentiable functions. Local maximum and minimum points

4 Differentiable functions $f : A\ (A \subset \mathbf{R}^n) \to \mathbf{R}$

4.1 Directional derivatives and partial derivatives

A vector $v \in \mathbf{R}^n$ is said to be a *direction, (or, versor)* if and only if its euclidean norm equals 1: in symbols, $||v|| = 1$.

if $\underline{x} \in \mathbf{R}^n$ e v is a direction, the set

$$r_{\underline{x},v} = \{x \in \mathbf{R}^n; x = \underline{x} + tv,\ t \in \mathbf{R}\}$$

is a LINE, in particular the(*unique*) line passing through the point $\underline{x}$ and with direction v.

Let$A$ be an open set of $\mathbf{R}^n$, $\underline{x}$ A, and let us consider a function

$$f : A \to \mathbf{R}.$$

We say that f e' has the *directional derivative in the direction* v *at the point* $\underline{x}$ A if and only if the limit

$$\lim_{t \to 0 \in \mathbf{R}} \frac{f(\underline{x} + tv) - f(\underline{x})}{t}$$

EXISTS and is FINITE (i.e., is a real number).

This limit, when it exists and is finite, is usually denoted by the symbol

$$\frac{\partial\, f}{\partial\, v}\ (\underline{x}),$$

and is called the directional derivative in the direction v at the point $\underline{x}$ A.

If v is a vector of the "canonical basis" $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$, the directional derivative in the direction $\mathbf{e}_i$ is called the $i-$th *PARTIAL DERIVATIVE* and is usually denote by the symbol

$$\frac{\partial\, f}{\partial\, x_i}\ (\underline{x}), D_i(f)(\underline{x}).$$

or, (short notation)

$$D_i(f)(\underline{x}).$$

OSSERVAZIONE FONDAMENTALE. In generale, se $n > 1$, le derivate direzionali (se esistono) sono *INFINITE*. Invece le derivate parziali (se esistono) sono in numero *FINITO* (al massimo n), cioe' la dimensione dello spazio dominio della funzione, o, equivalentemente, il numero delle variabili.

Consider a point (vector)

$$\underline{x} = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n) \in \mathbf{R}^n;$$

thus,

$$\underline{x} + t\mathbf{e}_i = (\underline{x}_1, \ldots, \underline{x}_i + t, \ldots, \underline{x}_n).$$

Given $i = 1, \ldots, n$, we have (direct computation)

$$\frac{\partial\, f}{\partial\, x_i}\ (\underline{x}) \ = \lim_{t \to 0 \in \mathbf{R}} \frac{f(\underline{x} + t\mathbf{e}_i) - f(\underline{x})}{t} =$$

$$\lim_{t \to 0 \in \mathbf{R}} \frac{f(\underline{x}_1, \ldots, \underline{x}_i + t, \ldots, \underline{x}_n) - f(\underline{x}_1, \ldots, \underline{x}_i, \ldots, \underline{x}_n)}{t}.$$

Therefore, the $i-$th partial derivative $\frac{\partial\, f}{\partial\, x_i}$ $(\underline{x})$ (at the point $\underline{x}$) is computed by regarding the variables $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ as *COSTANTS*, and by differentiating (as in the one variable case) with respect to the variable x_i.

4.2 Differentiable functions $f : A \to \mathbf{R}$

Given a function $f : A \to \mathbf{R}$, A open subset of $\mathbf{R}^n$, f is said to be *DIFFERENTIABLE* at the point $\underline{x} \in A$ if and only if there exists a *linear functional*

$$L_{\underline{x}} : \mathbf{R}^n \to \mathbf{R}$$

(depending from the point $\underline{x}$) such that

$$\lim_{h \to \underline{0} \in \mathbf{R}^n} \frac{f(\underline{x}+h) - f(\underline{x}) - L_{\underline{x}}(h)}{||h||} \ = \ 0.$$

The linear functional $L_{\underline{x}}$ is called the *differential* of f at the point $\underline{x}$, and it is also denoted by the symbol $df(\underline{x})$.

If f is differentiable at every point of A, we will say that f is differentiable in A.

4.3 "Geometric" interpretation of the differentiability condition

Let $f : A \to \mathbf{R}$ be a differentiable function at the point $\underline{x} \in A$,

and let $L_{\underline{x}}$ denote its its differential at the point $\underline{x}$.

Consider the numerator of the fraction that appears in the preceding limit, that is

$$E_{\underline{x}}(h) \ = \ f(\underline{x}+h) - f(\underline{x}) - L_{\underline{x}}(h),$$

regarded as a function of the "vector increment" $h \in \mathbf{R}^n$.

The value $E_{\underline{x}}(h)$ may be interpreted as the *error* (again, as a function of the "vector increment" $h \in \mathbf{R}^n$) that we make when we "approximate" that value of the function f (at the point $\underline{x}+h$) with the value

$$f(\underline{x}) + L_{\underline{x}},$$

the sum of a constant $f(\underline{x})$ and the evaluation of a linear function $L_{\underline{x}}$.

On the other hand, the denominator $||h|| = ||(\underline{x}+h) - h||$ is the distance of the point $\underline{x}+h$ from the "base point" $\underline{x}$.

Thus, the differentiability condition for f at the point $\underline{x}$ may be rewritten in the following way:

$$\exists L_{\underline{x}} : \mathbf{R}^n \to \mathbf{R} \quad \textit{linear}$$

and

$$E_{\underline{x}} \quad \textit{function of } h \textit{ in a neighbourhood of } \underline{x}$$

such that

$$f(\underline{x}+h) = (f(\underline{x}) + L_{\underline{x}}(h)) + E_{\underline{x}}(h),$$

with

$$\lim_{h \to \underline{0} \in \mathbf{R}^n} \frac{E_{\underline{x}}(h)}{||h||} = 0.$$

IN PLAIN WORS: near $\underline{x}$, we may approximate f by the polynomial $f(\underline{x}) + L_{\underline{x}}$ by making an error $E_{\underline{x}}$ (function of $||h|| = d(\underline{x}+h, \underline{x})$) that, if $h \to \underline{0}$, "goes to 0 more quickly" than the norm of h.

4.4 Differentials and directional derivatives

Proposition 4. *Let A be an open subset of $\mathbf{R}^n$, and let $f : A \to \mathbf{R}$ be a function differentiable at $\underline{x} \in A$. Then f admits, at $\underline{x}$, all the directional derivatives $\frac{\partial f}{\partial v}(\underline{x})$, for every direction v. Furthermore*

$$\frac{\partial f}{\partial v}(\underline{x}) = L_{\underline{x}}(v),$$

for every direction v.

DIMOSTRAZIONE. Let v be a given direction in $\mathbf{R}^n$. *By specializing* the differentiability condition to the case $h = tv$, we get

$$\lim_{t \to 0 \in \mathbf{R}} \frac{f(\underline{x} + tv) - f(\underline{x}) - L_{\underline{x}}(tv)}{|t|} = 0.$$

This condition is EQUIVALENT to the condition

$$\lim_{t \to 0 \in \mathbf{R}} \frac{f(\underline{x} + tv) - f(\underline{x}) - L_{\underline{x}}(tv)}{t} = 0.$$

(WHY? Use the definition of limit)

Since $L_{\underline{x}}$ is a linear functional, the last condition is equivalent to the condition

$$\lim_{t \to 0 \in \mathbf{R}} \frac{f(\underline{x} + tv) - f(\underline{x}) - t \cdot L_{\underline{x}}(v)}{t} = 0.$$

Hence,

$$\lim_{t \to 0 \in \mathbf{R}} \frac{f(\underline{x} + tv) - f(\underline{x})}{t} = \lim_{t \to 0 \in \mathbf{R}} \frac{t \cdot L_{\underline{x}(v)}}{t} = L_{\underline{x}(v)};$$

thus,

$$\lim_{t \to 0 \in \mathbf{R}} \frac{f(\underline{x} + tv) - f(\underline{x})}{t}$$

exists, is finite, and equals the evaluation $L_{\underline{x}}(v)$ of the differential on the direction v.

Therefore, the directional derivative $\frac{\partial f}{\partial v}(\underline{x})$ exists and, furthermore, we have proved the crucial identity

$$\frac{\partial f}{\partial v}(\underline{x}) = L_{\underline{x}}(v).$$

4.5 Evaluations of a differential, gradient vector, inner products

Let A be an open set of $\mathbf{R}^n$, and let $f : A \to \mathbf{R}$ be differentiable at $\underline{x} \in A$, $L_{\underline{x}}$ the differential of f at the point $\underline{x} \in A$.

We already know that f has all the directional derivatives at$\underline{x} \in A$ and then, a fortiori, all the n partial derivatives.

The vector

$$grad\ f(\underline{x}) \;=\; (\ \frac{\partial\ f}{\partial\ x_1}\ (\underline{x}), \frac{\partial\ f}{\partial\ x_2}\ (\underline{x}), \ldots, \frac{\partial\ f}{\partial\ x_n}\ (\underline{x})\ ) \in \mathbf{R}^n$$

is called the *gradient vector* of f at $\underline{x}$.

Let $v \;=\; (v_1, v_2, \ldots, v_n) \;=\; \sum_{i=1}^n\ v_i \cdot \mathbf{e}_i$ be a vector in $\mathbf{R}^n$.

(Recall: $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ denotes the canonical basis of $\mathbf{R}^n$)

By linearity, the evaluation on $v \in \mathbf{R}^n$ of the differential f at the point $\underline{x} \in A$ equals

$$L_{\underline{x}}(v) = \sum_{i=1}^{n}\ v_i \cdot L_{\underline{x}}(\mathbf{e}_i) = \sum_{i=1}^{n}\ v_i \cdot\ \frac{\partial\ f}{\partial\ x_i}\ (\underline{x}).$$

Therefore, the evaluation of the differential $L_{\underline{x}}$ on a given vector $v \in \mathbf{R}^n$ may be expressed, in quite concise way, as the inner product:

$$< gradf(\underline{x}), v > = \sum_{i=1}^{n}\ v_i \cdot\ \frac{\partial\ f}{\partial\ x_i}\ (\underline{x}) \;\;=\;\; L_{\underline{x}}(v).$$

Corollary 3. *Let A be an open set of $\mathbf{R}^n$, e sia $f : A \to \mathbf{R}$ differentiable at $\underline{x} \in A$. Then*

$$\frac{\partial\ f}{\partial\ v}\ (\underline{x}) \;= < gradf(\underline{x}), v >,$$

for every direction $v \in \mathbf{R}^n$.

4.6 Directional derivatives and differentials

If $n > 1$, the converse of Proposition 4 is FALSE.

A function f may have all the directional derivatives at a point $\underline{x}$ without being differentiable in such a point!

The differentiability condition is *strictly stronger* than the condition of having all the directional derivatives at a point (A simple way to understand and remember this important fact is the following: the differentiability condition at a point $\underline{x}$ implies the continuity at this point - as we shall see in a while -, but a function may have all the directional derivatives at a point $\underline{x}$ without being continuous in such point.

4.7 Differentiability and continuity

Proposition 5. *Let A be an open subset of $\mathbf{R}^n$, and let $f : A \to \mathbf{R}$ be a function differentiable at $\underline{x} \in A$. Then f is continuous in $\underline{x}$.*

Proof. Since f is differentiable at $\underline{x}$, we ahve

$$f(\underline{x} + h) = f(\underline{x}) + L_{\underline{x}}(h) + E_{\underline{x}}(h),$$

with

$$\lim_{h \to \underline{0} \in \mathbf{R}^n} \frac{E_{\underline{x}}(h)}{||h||} = 0.$$

Now

$$f(\underline{x}+h) - f(\underline{x}) = L_{\underline{x}}(h) + E_{\underline{x}}(h) =< grad\ f(\underline{x}), h > + E_{\underline{x}}(h);$$

thus,

$$|f(\underline{x}+h) - f(\underline{x})| \leq |< grad\ f(\underline{x}), h >| + |E_{\underline{x}}(h)|.$$

From the Cauchy-Schwarz inequality, it follows

$$|< grad\ f(\underline{x}), h >| \leq ||grad\ f(\underline{x})|| \cdot ||h||;$$

ne segue

$$|f(\underline{x}+h) - f(\underline{x})| \leq ||grad\ f(\underline{x})|| \cdot ||h|| + |E_{\underline{x}}(h)|.$$

Clearly $0 \leq |f(\underline{x}+h) - f(\underline{x})|$ and it is *less or equal to the sum of two functions that tend to* 0 for $h \to \underline{0} \in \mathbf{R}^n$. Then, $|f(\underline{x}+h) - f(\underline{x})| \to 0$ for $h \to \underline{0} \in \mathbf{R}^n$, that is f is continuous in $\underline{x}$.

The case of $||grad\ f(\underline{x})|| \cdot ||h||$ is trivial, by definition. For $|E_{\underline{x}}(h)|$, we argue as follows:

$$\lim_{h \to \underline{0} \in \mathbf{R}^n} \frac{E_{\underline{x}}(h)}{||h||} = 0 \Rightarrow \lim_{h \to \underline{0} \in \mathbf{R}^n} E_{\underline{x}}(h) = 0 \Leftrightarrow \lim_{h \to \underline{0} \in \mathbf{R}^n} |E_{\underline{x}}(h)| = 0.$$

□

4.8 The Total Differential Theorem

Therem 6. *Let A be an open subset of $\mathbf{R}^n$, and let $f : A \to \mathbf{R}$ be a function differentiable at $\underline{x} \in A$.*

Assume that f has all its partial derivatives $\underline{i}$n a neighbourhood of the point $\underline{x}$ and that these partial derivatives are continuous in $\underline{x}$.

Then f is differentiable at $\underline{x}$.

4.9 "Locally good" algebraic operations among differentiable functions

Proposition 6. *Let A be an open subset of $\mathbf{R}^n$, and let $f, g : A \to \mathbf{R}$ be differentiable functions at $\underline{x} \in A$. Then :*

- *$f + g$ is a differentiable function at $\underline{x}$. Furthermore*

$$d(f+g)(\underline{x}) \ = \ d(f)(\underline{x}) + d(g)(\underline{x}).$$

- *$f \cdot g$ be a differentiable function at $\underline{x}$. Furthermore*

$$d(f \cdot g)(\underline{x}) \ = \ d(f)(\underline{x}) \cdot g(\underline{x}) + f(\underline{x}) \cdot d(g)(\underline{x}).$$

- *or every $\lambda \in \mathbf{R}$, λf is a differentiable function at $\underline{x}$. Furthermore*

$$d(\lambda f)(\underline{x}) \ = \lambda \cdot d(f)(\underline{x}).$$

4.10 Mixed (successive) partial derivatives

We start with an elementary example. Consider the function:

$$f : \mathbf{R}^2 \to \mathbf{R}, \quad f(x,y) = xy + |y|.$$

Note that the partial derivative $\frac{\partial f}{\partial x} = y$ is defined at every point of $\mathbf{R}^2$, it is continuous in any point (as a function of the variables x e y) and admits *derivative* derivative with respect to the variable y.

Specifically, we have:

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = 1$$

at any point of the domain $\mathbf{R}^2$.

HOWEVER, the partial derivative

$\frac{\partial}{\partial y}$ with respect to the variable y DOESN'T EXIST in the points such that $y = 0$ and, therefore, it doesn't exist the mixed (successive) derivative

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right).$$

Therefore, the identity

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)$$

is, in general, FALSE.

In plain words, it is not true - in general - that if we "exchange the order of successive partial derivation" we obtain the same result; it may happen that a successive partial derivative exist in some points (in the previous example, $\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)$), the other one doesn't exist (in our example, $\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)$, per $y = 0$).

However, under suitable hypothesis, a fundamental "exchangeability result" holds:

Therem 7. *(Schwartz) Let A be an open subset of $\mathbf{R}^n$, and let $f : A \to \mathbf{R}$ be a differentiable functions at $\underline{x} \in A$, $\underline{x} = (\underline{x}_1, \ldots, \underline{x}_n) \in A$.*

If, given $i, j = 1, 2; \ldots, , n$, the successive derivatives

$$\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right), \quad \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)$$

exist in a NEIGHBOURHOOD of $\underline{x}$ and they are continuous in $\underline{x}$, then their evaluations at the point $\underline{x}$ are the same. In symbols:

$$\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)(\underline{x}) = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(\underline{x}).$$

4.11 Derivation of iterated functions

Let A be an open set of $\mathbf{R}^n$, $f : A \to \mathbf{R}$. Let $r : I \to \mathbf{R}^n$, $I \subset \mathbf{R}$ be an open interval,

$$r(t) = (r_1(t), r_2(t), \ldots, r_n(t))$$

such that the image of r

$$r[I] = \{(r_1(t), r_2(t), \ldots, r_n(t)) \in \mathbf{R}^n; t \in I\}$$

is contained in A.

Let $t_0 \in I$ be a point such that:

- $r_i(t)$ admit the (standard) derivative at t_0, for every $i = 1, 2, \ldots, n$.
- f is differentiable at $r(t_0) \in A$.

Then, setting

$$g(t) = f(r_1(t), r_2(t), \ldots, r_n(t))$$

for every $t \in I$, the *iterated function*

$$g = f \circ r :\to \mathbf{R}, \quad g : t \mapsto g(t)$$

admit the (standard) derivative at t_0. Furthermore, we have:

$$g'(t_0) =< grad\ f(r(t_0)), (r_1'(t_0), r_2'(t_0), \ldots, r_n'(t_0)) >= \sum_{i=1}^{n} \frac{\partial f(r(t_0))}{\partial x_i} r_i'(t_0).$$

4.12 Polinomi di Taylor di grado k

Siano A un aperto di $\mathbf{R}^n$, $f : A \to \mathbf{R}$. Sia f di classe $C^{(k+1)}$.

Fissato il punto base $\underline{x}$, ci chiediamo come costruire il $k-$esimo polinomio approssimante di Taylor $T_k(\underline{x} + h)$ (di grado k e relativo alla scelta del punto base $\underline{x}$).

Tale polinomio e' univocamente determintato, ed e' il polinomio:

$$T_k(\underline{x} + h) = \sum_{p=0}^{k} \sum_{(i_1, i_2, \ldots, i_n),\ i_1+i_2+\ldots+i_n=p} \frac{1}{(i_1)!(i_2)!\cdots(i_n)!} \frac{\partial^{(p)}\ f(\underline{x})}{\partial x_1^{i_1} \partial x_2^{i_2} \cdots \partial x_n^{i_n}} h_1^{i_1} h_2^{i_2} \cdots h_n^{i_n}.$$

Proposition 7. *Per ogni $p \leq k$, il polinomio di Taylor T_k ha in $h = 0$ (o, equivalentemente, riguardato come funzione di $\underline{x} + h$, cioe' in $\underline{x}$) le stesse derivate seccessive di ordine p di f in $\underline{x}$, fino ad ordine $p \leq k$.*

Proposition 8. *Sia f di classe $C^{(k)}$.*

Il resto $k-$esimo $R_k(\underline{x}; h)$ soddisfa la seguente proprieta' di andamento a ZERO:

$$\frac{R_k(\underline{x}; h)}{||h||^k} \to 0 \quad per \quad h \to \underline{0} \in \mathbf{R}^n$$

5 Local (or, relative) maximum and minimum points for a function $f : A \to \mathbf{R}$ on its domain A (A open set of $\mathbf{R}^n$)

5.1 Local (or, relative) maximum and minimum points for a function i several variables

Let $f : A \to \mathbf{R}$, $A \subset \mathbf{R}^n$ be a function in n variables.

Let $\underline{x} \in A$ be a point.

We say that $\underline{x}$ is a *local (relative maximum point* for f if and only if *there exist a neighbourhood* $U_{\underline{x}}$ of $\underline{x}$ in $\mathbf{R}^n$ such that

$$f(x) \leq f(\underline{x}),$$

for every $x \in A \cap U_{\underline{x}}$.

Analogously, we say that $\underline{x}$ is a *local (relative minimum point* for f if and only if *there exist a neighbourhood* $U_{\underline{x}}$ of $\underline{x}$ in $\mathbf{R}^n$ such that

$$f(x) \geq f(\underline{x}),$$

for every $x \in A \cap U_{\underline{x}}$.

5.2 Hessian Matrices

Let A be an open set of $\mathbf{R}^n$, $f : A \to \mathbf{R}$, and let $\underline{x} \in A$.Asumme that f is of class $C^{(2)}$.

The matrix

$$\mathbf{H}_{f(\underline{x})} = \begin{pmatrix} \frac{\partial^{(2)} f(\underline{x})}{\partial x_1^2} & \cdots\cdots & \frac{\partial^{(2)} f(\underline{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^{(2)} f(\underline{x})}{\partial x_2 \partial x_1} & \cdots\cdots & \frac{\partial^{(2)} f(\underline{x})}{\partial x_2 \partial x_n} \\ \vdots & & \\ \frac{\partial^{(2)} f(\underline{x})}{\partial x_n \partial x_1} & \cdots\cdots & \frac{\partial^{(2)} f(\underline{x})}{\partial x_n^2} \end{pmatrix} = \left(\frac{\partial^{(2)} f(\underline{x})}{\partial x_i \partial x_j} \right)_{i,j=1,\ldots,n}$$

is a *symmetric* matrix (Schwartz Theorem) and it is called the *HESSIAN MATRIX of the function* f *at the point* $\underline{x}$.

5.3 Some preliminary remarks and definitions

Let A be an open set of $\mathbf{R}^n$, $f : A \to \mathbf{R}$, and let $\underline{x} \in A$.

Let $\delta > 0$ be a real number such that $I(\underline{x}, \delta) \subseteq A$.

Given a vector (direction) $\underline{v} \in \mathbf{R}^n$, $\underline{v} \neq \underline{0}$, Consider the function (of the unique real variable t, with $|t| < \frac{\delta}{||\underline{v}||}$):

$$F(t) = f(\underline{x} + t\underline{v}) = (f \circ r)(t),$$

$$r(t) :] - \frac{\delta}{||\underline{v}||}, \frac{\delta}{||\underline{v}||}[\to \mathbf{R}, \quad r(t) = (r_1(t), \ldots, r_n(t))$$

where

$$r_i(t) = \underline{x}_i + t\underline{v}_i$$

$(\underline{x}_i = (\underline{x}_1, \ldots, \underline{x}_n), \quad \underline{v} = (\underline{v}_1, \ldots, \underline{v}_n))$.

- Let f of class $C^{(1)}$. Then F has the derivative at every point $t \in] - \frac{\delta}{||\underline{v}||}, \frac{\delta}{||\underline{v}||}[$ and, furthermore:

$$F'(t) =< grad\ f(r(t)), (r_1'(t), r_2'(t), \ldots, r_n'(t)) >= \sum_{i=1}^{n} \frac{\partial f(\underline{x} + t\underline{v})}{\partial x_i} \underline{v}_i. \qquad (*)$$

- Let f of class $C^{(2)}$. Then F has the second order derivative at every point $t \in] - \frac{\delta}{||\underline{v}||}, \frac{\delta}{||\underline{v}||}[$ and, furthermore:

$$F''(t) = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{x} + t\underline{v}))v_i v_j =< \underline{v} \times \mathbf{H}_{f(\underline{x}+t\underline{v})}, \underline{v} > .$$

Indeed, since f is of class $C^{(2)}$, the partial derivatives $\frac{\partial f}{\partial x_j}$ are of class $C^{(1)}$ and, therefore, they are differenziable; the assertion follows by applying the theorem about derivation of iterated functions to all summands in formula $(*)$.

5.4 Necessary conditions

Proposition 9. *Let A be an open set of $\mathbf{R}^n$, $f : A \to \mathbf{R}$, and let $\underline{x} \in A$.*

Lat $\underline{x} \in A$ be a local maximum (minimum) point for the fuction f.

If f is differentiable at $\underline{x}$, then the differential of f at $\underline{x}$ is the identically zero linear functional: in simbols

$$df(\underline{x}) \equiv 0.$$

Proof. If the point ia $\underline{x}$ a local maximum (minimum) point for the fuction f (and the function is differentiable at $\underline{x}$). then the function f has all the directional derivatives $\frac{\partial\ f}{\partial\ v}\ (\underline{x})$ at $\underline{x}$, and, furthermore

$$\frac{\partial\ f}{\partial\ v}\ (\underline{x})\ = 0,$$

foe *every direction* $v \in \mathbf{R}^n$.

Since $\frac{\partial\ f}{\partial\ v}\ (\underline{x})\ =\ df(\underline{x})(v)$, it follows that $df(\underline{x}) \equiv 0$. □

NB The points $\underline{x} \in A$ such that $df(\underline{x}) \equiv 0)$ are called *CRITICAL POINTS* of the fuction f.

Proposition 10. *Let A be an open set of $\mathbf{R}^n$, $f : A \to \mathbf{R}$, and let $\underline{x} \in A$. Assume that f is of class $C^{(2)}$.*

Let $\underline{x}$ be a critical point for f.

If $\underline{x}$ is a local maximum (minimum) point for the function f, then:

for every vector $v \in \mathbf{R}^n$,

$$< v \times \mathbf{H}_{f(\underline{x})}, v > \geq \ 0, \quad \left(< v \times \mathbf{H}_{f(\underline{x})}, v > \leq \ 0\right),$$

That is the Hessian matrix $\mathbf{H}_{f(\underline{x})}$ is positive semidefinite (negative semidefinite).

Proof. Let $\delta > 0$ be a real nmber such that $I(\underline{x}, \delta) \subseteq A$, and

$$f(x) \geq f(\underline{x}), \quad \forall x \in I(\underline{x}, \delta).$$

Given $\underline{v} \in \mathbf{R}^n$, $\underline{v} \neq \underline{0}$, the function (of the unique real variable t, with $|t| < \frac{\delta}{||\underline{v}||}$):

$$F(t) = f(\underline{x} + t\underline{v})$$

is of class $C^{(2)}$ and

$$F(t) = f(\underline{x} + t\underline{v}) \geq F(0) = f(\underline{x}), \quad \forall t \in] - \frac{\delta}{||\underline{v}||}, \frac{\delta}{||\underline{v}||}[.$$

Then 0 is a local minimum point for F, and, hence, $F'(0) = 0$ e $F''(0) \geq 0$.

We already know that $F'(0) = \sum_{i=1}^n \frac{\partial f(\underline{x})}{\partial x_i} \underline{v}_i = df(\underline{x})(\underline{v}) = 0$.

Thus,

$$F''(0) = \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{x}) v_i v_j =< \underline{v} \times \mathbf{H}_{f(\underline{x})}, \underline{v} >\geq 0;$$

since $\underline{v}$ may be any vector, the assertion is proved. □

5.5 Sufficient conditions

Therem 8. *Let A be an open set of $\mathbf{R}^n$, $f : A \to \mathbf{R}$, and let $\underline{x} \in A$. Assume that f is of class $C^{(2)}$.*

Let $\underline{x}$ be a critical point for f.

If, for every vector $v \in \mathbf{R}^n, v \neq \underline{0} \in \mathbf{R}^n$, we have

$$< v \times \mathbf{H}_{f(\underline{x})}, v > > \ 0, \quad \left(< v \times \mathbf{H}_{f(\underline{x})}, v > < \ 0\right),$$

that is the Hessian matrix is positive (negative) definite , then $\underline{x}$ is a local minimum (maximum) point for f.

The proof is omitted.

6 Exercises/examples on local maximum and minimum points

Let f be a two variables function f (we will denote by x e y the two variables). In the following section, for the sake simplicity of natation, we will denote by f_x, f_y, f_{xx}, f_{xy},f_{yx}, f_{yy} the partial derivatives $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{\partial^{(2)} f}{\partial x^2}$, $\frac{\partial^{(2)} f}{\partial x \partial y}$, $\frac{\partial^{(2)} f}{\partial y \partial x}$, $\frac{\partial^{(2)} f}{\partial y^2}$ at the "generic" point (x, y) of the domain (if they exist).

Given a point$\underline{x}$ of the domain in which these derivatives exist, we also write $f_x(\underline{x})$, $f_y(\underline{x})$, $f_{xx}(\underline{x})$, $f_{xy}(\underline{x})$, $f_{yx}(\underline{x})$, $f_{yy}(\underline{x})$ for the *evaluations* of them at the point $\underline{x}$.

1)

Consider the function $f : R^2 \to R$,

$$f(x, y) = x^3 - 3x^2 - 9x + \frac{1}{y^2 + 1}.$$

- The function $f$is differentiable on R^2?
- Write the gradient of f at the point$\alpha = (0, 0)$.
- Compute the directional derivative f at the point $\alpha = (0, 0)$, in the direction $v = (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$.
- Determine the local minimum and maximum points of f (if they exist).

SOLUTION.

The partial derivatives at the "generic" point (x, y) are

$$f_x = 3x^2 - 6x - 9, \quad f_y = \frac{-2y}{(y^2 + 1)^2}.$$

Then, the function f is differentiable at every point, since the partial derivatives exist and are continuous at every point of the domain (cfr., Total Differential Theorem).

We have $grad\ f(0, 0) = (-9, 0)$; then

$$\frac{\partial f}{\partial v}(0, 0) =< (-9, 0), (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}) > .$$

In order to determine local extremum points, we must determine first the critical points. Thus, we set $f_x = 3x^2 - 6x - 9 = 0$; we find two solutions $x = 3, -1$ e $f_y = \frac{-2y}{(y^2+1)^2} = 0$ and, hance, $y = 0$.

Thus, we found two critical points:

$$\underline{x} = (3, 0), \quad \underline{x'} = (-1, 0).$$

The entries of the Hessian matrix (at the generic point (x, y)) are:

$$f_{xx} = 6x - 6, \quad f_{yx} = f_{xy} = 0, \quad f_{yy} = \frac{6y^2 - 2}{(y^2 + 1)^3},$$

and, hence, by evaluating at the critical points, we get

$$f_{xx}(\underline{x}) = f_{xx}((3,0)) = 12 \quad e \quad f_{xx}(\underline{x}') = f_{xx}((-1,0)) = -12$$

e

$$f_{yy}(\underline{x}) = f_{yy}((3,0)) = f_{yy}(\underline{x}') = f_{yy}((-1,0)) = -2.$$

Thus, the Hessian matrices at the critical points are:

$$\mathbf{H}_{f(\underline{x})} = \begin{pmatrix} 12 & 0 \\ 0 & -2 \end{pmatrix}$$

e

$$\mathbf{H}_{f(\underline{x}')} = \begin{pmatrix} -12 & 0 \\ 0 & -2 \end{pmatrix}.$$

The first matrix is NOT SEMIDEFINITE, and $\underline{x}$ is a Saddle point; the second matrix is NEGATIVE DEFINITE, and $\underline{x}'$ is a MAXIMUM POINT.

2)

- Consider the function $f : R^2 \to R$, $f(x,y) = (x^2+1)(y^3-3y)$.
- (*) f is differentiable at every point of R^2?
- (*) Compute the gradient of f at $\alpha = (0,0)$ and the directional derivative f in the direction $v = (\frac{3}{5}, -\frac{4}{5})$ at the point α.
- (*) Consider the points α, $\beta = (0,1)$ e $\gamma = (0,-1)$: are they local extremum points fro f?

SOLUTION.

The partial derivatives at the "generic" point (x,y) are

$$f_x = 2x(y^3-3y), \quad f_y = (x^2+1)(3y^2-3).$$

The function f is differentiable at every point, since the partial derivatives exist and are continuous at every point of the domain (cfr., Total Differential Theorem).

We have $grad\ f(0,0) = (0,-3)$, and, hence,

$$\frac{\partial f}{\partial v}(0,0) = <(0,-3), (\frac{3}{5}, -\frac{4}{5})>.$$

The point $\alpha = (0,0)$ isn't a critical point, then it isn't a local extremum point.

The points $\beta = (0,1)$ e $\gamma = (0,-1)$ are critical points and, therefore, we have to study the Hessian matrices.

The entries of the Hessian matrix (at the "generic" point (x,y)) are:

$$f_{xx} = 2(y^3-3y), \quad f_{yx} = f_{xy} = 2x(3y^2-3), \quad f_{yy} = 6y(x^2+1).$$

Thus, the Hessian matrices at the points β e γ are:

$$\mathbf{H}_{f(\beta)} = \begin{pmatrix} -4 & 0 \\ 0 & 6 \end{pmatrix}$$

and

$$\mathbf{H}_{f(\gamma)} = \begin{pmatrix} 4 & 0 \\ 0 & -6 \end{pmatrix}.$$

Both the matrices are NOT SEMIDEFINITE; hence, β and γ are SADDLE points.

3)

- Consider the function

$$f : R^2 \to R$$

,

$$f(x, y) = (x^2 + 1)(y^3 + 3y^2 + 1).$$

- The function f is differentiable on R^2?
- Compute the gradient of f at $(1, 1)$.
- Determine the local minimum and maximum points of f (if they exist).

SOLUTION.

The partial derivatives of f at the "generic" point (x, y) are

$$f_x = 2x(y^3 + 3y + 1), \quad f_y = (x^2 + 1)(3y^2 + 6y).$$

The function f is differentiable at every point, since the partial derivatives exist and are continuous at every point of the domain (cfr., Total Differential Theorem).

We have $grad\ f(1, 1) = (10, 18)$.

The partial derivative f_y annihilates at $y = 0, -2$. Since, for $y = 0, -2$, we have $y^3 + 3y + 1 > 0$, the condition $grad\ f = (0, 0)$ implies $x = 0$.

Thus, the critical points are $\underline{x} = (0, 0)$ e $\underline{x}' = (0, -2)$.

The entries of the Hessian matrix (at the "generic" point (x, y)) are:

$$f_{xx} = 2(y^3 - 3y + 1) > 0.$$

At the critical points,

$$f_{yx} = f_{xy} = 0,$$

and

$$f_{yy} = (x^2 + 1)(6y - +6),$$

that is 6 in $\underline{x} = (0, 0)$ and -6 at $\underline{x}' = (0, -2)$.

Thus, $\underline{x}$ is a minimum point, $\underline{x}'$ is a saddle point.

4)

Let

$$f : R^2 \to R, \qquad f(x, y) = x^2 + xy - y^2.$$

- f is differentiable on R^2?
- Compute the gradient of f at the point $x_0 = (1, 1)$.
- Compute the directional derivative of f in $x_0 = (1, 1)$ in the direction $v = (\sqrt{2}/2, \sqrt{2}/2)$.
- The function f has local extremum points?

SOLUTION.

The partial derivatives at the "generic" point(x, y) are

$$f_x = 2x + y, \quad f_y = x - 2y.$$

The function f is differentiable at every point, since the partial derivatives exist and are continuous at every point of the domain (cfr., Total Differential Theorem).

We have $grad\ f(1, 1) = (3, -1)$.

The unique critical point is $\underline{x} = (0, 0)$.

The entries of the Hessian matrix (at the "generic" point (x, y)) are:

$$f_{xx} = 4 \quad f_{yx} = f_{xy} = 1, \quad f_{yy} = -2.$$

Note that the determinant of the Hessian matrix equals -9, at every point. Since the determinant gives the product of the eigenvalues, they have different sign and, then, the matrix is NOT SEMIDEFINITE. Thus, $\underline{x} = (0, 0)$ is a saddle point.

5)

Let

$$f : R^2 \to R, \qquad f(x, y) = 3x^3 + 3y^2 - x.$$

Determine the critical points of f. Among them, are there extremum points for f?

SOLUTION.

The partial derivatives at the "generic" point(x, y) are

$$f_x = 9x^2 - 1, \quad f_y = 6y.$$

The function f is differentiable at every point, since the partial derivatives exist and are continuous at every point of the domain (cfr., Total Differential Theorem).

The critical points are the points $\underline{x} = (\frac{1}{3}, 0)$ e $\underline{x}' = (-\frac{1}{3}, 0)$.

The entries of the Hessian matrix (at the "generic" point (x, y)) are:

$$f_{xx} = 18x, \quad f_{yx} = f_{xy} = 0, \quad f_{yy} = 6.$$

Therefore, $\underline{x}$ is a local minimum point, $\underline{x}'$ is a saddle point.

6)

Consider the function $f = R^2 \to R$,

$$f(x, y) = x\sin(y).$$

- f is differentiable on R^2?
- Compute the gradient of f at the point $\alpha = (\pi, \pi)$.
- Compute the directional derivative of f at the point α, in the direction $v = (-1/2, \sqrt{3}/2)$.
- Determine the critical points of f. Among them, are there extremum points for f?

SOLUTION.

The partial derivatives at the "generic" point(x, y) are

$$f_x = sin(y), \quad f_y = xcos(y).$$

The function f is differentiable at every point, since the partial derivatives exist and are continuous at every point of the domain (cfr., Total Differential Theorem).

The critical points are those of the form:

$$(0, k\pi), \quad k \in \mathbb{Z}.$$

The entries of the Hessian matrix (at the "generic" point (x, y)) are:

$$f_{xx} = 0, \quad f_{yx} = f_{xy} = cos(y), \quad f_{yy} = -xsin(y).$$

At *any* critical point the Hessian matrix is:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix};$$

since its determinant is NEGATIVE, the matrix is NOT SEMIDEFINITE. Thus, all the critical points are saddle points.

Part III

Maxima and minima subject to constraints

7 Differentiable functions $f : A(A \subset \mathbf{R}^r) \to \mathbf{R}^n$

7.1 Vector- valued functions

Let $A \subset \mathbf{R}^r$, $f : A \to \mathbf{R}^n$. Again, for the sake of simplicity, we always assume that A is an open set of $\mathbf{R}^r$.

Given a vector $\underline{x} = (x_1, \ldots, x_r) \in A$, its *image under* f (or, the "evaluation of f at $\underline{x}$") is a vector $f(\underline{x})$ in $\mathbf{R}^n$; thus, with respect to the canonical basis of $\mathbf{R}^n$, $f(\underline{x})$ is written as an $n-$tuple

$$f(\underline{x}) = (f_1(\underline{x}), f_2(\underline{x}), \ldots, f_n(\underline{x}));$$

the $f_i(\underline{x})$ are the *coordinates* of $f(\underline{x})$.

Hence, the function $f : A \to \mathbf{R}^n$ *determines and its determined* by n functions

$$f_1, f_2, \ldots, f_n$$

with values in $\mathbf{R}$; these $f_1, f_2, \ldots, f_n$ are said to be the *scalar components of the vector-valued function* f.

In general, we will identify $f : A \to \mathbf{R}^n$, $A \subset \mathbf{R}^r$ with $n-$tuple of its scalar components, and write

$$f \equiv (f_1, f_2, \ldots, , f_n).$$

7.2 Vector-valued differentiable functions

Let $f : A \to \mathbf{R}^n$, A open subset of $\mathbf{R}^r$.

We say that f is *differentiable at the point* x_0 if and only if there exists a linear operator

$$L_{x_0} : \mathbf{R}^r \to \mathbf{R}^n$$

(*depending from* x_0) such that

$$\lim_{h \to \underline{0} \in \mathbf{R}^r} \frac{f(x_0 + h) - f(x_0) - L_{x_0}(h)}{||h||} = \underline{0} \in \mathbf{R}^n,$$

where $||h||$ denotes the euclidean norm of the vector $h \in \mathbf{R}^r$.

The operator $h \mapsto L_{x_0}(h)$ is called again the*differential* of f at x_0 and is also denoted by the symbol $df(x_0)$; its evaluations are denoted by $df(x_0)(h), h \in \mathbf{R}^r$.

From the definition, it follows that *the vector-valued function* $f \equiv (f_1, f_2, \ldots, , f_n)$ *is differentiable at* x_0 *if and only if its scalar components* $f_1, f_2, \ldots, , f_n$ *are real-valued differentiable functions at* x_0.

Furthermore, for every $j = 1, 2, \ldots, n$, la $j-$th scalar component $L_{x_0,j}$ of the operator L_{x_0} equals the differential of the la $j-$th scalar component of the function f.

More explicitly, for every $h = (h_1, h_2, \ldots, h_r) \in \mathbf{R}^r$, we have

$$L_{x_0,j}(h) = \sum_{k=1}^{r} \frac{\partial f_j(x_0)}{\partial x_k} \cdot h_k =< grad\ f(x_0), h > .$$

It follows that the matrix of the linear operator L_{x_0} (with respect to the canonical bases of $\mathbf{R}^r$ e $\mathbf{R}^n$, respectively) is the matrix hose rows are the *gradients* of the scalar components $f_1, f_2, \dots, , f_n$; thus, we get the matrix

$$\mathbf{J}_{f(x_0)} = \begin{pmatrix} \frac{\partial f_1(x_0)}{\partial x_1} & \dots\dots & \frac{\partial f_1(x_0)}{\partial x_r} \\ \frac{\partial f_2(x_0)}{\partial x_1} & \dots\dots & \frac{\partial f_2(x_0)}{\partial x_r} \\ \vdots & & \\ \frac{\partial f_n(x_0)}{\partial x_1} & \dots\dots & \frac{\partial f_n(x_0)}{\partial x_r} \end{pmatrix}.$$

This matrix is called the *JACOBIAN MATRIX of the function f at x_0.*

Hence, the differentiability condition for the vector-valued function f may be rewritten in the following form

$$\lim_{h \to \underline{0} \in \mathbf{R}^r} \frac{f(x_0 + h) - f(x_0) - \mathbf{J}_{f(x_0)} \times h^t}{||h||} = \underline{0} \in \mathbf{R}^n.$$

By using the analogous results for real valued functions, we have:

Proposition 11. *If f is differentiable at $x_0 \in A$, then f is continuous in x_0; f is differentiable on A, then f is continuous in A. If $f \in C_A^{(1)}$, then f is differentiable on A.*

N.B. The last assertion is a (weak) form of *Total Differential Theorem.*

Therem 9. *(on the composition of differentiable functions)*

Let $A \subset \mathbf{R}^r$, $B \subset \mathbf{R}^n$, a internal point of A, b internal point of B.

Let $g : A \to B$, $g(a) = b$, $f : B \to \mathbf{R}^p$, g differentiable a, f differentiable at b.

Then, the iterated (composite) function

$$f \circ g : A \to \mathbf{R}^p$$

is differentiable at a.

Furthermore

$$d(f \circ g)(a) = df(b) \circ dg(b) \qquad (*).$$

N.B. The identity $(*)$ must be read: *the linear operator $d(f \circ g)(a) : \mathbf{R}^r \to \mathbf{R}^p$ is obtained by iterating (composing) the linear operators $dg(a) : \mathbf{R}^r \to \mathbf{R}^n$ e $df(b) : \mathbf{R}^r n \to \mathbf{R}^p$.*

In the language of Jacobian matrices, identity has a quite simple form:

Corollary 4.

$$\mathbf{J}_{(f \circ g)(a)} = \mathbf{J}_{f(b)} \times \mathbf{J}_{g(a)}.$$

Example 2. *Consider the functions (differentiable functions, thanks to the preceding remarks - they are polynomial functions and, hence, of class $C^{(\infty)}$)*

$$g : \mathbf{R}^2 \to \mathbf{R}^3, \quad g(x_1, x_2) = (x_1^2 - x_2, x_1 + 1, x_1 - x_2 + 2),$$

and

$$f : \mathbf{R}^3 \to \mathbf{R}^3, \quad f(x_1, x_2, x_3) = (x_2, x_3 - 1, x_1 x_2).$$

Set

$$F = f \circ g : \mathbf{R}^2 \to \mathbf{R}^3;$$

we have

$$F(x_1, x_2) = (g_2(x_1, x_2), g_3(x_1, x_2) - 1, g_1(x_1, x_2) g_2(x_1, x_2)) = (x_1 + 1, x_1 - x_2 + 1, x_1^3 + x_1^2 - x_1 x_2 - x_2).$$

By setting $a = (1, 2) \in \mathbf{R}^2$*, we have* $b = g(a) = (-1, 2, 1)$*.*

For a "generic" point $(x_1, x_2) \in \mathbf{R}^2$*, Jacobian matrix of* g *is*

$$\mathbf{J}_{g(x_1,x_2))} = \begin{pmatrix} 2x_1 & -1 \\ 1 & 0 \\ 1 & -1 \end{pmatrix};$$

thus, if we evaluate the Jacobian matrix at the point $a = (1, 2)$*, we get*

$$\mathbf{J}_{g(1,2))} = \begin{pmatrix} 2 & -1 \\ 1 & 0 \\ 1 & -1 \end{pmatrix}.$$

For a "generic" point $(x_1, x_2, x_3) \in \mathbf{R}^3$*, we have*

$$\mathbf{J}_{f(x_1,x_2,x_3))} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ x_2 & x_1 & 0 \end{pmatrix},$$

thus, if we evaluate the Jacobian matrix at the point $b = (-1, 2, 1)$*, we get*

$$\mathbf{J}_{f(-1,2,1))} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -1 & 0 \end{pmatrix};$$

On the other hand, we have:

at the point $(x_1, x_2) \in \mathbf{R}^2$*, we have*

$$\mathbf{J}_{F(x_1,x_2)} = \begin{pmatrix} 1 & 0 \\ 1 & -1 \\ 3x_1^2 + 2x_2 - x_2 & -x_1 - 1 \end{pmatrix};$$

hence,by evaluating the Jacobian matrix at the point $a = (1, 2)$*, we have*

$$\mathbf{J}_{F(1,2)} = \begin{pmatrix} 1 & 0 \\ 1 & -1 \\ 3 & -2 \end{pmatrix}.$$

As asserted by the Theorem, we have:

$$\mathbf{J}_{f(b)} \times \mathbf{J}_{g(a)} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -1 & 0 \end{pmatrix} \times \begin{pmatrix} 2 & -1 \\ 1 & 0 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & -1 \\ 3 & -2 \end{pmatrix} = \mathbf{J}_{F(1,2)}.$$

8 Continuous curves in $\mathbf{R}^n$

A *(continuous) curve* in $\mathbf{R}^n$ is a continuous map φ from a closed interval $J = [a,b] \subset \mathbf{R}$ to $\mathbf{R}^n$.

Given $\varphi \equiv (\varphi_1, \ldots, \varphi_n)$, $\varphi_i : [a,b] \to \mathbf{R}$, the curve φ is said to be of class $C^{(k)}$ if and only if its scalar components φ_i are of class $C^{(k)}$, for every $i = 1, \ldots, n$.

Given a curve $\varphi : [a,b] \to \mathbf{R}$, the set $\varphi([a,b]) = \{\varphi(t); t \in [a,b]\} \subset \mathbf{R}^n$ is called the *image* of the curve φ.

A curve $\varphi : [a,b] \to \mathbf{R}$ is called *regular* if and only if:

- The map φ is of class $C^{(1)}$.
- $\varphi'(t) \equiv (\varphi_1'(t), \ldots, \varphi_n'(t))$ is different from the zero vector of $\mathbf{R}^n$, for every $t \in]a,b[$ and, furthermore, if $\varphi(a) = \varphi(b)$ then $(\varphi_1'(a), \ldots, \varphi_n'(a)) = (\varphi_1'(b), \ldots, \varphi_n'(b)) \neq \underline{0}$.

A curve $\varphi : [a,b] \to \mathbf{R}^n$ is said to be *simple open* if and only if it defines a *homeomorphism* from $[a,b]$ to $\varphi([a,b]) = \{\varphi(t); t \in [a,b]\} \subset \mathbf{R}^n$.

REMARK 3. *We recall the following general result about continuous functions between metric spaces: let $f : X \to Y$ be a bijective continuous function, X a compact metric space. Then f is a homeomorphism. Since $[a,b]$ is a compact space, a continuous curve $\varphi : [a,b] \to \mathbf{R}^n$ is simple open if and only if it is INJECTIVE.*

8.1 Tangent vectors

For the sake of simplicity, from now on we limit ourselves to consider regular simple open curves $\varphi : [a,b] \to \mathbf{R}^n$. Given such a curve and $t_0 \in]a,b[$, the *tangent versor* to the curve a the point $\varphi(t_0)$ is the versor (direction)

$$T(t_0) = (\varphi_1'(t_0)^2 + \cdots + {\varphi_n}'(t_0)^2)^{-\frac{1}{2}}(\varphi_1'(t_0), \ldots, \varphi_n'(t_0)).$$

Consistently,the line

$$\{\varphi(t_0) + \lambda T(t_0); \lambda \in \mathbf{R}\}$$

is called the "tangent line" to φ at the point $\varphi(t_0)$, and any (non zero) vector proportional to $T(t_0)$ is called a "tangent vector" to the curve at the point $\varphi(t_0)$.

9 Varieties in $\mathbf{R}^n$

9.1 Jacobian matrices

Let F be a function

$$F \equiv (F_1, \ldots, F_{n-r}) : I(\underline{\alpha}, \delta) \to \mathbf{R}^{n-r},$$

where $I(\underline{\alpha}, \delta) \subset \mathbf{R}^n$ is a spherical open neighbourhood of the point $\underline{\alpha} \in \mathbf{R}^n$.

Assume that F is of class $C^{(k)}$, $k \geq 1$ (that is F_i of class $C^{(k)}$, $k \geq 1$, for every $i = 1, \ldots, n-r$).

We recall that the *JACOBIAN MATRIX* of F at the point $\underline{\alpha}$ $(n-r) \times n-$matrix:

$$\mathbf{J}_{F(\underline{\alpha})} = \begin{pmatrix} \frac{\partial F_1(\underline{\alpha})}{\partial x_1} & \cdots\cdots & \frac{\partial F_1(\underline{\alpha})}{\partial x_n} \\ \frac{\partial F_2(\underline{\alpha})}{\partial x_1} & \cdots\cdots & \frac{\partial F_2(\underline{\alpha})}{\partial x_n} \\ \vdots & & \\ \frac{\partial F_{n-r}(\underline{\alpha})}{\partial x_1} & \cdots & \frac{\partial F_{n-r}(\underline{\alpha})}{\partial x_n} \end{pmatrix}.$$

9.2 (Differentiable) Varieties. Regular points

let $1 \leq r \leq n$, $k \geq 1$, amd let V be a subset of $\mathbf{R}^n$.

V is said to be a *VARIETY OF* $\mathbf{R}^n$ *of DIMENSION* r *and CLASS* $C^{(k)}$ if and only if, for every $\underline{\alpha} \in V$, there exist a spherical open neighbourhood $I(\underline{\alpha}, \delta) \subset \mathbf{R}^n$ of $\underline{\alpha}$ and a function

$$F \equiv (F_1, \ldots, F_{n-r}) : I(\underline{\alpha}, \delta) \to \mathbf{R}^{n-r},$$

F of class $C^{(k)}$, such that:

1. $$V \cap I(\underline{\alpha}, \delta) = \{x \in I(\underline{\alpha}, \delta); F(x) = \underline{0}\}.$$
2. The Jacobian matrix $\mathbf{J}_{F(\underline{\alpha})}$ has rank $n-r$.

In plain words, we say that $F \equiv (F_1, \ldots, F_{n-r})$ is a function that provides ("locally", in a neighbourhood of $\underline{\alpha}$) the "equations" of the variety V.

In general, given a subset $V \subset \mathbf{R}^n$, the points $\underline{\alpha}$ for which the preceding conditions hold are called *REGULAR POINTS*. Therefore, a variety is a subset $V \subset \mathbf{R}^n$ whose points are regular points.

9.3 Curves, varieties, tangent and normal spaces

Let V be a variety of $\mathbf{R}^n$ of dimension r $(r \leq n)$ and class $C^{(k)}$ $(k \geq 1)$.

A vector $h \in \mathbf{R}^n$ is said to be *TANGENT to* V *at* $\underline{\alpha}$ if and only if there exist a positive real number $\delta \in \mathbf{R}^+$ and a *regular simple open curve*

$$\varphi : [-\delta, \delta] \to \mathbf{R}^n$$

such that

- i) $\varphi([-\delta, \delta]) \subset V$.
- *ii)* $\varphi(0) = \underline{\alpha}$ *e* $\varphi'(0) = h$.

Geometrically, condition i) means that the image of the curve φ "describe a path" on the variety V "passing through $\underline{\alpha}$" (for $t = 0$); condition ii) means that the vector h is proportional to the "tangent versor" to φ at the point $\varphi(0) = \underline{\alpha}$.

The *TANGENT SPACE* to the variety V at the point $\underline{\alpha}$ is the set

$$\mathbf{T}(\underline{\alpha}) = \{h \in \mathbf{R}^n ; h \quad tangente\ a\ V\ in\ \underline{\alpha}\} \cup \{\underline{0}\}.$$

Therem 10. *Let V be a variety of $\mathbf{R}^n$ of dimension r and class $C^{(k)}$, $k \geq 1$.*

Let $\underline{\alpha} \in V$ and F a function

$$F \equiv (F_1, \ldots, F_{n-r}) : I(\underline{\alpha}, \delta) \to \mathbf{R}^{n-r},$$

F of class $C^{(k)}$, such that:

1. $$V \cap I(\underline{\alpha}, \delta) = \{x \in I(\underline{\alpha}, \delta); F(x) = \underline{0}\}.$$

2. *The Jacobian matrix $\mathbf{J}_{F(\underline{\alpha})}$ has rank $n - r$.*

Then $\mathbf{T}(\underline{\alpha}) = Ker\ dF(\underline{\alpha})$, where $dF(\underline{\alpha})$ denotes, as usual, the differential of F at the point $\underline{\alpha}$.

Corollary 5. *The tangent space $\mathbf{T}(\underline{\alpha})$ is the vector subspace of dimension r given by:*

$$\{h \in \mathbf{R}^n ; \mathbf{J}_{F(\underline{\alpha})} \times h^t = \underline{0}\} \subset \mathbf{R}^n.$$

Since the $i-$th row of the Jacobian matrix $\mathbf{J}_{F(\underline{\alpha})}$ is the gradient *grad* $F_i(\underline{\alpha})$ of the $i-$th scalar component F_i of the function F, $i = 1, \ldots, n - r$, the preceding result can rephrased as follows.

Corollary 6. *The tangent space $\mathbf{T}(\underline{\alpha})$ is the space of all vectors $h \in \mathbf{R}^n$ such that*

$$< grad\ F_i(\underline{\alpha}), h >= 0, \quad per\ ogni\ i = 1, \ldots, n - r,$$

that is, the space of the vectors that are orthogonal to the gradients of the $n - r$ scalar components of F, evaluated at the point $\underline{\alpha}$.

Let V be a variety of $\mathbf{R}^n$ of dimension r and class $C^{(k)}$, $k \geq 1$.

The *NORMAL SPACE* to the variety V at the point $\underline{\alpha}$ is the *orthogonal complement* (in $\mathbf{R}^n$) of the tangent space $\mathbf{T}(\underline{\alpha})$, that is, the vector subspace

$$\mathbf{N}(\underline{\alpha}) = \{u \in \mathbf{R}^n ; < u, h >= 0 \quad per\ ogni\ h \in \mathbf{T}(\underline{\alpha})\} \subset \mathbf{R}^n.$$

REMARK 4. *If V has dimension r, we already know that the tangent space $\mathbf{T}(\underline{\alpha})$ has dimension r (as a vector subspace of $\mathbf{R}^n$): therefore, the normal space $\mathbf{N}(\underline{\alpha})$ is a vector subspace of dimension $n - r$.*

We also know that the $n - r$ vectors

$$grad\ F_i(\underline{\alpha}), \quad i = 1, \ldots, n - r$$

belong to the normal space $\mathbf{N}(\underline{\alpha})$.

Furthermore, since the Jacobian matrix $\mathbf{J}_{F(\underline{\alpha})}$ has rank $n - r$, these vectors are linearly independent.

Thus, we have

Corollary 7. *the set*

$$\{grad\ F_i(\underline{\alpha}); \quad i = 1, \ldots, n-r\}$$

is a basis of the normal space $\mathbf{N}(\underline{\alpha})$.

10 Maxima and minima subject to constraints. Lagrange multipliers

Let V be a variety of $\mathbf{R}^n$ of dimension r and class $C^{(k)}$, $k \geq 1$.

Let A be an open subset $\mathbf{R}^n$, $V \subset A$; consider a function $\Phi : A \to \mathbf{R}$ of class $C^{(1)}$ on A.

A local maximum (minimum) point of the "restriction" function Φ/V of Φ to V is said to be a *CONSTRAINED LOCAL MAXIMUM (MINIMUM) POINT* of Φ with respect to the *"constraint" V*.

Therem 11. *(Lagrange multipliers theorem)*

If Φ has a constrained maximum (minimum) point in $\underline{\alpha} \in V$ (with respect to the constraint V), Then there exist scalars

$$\lambda_1, \lambda_2, \ldots, \lambda_{n-r}$$

such that $\underline{\alpha}$ turns out to be a critical point for the function

$$\Phi - \sum_{j=1}^{n-r} \lambda_j f_j,$$

where

$$f \equiv (f_1, \ldots, f_{n-r}) : I(\underline{\alpha}, \delta) \to \mathbf{R}^{n-r}$$

is the function that provides ("locally", in a neighbourhood of $\underline{\alpha}$) the "equations" of the variety V.

Proof. Let h be a tangent vector to V at $\underline{\alpha}$ and let $\varphi : [-\delta, \delta] \to \mathbf{R}^n$ be a regular simple open curve such that

- i) $\varphi([-\delta, \delta]) \subset V$.
- *ii)* $\varphi(0) = \underline{\alpha} \quad e \quad \varphi'(0) = h$.

Set $\omega(t) = \Phi(\varphi)(t)$; since $\varphi(t) \in V$ and Φ/V has a local extremum (maximum or minimum point) in $\underline{\alpha}$, it follows that ω has a local extremum in $t = 0$, and, hence, $\omega'(0) = 0$.

On the other hand, we have

$$\omega'(0) = \sum_{j=1}^{n} \frac{\partial \Phi(\varphi(0))}{\partial x_j} \cdot \varphi_j'(0) = < grad\ \Phi(\varphi(0)), \varphi'(0) > = < grad\ \Phi(\underline{\alpha}), h > = 0.$$

(Note that this relation holds*for every* tangent vector h to V at $\underline{\alpha}$.)

Thus, *grad* $\Phi(\underline{\alpha})$ *belongs to the normal space to* V *in* $\underline{\alpha}$ and, since, this space is spanned by the vectors

$$grad\ f_1(\underline{\alpha}), grad\ f_2(\underline{\alpha}), \ldots, grad\ f_{n-r}(\underline{\alpha}),$$

there exist scalars

$$\lambda_1, \lambda_2, \ldots, \lambda_{n-r}$$

such that

$$grad\ \Phi(\underline{\alpha}) = \sum_{j=1}^{n-r} \lambda_j\ grad\ f_j(\underline{\alpha}).$$

Hence, $grad\ (\Phi - \sum_{j=1}^{n-r} \lambda_j\ grad\ f_j)(\underline{\alpha}) = 0$, that is, $\underline{\alpha}$ is a critical point for $\Phi - \sum_{j=1}^{n-r} \lambda_j\ grad\ f_j$. □

11 Examples on Lagrange multipliers

1) Let$\Phi : \mathbf{R}^2 \to \mathbf{R}$, $\Phi(x_1, x_2) = x_1 + x_2$ and let

$$V = \{(x_1, x_2) \in \mathbf{R}^2; f(x_1, x_2) = x_1^2 + x_2^2 - 1 = 0\}.$$

The set V is a variety of $\mathbf{R}^2$ of dimension 1

The critical points in V are the points where *grad* $\Phi(x_1, x_2) = (1, 1)$ is proportional to *grad* $f = (2x_1, 2x_2)$; thus, it must be $x_1 = x_2$ and, hence, since the points belong to V we have two cases:

$$(x_1, x_2) = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}),\ (-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}).$$

Since V is compact, by the theorem of Weierstrass, these two points will be a minimum and a maximum (absolute, with respect to the constraint V).

2) Let $\Phi : \mathbf{R}^3 \to \mathbf{R}$, $\Phi(x_1, x_2, x_3) = x_1 + x_2 + x_3^2$ and let

$$V = \{(x_1, x_2, x_3) \in \mathbf{R}^3; f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - 1 = 0\}.$$

The set V is a variety of $\mathbf{R}^2$ of dimension 1.

The critical points in V are the points where *grad* $\Phi(x_1, x_2, x_3) = (1, 1, 2x_3)$ is proportional to *grad* $f(2x_1, 2x_2, x_3) = (2x_1, 2x_2, 2x_3)$; thus,

$$1 = \lambda 2x_1, \quad 1 = \lambda 2x_2, \quad 2x_3 = \lambda 2x_3;$$

If $x_3 \neq 0$, then $\lambda = 1$, and $x_1 = \frac{1}{2}$, $x_2 = \frac{1}{2}$; since the points must belong to V, we get $x_3 = \pm\frac{1}{\sqrt{2}}$.

Thus, we found two critical points

$$(\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}), \quad (\frac{1}{2}, \frac{1}{2}, -\frac{1}{\sqrt{2}}),$$

with multiplier $\lambda = 1$.

If $x_3 = 0$, then $x_1^2 + x_2^2 = 1$, but also $x_1 = x_2$; thus, $x_1 = x_2 = \pm\frac{1}{\sqrt{2}}$.
We found two other critical points

$$(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0), \quad (-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0)$$

with multipliers $\lambda = \frac{\sqrt{2}}{2}$ e $\lambda = -\frac{\sqrt{2}}{2}$, respectively.
Since V is compact, by the theorem of Weierstrass, the function Φ admits a minimum and a maximum (absolute, with respect to the constraint V).
By evaluating the function in the four critical points, we find that the absolute minimum is in the point

$$(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0),$$

and the absolute maximum is in the two points

$$(\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}), \quad (\frac{1}{2}, \frac{1}{2}, -\frac{1}{\sqrt{2}}).$$

3) Let $\Phi : \mathbf{R}^3 \to \mathbf{R}$, $\Phi(x_1, x_2, x_3) = x_1 + x_2 + x_3$ and let

$$f \equiv (f_1, f_2) : \mathbf{R}^3 \to \mathbf{R}^2,\ f_1(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - 4,\ f_2(x_1, x_2, x_3) = x_1 - 1.$$

Let

$$V = \{(x_1, x_2, x_3) \in \mathbf{R}^3; f(x_1, x_2, x_3) = 0\}.$$

The set V is a variety of $\mathbf{R}^3$ of dimension 1.
We have *grad* $\Phi(x_1, x_2, x_3) = (1, 1, 1)$, *grad* $f_1 = (2x_1, 2x_2, 2x_3)$, *grad* $f_2 = (1, 0, 0)$ and we have to determine the points of V for which there exist $\lambda_1, \lambda_2 \in \mathbf{R}$ such that

$$(1, 1, 1) = \lambda_1(2x_1, 2x_2, 2x_3) + \lambda_2(1, 0, 0).$$

It follows that $x_2 = x_3$; since the points must belong to V, we also have $x_1 = 1$ and, then $x_2 = x_3 = \pm\frac{\sqrt{3}}{2}$. Furthermore, by evaluating the first coordinate we get $\lambda_2 = 1 - 2\lambda_1$.
Therefore, the critical points are $(1, \frac{\sqrt{3}}{2}, \frac{\sqrt{3}}{2})$ e $(1, -\frac{\sqrt{3}}{2}, -\frac{\sqrt{3}}{2})$ with multipliers

$$\lambda_1 = \frac{1}{\sqrt{3}},\ \lambda_2 = 1 - \frac{2}{\sqrt{3}}$$

e

$$\lambda_1 = -\frac{1}{\sqrt{3}},\ \lambda_2 = 1 + \frac{2}{\sqrt{3}},$$

respectively.
Since V is compact, by the theorem of Weierstrass, these two points will be a minimum and a maximum (absolute, with respect to the constraint V).
4) Let $\Phi : \mathbf{R}^2 \to \mathbf{R}$, $\Phi(x_1, x_2) = x_1 x_2$ and let

$$V = \{(x_1, x_2) \in \mathbf{R}^2; f(x_1, x_2) = x_1 + 2x_2 - 3 = 0\}.$$

Since $grad\ \Phi(x_1, x_2) = (x_2, x_1)$ e $grad\ f(x_1, x_2) = (1, 2)$, it follows

$$x_2 = \lambda, \quad x_1 = 2\lambda.$$

Thus, the critical points are of the form $(x_1, x_2) = (2\lambda, \lambda)$; since they must belong to V, we get $2\lambda + 2\lambda - 3 = 0$ and, hence, $\lambda = \frac{3}{4}$.

Therefore, the only critical point is $(\frac{3}{2}, \frac{3}{4})$, with multiplier $\lambda = \frac{3}{4}$.

5) Let $\Phi : \mathbf{R}^3 \to \mathbf{R}$, $\Phi(x_1, x_2, x_3) = x_1x_2 + x_3^2$ and let

$$f \equiv (f_1, f_2) : \mathbf{R}^3 \to \mathbf{R}^2, \ f_1(x_1, x_2, x_3) = x_1 + x_2 - 1, \ f_2(x_1, x_2, x_3) = x_2^2 + x_3^2 - 1.$$

Let

$$V = \{(x_1, x_2, x_3) \in \mathbf{R}^3; f(x_1, x_2, x_3) = 0\}.$$

The set V is a variety of $\mathbf{R}^3$ of dimension 1.

We have $grad\ \Phi(x_1, x_2, x_3) = (x_2, x_1, 2x_3)$, $grad\ f_1 = (1, 1, 0)$, $grad\ f_2 = (0, 2x_2, 2x_3)$ and we have to determine the points of V for which there exist $\lambda_1, \lambda_2 \in \mathbf{R}$ such that

$$(x_2, x_1, 2x_3) = \lambda_1(1, 1, 0) + \lambda_2(0, 2x_2, 2x_3).$$

Thus, $\lambda_1 = x_2$ and, if $x_3 \neq 0$, $\lambda_2 = 1$, then $x_1 = \lambda_1 + 2x_2 = 3x_2$; from the condition $f_1(x_1, x_2, x_3) = 0$ we infer that $x_2 = \frac{1}{4}$ e $x_1 = \frac{3}{4}$.

From the condition $f_1(x_1, x_2, x_3) = 0$ it follows $x_3 = \pm\frac{\sqrt{15}}{4}$.

Therefore, there are two critical points with $x_3 \neq 0$, specifically

$$(\frac{3}{4}, \frac{1}{4}, \frac{\sqrt{15}}{4}) \quad (\frac{3}{4}, \frac{1}{4}, -\frac{\sqrt{15}}{4}).$$

If $x_3 = 0$, then$x_2 = \pm 1$. We get two other critical points, specifically

$$(2, -1, 0)$$

with multipliers $\lambda_1 = -1$ e $\lambda_2 = -\frac{3}{2}$ e

$$(0, 1, 0)$$

with multipliers $\lambda_1 = 1$ e $\lambda_2 = -\frac{1}{2}$. Since V is compact, by the theorem of Weierstrass, the function Φ admits a minimum and a maximum (absolute, with respect to the constraint V).

By evaluating the function in the four critical points, we find that the absolute minimum is in the point

$$(2, -1, 0),$$

and the absolute maximum is in the two points

$$(\frac{3}{4}, \frac{1}{1}, \frac{\sqrt{15}}{4}), \quad (\frac{3}{4}, \frac{1}{1}, -\frac{\sqrt{15}}{4}).$$

6) Let $\Phi : \mathbf{R}^3 \to \mathbf{R}$, $\Phi(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$ and let

$$f \equiv (f_1, f_2) : \mathbf{R}^3 \to \mathbf{R}^2, \ f_1(x_1, x_2, x_3) = x_1^2 + x_2 - 1, \ f_2(x_1, x_2, x_3) = x_2 + x_3 - 1.$$

Let

$$V = \{(x_1, x_2, x_3) \in \mathbf{R}^3; f(x_1, x_2, x_3) = 0\}.$$

The set V is a variety of $\mathbf{R}^3$ of dimension 1.

We have $grad\ \Phi(x_1, x_2, x_3) = (2x_1, 2x_2, 2x_3)$, $grad\ f_1 = (2x_1, 1, 0)$, $grad\ f_2 = (0, 1, 1)$ and we have to determine the points of V for which there exist $\lambda_1, \lambda_2 \in \mathbf{R}$ such that

$$(2x_1, 2x_2, 2x_3) = \lambda_1(2x_1, 1, 0) + \lambda_2(0, 1, 1).$$

If $x_1 \neq 0$, then $\lambda_1 = 1$. Furthermore

$$2x_2 = \lambda_1 + \lambda_2, \quad 2x_3 = \lambda_2;$$

and, then

$$2x_2 - 2x_3 = 1, \quad x_2 + x_3 = 1;$$

thus $x_3 = \frac{1}{4}$ e $x_2 = \frac{3}{4}$.

Since the point must belong to a V, it follows that $x_1 = \pm\frac{1}{2}$.

Therefore, we found two critical points

$$(\frac{1}{2}, \frac{3}{4}, \frac{1}{4}), \quad (-\frac{1}{2}, \frac{3}{4}, \frac{1}{4}),$$

with multipliers

$$\lambda_1 = 1, \quad \lambda_2 = \frac{1}{2}.$$

If $x_1 = 0$, then $x_2 = 1$ e $x_3 = 0$. Then, we found a third critical point

$$(0, 1, 0)$$

with multipliers

$$\lambda_1 = 1, \quad \lambda_2 = 0.$$

Part IV

Appendix

12 Dual spaces and differentials

12.1 How to write a differential in "intrinsic way"?

Given the vector space $\mathbf{R}^n$, its *duale space* is, by definition, the set

$$(\mathbf{R}^n)^* = \{\varphi : \mathbf{R}^n \to \mathbf{R}; \quad \varphi \quad funzionale \quad lineare\},$$

endowed with thw *sum* operation:

$$(\varphi_1 + \varphi_2)(v) = \varphi_1(v) + \varphi_2(v), \quad \forall v \in \mathbf{R}^n,$$

and of *external multiplication* by a scalar:

$$(\lambda\varphi_1)(v) = \lambda\varphi(v), \quad \forall v \in \mathbf{R}^n, \quad \forall \lambda \in \mathbf{R}.$$

The resulting algebraic structure is clearly a vector space.

Differential are linear functionals, and, hence, they are elements of the dual space $(\mathbf{R}^n)^*$; therefore, if we detemine a "canonical" basis of the dual space $(\mathbf{R}^n)^*$ we will have a "canonical way" to express the differentials..

12.2 A "canonical basis" of the dual space $(\mathbf{R}^n)^*$

Denoted by $\{\underline{e}_1, \ldots, \underline{e}_n\}$ the canonical basis of $\mathbf{R}^n$, consider the set of linear functionals

$$\{dx_1, \ldots, dx_n\}$$

defined as follows:

$$dx_i : \mathbf{R}^n \to \mathbf{R},$$

where

$$dx_i(\underline{e}_i) = 1 \quad e \quad dx_i(\underline{e}_j) = 0 \ se \ i \neq j.$$

FUNDAMENTAL REMARK.

Let $v = (v_1, \ldots, v_n) = \sum_{i=1}^n \ v_i\underline{e}_i \in \mathbf{R}^n$.

Then

$$dx_i(v) = v_i;$$

this is the reason why the linear functionals dx_i are also called the *coordinate functionals.*

Therem 12. *The set*

$$\{dx_1, \ldots, dx_n\}$$

is a basis of the dual space $(\mathbf{R}^n)^*$*.*

In particular, the dual space $(\mathbf{R}^n)^*$ *is a vector space of finite dimension* n*.*

Proof. We have to prove that $\{dx_1, \ldots, dx_n\}$ is a system of generators and it is a linearly independent set in the vector space $(\mathbf{R}^n)^*$.

system of generators)

Let $\varphi \in (\mathbf{R}^n)^*$, and let $v = (v_1, \ldots, v_n) = \sum_{i=1}^n \ v_i\underline{e}_i \in \mathbf{R}^n$. Then

$$\varphi(v) = \varphi(\sum_{i=1}^n \ v_i\underline{e}_i) = \sum_{i=1}^n \ v_i\varphi(\underline{e}_i) = \sum_{i=1}^n \ \varphi(\underline{e}_i)dx_i(v), \quad \forall v \in \mathbf{R}^n.$$

This (infinite) family of identities among evaluations implies the following identity in the dual space $(\mathbf{R}^n)^*$:

$$\varphi = \sum_{i=1}^n \ \varphi(\underline{e}_i) \ dx_i.$$

linearly independent set)

We have to prove that the condition

$$\sum_{i=1}^{n} c_i \ dx_i \equiv 0$$

implies

$$c_1 = c_2 = \ldots = c_n = 0.$$

Now

$$\sum_{i=1}^{n} c_i \ dx_i(\underline{e}_1) = c_1 = 0,$$

$$\sum_{i=1}^{n} c_i \ dx_i(\underline{e}_2) = c_2 = 0,$$

....

$$\sum_{i=1}^{n} c_i \ dx_i(\underline{e}_n) = c_n = 0,$$

and the assertion follows. □

Corollary 8. *Sia $f : A \to \mathbf{R}$, A aperto di $\mathbf{R}^n$, $\underline{x} \in A$, e sia f differenziabile in $\underline{x}$. Denotiamo con $L_{\underline{x}}$ il differenziale di f in $\underline{x} \in A$.*

Allora il differenziale si scrive (in modo unico nello spazio duale $(\mathbf{R}^n)^*$) come segue:

$$L_{\underline{x}} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\underline{x}) dx_i \in (\mathbf{R}^n)^*.$$

Proof. It immediately follows from the equivalence between the (infinite) set of identities (on the pointwise evaluations):

$$L_{\underline{x}}(v) = < grad \ f(\underline{x}), v > = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\underline{x}) v_i \ = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\underline{x}) dx_i(v), \quad \forall v \in \mathbf{R}^n$$

and the "vectorial" identity:

$$L_{\underline{x}} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\underline{x}) dx_i \in (\mathbf{R}^n)^*.$$

□

In plain words, the differential $L_{\underline{x}}$ is written, in a unique way, *as a linear combination of the coordinate linear functional dx_i, where the coefficients are the partial derivatives* $\frac{\partial f}{\partial x_i}(\underline{x})$.

13 On the differential of composite functions: some details and proof

Consider two functions $g \equiv (g_1, \ldots, g_n) : A \to B \subseteq \mathbb{R}^n$ and $f \equiv (f_1, \ldots, f_p) : B \to \mathbb{R}^p$. Let $a \in A$ and $b = g(a) \in B$, g differentiable at a and f differentiable at $b = g(a)$.

We know that the composite function $f \circ g$ is differentiable at $a \in A$.

The Jacobian matrix $J_{(f \circ g)(a)}$ is a matrix of order $p \times r$. Given $i = 1, \ldots, p$ e $k = 1, \ldots, r$, we want to compute the entry in the position (i, k) in $J_{(f \circ g)(a)}$:

$$\frac{\partial (f \circ g)_i(a)}{\partial x_k}.$$

Our aim is to compute this entry in terms of the Jacobian matrices

$$J_{f(g(a))}, \quad J_{g(a)}.$$

First, we notice that the $i-$th scalar component $i-$esima $(f \circ g)_i$ of $f \circ g$ coincides with the composite function $(f_i \circ g)$.

Furthermore, the partial derivative

$$\frac{\partial (f \circ g)_i(a)}{\partial x_k} = \frac{\partial (f_i \circ g)(a)}{\partial x_k}$$

may be written as the standard derivative of the function $f_i \circ g$ *regarded as a function of a single variable* x_k, that is, in turn, the composition of the function f_i with the function $g \equiv (g_1, \ldots, g_n)$ *regarded as a function of the single variable* x_k.

Sinc g is differentiable at a, the partial derivatives $g_1, \ldots, g_n$ with respect to the variables x_k exist at a and, hence, - since f_i is differentiable at $g(a)$ - the partial derivative

$$\frac{\partial (f_i \circ g)(a)}{\partial x_k}$$

exists at a; therefore, we have:

$$\frac{\partial (f_i \circ g)(a)}{\partial x_k} = \sum_{j=1}^{n} \frac{\partial f_i(g(a))}{\partial x_j} \, \frac{\partial g_j(a)}{\partial x_k}.$$

Thus, the entry

$$\frac{\partial (f \circ g)_i(a)}{\partial x_k} = \frac{\partial (f_i \circ g)(a)}{\partial x_k} \in J_{(f \circ g)(a)}$$

coincide with the product between the $i-$th row of $J_{f(g(a))}$ and the $k-$th column of $J_{g(a)}$.

Andrea Brini

Dipartimento di Matematica, Universita di Bologna

40126 Bologna, Italy

E-mail: brini@dm.unibo.it