Chapter 1

Topology and calculus background

We consider $\mathbb{R}^n$ endowed with the scalar (or inner) product

$$x^T y = \sum_{i=1}^{n} x_i y_i$$

which induces the Euclidean norm

$$\|x\|_2 = \sqrt{x^T x} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

The following properties hold for any $x, y \in \mathbb{R}^n$ and any $\alpha \in \mathbb{R}$:

$$\begin{aligned}
&\|x\|_2 \geq 0 \\
&\|\alpha x\|_2 = |\alpha| \|x\|_2 \\
&\|x\|_2 = 0 \iff x = 0 \\
&\|x + y\|_2 \leq \|x\|_2 + \|y\|_2 \\
&(\|x - y\|_2 \leq \|x\|_2 + \|y\|_2) \\
&|x^T y| \leq \|x\|_2 \|y\|_2. \text{ (Schwarz inequality).}
\end{aligned}$$

In turn, the Euclidean norm induces the well-known Euclidean distance between the points $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$:

$$d(x, y) = \|x - y\|_2$$

and the following properties can be deduced from the above ones:

$$\begin{aligned}
&d(x, y) \geq 0 \\
&d(x, y) = 0 \iff x = y \\
&d(x, y) \leq d(x, z) + d(z, x).
\end{aligned}$$

1.1 Sequences

A family of points $\{x^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^n$ (i.e., $\{x^1, x^2, \dots, x^k, \dots\}$) is called a *sequence*. For instance, the family of points $x^k = (1/k, 1/k^2)$ is a sequence in $\mathbb{R}^2$.

Definition 1.1. $\bar{x} \in \mathbb{R}^n$ is the *limit of a sequence* $\{x^k\}_{k\in\mathbb{N}}$ if for each $\varepsilon > 0$ there exists $\bar{k} \in \mathbb{N}$ such that $d(x^k, \bar{x}) \le \varepsilon$ for all $k \ge \bar{k}$, or equivalently

$$\forall\ \varepsilon > 0 \quad \exists\ \bar{k} \in \mathbb{N} \quad \text{s.t.} \quad \|x^k - \bar{x}\|_2 \le \varepsilon \quad \forall\ k \ge \bar{k}.$$

If it exists, the limit of a sequence is unique. Standard notations to denote a limit are the following: $\lim_{k\to+\infty} x^k = \bar{x}$, $x^k \longrightarrow \bar{x}$ ($k \to +\infty$ below the arrow is often omitted).

Example 1.1. The limit of the sequence $(1/k, 1/k^2)$ is $\bar{x} = (0, 0)$, while the sequence $x^k = (1/k, (-1)^k)$ does not have a limit. Take the sequence obtained just considering odd indices: $x^1, x^3, x^5, \dots$ This sequence converges to $(0, -1)$. Analogously, the sequence obtained considering just even indices converges to $(0, 1)$.

Definition 1.2. $\{x^{k_j}\}_{j\in\mathbb{N}} \subseteq \{x^k\}_{k\in\mathbb{N}}$ is a *subsequence* if $k_j \to +\infty$ as $j \to +\infty$.

Definition 1.3. $\bar{x} \in \mathbb{R}^n$ is a *cluster point* of $\{x^k\}_{k\in\mathbb{N}}$ if there exists a subsequence $\{x^{k_j}\}_{j\in\mathbb{N}}$ such that $\bar{x}$ is its limit, i.e., $\lim_{j\to+\infty} x^{k_j} = \bar{x}$, or equivalently

$$\forall\ \varepsilon > 0 \quad \forall\ k \in \mathbb{N} \quad \exists\ \bar{k} \ge k \quad \text{s.t.} \quad \|x^{\bar{k}} - \bar{x}\|_2 \le \varepsilon.$$

If a sequence has a limit, then it is the unique cluster point of the sequence.

Example 1.2. The last sequence of Example 1.1 has 2 cluster points: $(0, 1)$ and $(0, -1)$, while the sequence $y^k = (k, 1/k)$ does not have any cluster point.

Theorem 1.1. (Bolzano-Weierstrass) *If the norm of all the points of a sequence* $\{x^k\}_{k\in\mathbb{N}}$ *do not exceed a threshold value, i.e., there exists* $M > 0$ *such that* $\|x^k\|_2 \le M$ *holds for all* $k \in \mathbb{N}$*, then the sequence has at least one cluster point.*

1.2 Topological properties in the Euclidean space

The open ball of centre $x \in \mathbb{R}^n$ and radius $\varepsilon > 0$ is the set

$$B(x, \varepsilon) = \{y \in \mathbb{R}^n \ : \ \|y - x\|_2 < \varepsilon\}.$$

Definition 1.4.

(*i*) $D \subseteq \mathbb{R}^n$ is called *open* if

$$\forall\ x \in D \quad \exists\ \varepsilon > 0 \quad \text{s.t.} \quad B(x, \varepsilon) \subseteq D.$$

(*ii*) $x \in D$ is called an *interior point* of D if

$$\exists\ \varepsilon > 0 \quad \text{s.t.} \quad B(x, \varepsilon) \subseteq D.$$

The set of the interior points of D is called *the interior of* D and it is generally denoted by $\operatorname{int} D$. Notice that a set D is open if and only if $D = \operatorname{int} D$.

Example 1.3. $B(x, \varepsilon)$, $\mathbb{R}^n$, $\emptyset$ are open sets in $\mathbb{R}^n$ while the interval $]-1, 1[$ is an open set in $\mathbb{R}$.

Proposition 1.1.

(*i*) *The union of a family of open sets is an open set.*

(*ii*) *The intersection of a finite family of open sets is an open set.*

The finiteness of the family is crucial for the intersection property:

$$\bigcap_{k=1}^{+\infty} B(0, 1/k) = \{0\}.$$

Definition 1.5.

(*i*) $D \subseteq \mathbb{R}^n$ is called *closed* if $\mathbb{R}^n \setminus D = \{x \in \mathbb{R}^n \ : \ x \notin D\}$ is open.

(*ii*) $x \in \mathbb{R}^n$ is called an *closure point* of D if

$$\forall\, \varepsilon > 0 \ : \ B(x, \varepsilon) \cap D \neq \emptyset.$$

The set of the closure points of D is called *the closure of* D and it is generally denoted by $\operatorname{cl} D$ or $\overline{D}$.

Proposition 1.2.

(*i*) *D is closed if and only if $D = \operatorname{cl} D$.*

(*ii*) *D is closed if and only if the limit of any convergent sequence contained in D belongs to D as well, i.e.,*

$$\forall\, \{x^k\}_{k\in\mathbb{N}} \subseteq D \quad s.t. \quad \exists\, \bar{x} \in \mathbb{R}^n \quad s.t. \quad x^k \longrightarrow \bar{x} \ : \ \bar{x} \in D.$$

Example 1.4. $\mathbb{R}^n$, $\emptyset$, $\{y \in \mathbb{R}^n \ : \ \|y - x\|_2 \leq \varepsilon\} = \overline{B(x, \varepsilon)}$ are closed sets in $\mathbb{R}^n$ while the interval $[-1, 1]$ is a closed set in $\mathbb{R}$. There exist sets which are neither closed nor open, for instance the interval [-1,1[in $\mathbb{R}$ and

$$D = [-1, 0] \times [-1, 1] \cup B(0, 1) \subseteq \mathbb{R}^2.$$

In fact, $(-1 - \varepsilon, 0) \notin D$ but $(-1 - \varepsilon, 0) \in B((-1, 0), \varepsilon)$ for any $\varepsilon > 0$ so that D is not open, and $x^k = (1 - 1/k, 0) \in D$ for any $k \in \mathbb{N}$ while $x^k \to (1, 0) \notin D$ so that D is not closed.

Proposition 1.3.

(*i*) *The union of a finite family of closed sets is an closed set.*

(*ii*) *The intersection of a family of closed sets is a closed set.*

The finiteness of the family is crucial for the union property:

$$\bigcup_{k=2}^{+\infty} \overline{B(0, 1 - 1/k)} = B(0, 1).$$

Definition 1.6. $x \in \mathbb{R}^n$ is called a *boundary point* of D if both

$$B(x, \varepsilon) \cap D \neq \emptyset \quad \text{and} \quad B(x, \varepsilon) \not\subseteq D$$

hold for any $\varepsilon > 0$.

The set of the boundary points of D is called *the boundary (or frontier) of* D and it is generally denoted by ∂D. Notice that $\partial D = \overline{D} \cap \overline{(\mathbb{R}^n \setminus D)}$.

Proposition 1.4. *$D \subseteq \mathbb{R}^n$ is both closed and open if and only if $D = \mathbb{R}^n$ or $D = \emptyset$.*

Definition 1.7.

(*i*) $D \subseteq \mathbb{R}^n$ is called *bounded* if

$$\exists\, M > 0 \quad \text{s.t.} \quad \forall\, x \in D \ : \ \|x\|_2 \leq M.$$

(*ii*) $D \subseteq \mathbb{R}^n$ is called *compact* if it is bounded and closed.

The set D in Example 1.4 is bounded but it is not compact (since it is not closed).

The Bolzano-Weierstrass' theorem can be enhanced in the following way.

Theorem 1.2. (Bolzano-Weierstrass) *A set is compact if and only if any sequence contained in the set has at least one cluster point and all its cluster points belong to the set.*

1.3 Functions of several variables

1.3.1 Continuity

Definition 1.8. $f : \mathbb{R}^n \to \mathbb{R}$ is called *continuous at* $\bar{x} \in \mathbb{R}^n$ if $f(\bar{x})$ is the limit of $f(x)$ as $x \to \bar{x}$, i.e.,

$$\forall\, \varepsilon > 0 \quad \exists\, \delta > 0 \quad \text{s.t.} \quad \|x - \bar{x}\|_2 \leq \delta \implies |f(x) - f(\bar{x})| \leq \varepsilon.$$

f is continuous on a set $D \subseteq \mathbb{R}^n$ if it is continuous at every $x \in D$.

Proposition 1.5. *f is continuous at $\bar{x} \in \mathbb{R}^n$ if and only if any sequence $\{x^k\}_{k \in \mathbb{N}}$ such that $x^k \longrightarrow \bar{x}$ satisfies $f(x^k) \longrightarrow f(\bar{x})$.*

Example 1.5. $f(x) = \|x\|_2$ is a continuous function on $\mathbb{R}^n$, $f(x_1, x_1) = \sin(\pi x_1 x_2)$ is a continuous function on $\mathbb{R}^2$.

Theorem 1.3. (Weierstrass) *Let $D \subseteq \mathbb{R}^n$ be compact and $f : \mathbb{R}^n \to \mathbb{R}$ continuous on D. Then, there exist at least one minimum point $\bar{x} \in D$ and one maximum point $\hat{x} \in D$ for f over D, i.e.,*

$$f(\bar{x}) = \min\{f(x) \ : \ x \in D\} \quad \textit{and} \quad f(\hat{x}) = \max\{f(x) \ : \ x \in D\}.$$

Proof. Let $\ell = \inf\{f(x) \ : \ x \in D\} \in [-\infty + \infty[$ and consider any minimizing sequence, that is any $\{x^k\}_{k\in\mathbb{N}}$ such that $f(x^k) \to \ell$. Since D is compact, there exist a subsequence $\{x^{k_j}\}_{j\in\mathbb{N}}$ and $\bar{x} \in D$ such that $x^{k_j} \to \bar{x}$ (as $j \to +\infty$) by Theorem 1.2. Since f is continuous, $f(x^{k_j}) \to f(\bar{x})$ and therefore $f(\bar{x}) = \ell$ by the uniqueness of the limit. As a consequence, $\ell \neq -\infty$ and $f(\bar{x}) = \min\{f(x) \ : \ x \in D\}$. The existence of $\hat{x}$ can be proved analogously. □

Example 1.6. Take $n = 1$, $f(x) = e^{-x}$ and $D = \mathbb{R}_+$: f is continuous on D, $\inf\{f(x) \ : \ x \in D\} = 0$ but there exists no $x \in D$ such that $f(x) = 0$. Indeed, D is not compact as it is not bounded.

1.3.2 Partial derivatives and differentiability

A point $d \in \mathbb{R}^n$ such that $\|d\|_2 = 1$ is also called a *direction*, and the set

$$\{\bar{x} + td \ : \ t \in \mathbb{R}\}$$

describes the line of direction d passing through $\bar{x} \in \mathbb{R}^n$. If only $t \in \mathbb{R}_+$ are considered, the set describes the corresponding half-line.

Just like the case $n = 1$, the key tool for developing calculus for a function $f : \mathbb{R}^n \to \mathbb{R}$ is the *incremental ratio*

$$icr_{(f,x,d)}(t) = [f(x + td) - f(x)]/t.$$

Definition 1.9. f has a derivative at $\bar{x}$ in the direction d if the derivative of the function of one variable $icr_{(f,\bar{x},d)}$ at $t = 0$ exists, that is $\lim\limits_{t\to 0} [f(\bar{x} + td) - f(\bar{x})]/t$ exists. In that case

$$\frac{\partial f}{\partial d}(\bar{x}) = \lim_{t\to 0} \frac{f(\bar{x} + td) - f(\bar{x})}{t}$$

is called the *(directional) derivative of f at $\bar{x}$ in the direction d*. For $n = 1$ there exists a unique (up to the sign) direction and the directional derivative coincides with the (usual) derivative and it is also denoted by $f'(\bar{x})$.

If d is one of the vectors of the canonical basis $\{e_1, \dots, e_n\}$ of $\mathbb{R}^n$, namely $d = e_i$, then the corresponding directional derivative is called *partial derivative* and denoted by $\partial f(x)/\partial x_i$ rather than $\partial f(x)/\partial e_i$. Indeed, the derivative can be computed considering f as a function of x_i while the other variables are kept fixed like parameters:

$$\frac{\partial f}{\partial x_i}(\bar{x}) = \lim_{t\to 0} \frac{f(\bar{x}_1, \dots, \bar{x}_{i-1}, \bar{x}_i + t, \bar{x}_{i+1}, \dots, \bar{x}_n) - f(\bar{x})}{t}$$

Definition 1.10. If f has all the partial derivatives at $\bar{x} \in \mathbb{R}^n$, the vector

$$\nabla f(\bar{x}) = \Big(\frac{\partial f}{\partial x_1}(\bar{x}), \frac{\partial f}{\partial x_2}(\bar{x}), \dots, \frac{\partial f}{\partial x_n}(\bar{x})\Big)^T$$

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

Example 1.7. Take $n = 2$ and $f(x_1, x_2) = \sin(\pi x_1 x_2)$:

$$\frac{\partial f}{\partial x_1}(x) = \pi x_2 \cos(\pi x_1 x_2), \quad \frac{\partial f}{\partial x_2}(\bar{x}) = \pi x_1 \cos(\pi x_1 x_2).$$

Other directional derivatives can be defined just considering the limit of the incremental ratio as $t \to 0^+$, that is $t \to 0$ for only positive t ($t > 0$).

Definition 1.11. The limit

$$f'(\bar{x}; d) = \lim_{t \to 0^+} \frac{f(\bar{x} + td) - f(\bar{x})}{t}$$

is called the *one-sided directional derivative of f at $\bar{x}$ in the direction d.*

Clearly, $f'(\bar{x}; d) = \partial f(\bar{x})/\partial d$ if the latter exists but this is not always the case.

Example 1.8. Consider $f(x) = \|x\|_2$ and take $\bar{x} = 0$:

$$[f(\bar{x} + td) - f(\bar{x})]/t = \|td\|_2/t = |t|\|d\|_2/t = \mathrm{sgn}(t)\|d\|_2$$

where $\mathrm{sgn}(t)$ denotes the sign of t ($\mathrm{sgn}(t) = 1$ if $t \geq 0$ and $\mathrm{sgn}(t) = -1$ if $t < 0$). Therefore, $f'(\bar{x}; d) = \|v\|_2 = 1$ while $\partial f(\bar{x})/\partial d$ does not exist.

Unlike the case $n = 1$, the existence of the directional/partial derivatives does not guarantee the continuity of the function.

Example 1.9. Take $n = 2$ and

$$f(x_1, x_2) = \begin{cases} [x_1^2 x_2/(x_1^4 + x_2^2)]^2 & \text{if } (x_1, x_2) \neq (0, 0) \\ 0 & \text{if } (x_1, x_2) = (0, 0). \end{cases}$$

Consider the parabola $x_2 = \alpha x_1^2$ for $x_1 \neq 0$:

$$f(x_1, \alpha x_1^2) = [\alpha x_1^4/(x_1^4 + \alpha^2 x_1^4)] = \alpha^2/(1 + \alpha^2)^2.$$

Therefore, f is not continuous at $\bar{x} = (0, 0)$: take the sequence $x^k = (1/k, 1/k^2)$ to get $x^k \to \bar{x}$ while $f(x^k) \equiv 1/4$. On the other hand, f has the directional derivative at $\bar{x}$ in each direction d:

$$\frac{\partial f}{\partial d}(\bar{x}) = \lim_{t \to 0} [t^3 d_1^2 d_2/t^2(t^2 d_1^4 + d_2^2)]^2/t = \lim_{t \to 0} t d_1^4 d_2^2/((t^2 d_1^4 + d_2^2)^2 = 0.$$

Definition 1.12. f is called *differentiable at* $\bar{x} \in \mathbb{R}^n$ if there exists a linear function $L : \mathbb{R}^n \to \mathbb{R}$ such that

$$\forall\ v \in \mathbb{R}^n\ :\ f(\bar{x} + v) = f(\bar{x}) + L(v) + r(v)$$

for some residual function r such that $r(v)/\|v\|_2 \to 0$ as $\|v\|_2 \to 0$. If f is differentiable at $\bar{x}$, L is called the *differential of* f *at* $\bar{x}$. Notice that both L and r depend not only on f but also on the considered point $\bar{x}$.
f is differentiable on a set $D \subseteq \mathbb{R}^n$ if it is differentiable at every $x \in D$.

Recall that $L : \mathbb{R}^n \to \mathbb{R}$ is linear if

$$\forall\ x, y \in \mathbb{R}^n\ \forall\ \alpha, \beta \in \mathbb{R}\ :\ L(\alpha x + \beta y) = \alpha L(x) + \beta L(y).$$

L is linear if and only if there exists $\ell \in \mathbb{R}^n$ such that $L(x) = \ell^T x$ for all $x \in \mathbb{R}^n$.

Proposition 1.6. *Suppose f is differentiable at $\bar{x} \in \mathbb{R}^n$. Then,*

(i) f is continuous at $\bar{x}$;

(ii) f has directional derivatives at $\bar{x}$ in each direction d and $\dfrac{\partial f}{\partial d}(\bar{x}) = L(d)$;

(iii) $L(d) = \nabla f(\bar{x})^T d$.

Proof. (i) It is enough to apply Definition 1.12 just taking $h = x - \bar{x}$ as $x \to \bar{x}$.
(ii) Take any direction $d \in \mathbb{R}^n$. Then, Definition 1.12 implies

$$\begin{aligned}\frac{\partial f}{\partial d}(\bar{x}) &= \lim_{t \to 0}\ (f(\bar{x} + td) - f(\bar{x}))/t \\ &= \lim_{t \to 0}\ (L(td) + r(td))/t \\ &= \lim_{t \to 0}\ (tL(d) + r(td))/t \\ &= L(d) + \lim_{t \to 0}\ r(td)/t \\ &= L(d) + \lim_{t \to 0}\ \operatorname{sgn}(t)\,(r(td))/\|td\|_2) = L(d).\end{aligned}$$

(iii) Since $d = \sum_{i=1}^{n} d_i e_i$, (ii) implies

$$\frac{\partial f}{\partial d}(\bar{x}) = L(d) = L\Big(\sum_{i=1}^{n} d_i e_i\Big) = \sum_{i=1}^{n} d_i L(e_i) = \sum_{i=1}^{n} d_i \frac{\partial f}{\partial x_i}(\bar{x}) = \nabla f(\bar{x})^T d. \qquad \square$$

Proposition 1.6 (iii) allows to restate the definition of differentiability through (the first order) Taylor's formula:

Taylor's formula $f(\bar{x} + v) = f(\bar{x}) + \nabla f(\bar{x})^T v + r(v) \quad (r(v)/\|v\|_2 \to 0)$

Considering any $v = x - \bar{x} \approx 0$, Taylor's formula states that $f(x)$ can be approximated by an affine function, namely $f(x) \approx f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x})$, and the closer x is to $\bar{x}$ the better the approximation is. Indeed, the set

$$\big\{\big(x, f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x})\big)\ :\ x \in \mathbb{R}^n\big\}$$

is the tangent hyperplane to the graph $\{(x, f(x))\ :\ x \in \mathbb{R}^n\}$ of f at $(\bar{x}, f(\bar{x}))$.

Theorem 1.4. *Let $\bar{x} \in \mathbb{R}^n$ and suppose f has all the partial derivatives at each $x \in B(\bar{x}, \varepsilon)$ for some $\varepsilon > 0$. Then, if the functions $x \mapsto \partial f(x)/\partial x_i$ are continuous at $\bar{x}$ for all $i = 1, ..., n$, then f is differentiable at $\bar{x}$.*

Example 1.10. Take $n = 2$ and

$$f(x_1, x_2) = \begin{cases} x_1^2 x_2/(x_1^2 + x_2^2) & \text{if } (x_1, x_2) \neq (0,0) \\ 0 & \text{if } (x_1, x_2) = (0,0) \end{cases}$$

and consider $\bar{x} = (0,0)$: f is continuous but not differentiable at $\bar{x}$. In fact, the derivative of f at $\bar{x}$ in the direction d is

$$\frac{\partial f}{\partial d}(\bar{x}) = \lim_{t \to 0} [t^3 d_1^2 d_2/t^2(d_1^2 + d_2^2)]/t = d_1^2 d_2$$

since $1 = \|d\|_2^2 = d_1^2 + d_2^2$. As a consequence, $\partial f(\bar{x})/\partial x_1 = \partial f(\bar{x})/\partial x_2 = 0$ while $\partial f(\bar{x})/\partial d \neq 0$ for all $d \neq e_1, e_2$ so that $\partial f(\bar{x})/\partial d \neq \nabla f(\bar{x})^T d$ (see Proposition 1.6).

Notice that

$$\frac{\partial f}{\partial x_1}(x) = 2x_1 x_2^3/(x_1^2 + x_2^2)^2 \qquad (x \neq \bar{x})$$

is not continuous at $\bar{x}$ (in accordance with Theorem 1.4): $x^k = (1/k, 1/k) \to \bar{x}$ while $\partial f(x^k)/\partial x_1 \equiv 1/2$ and $\partial f(\bar{x})/\partial x_1 = 0$.

Definition 1.13. f is called *continuously differentiable at $\bar{x} \in \mathbb{R}^n$* if there exists $\varepsilon > 0$ such that f is differentiable at each $x \in B(\bar{x}, \varepsilon)$ and the partial derivatives are continuous at $\bar{x}$. f is continuously differentiable on a set $D \subseteq \mathbb{R}^n$ if it is continuously differentiable at every $x \in D$.

Theorem 1.5. (mean value) *Suppose f is continuously differentiable (on $\mathbb{R}^n$). Given any $\bar{x}, v \in \mathbb{R}^n$, there exists $t \in]0,1[$ such that*

$$f(\bar{x} + v) = f(\bar{x}) + \nabla f(\bar{x} + tv)^T v.$$

Theorem 1.6. (upper estimate) *Suppose f is continuously differentiable (on $\mathbb{R}^n$) and the gradient mapping ∇f is Lipschitz with modulus $L > 0$, i.e.,*

$$\forall\, x, v \in \mathbb{R}^n \;:\; \|\nabla f(x) - \nabla f(v)\|_2 \leq L\|x - v\|_2.$$

Then, any $x, v \in \mathbb{R}^n$ satisfy $f(x + v) \leq f(x) + \nabla f(\bar{x} + v)^T v + L\|v\|_2^2/2$.

Proposition 1.7. (chain rules)

(i) If $g : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\bar{x} \in \mathbb{R}^n$ and $h : \mathbb{R} \to \mathbb{R}$ has a derivative at $f(\bar{x})$, then $f = h \circ g$ is differentiable at $\bar{x}$ and $\nabla f(\bar{x}) = h'(g(\bar{x}))\nabla g(\bar{x})$.

(ii) Let $h = (h_1, \ldots, h_n) : \mathbb{R} \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}$. If the functions $h_i : \mathbb{R} \to \mathbb{R}$ have a derivative at $\bar{t} \in \mathbb{R}$ for all $i = 1, ..., n$ and g is differentiable at $h(\bar{t}) \in \mathbb{R}^n$, then $g \circ h$ has a derivative at $\bar{t}$ and $(g \circ h)'(\bar{t}) = \nabla g(h(\bar{t}))^T h'(\bar{t})$ where $h'(\bar{t}) = (h_1'(\bar{t}), \ldots, h_n'(\bar{t}))^T$.

Definition 1.14. Let $F = (f_1, \dots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$. If the functions $f_i : \mathbb{R}^n \to \mathbb{R}$ have all the partial derivatives at $\bar{x} \in \mathbb{R}^n$ for all $i = 1, ..., n$, then

$$JF(\bar{x}) = \begin{bmatrix} \nabla f_1(\bar{x})^T \\ \vdots \\ \nabla f_n(\bar{x})^T \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\bar{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\bar{x}) \\ \vdots & \vdots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\bar{x}) & \cdots & \frac{\partial f_m}{\partial x_n}(\bar{x}) \end{bmatrix} \in \mathbb{R}^{m \times n}$$

is called the *Jacobian matrix of F at $\bar{x}$.*

1.3.3 Second-order derivatives

If a function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable on the whole $\mathbb{R}^n$, then each directional derivative exists at each point $x \in \mathbb{R}^n$. In this case, the derivative in the direction d is the function $\partial f/\partial d : \mathbb{R}^n \to \mathbb{R}$ such that $(\partial f/\partial d)(x) = \partial f(x)/\partial d$. If it has a derivative in the direction v, then

$$\frac{\partial}{\partial v}\Big(\frac{\partial f}{\partial d}\Big)(x) = \lim_{t \to 0} \Big[\frac{\partial f}{\partial d}(x + tv) - \frac{\partial f}{\partial d}(x)\Big]/t$$

is generally denoted by $\partial^2 f(x)/\partial v \partial d$.

Definition 1.15. f has *second-order partial derivatives* at $\bar{x} \in \mathbb{R}^n$ if it has the (first-order) partial derivatives at each $x \in B(\bar{x}, \varepsilon)$ for some $\varepsilon > 0$ and they have partial derivatives at $\bar{x}$ as well, namely

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \lim_{t \to 0} \Big[\frac{\partial f}{\partial x_j}(\bar{x} + tv) - \frac{\partial f}{\partial x_j}(\bar{x})\Big]/t$$

for all $i, j = 1, ..., n$. If $i = j$, then the derivative is generally denoted by $\partial^2 f(\bar{x})/\partial x_i^2$. For $n = 1$ there exists a unique second-order directional derivative which coincides with the (usual) second-order derivative and it is also denoted by $f''(\bar{x})$.

Example 1.11. Take the function of Example 1.7:

$$\frac{\partial f}{\partial x_1}(x) = \pi x_2 \cos(\pi x_1 x_2), \quad \frac{\partial f}{\partial x_2}(\bar{x}) = \pi x_1 \cos(\pi x_1 x_2),$$

$$\frac{\partial^2 f}{\partial x_2 \partial x_1}(x) = \pi \cos(\pi x_1 x_2) - \pi^2 x_1 x_2 \sin(\pi x_1 x_2) = \frac{\partial^2 f}{\partial x_1 \partial x_2}(x)$$

$$\frac{\partial^2 f}{\partial x_1^2}(x) = -\pi^2 x_2^2 \sin(\pi x_1 x_2), \quad \frac{\partial^2 f}{\partial x_2^2}(x) = -\pi^2 x_1^2 \sin(\pi x_1 x_2).$$

Theorem 1.7. (Schwarz) *Let $\bar{x} \in \mathbb{R}^n$ and suppose f has the second-order partial derivatives $\partial^2 f/\partial x_i \partial x_j$ and $\partial^2 f/\partial x_j \partial x_i$ at each $x \in B(\bar{x}, \varepsilon)$ for some $\varepsilon > 0$. If both the derivatives are continuous at $\bar{x}$, then*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\bar{x}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\bar{x}).$$

Definition 1.16. If f has *second-order partial derivatives* at $\bar{x} \in \mathbb{R}^n$, then

$$\nabla^2 f(\bar{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\bar{x}) & \cdots & \frac{\partial f}{\partial x_1 \partial x_n}(\bar{x}) \\ \vdots & \vdots & \vdots \\ \frac{\partial f}{\partial x_n \partial x_1}(\bar{x}) & \cdots & \frac{\partial f}{\partial x_n^2}(\bar{x}) \end{bmatrix} \in \mathbb{R}^{n \times n}$$

is called the *Hessian matrix of f at $\bar{x}$*.

Definition 1.17. f is called *twice continuously differentiable* at $\bar{x} \in \mathbb{R}^n$ if it has second-order partial derivatives at each $x \in B(\bar{x}, \varepsilon)$ for some $\varepsilon > 0$ and they are continuous at $\bar{x}$. f is twice continuously differentiable on a set $D \subseteq \mathbb{R}^n$ if it is twice continuously differentiable at every $x \in D$.

Notice that the Hessian matrix of a twice continuously differentiable function is symmetric and therefore all its eigenvalues are real numbers.

Theorem 1.8. (Taylor's formulas) *Suppose f is twice continuously differentiable (on $\mathbb{R}^n$). The following statements hold for any $\bar{x} \in \mathbb{R}^n$:*

(i) $\forall\ v \in \mathbb{R}^n\ \exists\ t \in]0,1[$ *such that* $f(\bar{x}+v) = f(\bar{x}) + \nabla f(\bar{x})^T v + \frac{1}{2} v^T \nabla^2 f(\bar{x}+tv) v$;

(ii) $\forall\ v \in \mathbb{R}^n\ :\ f(\bar{x}+v) = f(\bar{x}) + \nabla f(\bar{x})^T v + \frac{1}{2} v^T \nabla^2 f(\bar{x}) v + r(v)$
for some residual function r such that $r(v)/\|v\|_2^2 \to 0$ *as* $\|v\|_2 \to 0$.

Definition 1.18. f is called *quadratic* if there exist $Q \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$ such that

$$f(x) = \frac{1}{2} x^T Q x + b^T x + c = \frac{1}{2} \sum_{k=1}^{\ell} \sum_{\ell=1}^{n} q_{k\ell} x_k x_\ell + \sum_{k=1}^{n} b_k x_k + c.$$

Without loss of generality, Q can be taken symmetric, eventually replacing it by $(Q+Q^T)/2$ since $q_{k\ell} x_k x_\ell + q_{\ell k} x_\ell x_k = (q_{k\ell} + q_{\ell k}) x_k x_\ell / 2 + (q_{k\ell} + q_{\ell k}) x_\ell x_k / 2$.

The partial derivatives of a quadratic function can be easily computed:

$$\frac{\partial f}{\partial x_i}(x) = \frac{1}{2}\Big(\sum_{\ell=1}^{n} q_{i\ell} x_\ell + \sum_{k=1}^{n} q_{ki} x_k \Big) + b_i = \Big(\sum_{\ell=1}^{n} q_{i\ell} x_\ell \Big) + b_i = (Qx)_i + b_i$$

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(x) = \frac{\partial}{\partial x_j}\Big(\frac{\partial f}{\partial x_i}\Big)(x) = \frac{\partial f}{\partial x_j}\Big(\sum_{\ell=1}^{n} q_{i\ell} x_\ell + b_i \Big) = q_{ij}.$$

Therefore, $\nabla f(x) = Qx + b$ and $\nabla^2 f(x) = Q$.

Considering any $v = x - \bar{x} \approx 0$, the second-order Taylor's formula states that $f(x)$ can be approximated by a quadratic function, namely $f(x) \approx q(x)$ with

$$q(x) = f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + \frac{1}{2} (x - \bar{x})^T \nabla^2 f(\bar{x}) (x - \bar{x}),$$

that is

$$q(x) = \frac{1}{2}x^T\nabla^2 f(\bar{x})x + \big(\nabla f(\bar{x}) - \nabla^2 f(\bar{x})\bar{x}\big)^T x + \big(f(\bar{x}) - \nabla f(\bar{x})^T\bar{x} + \frac{1}{2}\bar{x}^T\nabla^2 f(\bar{x})\bar{x}\big).$$

Example 1.12. Take $n = 2$ and $f(x_1, x_2) = -x_1^4 - x_2^2$:

$$\nabla f(x) = \begin{pmatrix} -4x_1^3 \\ -2x_2 \end{pmatrix}, \quad \nabla^2 f(x) = \begin{bmatrix} -12x_1^2 & 0 \\ 0 & -2 \end{bmatrix}.$$

Considering $\bar{x} = (0, -2/5)$ the quadratic approximation of $f(x)$ near $\bar{x}$ is given by

$$q(x) = -2x_2^2 - 12x_2/5 - 20/25.$$