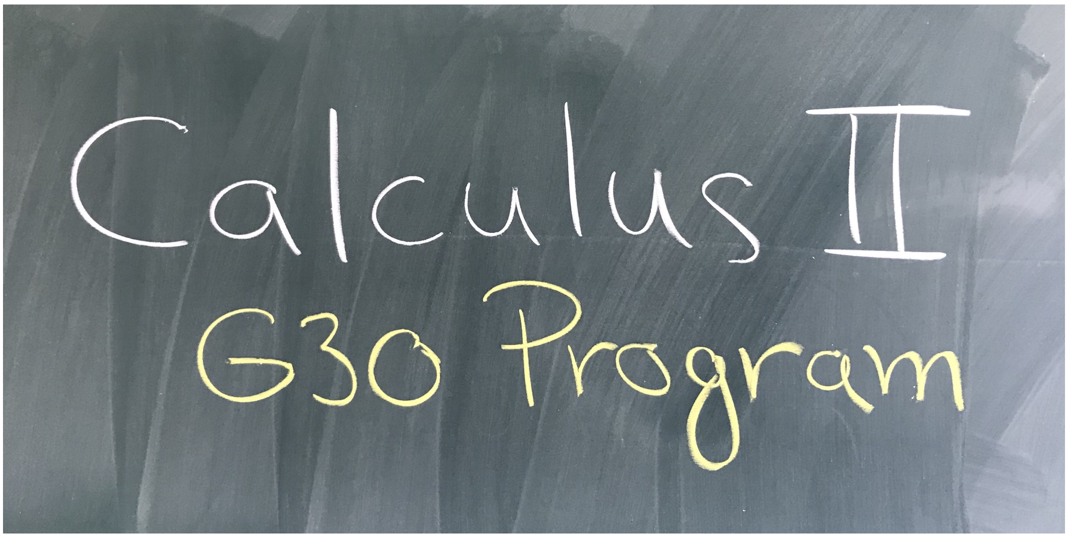
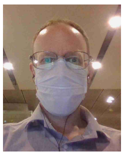


Spring Semester 2020



S. Richard

Revision: H. S. Nguyen and H. T. Bui

Contents

Chapter 1

Geometric setting

In this chapter we recall some basic notions on points and vectors in $\mathbb{R}^n$. The norm of a vector and the scalar product between two vectors are also introduced.

1.1 The Euclidean space $\mathbb{R}^n$

We set $\mathbb{N} := \{0, 1, 2, 3, \dots\}$ for the set of *natural numbers*, and let $\mathbb{R}$ be the set of all real numbers.

Definition 1.1. *One sets*

$$\mathbb{R}^n = \big\{(x_1, x_2, \dots, x_n) \mid x_j \in \mathbb{R} \textit{ for all } j \in \{1, 2, \dots, n\}\big\}^1.$$

Alternatively, an element of $\mathbb{R}^n$*, also called* a n-tuple *or* a vector*, is a collection of* n *numbers* $(x_1, x_2, \dots, x_n)$ *with* $x_j \in \mathbb{R}$ *for any* $j \in \{1, 2, \dots, n\}$*. The number* n *is called* the dimension of $\mathbb{R}^n$*.*

In the sequel, we shall often write $X \in \mathbb{R}^n$ for the vector $X = (x_1, x_2, \dots, x_n)$. With this notation, the values $x_1, x_2, \dots, x_n$ are called *the components* or *the coordinates* of X. Note that one often writes (x, y) for elements of $\mathbb{R}^2$ and (x, y, z) for elements of $\mathbb{R}^3$. However this notation is not really convenient in higher dimensions.

The set $\mathbb{R}^n$ can be endowed with two operations, *the addition* and *the multiplication by a scalar*.

Definition 1.2 (Addition and scalar multiplication)**.** *For any* $X, Y \in \mathbb{R}^n$ *with* $X = (x_1, x_2, \dots, x_n)$ *and* $Y = (y_1, y_2, \dots, y_n)$ *and for any* $\lambda \in \mathbb{R}$ *one defines the addition of* X *and* Y *by*

$$X + Y := (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n) \in \mathbb{R}^n$$

and the multiplication of X *by the scalar* λ *by*

$$\lambda X := (\lambda x_1, \lambda x_2, \dots, \lambda x_n) \in \mathbb{R}^n.$$

[1] The vertical line | has to be read "such that".

Examples 1.3. *(i)* $(1,3)+(2,4)=(3,7)\in\mathbb{R}^2$,

(ii) $(1,2,3,4,5)+(5,4,3,2,1)=(6,6,6,6,6)\in\mathbb{R}^5$,

(iii) $3(1,2)=(3,6)\in\mathbb{R}^2$,

(iv) $\pi(0,0,1)=(0,0,\pi)\in\mathbb{R}^3$.

One usually sets

$$\mathbf{0}=(0,0,\dots,0)\in\mathbb{R}^n$$

and this element satisfies $X+\mathbf{0}=\mathbf{0}+X=X$ for any $X\in\mathbb{R}^n$. If $X=(x_1,x_2,\dots,x_n)$ one also writes $-X$ for the element $-1X=(-x_1,-x_2,\dots,-x_n)$. Then, by an abuse of notation, one writes $X-Y$ for $X+(-Y)$ if $X,Y\in\mathbb{R}^n$, and obviously one has $X-X=\mathbf{0}$. Note that $X+Y$ is defined if and only if X and Y belong to $\mathbb{R}^n$, but has no meaning if $X\in\mathbb{R}^n$ and $Y\in\mathbb{R}^m$ with $n\neq m$.

Properties 1.4. *If $X,Y,Z\in\mathbb{R}^n$ and $\lambda,\mu\in\mathbb{R}$ then one has*

(i) $X+Y=Y+X$, (commutativity)

(ii) $(X+Y)+Z=X+(Y+Z)$, (associativity)

(iii) $\lambda(X+Y)=\lambda X+\lambda Y$, (distributivity)

(iv) $(\lambda+\mu)X=\lambda X+\mu X$,

(v) $(\lambda\mu)X=\lambda(\mu X)$.

1.2 Scalar product and norm

Definition 1.5 (Scalar product)**.** *For any $X,Y\in\mathbb{R}^n$ with $X=(x_1,x_2,\dots,x_n)$ and $Y=(y_1,y_2,\dots,y_n)$ one sets*

$$X\cdot Y:=x_1y_1+x_2y_2+\dots+x_ny_n=\sum_{j=1}^n x_jy_j$$

and calls this number the scalar product *between X and Y.*

For example, if $X=(1,2)$ and $Y=(3,4)$, then $X\cdot Y=1\cdot3+2\cdot4=3+8=11$, but if $X=(1,3)$ and $Y=(6,-2)$, then $X\cdot Y=6-6=0$. Be aware that the previous notation is slightly misleading since the dot $\cdot$ between X and Y corresponds to the scalar product while the dot between numbers just corresponds to the usual multiplication of numbers.

Properties 1.6. *For any $X,Y,Z\in\mathbb{R}^n$ and $\lambda\in\mathbb{R}$ one has*

(i) $X\cdot Y=Y\cdot X$,

(ii) $X \cdot (Y + Z) = X \cdot Y + X \cdot Z$,

(iii) $(\lambda X) \cdot Y = X \cdot (\lambda Y) = \lambda(X \cdot Y)$,

(iv) $X \cdot X \geq 0$, *and* $X \cdot X = 0$ *if and only if* $X = \mathbf{0}$.

Definition 1.7 (Euclidean norm)**.** *The* Euclidean norm *or simply* norm *of a vector* $X \in \mathbb{R}^n$ *is defined by*

$$\|X\| := \sqrt{X^2} \equiv \sqrt{X \cdot X} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

The positive number $\|X\|$ *is also referred to as the* magnitude *of* X. *A vector of norm* 1 *is called a* unit vector.

Example 1.8. *If* $X = (-1, 2, 3) \in \mathbb{R}^3$, *then* $X \cdot X = (-1)^2 + 2^2 + 3^2 = 14$ *and therefore* $\|X\| = \sqrt{14}$.

Remark 1.9. *If* $n = 2$ *or* $n = 3$ *this norm corresponds to our geometric intuition.*

Properties 1.10. *For any* $X \in \mathbb{R}^n$ *and* $\lambda \in \mathbb{R}$ *one has*

(i) $\|X\| = 0$ *if and only if* $X = \mathbf{0}$,

(ii) $\|\lambda X\| = |\lambda| \, \|X\|$,

(iii) $\| - X\| = \|X\|$.

Note that the third property is a special case of the second one.

1.3 Vectors and located vectors

Definition 1.11 (Located vector)**.** *For any* $A, B \in \mathbb{R}^n$ *we set* $\overrightarrow{AB}$ *for the arrow starting at* A *and ending at* B, *and call it the* located vector $\overrightarrow{AB}$, *see Figure 1.1.*

With this definition and for any $A \in \mathbb{R}^n$ the located vector $\overrightarrow{\mathbf{0}A}$ corresponds to the arrow starting at $\mathbf{0}$ and ending at A. This located vector is simply called *a vector* and is often identified with the element A of $\mathbb{R}^n$. This identification should not lead to any confusion in the sequel.

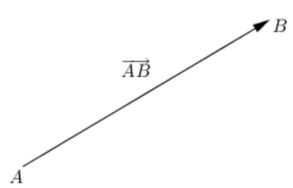


Fig. 1.1. A located vector

Definition 1.12. *Let* $A, B, C, D \in \mathbb{R}^n$ *and consider the located vectors* $\overrightarrow{AB}$ *and* $\overrightarrow{CD}$.

(i) $\overrightarrow{AB}$ *is* equivalent to $\overrightarrow{CD}$ *if* $B - A = D - C$,

(ii) $\overrightarrow{AB}$ *is* parallel to $\overrightarrow{CD}$ *if* $B - A = \lambda(D - C)$ *for some* $\lambda \in \mathbb{R}$,

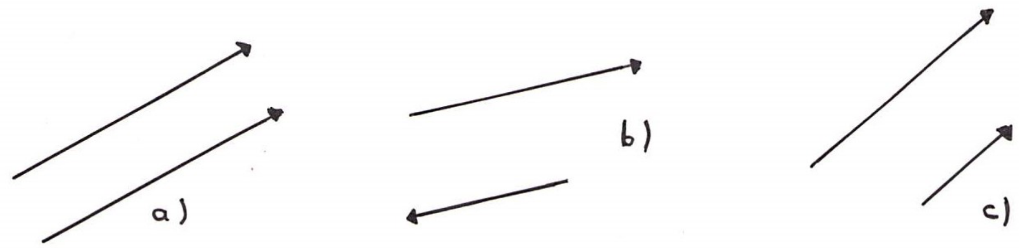


Fig. 1.2. Located vectors: a) equivalent, b) parallel, c) of same direction

(iii) $\overrightarrow{AB}$ *and* $\overrightarrow{CD}$ have the same direction *if* $B - A = \lambda(D - C)$ *for some* $\lambda > 0$.

Note that the located vector $\overrightarrow{AB}$ is always equivalent to the located vector $\overrightarrow{\mathbf{0}(B-A)}$ which is located at *the origin* $\mathbf{0}$. This fact follows from the equality

$$(B - A) - \mathbf{0} = (B - A) = B - A.$$

Definition 1.13. *The* norm *of a located vector* $\overrightarrow{XY}$ *is defined by* $\|\overrightarrow{XY}\| := \|Y - X\|$.

Note that this definition corresponds to our intuition in dimension $n = 2$ or $n = 3$ for the length of an arrow. It should also be observed that two equivalent located vectors have the same norm.

Definition 1.14. *Let* $A, B, C, D \in \mathbb{R}^n$.

(i) The vectors $\overrightarrow{\mathbf{0}A}$ *and* $\overrightarrow{\mathbf{0}B}$ *are* orthogonal *(or* perpendicular*) if* $A \cdot B = 0$. *In this case one writes* $\overrightarrow{\mathbf{0}A} \perp \overrightarrow{\mathbf{0}B}$ *or simply* $A \perp B$,

(ii) The located vectors $\overrightarrow{AB}$ *and* $\overrightarrow{CD}$ *are* orthogonal *if* $\overrightarrow{\mathbf{0}(B-A)}$ *and* $\overrightarrow{\mathbf{0}(D-C)}$ *are orthogonal, namely if* $(B - A) \cdot (D - C) = 0$. *In this case one writes* $\overrightarrow{AB} \perp \overrightarrow{CD}$, *see Figure 1.3.*

Fig. 1.3. Orthogonal vectors and orthogonal located vectors

Let us emphasize again that this notion corresponds to our intuition in dimension $n = 2$ or $n = 3$.

Example 1.15. *In $\mathbb{R}^n$ let us set $E_1 = (1, 0, \ldots, 0)$, $E_2 = (0, 1, 0, \ldots, 0)$, $\ldots$, $E_n = (0, \ldots, 0, 1)$ the n different vectors obtained by assigning a 1 at the coordinate j of E_j and 0 for all its other coordinates. Then, one easily checks that*

$$E_j \cdot E_k = 0 \text{ whenever } j \neq k \quad \text{and} \quad E_j \cdot E_j = 1 \text{ for any } j \in \{1, 2, \ldots, n\}.$$

These n vectors are said to be mutually orthogonal. *They generate a basis of* $\mathbb{R}^n$.

Theorem 1.16 (General Pythagoras theorem)**.** *For any $X, Y \in \mathbb{R}^n$ one has*

$$\|X + Y\|^2 = \|X\|^2 + \|Y\|^2 \quad \Longleftrightarrow \quad X \cdot Y = 0.$$

In other words, the equality on the left-hand side holds if and only if X and Y are orthogonal.

Let us add some other useful properties:

Lemma 1.17. *For any $X, Y \in \mathbb{R}^n$:*

(i) $|X \cdot Y| \leq \|X\| \, \|Y\|$,

(ii) $\|X + Y\| \leq \|X\| + \|Y\|$,

(iii) Let us recall from plane geometry that if one considers the triangle with vertices the points $\mathbf{0}$, X and Y, then the angle θ at the vertex $\mathbf{0}$ satisfies

$$\cos(\theta) = \frac{X \cdot Y}{\|X\| \, \|Y\|}.$$

1.4 Lines and hyperplanes

Let us consider $P, N \in \mathbb{R}^n$ with $N \neq \mathbf{0}$.

The set

$$L_{P,N} := \big\{P + tN \mid t \in \mathbb{R}\big\}$$

defines the line in $\mathbb{R}^n$ passing through P and having the direction parallel to $\overrightarrow{\mathbf{0}N}$.

The set

$$H_{P,N} := \big\{X \in \mathbb{R}^n \mid X \cdot N = P \cdot N\big\}$$

defines a *hyperplane* passing through P and normal to $\overrightarrow{\mathbf{0}N}$. Note that if $P = (p_1, p_2, \ldots, p_n)$ and if $N = (n_1, n_2, \ldots, n_n)$, then

$$H_{P,N} = \Big\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \mid n_1 x_1 + n_2 x_2 + \cdots + n_n x_n = \sum_{j=1}^{n} p_j n_j\Big\}.$$

One also observes that if $X \in H_{P,N}$ then the two located vectors $\overrightarrow{PX}$ and $\overrightarrow{ON}$ are orthogonal. Indeed, one has

$$\overrightarrow{PX} \perp \overrightarrow{\mathbf{0}N} \Longleftrightarrow \overrightarrow{\mathbf{0}(X - P)} \perp \overrightarrow{\mathbf{0}N} \Longleftrightarrow (X - P) \perp N = 0 \Longleftrightarrow X \cdot N = P \cdot N$$

with the last equality valid for any $X \in H_{P,N}$.

Chapter 2

Curve in $\mathbb{R}^d$

A special example of a function f from $\mathbb{R}^n$ to $\mathbb{R}^d$ is provided by a parametric curve. It corresponds to the special case $n = 1$. In the sequel, I is always denoted as an interval in $\mathbb{R}$. Whenever it is necessary, we shall mention a closed interval for $i = [a, b]$ or an open interval for $I = (a, b)$.

Definition 2.1 (Parametric curve)**.** *A* parametric curve *is a map $f : I \to \mathbb{R}^d$, where I is an interval in $\mathbb{R}$.*

Observe that, since for any $t \in I$ one has $f(t) \in \mathbb{R}^d$, it means

$$f(t) = \big(f_1(t), f_2(t), \dots, f_d(t)\big)$$

with $f_j(t) \in \mathbb{R}$ for any $j \in \{1, 2, \dots, d\}$. $f_j : I \to \mathbb{R}$ is called the j^{th}-component of f. Thus a parametric curve consists in a collection of d functions from I to $\mathbb{R}$.

Let us stress the difference between a parametric curve, as defined above, and the *curve* defined by

$$f(I) := \big\{f(t) \in \mathbb{R}^d \mid t \in I\big\}.$$

This curve is associated with the parametric curve f. In fact, one curve can be associated with several parametric curves, or in other terms a curve can be described by several parametrizations.

Example 2.2. *The functions*

$$f : \mathbb{R}_+ \to \mathbb{R}^2, \quad t \mapsto \big(\sin(t), t\big)$$

and

$$g : \mathbb{R}_+ \to \mathbb{R}^2, \quad t \mapsto \big(\sin(t^2), t^2\big)$$

define the same curve in $\mathbb{R}^2$, even if $f \neq g$.

Definition 2.3. *A parametric curve $f : I \to \mathbb{R}^d$, with I an open interval of $\mathbb{R}$, is of class C^k for some $k \in \mathbb{N}$ if each component $f_j : I \to \mathbb{R}$ of f is of class C^k, which means that each f_j is k-times differentiable, with its k-derivative continuous.*

If $f : I \to \mathbb{R}^d$ is of class C^1, we denote by f' of by $\dot{f}$ for its first derivative, which means $f' = (f_1', f_2', \dots, f_d')$. The vector $f'(t)$ is often interpreted as a located vector tangent to the curve at the point $f(t)$. The norm of $f'(t)$, namely $\|f'(t)\| \equiv \sqrt{f'(t) \cdot f'(t)}$ is called *the speed of the curve at* $f(t)$. Let us observe that this quantity depends on the parametrization. Indeed, in the Example 2.2 one has

$$\|f'(0)\| = \sqrt{\cos(0)^2 + 1} = \sqrt{2}$$

while

$$\|g'(0)\| = \sqrt{\big(2 \cdot 0 \cdot \cos(0)\big)^2 + (2 \cdot 0)^2} = 0.$$

In the sequel, we would like to have information on the curve which do not depend on the parametrization. First of all, we shall define a useful concept for reparametrizing a curve.

Definition 2.4 (Diffeomorphism)**.** *Let I, J be open intervals in $\mathbb{R}$, and let $k \in \mathbb{N}$. A map $\varphi : J \to I$ is a* diffeomorphism of class C^k *if*

(i) φ is of class C^k,

(ii) φ is invertible, with inverse denoted by $\varphi^{-1} : I \to J$,

(iii) φ^{-1} is of class C^k.

As an example of such a map, let us consider $\varphi : (0, \infty) \to \mathbb{R}$ with $\varphi(t) = t^3$. Then φ is a diffeomorphism of class C^k for any $k \in \mathbb{N}$.

Lemma 2.5. *Let $f : I \to \mathbb{R}^d$ be a parametric curve of class C^k and let $\varphi : J \to I$ be a diffeomorphism of class C^k. Then the composed map $f \circ \varphi : J \to \mathbb{R}^d$ is a parametric curve of class C^k describing the same curve as f.*

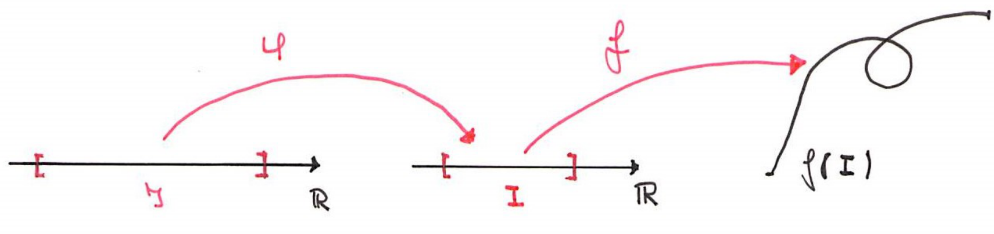


Fig. 2.1. Composition of the two maps

In the framework of the previous lemma, we say that the composed map is a *reparametrization* of the curve $f(I)$. The composed map belonging to class C^k is a simple consequence of the fact that the d functions $f_j \circ \varphi : J \to \mathbb{R}^d$ are of class C^k, for $j \in \{1, 2, \dots, d\}$.

Definition 2.6 (Length of a curve)**.** *Let $f : (a,b) \to \mathbb{R}^d$ be a parametric curve of class C^1. The length L_f of the corresponding curve $f\big((a,b)\big)$ is defined by*

$$L_f := \int_a^b \|f'(t)\| \mathrm{d}t. \tag{2.1}$$

Observe that in dimension 2 or 3, this above quantity corresponds to our intuition. If d is bigger than 3, we can consider this as a natural generalization. As we shall see in the next statement, the length of a curve is a quantity which is independent of the parametrization.

Proposition 2.7. *Let $f : (a,b) \to \mathbb{R}^d$ be a parametric curve of class C^1, and let $\varphi : (c,d) \to (a,b)$ be a diffeomorphism of class C^1. Then one has*

$$L_f = L_{f\circ\varphi}.$$

This statement is proved during the tutorial.

We shall now look at a special parametrization for a class of parametric curves which are called regular.

Definition 2.8 (Regular curve)**.** *A parametric curve $f : I \to \mathbb{R}^d$ is called* regular *at $t \in I$ if $f'(t) \neq \mathbf{0}$. f is* regular on I *if $f'(t) \neq \mathbf{0}$ for any $t \in I$.*

Let us now consider a parametric curve $f : [a,b] \to \mathbb{R}^d$ with is of class C^1 on (a,b) and regular on (a,b). Let us also define $\psi : (a,b) \to \mathbb{R}$ given for any $t \in (a,b)$ by

$$\psi(t) := \int_a^t \|f'(s)\| \mathrm{d}s.$$

Then, one observes that ψ is strictly increasing on $[a,b]$ and differentiable, with $\psi'(t) = \|f'(t)\| > 0$ for any $t \in (a,b)$, by assumption. Observe also that ψ has image equal to $[0, L_f]$ with L_f defined in (2.1). It then follows (see Calculus I) that ψ has an inverse, denoted by $\psi^{-1} : [0, L_f] \to [a,b]$ with ψ^{-1} differentiable on $(0, L_f)$ and with derivative given for any $s \in (0, L_f)$ by

$$\psi^{-1}(s)' = \frac{1}{\psi'\big(\psi^{-1}(s)\big)} = \frac{1}{\big\|f'\big(\psi^{-1}(s)\big)\big\|}.$$

Thus, if one sets

$$\varphi : [0, L_f] \to [a,b], \quad \varphi(s) := \psi^{-1}(s),$$

then φ is a diffeomorphism of class C^1 on $(0, L_f)$, and the composed map $f\circ\varphi : [0, L_f] \to \mathbb{R}^d$ satisfies

$$(f\circ\varphi)'(s) = f'\big(\varphi(s)\big)\,\varphi'(s) = f'\big(\varphi(s)\big)\frac{1}{\big\|f'\big(\psi^{-1}(s)\big)\big\|}$$

with norm

$$\Big\|(f\circ\varphi)'(s)\Big\| = \bigg\|f'\big(\varphi(s)\big)\frac{1}{\big\|f'\big(\psi^{-1}(s)\big)\big\|}\bigg\| = \frac{1}{\big\|f'\big(\varphi(s)\big)\big\|}\Big\|f'\big(\varphi(s)\big)\Big\| = 1.$$

In other words, the speed of $f\circ\varphi$ is always 1. We have thus proved the following statement:

Lemma 2.9. *For any parametric curve $f : [a, b] \to \mathbb{R}^d$ of class C^1 and regular on (a, b), there exists a new parametrization of the curve $f\big([a, b]\big)$ with constant speed equal to* 1.

The parametrization described in the previous statement and constructed above is called the *arc length parametrization*, or *the parametrization of the curve by its arc length*.

Chapter 3

Real functions of several variables

3.1 Open and closed sets in $\mathbb{R}^n$

The notions of open and closed sets play a minor but important role in this course. Let us start by introducing the notion of a ball.

Definition 3.1 (Open ball)**.** *For any real number $r > 0$ and any $X \in \mathbb{R}^n$ we define the ball*

$$\mathcal{B}_r(X) := \big\{Y \in \mathbb{R}^n \mid \|Y - X\| < r\big\}. \tag{3.1}$$

This notion corresponds to a ball in $\mathbb{R}^n$ of center X and of radius r. In fact this ball is *open* as we shall see with the next definition. The letter Ω (this is read *Omega*) will often be used for subset of $\mathbb{R}^n$.

Definition 3.2 (Open set)**.** *A subset $\Omega \subset \mathbb{R}^n$ is* open *if whenever $X \in \Omega$ there exists $r > 0$ such that $\mathcal{B}_r(X) \subset \Omega$.*

In other words, an open set Ω has the property that around any of its point X, one can put a small ball of center X which is included in the set Ω. For example, let us check that $\mathcal{B}_1(\mathbf{0})$ is open. Take any $X \in \mathcal{B}_1(\mathbf{0})$, which means that $\|X\| < 1$. Set now $r := \frac{1-\|X\|}{2} > 0$. Then one observes that $\mathcal{B}_r(X) \subset \mathcal{B}_1(\mathbf{0})$, which means that we have constructed a small ball of center X and of radius r which is entirely inside $\mathcal{B}_1(\mathbf{0})$. The same construction is valid for any open ball $\mathcal{B}_R(Y)$ for any $R > 0$ and any $Y \in \mathbb{R}^n$.

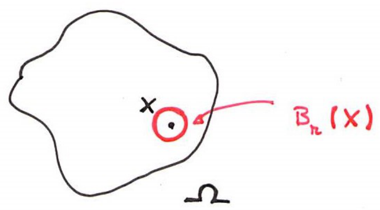


Fig. 3.1. An open set

The complementary notion is the notion of a closed set.

Definition 3.3 (Closed set)**.** *A subset $\Omega \subset \mathbb{R}^n$ is* closed *whenever its complement is open in $\mathbb{R}^n$, which means whenever*

$$\mathbb{R}^n \setminus \Omega := \{Y \in \mathbb{R}^n \mid Y \notin \Omega\}$$

is open.

First of all, one easily checks that if $n = 1$ the usual set (a, b) is open while the set $[a, b]$ is closed. On the other hand, the set $(a, b]$ is neither open nor closed. Indeed, there are plenty of sets which are neither open nor closed.

We provide some examples of open or closed sets. You are encourage to show that these sets are indeed open or closed.

Examples 3.4. *(i) The* closed ball

$$\overline{\mathcal{B}_r(X)} := \big\{Y \in \mathbb{R}^n \mid \|Y - X\| \leq r\big\} \tag{3.2}$$

is closed,

(ii) The set $\big\{(x, 0) \subset \mathbb{R}^2 \mid x \in [0, 1]\big\}$ *is closed in* $\mathbb{R}^2$*,*

(iii) The box $\big\{(x_1, \dots, x_n) \mid x_j \in [0, 1]$ *for* $j \in \{1, \dots, n\}\big\}$ *is closed in* $\mathbb{R}^n$*.*

In the sequel, our interest in open sets is due to the following heuristic property: the description of an open set in $\mathbb{R}^n$ uses the n variables. In other words, an open subset of $\mathbb{R}^n$ is really n-dimensional, while this is not always the case for arbitrary subsets of $\mathbb{R}^n$, like in the above example (ii).

3.2 Graph and level sets

In the sequel, we shall often use the notation Ω for a subset of $\mathbb{R}^n$, like I was used for an interval in the previous chapter.

Definition 3.5. *Let* $\Omega \subset \mathbb{R}^n$ *be open. A map* $f : \Omega \to \mathbb{R}$ *is called a (real) function of* n *variables.* Ω *is sometimes called the domain of* f*, and is denoted by* $\mathrm{Dom}(f)$*. The set* $f(\Omega) \subset \mathbb{R}$ *is called* the range *or* the image *of* f*.*

Examples 3.6.

(i) $f : \mathbb{R}^2 \to \mathbb{R}$ *with* $f(x, y) = x^2 + y^2$*,*

(ii) $f : \mathcal{B}_1(\mathbf{0}) \to \mathbb{R}$ *with* $f(x_1, \dots, x_n) := \frac{x_1}{1-\sqrt{x_1^2+\cdots+x_n^2}} = \frac{x_1}{1-\|X\|}$*,*

(iii) $f : \mathbb{R}^2 \setminus \{\mathbf{0}\} \to \mathbb{R}$ *with* $f(x, y) = \frac{x-y}{x^2+y^2}$*.*

In the previous chapter, we make a distinction between a curve and a parametric curve, in the present framework this consists in doing a distinction between a function and its graph.

Definition 3.7 (Graph of f)**.** *Let* $\Omega \subset \mathbb{R}^n$ *be open, and* $f : \Omega \to \mathbb{R}$*. The set*

$$\big\{\big(X, f(X)\big) \in \mathbb{R}^{n+1} \mid X \in \Omega\big\}$$

is called the graph of f*.*

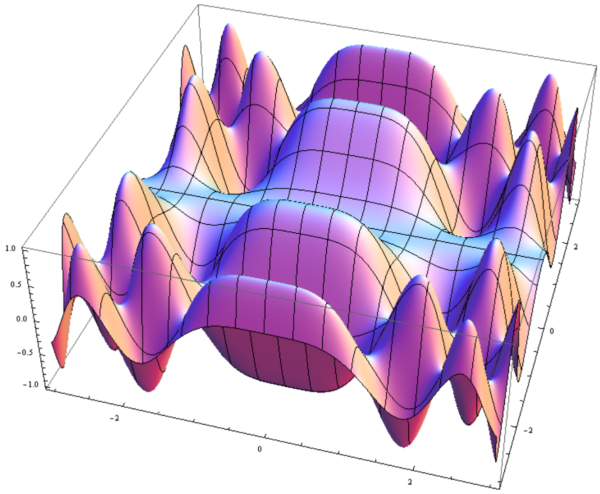


Fig. 3.2. The graph of $f(x,y) = \sin(x^2) \cdot \cos(y^2)$, see [2]

Let us emphasize that the graph of f is a subset of $\mathbb{R}^{n+1}$, not of $\mathbb{R}^n$. This notion of graph of a function corresponds to the one in Calculus I in the case $n = 1$, namely $\big\{\big(x, f(x)\big) \subset \mathbb{R}^2 \mid x \in \mathrm{Dom}(f)\big\}$. In the general case, the graph of a function looks like in Figure 3.2.

We now introduce a rather natural concept which is visible on many maps.

Definition 3.8 (Level sets)**.** *Let $\Omega \subset \mathbb{R}^n$ be open, let $f : \Omega \to \mathbb{R}$, and let $k \in \mathbb{R}$. We set*

$$L_k := \big\{X \in \Omega \mid f(X) = k\big\}$$

and call this set the k-level set of f*.*

Note that if $\Omega \subset \mathbb{R}^2$ one speaks sometimes about k-level curve, or if $\Omega \subset \mathbb{R}^3$ about k-level surface, but these terms might be misleading. Indeed, if $f : \mathbb{R}^2 \to \mathbb{R}$ is given by $f(x,y) = 3$, then $L_3 = \mathbb{R}^2$ which is not a curve, and $L_k = \emptyset$ for any $k \neq 3$.

Since L_k is a subset of Ω, which is in $\mathbb{R}^n$, the k-level sets can sometimes be more easily visualized than the graph of a function. For example, in Figures 3.3 and 3.4 the Himmelblau's function[1] and its k-level sets are provided. For information, the Himmelblau's function $f : \mathbb{R}^2 \to \mathbb{R}$ is given by

$$f(x,y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2.$$

Remark 3.9. *The k-level sets are often used, as for example in topographic maps where the curves represent places of equal elevation. In other fields, they are sometimes called equipotentials, or isothermal lines, see Figure 3.5.*

[1]The function is named after David Mautner Himmelblau (1924–2011) who introduced it, see [3].

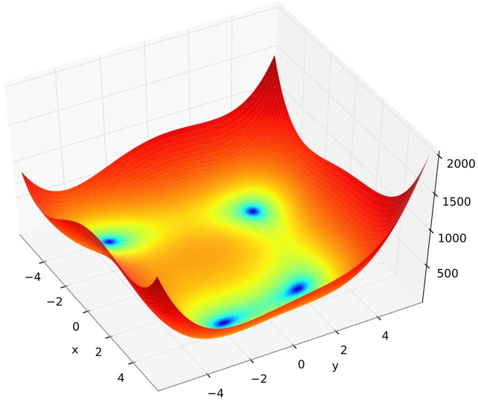


Fig. 3.3. The Himmelblau's function, see [3]

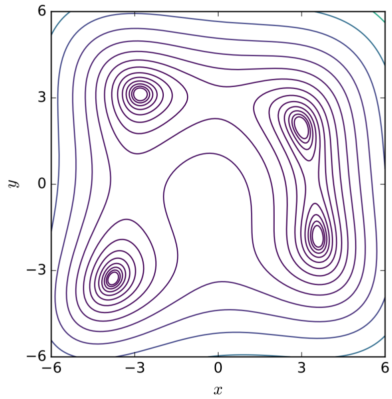
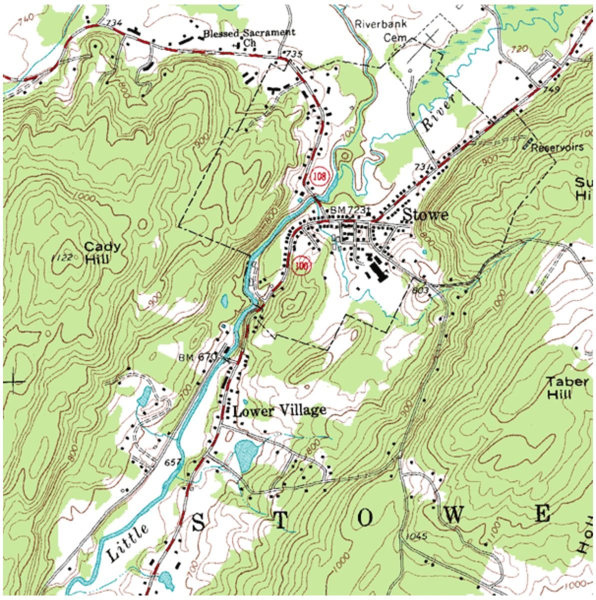
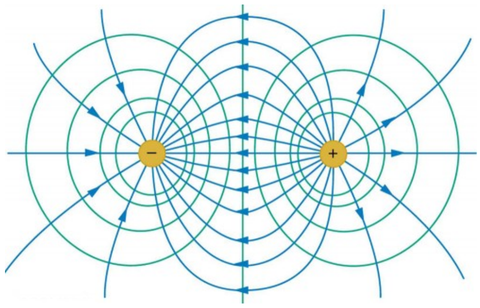


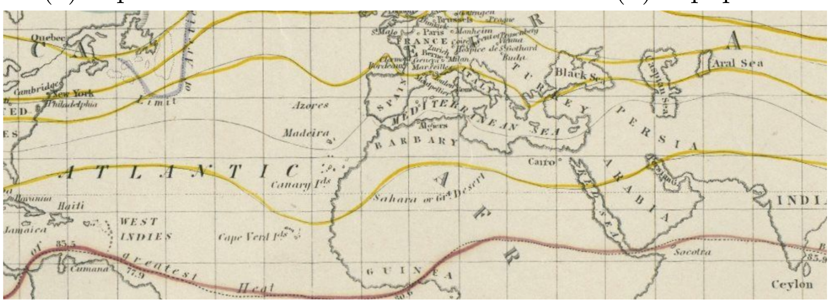
Fig. 3.4. Its k-level sets



(a) Equal elevation



(b) Equipotential



(c) Isothermal, see [4]

Fig. 3.5. Different k-level sets

3.3 Limits and continuity

Consider a function $f : \mathbb{R}^2 \setminus \{\mathbf{0}\} \to \mathbb{R}$. What can be the meaning of $\lim_{X\to\mathbf{0}} f(X)$? Which path converging to $\mathbf{0}$ do we consider ? In fact, one has to consider all of them.

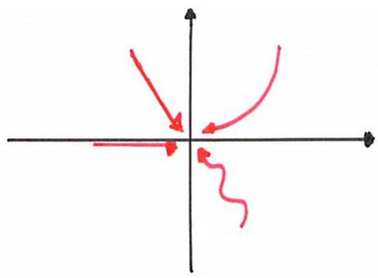


Fig. 3.6. Different paths to $\mathbf{0}$

Definition 3.10 (Limit). *Let $\Omega \subset \mathbb{R}^n$ be open with $X_0 \in \Omega$, and let $f : \Omega \setminus \{X_0\} \to \mathbb{R}$. The function f* has a limit at X_0 *if there exists a value $f(X_0) \in \mathbb{R}$ such that for any $\varepsilon > 0$ there exists a $\delta > 0$ with*

$$|f(X) - f(X_0)| \leq \varepsilon \qquad \text{for all } X \in \mathcal{B}_\delta(X_0).$$

If f has a limit at X_0 we write $\lim_{X\to X_0} f(X) = f(X_0)$.

Based on this definition, the notion of continuity is now quite clear.

Definition 3.11 (Continuity). *Let $\Omega \subset \mathbb{R}^n$ be open with $X_0 \in \Omega$, and let $f : \Omega \to \mathbb{R}$. The function f* is continuous at X_0 *if $\lim_{X\to X_0} f(X) = f(X_0)$, or more precisely if*

$$\forall \varepsilon > 0,\ \exists \delta > 0 \ \text{ such that } \ |f(X) - f(X_0)| \leq \varepsilon \ \text{ for all } X \in \mathcal{B}_\delta(X_0).$$

It is easily observed that in the special case of dimension 1, namely when $n = 1$, the previous two definitions correspond to the ones of Calculus I for limits and continuity of a function. Indeed, an open ball in dimension 1 is nothing but an open interval.

3.4 Partial derivatives

There are several natural notions of derivative for functions of more than one variable. We start with the simplest one, which consider the variables separately.

Definition 3.12 (Partial derivatives). *Let $\Omega \subset \mathbb{R}^n$ be open, and let $f : \Omega \to \mathbb{R}$.* The partial derivatives of f at *$X \in \Omega$ are defined by*

$$\lim_{\varepsilon\to 0} \frac{f(x_1 + \varepsilon, x_2, \dots, x_n) - f(x_1, x_2, \dots, x_n)}{\varepsilon}$$
$$\lim_{\varepsilon\to 0} \frac{f(x_1, x_2 + \varepsilon, \dots, x_n) - f(x_1, x_2, \dots, x_n)}{\varepsilon}$$
$$\vdots$$
$$\lim_{\varepsilon\to 0} \frac{f(x_1, x_2, \dots, x_n + \varepsilon) - f(x_1, x_2, \dots, x_n)}{\varepsilon}$$

whenever these limits exist.

If the partial derivatives of f at X exist, they are denoted by $\partial_j f(X)$ for $j \in \{1, \dots, n\}$, or also by $\frac{\partial f}{\partial x_j}(x_1, \dots, x_n)$. Let us emphasize that for each partial derivative, only one variable plays a role, the other variables are just spectators. For example, in the definition of $\partial_1 f$, only the variable x_1 is really important, the other ones play no role. Note finally that there exist n different partial derivatives of f at X, and that each $\partial_j(f)$ is a function from Ω to $\mathbb{R}$.

Examples 3.13. *(i) If $f : \mathbb{R}^3 \to \mathbb{R}$ is defined by $f(x_1, x_2, x_3) = x_1^3 x_2^2 x_3$, then $\frac{\partial f}{\partial x_1}(x_1, x_2, x_3) = 3x_1^2 x_2^2 x_3$, $\frac{\partial f}{\partial x_2}(x_1, x_2, x_3) = 2x_1^3 x_2 x_3$, and $\frac{\partial f}{\partial x_3}(x_1, x_2, x_3) = x_1^3 x_2^2$.*

(ii) If $f : \mathbb{R}^2 \to \mathbb{R}$ with $f(x, y) = \sin(xy) + \cos(x)$, then $\frac{\partial f}{\partial x}(x, y) = y \cos(xy) - \sin(x)$, and $\frac{\partial f}{\partial y}(x, y) = x \cos(xy)$.

Let's take a look at the second example. Since the variables are denoted by (x, y) and not (x_1, x_2), the notation for the partial derivatives has to be adapted. Also, since the second term does not depend on the variable y, this term disappear when the partial derivative with respect to y is taken.

There is one convenient notation for putting all the partial derivatives in one object, namely the gradient.

Definition 3.14 (Gradient)**.** *Let $\Omega \subset \mathbb{R}^n$ be open, and suppose that $f : \Omega \to \mathbb{R}$ admits n partial derivatives at $X \in \Omega$. Then one sets*

$$\nabla f(X) = \begin{pmatrix} \partial_1 f(X) \\ \partial_2 f(X) \\ \vdots \\ \partial_n f(X) \end{pmatrix}. \tag{3.3}$$

If the partial derivatives of f exist at every $X \in \Omega$ one simply writes $\nabla f = \begin{pmatrix} \partial_1 f \\ \partial_2 f \\ \vdots \\ \partial_n f \end{pmatrix}$, and observe that $\nabla f : \Omega \to \mathbb{R}^n$. This function ∇f is called the gradient of f.

Remark 3.15 (Column and row vectors)**.** *With* (3.3) *it is the first time that a column vector appears in these notes, so far only row vectors were used. Clearly, row vectors are simpler for the notations (and take less space), but from now on we shall make a difference between column vectors and row vectors, and try to be consistent. Note that for convenience we shall use the transpose, namely*

$${}^t(x_1, x_2, \dots, x_n) := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}. \tag{3.4}$$

Once a distinction is made between column and row vectors is made, one has to change the convention used so far: **All vectors which have been written as row vectors are in fact column vectors**. *It means that $\mathbb{R}^n$ has to be considered as made of column vectors, and so is the image of a parametric curve in $\mathbb{R}^d$. It also means that the scalar product is taken between two column vectors, and no two row vectors.*

The next statement follows directly from the definition of the partial derivatives and from the linearity of the derivativation process.

Lemma 3.16. *Let $\Omega \subset \mathbb{R}^n$ be open, and suppose that $f, g : \Omega \to \mathbb{R}$ admit n partial derivatives at any $X \in \Omega$. Then, for any $\lambda \in \mathbb{R}$ one has*

$$\nabla(f + \lambda g) = \nabla f + \lambda \nabla g.$$

As mentioned above, the partial derivatives do not really consider the n variables together. In that respect, this notion is too weak, and one has to define a slightly stronger notion.

3.5 Differentiability

First of all, recall that the natural basis $\{E_j\}_{j=1}^n$ of $\mathbb{R}^n$ has been introduced in Example 1.15. With this notation, the partial derivative can be rewritten:

$$\partial_j f(X) = \lim_{\varepsilon \to 0} \frac{f(X + \varepsilon E_j) - f(X)}{\varepsilon}$$

whenever this limit exists. Thus, one slightly more general notion can be defined, as in the next definition.

Definition 3.17 (Directional derivative)**.** *Let $\Omega \subset \mathbb{R}^n$ be open, let $f : \Omega \to \mathbb{R}$, and consider $V \in \mathbb{R}^n$ with $\|V\| = 1$. For any $X \in \Omega$ one sets*

$$D_V f(X) := \lim_{\varepsilon \to 0} \frac{f(X + \varepsilon V) - f(X)}{\varepsilon}$$

if this limit exists. If this limit exists, it is called the derivative of f in the direction V at X, *or simply* directional derivative.

It is clear that the notion above is a generalization of the partial derivatives. Indeed, if we fix $V = E_j$ for some $j \in \{1, \dots\}$, then $D_{E_j} f = \partial_j f$. Also note that the normalization $\|V\| = 1$ is not really necessary, some authors do not assume it, but only assume $V \neq \mathbf{0}$.

Even though the directional derivative is more general than partial derivatives, the more useful concept for functions of n variables is presented in the next definition.

Definition 3.18 (Differentiability). *Let $\Omega \subset \mathbb{R}^n$ be open, let $f : \Omega \to \mathbb{R}$. The function f is* differentiable at $X_0 \in \Omega$ *if f admits n partial derivatives at X_0 and if*

$$\lim_{X \to X_0} \frac{f(X) - f(X_0) - \nabla f(X_0) \cdot (X - X_0)}{\|X - X_0\|} = 0. \tag{3.5}$$

If f is differentiable at any $X \in \Omega$ we say that f is differentiable on Ω*.*

First of all, let us observe that (3.5) is meaningful. Indeed, if f admits n partial derivatives, then the gradient is well defined and $\nabla f(X_0)$ corresponds to a column vector. According to the convention made in Remark 3.15 the vector $X - X_0$ is also a column vector, and therefore the scalar product between two column vectors is well defined. It means that the numerator is indeed a difference of three numbers, and it is divided by a number. The condition (3.5) is that this ration goes to 0 as X approaches X_0.

Let us still observe that this condition corresponds to the usual one in Calculus I. Indeed, let $f : I \to \mathbb{R}$ be differentiable at x_0, with I an interval in $\mathbb{R}$ and $x_0 \in I$. This means that there exists $f'(x_0) \in \mathbb{R}$ with

$$\lim_{\varepsilon \to 0} \frac{f(x_0 + \varepsilon) - f(x_0)}{\varepsilon} - f'(x_0) = 0.$$

By multiplying by ε and by a trivial observation about the denominator the above condition is equivalent to

$$\lim_{\varepsilon \to 0} \frac{f(x_0 + \varepsilon) - f(x_0) - f'(x_0)\varepsilon}{|\varepsilon|} = 0.$$

Finally, if one consider x approaching x_0 and set $x - x_0 = \varepsilon$ this condition is equivalent to

$$\lim_{x \to x_0} \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{\|x - x_0\|} = 0,$$

which corresponds to (3.5) in the special case $n = 1$. It means that (3.5) is a natural generalization of the usual definition of differentiability in Calculus I, when more than one variable are involved.

Let us now show that the above notion of differentiability is stronger than the directional derivative.

Lemma 3.19. *Let $\Omega \subset \mathbb{R}^n$ be open, and let $f : \Omega \to \mathbb{R}$ be differentiable at $X_0 \in \Omega$. For any $V \in \mathbb{R}^n$ with $\|V\| = 1$ the directional derivative $D_V f(X_0)$ at X_0 exists and satisfies*

$$D_V f(X_0) = V \cdot \nabla f(X_0). \tag{3.6}$$

Before providing the proof, let us emphasize that the equality (3.6) might fail if f is not differentiable, even if f admits directional derivatives in all directions.

Proof. By the definition of the differentiability, and by setting $X := X_0 + \varepsilon V$ for $\varepsilon \in \mathbb{R}$ small enough, one has

$$\begin{aligned}
&\lim_{\varepsilon\to 0} \frac{f(X_0+\varepsilon V) - f(X_0) - \nabla f(X_0)\cdot \varepsilon V}{\|\varepsilon V\|} = 0 \\
\iff &\lim_{\varepsilon\to 0} \frac{f(X_0+\varepsilon V) - f(X_0) - \varepsilon V\cdot \nabla f(X_0)}{|\varepsilon|} = 0 \\
\iff &\lim_{\varepsilon\to 0} \frac{f(X_0+\varepsilon V) - f(X_0) - \varepsilon V\cdot \nabla f(X_0)}{\varepsilon} = 0 \\
\iff &\lim_{\varepsilon\to 0} \frac{f(X_0+\varepsilon V) - f(X_0)}{\varepsilon} - V\cdot \nabla f(X_0) = 0 \\
\iff &\lim_{\varepsilon\to 0} \frac{f(X_0+\varepsilon V) - f(X_0)}{\varepsilon} = V\cdot \nabla f(X_0).
\end{aligned}$$

Now, the l.h.s. corresponds to the definition of the directional derivative, which therefore exists, and it is equal to $V \cdot \nabla f(X_0)$, as mentioned in the statement. □

Let us also observe a rather trivial fact: any differentiable function at X_0 is continuous at X_0. Indeed one has

$$\begin{aligned}
&f(X) - f(X_0) \\
&= \|X - X_0\| \frac{f(X) - f(X_0) - \nabla f(X_0)\cdot (X - X_0) + \nabla f(X_0)\cdot (X - X_0)}{\|X - X_0\|} \\
&= \|X - X_0\| \left\{ \frac{f(X) - f(X_0) - \nabla f(X_0)\cdot (X - X_0)}{\|X - X_0\|} + \nabla f(X_0)\cdot \frac{X - X_0}{\|X - X_0\|} \right\}.
\end{aligned}$$

Then, it is enough to observe that $\frac{f(X)-f(X_0)-\nabla f(X_0)\cdot(X-X_0)}{\|X-X_0\|} \to 0$ as $X \to X_0$, and that $\frac{X-X_0}{\|X-X_0\|}$ is of norm 1. When $X \to X_0$, the factor $\|X - X_0\|$ vanishes, which implies that $f(X) - f(X_0) \to 0$ as $X \to X_0$. ♡

Let us still state one important result of differentiability, but refer to [1, Thm. 2, p. 68] for its proof.

Theorem 3.20. *Let $\Omega \subset \mathbb{R}^n$ be open, and let $f : \Omega \to \mathbb{R}$. Assume that the partial derivatives of f exist and are continuous on Ω. Then f is differentiable on Ω.*

Once the notion of one time differentiable has been defined, it is natural to wonder about higher order derivatives.

Definition 3.21 (Higher derivatives)**.** *Let $\Omega \subset \mathbb{R}^n$ be open, and let $f : \Omega \to \mathbb{R}$. The function f* is of class C^1 *if its partial derivatives $\partial_j f$ exist and are continuous on Ω. Similarly, the function f* is of class C^2 *if its n partial derivatives $\partial_j f$ are of class C^1. These derivatives are denoted by $\partial_k\partial_j f$ for $j, k \in \{1, \dots, n\}$. Recursively, the function f is of class C^k for some $k \in \mathbb{N}^*$ if it is of class C^{k-1} and the functions $\partial_{j_1}\partial_{j_2}\dots\partial_{j_{k-1}} f$ are of class C^1, for $j_1, j_2, j_{k-1} \in \{1, \dots, n\}$.*

Before moving to the next statement, think about the notation: $\partial_{j_1}\partial_{j_2}\dots\partial_{j_{k-1}}f$ for $j_1, j_2, j_{k-1} \in \{1,\dots,n\}$ is the only way to speak about all possible $k-1$ derivatives for a function of n variables. Then, there is a natural question, namely does the following equality holds: $\partial_1\partial_2 f = \partial_2\partial_1 f$? And more generally, does the equality

$$\partial_j\partial_k f = \partial_k\partial_j f$$

holds, for any $j,k \in \{1,\dots,n\}$? In general, the answer is no, and counterexamples exist.

Exercise 3.22. *Study the second derivatives of the function $f:\mathbb{R}^2\to\mathbb{R}$ defined by*

$$f(x,y) = \begin{cases} \frac{xy(x^2-y^2)}{x^2+y^2} & \text{if } (x,y)\neq(0,0), \\ 0 & \text{if } (x,y)=(0,0). \end{cases}$$

However, under suitable conditions the answer to the previous question is yes, as mentioned in the following statement.

Theorem 3.23. *Let $\Omega\subset\mathbb{R}^n$ be open, and let $f:\Omega\to\mathbb{R}$ be of class C^k for some $k\in\mathbb{N}$ with $k>1$. Then*

$$\partial_{j_1}\partial_{j_2}\dots\partial_{j_k}f = \partial_{\sigma(j_1)}\partial_{\sigma(j_2)}\dots\partial_{\sigma(j_k)}f$$

where σ is any permutation of the k indices.

In simpler terms, the above statement means that if f is of class C^k then k derivatives can be taken in an arbitrary order. The proof of this statement is provided (in a simpler form) in Section VI.1 page 109 of [1]. When $n=1$, the second derivative of a function plays an important role when one studies the local maximum or minimum of a function. What is the related concept for functions of several variables ?

Definition 3.24 (Hessian matrix)**.** *Let $\Omega\subset\mathbb{R}^n$ be open, and let $f:\Omega\to\mathbb{R}$ be of class C^2. The matrix $\mathcal{H}_f$ of second derivatives of f is called* the Hessian matrix *and is given by*

$$\mathcal{H}_f := \begin{pmatrix} \partial_1\partial_1 f & \partial_2\partial_1 f & \dots & \partial_n\partial_1 f \\ \partial_1\partial_2 f & \partial_2\partial_2 f & \dots & \partial_n\partial_2 f \\ \vdots & \vdots & \ddots & \vdots \\ \partial_1\partial_n f & \partial_2\partial_n f & \dots & \partial_n\partial_n f \end{pmatrix}, \tag{3.7}$$

or with a shorter notation: $\big(\mathcal{H}_f\big)_{jk} = \partial_k\partial_j f$.

Note that since f is of class C^2, it follows from Theorem 3.23 that $\partial_k\partial_j f = \partial_j\partial_k f$, which implies that the Hessian matrix is symmetric. For simplicity, we often write $\partial_j^2 f$ for $\partial_j\partial_j f$.

3.6 Taylor expansion

Before studying the meaning of Taylor expansion for functions of several variables, let us come back to the one dimensional case. In Calculus I one has learned that for a sufficiently differentiable function $f : I \to \mathbb{R}$ one has

$$f(x+h) = f(x) + \sum_{j=1}^{N} \frac{1}{j!} f^{(j)}(x) h^j + \frac{1}{(N+1)!} f^{(N+1)}(x+s) h^{N+1} \tag{3.8}$$

for some $s \in (0, h)$. Here we have assumed that x and $x+h$ belong to the open interval I, and used the notation $f^{(j)}$ for the j^{th} derivative of f. What is now the corresponding formula for $f : \Omega \to \mathbb{R}$? The main idea in the forthcoming construction is to come back to formula (3.8).

Let $\Omega \subset \mathbb{R}^n$ be an open set, and let $f : \Omega \to \mathbb{R}$ be of class C^{N+1} for some $N \in \mathbb{N}$. Let $X \in \Omega$ and consider $H \in \mathbb{R}^n$ with $H \neq \mathbf{0}$ such that for any $t \in [0,1]$ the point $X + tH$ belong to Ω. Note that such H always exists since Ω is open and since around any $X \in \Omega$ we can construct a small ball entirely included in Ω. We shall now study the function

$$g : [0,1] \ni t \mapsto g(t) := f(X+tH) \in \mathbb{R}.$$

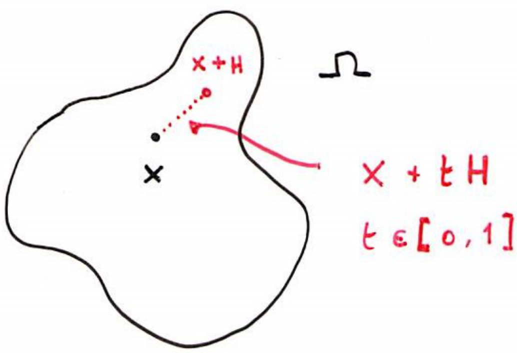


Fig. 3.7. Path between X and $X + H$

The main point with this function is that it is a function of one variable, defined on the interval $[0,1]$ and taking values in $\mathbb{R}$. It means that the Taylor expansion recalled in (3.8) applies to the function g. If we formally copy and paste the expansion, and then set $x = 0$ and $h = 1$, we get

$$g(1) = g(0) + \sum_{j=1}^{N} \frac{1}{j!} g^{(j)}(0) + \frac{1}{(N+1)!} g^{(N+1)}(s)$$

for some $s \in (0,1)$, or equivalently

$$f(X+H) = f(X) + \sum_{j=1}^{N} \frac{1}{j!} \frac{\mathrm{d}^j f(X+tH)}{\mathrm{d}t^j}\Big|_{t=0} + \frac{1}{(N+1)!} \frac{\mathrm{d}^{N+1} f(X+tH)}{\mathrm{d}t^{N+1}}\Big|_{t=s} \tag{3.9}$$

for some $s \in (0,1)$. However, it remains to compute explicitly the terms $\frac{\mathrm{d}^j f(X+tH)}{\mathrm{d}t^j}$. We provide the necessary information in a slightly more general lemma.

Lemma 3.25 (Chain rule). *Let $\Omega \subset \mathbb{R}^n$ be an open set, and let I be an open interval of $\mathbb{R}$. Let $f : \Omega \to \mathbb{R}$ be a differentiable function, and let $\varphi : I \to \Omega$ be a parametric curve which is also differentiable. Then the composed map $f \circ \varphi : I \to \mathbb{R}$ is differentiable and one has*

$$f\big(\varphi(t)\big)' = \frac{\mathrm{d}f\big(\varphi(t)\big)}{\mathrm{d}t} = \nabla f\big(\varphi(t)\big) \cdot \varphi'(t). \tag{3.10}$$

Before giving the proof, let us just observe that the scalar product is taking place between two column vectors, as it should be. For this, we recall the convention taken in Remark 3.15, and in particular that the image of any parametric curve corresponds to column vectors.

Proof. Let us first observe that the condition of differentiability provided in (3.5) can be rewritten equivalently in the following form

$$f(X+H) - f(X) - \nabla f(X) \cdot H = \|H\| \, q(H)$$

for some $q : \mathbb{R}^n \to \mathbb{R}$ with $q(H) \to 0$ as $H \to \mathbf{0}$. With this notation and by setting $X := \varphi(t)$ and : $H = \varphi(t+h) - \varphi(t)$ one has

$$\begin{aligned}
f\big(\varphi(t)\big)' &= \lim_{h\to 0} \frac{f\big(\varphi(t+h)\big) - f\big(\varphi(t)\big)}{h} \\
&= \lim_{h\to 0} \frac{f(X+H) - f(X)}{h} \\
&= \lim_{h\to 0} \frac{\nabla f(X) \cdot H + \|H\| q(H)}{h} \\
&= \lim_{h\to 0} \nabla f\big(\varphi(t)\big) \cdot \frac{\varphi(t+h) - \varphi(t)}{h} + \frac{\|\varphi(t+h) - \varphi(t)\|}{h} q(H) \\
&= \nabla f\big(\varphi(t)\big) \cdot \varphi'(t),
\end{aligned}$$

since $\lim_{h\to 0} \frac{\varphi(t+h)-\varphi(t)}{h} = \varphi'(t)$ and since $H \to 0$ as $h \to 0$. Note that

$$\lim_{h\to 0_\pm} \frac{\|\varphi(t+h) - \varphi(t)\|}{h} = \pm\|\varphi'(t)\|$$

but that this sign does not play any role since this factor is multiplied by $q(H)$ which vanishes as H goes to $\mathbf{0}$. □

Let now come back to the initial Taylor expansion, and state the result precisely.

Theorem 3.26. *Let $\Omega \subset \mathbb{R}^n$ be an open set, and let $f : \Omega \to \mathbb{R}$ be of class C^{N+1} for some $N \in \mathbb{N}$. Let $H \in \mathbb{R}^n$ such that $X + tH \in \Omega$ for any $t \in [0,1]$. Then one has*

$$f(X+H) = f(X) + \sum_{j=1}^{N} \frac{1}{j!}\big[(H\cdot\nabla)^j f\big](X) + \frac{1}{(N+1)!}\big[(H\cdot\nabla)^{N+1} f\big](X+sH) \quad (3.11)$$

for some $s \in (0,1)$.

Let us explain the meaning of the notation $(H \cdot \nabla)^j f$. For $j = 1$ and if $H = {}^t(h_1, \dots, h_n)$ one has

$$H \cdot \nabla f = \sum_{j=1}^{n} h_j \partial_j f$$

which is again a function from Ω to $\mathbb{R}$. It means that we can again take the gradient of this function and consider $\nabla\big(H\cdot\nabla f\big)$, which is nothing but the gradient of the function $H\cdot\nabla f:\Omega\to\mathbb{R}$. Then, the expression

$$\big(H\cdot\nabla\big)^2 f \equiv H\cdot\nabla\big(H\cdot\nabla f\big)$$

is again well defined and is a function from Ω to $\mathbb{R}$. The construction can continue, and provides a meaning to $(H\cdot\nabla)^j f$ for any j.

Proof of Theorem 3.26. The proof consists simply in defining the parametric curve

$$\varphi:[0,1]\ni t\mapsto\varphi(t):=X+tH\in\Omega$$

and to compute the different terms in (3.9). Clearly, by the previous lemma one has

$$\begin{aligned}\frac{\mathrm{d}f(X+tH)}{\mathrm{d}t}&=\frac{\mathrm{d}f\big(\varphi(t)\big)}{\mathrm{d}t}\\&=\nabla f\big(\varphi(t)\big)\cdot\varphi'(t)\\&=\nabla f(X+tH)\cdot H\\&=H\cdot\nabla f(X+tH).\end{aligned}$$

Similarly, by applying several times the same lemma one also gets

$$\frac{\mathrm{d}^j f(X+tH)}{\mathrm{d}t^j}=\big[(H\cdot\nabla)^j f\big](X+tH).$$

Then, the statement is obtained either by evaluating these terms at $t=0$, namely $\frac{\mathrm{d}^j f(X+tH)}{\mathrm{d}t^j}\big|_{t=0}=\big[(H\cdot\nabla)^j f\big](X)$ or by evaluating the last term at $t=s$ for some $s\in(0,1)$. $\square$

As a final observation for this section, let us rewrite the first few terms of this expansion in a slightly different form. Consider first the term corresponding to $j=2$ and for $H={}^t(h_1,\dots,h_n)$, namely

$$\begin{aligned}\big[(H\cdot\nabla)^2 f\big](X)&=\Big[(H\cdot\nabla)\Big(\sum_{j=1}^n h_j\partial_j f\Big)\Big](X)\\&=\Big[\sum_{k=1}^n h_k\partial_k\Big(\sum_{j=1}^n h_j\partial_j f\Big)\Big](X)\\&=\Big[\sum_{j,k=1}^n j_j h_k\partial_j\partial_k f\Big](X)\\&={}^tH\mathcal{H}_f(X)H\end{aligned}$$

with $\mathcal{H}_f$ the Hessian matrix defined in (3.7). Note that ${}^tH\mathcal{H}_f(X)H$ can be understood as the product of three matrices: ${}^tH\in M_{1n}(\mathbb{R})$, $\mathcal{H}_f(X)\in M_{nn}(\mathbb{R})$ and $H\in M_{n1}(\mathbb{R})$.

Here we have used the notation $M_{mn}(\mathbb{R})$ for any $m \times n$ matrix with entries in $\mathbb{R}$. With this notation, then the first few terms of the Taylor expansion (3.11) can be rewritten as

$$f(X + H) = f(X) + H \cdot \nabla f(X) + \tfrac{1}{2}{}^{t}H\mathcal{H}_f(X)H + \dots . \tag{3.12}$$

This notation will be useful in the next section, when local minimum and local maximum are studied.

3.7 Maximum and minimum

Let us start by recalling some known facts from Calculus I. We consider a function $f : I \to \mathbb{R}$. For this function, $x \in I$ is a local maximum if there exists $\varepsilon > 0$ with $f(x) \geq f(y)$ for any $y \in (x - \varepsilon, x + \varepsilon) \cap I$, while x is a local minimum if there exists $\varepsilon > 0$ with $f(x) \leq f(y)$ for any $y \in (x - \varepsilon, x + \varepsilon) \cap I$. In addition, if I is open, and the function f is differentiable on I, then the set of critical points of f is given by $\{x \in I \mid f'(x) = 0\}$. In this framework, it has been proved in Calculus I that a local maximum or a local minimum belongs to the set of critical points. However, let us also recall that not all critical points are local maximum or local minimum. For example, if $f(x) = x^3$, the point 0 is a critical point, but not a local extremum.

It is now natural to wonder what is the corresponding situation for $f : \Omega \to \mathbb{R}$ when Ω is an open subset of $\mathbb{R}^n$?

Definition 3.27. *Let Ω be a subset of $\mathbb{R}^n$, and let $f : \Omega \to \mathbb{R}$. A point $X \in \Omega$ is* a local maximum for f *if there exists $r > 0$ such that $f(X) \geq f(Y)$ for any $Y \in \mathcal{B}_r(X) \cap \Omega$, while $X \in \Omega$ is* a local minimum for f *if there exists $r > 0$ such that $f(X) \leq f(Y)$ for any $Y \in \mathcal{B}_r(X) \cap \Omega$.*

Observe that in the previous definition, no condition is imposed on Ω, and no condition of continuity or differentiability is imposed on f. An illustration of these extrema is provided in Figure 3.8. On the other hand, the notion of critical points require that Ω is open (in order to take the derivatives) and that f is differentiable.

Definition 3.28 (Critical points)**.** *Let $\Omega \subset \mathbb{R}^n$ be an open set, and let $f : \Omega \to \mathbb{R}$ be of differentiable function. A point $X \in \Omega$ is* a critical point for f *if $\nabla f(X) = \mathbf{0}$, which means that $\partial_j f(X) = 0$ for all $j \in \{1, \dots, n\}$.*

We stress that the condition holds if all partial derivatives of f at X are equal to 0. It is not sufficient that some of them are 0. The statement similar to the one dimensional case now reads:

Lemma 3.29. *Let $\Omega \subset \mathbb{R}^n$ be an open set, and let $f : \Omega \to \mathbb{R}$ be of differentiable function. If $X \in \Omega$ is a local extremum, then X is a critical point.*

Proof. We suppose that X is a local maximum, the proof for a local minimum being similar. Consider $r > 0$ such that $\mathcal{B}_r(X) \subset \Omega$, which is always possible since Ω is open.

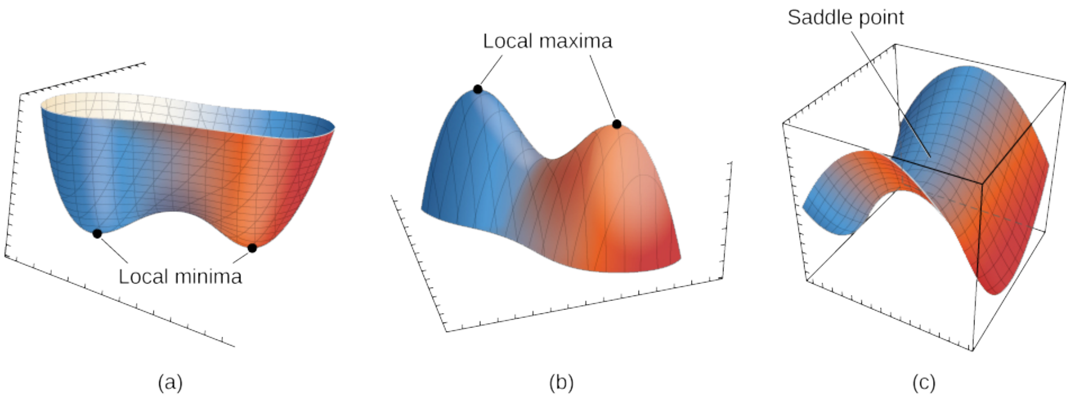


Fig. 3.8. Various critical points, see [5]

Choose $j \in \{1, \dots, n\}$ and consider the function $\varphi_j : (-r, r) \ni t \mapsto X + tE_j \in \Omega$. Then the maps

$$f \circ \varphi_j : (-r, r) \ni t \mapsto f(X + tE_j) \in \mathbb{R}$$

are differentiable and have a local maximum at $t = 0$. By what has been recalled for a function of 1 variable, one infers that $\frac{\mathrm{d}f(\varphi_j(t))}{\mathrm{d}t}\big|_{t=0} = 0$. By the chain rule presented in Lemma 3.25, one infers that

$$0 = \frac{\mathrm{d}f\,(\varphi_j(t))}{\mathrm{d}t}\Big|_{t=0} = E_j \cdot \nabla f(X) = \partial_j f(X).$$

Since this equality holds for any $j \in \{1, \dots, n\}$, one deduces that $\nabla f(X) = \mathbf{0}$, as stated. □

The above statement gives some information about the local maximum or minimum which is inside Ω. A situation similar to the one already seen in Calculus I can also take place: If Ω is not open, then local maxima or minima can appear on the boundary of Ω, and these points are not necessarily critical points. The study of these points can only be done on a case-by-case basis. We shall mention this situation again at the end of this section.

Let us come back to the one dimensional case $f : I \to \mathbb{R}$ which is twice differentiable on the open interval I. Once the critical points are listed, their nature can be studied by looking at the second derivative of f. Indeed, it is known that if $x \in I$ is a local minimum if $f''(x) \geq 0$, while x is a local maximum if $f''(x) \leq 0$. This property can easily be inferred by looking at the Taylor expansion provided in 3.8, namely

$$\begin{aligned} & f(x+h) = f(x) + f'(x)h + \tfrac{1}{2}f''(x)h^2 + \dots \\ \Longleftrightarrow\, & f(x+h) - f(x) = \tfrac{1}{2}f''(x)h^2 + \dots \end{aligned} \tag{3.13}$$

since $f'(x) = 0$ (x is supposed to be a critical point). Clearly, $f(x+h)$ is bigger than $f(x)$ if $f''(x) \geq 0$, which means that f takes locally the smallest value at x.

Remark 3.30. *Let us take this opportunity to introduce one more concept. The notation "…" at the end of the previous argument is not really precise. One can do better by using for any $k \geq 0$ the notation $O(|h|^k)$ and $o(|h|^k)$. A function f belongs to $O(|h|^k)$ near 0 if $\frac{1}{|h|^k}|f(h)| \leq c$ for all h near 0 and a fixed constant $c < \infty$, while $f \in o(|h|^k)$ if $\lim_{h\to 0} \frac{1}{|h|^k} f(h) = 0$. With this notation, if the function f in (3.8) is of class C^{N+1}, then the remainder term belongs to $O(|h|^{N+1})$. Similarly, if f is of class C^3 in (3.13) then "…" could be replaced by $O(|h|^3)$.*

Let us now come to the n-dimensional case. We consider $f : \Omega \to \mathbb{R}$ with Ω an open subset of $\mathbb{R}^n$ and f of class C^3. Starting from the expression provided in (3.12), and by assuming that $X \in \Omega$ is a critical point for f one gets

$$
\begin{aligned}
&f(X+H) = f(X) + H \cdot \nabla f(X) + \tfrac{1}{2}{}^t H \mathcal{H}_f(X) H + O(\|H\|^3) \\
\Longleftrightarrow\, &f(X+H) - f(X) = \tfrac{1}{2}{}^t H \mathcal{H}_f(X) H + O(\|H\|^3).
\end{aligned}
\tag{3.14}
$$

The term $\frac{1}{2}{}^t H \mathcal{H}_f(X) H$ is called *quadratic term* at the critical point, and the nature of the critical point X will depend on the sign that this quadratic term can take. We shall now have a look at this trough a few examples.

Examples 3.31. *In these examples, we always assume that $X \in \Omega$ is a critical point for a function $f : \Omega \to \mathbb{R}$ with Ω an open subset of $\mathbb{R}^n$.*

(i) For $n = 2$ and $\mathcal{H}_f(X) = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$, and if we set $H = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$ then one has

$$
{}^t H \mathcal{H}_f(X) H = \begin{pmatrix} h_1 & h_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = h_1^2 + 2h_2^2
$$

with $h_1^2 + 2h_2^2 > 0$ whenever $\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix} \equiv \mathbf{0}$. It then follows from (3.14) that $f(X+H) - f(X) > 0$ for any $H \neq \mathbf{0}$ with $\|H\|$ small enough. Thus X is a local minimum in this case.

(ii) For $n = 3$ and $\mathcal{H}_f(X) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -3 \end{pmatrix}$, if we set $H = \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix}$ then one has

$$
{}^t H \mathcal{H}_f(X) H = \begin{pmatrix} h_1 & h_2 & h_3 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -3 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} = -h_1^2 - 2h_2^2 - 3h_3^2
$$

with $-h_1^2 - 2h_2^2 - 3h_3^2 < 0$ whenever $\begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} \neq \mathbf{0}$. It then follows from (3.14) that $f(X+H) - f(X) < 0$ for any $H \neq \mathbf{0}$ with $\|H\|$ small enough. Thus X is a local maximum in this case.

(iii) *For* $n = 2$ *and* $\mathcal{H}_f(X) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, *and if we set* $H = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$ *then one has*

$$^tH\mathcal{H}_f(X)H = \begin{pmatrix} h_1 & h_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = h_1^2 - h_2^2.$$

In this case, the term $^tH\mathcal{H}_f(X)H$ *can take positive values, as for example for* $H = \begin{pmatrix} \varepsilon \\ 0 \end{pmatrix}$, *or negative values, as for example for* $H = \begin{pmatrix} 0 \\ \varepsilon \end{pmatrix}$ *for any* $\varepsilon \neq 0$. *It means that* $f(X + H) - f(X)$ *can be positive or negative, depending on* H. *In this case, we say that* X *is* a saddle point *or a* mountain pass point. *Indeed, starting from* X *the function* f *increases in some direction, and decreases in other directions, see Figure 3.8.*

Based on these examples, what can be said in the general case ?

First of all, a result usually proved in Linear algebra has be recalled. Note that $M_n(\mathbb{R})$ denotes the set of $n \times n$ matrices with entries in $\mathbb{R}$.

Theorem 3.32. *For any symmetric matrix* $\mathcal{A} \in M_n(\mathbb{R})$ *there exists an orthogonal matrix* $\mathcal{B} \in M_n(\mathbb{R})$ *such that*

$$\mathcal{B}\mathcal{A}\mathcal{B}^t = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{pmatrix} \tag{3.15}$$

with $\lambda_j \in \mathbb{R}$ *the eigenvalues of the matrix* $\mathcal{A}$, *and* $\mathcal{B}^t$ *the transpose of* $\mathcal{B}$.

Based on this result, the general situation of a critical point can be studied. We recall that for a function of class C^2 the Hessian matrix $\mathcal{H}_f$ is a symmetric matrix. This fact means that the above theorem can be applied to $\mathcal{H}_f(X)$.

Theorem 3.33. *Let* $\Omega \subset \mathbb{R}^n$ *be open, and let* $f : \Omega \to \mathbb{R}$ *be of class* C^2. *Let* $X \in \Omega$ *be a critical point of* f. *The following situations can take place:*

(i) *If all eigenvalues of* $\mathcal{H}_f(X)$ *are strictly positive, then* X *is a local minimum,*

(ii) *If all eigenvalues of* $\mathcal{H}_f(X)$ *are strictly negative, then* X *is a local maximum,*

(iii) *If some eigenvalues of* $\mathcal{H}_f(X)$ *are strictly negative, and the remaining ones are strictly positive, then* X *is a saddle point (mountain pass point),*

(iv) *If* $\mathcal{H}_f(X)$ *admits some* 0 *eigenvalues, futher ananlysis is necessary.*

The first three situations have already been illustrated in the examples above. For the last case, this situation already appears for function of 1 variable. For example, the function $f : x \mapsto x^4$ has a local minimum at 0, even if $f'(x) = f''(0) = 0$, while the function $g : x \mapsto x^3$ also satisfies $g'(0) = g''(0) = 0$ but possesses a saddle point at 0.

Idea of the proof. Set $\mathcal{A} := \mathcal{H}_f(X)$ and call $\mathcal{D}$ the diagonal matrix mentioned in (3.15). Observe then that

$$^tH\mathcal{A}H = {}^tH\mathcal{B}^t\mathcal{D}\mathcal{B}H = {}^t(\mathcal{B}H)\mathcal{D}(\mathcal{B}H) = {}^t(\mathcal{B}H)\begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{pmatrix}(\mathcal{B}H).$$

Since $\|\mathcal{B}H\| \to 0$ as $\|H\| \to 0$, the analysis performed in Examples 3.31 with the diagonal matrices can be mimicked, and one gets the statement. □

Let us end this section with an additional analogy with functions of 1 variable. Recall that if $f : [a, b] \to \mathbb{R}$ is differentiable on (a, b) and continuous on $[a, b]$, then its exterma can be taken at some $x \in (a, b)$ when $f'(x) = 0$, or at a or at b. What happens in n dimensions ?

Let us first introduce the boundary of a subset $\Omega \subset \mathbb{R}^n$. We say that $X \in \mathbb{R}^n$ belongs to the boundary $\partial\Omega$ of Ω if for any $r > 0$, $\mathcal{B}_r(X) \cap \Omega \neq \emptyset$ and $\mathcal{B}_r(X) \cap (\mathbb{R}^n \setminus \Omega) \neq \emptyset$. In other words, any ball centered at a point of the boundary of Ω has a non-empty intersection with Ω and with the complement of Ω. Note that a set Ω is closed precisely when $\partial\Omega \subset \Omega$ while Ω is an open set precisely when $\partial\Omega \cap \Omega = \emptyset$. Furthermore, the interior of Ω is defined by $\{X \in \Omega \mid \mathcal{B}_r(X) \subset \Omega \text{ for some } r > 0\}$ and is often denoted by Ω^o. Clearly, the equality $\partial\Omega \cap \Omega^o = \emptyset$ always holds.

Consider now $f : \Omega \to \mathbb{R}$ with no special condition on Ω. Then the extrema of f are located either at some critical points $X \in \Omega^o$ or on the boundary $\partial\Omega \cap \Omega$. Here, we take the intersection with Ω since the function f is only defined on Ω and might not be defined at all points of $\partial\Omega$. Depending on the situation, the study of the extrema on $\partial\Omega \cap \Omega$ might then be rather complicated, and one might have to look at the boundary of this set, and then the boundary of the boundary of this set... On concrete examples, the situation is usually simpler, as for example in the following exercise.

Exercise 3.34. *Study the extrema of the function $f : \Omega \to \mathbb{R}$ defined for $\Omega = [-1, 1] \times [-1, 1]$ by $f(x, y) = x^2 + y^2 - 2x - 2y + 2$.*

3.8 Lagrange multipliers

Let us introduce one more concept related to local extrema: an extremum under a constraint given by an equality. The approach is based on the so-called *Lagrange multipliers*, a method which has been extensively developed and which is often used in optimization problems. Here we present the simplest situation.

The problem we consider is expressed in terms of two functions f and g. The function f corresponds to a constraint, while the function g is related to the extremum we look for. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function and consider its k-level set, namely

$$L_k := \big\{X \in \mathbb{R}^n \mid f(X) = k\big\}.$$

In this framework, the following result can be proved:

Exercise 3.35. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function and let $X \in \mathbb{R}^n$ such that $[\nabla f](X) \neq \mathbf{0}$. Let $k \in \mathbb{R}$ be given by $k := f(X)$, and suppose that the k-level set can be considered (at least locally around X) as a surface of dimension $n-1$ in $\mathbb{R}^n$. Show that $[\nabla f](X)$ is perpendicular to the surface L_k. For that purpose, we can consider any differentiable parametric curve $\varphi : (-1,1) \to L_k$ with $\varphi(0) = X$ and show that $[\nabla f](X)$ is perpendicular to it at the point X. More precisely, show that $[\nabla f](x) \cdot \varphi'(0) = 0$.*

From now on, we consider the 0-level set of f, but of course the argument is similar for any k-level set. Let us consider a differentiable function $g : \mathbb{R}^n \to \mathbb{R}$. Our aim is to look for a point $X_{\max} \in L_0$ such that $g(X_{\max}) \geq g(X)$ for any $X \in L_0$. Clearly, looking for a local maximum of g in $\mathbb{R}^n$ is not relevant in this context, because $X_{\max}$ is usually not a critical point of g without the constraint provided by f. For this reason, it is natural to call this problem *an extremum under a constraint.*

Suppose that L_0 is not reduced to a point (in which case the problem is trivial), and let $\varphi : (-1,1) \to L_0$ be a differentiable parametric curve satisfying $\varphi(0) = X_{\max}$. By looking at the function

$$(-1,1) \ni t \mapsto g\big(\varphi(t)\big) \in \mathbb{R}$$

one observes that this function has local maximum at $t = 0$. Since this function is differentiable, one gets

$$\frac{\mathrm{d}}{\mathrm{d}t} g\big(\varphi(t)\big)\Big|_{t=0} = 0 \Longleftrightarrow [\nabla g](X_{\max}) \cdot \varphi'(0) = 0$$

which means that $[\nabla g](X_{\max})$ is perpendicular to any vector tangent to the surface L_0 at $X_{\max}$. But this condition is the same as the one already observed for ∇f in the above exercise. By putting the two information together one gets:

Theorem 3.36 (Lagrange multiplier)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable and let*

$$S_0 := \big\{X \in \mathbb{R}^n \mid f(X) = 0 \text{ and } [\nabla f](X) \neq \mathbf{0}\big\} \subset L_0.$$

Let $g : \mathbb{R}^n \to \mathbb{R}$ be differentiable, and assume that there exists $X_{\max} \in S_0$ such that $g(X_{\max}) \geq g(X)$ for any $X \in L_0$. Then there exists $\lambda \in \mathbb{R}$ such that

$$[\nabla g](X_{\max}) = \lambda\, [\nabla f](X_{\max}). \tag{3.16}$$

The constant λ is called *the Lagrange multiplier*. Our interest in this theorem is that the condition (3.16) is often easy to check, and that it provides the candidates for the points $X_{\max}$ we are looking for.

Let us look at one example: We consider the following problem. Consider the function $g : \mathbb{R}^2 \to \mathbb{R}$ given by $g(x, y) = x + y$. What is the maximum value of g on the unit circle ? This constraint can be encoded with the following function $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x, y) = x^2 + y^2 - 1$. Indeed, the unit circle corresponds to the 0-level set of the function f. Then, the problem consists in looking for $(x, y) \in \mathbb{R}^2$ satisfying

$$[\nabla g](x, y) = \lambda\,[\nabla f](x, y) \Longleftrightarrow \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \lambda \begin{pmatrix} 2x \\ 2y \end{pmatrix} \Longleftrightarrow \begin{cases} \lambda \neq 0 \\ x = y = \frac{1}{2\lambda} \end{cases} . \tag{3.17}$$

Note that the value of λ is not important, but we infer from the equality $x = y$ that $0 = f(x, x) = 2x^2 - 1$, which means that $x = \pm\frac{\sqrt{2}}{2}$. Since we look for a maximum of g, the only solution is $X_{\max} = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$.

Chapter 4

Functions from $\mathbb{R}^n$ to $\mathbb{R}^d$

4.1 Interlude

So far, we have considered the norm in $\mathbb{R}^n$ provided by the formula

$$\|X\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{\sum_{j=1}^n x_j^2} =: \|X\|_2.$$

We can also consider other expressions, for example:

$$\|X\|_1 := |x_1| + |x_2| + \cdots + |x_n| = \sum_{j=1}^n |x_j|,$$

or

$$\|X\|_\infty := \max_{j\in\{1,\dots,n} |x_j|.$$

It is easy to check that these two alternative expressions also define norms. Namely, they satisfy the three conditions for any norm $|||\cdot|||$:

(i) $|||X||| \geq 0$ for any $X \in \mathbb{R}^n$, and $|||X||| = 0$ if and only if $X = \mathbf{0}$,

(ii) $|||\lambda X||| = |\lambda|\, |||X|||$, for any $\lambda \in \mathbb{R}$ and $X \in \mathbb{R}^n$,

(iii) $|||X + Y||| \leq |||X||| + |||Y|||$, for any $X, Y \in \mathbb{R}^n$.

The relations between these three norms is provided in the following statement, which can be proved as an exercise.

Lemma 4.1. *For any $X \in \mathbb{R}^n$ one has*

$$\|X\|_2 \leq \sqrt{n}\|X\|_\infty \leq \sqrt{n}\|X\|_1 \leq n\|X\|_2.$$

Whenever two norms $\|\cdot\|$ and $|||\cdot|||$ satisfy $c\|X\| \leq |||X||| \leq d\|X\|$ for some constants $c, d > 0$ and all $X \in \mathbb{R}^n$ we say that these two norms are equivalent. Note that the inequalities $\frac{1}{d}|||X||| \leq \|X\| \leq \frac{1}{c}|||X|||$ also hold. As a consequence of the previous lemma, it follows that the three norms $\|\cdot\|_2$, $\|\cdot\|_1$, and $\|\cdot\|_\infty$ are equivalent. In fact, this observation is a special instance of a deeper result.

Theorem 4.2. *All norms on $\mathbb{R}^n$ are equivalent.*

A proof of this statement can be found in many textbooks or over internet. Note that in the sequel we shall continue using the notation $\|\cdot\|$ for the usual Euclidean norm $\|\cdot\|_2$.

4.2 Limits and continuity

From now on we consider $\Omega \subset \mathbb{R}^n$ an open subset, and a function $f : \Omega \to \mathbb{R}^d$. Note that n and d can be different, or the same integers. Ω is sometimes called the domain of f, and is denoted by $\mathrm{Dom}(f)$. The set $f(\Omega) \subset \mathbb{R}^d$ is again called *the range* or *the image* of f. Since for any $X \in \Omega$ one has $f(X) = \begin{pmatrix} f(X)_1 \\ f(X)_2 \\ \vdots \\ f(X)_d \end{pmatrix} \in \mathbb{R}^d$ we set $f_j(X) := f(X)_j \in \mathbb{R}$, and thus get a collection of d functions $\{f_j\}_{j=1}^d$ with $f_j : \Omega \to \mathbb{R}$ and satisfying $f = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_d \end{pmatrix}$. The d functions f_j are called *the components* of f, and, as functions from Ω to $\mathbb{R}$, they have already been studied in the previous chapter. This means that several notions can be mimicked from Chapter 3.

Definition 4.3 (Limit)**.** *Let $\Omega \subset \mathbb{R}^n$ be open with $X_0 \in \Omega$, and let $f : \Omega \setminus \{X_0\} \to \mathbb{R}^d$. The function f* has a limit at X_0 *if there exists $f(X_0) \in \mathbb{R}^d$ such that for any $\varepsilon > 0$ there exists a $\delta > 0$ with*

$$\|f(X) - f(X_0)\| \leq \varepsilon \qquad \textit{for all } X \in \mathcal{B}_\delta(X_0).$$

If f has a limit at X_0 we write $\lim_{X \to X_0} f(X) = f(X_0)$.

We can now easily relate this notion with a similar property for each component of f.

Lemma 4.4. *In the setting of the previous definition, f has a limit at X_0 if and only if each component $f_j : \Omega \setminus \{X_0\} \to \mathbb{R}$ has a limit at X_0.*

The proof is left as an exercise, but let us mention that the following two inequalities (based on the equivalence of the norms seen in the previous section) can be used:

$$|f_j(X) - f_j(X_0)| \leq \|f(X) - f(X_0)\|_\infty \leq \sqrt{d}\|f(X) - f(X_0)\|,$$

and

$$\|f(X) - f(X_0)\| \leq \sqrt{d}\|f(X) - f(X_0)\|_\infty.$$

Based on the previous definition, the notion of continuity is now quite clear.

Definition 4.5 (Continuity)**.** *Let $\Omega \subset \mathbb{R}^n$ be open with $X_0 \in \Omega$, and let $f : \Omega \to \mathbb{R}^d$. The function f* is continuous at X_0 *if $\lim_{X \to X_0} f(X) = f(X_0)$, or more precisely if*

$$\forall \varepsilon > 0,\ \exists \delta > 0 \ \text{ such that } \ \|f(X) - f(X_0)\| \leq \varepsilon \ \text{ for all } X \in \mathcal{B}_\delta(X_0).$$

By using Lemma 4.4, one easily infers the following result:

Lemma 4.6. *Let $\Omega \subset \mathbb{R}^n$ be open with $X_0 \in \Omega$, and let $f : \Omega \to \mathbb{R}^d$. The function f* is continuous at X_0 *if and only if each component function $f_j : \Omega \to \mathbb{R}$ is continuous at X_0.*

4.3 Differentiability

For an open set $\Omega \subset \mathbb{R}^n$, let us recall that a function $f : \Omega \to \mathbb{R}$ is differentiable at $X \in \Omega$ if

$$\frac{f(X + H) - f(X) - \nabla f(X) \cdot H}{\|X\|} = g(H).$$

with $g(H) \to 0$ as $H \to \mathbf{0}$. Note that this condition can be rewritten as

$$f(X + H) = f(X) + {}^t\big(\nabla f(X)\big)H + \|H\|g(H)$$

with $g(H) \to 0$ as $H \to \mathbf{0}$. The term ${}^t\big(\nabla f(X)\big)$ can be seen as a matrix $M_{1n}(\mathbb{R})$, which means that the expression ${}^t\big(\nabla f(X)\big)H$ can be understood as a *linear map* applied to H. The same idea will be used to define the notion of differentiability for a function from Ω to $\mathbb{R}^d$.

Definition 4.7 (Differentiability and Jacobian matrix)**.** *Let $\Omega \subset \mathbb{R}^n$ be open, and let $f : \Omega \to \mathbb{R}^d$. The function f is* differentiable at $X \in \Omega$ *if there exists a matrix $\mathcal{J}_f(X) \in M_{dn}(\mathbb{R})$ such that*

$$f(X + H) = f(X) + \mathcal{J}_f(X)H + \|H\|g(H) \tag{4.1}$$

for some $g : \mathbb{R}^n \to \mathbb{R}^d$ with $g(H) \to \mathbf{0}$ as $H \to \mathbf{0}$. The matrix $\mathcal{J}_f(X)$ is called the Jacobian matrix *of f at X.*

Note that the notation $\mathcal{D}_f(X)$ is also often used for the Jacobian matrix. We also emphasize that the $\mathbf{0}$ in the expressions $g(H) \to \mathbf{0}$ and $H \to \mathbf{0}$ are not the same: the first one is $\mathbf{0} \in \mathbb{R}^d$ while the second one is $\mathbf{0} \in \mathbb{R}^n$. Since this ambiguity should not lead to any confusion, we shall keep the same notation for these two $\mathbf{0}$.

Remark 4.8. *According to the definition of a derivative in Calculus I and to the definition of differentiability of the previous chapter, the following equalities hold:*

For $n = d = 1:\ \mathcal{J}_f(x) = f'(x),$ *with* $x \in \mathbb{R}$,
For $d = 1:\ \mathcal{J}_f(X) = {}^t\big(\nabla f(X)\big) = \big(\partial_1 f(X)\ \partial_2 f(X)\ \dots\ \partial_n f(X)\big),$ *with* $X \in \mathbb{R}^n$.

In general situation, one has:

Lemma 4.9. *Let $\Omega \subset \mathbb{R}^n$ be open, and let $f : \Omega \to \mathbb{R}^d$ be differentiable at $X \in \Omega$. Then the components of f admit partial derivatives, and for any $j \in \{1, \dots, n\}$ the j^{th}-column of $\mathcal{J}_f(X)$ is given by* $\begin{pmatrix} \partial_j f_1(X) \\ \partial_j f_2(X) \\ \vdots \\ \partial_j f_d(X) \end{pmatrix}$.

Proof. Consider $H = hE_j$ with E_j be one vector of the canonical basis of $\mathbb{R}^n$ and $h \in \mathbb{R}$. Then, (4.1) reduces to

$$\begin{aligned} f(X + hE_j) &= f(X) + h\mathcal{J}_f(X)E_j + |h|g(hE_j) \\ &= f(X) + h\big(\mathcal{J}_f(X)\big)^j + |h|g(hE_j) \end{aligned}$$

with $\big(\mathcal{J}_f(X)\big)^j$ be the j^{th}-column of $\mathcal{J}_f(X)$. By rewriting this equality in a different form, one gets

$$\frac{f(X + hE_j) - f(X)}{h} = \big(\mathcal{J}_f(X)\big)^j + \operatorname{sgn}(h)g(hE_j)$$

with $\operatorname{sgn}(h)$ be the sign of h. Let us emphasize that the above equality in fact corresponds to d equalities. It can be rewritten for any $k \in \{1, \dots, d\}$ as

$$\frac{f_k(X + hE_j) - f_k(X)}{h} = \big(\mathcal{J}_f(X)\big)_{kj} + \operatorname{sgn}(h)g_k(hE_j)$$

with $\big(\mathcal{J}_f(X)\big)_{kj}$ be the entry at row k and column j of the matrix $\mathcal{J}_f(X)$. Since $g_k(hE_j) \to 0$ as $h \to 0$, one infers that each component function f_k admit n partial derivatives, and that $\partial_j f_k(X) = \big(\mathcal{J}_f(X)\big)_{kj}$. By collecting these expressions in a vector, one obtains the statement of the lemma. $\square$

If we summarize our findings, one gets that if $f : \Omega \to \mathbb{R}^d$ is differentiable at X, then all partial derivatives of the components f_k of f exist at X, and the Jacobian matrix is given by

$$\mathcal{J}_f(X) = \begin{pmatrix} \partial_1 f_1(X) & \partial_2 f_1(X) & \dots & \partial_n f_1(X) \\ \partial_1 f_2(X) & \partial_2 f_2(X) & \dots & \partial_n f_2(X) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_1 f_d(X) & \partial_2 f_d(X) & \dots & \partial_n f_d(X) \end{pmatrix} \tag{4.2}$$

which is a $d \times n$ matrix with entries in $\mathbb{R}$. Let us emphasize that in general this matrix is not a square matrix. In the special case $n = d$ this matrix becomes a square matrix, but it is not a symmetric matrix in general. Let us also mention that this matrix is also denoted by $f'(X)$, since it represents the derivative of f at X. In other words, the derivative of a function from $\mathbb{R}^n$ to $\mathbb{R}^d$, once evaluated at one point X, is an element of $M_{dn}(\mathbb{R})$.

Let us introduce one more object which will be very useful later on, namely the determinant of the Jacobian matrix, whenever it is defined.

Definition 4.10. *If $n = d$ and whenever the Jacobian matrix $\mathcal{J}_f(X) \in M_n(\mathbb{R})$ is defined at X, we write $|\mathcal{J}_f(X)|$ or $\det\big(\mathcal{J}_f(X)\big)$ for the determinant of this matrix.*

Let us illustrate the above construction with one example. In the sequel, we shall often use the notation $\mathbb{T}$ for $\mathbb{R}/2\pi\mathbb{Z}$. By this, we mean the set of numbers modulo 2π. The advantage of this notation, compared to $[0, 2\pi)$, is that the sum of any elements θ_1 and θ_2 of $\mathbb{T}$ still belongs to $\mathbb{T}$, while if we take $\theta_1, \theta_2 \in [0, 2\pi)$, their sum is not always in $[0, 2\pi)$ if we don't specify that we perform the addition *modulo* 2π. It is also useful to observe that $\mathbb{T}$ is an open set. ☺

Example 4.11 (Polar coordinates)**.** *Consider*

$$\phi : (0, \infty) \times \mathbb{T} \ni (r, \theta) \mapsto \phi(r, \theta) = \big(r\cos(\theta), r\sin(\theta)\big) \in \mathbb{R}^2.$$

We can observe that the image of this function is not equal to $\mathbb{R}^2$ but to $\mathbb{R}^2 \setminus \{\mathbf{0}\}$, which is an open set. For the function ϕ its component are $\phi_1(r, \theta) = r\cos(\theta)$ and $\phi_2(r, \theta) = r\sin(\theta)$. If one computes the Jacobian matrix, one finds

$$\mathcal{J}_\phi(r, \theta) = \begin{pmatrix} \frac{\partial \phi_1}{\partial r}(r, \theta) & \frac{\partial \phi_1}{\partial \theta}(r, \theta) \\ \frac{\partial \phi_2}{\partial r}(r, \theta) & \frac{\partial \phi_1}{\partial \theta}(r, \theta) \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{pmatrix},$$

and for the Jabodian determinant:

$$|\mathcal{J}_\phi(r, \theta)| = r\big(\cos^2(\theta) + \sin^2(\theta)\big) = r. \tag{4.3}$$

Let us finally state a useful and quite natural result. Its slightly technical proof is provided in [1, Thm. 2, p. 212].

Theorem 4.12. *Let $\Omega \subset \mathbb{R}^n$ be open, and let $f : \Omega \to \mathbb{R}^d$, with component functions $f_j : \Omega \to \mathbb{R}$ for $j \in \{1, \dots, d\}$. If all f_j are differentiable at $X \in \Omega$, then f is differentiable at X, and the equality* (4.2) *holds. In particular, if each f_j is of class C^1, then f is differentiable and* (4.2) *holds.*

4.4 Chain rule

In Lemma 3.25, when discussing reparametrization of parametric curves, we already introduced the concept of chain rule in a more general setting compared to Calculus I. Let us now introduce it in the general framework of functions from $\mathbb{R}^n$ to $\mathbb{R}^d$.

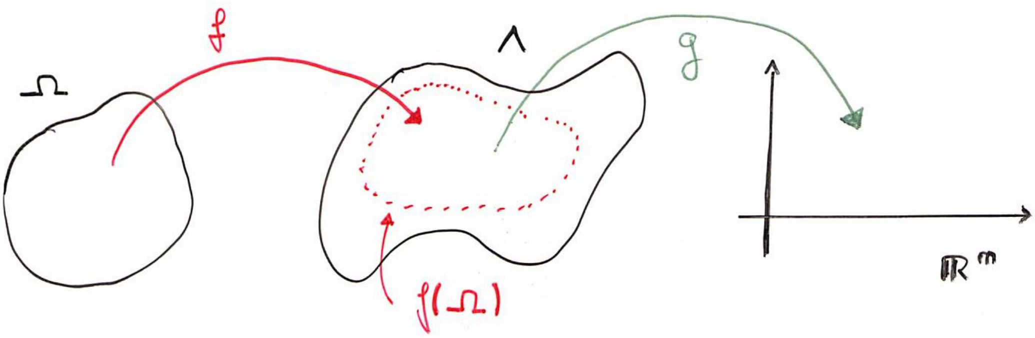


Fig. 4.1. Composition of two functions

Let $\Omega \subset \mathbb{R}^n$ be open, and consider $f : \Omega \to \mathbb{R}^d$ be differentiable. Let also $\Lambda \subset \mathbb{R}^d$ be open with $f(\Omega) \subset \Lambda$, and consider $g : \Lambda \to \mathbb{R}^m$ be differentiable. In this framework, what can be said about the composed function $g \circ f : \Omega \to \mathbb{R}^m$ (see Figure 4.1)? The following statement answers this question. Its proof is similar to the one of Lemma 3.25 but slightly more complicated because of the number of variables involved. We refer to [1, Thm.3 p. 216] for the proof.

Theorem 4.13. *In the above setting the composed map $g \circ f : \Omega \to \mathbb{R}^m$ is differentiable, with derivative at $X \in \Omega$*

$$\mathcal{J}_{g\circ f}(X) = \mathcal{J}_g\big(f(X)\big)\, \mathcal{J}_f(X). \tag{4.4}$$

Let us emphasize that the r.h.s. of (4.4) corresponds to a product of matrices. Indeed, $\mathcal{J}_f(X)$ belongs to $M_{dn}(\mathbb{R})$, while $\mathcal{J}_g(Y)$ belongs to $M_{md}(\mathbb{R})$ for any $Y \in \Lambda$. As a result, the product of these matrices belongs to $M_{mn}(\mathbb{R})$ as it should be, since $\mathcal{J}_{g\circ f}(X)$ is also a $M_{mn}(\mathbb{R})$ matrix.

Clearly, the above statement contains the simple chain rule of Calculus I, but also the one provided for reparametrization of parametric curves. To observe this, it is sufficient to take Remark 4.8 and the following equalities into account:

$$\nabla f\big(\varphi(t)\big) \cdot \varphi'(t) = {}^t\Big(\nabla f\big(\varphi(t)\big)\Big)\varphi'(t) = \mathcal{J}_f\big(\varphi(t)\big)\mathcal{J}_\varphi(t).$$

Let us finally look at one example, and compute the derivative by two methods.

Example 4.14. *Consider again*

$$\phi : (0,\infty) \times \mathbb{T} \ni (r,\theta) \mapsto \phi(r,\theta) = \big(r\cos(\theta), r\sin(\theta)\big) \in \mathbb{R}^2$$

and let $g : \mathbb{R}^2 \to \mathbb{R}$ given by $g(x,y) = xy^2$. If one computes separately the two Jacobian matrices, one has (see also Example 4.11)

$$\mathcal{J}_\phi(r,\theta) = \begin{pmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{pmatrix} \qquad \text{and} \qquad \mathcal{J}_g(x,y) = \begin{pmatrix} y^2 & 2xy \end{pmatrix},$$

which leads to $\mathcal{J}_g\big(\phi(r,\theta)\big) = \big(r^2\sin^2(\theta) \;\; 2r^2\cos(\theta)\sin(\theta)\big)$. Thus by the chain rule one gets:

$$\begin{aligned} \mathcal{J}_{g\circ\phi}(r,\theta) &= \big(r^2\sin^2(\theta) \quad 2r^2\cos(\theta)\sin(\theta)\big)\begin{pmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{pmatrix} \\ &= \Big(r^2\sin^2(\theta)\cos(\theta) + 2r^2\cos(\theta)\sin^2(\theta) \quad -r^3\sin^3(\theta) + 2r^3\cos^2(\theta)\sin(\theta)\Big) \\ &= \Big(3r^2\sin^2(\theta)\cos(\theta) \quad r^3\big(2\cos^2(\theta)\sin(\theta) - \sin^3(\theta)\big)\Big). \end{aligned}$$

Note that the same result can be obtained by a direct computation. Indeed, one has $g\big(\phi(r,\theta)\big) = r^3\cos(\theta)\sin^2(\theta)$, and the Jacobian matrix for the function $(r,\theta) \mapsto r^3\cos(\theta)\sin^2(\theta)$ is

$$\begin{aligned} &\left(\frac{\mathrm{d}\big(r^3\cos(\theta)\sin^2(\theta)\big)}{\mathrm{d}r} \quad \frac{\mathrm{d}\big(r^3\cos(\theta)\sin^2(\theta)\big)}{\mathrm{d}\theta}\right) \\ &= \Big(3r^2\sin^2(\theta)\cos(\theta) \quad r^3\big(2\cos^2(\theta)\sin(\theta) - \sin^3(\theta)\big)\Big), \end{aligned}$$

as before.

4.5 Invertibility

Let us consider an open interval $I \subset \mathbb{R}$. We recall that if $f : I \to \mathbb{R}$ is strictly increasing on I, then it is invertible on its image $J := f\big((a,b)\big)$, which means that there exists $f^{-1} : J \to I$ with $f^{-1} \circ f(x) = x$ for any $x \in I$ and $f \circ f^{-1}(y) = y$ for any $y \in J$. The same observation holds if f is strictly decreasing. In addition, if f is differentiable, note that the condition of strictly increasing or strictly decreasing corresponds to a derivative never equals to 0.

Now what is the analogue of this situation for functions of n variables taking values in $\mathbb{R}^d$? Clearly, the notion of strictly increasing or strictly decreasing does not mean anything in this general framework.

Definition 4.15 (Invertibility). *Let $\Omega \subset \mathbb{R}^n$ be open, and let $f : \Omega \to \mathbb{R}^n$. We set $\Lambda \subset \mathbb{R}^n$ for the image of Ω by f, namely $\Lambda := f(\Omega)$. The function f is* invertible *on Λ if Λ is open and if there exists $f^{-1} : \Lambda \to \Omega$ with $f^{-1} \circ f(X) = X$ for any $X \in \Omega$ and $f \circ f^{-1}(Y) = Y$ for any $Y \in \Lambda$.*

Let us stress that we have considered a function from $\Omega \subset \mathbb{R}^n$ to $\mathbb{R}^n$, and not to $\mathbb{R}^d$ with arbitrary d. This more general situation could also be considered, and is useful sometimes, but we shall stick to the above framework. As an example with different n and d, we can define $f : \mathbb{R}^2 \to \mathbb{R}^3$ with $f(x, y) = (x, y, 0)$ for any $(x, y) \in \mathbb{R}^2$. Note however that the image of f is not open in $\mathbb{R}^3$, which means that the condition about openness should be dropped.

Examples 4.16. *(i) Let $D \in \mathbb{R}^n$ be a fixed vector, and consider $f : \mathbb{R}^n \to \mathbb{R}^n$ the translation by D defined by $f(X) = X + D$ for any $X \in \mathbb{R}^n$. Clearly, this function is invertible with $f^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ given by $f^{-1}(Y) = Y - D$.*

(ii) Let $C, D \in \mathbb{R}^n$ be fixed vectors, and let $\mathcal{A} \in M_n(\mathbb{R})$ be a given $n \times n$ invertible matrix. Then the map $f : \mathbb{R}^n \to \mathbb{R}^n$ given by $f(X) = \mathcal{A}(X - C) + D$ is an invertible map, with its inverse $f^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ given by $f^{-1}(Y) = \mathcal{A}^{-1}(Y - D) + C$.

(iii) Consider f defined by

$$f : (0, \infty) \times (0, \pi/2) \ni (r, \theta) \to \big(r\cos(\theta), r\sin(\theta)\big) \in \mathbb{R}^2.$$

Clearly, the range of f is $\Lambda := \{(x, y) \in \mathbb{R}^2 \mid x > 0, y > 0\}$, and for any $(x, y) \in \Lambda$ one can define $f^{-1}(x, y) = \Big(\sqrt{x^2 + y^2}, \arccos\big(x/\sqrt{x^2 + y^2}\big)\Big)$, and check that f^{-1} is indeed an inverse for f on Λ.

So far so good, but let us consider the function $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x, y) = \big(\sin(x), \sin(y)\big)$. The range of f is the subset $[-1, 1] \times [-1, 1]$ of $\mathbb{R}^2$ and this function is clearly not invertible on this set. However, if we restrict f to a small open set Ω around $\mathbf{0}$, then this restriction might define a new function which is invertible on its image. This means that the notion of invertibility introduced so far is too global, a local version is necessary. The concept of *local* is going to be more and more important in the sequel.

Definition 4.17 (Local invertibility)**.** *Let $\Omega \subset \mathbb{R}^n$ be open, let $f : \Omega \to \mathbb{R}^n$, and let $X_0 \in \Omega$. The function f is* invertible around $f(X_0) \in f(\Omega)$ *if there exists $\varepsilon > 0$ such that f is invertible on $f\big(\mathcal{B}_\varepsilon(X_0)\big)$,* i.e. *$f\big(\mathcal{B}_\varepsilon(X_0)\big)$ is an open set in $\mathbb{R}^n$ and there exists $f^{-1} : f\big(\mathcal{B}_\varepsilon(X_0)\big) \to \mathcal{B}_\varepsilon(X_0)$ such that $f^{-1} \circ f(X) = X$ for all $X \in \mathcal{B}_\varepsilon(X_0)$ and $f \circ f^{-1}(Y) = Y$ for any $Y \in f\big(\mathcal{B}_\varepsilon(X_0)\big)$.*

We can now state an important technical result, and sketch its proof.

Theorem 4.18. *Let $\Omega \subset \mathbb{R}^n$ be open, let $f : \Omega \to \mathbb{R}^n$ be of class C^1, and let $X_0 \in \Omega$. Then the function f is invertible around $f(X_0) \in f(\Omega)$ if its Jacobian matrix $\mathcal{J}_f(X_0)$ at X_0 is invertible.*

Note that since $f : \Omega \to \mathbb{R}^n$, its Jacobian matrix $\mathcal{J}_f(X_0)$ is a square matrix, and therefore the notion of invertibility is well defined. In addition, this matrix is invertible if and only if its determinant is not 0.

Sketch of the proof. By the differentiability of f there exists a function $g : \mathbb{R}^n \to \mathbb{R}^n$ with $g(H) \to \mathbf{0}$ as $H \to \mathbf{0}$ such that

$$f(X) - f(X_0) - \mathcal{J}_f(X_0)(X - X_0) = \|X - X_0\|\, g(X - X_0).$$

If we set $\mathcal{A} := \mathcal{J}_f(X_0)$ this can be rewritten as

$$f(X) = \mathcal{A}(X - X_0) + f(X_0) + o(X - X_0) \tag{4.5}$$

with $o(X - X_0)$ denoting a function from $\mathbb{R}^n$ to $\mathbb{R}^n$ satisfying $\frac{1}{\|X-X_0\|} o(X - X_0) \to \mathbf{0}$ as $X \to X_0$. Thus, modulo this small remainder term, the equality (4.5) is similar to the example (ii) in Examples 4.16, with an invertible matrix $\mathcal{A}$, and two fixed vectors X_0 and $f(X_0)$. Thus, we can use the fact that the function given by $f(X) = \mathcal{A}(X - X_0) + f(X_0)$ is invertible. The precise proof consists in showing that the remainder term $o(X - X_0)$ does not destroy this picture, if X is close enough to X_0. We leave this for a very motivated reader. $\square$

Remark that in the special case $n = 1$, the condition of invertibility of $\mathcal{J}_f(X_0)$ corresponds to the condition $\mathcal{J}_f(X_0) \equiv f'(X_0) \neq 0$. Thus, this condition means that f is either strictly increasing or is strictly decreasing in a neighbourhood of X_0.

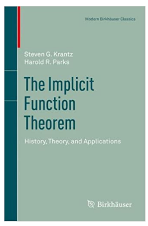
Let us make another important observation. Suppose that f is of class C^1 and invertible around $f(X_0)$. Let $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$ denote the identity function, namely $\mathrm{id}(X) = X$. Then, from the idendity $f^{-1} \circ f(X) = X = \mathrm{id}(X)$, valid for any X close to X_0, one can take the derivative on both sides of the equality, use the chain rule, and obtain

$$\mathcal{J}_{f^{-1}}\big(f(X_0)\big)\, \mathcal{J}_f(X_0) = 1_{nn},$$

with 1_{nn} the $n \times n$ identity matrix. Thus, by multiplying by $\big(\mathcal{J}_f(X_0)\big)^{-1}$, which exists by our assumption, and by setting $Y := f(X_0)$ one obtains

$$\mathcal{J}_{f^{-1}}(Y) = \Big(\mathcal{J}_f\big(f^{-1}(Y)\big)\Big)^{-1}.$$

In Calculus I, this equality corresponds to the derivative of the inverse function. Now, the derivative of the inverse of a function of several variables evaluated at Y also exists, but it is a matrix and this matrix is given by the inverse of the Jacobian matrix of f at $f^{-1}(Y)$. Be aware that the operation here is the inverse of a matrix, it is not 1 over a matrix, which is something that does not exist.



4.6 Implicit function theorem

Let us first consider a function $f : \mathbb{R}^2 \to \mathbb{R}$ and consider the set

$$L_0 := \{(x, y) \in \mathbb{R}^2 \mid f(x, y) = 0\}.$$

Clearly, this set corresponds to the 0-level set, as introduced in Section 3.2. In general, this set describes a curve in $\mathbb{R}^2$, but degenerated cases can also exist, for which this set is empty, or it includes some larger sets in $\mathbb{R}^2$. A few examples are here:

Examples 4.19. *We consider examples of functions $f : \mathbb{R}^2 \to \mathbb{R}$.*

(i) *If $f(x, y) = \sin(x) \cos(y)$, then L_0 is represented by a square lattice in $\mathbb{R}^2$,*

(ii) *If $f(x, y) = x^2 + y^2$, then L_0 corresponds to just one point, the origin $\mathbf{0}$,*

(iii) *If $f(x, y) = \mathrm{e}^{xy}$, then L_0 is the empty set,*

(iv) *If $f(x, y) = |\lfloor x \rfloor| + |\lfloor y \rfloor|$, then L_0 corresponds to the set $[0, 1) \times [0, 1)$ in $\mathbb{R}^2$. Here, $\lfloor x \rfloor$ denotes the floor function and gives the greatest integer less than or equal to x,*

(v) *If $f(x, y) = x^2 + y^2 - 1$, then L_0 corresponds to the unit circle in $\mathbb{R}^2$.*

Suppose now that L_0 is a "nice" curve, is it possible to express y as a function of x, at least locally ? In other words, can one solve the equation $f(x, y) = 0$ and get $y = \phi(x)$? As clearly visible on Figure 4.2 the answer is yes, except at a few points (represented in black on this figure). Note that a similar question was discussed in Calculus I, and is related to implicit differention. However, what is the analogue question for more than 2 variables ? If we have the equation $f(x_1, x_2, \dots, x_n) = 0$, can one express x_n in terms of the other variables ? Or not only x_n, but also x_{n-1}, x_{n-2}, and so on ? All the questions are related to the Implicit function theorem which is presented below. It is a rather deep result, which proof is based on the invertibility results provided in the previous section.

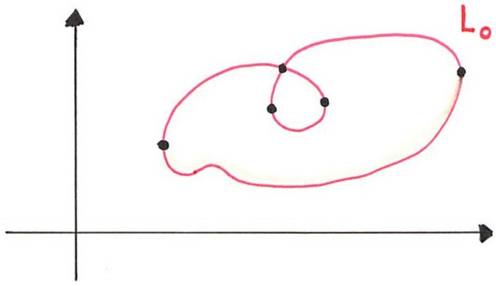


Fig. 4.2. A nice curve

The framework is the following: We consider Ω an open subset of $\mathbb{R}^n$ and let $f : \Omega \to \mathbb{R}^d$ with $d < n$. We emphasize that such a function corresponds to d functions $f_j : \Omega \to \mathbb{R}$. Each of these d functions can be considered as an equation, or as a relation between the n variables. The system we shall consider is $f(x_1, x_2, \dots, x_n) = \mathbf{0}$, or equivalently $f_j(X) = 0$ for $j \in \{1, \dots, d\}$ and for $X = (x_1, x_2, \dots, x_n)$. With these d equations our aim is to express d variables in terms of the $n - d$ remaining variables. By convention, we choose to express the last d variables in terms of the other ones, but this is purely by convention. Note also that the choice of putting $f(X) = \mathbf{0}$, and not another vector on the right hand side, is also a matter of convention. Any other choice can be recast in the above framework by changing the function f.

We state now the main result of this section. It is a complicated statement in the general case, but we provide after it one simpler version which can be more easily understood.

Theorem 4.20 (Implicit function theorem)**.** *Let $\Omega \subset \mathbb{R}^n$ be open, let $d < n$ and consider $f : \Omega \to \mathbb{R}^d$ of class C^k for some $k \geq 1$. Let $X \equiv (x_1, \dots, x_n) \in \Omega$ verifying*

$f(X) = \mathbf{0}$, and assume that

$$\underline{\mathcal{J}}_f(X) := \begin{pmatrix} \partial_{n-d+1} f_1 & \partial_{n-d+2} f_1 & \dots & \partial_n f_1 \\ \partial_{n-d+1} f_2 & \partial_{n-d+2} f_2 & \dots & \partial_n f_2 \\ \vdots & \vdots & \ddots & \vdots \\ \partial_{n-d+1} f_d & \partial_{n-d+2} f_2 & \dots & \partial_n f_d \end{pmatrix} (X)$$

is an invertible $d \times d$ matrix. Then there exists $\varepsilon > 0$ and $\phi : \mathcal{B}_\varepsilon\big((x_1, \dots, x_{n-d})\big) \to \mathbb{R}^d$ of class C^k, with $\phi(x_1, \dots, x_{n-d}) = (x_{n-d+1}, \dots, x_n)$, such that

$$f\big(y_1, \dots, y_{n-d}, \phi_1(y_1, \dots, y_{n-d}), \dots, \phi_d(y_1, \dots, y_{n-d})\big) = \mathbf{0}$$

for all $(y_1, \dots, y_{n-d}) \in \mathcal{B}_\varepsilon\big((x_1, \dots, x_{n-d})\big)$. In addition, ϕ is the unique function satisfying

$$\begin{aligned} &\big\{Y \in \mathcal{B}_\varepsilon(X) \mid f(Y) = \mathbf{0}\big\} \\ &\iff \big\{Y \in \mathcal{B}_\varepsilon(X) \mid y_{n-d+\ell} = \phi_\ell(y_1, \dots, y_{n-d}) \text{ for } \ell \in \{1, \dots, d\}\big\}. \end{aligned}$$

The function ϕ described in this theorem is called *the implicit function*. As easily seen by the above statement, this theorem is not easy to understand in this generality. Let us then restate it in a simpler case, namely when $d = 1$ and $n \geq 2$. This means that we are going to deal with only 1 equation, and try to express the last variable in terms of the $n - 1$ other variables.

Theorem 4.21 (Simpler version)**.** *Let $\Omega \subset \mathbb{R}^n$ be open, let $d < n$ and consider $f : \Omega \to \mathbb{R}$ of class C^k for some $k \geq 1$. Let $X \equiv (x_1, \dots, x_n) \in \Omega$ verifying $f(X) = 0$, and assume that $\partial_n f(X) \neq 0$. Then there exists $\varepsilon > 0$ and $\phi : \mathcal{B}_\varepsilon\big((x_1, \dots, x_{n-1})\big) \to \mathbb{R}$ of class C^k with $\phi(x_1, \dots, x_{n-1}) = x_n$, such that*

$$f\big(y_1, \dots, y_{n-1}, \phi(y_1, \dots, y_{n-1})\big) = 0$$

for all $(y_1, \dots, y_{n-1}) \in \mathcal{B}_\varepsilon\big((x_1, \dots, x_{n-1})\big)$. In addition, ϕ is the unique function satisfying

$$\big\{Y \in \mathcal{B}_\varepsilon(X) \mid f(Y) = 0\big\} \iff \big\{Y \in \mathcal{B}_\varepsilon(X) \mid y_n = \phi(y_1, \dots, y_{n-1})\big\}.$$

Let us now look at two interpretations of this theorem. For simplicity, we shall switch from the variables $(x_1, \dots, x_n)$ to the variables (x, y) or (x, y, z). As a first example, we let $f : \mathbb{R}^2 \to 0$ and consider the 0-level set L_0. Suppose that we know one point $X = (x_0, y_0)$ belonging to L_0. Our aim is to describe the points of L_0 which are close to (x_0, y_0). The theorem tells us that if $\partial_y f(x_0, y_0) \neq 0$, then there exists a function ϕ which describes L_0 close to (x_0, y_0), see Figure 4.3. More precisely, there exists $\varepsilon > 0$ and

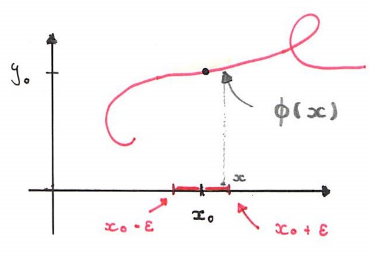


Fig. 4.3. Example in dim. 2

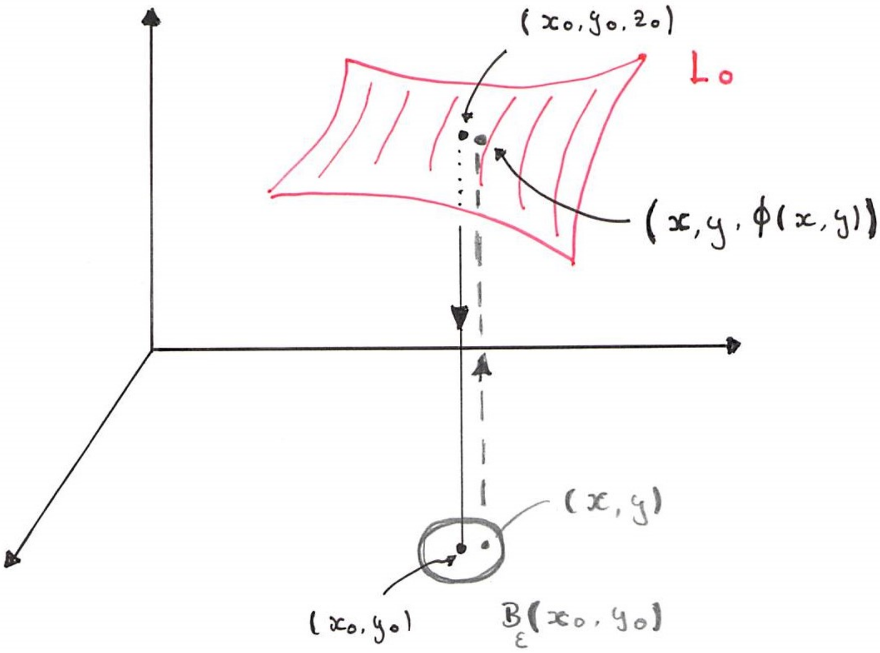
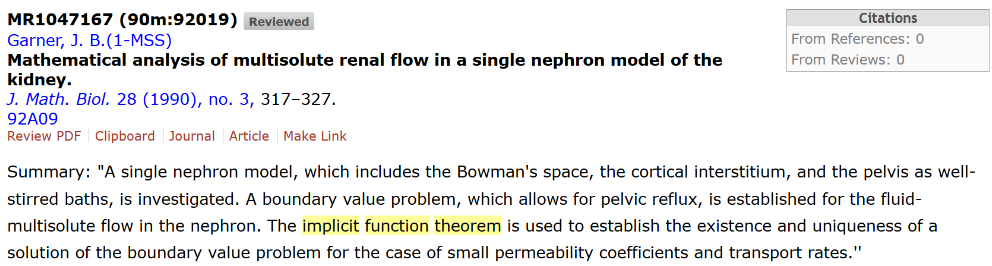


Fig. 4.4. Example in dim. 3

$\phi : (x_0-\varepsilon, x_0+\varepsilon) \to \mathbb{R}$ such that $\phi(x_0) = y_0$ and $f\big(x, \phi(x)\big) = 0$ for all $x \in (x_0-\varepsilon, x_0+\varepsilon)$. The last statement in the theorem says that all the points on L_0 which are close to (x_0, y_0) are described by the pairs $\big(x, \phi(x)\big)$ for all $x \in (x_0 - \varepsilon, x_0 + \varepsilon)$.

For the second example we consider $f : \mathbb{R}^3 \to \mathbb{R}$ and again want to study the surface defined by $f(x, y, z) = 0$. Suppose that (x_0, y_0, z_0) belongs to L_0 and that $\partial_z f(x_0, y_0, z_0) \neq 0$. Then there exists $\varepsilon > 0$ and $\phi : \mathcal{B}_\varepsilon\big((x_0, y_0)\big) \to \mathbb{R}$ such that $\phi(x_0, y_0) = z_0$ and $f\big(x, y, \phi(x, y)\big) = 0$ for all $(x, y) \in \mathcal{B}_\varepsilon\big((x_0, y_0)\big)$. This means that the point $\big(x, y, \phi(x, y)\big)$ belongs to L_0, and the last statement of the theorem implies that all points of L_0 close to (x_0, y_0, z_0) can be described by such triplets.

The Implicit function theorem is of fundamental importance in mathematics, but has also application in other fields, see Figure 4.5.



MR1047167 (90m:92019) Reviewed
Garner, J. B.(1-MSS)
Mathematical analysis of multisolute renal flow in a single nephron model of the kidney.
J. Math. Biol. 28 (1990), no. 3, 317–327.
92A09
Review PDF | Clipboard | Journal | Article | Make Link

Citations
From References: 0
From Reviews: 0

Summary: "A single nephron model, which includes the Bowman's space, the cortical interstitium, and the pelvis as well-stirred baths, is investigated. A boundary value problem, which allows for pelvic reflux, is established for the fluid-multisolute flow in the nephron. The implicit function theorem is used to establish the existence and uniqueness of a solution of the boundary value problem for the case of small permeability coefficients and transport rates."

Fig. 4.5. Application in biology

Chapter 5

Curve integrals

With this chapter we start the study of integrals of functions from $\mathbb{R}^n$ to $\mathbb{R}^d$. These exist several types of integrals, and one should not mix them up.

5.1 Line integrals

The first type of integral corresponds to integral of vector fields along curves, as we describe now.

Definition 5.1 (Vector field)**.** *Let $\Omega \subset \mathbb{R}^n$ be open. A* vector field *is a function $f : \Omega \to \mathbb{R}^n$.*

Compared to a genral function from $\mathbb{R}^n$ to $\mathbb{R}^d$, what characterizes a vector field is that $n = d$. Vector fields can be quite easily represented since at each point of Ω in $\mathbb{R}^n$ one attache a vector in $\mathbb{R}^n$, see Figure 5.1.

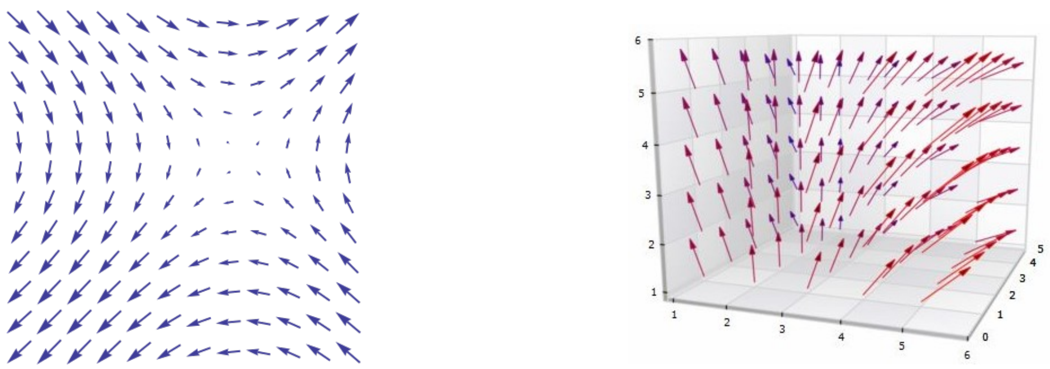


Fig. 5.1. Two vector fields

In addition to a vector field, we shall also need a parametric curve in Ω as introduced in Chapter 2. In the present framework, a curve will be denoted by $c : [a, b] \to \Omega$. So, let us consider a vector field $f : \Omega \to \mathbb{R}^n$ which is continuous, and a parametric curve

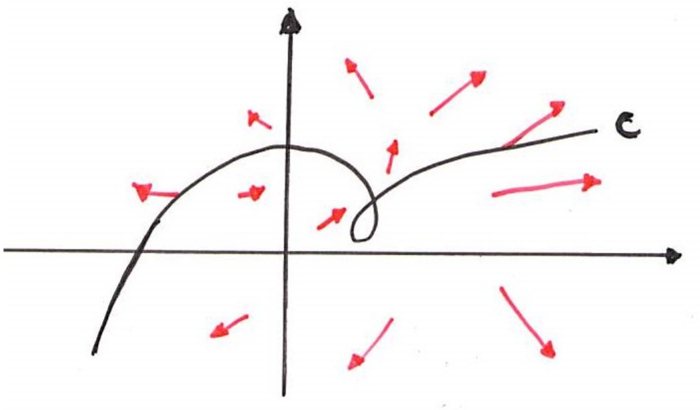


Fig. 5.2. A vector field (in red) and a curve in $\mathbb{R}^2$

$c : [a, b] \to \Omega$ of class C^1 in (a, b), see Figure 5.2. With these ingredients we can consider the following function

$$(a, b) \ni \quad t \mapsto f\big(c(t)\big) \cdot c'(t) \quad \in \mathbb{R}$$

It is easy to observe that this function is continuous, and corresponds to a function of one variable with values in $\mathbb{R}$, similar to the ones studied in Calculus I. Thus, the integral of this function is well defined and can be written

$$\int_a^b f\big(c(t)\big) \cdot c'(t)\,\mathrm{d}t.$$

Since such integrals appear in several contexts, let us give a name to them.

Definition 5.2. *Let $\Omega \subset \mathbb{R}^n$ be open, and let $f : \Omega \to \mathbb{R}^n$ be continuous. Let also $c : [a, b] \to \Omega$ be of class C^1 in (a, b). Then the integral*

$$\int_c f := \int_a^b f\big(c(t)\big) \cdot c'(t)\,\mathrm{d}t$$

is called the integral of f along c, *or* the line integral. *It is also denoted by $\int_a^b f \cdot \mathrm{d}c$.*

Let us observe that if $n = 1$ and $c(t) = t$, then the integral presented in this definition reduce to the usual integral $\int_a^b f(t)\,dt$. Also if we set $f(t) = \begin{pmatrix} f_1(t) \\ \vdots \\ f_n(t) \end{pmatrix}$ and if we use the notation $c(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix}$ then one has

$$\begin{aligned}
&\int_a^b f\big(c(t)\big) \cdot c'(t)\,\mathrm{d}t \\
&= \int_a^b \Big\{ f_1\big(x_1(t), \ldots, x_n(t)\big) x_1'(t) + \cdots + f_n\big(x_1(t), \ldots, x_n(t)\big) x_n'(t) \Big\} \mathrm{d}t \\
&\equiv \int_a^b f_1\,\mathrm{d}x_1 + f_2\,\mathrm{d}x_2 + \cdots + f_n\,\mathrm{d}x_n.
\end{aligned}$$

This last notation is often used in the literature, and has a special meaning in a more general context (in the framework of differential forms). We shall not touch this subject in these notes.

Example 5.3. *Consider $f : \mathbb{R}^2 \to \mathbb{R}^2$ with $f(x,y) = \begin{pmatrix} x^2 y \\ y^3 \end{pmatrix}$, and let $c : [0,1] \to \mathbb{R}^2$ with $c(t) = \begin{pmatrix} t \\ t \end{pmatrix}$. Then one has*

$$\int_c f = \int_0^1 \begin{pmatrix} t^2 t \\ t^3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \mathrm{d}t = \int_0^1 2t^3\,\mathrm{d}t = \frac{1}{2}.$$

Obseve that if we use a different parametrization of the same curve, namely $c(t) = \begin{pmatrix} t^2 \\ t^2 \end{pmatrix}$, then one obtains

$$\int_c f = \int_0^1 \begin{pmatrix} t^4 t^2 \\ t^6 \end{pmatrix} \cdot \begin{pmatrix} 2t \\ 2t \end{pmatrix} \mathrm{d}t = \int_0^1 4t^7\,\mathrm{d}t = \frac{1}{2}.$$

In the previous example, we have obtained the same result with two different parametrizations of the same curve. Is this an accident, or is there a general result ? For the next statement, we recall that the notion of diffeomorphism has been introduced in Definition 2.4.

Lemma 5.4 (Independence of the parametrization)**.** *Let $\Omega \subset \mathbb{R}^n$ be open, and let $f : \Omega \to \mathbb{R}^n$ be continuous. Let also $c : [a,b] \to \Omega$ be of class C^1 in (a,b), and let $\varphi : [c,d] \to [a,b]$ be a diffeomorphism of class C^1 in (c,d). Assume that $\varphi(c) = a$ and $\varphi(d) = b$. Then the following equality hold:*

$$\int_{c \circ \varphi} f = \int_c f. \tag{5.1}$$

In simpler tems, the above statement means that a line integral is independent of the parametrization but depends only on the curve itself, and of the function f, of course.

Proof. Let us consider the parametric curve $c \circ \varphi : [c, d] \to \Omega$ which is of class C^1 on (c, d). Then one has

$$\begin{aligned}\int_{c\circ\varphi} &= \int_c^d f\big(c(\varphi(t))\big) \cdot \frac{\mathrm{d}c(\varphi(t))}{\mathrm{d}t}\,\mathrm{d}t \\ &= \int_c^d f\big(c(\varphi(t))\big) \cdot c'\big(\varphi(t)\big)\varphi'(t)\,\mathrm{d}t.\end{aligned}$$

By the change of variable $s = \varphi(t)$ with $\mathrm{d}s = \varphi'(t)\mathrm{d}t$ the previous expression is equal to

$$\begin{aligned}&\int_{\varphi(c)=a}^{\varphi(d)=b} f\big(c(s)\big) \cdot c'(s)\,\mathrm{d}s \\ &\int_a^b f\big(c(t)\big) \cdot c'(t)\,\mathrm{d}t \\ &= \int_c f,\end{aligned}$$

which finishes the proof. □

Let us observe that the definition of a line integral can be extended to a slightly more general context. Indeed, it is known from Calculus I that

$$\int_a^c g(x)\,\mathrm{d}x = \int_a^b g(x)\,\mathrm{d}x + \int_b^c g(x)\,\mathrm{d}x$$

whenever $g : [a, c] \to \mathbb{R}$ is continuous and b is a point in the interval $[a, c]$. The same idea can be applied here. Let us consider a curve c given by a finite sequence of parametric curves c_j taking values in an open set $\Omega \subset \mathbb{R}^n$, each of them of class C^1. Then we can set $c = c_1 \cup c_2 \cup \cdots \cup c_N$ and set

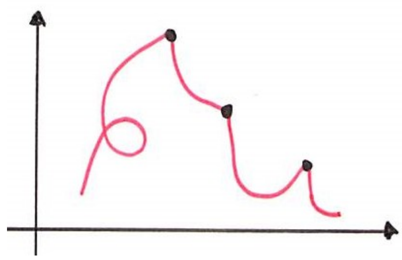


Fig. 5.3. Curve with singular points (black dots)

$$\int_c f = \int_{c_1} f + \int_{c_2} f + \cdots + \int_{c_N} f$$

whenever f is a continuous function from Ω to $\mathbb{R}^n$. We say that such a curve has some *singular points*, see Figure 5.3. The curve c is also called *a path*, and this path is *closed* if the starting point of c_1 coincides with the final point of c_N, see Figure 5.4.

Note that so far we have not considered the notion of orientation, and in fact Lemma 5.4 is slightly misleading in the following sense: In the statement, we have assumed that $\varphi(c) = a$ and $\varphi(d) = b$, and this has been used in the proof. What would have happened if we had only assumed that φ is a diffeomorphism from $[c, d]$ to $[a, b]$, without specifying the image of c and d ? Then the equality (5.1) might not always be correct, it would depend on the orientation of the parametrization.

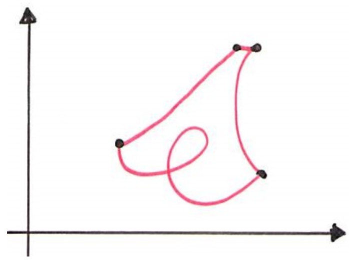


Fig. 5.4. Closed curve with singular points

Definition 5.5. *For a parametric curve $c : [a,b] \to \mathbb{R}^n$, the* opposite curve *or* reverse curve c^- *is defined by*

$$c^- : [a,b] \to \mathbb{R}^n, \qquad c^-(t) = c(a+b-t).$$

The following statement can then easily be proved:

Lemma 5.6. *Let $\Omega \subset \mathbb{R}^n$ be open, and let $f : \Omega \to \mathbb{R}^n$ be continuous. Let also $c : [a,b] \to \Omega$ be of class C^1 in (a,b). Then*

$$\int_{c^-} f = -\int_c f.$$

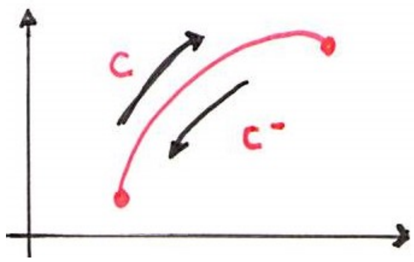


Fig. 5.5. A reverse curve

The result together with the content of Lemma 5.4 means that line integrals do not depend on the precise parametrization of a curve, but they depend on the choice of the initial point and the final point on the curve. More precisely, they depend on which direction one integrates along the curve. Only two choices are possible, and the two different results differ only by a minus sign.

Proof. One has

$$\begin{aligned}
\int_{c^-} f &= \int_a^b f\big(c^-(t)\big) \cdot \frac{\mathrm{d}c^-(t)}{\mathrm{d}t}\,\mathrm{d}t \\
&= \int_a^b f\big(c(a+b-t)\big) \cdot \frac{\mathrm{d}c(a+b-t)}{\mathrm{d}t}\,\mathrm{d}t \\
&= -\int_a^b f\big(c(a+b-t)\big) \cdot c'(a+b-t)\,\mathrm{d}t \\
&= -\int_b^a f\big(c(s)\big) \cdot c'(s)\big(-\mathrm{d}s\big) \\
&= -\int_a^b f\big(c(s)\big) \cdot c'(s)\,\mathrm{d}s,
\end{aligned}$$

where we have performed the changed of variable $s = a+b-t$ with $\mathrm{d}s = -\mathrm{d}t$. Note that at the last step one has exchanged the boundaries of the integral, which produces an additional minus sign. We end up with three minus signs which reduce to one minus sign. $\square$

5.2 Curve integrals and potential functions

In the previous section we have seen that line integrals do not depend on the parametrization of a curve (just in the orientation of the path). In this section, we shall see that in some cases, the integral does not even depend on the choice of the curve, it depends only on the two end points.

Definition 5.7. *Let $\Omega \subset \mathbb{R}^n$ be open and let $f : \Omega \to \mathbb{R}^n$. A function $\phi : \Omega \to \mathbb{R}$ is called* a potential function for f *if ϕ is diffentiable on Ω and if the equality $\nabla\phi = f$ holds on Ω.*

Observe that if we ask ϕ to be of class C^1, then f would be a continuous function. Similarly, if ϕ is of class C^k on Ω for some $k \geq 1$, then f is of class C^{k-1} on Ω.

The main result of this section is provided in the following statement.

Proposition 5.8. *Let $\Omega \subset \mathbb{R}^n$ be open, let $\phi : \Omega \to \mathbb{R}$ be of class C^1, and set $f := \nabla\phi$. Consider a parametric curve $c : [a, b] \to \Omega$ of class C^1. Then,*

$$\int_c f \equiv \int_a^b f\big(c(t)\big) \cdot c'(t)\,\mathrm{d}t = \phi\big(c(b)\big) - \phi\big(c(a)\big).$$

This equality means clearly that the curve c does not play any role. Only the end points $c(b)$ and $c(a)$ are important. However, the condition that $f = \nabla\phi$ is a very strong condition and does not hold for all vector fields f, see Definition 5.1.

Proof. One has

$$\begin{aligned}
&\int_a^b f\big(c(t)\big) \cdot c'(t)\,\mathrm{d}t \\
&= \int_a^b \nabla\phi\big(c(t)\big) \cdot c'(t)\,\mathrm{d}t \\
&= \int_a^b \underbrace{{}^t\big(\nabla\phi\big(c(t)\big)\big)c'(t)}_{\text{product of 2 Jacobian matrices}}\,\mathrm{d}t \\
&= \int_a^b \mathcal{J}_\phi\big(c(t)\big)\mathcal{J}_c(t)\,\mathrm{d}t \\
&= \int_a^b \mathcal{J}_{\phi\circ c}(t)\,\mathrm{d}t \\
&= \int_a^b \frac{\mathrm{d}(\phi\circ c(t))}{\mathrm{d}t}\,\mathrm{d}t \\
&= [\phi\circ c](b) - [\phi\circ c](a) \\
&= \phi\big(c(b)\big) - \phi\big(c(a)\big),
\end{aligned}$$

as expected. ☕

From the previous statement one directly infers the next corollary, if one remembers that a closed curve means precisely that $c(b) = c(a)$.

Corollary 5.9. *In the previous setting, if the curve c is closed then $\int_c f = 0$.*

Let us provide a rather famous example:

Example 5.10. *Let us consider $\Omega := \mathbb{R}^3 \setminus \{\mathbf{0}\}$, and for set $r := \|X\|$ for any $X \in \Omega$. Clearly, one has $r > 0$ for any $X \in \Omega$. We now define the function $F : \Omega \to \mathbb{R}^3$ given by $F(X) := -\frac{GMm}{r^2}\frac{X}{r}$, where G, M and m are some fixed constants. For example, M and m could be the masses of some objects, and G the universal gravitational constant. Observe also that the expression $\frac{X}{r}$ corresponds to the unit vector in the direction by X, which is sometimes denoted by $\hat{X}$. Let us also define $\phi : \Omega \to \mathbb{R}$ by $\phi(X) := \frac{GMm}{r}$. Then, it is easily observed that $F = \nabla\phi$. Based on Proposition 5.8 one infers that if c denotes any curve between two points A and B of Ω, one has*

$$\int_c F = \phi(B) - \phi(A) = GMm\Big(\frac{1}{\|B\|} - \frac{1}{\|A\|}\Big) = -GMm\Big(\frac{1}{\|A\|} - \frac{1}{\|B\|}\Big).$$

In particular, if we set $R := \|A\|$ and $\|B\| = R + h$ with h small with respect to R one gets

$$\begin{aligned} W &:= -GMm\Big(\frac{1}{\|A\|} - \frac{1}{\|B\|}\Big) \\ &= -GMm\Big(\frac{1}{R} - \frac{1}{R+h}\Big) \\ &= -\frac{GMm}{R}\Big(1 - \frac{1}{1+\frac{h}{R}}\Big) \\ &\cong -\frac{GMm}{R}\Big(1 - 1 + \frac{h}{R}\Big) \\ &= -\frac{GM}{R^2}mh, \end{aligned}$$

where $\frac{GM}{R^2}$ is often denoted by g when M is the mass of the earth and R its radius. Note that a Taylor expansion up to the first term has been used for the approximate equality.

Remark 5.11. *Let us mention that there exist additional relations between the existence of a potential function ϕ and the equality $\int_c \nabla\phi = 0$. Such relations also play an important role in complex analysis. We shall not develop this any further here.*

As a final remark, let us stress that there exist other types of integrals which can be computed along a curve and which depend on the parametrization of the curve. For example, if $f : \Omega \to \mathbb{R}^n$ and $c : [a, b] \to \Omega$ is a parametric curve of class C^1, then the integrals $\int_a^b \big\|f\big(c(t)\big)\big\|\mathrm{d}t$, or $\int_a^b \big\|f\big(c(t)\big)\big\|\|c(t)\|\mathrm{d}t$, do depend on the parametrization. Many other examples can be exhibited, but it is also possible to exhibit examples which do not depend on the precise parametrization (but still on the orientation).

Chapter 6

Integrals in $\mathbb{R}^n$

In this chapter we shall generalize the Riemann sums which have been introduced in Calculus I.

6.1 Motivation and definition

First of all, let us recall the construction of the Riemann sums. We consider $f : [a, b] \to \mathbb{R}$ and fix $N \in \mathbb{N}$. The interval $[a, b]$ is then divided into N subintervals, and this defines a sequence of point in $[a, b]$ satisfying

$$a = x_0 < x_1 < x_2 < \cdots < x_{N-1} < x_N = b.$$

The points x_{j-1} and x_j define the j^{th} subinterval $[x_{j-1}, x_j]$. We also fix an additional point x_j^* in each subinterval, namely $x_j^* \in [x_{j-1}, x_j]$. The choice of $\{x_j\}_{j=0}^N$ together with the choice of $\{x_j^*\}_{j=1}^N$ define a partition $\mathcal{P}$ of $[a, b]$. We can then define the Riemann sum based on $\mathcal{P}$

$$\sum_{\mathcal{P}} f := \sum_{j=1}^{N} (x_j - x_{j-1}) f(x_j^*). \tag{6.1}$$

By definition, we say that the Riemann integral of f exists if by taking finer and finer partitions of $[a, b]$ (with $|x_j - x_{j-1}|$ becoming smaller and smaller) the quantity defined in (6.1) converges independently of the choice of the partitions. Whenever it converges, we write $\int_a^b f(x)\,\mathrm{d}x$ for the limiting value.

Note that the process is rather complicated since one has to consider all possible partitions of the interval $[a, b]$. Fortunately, a theorem seen in Calculus I says that this process is converging if the function f is continuous on $[a, b]$. In summary, it means that Riemann integrals exist for all continuous functions on closed intervals (but more general functions also admit a Riemann integral). Figure 6.1 is an illustration of this process.

It is now natural to wonder what is the analogue construction for functions $f : \Omega \to \mathbb{R}$ with $\Omega \subset \mathbb{R}^n$? Before looking at the general theory, let us start with $n = 2$ and with

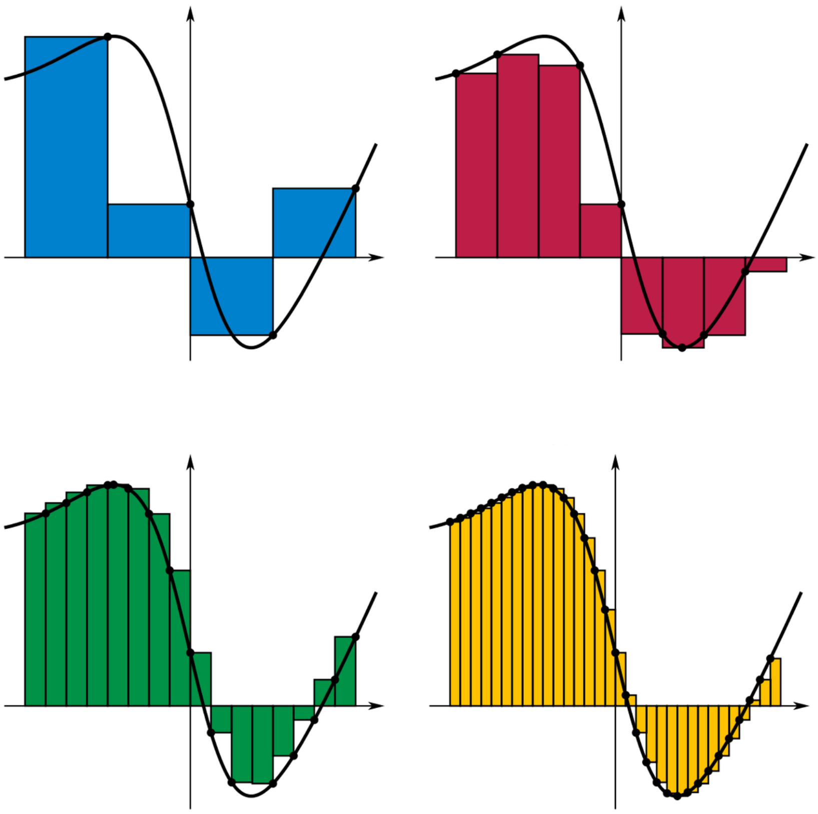


Fig. 6.1. Various Riemann sums, see [6]

a rather simple domain Ω, namely a rectangle. More explicitly, let us consider

$$f : [a, b] \times [c, d] \to \mathbb{R}$$

with $\Omega := [a, b] \times [c, d]$ a rectangle in $\mathbb{R}^2$. What is the meaning of $\int_\Omega f(x, y)\,\mathrm{d}x\,\mathrm{d}y$?

First of all, let us subdivide the rectangle into smaller rectangles. For this we consider two families of points $\{x_i\}_{i=0}^N$ and $\{y_j\}_{j=0}^M$ satisfying

$$a = x_0 < x_1 < x_2 < \cdots < x_{N-1} < x_N = b$$

and

$$c = y_0 < y_1 < y_2 < \cdots < y_{M-1} < y_M = d.$$

Inside each small rectangle $R_{ij} := [x_{i-1}, x_i] \times [y_{j-1}, y_j]$ one chooses a point $\big(x^*_{ij}, y^*_{ij}\big)$, as shown in Figure 6.2. Note that this double indexation can not be avoided. Again, the choice of $\{x_j\}_{j=0}^N$, $\{y_k\}_{k=0}^M$ together with the choice of $\big\{\big(x^*_{ij}, y^*_{ij}\big)\big\}_{i=1,j=1}^{i=N,j=M}$ define a partition $\mathcal{P}$ of $[a, b] \times [c, d]$. Then, by mimicking the expression provided in (6.1) it is natural to set

$$\sum_{\mathcal{P}} f := \sum_{i=1}^N \sum_{j=1}^M (x_i - x_{i-1})\,(y_j - y_{j-1})\, f\big(x^*_{ij}, y^*_{ij}\big) \tag{6.2}$$

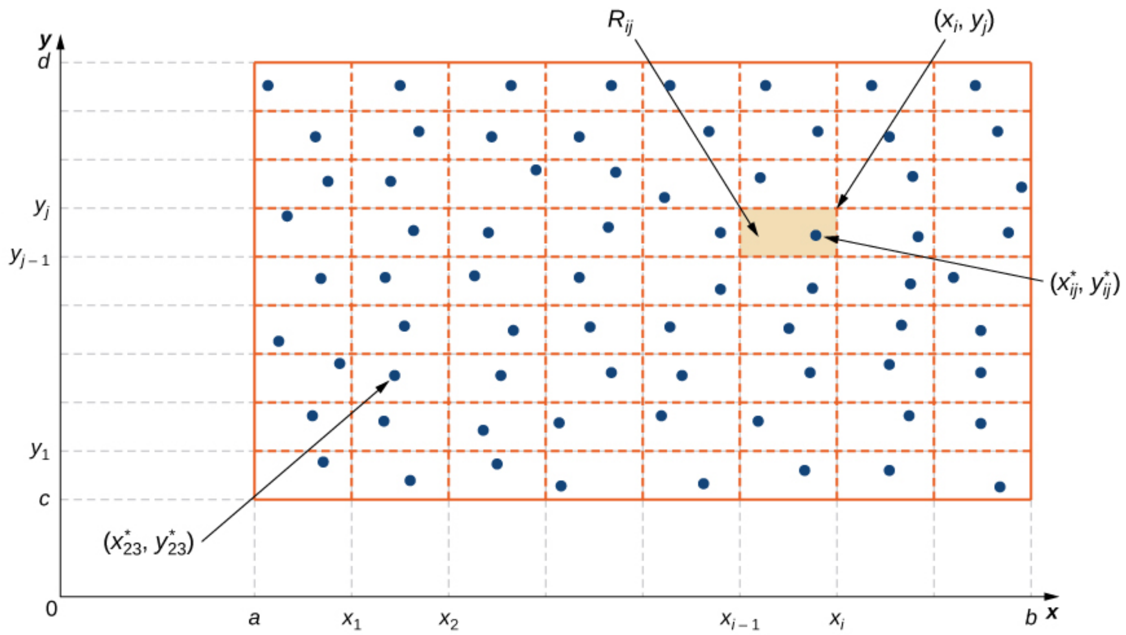


Fig. 6.2. Partition for $n = 2$, see [7]

and call this expression a Riemann sum based on $\mathcal{P}$. Note that in this expression, the product $(x_i - x_{i-1})(y_j - y_{j-1})$ represents the surface (area) of the small rectangle R_{ij}. The last step is again to consider finer and finer partitions of the rectangle $[a, b] \times [c, d]$ and to look for the convergence of the Riemann sums. If the r.h.s. of (6.2) converges independently of the choice of the partitions to a fixed quantity, we shall say that the Riemann integral of f over $[a, b] \times [c, d]$ exists.

Clearly, this construction can be extended to any hyperbox in $\mathbb{R}^n$, simply by considering n indexes instead of just i and j. Here, hyperbox means a domain Ω of the form

$$\Omega := [a_1, b_1] \times [a_2, b_2] \times \cdots [a_n, b_n] \quad \subset \mathbb{R}^n \tag{6.3}$$

which generalizes a rectangle in higher dimensions. Note that there will also be n sums instead of 2 in (6.2). and that the product

$$(x_i - x_{i-1})(y_j - y_{j-1}) \dots (z_k - z_{k-1})$$

appearing in the expression would simply denote the volume of a small hyperbox inside the domain Ω.

As in the one dimensional case recalled at the beginning of this section one important result can now be stated. The lengthy and slightly messy proof is not provided here but can be found in several classical books on functions of n variables.

Theorem 6.1. *Consider a hyperbox Ω in $\mathbb{R}^n$ of the form provided in* (6.3)*, and let $f : \Omega \to \mathbb{R}$ be continuous. Then the Riemann integral of f over Ω exists, and is denoted* *simply by $\int_\Omega f \mathrm{d}V$ (the letter V is for volume), or by*

$$\iint \cdots \int_\Omega f(x_1, x_2, \dots, x_n) \,\mathrm{d}x_1 \,\mathrm{d}x_2 \dots \mathrm{d}x_n. \tag{6.4}$$

Note that the above result means that if f is a continuous function on Ω the limit of the Riemann sums (6.2) over finer and finer partitions exist and is independent of the limiting process. We then get a number which is denoted by the expression provided in (6.4). Note that the number of integral signs is equal to n.

Before moving to more general domains Ω let us discuss two possible interpretations of such integrals. In dimension 1 and for a positive function an integral can be understood as the area below the curve. In higher dimension, such an interpretation is also possible, see Figure 6.3. Indeed, in the sum (6.2), each term of the form $(x_i - x_{i-1})\,(y_j - y_{j-1})\, f\big(x^*_{ij}, y^*_{ij}\big)$ can be interpreted as the volume of a rectangular tube $(x_i - x_{i-1})\,(y_j - y_{j-1})$ and of height $f\big(x^*_{ij}, y^*_{ij}\big)$. By summing up these contributions one obtains the volume below the graph of the function f.

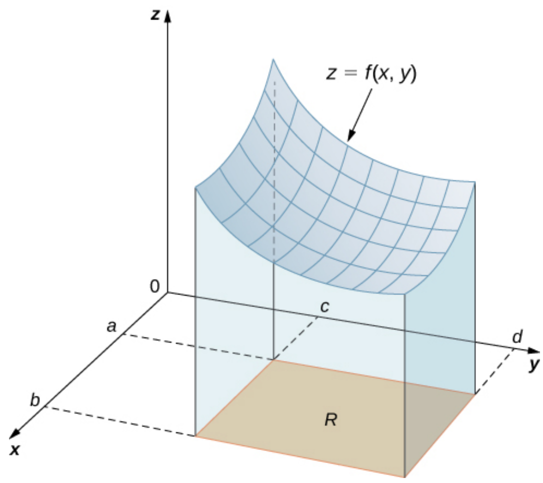


Fig. 6.3. Volume below the graph

The same integral in dimension 2 can also be interpreted as the computation of the mass of a 2-dimensional object, if the function f represents the area density denoted by σ in Figure 6.4.

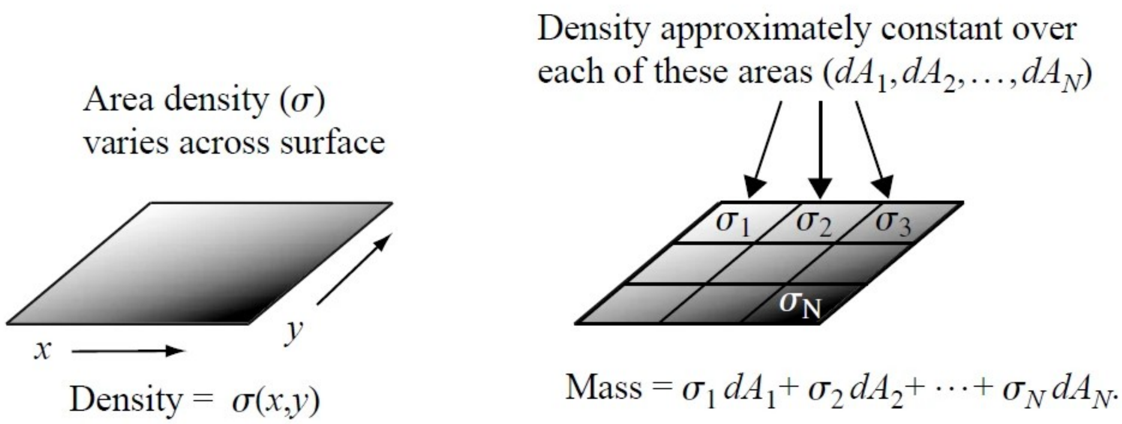


Fig. 6.4. Mass = Integral of a density

In this interpretation, what is called "Mass" below the second picture is in fact a Riemann sum for a given partition, and the Riemann integral is obtained by taking finer and finer partitions of the rectangle.

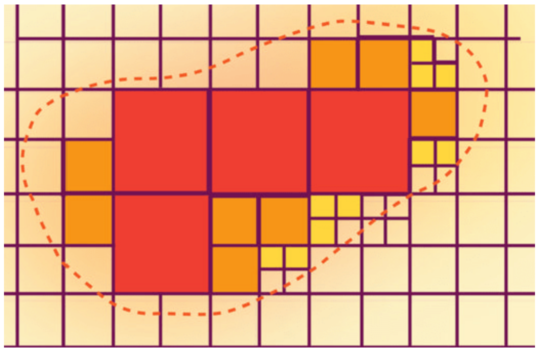


Fig. 6.5. Partition of Ω, see [9]

Now, what about more general domain than a rectangle in dimension 2, or a hyperbox in arbitrary dimension ? The general theory is rather complicated, but for domains Ω which are regular enough, the key idea is easy to understand. For example, if we consider the domain Ω presented in Figure 6.5. Then one can partition it by considering smaller and smaller rectangles (or hyperboxes in higher dimension) and proceed as before by considering finite sums. What could happen is that the boundary of Ω is very complicated (like the surface of a cauliflower) and that the successive approximations of the surface Ω does not not lead to a converging sequence of Riemann sums. ☹

Fortunately, if the boundary is piecewise smooth (made of a finite number of pieces which are infinitely many times differentiable), then the successive approximations will converge, and one can thus define a Riemann integral on any such domain. We haven't really defined the content of the previous sentence (what is the exact meaning of a piecewise smooth boundary) but in the applications this condition will be easy to check. The next statement is the generalization of Theorem 6.1 for more general domain Ω

Theorem 6.2. *Let Ω be an open and bounded subset in $\mathbb{R}^n$ with a boundary which is piecewise smooth. Let $f : \Omega \to \mathbb{R}$ which is continuous and bounded. Then the Riemann integral of f on Ω exists, and is denoted by $\int_\Omega f \mathrm{d}V$, or by the longer expression* (6.4).

Observe that we have taken f bounded on Ω in order to avoid any singularity of f on the boundary of Ω. For example, in dimension 1 such a singularity could appear if we consider $\Omega = (0, 1)$ and $f(x) = 1/x$. The function f is continuous on Ω, but f is not bounded on this domain since it takes arbitrarily large values. We have also taken Ω bounded in order to avoid large domain, like for example $(0, \infty) \times (0, \infty)$ in $\mathbb{R}^2$.

6.2 Properties and examples

In the next statement we gather a few results which can be obtained from the definition of a Riemann integral. Such properties have already been mentioned in the one dimensional case.

Lemma 6.3. *Let Ω be an open and bounded subset in $\mathbb{R}^n$ with a boundary which is piecewise smooth, and let $f, g : \Omega \to \mathbb{R}$ which are continuous and bounded. Let also $c \in \mathbb{R}$. Then*

(i) $\int_\Omega (f + g)\mathrm{d}V = \int_\Omega f \,\mathrm{d}V + \int_\Omega g \,\mathrm{d}V$,

(ii) $\int_\Omega cf \,\mathrm{d}V = c \int_\Omega f \,\mathrm{d}V$,

(iii) $\int_\Omega (fg)\,\mathrm{d}V \neq \big(\int_\Omega f\,\mathrm{d}V\big)\big(\int_\Omega g\,\mathrm{d}V\big)$ *in general.*

We shall now start computing such integrals on a few examples. Note that we first consider the domain Ω given by a rectangle, or more generally a hyperbox. More general domains are introduced later on.

Examples 6.4. *(i) Let* $\Omega := [0,1] \times [0,1] \subset \mathbb{R}^2$ *be a unit square, and consider* $\int_\Omega x^2y^2\,\mathrm{d}x\,\mathrm{d}y$*. Then one has*

$$\begin{aligned}\iint_{[0,1]\times[0,1]} x^2y^2\,\mathrm{d}x\,\mathrm{d}y &= \int_0^1 \Big[\int_0^1 x^2y^2\,\mathrm{d}y\Big]\mathrm{d}x \\ &= \int_0^1 \Big[x^2\frac{1}{3}y^3\big|_{y=0}^{y=1}\Big]\mathrm{d}x = \frac{1}{3}\Big[\frac{1}{3}x^3\big|_{x=0}^{x=1}\Big] = \frac{1}{9}.\end{aligned}$$

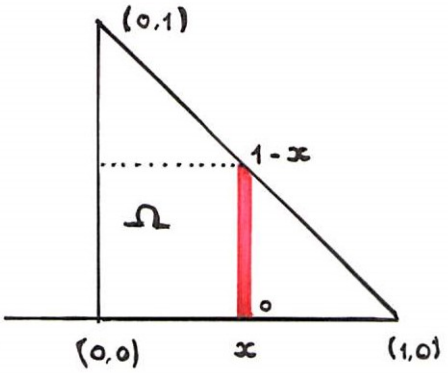
(ii) Let $\Omega := [0,1]\times[0,1]\times[0,1] \subset \mathbb{R}^3$ *be a unit cube, and consider* $\int_\Omega (x+y+z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$*. Then one has*

$$\begin{aligned}\iiint_{[0,1]\times[0,1]\times[0,1]} (x+y+z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z &= \int_0^1 \Big\{\int_0^1 \Big[\int_0^1 (x+y+z)\,\mathrm{d}z\Big]\mathrm{d}y\Big\}\mathrm{d}x \\ &= \int_0^1 \Big\{\int_0^1 \Big[\big(xz+yz+\frac{1}{2}z^2\big)\big|_{z=0}^{z=1}\Big]\mathrm{d}y\Big\}\mathrm{d}x = \int_0^1 \Big\{\int_0^1 \Big[\big(x+y+\frac{1}{2}\big)\Big]\mathrm{d}y\Big\}\mathrm{d}x \\ &= \int_0^1 \Big\{\big(xy+\frac{1}{2}y^2+\frac{1}{2}y\big)\big|_{y=0}^{y=1}\Big\}\mathrm{d}x = \int_0^1 (x+1)\,\mathrm{d}x \\ &= \big(\frac{1}{2}x^2+x\big)\big|_{x=0}^{x=1} = \frac{3}{2}.\end{aligned}$$

In these two examples, the integrations have been performed in a certain order, but one easily observes that this order is irrelevant. For example, one could have computed

$$\begin{aligned}\iint_{[0,1]\times[0,1]} x^2y^2\,\mathrm{d}x\,\mathrm{d}y &= \int_0^1 \Big[\int_0^1 x^2y^2\,\mathrm{d}x\Big]\mathrm{d}y \\ &= \int_0^1 \Big[y^2\frac{1}{3}x^3\big|_{x=0}^{x=1}\Big]\mathrm{d}y = \frac{1}{3}\Big[\frac{1}{3}y^3\big|_{y=0}^{y=1}\Big] = \frac{1}{9}.\end{aligned}$$

Obviously, one gets the same result. Note that the difference in the two processes can be visualized on Figure 6.2. In the first computation, by performing firstly the integration over y is like summing the different contributions in this figure along a column: the variable x is fixed, and we sum the contributions vertically. Then the integration over x does the sum over the contribution of the different columns. On the other hand, by performing firstly the integral over x is like summing along a row: the variable y is fixed, and we sum the different contributions horizontally. Then, the integration over y does the sum of the contribution over the different rows. In summary, the order of integration is irrelevant as long as one takes all small contributions into account.



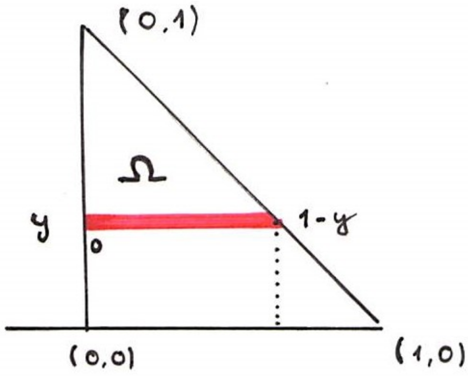


Fig. 6.6. Two representations of the domain of integration Ω

Let us now consider an example with a slightly more complicated domain Ω. In Figure 6.6 we present the domain Ω and emphasize the order chosen for the integration. The first picture corresponds to the first integral provided below, and the second picture to the second integral provided below.

Example 6.5. *Let* Ω *be the domain represented in Figure 6.6, and consider the integral* $\int_\Omega x^2y^2\,\mathrm{d}x\,\mathrm{d}y$*. Observe that for a fixed* x*, the domain of integration of* y *is the interval* $[0,1-x]$*, and therefore one has*

$$\begin{aligned}&\iint_\Omega x^2y^2\,\mathrm{d}x\,\mathrm{d}y = \int_0^1\Big[\int_0^{1-x} x^2y^2\,\mathrm{d}y\Big]\mathrm{d}x = \int_0^1\Big[\big(x^2\frac{1}{3}y^3\big)\big|_{y=0}^{y=1-x}\Big]\mathrm{d}x\\ &= \frac{1}{3}\int_0^1\big[x^2(1-x)^3\big]\mathrm{d}x = \frac{1}{3}\int_0^1\big(x^2-3x^3+3x^4-x^5\big)\mathrm{d}x = \frac{1}{3}\Big(\frac{1}{3}-\frac{3}{4}+\frac{3}{5}-\frac{1}{6}\Big)\\ &= \frac{1}{180}.\end{aligned}$$

Alternatively, for any fixed y*, the domain of integration of* x *is the interval* $[0,1-y]$*, and therefore one has*

$$\begin{aligned}&\iint_\Omega x^2y^2\,\mathrm{d}x\,\mathrm{d}y = \int_0^1\Big[\int_0^{1-y} x^2y^2\,\mathrm{d}x\Big]\mathrm{d}y = \int_0^1\Big[\big(y^2\frac{1}{3}x^3\big)\big|_{x=0}^{x=1-y}\Big]\mathrm{d}x\\ &= \frac{1}{3}\int_0^1\big[y^2(1-y)^3\big]\mathrm{d}y = \frac{1}{3}\int_0^1\big(y^2-3y^3+3y^4-y^5\big)\mathrm{d}y = \frac{1}{3}\Big(\frac{1}{3}-\frac{3}{4}+\frac{3}{5}-\frac{1}{6}\Big)\\ &= \frac{1}{180}.\end{aligned}$$

The next example is taking place in $\mathbb{R}^3$ and the domain Ω is deliminated by some planes.

Example 6.6. *Let* Ω *be the interior domain defined by the four planes*

$$y=1, \quad y=-x, \quad z=0, \quad z=-x,$$

and consider the integral $\int_\Omega \mathrm{e}^{x+y+z}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$. For this integral, the order of integration is motivated by the fact that two planes are defined in terms of the variable x. Thus one has

$$
\begin{aligned}
&\iiint_\Omega \mathrm{e}^{x+y+z}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = \int_{-1}^{0}\Big\{\int_{-x}^{1}\Big[\int_{0}^{-x}\mathrm{e}^{x+y+z}\mathrm{d}z\Big]\mathrm{d}y\Big\}\mathrm{d}x \\
&= \int_{-1}^{0}\Big\{\int_{-x}^{1}\mathrm{e}^{x+y}\Big[\mathrm{e}^{z}\big|_{z=0}^{z=-x}\Big]\mathrm{d}y\Big\}\mathrm{d}x = \int_{-1}^{0}\Big\{\mathrm{e}^{x}\big(\mathrm{e}^{-x}-1\big)\int_{-x}^{1}\mathrm{e}^{y}\mathrm{d}y\Big\}\mathrm{d}x \\
&= \int_{-1}^{0}\Big\{\big(1-\mathrm{e}^{x}\big)\big(\mathrm{e}-\mathrm{e}^{-x}\big)\Big\}\mathrm{d}x = \mathrm{e}\int_{-1}^{0}\big(1-\mathrm{e}^{x}\big)\mathrm{d}x - \int_{-1}^{0}\big(\mathrm{e}^{-x}-1\big)\mathrm{d}x \\
&= \mathrm{e}\big(x-\mathrm{e}^{x}\big)\big|_{-1}^{0} - \big(-\mathrm{e}^{-x}-x\big)\big|_{-1}^{0} = \mathrm{e}(0-1+1+\mathrm{e}^{-1}) - (-1-0+\mathrm{e}-1) \\
&= 3-\mathrm{e}.
\end{aligned}
$$

Note that a common mistake in this example would be to put some factors depending on x in front of the integral over x. This would clearly be a mistake since the final result should not depend on x.

Let us now motivate the next section. Consider the domain Ω given by the unit disk in $\mathbb{R}^2$, namely $\mathcal{B}_1(\mathbf{0})$. If we had to deal with the integral $\int_\Omega (x^2+y^2)\,\mathrm{d}x\,\mathrm{d}y$, what would be the best way to compute such an integral ? It seems that polar coordinates would be useful, but how can one implement new coordinates in an integral ?

6.3 Change of variables

Let us first recall the change of variable in the framework of an integral on an interval. Consider a continuous function $f:[a,b]\to\mathbb{R}$ and let $\phi:[c,d]\to[a,b]$ be a bijective function which is C^1 on (c,d) with $\phi'(x)>0$ for any $x\in(c,d)$. Clearly, $\phi(c)=a$ and $\phi(d)=b$, and ϕ is nothing but a diffeomorphism of class C^1, as introduced in Definition 2.4. Then one has

$$\int_a^b f(x)\,\mathrm{d}x = \int_c^d f\big(\phi(y)\big)\phi'(y)\,\mathrm{d}y. \tag{6.5}$$

What it the natural generalization of this procedure in n dimensions, in particular what is $\phi'(y)$ in the more general framework ?

First of all, let us generalize the function ϕ. Consider Λ and Ω two open and bounded subset of $\mathbb{R}^n$, and let $\phi:\Lambda\to\Omega$ be a diffeomorphism of class C^1, which means that ϕ is bijective, of class C^1, that its inverse exists and is also of class C^1. Let us also consider a continuous and bounded function $f:\Omega\to\mathbb{R}$. If we impose the necessary conditions on the boundaries of Ω and Λ, what would be the equality

$$
\begin{aligned}
&\iint\cdots\int_\Omega f(x_1,x_2,\dots,x_n)\,\mathrm{d}x_1\,\mathrm{d}x_2\dots\mathrm{d}x_n \\
&= \iint\cdots\int_\Lambda f\big(\phi(y_1,y_2,\dots,y_n)\big)\dots ? \dots\mathrm{d}y_1\,\mathrm{d}y_2\dots\mathrm{d}y_n \quad ?
\end{aligned}
$$

The answer to this question and the precise formula are provided in the following statement. We recall that the Jacobian matrix has been introduced in Definition 4.7.

Theorem 6.7. *Let Ω be an open and bounded set in $\mathbb{R}^n$ with a boundary which is piecewise smooth. Let $f : \Omega \to \mathbb{R}$ be continuous and bounded. Consider also an open and bounded set Λ with a boundary which is piecewise smooth, and assume that there exists $\phi : \Lambda \to \Omega$ a diffeomorphism of class C^1. Then the following equality holds:*

$$\begin{aligned}&\iint \cdots \int_\Omega f(x_1, x_2, \dots, x_n)\,\mathrm{d}x_1\,\mathrm{d}x_2 \dots \mathrm{d}x_n \\ &= \iint \cdots \int_\Lambda f\big(\phi(y_1, y_2, \dots, y_n)\big)\det\big(\mathcal{J}_\phi(y_1, y_2, \dots, y_n)\big)\,\mathrm{d}y_1\,\mathrm{d}y_2 \dots \mathrm{d}y_n\end{aligned}$$

where $\det(\mathcal{J}_\phi)$ *denotes the determinant of the Jacobian matrix of* ϕ.

Let us make some comments about this equality. First of all, observe that since ϕ maps the subset Λ of $\mathbb{R}^n$ to the subset Ω of $\mathbb{R}^n$, its Jacobian matrix is a square matrix and its determinant is well defined. In the special case $n = 1$, then we get nothing but the expression ϕ' of (6.5). There is also one ambiguity in the above formula, which is already visible for $n = 1$, namely the orientation. Indeed, if instead of writing

$$\int_c^d f\big(\phi(y)\big)\phi'(y)\,\mathrm{d}y$$

in (6.5) we had written $\int_{(c,d)} f\big(\phi(y)\big)\phi'(y)\,\mathrm{d}y$ we would not do a mistake in the above example. But if ϕ was still bijective, but with $\phi(c) = b$ and $\phi(d) = a$, then it would follow that $\phi'(y) < 0$ and the notation $\int_{(c,d)} f\big(\phi(y)\big)\phi'(y)\,\mathrm{d}y$ would lead to an ambiguity. Indeed, one has to be careful with the orientation, and in the present case one would have to compute $\int_d^c f\big(\phi(y)\big)\phi'(y)\,\mathrm{d}y$. Thus, one way to resolve this problem is to speak about the orientation of Λ and of Ω. Since this problem is only a difference of sign, and since in the applications one rarely does a mistake, we do not provide additional information for this small problem, but just warn about a possible mistake in the sign of the resulting integral.

Let us still mention that the appearance of the determinant is rather natural. Indeed, the determinant has some properties which connect it directly to the notion of volume, as presented in the following exercise:

Exercise 6.8. *Let X, Y be two vectors in $\mathbb{R}^2$. Check that the area of the parallelogram spanned by X and Y is equal to the absolute value of the determinant of the matrix $(X\ Y) \in M_2(\mathbb{R})$. More generally, if $X_1, \dots, X_n$ are n vectors of $\mathbb{R}^n$, one writes $\mathsf{Vol}(X_1, \dots, X_n)$ for the volume of the n-dimensional box spanned by $X_1, \dots, X_n$. Why is it natural to have*

$$\mathsf{Vol}(X_1, \dots, X_n) = |\det(X_1 \dots X_n)| \ ?$$

In the following examples, we compute some integrals with respect to classical coordinates.

Example 6.9 (Polar coordinates). *Let us consider $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x,y) = \mathrm{e}^{-(x^2+y^2)}$. Our aim is to compute the integral*

$$\iint_{\mathbb{R}^2} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \iint_{\mathbb{R}^2} \mathrm{e}^{-(x^2+y^2)}\,\mathrm{d}x\,\mathrm{d}y.$$

Because of the symmetry of the function, and of the domain of integration, it is rather natural (and in fact necessary) to use polar coordinates, as introduced in Example 4.11. The determinant of the Jacobian matrix has been computed in 4.3. With this result one gets

$$\iint_{\mathbb{R}^2} \mathrm{e}^{-(x^2+y^2)}\,\mathrm{d}x\,\mathrm{d}y = \int_0^{2\pi} \Big[\int_0^{\infty} \mathrm{e}^{-r^2} r\,\mathrm{d}r\Big]\mathrm{d}\theta$$
$$= \int_0^{2\pi} \Big[-\frac{1}{2}\mathrm{e}^{-r^2}\big|_0^{\infty}\Big]\mathrm{d}\theta = 2\pi\Big(-\frac{1}{2}(0-1)\Big) = \pi.$$

Note that if we had consider the set $\Omega = [0,\infty)\times[0,\infty)$, which is the first quadrant, one would have obtained

$$\iint_{\Omega} \mathrm{e}^{-(x^2+y^2)}\,\mathrm{d}x\,\mathrm{d}y = \int_0^{\pi/2} \Big[\int_0^{\infty} \mathrm{e}^{-r^2} r\,\mathrm{d}r\Big]\mathrm{d}\theta = \frac{\pi}{2}\frac{1}{2} = \frac{\pi}{4}.$$

Example 6.10 (Spherical coordinates). *In this example we shall use spherical coordinates. Consider*

$$\phi : (0,\infty)\times\mathbb{T}\times[0,\pi) \ni (r,\theta,\varphi) \mapsto \phi(r,\theta,\varphi) \in \mathbb{R}^3$$

with

$$\phi(r,\theta,\varphi) = \big(r\cos(\theta)\sin(\varphi), r\sin(\theta)\sin(\varphi), r\cos(\varphi)\big). \tag{6.6}$$

We can observe that the image of this function is not equal to $\mathbb{R}^3$ but to $\mathbb{R}^3\setminus\{\mathbf{0}\}$, which is an open set. For the Jacobian matrix, one finds

$$\mathcal{J}_\phi(r,\theta,\varphi) = \begin{pmatrix} \cos(\theta)\sin(\varphi) & -r\sin(\theta)\sin(\varphi) & r\cos(\theta)\cos(\varphi) \\ \sin(\theta)\sin(\varphi) & r\cos(\theta)\sin(\varphi) & r\sin(\theta)\cos(\varphi) \\ \cos(\varphi) & 0 & -r\sin(\varphi) \end{pmatrix}, \tag{6.7}$$

and for the Jacobian determinant the above matrix leads to the expression $-r^2\sin(\varphi)$... We are precisely in one of the examples mentioned above in which the orientation of the domain has been changed. No problem takes place if we keep using the three coordinates (r,θ,φ) on the domain $(0,\infty)\times\mathbb{T}\times[0,\pi)$ but use the expression

$$\big|\det\big(\mathcal{J}_\phi(r,\theta,\varphi)\big)\big| = r^2\sin(\varphi).$$

A simple application of this change of coordinates allows us to compute the volume of a ball of radius R, namely

$$\iiint_{\text{sphere of radius } R} 1\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = \int_0^R \Big[\int_0^{2\pi}\Big\{\int_0^{\pi} 1r^2\sin(\varphi)\mathrm{d}\varphi\Big\}\mathrm{d}\theta\Big]\mathrm{d}r$$
$$= \frac{2\pi}{3}R^3\big(-\cos(\varphi)\big)\big|_0^{\pi} = \frac{4\pi}{3}R^3.$$

Let us take this opportunity for introducing a notation which is often used when several variables are integrated. Whenever the domain of integration Ω has the shape of a box, namely $\Omega = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$ one often uses the notation

$$\begin{aligned}
&\iint \cdots \int_\Omega f(x_1, x_2, \ldots, x_n)\, \mathrm{d}x_1\, \mathrm{d}x_2 \ldots \mathrm{d}x_n \\
&= \int_{a_1}^{b_1} \mathrm{d}x_1 \int_{a_2}^{b_2} \mathrm{d}x_2 \cdots \int_{a_n}^{b_n} \mathrm{d}x_n\ f(x_1, x_2, \ldots, x_n).
\end{aligned}$$

We emphasize that this is just a convenient notation which reduces the number of parenthesis (compare it with the expression in the previous example) but which has to be used with some care: the function f contains the dependence on all variables, even if this function is located after the sign $\mathrm{d}x_j$ for any j.

For completeness, let us conclude with the computation of the Jacobian matrix and determinant for another useful change of coordinates.

Example 6.11 (Cylindrical coordinates)**.** *Whenever a problem has a cylindrical symmetry, the following change of variable is useful:*

$$\phi : (0, \infty) \times \mathbb{T} \times \mathbb{R} \ni (r, \theta, z) \mapsto \phi(r, \theta, z) \in \mathbb{R}^3$$

with

$$\phi(r, \theta, z) = \big(r \cos(\theta), r \sin(\theta), z\big).$$

This change of coordinates corresponds to a mixture of polar coordinates for the first two variables, and cartesian coordinate for the third one. In this setting one finds

$$\mathcal{J}_\phi(r, \theta, z) = \begin{pmatrix} \cos(\theta) & -r\sin(\theta) & 0 \\ \sin(\theta) & r\cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{6.8}$$

and $\det\big(\mathcal{J}_\phi(r, \theta, z)\big) = r$.

Chapter 7

Green's theorem

In this chapter we construct a link between two of the integrals previously introduced. The two main ingredients are again a vector field and a curve. Note that the present construction takes place in $\mathbb{R}^2$ only. Generalizations will be provided later on.

Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be a vector field of class C^1, and let $c : [0,1] \to \mathbb{R}^2$ be a closed parametric curve of class C^1, see Figure 7.1. As seen in Definition 5.2, the integral of f along c is well defined and is given by

$$\int_0^1 f\big(c(t)\big) \cdot c'(t)\,\mathrm{d}t. \tag{7.1}$$

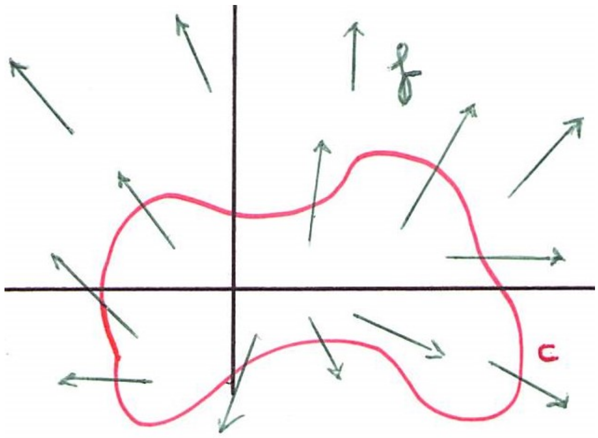


Fig. 7.1. A curve in a vector field

On the other hand, since $f = (f_1, f_2)$ with $f_j : \mathbb{R}^2 \to \mathbb{R}$ also of class C^1, we can consider the following expression

$$\mathrm{curl}_3 f := \partial_1 f_2 - \partial_2 f_1 \equiv \partial_x f_2 - \partial_y f_1$$

with $\mathrm{curl}_3 f : \mathbb{R}^2 \to \mathbb{R}$. If we denote by Ω the interior of the closed curve defined by c, then we can consider the integral

$$\iint_\Omega [\mathrm{curl}_3 f](x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{7.2}$$

as presented in Theorem 6.2. Thus, the two expressions (7.1) and (7.2) are defined only in terms of the vector field f and the curve c. A natural question is about the relation between these two expressions. In fact, once sufficient conditions are imposed on c, the two expressions are equal, and this is the content of Green's theorem. For its precise statement we recall that the boundary $\partial\Omega$ of a domain Ω and its interior Ω^o have been introduced at the end of Section 3.7. We also recall that a closed curve can be composed of a finite family of parametric curves, all of class C^1, see Figure 5.4 and the explanations before this figure.

Theorem 7.1 (Green's theorem). *Let $\Omega \subset \mathbb{R}^2$ be a bounded and closed domain whose boundary consists in a finite number of curves $c_1 \cup c_2 \cup \dots \cup c_N$ defined by parametric curves of class C^1, and such that Ω is on the left of each of these parametric curves. Let $f = (f_1, f_2) : \Omega \to \mathbb{R}^2$ be a continuous vector field of class C^1 on Ω^o, with its derivatives continuous on Ω. Then the following equality holds:*

$$\int_c f = \iint_\Omega [\partial_x f_2 - \partial_y f_1](x, y)\, \mathrm{d}x\, \mathrm{d}y \tag{7.3}$$

with $c = c_1 \cup c_2 \cup \dots \cup c_N$.

Before the proof of this theorem, let us provide a few explanations. Firstly, note that we have imposed the condition that Ω *is on the left of each parametric curve*. This imposes an orientation on the parametric curves. As seen in the previous chapter, not choosing an orientation could lead to a different sign in the result of the line integral.

Secondly, we have imposed that f is defined on the closed domain Ω such that its values on the boundary $\partial\Omega$ are well defined. This is important for the l.h.s. of (7.3). We have also assumed that f is continuously differentiable with derivatives continuous on Ω. This requirement is slightly stronger than $f \in C^1(\Omega^o)$ and is necessary in the proof. Let us make an analogue of this condition in one dimension: the function $x \mapsto x^{1/2}$ is $C^1\big((0, 1)\big)$ but its derivative does not admit an extension on $[0, 1]$. In fact, its derivative is not bounded near 0, and this would generate a problem in the subsequent proof. In fact the conditions that we impose on f imply that $\partial_x f_2$ and $\partial_y f_1$ are bounded on Ω. It would be sufficient to impose that the difference $\partial_x f_2 - \partial_y f_1$ is bounded on Ω, but we do not consider this refinement here ♥.

Proof. We provide the main ingredients of the proof. The proof consists in three steps: 1) we check the statement on a very simple domain, 2) we show how simple domains can be patched together, 3) we cover Ω with such simple elements and conclude.

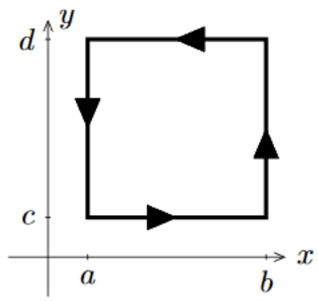


Fig. 7.2. Simplest domain

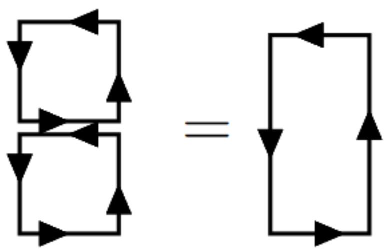


Fig. 7.3. Juxtaposition of 2 domains

We first consider the domain Q defined by Figure 7.2, and provide a parametrization of its boundary. The boundary consists in four segments which can be described by the following four parametric curves:

(i) $c_1 : [0,1] \ni t \mapsto \big(a + t(b-a), c\big) \in \mathbb{R}^2$,

(ii) $c_2 : [0,1] \ni t \mapsto \big(b, c + t(d-c)\big) \in \mathbb{R}^2$,

(iii) $c_3 : [0,1] \ni t \mapsto \big(b + t(a-b), d\big) \in \mathbb{R}^2$,

(iv) $c_4 : [0,1] \ni t \mapsto \big(a, d + t(c-d)\big) \in \mathbb{R}^2$.

If we set $c = c_1 \cup c_2 \cup c_3 \cup c_4$ one easily gets

$$\begin{aligned}
\int_c f &= \sum_{j=1}^4 \int_0^1 f\big(c_j(t)\big) \cdot c_j'(t) \,\mathrm{d}t \\
&= \int_0^1 f_1\big(a + t(b-a), c\big)(b-a) \,\mathrm{d}t + \int_0^1 f_2\big(b, c + t(d-c)\big)(d-c) \,\mathrm{d}t \\
&\quad + \int_0^1 f_1\big(b + t(a-b), d\big)(a-b) \,\mathrm{d}t + \int_0^1 f_2\big(a, d + t(c-d)\big)(c-d) \,\mathrm{d}t \\
&= \int_a^b \big[f_1(x,c) - f_1(x,d)\big] \mathrm{d}x + \int_c^d \big[f_2(b,y) - f_2(a,y)\big] \,\mathrm{d}y. \qquad (7.4)
\end{aligned}$$

On the other hand, one can also compute

$$\begin{aligned}
&\iint_Q [\partial_x f_2 - \partial_y f_1](x,y) \,\mathrm{d}x \,\mathrm{d}y \\
&= \int_c^d \Big[\int_a^b \partial_x f_2(x,y) \,\mathrm{d}x\Big] \mathrm{d}y - \int_a^b \Big[\int_c^d \partial_y f_1(x,y) \,\mathrm{d}y\Big] \mathrm{d}x \\
&= \int_c^d \Big[f_2(b,y) - f_2(a,y)\Big] \mathrm{d}y - \int_a^b \Big[f_1(x,d) - f_1(x,c)\Big] \mathrm{d}x \\
&= \int_a^b \big[f_1(x,c) - f_1(x,d)\big] \mathrm{d}x + \int_c^d \big[f_2(b,y) - f_2(a,y)\big] \mathrm{d}y. \qquad (7.5)
\end{aligned}$$

By comparing (7.4) and (7.5), one observes that the two expressions are the same, which provides a proof of the equality (7.3) in the special case $\Omega = Q$.

The next step consists in combining different squares together. By looking at Figure 7.3 and by recalling that the equality (7.3) holds separately for both squares, one observes that the equality (7.3) also holds for the union of these two squares. Indeed, the r.h.s. of (7.3) is additive, which means that the contributions of the two squares are added. On the other hand, for the l.h.s. one observes that two contributions are going to cancel each others, and therefore the resulting integral corresponds to the integral around the union of the two squares. This patching of squares can be repeated, and one gets that for any domain Ω obtained by patching squares, the equality (7.3) holds. As an example, if Ω corresponds the yellow domain presented in Figure 7.4, then the equality (7.3) holds for this domain.

For the final step, one has to consider smaller and smaller squares, and approximate the surface on the left of Figure 7.4 with a domain like the yellow domain on the right. As mentioned above, the statement holds for the yellow domain. Then, it remains to show that the difference between $\int_c f$ and the integral computed on the boundary of the yellow domain can be made very small, by choosing smaller and smaller squares. Similarly, one has to show that the integral $\iint_\Omega [\partial_x f_2 - \partial_y f_1](x, y)\, \mathrm{d}x\, \mathrm{d}y$ can be well approximated by the similar expression on the yellow domain. For the first approximation, one has to use that the function f is continuous and bounded on Ω, while for the second approximation one uses that the expression $\partial_x f_2 - \partial_y f_1$ is also continuous and bounded on the domain Ω. The details are left as an exercise. $\square$

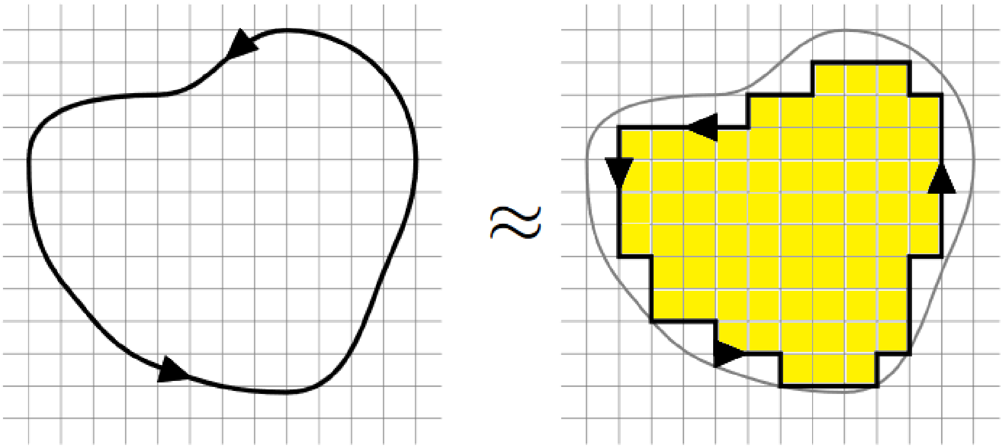


Fig. 7.4. Approximation with small squares

Let us look at two applications of the previous result.

1) Let Ω be the ellipse in $\mathbb{R}^2$ defined by the equation $x^2 + \left(\frac{y}{2}\right)^2 \leq 1$, and consider the vector field $f = (f_1, f_2) : \mathbb{R}^2 \to \mathbb{R}^2$ with $f_1(x, y) = y + 3x$ and $f_2(x, y) = 2y - x$. We consider a counterclockwise parametrization of the curve c defining the boundary of Ω. With this choice, the domain Ω remains on the left of the parameterized curve. Without having to explicitly exhibit such a parametrization, one infers from Green's theorem that:

$$\begin{aligned}\int_c f &= \iint_\Omega [\partial_x f_2 - \partial_y f_1](x, y)\, \mathrm{d}x\, \mathrm{d}y \\ &= -2 \iint_\Omega \mathrm{d}x\, \mathrm{d}y = -2\,\mathrm{Area}(\Omega) = -2\pi \cdot 1 \cdot 2 = -4\pi.\end{aligned}$$

2) Consider now a domain Ω as presented in Figure 7.5. The boundary of Ω consists in the two curves c_1 and c_2, and observe that their orientations have been chosen such that Ω is always on the left of the curves. It means that Green's theorem can be applied in the form

$$\int_c f = \int_{c_1} f + \int_{c_2} f = \iint_\Omega [\partial_x f_2 - \partial_y f_1](x, y)\, \mathrm{d}x\, \mathrm{d}y.$$

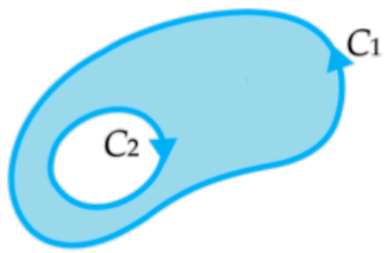


Fig. 7.5. A domain Ω with a hole

Suppose now that we choose a function f satisfying $[\partial_x f_2 - \partial_y f_1](x,y) = 0$ for any $(x,y) \in \Omega$. Then, one gets

$$\int_{c_1} f + \int_{c_2} f = 0 \Longleftrightarrow \int_{c_1} f = -\int_{c_2} f = \int_{c_2^-} f$$

with the notation c_2^- for the opposite curve introduced in Definition 5.5. In particular, if $f = \nabla\phi$ for a function ϕ which is regular enough on the domain Ω, the assumption $\partial_x f_2 = \partial_y f_1$ holds, and therefore the previous equality applies. An important application of this result is presented in the following exercise.

Exercise 7.2. *Consider the vector field $f : \mathbb{R}^2 \setminus \{(0,0)\} \ni (x,y) \mapsto \big(\frac{-y}{x^2+y^2}, \frac{x}{x^2+y^2}\big) \in \mathbb{R}^2$. Compute the curve integral for the following curves:*

(i) The curve defined by the circle centered at $(0,0)$ and of radius $\sqrt{2}$, taken in counterclockwise direction, from $(1,1)$ to $(-\sqrt{2},0)$,

(ii) The curve defined by the unit circle centered at $(0,0)$, taken in counterclockwise direction,

(iii) The curve defined by the circle centered at $(0,0)$ and of radius $r > 0$, taken in counterclockwise direction.

We shall now deduce another theorem from Green's theorem. For that purpose, let $c : [a,b] \to \mathbb{R}^2$ be a parametric curve of class C^1, and recall that the located vector $c'(t)$, once located at the point $c(t)$, is tangent to the curve. This holds for any $t \in (a,b)$. Since $c'(t) = \big(\begin{smallmatrix} c_1'(t) \\ c_2'(t) \end{smallmatrix}\big)$, the located vector $N(t)$ defined by $\big(\begin{smallmatrix} c_2'(t) \\ -c_1'(t) \end{smallmatrix}\big)$ is perpendicular to $c'(t)$, and therefore is perpendicular to the curve at $c(t)$. Let us also mention that if $c : [a,b] \to \mathbb{R}^2$ is closed, then one has $c(a) = c(b)$. With these notations one can state:

Theorem 7.3 (Divergence theorem in $2D$)**.** *Let $\Omega \subset \mathbb{R}^2$ be a bounded and closed domain whose boundary consists in a curve defined by a closed parametric curves $c : [a,b] \to \mathbb{R}^2$ of class C^1, and such that Ω is on the left of this parametric curve. Let $f = (f_1, f_2) : \Omega \to \mathbb{R}^2$ be a continuous vector field of class C^1 on Ω^o, with its derivatives continuous on Ω. Then the following equality holds:*

$$\int_a^b f\big(c(t)\big) \cdot N(t)\,\mathrm{d}t = \iint_\Omega [\mathrm{div} f](x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{7.6}$$

with $\mathrm{div} f = \partial_1 f_1 + \partial_2 f_2$ *the divergence of f.*

Let us already mention that a generalization of this result will take place in higher dimensions. In the more general framework, we shall provide an interpretation of this equality. Note also that we have written ∂_1 for ∂_x and ∂_2 for ∂_y. This is motivated by some generalizations later on.

Proof. Define the function $g : \Omega \to \mathbb{R}^2$ by $g = \binom{-f_2}{f_1}$, or equivalently $g_1 = -f_2$ and $g_2 = f_1$. We can then apply Green's theorem to the function g and infers that

$$\begin{aligned}\int_c g &= \iint_\Omega [\partial_1 g_2 - \partial_2 g_1](x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \iint_\Omega [\partial_1 f_1 + \partial_2 f_2](x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \iint_\Omega [\mathrm{div} f](x,y)\,\mathrm{d}x\,\mathrm{d}y. \end{aligned} \tag{7.7}$$

On the other hand, one also has

$$\begin{aligned}\int_c g &= \int_a^b \big[g_1\big(c(t)\big)c_1'(t) + g_2\big(c(t)\big)c_2'(t)\big]\mathrm{d}t \\ &= \int_a^b \big[-f_2\big(c(t)\big)c_1'(t) + f_1\big(c(t)\big)c_2'(t)\big]\mathrm{d}t \\ &= \int_a^b \begin{pmatrix} f_1(c(t)) \\ f_2(c(t)) \end{pmatrix} \cdot \begin{pmatrix} c_2'(t) \\ -c_1'(t) \end{pmatrix}\mathrm{d}t \\ &= \int_a^b f\big(c(t)\big)\cdot N(t)\,\mathrm{d}t. \end{aligned} \tag{7.8}$$

By comparing (7.7) and (7.8), one directly obtains the statement. $\square$

Chapter 8

Surface integrals

In this chapter, we shall consider surface in $\mathbb{R}^3$ and define integrals on such surface. More precisely, if Ω is an open subset of $\mathbb{R}^2$ and if $q : \Omega \to \mathbb{R}^3$ is regular enough, then this function defines a surface in $\mathbb{R}^3$. Indeed, the map

$$\Omega \ni (s,t) \mapsto \begin{pmatrix} q_1(s,t) \\ q_2(s,t) \\ q_3(s,t) \end{pmatrix} \in \mathbb{R}^3 \tag{8.1}$$

corresponds to a surface in $\mathbb{R}^3$. Note that the simplest example is constructed from a function $g : \Omega \to \mathbb{R}$ by setting

$$\Omega \ni (s,t) \mapsto \begin{pmatrix} s \\ t \\ g(s,t) \end{pmatrix} \in \mathbb{R}^3 \tag{8.2}$$

which correponds to the graph of g.

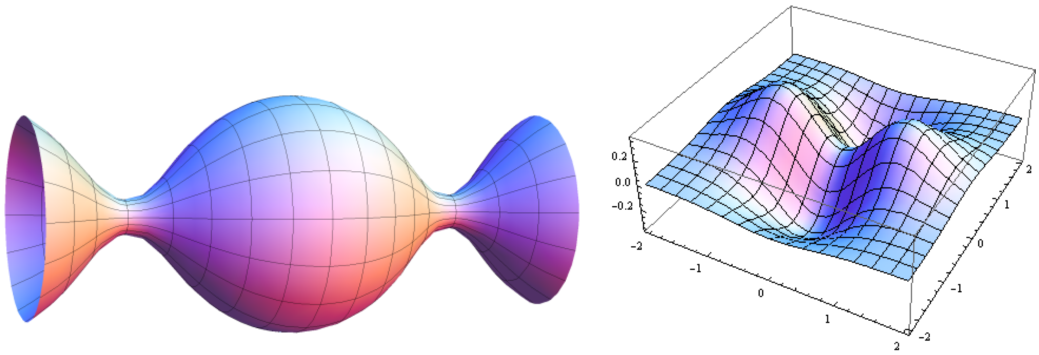


Fig. 8.1. Two surfaces in $\mathbb{R}^3$

The first example of Figure 8.1 corresponds to a general surface in $\mathbb{R}^3$ provided by a function like in (8.1), while the second example of Figure 8.1 corresponds to a surface of the type described in (8.2). Let us now provide a precise definition:

Definition 8.1 (Parametric surface)**.** *For any open set* $\Omega \subset \mathbb{R}^2$*, a map* $q : \Omega \to \mathbb{R}^3$ *with* $q(s,t) = \begin{pmatrix} q_1(s,t) \\ q_2(s,t) \\ q_3(s,t) \end{pmatrix}$ *is called* a parametric surface*. Its image* $q(\Omega)$ *is a* surface in $\mathbb{R}^3$.

Let us illustrate this definition with a few basic examples. The variables s and t will take different names according to the situation. These surfaces are illustrated in Figure 8.2.

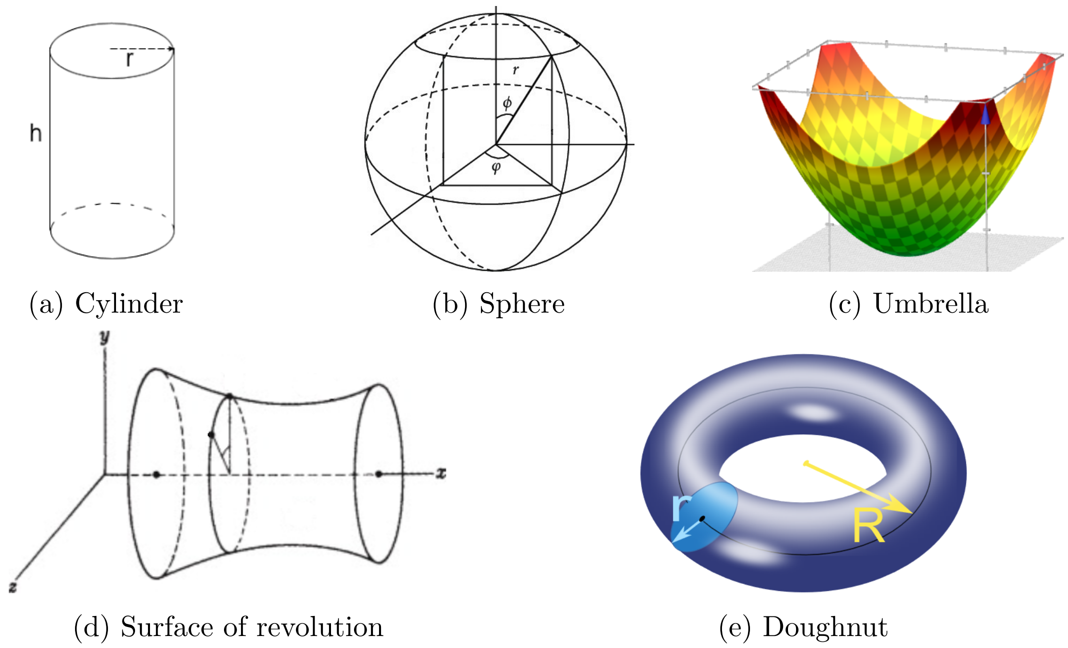
(a) For $r, h > 0$ the function $q : \mathbb{T} \times [0,h] \ni (\theta, z) \mapsto \begin{pmatrix} r\cos(\theta) \\ r\sin(\theta) \\ z \end{pmatrix} \in \mathbb{R}^3$ defines a vertical cylinder in $\mathbb{R}^3$.

(b) For $r > 0$ the function $q : \mathbb{T} \times [0,\pi] \ni (\varphi, \phi) \mapsto \begin{pmatrix} r\sin(\phi)\cos(\varphi) \\ r\sin(\phi)\sin(\varphi) \\ r\cos(\phi) \end{pmatrix} \in \mathbb{R}^3$ corresponds to a sphere of radius r in $\mathbb{R}^3$.

(c) $q : \mathbb{R}^2 \ni (x,y) \mapsto \begin{pmatrix} x \\ y \\ x^2+y^2 \end{pmatrix} \in \mathbb{R}^3$ corresponds to an upside down umbrella in $\mathbb{R}^3$.

(d) If $g : \mathbb{R} \to \mathbb{R}_+$, then $q : \mathbb{T} \times \mathbb{R} \ni (x,\theta) \mapsto \begin{pmatrix} x \\ g(x)\cos(\theta) \\ g(x)\sin(\theta) \end{pmatrix} \in \mathbb{R}^3$ corresponds to a surface of revolution in $\mathbb{R}^3$ along the x-axis.

(e) For $R > r > 0$ the surface $q : \mathbb{T} \times \mathbb{T} \ni (\theta, \varphi) \mapsto \begin{pmatrix} (R+r\cos(\varphi))\cos(\theta) \\ (R+r\cos(\varphi))\sin(\theta) \\ r\sin(\varphi) \end{pmatrix} \in \mathbb{R}^3$ corresponds to a doughnut in $\mathbb{R}^3$.



(a) Cylinder (b) Sphere (c) Umbrella

(d) Surface of revolution (e) Doughnut

Fig. 8.2. Five basic surfaces

We now consider one parametric surface given by a function $q : \Omega \to \mathbb{R}^3$ with $q(s,t) \in \mathbb{R}^3$, and let us freeze one of the variable, for example the variable t. In such a case, we can define the subset $\Omega_t = \{s \in \mathbb{R} \mid (s,t) \in \Omega\}$. A picture of the set Ω_t is provided in Figure 8.3. Then, always for fixed t, the map

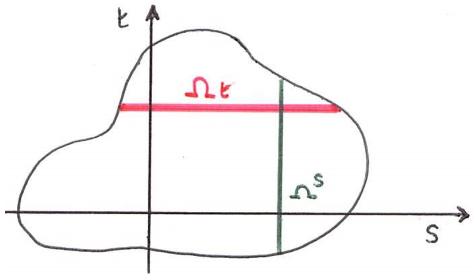


Fig. 8.3. Ω_t and Ω^s

$$\Omega_t \ni s \mapsto \begin{pmatrix} q_1(s,t) \\ q_2(s,t) \\ q_3(s,t) \end{pmatrix} \in \mathbb{R}^3$$

define a curve located on the surface of $q(\Omega)$. Thus, if we consider the derivative of this function with respect to the variable s, namely $[\partial_1 q](s,t)$, and if we locate this vector at the point $q(s,t) \in \mathbb{R}^3$, we get a located vector which is tangent to the surface $q(\Omega)$. Since the same procedure can be applied to the function $\Omega^s \ni t \mapsto q(s,t) \in \mathbb{R}^3$ for any fixed s and with Ω^s provided in Figure 8.3, one gets a second located vector $\partial_2 q(s,t)$ tangent to $q(\Omega)$ at $q(s,t)$. In other terms, one has obtained two located vectors tangent to the surface at $q(s,t)$. More precisely, one has obtained:

Definition 8.2 (Tangent plane)**.** *Let $\Omega \subset \mathbb{R}^2$ be open, and let $q : \Omega \to \mathbb{R}^3$ be of class C^1. The two located vectors $[\partial_1 q](s,t)$ and $[\partial_2 q](s,t)$, once located at $q(s,t)$, are tangent to the surface $q(\Omega)$ and generate a plane called* the tangent plane *at $q(s,t)$.*

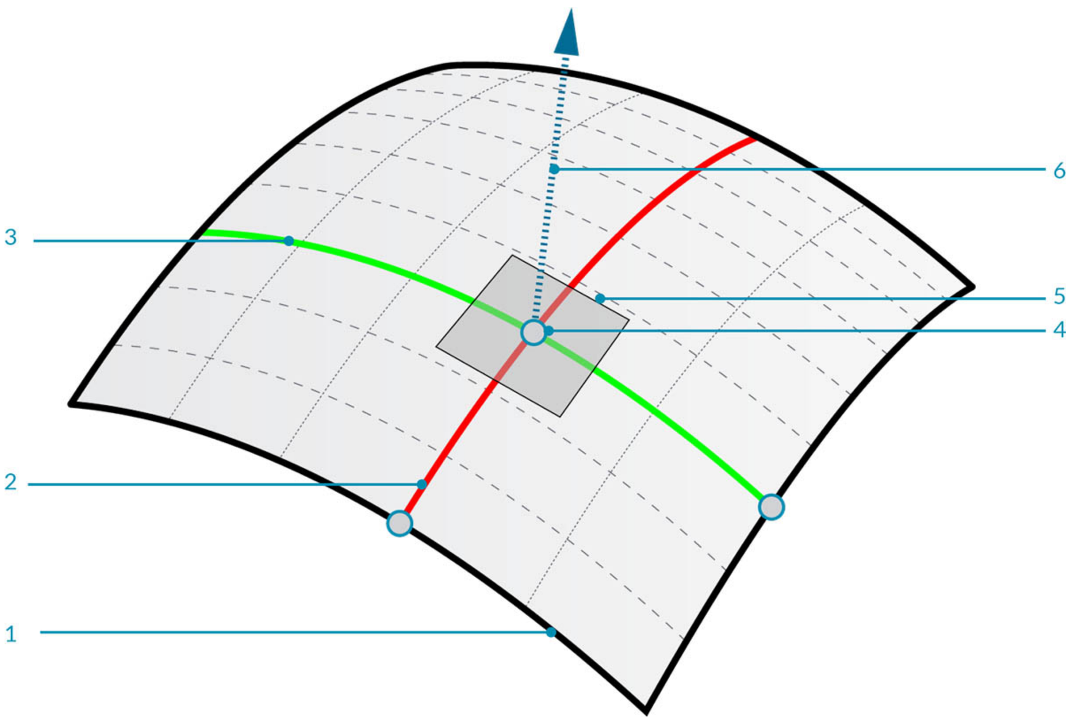


Fig. 8.4. A surface and the tangent plane at one point

A illustration of the tangent plane is provided in Figure 8.4. In this figure, one has: 1) the surface in $\mathbb{R}^3$; 2) the map $s \mapsto q(s,t)$; 3) the map $t \mapsto q(s,t)$; 4) the point $q(s_0,t_0)$, 5) the tangent plane at $q(s_0,t_0)$; the vector normal to the tangent plane at $q(s_0,t_0)$ Let us now provide more information about the normal vector. For that purpose, recall that if $X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3$ and $Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \in \mathbb{R}^3$, are linearly independent vectors, then the cross product $X \times Y$ is given by the expression

$$X \times Y = \begin{pmatrix} x_2y_3 - x_3y_2 \\ x_3y_1 - x_1y_3 \\ x_1y_2 - x_2y_1 \end{pmatrix} \in \mathbb{R}^3.$$

It is known that this vector is perpendicular to X and to Y, and therefore is also perpendicular to the plane generated by X and Y. It follows that the vector $\frac{X \times Y}{\|X \times Y\|}$ is perpendicular to this plane and is a vector of norm 1.

Let us now use this information for the parametric surface defined by $q : \Omega \to \mathbb{R}^3$. If q is of class C^1 and if $[\partial_1 q](s,t)$ and $[\partial_2 q](s,t)$ are linearly independent vector (which we shall assume in the sequel), then the vector

$$N_q(s,t) := [\partial_1 q \times \partial_2 q](s,t) \tag{8.3}$$

defines a located vector which is perpendicular for the surface $q(\Omega)$ at the point $q(s,t)$. This located vector plays a central role in the subsequent type of integrals.

Definition 8.3 (Surface area). *Let $\Omega \subset \mathbb{R}^2$ be open and bounded, and let $q : \Omega \to \mathbb{R}^3$ be a parametric surface of class C^1 (and injective). Then we set*

$$\iint_{q(\Omega)} \mathrm{d}\sigma := \iint_\Omega \|N_q(s,t)\| \,\mathrm{d}s\,\mathrm{d}t \tag{8.4}$$

and call this quantity the area of the surface $q(\Omega)$.

In Figure 8.5 the area of each small element of surface is proportional to $\|N_q(s,t)\|$. The total area is obtained by summing these small contributions, and this procedure directly leads to (8.4).

Observe first that in the very special case $q(s,t) = \begin{pmatrix} s \\ t \\ 0 \end{pmatrix}$, then $\|N_q(s,t)\| = 1$ and the quantity $\iint_\Omega 1 \ \mathrm{d}s \ \mathrm{d}t$ corresponds to the area of the flat surface Ω in $\mathbb{R}^3$. In the above definition we have also assumed that q is injective, meaning that $q(s,t) \neq q(s',t')$ whenever $(s,t) \neq (s',t')$. This is indeed a natural requirement. For example, if we consider the surface given by

$$(0,4\pi) \times (0,h) \ni (\theta,z) \mapsto q(\theta,z) = \big(\cos(\theta),\sin(\theta),z\big) \in \mathbb{R}^3$$

which is a cylinder in $\mathbb{R}^3$. The function q is not injective, and the image of q corresponds to the cylinder, but *covered twice*. Thus, its surface is $2\pi h$, but if we apply the above formula we would find $4\pi h$ because we would compute twice the surface. This is not really meaningful, and for this reason we ask the function to be injective. Note however that we could weaken this condition by allowing intersection of surface 0 but avoiding overlap (of surface different from 0).

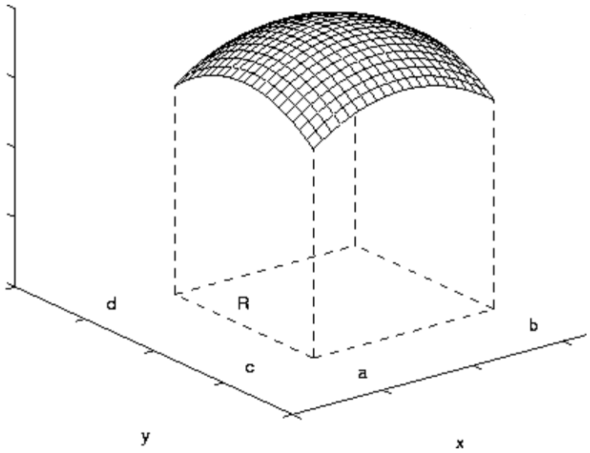


Fig. 8.5. A surface with the small elements represented

Remark 8.4. *It has been proved in Proposition 2.7 that the length of a curve is independent of its parametrization. More precisely, the same length is obtained for the curve c and the curve $c \circ \varphi$ if φ is a suitable diffeomorphism. It turns out that a similar property holds for the area of a surface, namely the area of a surface defined by q and the area of the surface obtained by $q \circ \varphi$, for a suitable diffeomorphism φ, are the same Since the proof is slightly more complicated, and use the special form of the vector N_q, we do not provide it here.*

Let us now look at one example based on Example 6.10. Consider the function

$$q : \mathbb{T} \times [0, \pi) \ni (\theta, \varphi) \mapsto q(\theta, \varphi) \in \mathbb{R}^3$$

with $q(\theta, \varphi) = \big(r\cos(\theta)\sin(\varphi), r\sin(\theta)\sin(\varphi), r\cos(\varphi)\big)$. This function describes a sphere in $\mathbb{R}^3$ of radius $r > 0$. The quantity $\|N_q(\theta, \varphi)\|$ can easily be computed, and one finds $\|N_q(\theta, \varphi)\| = r^2 \sin(\varphi)$. One then infers the known result:

$$\iint_{q(\Omega)} \mathrm{d}\sigma = \iint_{\mathbb{T}\times[0,\pi)} r^2 \sin(\varphi)\,\mathrm{d}\theta\,\mathrm{d}\varphi = 4\pi r^2$$

which is indeed the surface of a sphere of radius r.

In addition to the surface defined by $q : \Omega \to \mathbb{R}^3$, consider now another function defined on the surface $q(\Omega)$, namely consider a function $f : q(\Omega) \to \mathbb{R}$. Clearly, if f is defined on all $\mathbb{R}^3$, it is also defined on the surface $q(\Omega)$, but it is not necessary that the function f is defined outside of $q(\Omega)$. Then we can introduce one more natural quantity:

Definition 8.5 (Integral on a surface)**.** *Let $\Omega \subset \mathbb{R}^2$ be open and bounded, and let $q : \Omega \to \mathbb{R}^3$ be a parametric surface of class C^1 (and injective). Consider $f : q(\Omega) \to \mathbb{R}$ with the assumption that the composed map $f \circ q : \Omega \to \mathbb{R}$ is continuous. Then,* the integral of f over the surface $q(\Omega)$ *is defined by*

$$\iint_{q(\Omega)} f\,\mathrm{d}\sigma := \iint_{\Omega} f\big(q(s,t)\big)\|N_q(s,t)\|\,\mathrm{d}s\,\mathrm{d}t. \tag{8.5}$$

One immediately observes that the continuity condition is precisely the necessary one for having the integral of a continuous function. Also, a special case of this integral is provided by the computation of the area of $q(\Omega)$ provided in Definition 8.3. Indeed, it corresponds to the special choice $f = 1$. In that respect, one can consider the function f as a local density of the surface $q(\Omega)$, and then the integral $\iint_{q(\Omega)} f \,\mathrm{d}\sigma$ can be interpreted as the total mass of $q(\Omega)$. Of course, other interpretations are possible, depending on the nature of q.

In the previous construction, the function f was scalar valued, which means that it takes values in $\mathbb{R}$. What can one do if f takes values in $\mathbb{R}^3$. Of course, one could define $\iint_{q(\Omega)} f \,\mathrm{d}\sigma$ a *vector-valued integral*, but this would simply be three independent integrals of the type provided in (8.5). In the case f takes values in $\mathbb{R}^3$ and since N_q provides also a vector in $\mathbb{R}^3$, one can certainly construct something more interesting. ☺

Definition 8.6 (Flux through a surface)**.** *Let $\Omega \subset \mathbb{R}^2$ be open and bounded, and let $q : \Omega \to \mathbb{R}^3$ be a parametric surface of class C^1 (and injective). Let $f : q(\Omega) \to \mathbb{R}^3$ with the assumption that the composed map $f \circ q : \Omega \to \mathbb{R}^3$ is continuous. Then,* the flux of f through the surface $q(\Omega)$ *is defined by*

$$\begin{aligned} \iint_{q(\Omega)} f \cdot N_q \,\mathrm{d}\sigma &:= \iint_{\Omega} f\big(q(s,t)\big) \cdot N_q(s,t) \,\mathrm{d}s \,\mathrm{d}t \\ &= \iint_{\Omega} f\big(q(s,t)\big) \cdot [\partial_1 q \times \partial_2 q](s,t) \,\mathrm{d}s \,\mathrm{d}t. \end{aligned} \tag{8.6}$$

A representation of the framework for this integral is provided in Figure 8.6. Note also that this integral will play a crucial role in the last chapter of this course.

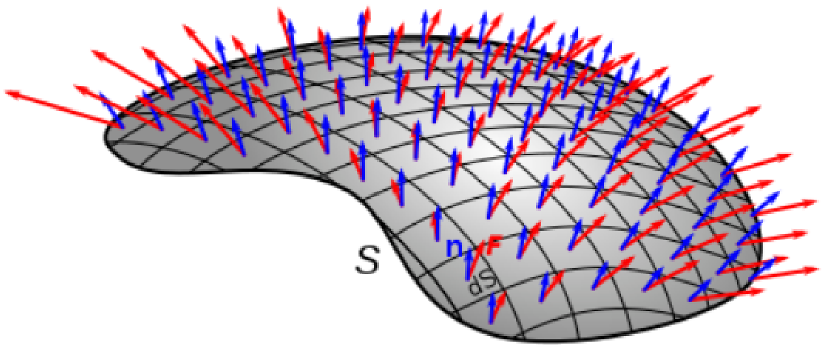


Fig. 8.6. A surface $S = q(\Omega)$ with N_q in blue and f in red

Chapter 9

Divergence and Stokes' theorems

Green's theorem and the divergence theorem seen in theorems 7.1 and 7.3 were taking place in $\mathbb{R}^2$. In this chapter we look for similar results in $\mathbb{R}^3$. These results are called Stokes' theorem and divergence theorem. We start with divergence theorem.

Recall that in $\mathbb{R}^2$ this statement is an equality of the form

$$\int_a^b f\big(c(t)\big) \cdot N(t)\,\mathrm{d}t = \iint_\Omega [\mathrm{div} f](x,y)\;\mathrm{d}x\,\mathrm{d}y$$

where $f : \Omega \to \mathbb{R}^2$ is a vector field. If we consider now a vector field $f : \mathbb{R}^3 \to \mathbb{R}^3$, the r.h.s. of this equality is still meaningful. Indeed, the divergence can be defined on a vector field of any dimension by the formula: $\mathrm{div} f = \sum_{j=1}^n \partial_j f_j$, with $\{f_j\}$ the components of f. Now, what would be the replacement of the l.h.s. ? The most natural choice is to replace the line integral by the integral of a flux through a surface, as defined in (8.6). Note however that we have to change our notations: the domain Ω can not be both a bounded volume in $\mathbb{R}^3$, and a subset of $\mathbb{R}^2$ used for parameterizing the surface enclosing the volume of $\mathbb{R}^3$. The next statement takes these necessary changes into account.

Theorem 9.1 (Divergence theorem in $\mathbb{R}^3$). *Let $\mathcal{U}$ be a bounded and closed domain in $\mathbb{R}^3$ whose boundary S consists in the union of a finite number of surfaces which are defined by parametric surfaces $S = S_1 \cup S_2 \cdots \cup S_n$ of class C^1 (and such that the normal vector N_q to these surface points outside of $\mathcal{U}$). Let $f : \mathcal{U} \to \mathbb{R}^3$ be a continuous vector field of class C^1 on $\mathcal{U}^o$, with its derivatives continuous on $\mathcal{U}$. Then the following equality holds:*

$$\iint_S f \cdot N_q\,\mathrm{d}\sigma = \iiint_{\mathcal{U}} [\mathrm{div} f]\;\mathrm{d}V. \tag{9.1}$$

In the previous equality, the r.h.s. corresponds to a volume integral while the l.h.s corresponds to the flux of f through the surface S. If this surface is made of n pieces, then one has to perform the sum over each contribution. It is important that the parametrization of each part of the surface defines a normal vector N_q which points outside of the surface. As a result, one computes the total flux of f leaving the volume $\mathcal{U}$.

We do not provide the proof of this theorem, but let us mention the key point. The process is similar to the proof of Green's theorem: one first considers elementary volumes, proves the statement for them, and then approximates $\mathcal{U}$ by these elementary volumes, see Figure 9.1. Here, the elementary volumes are small cubes. It is indeed easy to prove equality (9.1) on them, especially when the faces of these cubes are chosen parallel to the planes $x = 0$, or $y = 0$, or $z = 0$. For such faces the normal vectors N_q are quite easy to compute (and can even be guessed *a priori*). Once the statement is proved on these cubes, their juxtaposition leads to the cancellation of some contributions, and one easily realizes that the statement still holds for any volume made of juxtapositions of small cubes. It finally remains to approximate $\mathcal{U}$ by such a volume, and see that the errors between the computations realized on $\mathcal{U}$ and the computations performed on this special volume can be made as small as necessary. We leave the details for the interested reader.

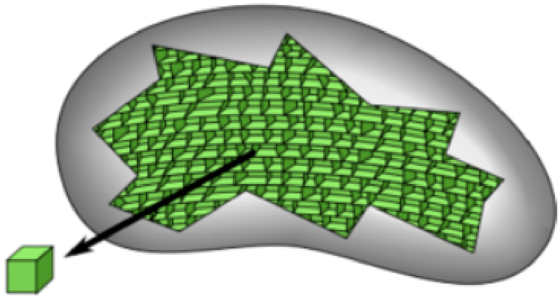


Fig. 9.1. A partition with small cubes

Let us now present one application of this theorem. We consider first the function $g : \mathbb{R}^3 \setminus \{\mathbf{0}\} \to \mathbb{R}$ given by $g(x, y, z) = \frac{q}{4\pi} \frac{1}{(x^2+y^2+z^2)^{1/2}}$, where q is a constant. We also define vector field $f : \mathbb{R}^3 \setminus \{\mathbf{0}\} \to \mathbb{R}^3$ given by $f = -\nabla g$, which gives

$$f(x, y, z) = \frac{q}{4\pi} \frac{1}{(x^2 + y^2 + z^2)^{3/2}} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

It is clear that this vector field is differentiable an arbitrary number of times on $\mathbb{R}^3 \setminus \{\mathbf{0}\}$. In addition, one can readily check that

$$\operatorname{div} f = \operatorname{div} \nabla g = 0 \quad \text{on } \mathbb{R}^3 \setminus \{\mathbf{0}\}. \tag{9.2}$$

It then follows from (9.2) and from the divergence theorem that $\iint_S f \cdot N_q \, \mathrm{d}\sigma = 0$ for any closed surface S which does not enclose $\mathbf{0}$. On the other hand, one also infers that for any surface S enclosing $\mathbf{0}$ one has $\iint_S f \cdot N_q \, \mathrm{d}\sigma = \text{cst}$, for a constant which has to be determined. The easiest example on which this computation can be performed is on a unit sphere centered at $\mathbf{0}$. Indeed, by parameterizing such a sphere with

$$q : \mathbb{T} \times [0, \pi) \ni (\theta, \varphi) \mapsto \begin{pmatrix} \cos(\theta) \sin(\varphi) \\ \sin(\theta) \sin(\varphi) \\ \cos(\varphi) \end{pmatrix}$$

one gets (after a few computations) that

$$N_q(\theta, \varphi) = \begin{pmatrix} -\cos(\theta) \sin^2(\varphi) \\ -\sin(\theta) \sin^2(\varphi) \\ -\sin(\varphi) \cos(\varphi) \end{pmatrix} = -\sin(\varphi) q(\theta, \varphi),$$

see also Example 6.10. Note that in order to have a vector pointing outside of the surface, we shall have to change the sign of the above expression. Thus one gets:

$$\begin{aligned}
&\iint_{\text{unit sphere}} f \cdot N_q \,\mathrm{d}\sigma \\
&= \iint_{\mathbb{T}\times[0,\pi)} f\big(\cos(\theta)\sin(\varphi), \sin(\theta)\sin(\varphi), \cos(\varphi)\big) \cdot \sin(\varphi) q(\theta,\varphi)\,\mathrm{d}\theta\,\mathrm{d}\varphi \\
&= \iint_{\mathbb{T}\times[0,\pi)} \frac{q}{4\pi} q(\theta,\varphi) \cdot \sin(\varphi) q(\theta,\varphi)\,\mathrm{d}\theta\,\mathrm{d}\varphi \\
&= \frac{q}{4\pi} \iint_{\mathbb{T}\times[0,\pi)} \sin(\varphi)\,\mathrm{d}\theta\,\mathrm{d}\varphi \\
&= q,
\end{aligned}$$

where we have used that $q(\theta,\varphi)$ belongs to the unit sphere, and therefore has norm 1. Observe that this result is nothing but Gauss's law in electromagnetism.

The next and final result is an extension of Green's theorem, called Stokes' theorem. It consists in looking at Green's theorem in $\mathbb{R}^3$ instead of $\mathbb{R}^2$. Recall that Green's theorem is an equality of the form

$$\int_c f = \iint_\Omega [\partial_x f_2 - \partial_y f_1](x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

Clearly, the l.h.s. is well defined even if the curve is not located on a plane. Thus, for any curve in $\mathbb{R}^3$ of class C^1 and for any continuous vector field f in $\mathbb{R}^3$, the curve integral can be defined. For the r.h.s. it has already been mentioned that the integral corresponds to $\mathrm{curl}_3 f$. Thus, it would be natural to use the notion of $\mathrm{curl} f$, namely

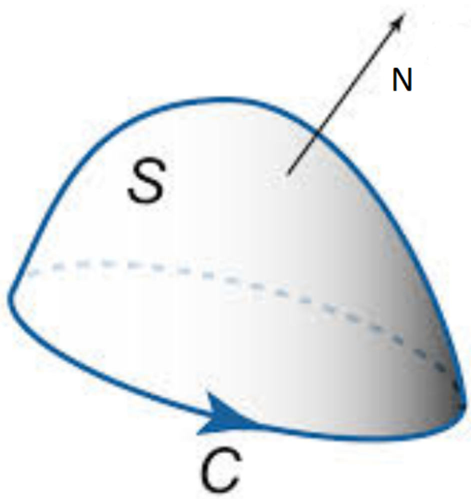


Fig. 9.2. Surface with boundary

$$\mathrm{curl} f = {}^t\big(\partial_y f_3 - \partial_z f_2, \partial_z f_1 - \partial_x f_3, \partial_x f_2 - \partial_y f_1\big).$$

Since $\mathrm{curl} f$ is a vector field, we can use it and consider a surface integral of the form proposed in (8.6). The surface considered has a boundary, and this boundary corresponds precisely to the curve used for the curve integral. A representation of the setting is provided in figure 9.2. Observe that if the surface is defined by a parametric surface $q : \Omega \to \mathbb{R}^3$ with $\Omega \subset \mathbb{R}^2$, and if the boundary of $q(\Omega)$ is a curve which corresponds to a parametric curve of class C^1, then one parametrization of this curve can be provided by q applied to the boundary $\partial\Omega$ of Ω. We can now state this theorem (the notion of orientability which appears in the statement will be discussed later on).

Theorem 9.2 (Stokes' theorem)**.** *Let $\Omega \subset \mathbb{R}^2$ be open and let $q : \Omega \to \mathbb{R}^3$ be a parametric surface in $\mathbb{R}^3$ of class C^1. Assume that the boundary $\partial q(\Omega)$ of $q(\Omega)$ consists* *in a curve which can be defined by a parametric curve* $c : [a,b] \to \mathbb{R}^3$ *of class* C^1. *Assume also that the surface* $q(\Omega)$ *is orientable, and that the surface lies on the left of the curve. Let* $f : \mathbb{R}^3 \to \mathbb{R}^3$ *be a vector field of class* C^1. *Then the following equality holds:*

$$\iint_{q(\Omega)} (\mathrm{curl} f) \cdot N_q \,\mathrm{d}\sigma = \int_c f. \tag{9.3}$$

Before discussing the above statement, let us write the equality with more details, namely

$$\iint_\Omega \big[\mathrm{curl} f\big]\big(q(s,t)\big) \cdot [\partial_1 q \times \partial_2 q](s,t)\,\mathrm{d}s\,\mathrm{d}t = \int_a^b f\big(c(\zeta)\big) \cdot c'(\zeta)\,\mathrm{d}\zeta. \tag{9.4}$$

Observe that we have used the variable ζ (instead of t) for the curve integral since the variable t is already used in the l.h.s. Observe also that our assumptions on f are in fact too strong. It is sufficient that f is defined on a small open set containing the surface $q(\Omega)$ such that the derivative of f is well defined. Accordingly, it is sufficient to require that the map $(s,t) \mapsto \big[\mathrm{curl} f\big]\big(q(s,t)\big)$ is continuous.

By looking at (9.3) it is clear that Stokes' theorem relates an integral on a surface with an integral on the boundary of this surface, which is a curve. We shall not prove it here, since the proof would take a long time, but also since a much shorter proof exists in a more advanced framework. Just for information, let us mention the formulation of Stokes' theorem in this more advanced framework. It can be summarized in *the integral of a differential form* ω *over the boundary of some orientable surface* S *is equal to the integral of its exterior derivative* $\mathrm{d}\omega$ *over the whole of* S, namely $\int_{\partial S} \omega = \int_S \mathrm{d}\omega$. Clearly, there is still a long path before one can fully understand such a statement.

Let us now consider a special application of Theorem 9.2. If $\Omega \subset \mathbb{R}^2$ is open and if we consider the parametric surface given by $q : \Omega \ni (s,t) \mapsto \left(\begin{smallmatrix} s \\ t \\ 0 \end{smallmatrix}\right)$, then $N_q(s,t) = \left(\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix}\right)$, and for any $f : \Omega \to \mathbb{R}^3$ of class C^1 with $f_3 = 0$ one has

$$\begin{aligned} &\iint_\Omega \big[\mathrm{curl} f\big]\big(q(s,t)\big) \cdot [\partial_1 q \times \partial_2 q](s,t)\,\mathrm{d}s\,\mathrm{d}t \qquad (9.5)\\ &= \iint_\Omega \big[\mathrm{curl} f\big](s,t,0) \cdot \left(\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix}\right) \mathrm{d}s\,\mathrm{d}t \\ &= \iint_\Omega \big[\partial_1 f_2 - \partial_2 f_1\big](s,t,0)\,\mathrm{d}s\,\mathrm{d}t. \end{aligned}$$

On the other hand, if $c : [a,b] \to \mathbb{R}^3$ is a parametric curve defining the boundary of Ω, one has $c_3(\zeta) = 0$, namely the third component of $c(\zeta)$ is 0 because Ω is located in the plane $z = 0$. And by Stokes' theorem, the expression (9.5) is also equal to

$$\begin{aligned} &\int_a^b f\big(c(\zeta)\big) \cdot c'(\zeta)\,\mathrm{d}\zeta \\ &= \int_a^b \begin{pmatrix} f_1(c_1(\zeta), c_2(\zeta), 0) \\ f_2(c_1(\zeta), c_2(\zeta), 0) \\ 0 \end{pmatrix} \cdot \begin{pmatrix} c_1'(\zeta) \\ c_2'(\zeta) \\ 0 \end{pmatrix} \mathrm{d}\zeta. \end{aligned}$$

Thus, if we write $\tilde{f}(x,y)$ for $f(x,y,0)$ for any $x,y \in \mathbb{R}$, and define $\tilde{c} := \left(\begin{smallmatrix} c_1 \\ c_2 \end{smallmatrix}\right)$, then one has obtained the equality

$$\iint_\Omega \big[\partial_1 \tilde{f}_2 - \partial_2 \tilde{f}_1\big](s,t)\,\mathrm{d}s\,\mathrm{d}t = \int_a^b \tilde{f}\big(\tilde{c}(\zeta)\big) \cdot \tilde{c}'(\zeta)\,\mathrm{d}\zeta$$

which is nothing but Green's theorem, as stated in Theorem 7.1. As a summary, it means that Green's theorem corresponds to a special case of Stokes' theorem, or reciprocally it means that Stokes' theorem is an extension of Green's theorem. Note that this was precisely the starting point of our discussion on Stokes's theorem. In that respect, it is not surprising that one proof of Stokes' theorem is based on Figure 9.3. The idea is similar to the proof of Green's theorem seen earlier.

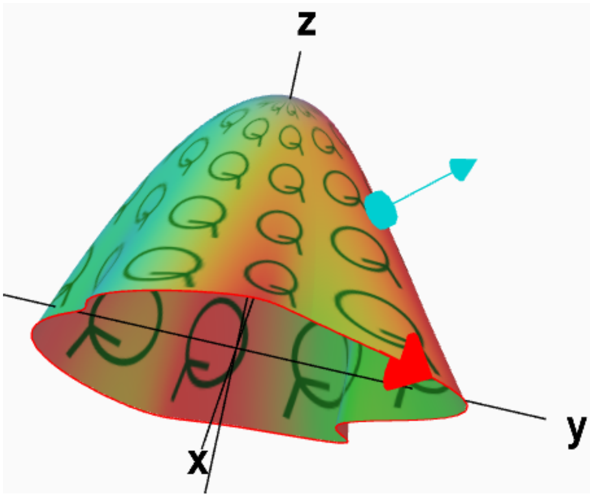


Fig. 9.3. The key idea for a proof of Stokes' theorem

It only remains to discuss the notion of *orientability*. Like the differential forms mentioned earlier, this notion can not be fully defined in this course, and understanding it should be a good motivation for further study. In fact, most of the common surfaces are orientable, but famous exceptions like the Möbius strip exist. Let us just mention the main idea behind orientability. Consider a parametetric surface $q : \Omega \to \mathbb{R}^3$ and the normal vector N_q introduced in (8.3) Consider also an arbitrary continuous and closed curve $c : [a,b] \to q(\Omega)$. One looks at the continuous map

$$[a,b] \ni t \mapsto N_q\big(c(t)\big) \in \mathbb{R}^3.$$

If $N_q\big(c(a)\big) = N_q\big(c(b)\big)$, then the surface is orientable. On the other hand, if for some curve one has $N_q\big(c(a)\big) = -N_q\big(c(b)\big)$, then one says that the surface is not orientable. One example is provided in Figure 9.4.

As a conclusion, let us observe one surprising consequence of Stokes' theorem. Consider first a parametric surface $q : \Omega \to \mathbb{R}^3$ and its boundary $\partial q(\Omega)$. Let also its boundary be parameterized by a curve c. The r.h.s. of (9.4) depends only on c and not on the

Fig. 9.4. A Möbius strip

parametric surface. Thus, if one deforms the surface by keeping its boundary constant, the surface on which the computation of the l.h.s. of (9.4) is modified, but nevertheless the result of the computation is constant (since the r.h.s. is not modified). This idea is illustrated with the three surfaces of Figure 9.5 having the same boundary. A nice result, among so many others !

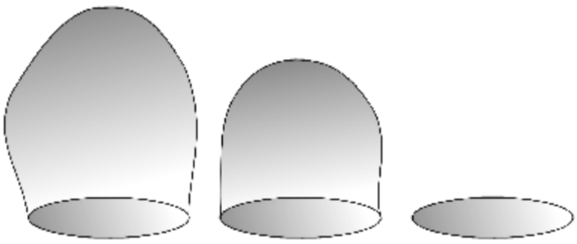


Fig. 9.5. Three surfaces with the same boundary

Bibliography

[1] S. Lang, *Calculus of several variables*, Addison-Wesley Publishing Company, 1973.

[2] https://commons.wikimedia.org/wiki/File:Three-dimensional_graph.png

[3] https://en.wikipedia.org/wiki/Himmelblau's_function

[4] https://www.antiquemapsandprints.com

[5] https://math.libretexts.org/Bookshelves/Calculus/Map%3A_University_Calculus_(Hass_et_al.)/13%3A_Partial_Derivatives/13.7%3A_Extreme_Values_and_Saddle_Points

[6] https://commons.wikimedia.org/wiki/File:Riemann_sum_convergence.png

[7] https://openstax.org/books/calculus-volume-3/pages/5-1-double-integrals-over-rectangular-regions

[8] https://www.fiberoptics4sale.com/blogs/math-for-physics/102498182-surface-integral

[9] F.S. Nelson, *A user-friendly introduction to Lebesgue measure and integration*, AMS, 2015.