# A Précis of Functional Analysis for Engineers

## DRAFT – NOT FOR DISTRIBUTION

Jean-François Hiller and Klaus-Jürgen Bathe

August 29, 2002

# 1 Introduction

The purpose of this précis is to review some classical results regarding Lebesgue integration theory, Sobolev spaces, the theory of distributions and the relevance of these theories in engineering analysis. The results are given in the context where they are most frequently encountered in engineering analysis, and in particular we focus on real valued functions.

With these objectives in mind, we first give some background on topology and vector spaces as well as inner products, and review briefly notions of functional analysis such as derivation in the sense of calculus and Riemann integration.

Mostly, we restrict ourselves to the study of vector spaces that use the real set as their scalar field. In mechanical engineering, reals are used almost exclusively, thus our choice to focus on this scalar field.

We contrast the results obtained employing these theories with results from the more classical Riemann integration theory and functional analysis.

# 2 Basic notions of topology

In this section, we recall some very basic notions of topology[1] that are necessary in our work.

**Definition of a set** We loosely define a **set** as an entity made of the gathering of individual elements. A set can be empty (if it does not contain any element); a set can be made of a finite number of elements (for instance $\{0; 1; 2\}$ is a set of three integers) or an infinite number of elements (for instance $[0; 1]$ is the set of all real numbers $x$ that verify $0 \leq x \leq 1$). We assume in this write-up that the reader is familiar with the basic concepts and notations employed in set theory, such as intersection (denoted by $\cap$), union (denoted by $\cup$), subsets (denoted by $\subset$), belonging to (denoted by $\in$), "there exists" (denoted by $\exists$), uniqueness (denoted by !), excluding (denoted by / ) etc.
We recall the standard notations for some particular sets: $\mathbb{N}$ denotes the set of positive integers, $\mathbb{Z}$ denotes the set of all integers, $\mathbb{Q}$ denotes the set of all rational numbers[2] , $\mathbb{R}$ denotes the set of all real numbers, $\mathbb{C}$ denotes the set of complex numbers. The standard modifiers are also employed: $\mathbb{R}_+$ denotes the set of real positive numbers, and $\mathbb{R}_+^*$ denotes the set of real strictly (i.e. excluding zero) positive numbers, etc.

**Definition of the product of sets and of an equivalence relation** Let us consider a set $A$ and $\Gamma$ a part[3] of $A \times A$. Given an element $(a, b)$ of $\Gamma$, we use the notation $a\mathcal{R}b$. We say that $\Gamma$ defines an equivalence relation if the following three conditions are satisfied:

- $\mathcal{R}$ is **reflexive**: for any $a \in A$, $a\mathcal{R}a$

[1] Topology is the branch of mathematics concerned with 'the study of the properties of geometric figures or solids that are not changed by homeomorphisms, such as stretching or bending' (American Heritage Dictionary of the English Language). Notions of topology are indispensable to discuss such notions as convergence, continuity, etc...

[2] rational numbers are real numbers that can be expressed in the form $a/b$ with $a$ and $b$ integers (with $b$ non-zero). Other real numbers are called irrationals. Examples of rationals are 1, 1.25, 0.3333.... Examples of irrationals are $\pi$, and $e \approx 2.71828$.

[3] $A^2$ denotes $A \times A$, where for any two sets $A$ and $B$, $A \times B$ denotes the set of all pairs of elements such that the first element is in A and the second element is in $B$. Following the same pattern, we define $A^3 = A \times A \times A$, $A^3 \times B^2 = A \times A \times A \times B \times B$, etc.

- $\mathcal{R}$ is **symmetric**: for any $(a, b) \in A \times A$, $a\mathcal{R}b$ implies $b\mathcal{R}a$
- $\mathcal{R}$ is **transitive** : for any $(a, b, c) \in A^3$, ($a\mathcal{R}b$ and $b\mathcal{R}c$) implies $a\mathcal{R}c$.

Considering one element $a \in A$, we denote by $[a]$ the set of all the elements $b \in A$ that verify $a\mathcal{R}b$. $[a]$ is called the equivalence class of $a$.

**Definition of an application** Given two sets $A$ and $B$, and a part $\Gamma$ of $A \times B$ such that for any $x$ in $A$ there exists one and only one $y$ in $F$ such that the pair $(x, y)$ is in $\Gamma$. Then $(\Gamma, A, B)$ is called an **application** from $A$ to $B$ of graph $\Gamma$. The application is said to be defined on $A$ with values in $B$.

**Definition of a metric** We call a **metric** over a set $A$ an application from $A^2$ into $\mathbb{R}_+$ such that

$$\forall (x, y, z) \in A^3 \begin{cases} d(x, y) \geq 0 \\ d(x, y) = d(y, x) \\ d(x, z) \leq d(x, y) + d(y, z) \\ d(x, y) = 0 \iff x = y \end{cases} \tag{1}$$

Note: A metric $d(x, y)$ is the formal definition of the idea of distance: the for properties of a metric are, in substance:

- The distance between two points $x$ and $y$ is always positive.
- The distance from $x$ to $y$ is the same as the distance from $y$ to $x$.
- The shortest way between $x$ and $z$ is in a straight line.
- The distance between $x$ and $y$ is zero if and only if $x$ and $y$ are the same.

**Definition of a metric space** The pair $(A, d)$ where $A$ is a set and $d$ is a metric over that set is called a **metric space**.

**Definition of the Euclidian metric** The Euclidian metric on the space $\mathbb{R}^n$ is defined by

$$d((x_1, x_2, ..., x_n), (y_1, y_2, ..., y_n)) = \sqrt{\sum_{i=1}^{i=n} (x_i - y_i)^2} \tag{2}$$

It can be easily checked that Equation 2 indeed defines a metric, as defined by Equation 1.

**Note** Frequently, when the metric considered is obvious, it is said that $A$ is a metric space.

**Example 2.1** *The set of real numbers equipped with $d(x, y) = |x - y|$ is a metric space. The set or complex numbers equipped with the Euclidian metric*

$$d(x_1 + ix_2, y_1 + iy_2) = ((x_1 - y_1)^2 + (x_2 - y_2)^2)^{1/2} \tag{3}$$

*is a metric space.*

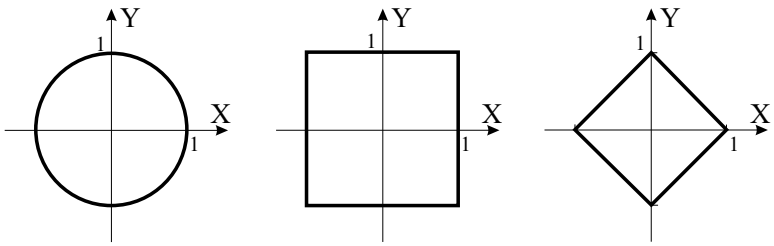



Figure 1: The balls $\mathcal{B}(0,1)$ corresponding to the metrics $d((x_o,y_o),(x_1,y_1))$ defined by [5] $\sqrt{(x_1-x_0)^2+(y_1-y_0)^2}$ , $max\{(x_1-x_0),(y_1,y_0)\}$ and $|(x_1-x_0)+(y_1-y_0)|$ are the areas inside the curves above (from left to right). Closed balls include their boundary, open balls do not.

**Definition of an open ball** We consider a subset $B \subset A$ of a metric space $(A,d)$, and an element (or "point") $x \in B$. We call **open ball** of radius $r$ and center $x$ the subset of $A$ defined by

$$\mathcal{B}(x,r) = \{y \in A, d(x,y) < r\} \tag{4}$$

See Figure 1.

**Definition of boundedness** A metric space $(A,d)$ is **bounded** if there exists an open ball that includes $A$.

**Definition of a closed ball** We consider a subset $B \subset A$ of a metric space $(A,d)$, and an element (or "point") $x \in B$. We call **closed ball** of radius $r$ and center $x$ the subset of $A$ defined by

$$\mathcal{B}(x,r) = \{y \in A, d(x,y) \leq r\} \tag{5}$$

Note that the notation $\bar{\mathcal{B}}(x,r)$ is sometimes also used for a closed ball to distinguish from an open ball.

**Definition of openness** A subset $B$ of a metric space $(A,d)$ is said to be **open** if for all $x \in B$ there exists an open ball of center $x$ that is a subset of $B$.

**Example 2.2** *In the metric space* $\mathbb{R}$ *(with its natural metric mentioned earlier), the set* $]a,b[$ *(where* $a < b$*) is an open set. Note that the sets* $[a,b]$, $[a,b[$ *and* $]a,b]$ *and the singleton* $\{a\}$ *are not open sets.*

**Definition of closedness** A subset $B$ of a metric space $(A,d)$ is said to be **closed** if its complementary in $A$ is open.

**Example 2.3** *In the metric space* $\mathbb{R}$*, the set* $[a,b]$ *(where* $a < b$*) and the singleton* $\{a\}$ *are closed sets. Note that the sets* $]a,b[$, $[a,b[$ *and* $]a,b]$ *are not closed sets.*

**Note 1** As shown by the previous examples, some sets are neither open nor closed.

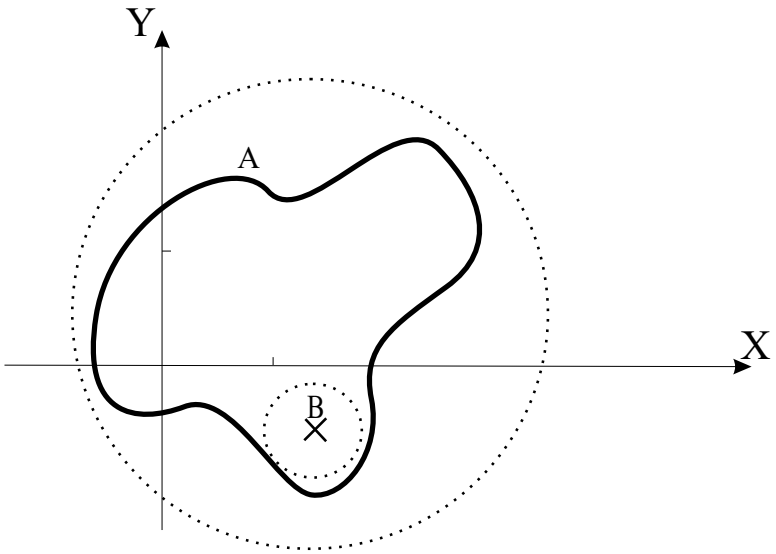



Figure 2: The set $A$ shown does not include its boundary. It is an open and bounded set because any point $B$ of $A$ belongs to an open ball that is included in $A$ and the whole set $A$ is included in an open ball.

**Note 2** In a metric space $E$, only two subset of $E$ are at the same time open and closed. Those two subsets are the entire metric space $E = \mathbb{R}$, and the empty set $\emptyset$.

**Definition of (strong) convergence of a sequence** A sequence[6] of elements $x_n$ of a metric space $(A, d)$ is said to converge to a limit $y \in A$ if there exists a $y \in A$ such that for any $\epsilon \in \mathbb{R}_*^+$ smaller that an $\epsilon_0 \in \mathbb{R}^+$ there exists an $N \in \mathbb{N}$ such that all the elements $x_n$ with $n > N$ are in the open ball of radius $\epsilon$ and center $y$.

**Note 1** A sequence can only have one or no limit.

**Note 2** Convergence is called "strong convergence" when it is necessary to distinguish it from the notion of weak convergence defined below.

**Example 2.4** *In the space $\mathbb{R}$ equipped with the Euclidian metric, the sequence defined by $x_n = 1/n$ converges to zero (see Fig 3).*

**Example 2.5** *We denote by $B$ the subset of $\mathbb{R}$ defined by*

$$B = \{x \in \mathbb{R}, \exists n \in \mathbb{N}^*, x = 1/n\} \tag{6}$$

[6]Note that we use the word "sequence" to denote a denumerable subset of a set . Sometimes the word "series" is used instead, although technically the word "series" should be reserved for the sums of the terms of a sequence.

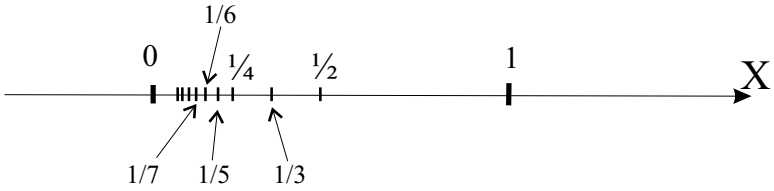



*Figure 3: The first few terms of the sequence $x_n = 1/n$ which converges to 0. Clearly for any strictly positive $\epsilon$ there exists an $N$ such that all terms $x_{n>N}$ of the sequence are in the ball of radius $\epsilon$ and center zero.*

*The sequence defined by $x_n = 1/(2n)$ does not converge in $B$ (because $B$ does not include the point 0), but regarded as a sequence of elements of $\mathbb{R}$ it converges to zero (because $\mathbb{R}$ includes zero).*

**Definition of compactness** A metric space $(A, d)$ is said to be compact if from any sequence of elements of $A$ a subsequence that converges can be extracted.

**Property** A compact space is closed.

This is intuitively natural. If a space $A$ is not closed, its complementary is not open, hence there exists at least a point $B$ of its complementary such that for any strictly positive $\epsilon_n$ there is a point $B_n$ of $A$ in the ball of radius $\epsilon$ and center $B$. Since this is true about any $\epsilon_n$, we can construct a sequence of points $B_n$ corresponding to the sequence $\epsilon_n = 1/n$. This sequence does not converge in $A$ and therefore $A$ is not compact.

**Property** The compacts in $\mathbb{R}^n$ are the closed and bounded sets.

## 3 Some background on normed vector spaces

In the analysis of finite element methods, a particular category of vector spaces named Hilbert spaces are of primary importance. In this section, we give some background on these vector spaces. First we recall some basic mathematical structures needed in the definition of a vector space:

**Definition of a group** We call a **group** a set $G$ equipped with an associative inner law such that there exists a neutral element and every element of $G$ is symmetrisable, i.e.

$$\exists \ \ 0 \in G, \qquad \forall x \in G, \qquad x + 0 = x \tag{7}$$

$$\forall \ \ x \in G, \qquad \exists x' \in G, \qquad x + x' = 0 \tag{8}$$

We have chosen the addition notation for the inner law, in which case the neutral element is denoted by 0. In the multiplication notation, the neutral element is denoted by 1.

**Definition of a ring** We call a **ring** a set $R$ equipped with two inner laws (addition and multiplication) such that:

1. For the addition, $R$ is a commutative group (i.e. a group with an inner law that satisfies $x + x' = x' + x$ for all $(x, x') \in R^2$)

2. The multiplication is associative and distributive with respect to the addition, i.e.

$$\forall(\alpha, \beta, \gamma) \in R, \begin{cases} (\alpha\beta)\gamma = \alpha(\beta\gamma) & \text{(associativity)} \\ (\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma & \text{(distributivity)} \end{cases} \tag{9}$$

3. The multiplication has a neutral element (or unit element), denoted by 1.

**Definition of a scalar field** We call a **scalar field** a ring $K$ that is not reduced to 0 (i.e. there exist at least an element in $K$ that is not a neutral element) and in which every non-zero element is invertible, i.e.

$$\forall\alpha \in K, \qquad \exists\beta \in K, \qquad \alpha.\beta = 1 \tag{10}$$

**Example 3.1** *The set of real numbers $\mathbb{R}$ and the set of complex numbers $\mathbb{C}$, equipped with the usual addition and multiplication are valid scalar fields. In the study of finite element methods, the scalar field we use is $\mathbb{R}$.* □

**Definition of a vector space** We call a **vector space** on the scalar field $K$ a set $E$ equipped with an inner law (addition) and an outer law (multiplication) such that:

1. For the addition, $E$ is a commutative group, i.e. it is a group and its inner law is commutative (i.e. $\forall(x, x') \in E^2, x + x' = x' + x$)

2. For all $(x, y, \alpha, \beta) \in E^2 \times K^2$ We have

$$1.x = x \tag{11}$$
$$\alpha.(\beta.x) = (\alpha.\beta).x \tag{12}$$
$$(\alpha + \beta).x = \alpha.x + \beta.x \tag{13}$$
$$\alpha.(x + y) = \alpha.x + \beta.y \tag{14}$$

In the study of finite element methods, we are only interested in vector spaces with $\mathbb{R}$ as their scalar field, and with real-valued inner products, so we restrict ourself to these conditions.

**Definition of a vector subspace** Consider a vector space $E$. We call $F \subset E$ a vector subspace of $E$ if $F$ equipped with the outer and inner laws of $E$ is a vector space.

**Example 3.2** *Consider the set $E = \mathbb{R}^3$, the scalar field $K = \mathbb{R}$ and the natural corresponding outer and inner products. Then the subset of $E$ defined by $F = \{(a, b, 0),\ \ (a, b) \in \mathbb{R}^2\}$ is a vector subspace of $E$.*

**Definition of an inner product or scalar product** We call an **inner product** or **scalar product** any bilinear, symmetric, positive definite application defined over a vector space and with values in $\mathbb{R}$, i.e. the application (.,.) is a scalar product over the vector space $E$ with scalar field $K$ if and only if

$$\forall(x,y,z,\alpha,\beta) \in E^3 \times K^2, \quad \begin{cases} (\alpha x + \beta y, z) = \alpha(x,z) + \beta(y,z) \\ (x,y) = (y,x) \\ (x,x) = 0 \iff x = 0 \end{cases} \tag{15}$$

**Example 3.3** *In $\mathbb{R}^2$, we have the traditional scalar product:*

$$(u,v) = u.v = ||u||||v||cos(\alpha) \tag{16}$$

*where alpha is the angle between the vectors $u$ and $v$.*

**Property** Consider the space $E = A \times B$ where $A$ and $B$ are two vector spaces equipped with the scalar products $(.,.)_A$ and $(.,.)_B$. It is easily seen that the application defined by

$$\begin{array}{ll} E & \to \mathbb{R} \\ (a,b) & \to (a,a)_A + (b,b)_B \end{array} \tag{17}$$

is a scalar product for the space $E$.

**Definition of a norm** Consider a vector space $E$ with $K = \mathbb{R}$ or $K = \mathbb{C}$ for scalar field. We call a **norm** on $E$ an application from $E$ into $\mathbb{R}_+$ denoted by $||.||$ such that

$$\forall(x,y,\alpha) \in E^2 \times K, \quad \begin{cases} ||\alpha x|| = |\alpha|||x|| \\ ||x+y|| \leq ||x|| + ||y|| \\ ||x|| = 0 \iff x = 0 \end{cases} \tag{18}$$

**Example 3.4** *Consider the vector space $E = \mathbb{R}^2$ with the scalar field $\mathbb{R}$. A possible norm for this space is defined by*

$$\begin{array}{ll} E & \to \mathbb{R} \\ (x_1,x_2) & \to |x_1| + |x_2| \end{array} \tag{19}$$

*Other norms are possible, for instance*

$$\begin{array}{ll} E & \to \mathbb{R} \\ (x_1,x_2) & \to \sqrt{x_1^2 + x_2^2} \end{array} \tag{20}$$

**Property** Consider the space $E = A \times B$ where $A$ and $B$ are two vector spaces equipped with the norms $||.||_A$ and $||.||_B$. It is easily seen that the application defined by

$$\begin{array}{ll} E & \to \mathbb{R} \\ (a,b) & \to ||a||_A + ||b||_B \end{array} \tag{21}$$

is a norm for the space $E$.

**Definition of a quadratic form** Given a bilinear application $A(.,.)$, we call **quadratic form** associated with $A(.,.)$ the application defined by $x \to A(x,x)$.

**Definition of a semi-norm** Consider a vector space $E$ with $K = \mathbb{R}$ or $K = \mathbb{C}$ for scalar field. We call a **semi-norm** on $E$ an application from $E$ into $\mathbb{R}_+$ denoted by $||.||$ such that

$$\forall (x,y,\alpha) \in E^2 \times K, \quad \begin{cases} ||\alpha x|| = |\alpha| ||x|| \\ ||x+y|| \leq ||x|| + ||y|| \end{cases} \tag{22}$$

i.e. unlike with a proper norm, with a semi-norm we can have that for $x \neq 0$ in fact $||x|| = 0$.

**Definition of the norm associated with a scalar product** Consider a vector space $E$ equipped with the scalar product $(.,.)$, and with its scalar field being either $\mathbb{R}$ or $\mathbb{C}$. Then the application $x \to \sqrt{(x,x)}$ is called the **norm associated with the scalar product** $(.,.)$. Hence the norm associated with a scalar product is the square root of the quadratic form associated with that scalar product.

**Example 3.5** *In mechanical engineering, scalar products and their associated norm appear naturally. For instance in the linear analysis of a three-dimensional medium when by use of the principle of virtual work we introduce the work done by a displacement field $U$ in the stress field created by another displacement field $v$, that work $A(u,v)$ can be shown to be a scalar product on the functional space of admissible displacements (admissible displacements must satisfy the boundary conditions and regularity properties). The strain energy associated with a displacement field $u$ is $A(u,u)/2$.*

*In the finite element analysis of this problem, the work done by a finite element displacement field $u_h$ in the stress field created by another finite element displacement field $v_h$, denoted either $A_h(u_h, v_h)$ or $V^T K U$ (in matrix form, where $U$ and $V$ are vector that store the nodal displacements and $K$ is the stiffness matrix), generally defines a scalar product. It may happen though that $Ah(.,.)$ is not definite positive but only positive: this is the case when the finite element formulation features spurious zero-energy modes.*

**Theorem 3.1 Schwarz's inequality** *Consider a vector space $E$ on the scalar field $K$ and equipped with a scalar product $(.,.)$ and the induced norm $||.||$. The for any $(x,y) \in E^2$ we have:*

$$(x,y) \leq ||x||\ ||y|| \tag{23}$$

*The equality $(x,y) = ||x||\ ||y||$ holds if and only if there exists $\alpha \in K$ such that $y = \alpha x$.*

**Example 3.6** *The* **metric induced by a norm** $||.||$*, denoted by $d(.,.)$, is defined by $d(x,y) = ||x-y|| \quad \forall (x,y) \in E^2$. We can easily check that this definition assures that $d(.,.)$ is indeed a metric.* □

The definition above shows that we can define the notion of convergence in any normed vector space as follows:

**Definition of convergence and absolute convergence in a vector space** Consider a sequence $x_n$ of elements of a vector space $E$ equipped with the norm $||.||$. The sequence $\sum_{n=1}^{n=\infty} x_n$ is said to **converge** if there exists $x \in E$ such that $\lim_{j\to\infty} ||\sum_{n=1}^{n=j} x_n - x|| = 0$. If $\sum_{n=1}^{n=\infty} ||x_n|| < \infty$, then the sequence is said to be **absolutely convergent**.

**Note** In general, a sequence can be absolutely convergent without being convergent.

**Example 3.7** *Let E denote the vector space of polynomials defined over [-1,1].* $||p|| = \max_{[-1,1]} |p(x)|$ *is a norm for this space. Consider the sequence* $p_n(x) = \sum_{i=1}^{i=n} \frac{x^{i-1}}{(i-1)!}$. *This sequence is not convergent in E because its limit is the exponential functional which is not a polynomial, but the sequence* $\sum_{i=1}^{i=n} ||p|| = \sum_{i=1}^{i=n} \frac{1}{(i-1)!}$ *converges to* $e \approx 2.718$ *and therefore the sequence* $p_n$ *is absolutely convergent but not convergent.*

**Definition of continuity** Let us consider two metric spaces $(E, d)$ and $(F, d')$ and an application $f$ from $E$ with values in $F$, and $x_0$ a point of $E$. The application $f$ is said to be **continuous** at $x_0$ if

$$\forall \epsilon \in \mathbb{R}_+^* \quad \forall x \in E \quad \exists \eta \in \mathbb{R}_+^* \quad d(x, x_0) \leq \eta \Rightarrow d'(f(x), f(x_0)) \leq \epsilon \tag{24}$$

**Definition of the derivative of a function** The calculus definition of the **derivative of a function** $f$ at a point $x_0$ of the interior of a domain of definition is (if this limit exists)

$$f'(x_0) = \lim_{h\to 0} \frac{f(x_0 + h) - f(x_0)}{h} \tag{25}$$

This notion is readily extended to include functions of several variables.

An application is differentiable in the sense of calculus over its domain of definition if it has a derivative at every point of the interior of its domain of definition.

**Example 3.8** *The function* $f(x) = |x|$ *is not differentiable in the sense of calculus over the interval* $[-1; 1]$ *because it is not differentiable at* $x = 0$.

**Definition of the spaces $C^0$, $C^1$, etc** An application that is continuous at all points of the set $A$ is said to be of **class** $C^0(A)$. If its derivative exists and is continuous at all points of $A$, the function is said to be of class $C^1(A)$, etc. If all derivatives of the application exist and are continuous at all points of $A$, the application is said to be of class $C^\infty(A)$.

**Definition of equivalence between norms** Consider a vector space equipped with two different norms which we denote by $||.||_a$ and $||.||_b$. These two norms are said to be **equivalent** if

$$\forall x \in E, \quad \exists (c, C) \in {\mathbb{R}_+^*}^2, \qquad c||x||_a \leq ||x||_b \leq C||x||_a \tag{26}$$

**Property** On a finite dimensional vector space, any two norms are equivalent. On an infinite dimensional vector space, not all norms are equivalent.

**Property** If two norms are equivalent, they define the same notion of convergence: any sequence that converges for one norm also converges for the other norm.

**Note** The property above is what makes the notion of equivalent norms very useful in engineering analysis. Say for some reason we want to show that a sequence in convergent for a certain norm $||.||_a$, and say this norm is difficult, or computationally expensive to evaluate but we know a simpler norm $||.||_b$ that is equivalent to $||.||_a$. Then we can simply show that the sequence is convergent for $||.||_b$ and automatically this will ensure that the sequence is also convergent for $||.||_a$.

**Example 3.9** *Let us consider the space $C^0([0,1])$ of continuous functions over $\Omega = [0,1]$. Two possible norms for this space* [7] *are*

$$||f||_\infty = \sup_{x\in\Omega} |f(x)| \tag{27}$$

*which is called the uniform norm and*

$$||f||_{L^1(\Omega)} = \int_0^1 |f(x)|dx \tag{28}$$

*which is called the $L^1$ norm* [8]*. The sequence of functions defined over $\Omega = [0,1]$ by*

$$\begin{cases} f_n(x) = nx & \text{if } x \in [0, 1/n] \\ f_n(x) = 2 - nx & \text{if } x \in [1/n, 2/n] \\ f_n(x) = 0 & \text{if } x \in [2/n, 1] \end{cases} \tag{29}$$

*(see Figure 4) is in the space of continuous functions over $\Omega$. However, $\lim_{n\to\infty} ||f_n||_{L^1(\Omega)} = 0$ whereas $\lim_{n\to\infty} ||f_n||_\infty = 1$. This proves that the two norms above are not equivalent.* □

**Definition of pointwise convergence** Considering a sequence of functions $g_n$ in $C^0([0,1])$, and if there exists a function $g$ such that for any $x \in [0,1]$ $\lim_{n\to\infty} g_n(x) = g(x)$ then it is said that the sequence of functions $g_n$ is **pointwise convergent** to $g$.

**Theorem 3.2** *There is no norm which defines pointwise convergence, i.e. there is no norm $||.||$ such that $||g_n - g|| \to 0$ if and only if $g_n(x) \to g(x)$ for all $x \in [0,1]$.*

**Proof** (adapted from [4]) Let us assume that such a norm exists and let us denote it by $||.||_{PCN}$. Consider the sequence of functions $f_n$ introduced in the previous example, and define $g_n = f_n/||f_n||_{PCN}$ for all $n \in \mathbb{N}$.

Since $||g_n||_{PCN} = 1$ the sequence of functions $g_n$ does not converge to the zero function for the norm $||.||_{PCN}$, although $g_n$ is trivially pointwise convergent to the zero function. This results in a contradiction, showing that such a norm does not exist. □

[7] *We recall that the supremum of a set $E$ of scalars, denoted by $\sup_E$ is the lowest upper bound to this set, and that its infimum, denoted by $\inf_E$ is the greatest lower bound to this set.*

[8] *The rigorous definition of the $L^1$ norm requires that the integral sign in Equation 28 be a Lebesgue integral, see Section 6.*

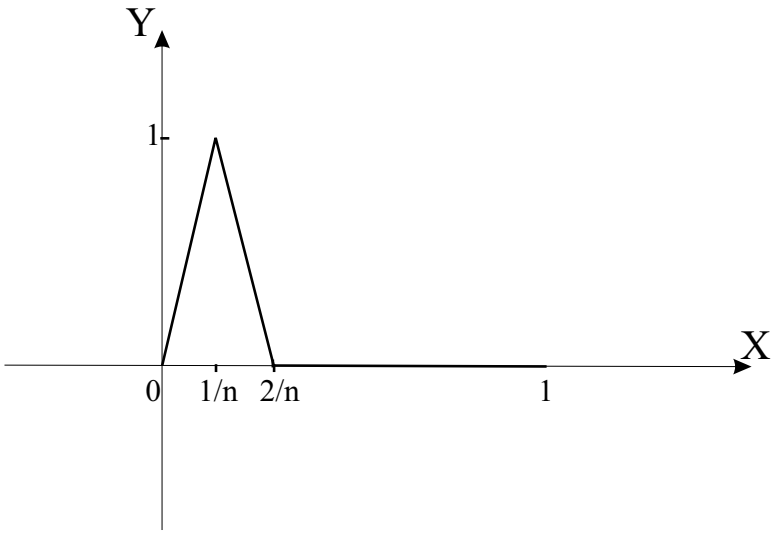



Figure 4: Functions $f_n$

**Definition of uniform convergence** Consider a sequence of functions $f_n$ defined over a domain $\Omega$ with values in the normed space $E$. This sequence is said to **converge uniformly** (or to be uniformly convergent) to a limit function $f$ if there exist a function $f$ such that

$$\lim_{n\to\infty} \sup_{x\in\Omega} ||f(x) - f_n(x)||_E = 0 \tag{30}$$

The difference between uniform convergence and pointwise convergence is illustrated in Figure 5.

**Definition of a Cauchy sequence** Consider a vector space $E$ equipped with the norm $||.||$. A sequence $u_n$ of elements of $E$ is called a **Cauchy sequence** (for that norm) if it satisfies

$$\lim_{i\to\infty,\ j\to\infty} ||u_i - u_j|| = 0 \tag{31}$$

i.e.

$$\forall \epsilon > 0, \quad \exists N \in \mathbb{N}, \quad \forall i > N, \quad \forall j > N, \quad ||u_i - u_j|| < \epsilon \tag{32}$$

**Example 3.10** *Using the absolute value as a norm, the sequence of elements of* $\mathbb{R}$ *defined by*

$$u_n = 1/n \tag{33}$$

*is a Cauchy sequence because*

$$\lim_{i\to\infty,\ j\to\infty} |u_i - u_j| = \lim_{i\to\infty,\ j\to\infty} |1/i - 1/j| = 0 \tag{34}$$

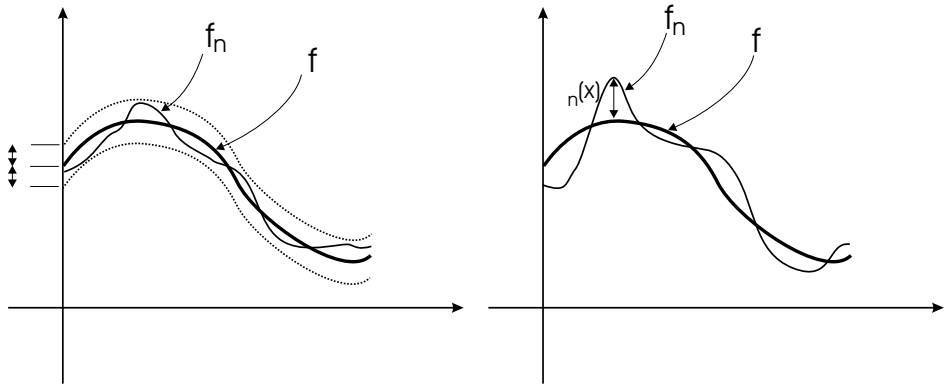



Figure 5: Uniform vs pointwise convergence

When a sequence of functions $f_n$ is pointwise convergent (on the right) to a limit function $f$, the distance at each point $s$, $\epsilon_n(x) = |f_n(x) - f(x)|$, between an element of the sequence and the limit decreases as $n$ goes to infinity, but for two points $x_1$ an $x_2$, $\epsilon_n(x_1)$ can go to zero much slower than $\epsilon_n(x_2)$ as $n$ goes to infinity. When a sequence of functions $f_n$ is uniformly convergent (on the left) to a limit function $f$, we can pick any strictly positive real $\epsilon$ we want and for $N$ large enough all functions $f_n$ with $n > N$ will be in the "ribbon" defined by $f$ and $\epsilon$.

*but the sequence*

$$u_n = n \tag{35}$$

*is not a Cauchy sequence because the limit*

$$\lim_{i\to\infty,\ j\to\infty} |u_i - u_j| = \lim_{i\to\infty,\ j\to\infty} |i - j| \tag{36}$$

*does not exist.*

**Theorem 3.3** *Every convergent sequence is a Cauchy sequence.*

**Definition of completeness** Consider a vector space $E$ equipped with the norm $||.||$. This space is said to be **complete** for this norm if *any* Cauchy sequence (for that norm) converges (for that norm) to a limit that is in $E$.

Note that completeness is a property of the vector space and the norm considered. A vector space that is complete for one norm may not be complete for another norm. In Example 3.11 below, we give an instance of a space that is complete with one norm and not complete for a second norm.

**Theorem 3.4** *A normed space is complete if and only if every absolutely convergent sequence converges.*

**Theorem 3.5** *The spaces $\mathbb{R}^n$ are complete for all $n \in \mathbb{N}$ (Note that we do not need to specify the norm for which these spaces are complete, because in a finite dimensional space all norms are equivalent and if a space is complete for one norm it is also complete for any equivalent norm).*

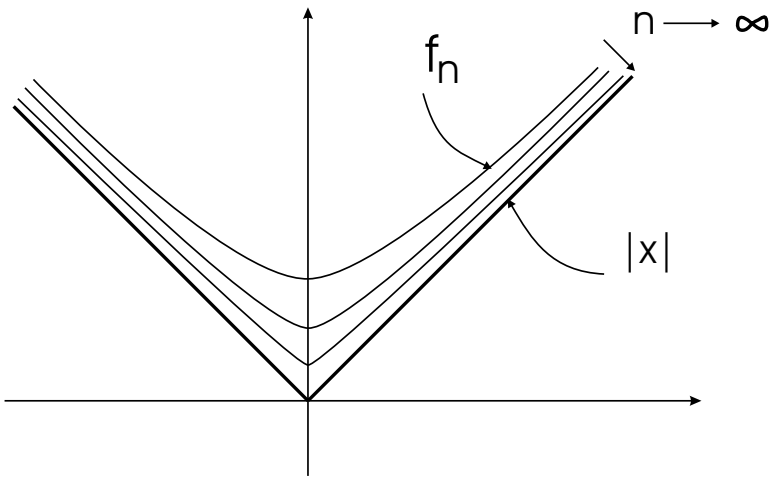



*Figure 6: Example functions for Example 3.11*
*As n goes to infinity, the functions $f_n$ approach the function $|x|$, which, unlike the functions $f_n$, is not of class $C^1([-1;+1])$.*

**Theorem 3.6** *Consider a Cauchy sequence $x_n$ in the normed vector space $E$. The sequence (of real numbers) $||x_n||$ is convergent.*

**Proof** $|\ ||x_n|| - ||x_m||\ | \leq ||x_n - x_m|| \to 0$ as $n$ and $m$ go to $\infty$, therefore the sequence $||x_n||$ is a Cauchy sequence of real numbers, hence it converges.

**Example 3.11** *(taken from [1]) Consider the vector space $E$ of functions of class $C^1([-1;1])$. The application $f \to \sup_{x\in[-1;1]} |f(x)|$ is a norm for $E$ called the uniform norm, denoted by $||.||_\infty$.*
*Consider the sequence of functions of $E$ defined by $f_n(x) = \sqrt{x^2 + n^{-2}}$. It is a Cauchy sequence for the uniform norm, because (assuming $i < j$) $\forall x \in [-1;1]$, $\sup_{x\in[-1;1]} |f_i(x) - f_j(x)| = 1/j - 1/i$ (this minimum being reached at $x = 0$) as revealed by a straightforward analysis of the variations of $f_i(x) - f_j(x)$. However, the limit of the sequence is the function $f(x) = |x|$ which is not in $E$ because $f$ is not differentiable at $x = 0$, so that $E$ is not complete for the uniform norm. See Figure 6*

*If for some reason we want to work in a complete vector space, we must either change norm, or change vector space:*

1. *Consider the vector space $E^*$ of functions of class $C^0([-1;1])$. It is complete for the uniform norm (We will not prove this result here).*

2. *Considering the application $f \to ||f||_a = \sup_{x\in[-1;1]} |f(x)| + \sup_{x\in[-1;1]} |f'(x)|$ (where the "prime" denotes the calculus derivative). This application is also a norm on $E$, and $E$ equipped with this norm is complete for this norm. This can be deduced from the fact that $E^*$ above is complete for the uniform norm as follows.*
   *Consider a Cauchy sequence $f_n$. We have (by definition of a Cauchy sequence) that $||f_i - f_j||_a$ goes to zero as $i$ and $j$ go to infinity. For that to happen, we need to have that separately the*

*sequence $f_n$ and $f'_n$ are Cauchy sequence for the uniform norm. Since the space $E*$ above is complete for the uniform norm, these two sequence have limits that are in $C^0([-1,1])$ and that we denote $f$ and $g$. To conclude, we need to prove that $f$ is differentiable and that $g$ is the derivative of $f$. To prove so, consider the sequence defined by*

$$f_n(x) = \int_{-1}^{x} f'_n(t)dt + f_n(-1) \tag{37}$$

*Because the sequence of functions $f'_n$ converges uniformly to $g$ and $f_n$ converges to $f$, we have that $f_n$ converges to $\int_{-1}^{x} g(t)dt + f(-1)$. Therefore we have*

$$f(x) = \int_{-1}^{x} g(t)dt + f(-1) \tag{38}$$

*This proves that $f$ is differentiable and that its derivative is $g$. We conclude that the sequence $f_n$ converges in the sense of $||.||_a$ to the function $f$ that is in $C^1([-1,1])$ and therefore the space $E$ is complete for the $||.||_a$ norm.*

□

Sometime the normed space that arises naturally from an application happens not to be complete, and we may need to enlarge this space to a complete space. The following Theorem indicates that it is possible. First we need to define density.

**Definition of density** The set $A$ is said to be dense in the set $B$ if $B = A \cup L$, where $L$ is the set of limits of sequence in $A$. In other words, $A$ is dense in $B$ if any point of $B$ is the limit of a sequence of $A$.

**Example 3.12** *The set $\mathbb{Q}$ is dense in $\mathbb{R}$. Consider for instance $\pi$. It is for instance the limit of the sequence of elements of $\mathbb{Q}$ defined by $u_1 = 3$; $u_2 = 3.1$; $u_3 = 3.14$; $u_4 = 3.141$; etc...*

**Theorem 3.7** *Let us consider the vector space $E$ with the norm $||.||_E$. It is possible to construct a vector space $\tilde{E}$ and a norm $||.||_{\tilde{E}}$ for that space such that the following four properties hold:*

- *(i) $E$ can be identified with a vector subspace of $\tilde{E}$.*
- *(ii) For any $x \in E$, $||x||_E = ||x||_{\tilde{E}}$.*
- *(iii) $E$ is dense in $\tilde{E}$.*
- *(iv) $\tilde{E}$ equipped with the norm $||.||_{\tilde{E}}$ is complete.*

*$\tilde{E}$ is then called the* **completion** *of $E$.*

**Proof** (adapted from [4]) Consider the space of the Cauchy sequence of elements of $E$. Let us call this space $F$. Now consider the part $\Gamma$ of $F^2$ defined by

$$\Gamma = \{(a_n, b_n) \in F^2, \ \lim_{n\to\infty} ||a_n - b_n|| = 0\} \tag{39}$$

It is easily seen that $\Gamma$ defines an equivalence relation. The set of all equivalence classes, equipped with the natural outer and inner product is trivially a vector space, which we denote by $\tilde{E}$. $|||[x_n]|||_{\tilde{E}} = \lim_{n\to\infty} ||x_n||$ is a norm for the space $\tilde{E}$. Let us show that this choice of $\tilde{E}$ satisfies the four conditions above.

Since every element $x \in E$ can be identified with the sequence $\{x, x, x, ...\}$ which is a Cauchy sequence, $E$ can be considered as a subspace of $\tilde{E}$ (Hence (i) is proven). Trivially, $|||[\{x, x, x, ...\}]|||_{\tilde{E}} = ||x||_E$ (and (ii) is proven).

$E$ is dense in $\tilde{E}$ because any element $[(x_1, x_2, x_3, ...)]$ of $\tilde{E}$ is the limit of the following sequence of elements of $E$: $(x_1, x_1, x_1, ...)$, $(x_2, x_2, x_2, ...)$, etc, hence (iii) is proven.

Consider a Cauchy sequence $X_n$ in $\tilde{E}$. Since $E$ is dense in $\tilde{e}$, for any $n \in \mathbb{N}$ there exists $x_n \in E$ that satisfies $||x_n - X - n||_{\tilde{E}} < 1/n$. Using the triangular inequality, we obtain

$$||x_n - x_m||_E = ||x_n - x_m||_{\tilde{E}} \leq ||x_n - X_n||_{\tilde{E}} + ||X_n - X_m||_{\tilde{E}} + ||X_m - x_m||_{\tilde{E}} \leq 1/n + 1/m + ||X_n - X_m||_{\tilde{E}} \tag{40}$$

and therefore $x_n$ is a Cauchy sequence in $E$. Denoting by $X = [x_n]$, we have

$$||X_n - X||_{\tilde{E}} \leq ||X_n - x_n||_{\tilde{E}} + ||x_n - X||_{\tilde{E}} < ||x_n - X_n||_{\tilde{E}} + 1/n \tag{41}$$

and the right hand side tends to zero as $n$ goes to infinity because $\lim_{n\to\infty} ||x_n - X||_{\tilde{E}} = 0$ and therefore $X_n$ converges to $X$ in $\tilde{E}$ which proves the completeness of $\tilde{E}$. □

**Definition of a Banach space** A **Banach space** is a normed vector space that is complete.

**Definition of a Hilbert space** A **Hilbert space** is a Banach space with the scalar field $\mathbb{R}$ and in which the norm is derived from an inner product.

Since a Hilbert space is always equipped with a norm, the notion of strong convergence in such a space is always well defined. We introduce a second type of convergence, called "weak convergence".

**Definition of weak convergence of a sequence** A sequence of elements $(x_n)$ of elements of a space $E$ equipped with an inner product (such as a Hilbert space) is said to be **weakly convergent** to a limit $x \in E$ if $lim_{n\to\infty}(x_n, y) = (x, y)$ for all $y \in E$.

**Property** Strong convergence implies weak convergence: If a sequence converges strongly to a limit $x$, it also converges to $x$ weakly. A weakly convergent sequence however is not necessarily strongly convergent.

**Proof** Consider a sequence $(x_n)$ of elements of a Hilbert space $E$, such that if $(x_i, x_j) = \delta_{ij}$ (where $\delta_{ij}$ denotes the Kronecker delta). An equality known as Bessel's equality states that for all $x \in E$, we have

$$||x - \sum_{i=1}^{N} (x, x_i)x_i||^2 = ||x||^2 - \sum_{i=1}^{N} |(x, x_i)|^2 \tag{42}$$

Letting $n$ go to $\infty$, we get that

$$\sum_{i=1}^{\infty} |(x, x_i)|^2 \leq ||x||^2 \tag{43}$$

which establishes that for every $x \in E$ the sum $\sum_{i=1}^{\infty} |(x, x_i)|^2$ is bounded. Therefore, for any $x \in E$, $\lim_{n\to\infty}(x, x_n) = 0$, and the sequence $(x_n)$ is weakly convergent to zero. But since $||x_n|| = \sqrt{(x_n, x_n)} = 1$ for all values of $n$ the sequence $(x_n)$ is not strongly convergent to zero because $\lim_{n\to\infty} ||x_n - 0|| \neq 0$. □

**Definition of a dual space** The space of linear continuous forms (applications with values in $\mathbb{R}$) on a normed vector space $E$ is called the **dual space** of $E$, and is usually denoted by $E'$. With the usual calculus addition and multiplication, $E'$ is a vector space. We can equip the dual of $E$ with the following **dual norm** (which is indeed a norm for the space $E'$ by the definition 18)

$$||x'||_{E'} = \sup_{x\in E,\ x\neq 0} \frac{x'(x)}{||x||_E} \tag{44}$$

As with any space, other norms are possible, but this one has the useful property that $x'(x) \leq ||x'||_{E'}||x||_E$. Note that the action of an element $x'$ of the dual space $E'$ on an element $E$ of the space $E$ can be denoted $x'(x)$ (as above), or $(x', x)$, or $< x', x >$. The notation $(x', x)$ can be ambiguous as is easily confused with a scalar product. The action of an element of $E'$ on an element of $E$ is referred to as a **duality pairing** .

**Riesz representation theorem** Consider a Hilbert space $H$. To every element $v_f \in H$ corresponds a linear continuous form $L_{v_f}$ defined by $L_{v_f}(v) = (v_f, v)$. The Riesz representation theorem states that reciprocally, to every given linear continuous form $L$ over $H$ there corresponds a unique $v_f \in H$ such that $L(.) = (v_f, .)$.

This allows to identify a Hilbert space with its dual space by a **Riesz map**. A Riesz map is the application $K : H \to H'$ that allows to find the element $v_f \in H$ such that $L = L_{v_f}$ by solving the equation $L = K(v_f)$. We then have $||L||_{H'} = ||v_f||_H$ because

$$||L||_{H'} = ||L_{v_f}||_{H'} = \sup_{x\in E,\ x\neq 0} \frac{(v_f, x)}{||x||_E} = \frac{(v_f, v_f)}{||v_f||_H} = ||v_f||_H. \tag{45}$$

Another way of formulating the Riesz representation theorem is the following:

$$\forall L \in H', \quad \exists v_f \in H \ \text{ such that } \ L(v) = (v_f, v) \quad \forall v \in H \tag{46}$$

$$\text{and } v_f \text{ is obtained by solving} \quad L = K(v_f). \tag{47}$$

The application $K$ depends of course on the space $H$ considered.

**Example 3.13** *As a basic example, we consider $E = \mathbb{R}^2$ equipped with the ordinary scalar product*

*and with the basis vectors* $\mathbf{x}$ *and* $\mathbf{y}$*, and the linear form* $L$ *defined by*

$$\begin{cases} L(\mathbf{x}) = \mathbf{4} \\ L(\mathbf{y}) = \mathbf{3} \end{cases} \tag{48}$$

*We can immediately see that the application defined by*

$$\mathbf{u} \in \mathbb{R}^2 \rightarrow (4\mathbf{x} + 3\mathbf{y}, \mathbf{u}) \tag{49}$$

*is identical to the form* $L$*, and therefore the element* $\mathbf{v_f} \in \mathbf{E}$ *such that* $L = L_{\mathbf{v_f}}$ *is* $\mathbf{v_f} = 4\mathbf{x} + 3\mathbf{y}$*.*

**Example 3.14** *We give a more advanced example in Section 8.* □

In the case where the vector space $E$ is a Hilbert space, we can write

$$||L_y||_{E'} = \sup_{x \in E,\ x \neq 0} \frac{(y, x)}{||x||_E} \tag{50}$$

where we have denoted by $y$ the element of $E$ identified by the Riesz map to the element $L_y : x \in E \rightarrow L_y(x) \in \mathbb{R}$ of $E'$, and where $||.||_E$ denotes the norm of the vector space $E$.

We have said that the dual $E'$ of the normed vector space $E$ is itself a normed vector space. It is therefore possible to consider its own dual, which we naturally denote by $E''$. $E''$ is called the second dual or bidual of $E$. A linear injection [9] of a normed space $E$ into its bidual $E''$ is provided by the mapping J defined by:

$$x \rightarrow J(x) \in E'' \tag{51}$$

where

$$J(x)(x') = x'(x) \qquad \forall x' \in E' \tag{52}$$

With this definition, we can show that

$$||J(x)||_{E''} = ||x||_E \tag{53}$$

where

$$||x''||_{E''} = \sup_{x' \in E',\ x' \neq 0} \frac{(x'', x')}{||x'||_{E'}}. \tag{54}$$

Hence $J$ is an isometry [10] from $E$ to $E''$. If the range of $J$ is the entire space $E''$, we say that the normed space $E$ is a **reflexive space** . A reflexive space must be complete and hence a Banach space.

**Property** Consider a sequence $(v_n)$ in a reflexive space. If all weakly convergent subsequences of $(v_n)$ have the same limit, then the sequence $(v_n)$ is weakly convergent to that same limit.

[9] Let us recall that a part $\Gamma$ of $A \times B$ is an injection if for all elements $x$ in $A$ there exists a $y$ in $B$ such that $(x, y)$ is $\Gamma$.

[10] An isometry is an application that preserves distances: given two metric spaces $(A, d)$ and $(B, d')$, if for all $x$ and $y$ in $A$ we have $d(x, y) = d'(f(x), f(y))$ then $f$ is an isometry.

**Example 3.15** *Consider again the above example. For any* $\mathbf{u} \in E$*, we have, directly from the definition,* $J(\mathbf{u})(L) = (4\mathbf{x} + 3\mathbf{y}, \mathbf{u})$*. We can see that, by definition, we have*[11]

$$||J(4\mathbf{x} + 3\mathbf{y})||_{E''} = \sup_{\mathbf{a} \in E'_*} \frac{(4\mathbf{x} + 3\mathbf{y}, \mathbf{a})}{\sup_{\mathbf{b} \in E*} \frac{(\mathbf{a},\mathbf{b})}{||\mathbf{b}||_E}} = 5 = ||4\mathbf{x} + 3\mathbf{y}||_E \tag{55}$$

# 4 Review of the Riemann integral

Here we recall the definition of the Riemann integral of a bounded function $f$ from $[a, b]$ into $\mathbb{R}$.

We define

$$S_h^+(f) = h \sum_{j=-\infty}^{j=+\infty} \sup_{x \in [jh,(j+1)h] \cap [a,b]} f(x) \tag{56}$$

and

$$S_h^-(f) = h \sum_{j=-\infty}^{j=+\infty} \inf_{x \in [jh,(j+1)h] \cap [a,b]} f(x) \tag{57}$$

If there exist a number $\alpha$ that verifies

$$\alpha = \lim_{h \to 0} S_h^+(f) = \lim_{h \to 0} S_h^-(f) \tag{58}$$

then the function $f$ is said to be Riemann integrable and its Riemann integral is by definition $\int_{\mathbb{R}} f = \alpha$.

An interpretation of this definition is to say that the Riemann integral measures the limit value as $h$ goes to zero of the area of rectangles of width $h$ fitted under (and above) the curve $f(x)$.

The concept of the Riemann integral is extended to functions defined over $\mathbb{R}$ (not just a segment) and to functions that are not bounded (improper integrals) on the one hand, and on the other hand to functions that are defined over multi-dimensional domains.

For instance, consider the case of a bounded function $f$ defined over $\mathbb{R}$. This function is said to have an improper integral over $\mathbb{R}$ if the limit

$$\lim_{a \to -\infty \ , \ b \to +\infty} \int_a^b f(x)dx \tag{59}$$

exists, and this limit is then called the (improper) integral of $f$.

Consider now a function $f$ defined over an interval $]a, b]$ and assume that the function $f$ is bounded on any interval $[a_1, b]$ for any $a_1$ verifying $a < a_1 < b$. The function $f$ is said to have an

[11] *By* $E_*$ *we denote the set of non-zero elements of* $E$*.*

improper integral over $]a,b]$ if the limit

$$\lim_{a_1 \to a} \mathbb{R}\!\!\int_{a_1}^{b} f(x)dx \tag{60}$$

exists and this limit is then called the (improper) integral of $f$ over $]a,b]$.

Similarly, considering a function $f$ defined over $[a,b[\cup]b,c]$ that is bounded on any set $[a,b_1[\cup]b_2,c]$ where $a < b_1 < b < b_2 < c$, then if the limits

$$\begin{cases} \lim_{b_1 \to b} \mathbb{R}\!\!\int_a^{b_1} f(x)dx \\ \\ \lim_{b_2 \to b} \mathbb{R}\!\!\int_{b_2}^{c} f(x)dx \end{cases} \tag{61}$$

exist then the function $f$ is said to have an (improper) integral over $[a,c]$ equal to

$$\lim_{b_1 \to b} \mathbb{R}\!\!\int_a^{b_1} f(x)dx + \lim_{b_2 \to b} \mathbb{R}\!\!\int_{b_2}^{c} f(x)dx \tag{62}$$

This concept can be extended to consider functions with multiple singular points. However, if it is not possible to define intervals of $\mathbb{R}$ that do not contain a singular point, then the improper integral can not be considered. We will see that the Lebesgue integral does not have this limitation.

**Warning 1** Not all functions are Riemann integrable!

**Example 4.1** *Consider the function $f$ defined over $[0,1]$ by*

$$\begin{cases} f(x) = 1 & \text{if } x \in \mathbb{Q} \\ f(x) = 0 & \text{otherwise} \end{cases} \tag{63}$$

*This function is not Riemann integrable because $\lim_{h\to 0} S_h^+(f) = 1$ but $\lim_{h\to 0} S_h^-(f) = 0$ because $\forall h > 0, \forall j \in \mathbb{N}, \quad \sup_{x\in[jh,(j+1)h]\cap[0,1]} f(x) = 1$ and $\forall h > 0, \forall j \in \mathbb{N}, \quad \inf_{x\in[jh,(j+1)h]\cap[0,1]} f(x) = 0$*
□

**Property** All continuous functions and piecewise continuous functions are Riemann integrable. We recall that a function is said to be **piecewise continuous** if there is at most a countable number of points at which this function is not continuous (cf further for the definition of countability).

**Warning 2** Strange things can happen when you consider limits of integrals and integrals of limits. One thing that can happen is that the limit of integrable functions may not be integrable.

**Example 4.2** *Consider the sequence of functions $f_n$ defined over $[-1,1]$ by*

$$\begin{cases} f_n(x) = n & \text{if } x \in [-\frac{1}{2n}, +\frac{1}{2n}] \\ f_n(x) = 0 & \text{otherwise} \end{cases} \tag{64}$$

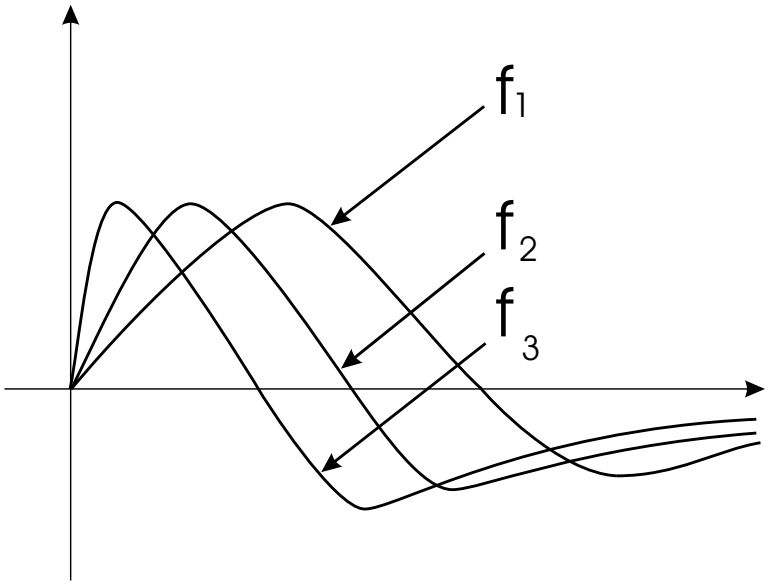



*Figure 7: Functions $f_n$*

*Then clearly*

$$\lim_{n\to\infty} \not{R}_{-1}^{1} f_n = 1 \tag{65}$$

*but $\lim_{n\to\infty} f_n$ does not have a proper Riemann integral because it is not bounded! The improper Riemann integral is given by*

$$\not{R}_{-1}^{1} \lim_{n\to\infty} f_n(x)dx \stackrel{def}{=} \lim_{\epsilon\in\mathbb{R}^+\to 0} (\not{R}_{+\epsilon}^{1} \lim_{n\to\infty} f_n(x)dx + \not{R}_{-1}^{-\epsilon} \lim_{n\to\infty} f_n(x)dx) = 0. \tag{66}$$

This shows that another problem is that "the limit of the integrals is not necessarily the integral of the limit".

**Example 4.3** *This example illustrates that the problem presented above occurs also with very smooth functions. Consider the sequence of functions $f_n$ defined over $\mathbb{R}$ by*

$$f_n(x) = -2n^2 x exp(-n^2x^2) + 2(n+1)^2 x exp(-(n+1)^2x^2) \tag{67}$$

*(see Figure 7) and the sequence of functions $g_n$ defined by*

$$g_n(x) = \sum_{j=1}^{j=n} f_j(x) \tag{68}$$

*We then have*

$$\lim_{n\to\infty} g_n(x) = -2x exp(-x^2) \tag{69}$$

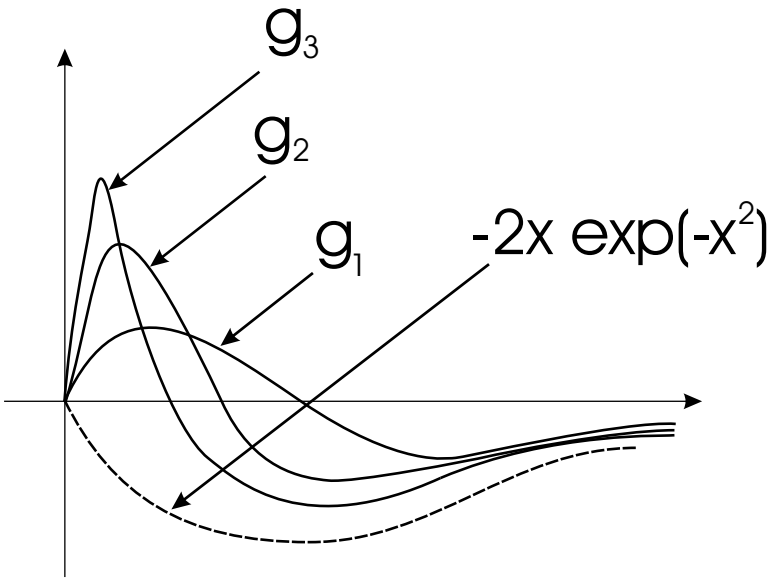



*Figure 8: Functions $g_n$ and their limit $-2xexp(-x^2)$. It should be noted that as $n$ increases the peak near the origin featured by $g_n$ becomes narrower but higher, while away from the origin $g_n$ approaches the limit function (dotted line).*

*(see Figure 8) and therefore*

$$\int_0^t \lim_{n\to\infty} g_n(x)dx = exp(-t^2) - 1. \tag{70}$$

*However*

$$\int_0^t g_n(x)dx = \int_0^t \sum_{j=1}^{j=n} f_j(x) = \sum_{j=1}^{j=n} \int_0^t f_j(x)dx = ... \tag{71}$$

$$... = \sum_{j=1}^{j=n} (exp(-n^2t^2) - exp(-(n+1)^2t^2)) = -exp(-(n+1)^2t^2) + exp(-t^2) \tag{72}$$

*therefore*

$$\lim_{n\to\infty} \int_0^t g_n(x)dx = exp(-t^2) \tag{73}$$

*where we see that*

$$\lim_{n\to\infty} \int_0^t g_n(x)dx \neq \int_0^t \lim_{n\to\infty} g_n(x)dx \tag{74}$$

□

**Example 4.4** *Consider the functions defined by*

$$\begin{cases} f_n(x) = 1 & \text{if } x \in [n, n+1] \\ f_n(x) = 0 & \text{otherwise} \end{cases} \tag{75}$$

*Clearly* $\lim_{n\to\infty} f_n(x) = 0 \ \forall x \in \mathbb{R}$*, so that* $\int_{\mathbb{R}} lim_{n\to\infty} f_n(x)dx = 0$ *whereas* $\lim_{n\to\infty} \int_{\mathbb{R}} f_n(x)dx = 1$
□

**Theorem 4.1** *Consider a compact set* $\Omega$ [12] *. If a sequence of functions* $f_n$ *converges uniformly to a limit* $f$ *(i.e.* $\exists f, \lim_{n\to\infty} \sup_{x\in\Omega} |f(x) - f_n(x)| = 0$*), and if the functions* $f_n$ *are Riemann integrable, then the function* $f$ *is Riemann integrable and we do have in this case that*

$$\lim_{n\to\infty} \int_{\Omega} f_n(x)dx = \int_{\Omega} \lim_{n\to\infty} f_n(x)dx \tag{76}$$

This result indicates that the functions defined by Equation 64 and 68 do not converge uniformly to their respective limits.

# 5 Lebesgue measure

In this section, we introduce the notion of Lebesgue measure. This notion will allow us in the next section to present Lebesgue integration, a more powerful form of integration than the Riemann integration. In particular, when we try to Riemann integrate a function $f$ which presents singularities, we need to play close attention to the behavior of $f$ at the points where these singularities occur. Lebesgue integration based on Lebesgue measure handles singularities in a very natural way: the value of $f$ at isolated points does not affect the Lebesgue integrals and these points can therefore be neglected a priori.

The definition of the Lebesgue integral involves domains that need to be "Lebesgue measurable", so this concept must be explained. In the following we only consider the Lebesgue measure (other measures can be defined. For a general introduction to measure theory see for instance [5]). Therefore, we can commit a slight abuse of notation and refer to "measurable sets" where in fact we should —rigorously speaking— always say "Lebesgue measurable sets" without risking any confusion.

We call a closed interval in $\mathbb{R}^n$ a set of the form $A = \{x \in \mathbb{R}^n; a_j \leq x \leq b_j; j \in \{1, 2, .., n\}\}$

We define the measure of such a closed interval as $\mu(A) = \prod_{j=1}^{j=n} |b_j - a_j|$.

A set is said to be "pavable" if it is the finite union of closed intervals with disjoint interiors (with no interior point in common). The measure of a "pavable" set is defined to be the sum of the measures of the closed intervals of which it is the union.

[12] *By definition, a set A is compact if from any sequence of elements of A a sub-sequence that converges in A can be extracted (Bolzano-Weierstrass property).*

The measure $\mu(A)$ of an open set $A$ is defined to be the supremum of the measures of pavable sets included in that open set.

The measure $\mu(A)$ of a compact set $A$ is defined to be the infimum of the measures of open sets that include that compact set.

**Definition of measurability of a set** A set $A$ is said to be **measurable** if, for any $\epsilon > 0$, one can find a sequence of compact sets $K_j$ and a sequence of open sets $\Omega_j$ such that

$$\cup_{j=0}^{\infty} K_j \subset A \subset \cup_{j=0}^{\infty} \Omega_j \tag{77}$$

and

$$\sum_{j=0}^{\infty} \mu(\Omega_j \backslash K_j) \leq \epsilon \tag{78}$$

The measure $\mu(A)$ of $A$ is then defined by

$$\mu(A) = \sup_{K \subset A, \ K \text{ compact}} \mu(K) = \inf_{\Omega \supset A, \ \Omega \text{ open}} \mu(\Omega) \tag{79}$$

This definition is illustrated in the following Figure 9.

We now report some results and definitions regarding countable sets, as they are required in the definition of Lebesgue integrals.

**Definition of countability** A set $A$ is **countable** if there exists a bijection [13] from $\mathbb{N}$ to $A$.

**Example 5.1** *It can be shown that $\mathbb{Q}$ is countable but $\mathbb{R}$ is not.*

The following results hold:

1. The intersection of a countable number of measurable sets is a measurable set. In particular the intersection of a finite number of measurable set is a measurable set.
2. The union of a countable number of measurable sets is a measurable set. In particular the union of a finite number of measurable sets is a measurable set.
3. The complementary set of a measurable set is a measurable set.

**Example 5.2** *Directly from the definitions, it is seen that because of these definitions a single point (in $\mathbb{R}^n$) is measurable, of measure zero, and that any finite collection of points is also measurable, of measure zero.*

$\mathbb{N}$ is measurable, of measure 0 in $\mathbb{R}$. $\mathbb{Q}$ is measurable, of measure 0 in $\mathbb{R}$.

[13] Let us recall that an application $\phi$ is a bijection from the set $A$ into the set $B$ if and only if it is both injective and surjective, i.e. $\forall b \in B, \quad \exists! a \in A, \quad \phi(a) = b$

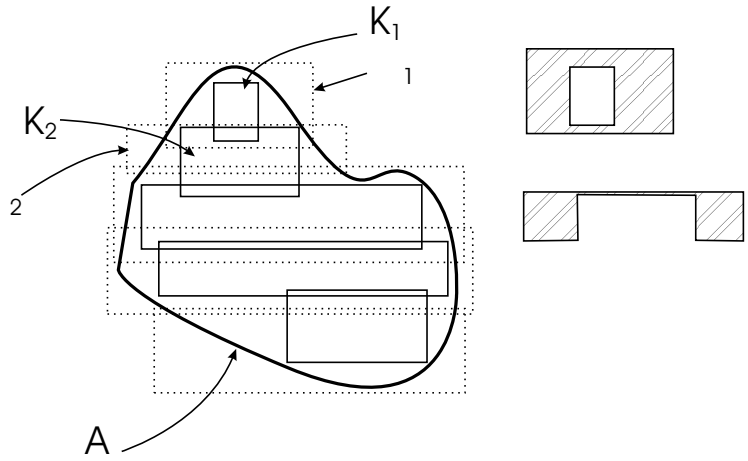



Figure 9: Measurability of a set. To determine if the set $A$ is measurable, we pick open sets $\Omega_j$ such that $A \subset \cup_j \Omega_j$ and compact sets $K_j$ such that $\cup_j K_j \subset A$, and essentially check whether "the volume of the open sets can be made arbitrarily close to the volume of the compact sets". This illustration shows this process, with 5 open sets (the dotted rectangles; their boundaries are not included in the sets) and 5 compact sets (the solid rectangles; their boundaries are included in the sets) shown. The measures $\mu(\Omega_1 \backslash K_1)$ and $\mu(\Omega_2 \backslash K_2)$ are the surface areas of the striped areas on the right. Of course, to show that the property holds for arbitrarily small values of $\epsilon$, we need to use many open sets and many compact sets. In general, the open and compact sets need not be rectangular in shape.

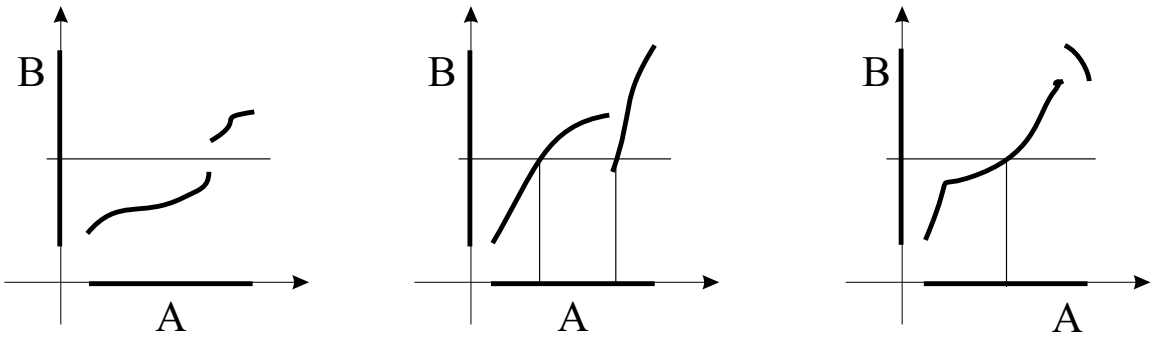



Figure 10: Left: an injective application: any element in $B$ is the image by the application of at most one element of $A$ – but may be the image of no element of $A$ as shown – . Center: a surjective application: any element in $B$ is the image of at least one element of $A$ – but may be the image of several elements of $A$ as shown – . Right: a bijective application: every element of $B$ is the image of exactly one element in $A$. Note that none of these concepts imply continuity, as shown.

**Note [1]** Whether there actually exist sets that are not measurable depends on fundamental logic axioms upon which mathematics are based. In engineering practice however, you will never encounter a set that is not measurable.

# 6 Measurable functions and their Lebesgue integral

## 6.1 Definitions and examples

**Definition of the reciprocal image of a set** The reciprocal image of a set $D$ by a function $f$ defined from $A$ with values in $B$ consists of the elements $x$ of $A$ such that $f(x) \in D$. It is denoted by $f^{-1}(D)$. It should be noted that this is only a notation, and the function $f$ may not be invertible (i.e. it may not be an injection), as shown by the following example.

**Example 6.1** *Consider the function $f$ defined over $\mathbb{R}$ by*

$$f(x) = x^2 \tag{80}$$

*Then we have $f^{-1}([0;+1]) = [-1;+1]$. Note that the function $f$ is not an injection, since for example $f(-1) = f(1) = 1$.*

**Definition of measurability of a function** A function $f$ from $\mathbb{R}^n$ into $\mathbb{R}_+ \cup \{+\infty\}$ is said to be **measurable** if the reciprocal image of any interval is a measurable set (i.e., if for all intervals $I$, $f^{-1}(I)$ is measurable, then $f$ is said to be measurable).

If the reciprocal image of $+\infty$ is a measurable set of non-zero measure, then the Lebesgue integral of $f$ is defined to be infinite. Otherwise, the Lebesgue integral of a measurable function is defined by:

$$\int\!\!\!\!\!\!\text{-} f(x)dx = \sup_{h>0} \sum_{j=0}^{\infty} \mu(f^{-1}(]jh,(j+1)h[))jh = \lim_{h\to 0} \sum_{j=0}^{\infty} \mu(f^{-1}(]jh,(j+1)h[))jh \tag{81}$$

Let us consider the case $n = 1$. The definition above can be interpreted by saying that the Lebesgue integral evaluates the limit value as $h$ goes to zero of the area of rectangles with a height that is a multiple of $h$ fitted under the curve $f(x)$ (See Figure 11), as opposed to the Riemann integral which evaluates the limit value as $h$ goes to zero of the area of rectangles with a width that is a multiple of $h$ fitted under and above the curve $f(x)$ (See Figure 12).

The definition of the Lebesgue integral can readily be extended to include the case of functions with values in $\mathbb{R}$ (instead of just $\mathbb{R}_+ \cup \{+\infty\}$).

**Definition of Lebesgue integrability** Consider a function $f$ from $\mathbb{R}^n$into $\mathbb{R}$. This function is said to be **Lebesgue integrable** if $\int\!\!\!\!\!\!\text{-}_{\mathbb{R}^n} |f(x)|dx < \infty$. The space of all Lebesgue integrable functions is denoted by $L^1(\mathbb{R}^n)$.

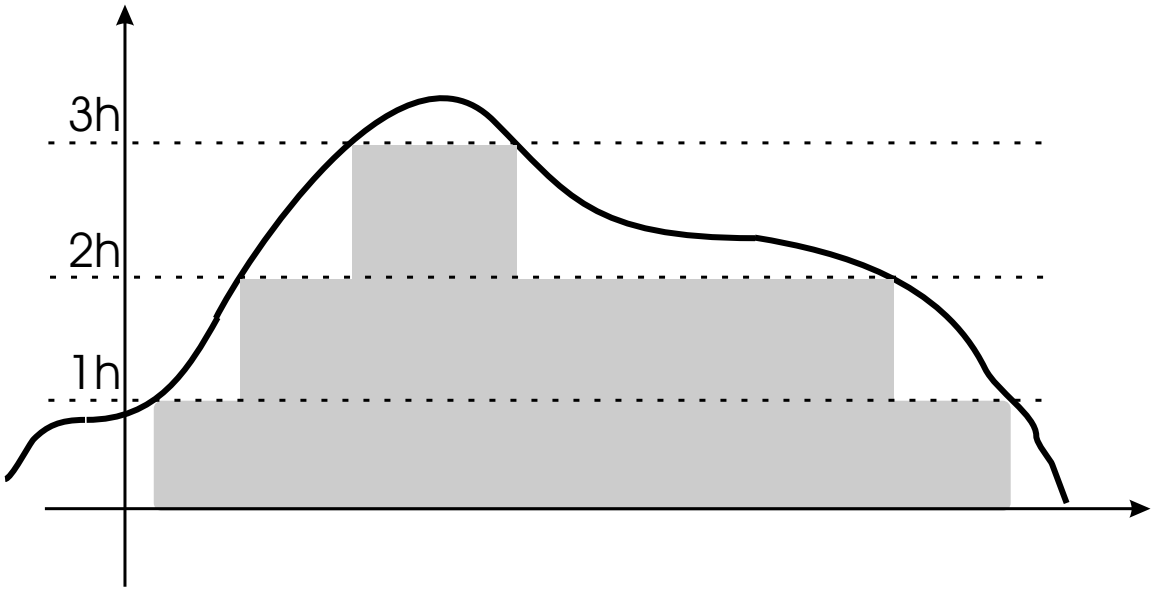



Figure 11: Lebesgue integral

**Equivalent definition of Lebesgue integrability** (in one dimension) A function $f$ from $\mathbb{R}$ into $\mathbb{R}$ is called **Lebesgue integrable** if there exists a sequence of step [14] functions $f_n$ such that the following two conditions are satisfied:

- $\sum_{n=1}^{n=\infty} \int_{\mathbb{R}} |f_n(x)|dx < \infty$
- f(x)=$\sum_{n=1}^{n=\infty} f_n(x)$ for every $x \in \mathbb{R}$ such that $\sum_{n=1}^{n=\infty} |f_n(x)| < \infty$

and the Lebesgue integral of $f$ is then defined by

$$\int_{\mathbb{R}} f(x)dx = \sum_{n=1}^{n=\infty} \int_{\mathbb{R}} f_n(x)dx \tag{82}$$

It can be shown that if a function is Riemann integrable then it is also Lebesgue integrable and its Riemann integral is equal to its Lebesgue integral.

**Example 6.2** *Going back to the sequence of functions defined by Equation 64, we see that* $\lim_{n\to\infty} f_n$ *is Lebesgue integrable (it was not Riemann integrable). The limit function* $\lim_{n\to\infty} f_n(x)$ *is zero almost everywhere (the origin* 0 *is of measure zero), so that it is zero everywhere except on a set of measure zero. Hence its Lebesgue integral is equal to* 0.

This last example shows that with the Lebesgue integral, we still have the problem that "the limit of the integral is in general not the integral of the limit".

[14] By step function we mean a function that is zero outside a segment $[a, b[$ where it is constant.

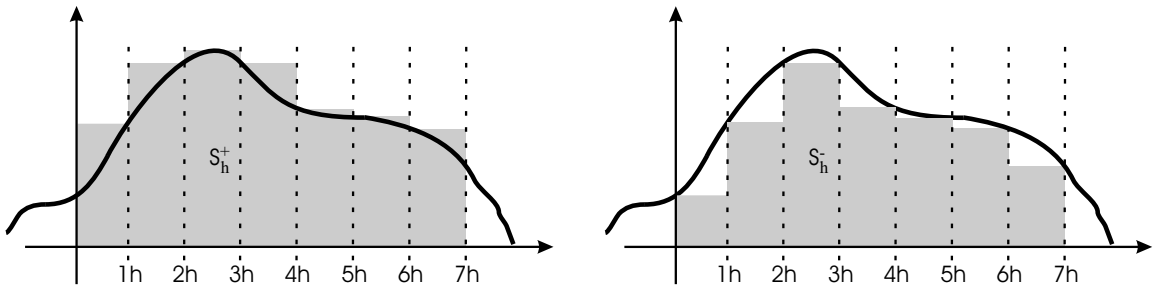



Figure 12: Riemann integral

## 6.2 Two theorems concerning sequence of Lebesgue integrals

We have seen earlier that for the Riemann integral, we had in general

$$\lim_{n\to\infty} \text{R}\!\int_\Omega f_n(x)dx \neq \text{R}\!\int_\Omega \lim_{n\to\infty} f_n(x)dx \tag{83}$$

and in the preceding section we showed that this is in general also the case when Lebesgue integral are considered. Here we state two theorems that give conditions under which we do have, for the Lebesgue integral,

$$\lim_{n\to\infty} \text{L}\!\int_\Omega f_n(x)dx = \text{L}\!\int_\Omega \lim_{n\to\infty} f_n(x)dx \tag{84}$$

**Theorem 6.1 Monotone convergence theorem** *Let us consider a sequence of functions $f_n(x)$ with values in $\mathbb{R} \cup +\infty$ that satisfy $\forall x \in \Omega \subset \mathbb{R}^n$ (where $\Omega$ is measurable), $0 \leq f_1(x) \leq f_2(x) \leq ...$ . Then*

$$\lim_{n\to\infty} \text{L}\!\int_{\mathbb{R}} f_n(x)dx = \text{L}\!\int_{\mathbb{R}} \lim_{n\to\infty} f_n(x)dx \leq +\infty \tag{85}$$

**Theorem 6.2 Lebesgue's dominated convergence theorem** *Let us consider a sequence of functions $f_n(x)$ that converge almost everywhere (i.e. except on a set of zero measure, also called a* **nullset***) to a limit function $f$. We also assume there exists a given function $h$ such that, $\forall n \in \mathbb{N}$, $|f_n(x)| \leq h(x)$ almost everywhere (i.e. for all $x \in \mathbb{R}^n$ except on a set of zero measure). Then*

$$\lim_{n\to\infty} \text{L}\!\int_{\mathbb{R}} f_n(x)dx = \text{L}\!\int_{\mathbb{R}} \lim_{n\to\infty} f_n(x)dx \tag{86}$$

*and*

$$\lim_{n\to\infty} \text{L}\!\int_{\mathbb{R}} |f(x) - f_n(x)|dx = 0 \tag{87}$$

This theorem is similar to the one given before for the Riemann integral, but it is more powerful in that we allow here the integral to be taken over a infinite domain and we do not restrict our discussion to bounded functions. Besides, uniform convergence of the sequence $f_n$ to $f$ is not required.

**Remark** The hypotheses of the dominated convergence theorem are not necessary for its conclusions to be verified, but are "fairly close" to being necessary, in the sense that if its hypotheses are not satisfied it is unlikely that its conclusions will hold.

**Example 6.3** *Consider the sequence of functions $f_n$ defined over $[-1,1]$ by*

$$\begin{cases} f_n(x) = \frac{sin(x)}{n} & if\ x < 0 \\ f_n(x) = \frac{1}{2} & if\ x = 0 \\ f_n(x) = 1 + \frac{sin(x)}{n} & if\ x > 0 \end{cases} \tag{88}$$

*and the functions $g_1$ and $g_2$ defined over the same domain by*

$$\begin{cases} g_1(x) = 0 & if\ x < 0 \\ g_1(x) = \frac{1}{2} & if\ x = 0 \\ g_1(x) = 1 & if\ x > 0 \end{cases} \tag{89}$$

*and*

$$\begin{cases} g_2(x) = 0 & if\ x \leq 0 \\ g_2(x) = 1 & if\ x > 0 \end{cases} \tag{90}$$

*We can see immediately that the sequence $f_n$ converges uniformly to $g_1$ but not to $g_2$. Therefore we can not directly employ the Theorem 4.1 to prove that*

$$\lim_{n\to\infty} \int_{-1}^{1} f_n(x)dx = \int_{-1}^{1} g_2(x)dx \tag{91}$$

*Equation 91 can be shown to hold as follows: first, prove that*

$$\lim_{n\to\infty} \int_{-1}^{1} f_n(x)dx = \int_{-1}^{1} g_1(x)dx \tag{92}$$

*and then show that*

$$\int_{-1}^{1} g_1(x)dx = \int_{-1}^{1} g_2(x)dx \tag{93}$$

*Now, we can show immediately, using Lebesgue's dominated convergence theorem that*

$$\lim_{n\to\infty} \int_{-1}^{1} f_n(x)dx = \int_{-1}^{1} g_2(x)dx \tag{94}$$

*by using the constant function equal to 2 as the h function required by the theorem.*

**Example 6.4** *Consider the sequence of functions $f_n$ defined over $[-1,1]$ by*

$$\begin{cases} f_n(x) = 0 & if\ x < 0 \\ f_n(x) = 1 & if\ 0 \leq x \leq 1/n \\ f_n(x) = 0 & if\ x > 1/n \end{cases} \tag{95}$$

*Although it is obvious that we have that*

$$\lim_{n\to\infty} \not{R}\int_{-1}^{1} f_n(x)dx = \not{R}\int_{-1}^{1} 0dx = 0 \tag{96}$$

*we can not obtain this result directly by using Theorem 4.1 because the sequence $f_n$ does not converge uniformly to any limit. Again, proving the equivalent result*

$$\lim_{n\to\infty} \not{L}\int_{-1}^{1} f_n(x)dx = \not{L}\int_{-1}^{1} 0dx = 0 \tag{97}$$

*is easy with Theorem 6.2.*

In the previous two examples, we can see that proving that a certain property holds for the Riemann integral is not as simple as proving that the equivalent property holds for the Lebesgue integral, because with the Lebesgue integration we have more powerful "tools".
There are also properties that hold for the Lebesgue integral that do not hold for the Riemann integral, and we will give an example of that in Example 7.4 of Section 7.4. Such examples need to involve "ill-behaved" functions, because for "well-behaved" functions there is no practical difference between the two integrations.

# 7 Introduction to Sobolev spaces

## 7.1 The space $\mathcal{D}$

Let $\Omega$ be a domain in $\mathbb{R}^n$. We denote by $\mathcal{D}(\Omega)$ or $C_0^\infty(\Omega)$ the set of $C^\infty(\Omega)$ functions with compact support in $\Omega$ (Note that $\mathcal{D}(\Omega)$ is defined based on the calculus definition of derivatives, see Section 7.3). When $\Omega$ is the space $\mathbb{R}^n$, one simply denotes this space by $\mathcal{D}$. There is a slight ambiguity due to the fact that the notation $\mathcal{D}$ does not explicitly indicate what value of $n$ is considered, but since in most practical applications $n$ is the dimensionality of the problem, this is not a serious issue.

## 7.2 Lebesgue spaces

The Sobolev spaces which are very generally used in the analysis of finite element procedures (and more generally in the analysis of differential equations) are based on Lebesgue integrals.

**Definition of the Lebesgue spaces** First we focus on the Lebesgue spaces, or $L^p$ spaces. These spaces are defined in the following way: the space $L^p(\Omega)$ (where $\Omega$ is a domain of $\mathbb{R}^n$ ) is defined as the space of the functions $f$ defined over $\Omega$ with values in $\mathbb{R}$ such that $(\int_\Omega |f(x)|^p dx)^{\frac{1}{p}} < +\infty$. Note that this definition is consistent with the definition of $L^1(\mathbb{R}^n)$ given in Section 6

Let us consider the functional $L^p$ from $L^p(\Omega)$ into $(R)$ defined by

$$L^p(f) = (\not{L}\int_\Omega |f(x)|^p dx)^{1/p} \tag{98}$$

We note that if two functions $f$ and $g$ are equal at every point of the domain $\Omega$ except on a set of zero measure (or nullset) then $L^p(f-g)=0$ and therefore the functional $L^p$ is not exactly a norm (It is easily seen that it satisfies all the other properties of a norm) . In technical terms, two functions $f$ and $g$ that satisfy $L^p(f-g)=0$ belong to the same equivalence class (as defined in Section 2) for the functional $L^p$, and the functional $L^p$ is a norm on the space of these equivalence classes. In practice, we identify functions that are equal almost everywhere and the functional $L^p$ is then also a norm on the space $L^p(\Omega)$ . The functional $L^p$ is therefore denoted by $||.||_{L^p(\Omega)}$.

Two functions that are equal except on a nullset are said to be equal almost everywhere (sometimes the notation "a.e." is used). Since two such functions are for all practical purposes equal, we frequently commit a slight abuse of language in equating them (It is important to understand that the value of a function over a set of zero measure is essentially meaningless, in the sense that the Lebesgue integral can not be used to distinguish between two functions that are equal almost everywhere).

It can be shown that $L^p$ is a Banach space, i.e. that it is complete with respect to the metric induced by the norm $||.||_{L^p(\Omega)}$ . Let us recall that $L^p$ is complete means that any Cauchy sequence of functions in $L^p$ converges to a limit which is itself in $L^p$.

We prove it here for $p=1$.

**Proof** We have said in Section 3 that a normed vector space where every absolutely convergent sequence converges is complete. Let us therefore consider an absolutely convergent sequence $f_n$ of elements of $L^1(\mathbb{R})$:

$$\sum_{n=1}^{n=\infty} \int\!\!\!\!\!\!-_{\mathbb{R}} |f_n(x)|dx < \infty \tag{99}$$

The function $f$ defined by

$$\begin{cases} f(x)=\sum_{n=1}^{n=\infty} f_n(x) & \text{if } \sum_{n=1}^{\infty} |f_n(x)| < \infty \\ f(x)=0 & \text{otherwise} \end{cases} \tag{100}$$

is in $L^1(\mathbb{R})$ and the sequence $\sum_{i=1}^{i=n} f_i$ converges to $f$ in the norm $L^1(\mathbb{R})$, i.e.

$$\lim_{n\to\infty} \; [\int\!\!\!\!\!\!-_{\mathbb{R}} |f(x) - \sum_{i=1}^{i=n} f_i(x)| \; dx \;] = 0 \tag{101}$$

Therefore the space $L^1(\mathbb{R})$ is complete.

**Property** The space $L^2(\Omega)$ can be equipped with the scalar product defined by

$$(f,g) = \int\!\!\!\!\!\!-_{\Omega} f(x)g(x)dx \tag{102}$$

This scalar product is associated with the norm $L^2(\Omega)$ defined earlier. This makes the space $L^2(\Omega)$ a Hilbert space.

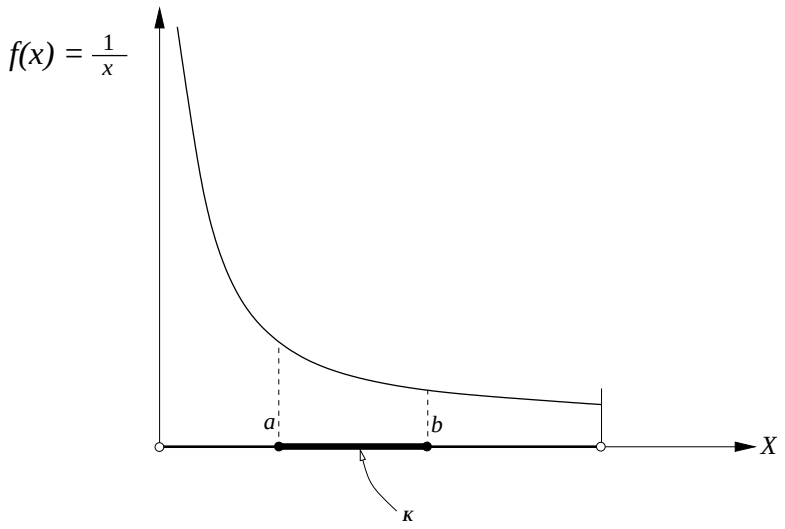



Figure 13: An $L^1_{\text{loc}}(]0;1[)$ function

**Definition of the space $L^1_{\text{loc}}(\Omega)$** We define the set of **locally integrable functions** over a domain $\Omega$ by

$$L^1_{\text{loc}}(\Omega) = \{f : f \in L^1(K) \quad \forall \text{ compact } K \subset \text{interior } \Omega\} \tag{103}$$

**Example 7.1** *Consider the function defined by $f(x) = 1/x$ over $]0;1[$. This function is not in $L^1(\Omega)$ but for any compact $K$ that is included in $]0;1[$ the integral $\int_K |f(x)|dx$ is finite (see Figure 13). For instance for $K = [a;b]$, we have $\int_K |f(x)|dx = ln(b) - ln(a)$. Therefore $f$ is in $L^1_{loc}(\Omega)$.*

## 7.3 Weak derivatives

Here we introduce the notion of weak derivatives, which extends the usual definition of derivatives.

**Definition of weak derivatives** Let us denote by $\alpha$ an n-tuple of non-negative integers (i.e. *alpha* is a vector with $n$ entries, where $n$ is the dimensionality of the problem, i.e. $\Omega \subset \mathbb{R}^n$) and define $|\alpha| = \sum_{i=1}^{i=n} \alpha_i$. We say that a function $f \in L^1_{\text{loc}}(\Omega)$ (where $\Omega \subset \mathbb{R}^n$) has a **weak derivative** $D^\alpha f$, if there exists a function $g \in L^1_{\text{loc}}(\Omega)$ such that

$$\int_\Omega g(x)\ \Phi(x)dx = (-1)^{|\alpha|}\int_\Omega f(x)\ \Phi^{(\alpha)}(x)dx \quad \forall \Phi \in \mathcal{D}(\Omega) \tag{104}$$

where $\Phi^{(\alpha)}$ denotes the calculus derivative of $\Phi$, i.e. $\frac{\partial^{|\alpha|}\Phi}{\partial^{\alpha_1}x_1\partial^{\alpha_2}x_2...\partial^{\alpha_n}x_n}$. $g$ is then called the weak derivative $D^\alpha f$ of $f$, or sometimes the derivative of $f$ in the sense of distributions , an expression which will be justified in Section 8 (See [2]). Very often the notation $\frac{\partial^{|\alpha|}f}{\partial^{\alpha_1}x_1\partial^{\alpha_2}x_2...\partial^{\alpha_n}x_n}$ is used for the weak derivative of $f$, which is ambiguous because this notation is usually employed to denote a derivative in the sense of calculus. However, since when a function has a derivative in the sense of

calculus it also has a weak derivative that is equal to its derivative in the sense of calculus, there is no risk of confusion.

It is important to note that the notation $D^\alpha f$ is valid because the order of differentiation is always irrelevant for weak derivatives (which is in general not the case for derivatives in the sense of calculus).

**Example 7.2** *Consider the function defined over* $\mathbb{R}$ *by* $f(x) = |x|$ *over* $]-1;+1[$. *We can calculate its weak derivative* $Df$ *using the definition above:*

$$\int\!\!\!\!\!\!\diagup_\Omega Df(x)\ \Phi(x)dx = (-1)^{|\alpha|}\int\!\!\!\!\!\!\diagup_\Omega f(x)\ \Phi'(x)dx \quad \forall\Phi \in \mathcal{D}(\Omega) \tag{105}$$

*We note that since* $f$ *and* $\Phi'$ *are Riemann integrable, the Lebesgue integral on the right-hand side is really the same as a Riemann integral, so we can write that* $Df$ *satisfies*

$$\int\!\!\!\!\!\!\diagup_\Omega Df(x)\ \Phi(x)dx = (-1)\int\!\!\!\!\!\!\diagup_\Omega f(x)\ \Phi'(x)dx \quad \forall\Phi \in \mathcal{D}(\Omega) \tag{106}$$

*The integral on the right can be subdivided into two integrals:*

$$\int\!\!\!\!\!\!\diagup_\Omega f(x)\ \Phi'(x)dx = \int\!\!\!\!\!\!\diagup_{\Omega+} x\Phi'(x)dx + \int\!\!\!\!\!\!\diagup_{\Omega-} -x\Phi'(x)dx \quad \forall\Phi \in \mathcal{D}(\Omega) \tag{107}$$

*where* $\Omega-$ *(respectively* $\Omega+$*) denotes the* $]-1;0[$ *(respectively* $]0;+1[$*), and we can immediately integrate by parts both of these integrals and get:*

$$\int\!\!\!\!\!\!\diagup_\Omega f(x)\ \Phi'(x)dx = \int\!\!\!\!\!\!\diagup_\Omega g(x)\ \Phi(x)dx \tag{108}$$

*where we have defined the function* $g$ *by* $g(x) = 1$ *if* $x > 0$ *and* $g(x) = -1$ *if* $x < 0$. *Note that we can specify any value for* $g$ *at* $x = 0$ *because integrals are not affected by values at points. Comparing the last equation to the definition of* $Df$, *we see that* $Df = g$.

**Example 7.3** *Consider the function defined over* $\mathbb{R}^2$ *by*

$$f(x, y) = |x| \tag{109}$$

*This function has a derivative in the sense of calculus* $\frac{\partial^2 f}{\partial y \partial x}$ *at all points (and it equals zero), but its derivative in the sense of calculus* $\frac{\partial^2 f}{\partial x \partial y}$ *is not defined along the line* $x = 0$. *The function* $f$ *has a weak derivative* $D^{(1,1)}f$ *which equals zero (almost everywhere). Here we see that due to the fact that in Lebesgue spaces the value of a function on a set of zero measure is meaningless (i.e. we can neglect sets of measure zero when we consider Lebesgue integrals), the difficulty that arises along* $x = 0$ *when using derivatives in the sense of calculus is automatically taken care of by the use of weak derivatives.*

The notation $D^\alpha f$ can of course be replaced by $\frac{\partial f}{\partial x_k}$ when $\alpha = (\alpha_1 = 0, \alpha_2 = 0, ..., \alpha_{k-1} =$

$0, \alpha_k = 1, \alpha_{k+1} = 0, ..., \alpha_n = 0)$ and no confusion is risked.

Note that if, over an interval $f$, has a derivative in the sense of calculus, then $f$ is also weakly differentiable on the interval, and its weak derivative is equal to the derivative in the sense of calculus. The weak derivative is therefore an extension of the concept of the differentiation in the sense of calculus (hence its name).

### 7.4 Sobolev spaces

Other functional spaces can be defined based on the previous definitions: the $H^k(\Omega)$ spaces are based on the weak derivatives and the Lebesgue integral. The space $H^k(\Omega)$ is defined by :

$$H^k(\Omega) = \{f : D^\alpha f \in L^2(\Omega) \quad \forall \alpha, |\alpha| \leq k\} \tag{110}$$

It can be proven that these spaces are Banach spaces for the metric induced by the norms $||.||_{H^k(\Omega)}$ for these spaces

$$||f||_{H^k(\Omega)} = \left( \sum_{|\alpha|=0}^{|\alpha|=k} ||D^\alpha f||^2_{L^2(\Omega)} \right)^{1/2} \tag{111}$$

**Property** We can equip the Sobolev spaces with scalar products associated with the norms above to make them Hilbert spaces . For instance, consider the space $H^1(\Omega)$, with say $\Omega \subset \mathbb{R}^2$. The scalar product associated with the norm $||.||_{H^1(\Omega)}$ is defined by

$$(f, g)_{H^1(\Omega)} = \int_\Omega (f(x)g(x) + \frac{\partial f}{\partial x_1}\frac{\partial g}{\partial x_1} + \frac{\partial f}{\partial x_2}\frac{\partial g}{\partial x_2})dx \tag{112}$$

where the derivatives involved are weak derivatives.

**Example 7.4** *In this example, we will show that the use of the Lebesgue integral is indispensable for the space $L^1$ to be complete (i.e. to show that if the Riemann integral were used to define the space $L^1(\Omega)$ then that space would not be complete). We will not show here that with the Lebesgue integral the space $L^1(\Omega)$ is complete.*

*We introduce the norm $||f||^1_R(\Omega) = \int_\Omega |f(x)|dx$ (which only differs from the $L^1(\Omega)$ norm in that it employs the Riemann integral instead of the Lebesgue integral) and the associated functional space defined by $R^1(\Omega) = \{f : \Omega \to \mathbb{R}, \ \ ||f||_{R^1(\Omega)} < \infty\}$.*

*We will show that a sequence of functions $f_n$ that is a Cauchy sequence for either of the $L^1(\Omega)$ or $R^1(\Omega)$ norms does not converge to limit that has a finite $R^1(\Omega)$ norm. That sequence of functions is defined as follows:*

*Consider a set of points $r_n$ that is dense in $[0, 1]$ and the sequence of functions ([2])*

$$f_n(x) = \sum_{j=1}^{j=n} 2^{-j} \log |x - r_j| \tag{113}$$

*These functions are measurable, and for all n satisfy*

$$\int_0^1 |f_n(x)|dx \leq 2\int_0^1 |log(x)|dx \tag{114}$$

*holds. This can be shown as follows. First we use the triangle inequality:*

$$\int_0^1 |f_n(x)|dx = \int_0^1 |\sum_{j=1}^{j=n} 2^{-j} \log|x - r_j||dx \leq \int_0^1 \sum_{j=1}^{j=n} 2^{-j} |\log|x - r_j||dx = \sum_{j=1}^{j=n} 2^{-j} \int_0^1 |\log|x - r_j||dx \tag{115}$$

*and then we bound each term in the sum on the right hand side by*

$$\forall r_j \in [0,1], \quad \int_0^1 |log|x - r_j||dx \leq 2\int_0^1 |log(x)|dx \tag{116}$$

*The latter bound comes from*

$$\forall r_j \in [0,1], \quad \int_0^{r_j} |log|x - r_j||dx \leq \int_0^1 |log(x)|dx \tag{117}$$

*and*

$$\forall r_j \in [0,1], \quad \int_{r_j}^1 |log|x - r_j||dx \leq \int_0^1 |log(x)|dx \tag{118}$$

*We can now show that the sequence is a Cauchy sequence for the $L^1$ norm. Assume $k > m$, then we have*

$$||f_k - f_m||_{L^1([0;1]} = \int_0^1 |f_k(x) - f_m(x)|dx = \int_0^1 |\sum_{j=1}^{j=k} 2^{-j} \log|x - r_j| - \sum_{j=1}^{j=l} 2^{-j} \log|x - r_j||dx \tag{119}$$

*Hence*

$$||f_k - f_m||_{L^1([0;1]} = \int_0^1 |\sum_{j=m+1}^{j=k} 2^{-j} \log|x - r_j||dx \leq \sum_{j=m+1}^{j=k} 2^{-j} \int_0^1 |\log|x - r_j||dx \tag{120}$$

*and we employ the upper bound of Equation 116 to conclude*

$$||f_k - f_m||_{L^1([0;1])} \leq \sum_{j=m+1}^{j=k} 2^{-j+1} \int_0^1 |\log(x)|dx \leq 2^{-m+2} \int_0^1 |\log(x)|dx \tag{121}$$

*Since the right hand side of Equation 121 goes to zero as k and m go to infinity (note that k does not actually appear in our upper bound), by definition, the sequence $f_n$ is a Cauchy sequence for the $L^1([0,1])$ norm. Since all the $f_n$'s are in $L^1([0,1])$ and $L^1([0,1])$ is a complete set the sequence must converge to a limit function $f \in L^1([0,1])$. Hence*

$$\int_0^1 |f(x)|dx < +\infty \tag{122}$$

*In the same way that we showed that $f_n$ is a Cauchy sequence for the $R^1(\Omega)$ norm, we can show that*

$$||f_k - f_m||_{R^1([0;1])} \leq 2^{-m+2} \int_0^1 |\log(x)| dx \tag{123}$$

*and therefore the sequence of functions $f_n$ is also a Cauchy sequence for the $R^1(\Omega)$.*

*We remark that if we had used Riemann integrals, the limit function $f = \sum_{j=1}^{j=\infty} 2^{-j} \log|x - r_j|$ would not be integrable (even with improper Riemann integrals), because $f$ is infinite at some point in any sub-interval of $[0, 1]$. This shows that the space $R^1(\Omega)$ is not complete and therefore it is not a Banach space (and therefore not a Hilbert space).*

*Considering the sequence of functions $g_n(x) = (|f_n(x)|)^{1/2}$ where the sequence $f_n$ is the one above, we can show that similarly if the space $L^2([0, 1])$ had been defined based on the Riemann integral (instead of the Lebesgue integral) that space would not be a Banach space.*

*We will see later that the completeness of the space $L^2$ is crucial in establishing theoretical results in mechanical engineering and finite element analysis in particular.*

**Important note** We have just shown that $|f|$ has a finite Lebesgue integral whereas its Riemann integral is not defined. This is the first such example function given in this document. The reader may feel that this example is a bit contrived. As a matter of fact, the definition of the function $f$ is rather complicated. It is also not very clear how such a function may appear in the context of mechanical engineering. In fact, in engineering, we are more used to encountering functions that are piecewise smooth. For such functions however, whether we use Riemann integrals or Lebesgue integrals makes no difference, because the two concepts coincide for those functions. To see a difference between the two types of integrals, we need to consider more irregular functions, such as the function $f$ above. We can not think of any mechanical engineering problem where a function would arise that has a Lebesgue integral that differs from its Riemann integral. This does not mean that Lebesgue integrals are useless in mechanical engineering though! As a matter of fact, because the Lebesgue integral enjoys some properties that the Riemann integral does not, there are a number of results such as the Lax-Milgram theorem that are critical in engineering that can be proven relatively easily using Lebesgue integrals but would not even hold if we used Riemann integral (see Section 9)

**Definition of homogeneous Sobolev spaces** We call **homogeneous Sobolev spaces** the following vector spaces:

$$H_0^k = \{f \in H^k, D^\alpha f|_{\partial\Omega} = 0, |\alpha| < k\} \tag{124}$$

where $\partial\Omega$ denotes the boundary of the domain $\Omega$ under consideration.

**Theorem 7.1 Poincaré (or Poincaré-Friedrichs) inequality** *Consider the case where the domain $\Omega$ is bounded, then we have:*

$$\forall f \in H_0^1(\Omega), \qquad \exists C \in \mathbb{R}, \qquad ||f||_{L^2(\Omega)} \leq C|f|_{H_0^1(\Omega)} \tag{125}$$

Note that $|f|_{H_0^1(\Omega)}$ denotes $\int_\Omega |\nabla f|^2 d\Omega$. The Poincaré-Friedrichs inequality implies that $\int_\Omega |\nabla f|^2 d\Omega$ is a norm for $H_0^1(\Omega)$ that is equivalent to the $H^1(\Omega)$ norm. Note also that $\int_\Omega |\nabla f|^2 d\Omega$ is a norm for $H_0^1(\Omega)$ but only a semi-norm for $H^1(\Omega)$.

The following Korn's inequality is effectively used in the proof of the existence and uniqueness of the solution to the equations of linear elasticity (see Section 9)

**Theorem 7.2 Korn's inequality** *Consider again a bounded finite open domain $\Omega \subset \mathbb{R}^N$. Then there exists a constant $C > 0$ such that*

$$\sum_{i=1}^{N} \int\!\!\!\!\!\!- _{\Omega} v_i(x)^2 dx + \sum_{i=1}^{N} \int\!\!\!\!\!\!- _{\Omega} \epsilon_{ij}(v)(x)^2 dx \geq C \sum_{i=1}^{N} \int\!\!\!\!\!\!- _{\Omega} (v_i(x)^2 + |\nabla v_i(x)|^2) dx \tag{126}$$

*for all $v \in H^1(\Omega)^N$, where*

$$\epsilon_{ij}(v) = \tfrac{1}{2}(\tfrac{\partial v_i}{\partial x_j} + \tfrac{\partial v_j}{\partial x_i}) \quad \textit{for all } 1 \leq i, j \leq N \tag{127}$$

*(i.e., if $v$ denotes a displacement field, $\epsilon$ is the corresponding strain tensor).*

**Property: Green's formula** Consider the case where the domain $\Omega \subset \mathbb{R}^n$ is bounded, then for all $i \in \{1, \cdots, n\}$, and for all $(u, v) \in (H^1(\Omega))^2$, we have:

$$\int\!\!\!\!\!\!- _{\Omega} \frac{\partial u}{\partial x_i}(x) v(x) dx = \int\!\!\!\!\!\!- _{\partial\Omega} u(x) v(x) n_i ds - \int\!\!\!\!\!\!- _{\Omega} u(x) \frac{\partial v}{\partial x_i}(x) dx \tag{128}$$

where $n$ denotes the unit outward normal vector and $\partial\Omega$ denotes the boundary of $\Omega$. Note that this formula reduces to integration by parts in one dimension.

**Definition of negative Sobolev spaces** We call **negative Sobolev spaces**, denoted by $H^{-k}(\Omega)$, the dual spaces of the Sobolev spaces, i.e. $H^{-k}(\Omega) = (H^k(\Omega))'$.

**Property** We have the following chain of inclusions:

$$\forall k \in \mathbb{N}_+, \quad H^k \subset H^{k-1} \subset \cdots \subset H^1 \subset L^2 = H^0 \subset H^{-1} \subset \cdots \subset H^{-k+1} \subset H^{-k}. \tag{129}$$

**Theorem 7.3 Rellich-Kondrachov (or Rellich-Kondrakov) Theorem** *When $\Omega$ is a bounded domain (which is usually the case in engineering applications), a weakly convergent sequence in $H^p(\Omega)$ is strongly convergent in $H^q(\Omega)$. This is summarized by saying that "the inclusion (or embedding) $H^p(\Omega) \subset H^q(\Omega)$ (for $0 \leq q < p$) is compact", or that "the injection from $H^1(\Omega)$ to $L^2(\Omega)$ is compact". Another way to state this result consists in saying that from any bounded sequence $(v_n)$ in $H^p(\Omega)$ we can extract a subsequence that is convergent in $H^q(\Omega)$.*

## 7.5 Sobolev spaces and smoothness

The following theorem, a 'simplified' version of Sobolev's inequality, characterizes the Sobolev spaces $H^k$ in terms of smoothness of the continuity of the functions that belong to these spaces.

First, we define the following :

$$||f||_{L^\infty(\Omega)} = \sup\{|f(x)|, x \in \Omega\} \tag{130}$$

We define the Lipschitz norm of a function $f$ by

$$||f||_{Lip(\Omega)} = ||f||_{L^\infty(\Omega)} + \sup\{|\frac{f(x)-f(y)}{x-y}| \ , (x,y) \in \Omega^2; \ \ x \neq y\} \tag{131}$$

If $||f||_{Lip(\Omega)}$ is finite, $f$ is said to be a **Lipschitz function**, or to a Lipschitz-continuous function .

**Property** A Lipschitz function over $\Omega$ is continuous over $\Omega$. However, not every continuous function is a Lipschitz function.

In lame terms, a differentiable function f(x) defined over a bounded domain is a Lipschitz function if and only if its slope $\frac{df}{dx}$ is bounded over the domain.

**Example 7.5** *Consider a function defined over* $[-1;+1]$ *by*

$$f(x) = x^2 \tag{132}$$

*It is a Lipschitz function. The same function defined over* $\mathbb{R}$ *is not a Lipschitz function. Consider the function defined over* $[0;2]$ *by*

$$f(x) = \sqrt{x} \tag{133}$$

*It is not Lipschitz continuous because its slope is not bounded. The same function defined over* $[1;2]$ *is a Lipschitz function though.*

Also, we introduce the notion of Lipschitz boundaries:

**Definition of a Lipschitz boundary** We say that a domain $\Delta$ has a **Lipschitz boundary** or **Lipschitz-continuous boundary** $\partial\Delta$ if there exists a collection of open sets $O_i$, a positive number $\epsilon$, an integer $N$ and a finite number $M$ such that the three following conditions hold:

1. For all $x \in \partial\Delta$, the ball of radius $\epsilon$ and center $x$ is included in one of the sets $O_i$.
2. No more than $N$ of the sets $O_i$ have non-zero intersections.
3. For each domain $O_i$ there exists a domain $\Omega_i$ such that $O_i \cap \Delta = O_i \cap \Omega_i$ and the boundary of $\Omega_i$ is the graph of a Lipschitz function $\Phi_i$ (i.e. $\Omega_i = \{(r,s) \in \mathbb{R}^n, \ \ r \in \mathbb{R}^{n-1}, \ \ s < \Phi_i(r)\}$) satisfying $||\Phi_i||_{Lip(\mathbb{R}^{n-1})} \leq M$.

**Interpretation** It is important to understand what domains do not have a Lipschitz boundary. Here we clarify the complex definition above with the emphasis on two dimensional sets for ease of drawing. The crucial condition for a set to satisfy to have a Lipschitz-continuous boundary is condition number 3. Note that by definition, the sets $\Omega_i$ are half-planes defined by the curves $\Phi(r)$ and the inequality $s > \Phi(r)$ and that we can choose the axes $r = constant$ and $s = constant$ that we like (see Example 7.6).

From the definition above, it can be seen that if a boundary has a cusped corner (i.e. a 360 degrees angle) it is not a Lipschitz boundary. For the same reason, domains with cracks do not have a Lipschitz boundary.

**Example 7.6** *We consider the three following domains, shown on Figure 14:*

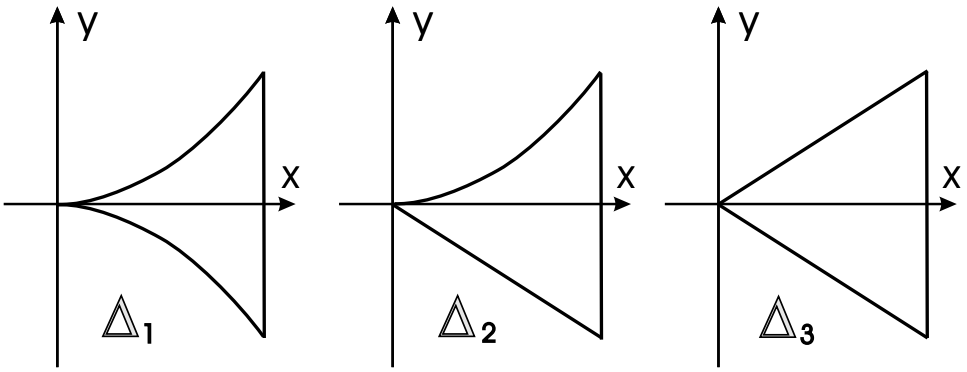



*Figure 14: Three domains: only $\Delta_1$ has a non-Lipschitz boundary.*

*1.*

$$\Delta_1 = \{(x, y) \in \mathbb{R}^2,\ \ 0 < x < 1; |y| < x^2\} \tag{134}$$

*2.*

$$\Delta_2 = \{(x, y) \in \mathbb{R}^2,\ \ 0 < x < 1; -x < y < x^2\} \tag{135}$$

*3.*

$$\Delta_3 = \{(x, y) \in \mathbb{R}^2,\ \ 0 < x < 1; -x < y < x\} \tag{136}$$

*The domains $\Delta_2$ and $\Delta_3$ have Lipschitz boundaries but $\Delta_1$ does not have a Lipschitz boundary.*

*One possible set of domains $\Omega_i$'s and subsequent $O_i$'s is shown in Figure 15 for $\Delta_3$ and it is easy to verify that they satisfy all three conditions above.*

*It can be seen that a similar set of domains $\Omega_i$'s would not work for $\Delta_1$ because the equation of the boundary of $\Omega_1$ would be something like*

$$\Phi_1(r) = \sqrt{|r|} \tag{137}$$

*where $r$ corresponds to $y$ in Figure 14 and this function can be seen not to be a Lipschitz function. Of course, to actually prove that $\Delta_1$ does not have a Lipschitz boundary, one would need to consider all possible sets of domains $\Omega_i$'s.*

*The domain $\Delta_2$ illustrates how the fact that we can choose the coordinate system $(r, s)$ is important. Consider the neighborhood of the point $x = y = 0$, shown on Figure 14 and again, rotated on Figures 16 and 17. Say we choose to use the set $O_1$ shown on the latter two figures. Since $\Omega_1$ must satisfy $O_1 \cap \Delta_2 = O_1 \cap \Omega_1$, the domain $\Omega_1$ that we need to use corresponds to the region above lines 1 and 2. With the $(r, s)$ coordinate system shown in Figure 16, the boundary $s = \Phi_1(r)$ of $\Omega_1$ is not a Lipschitz function because of the infinite slope at $r = 0$. If we use the coordinate system $(r', s')$ shown in figure 17, which is rotated by 20 degrees compared to $(r, s)$, then the function $\Phi_1'$ which describes the boundary $s' = \Phi_1'(r')$ is now a Lipschitz function. (Of course $\Phi_1'$ does not denote the derivative of $\Phi_1$ here)*

*Technical note: For the function $Phi_1'$ to be a Lipschitz, we must make sure that it does not go*

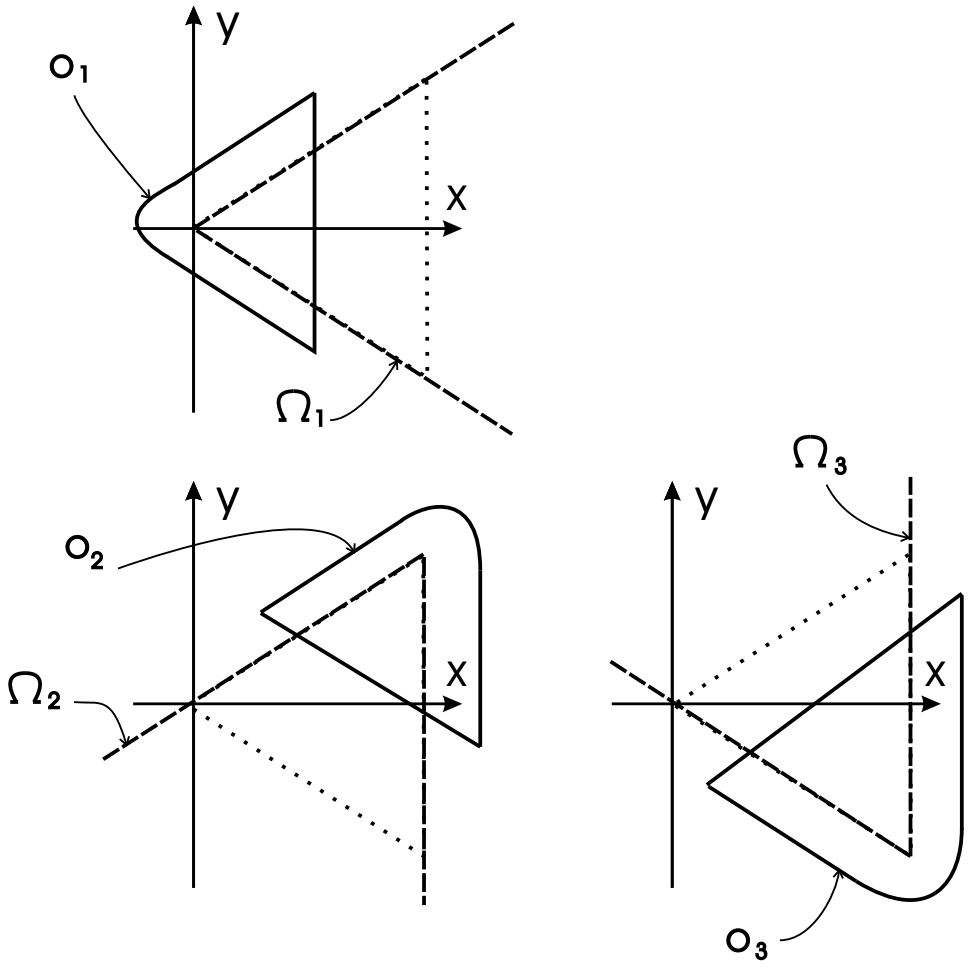



*Figure 15: Decomposition of a domain with a Lipschitz boundary: the $\Omega_i$ domains are triangularly shaped*

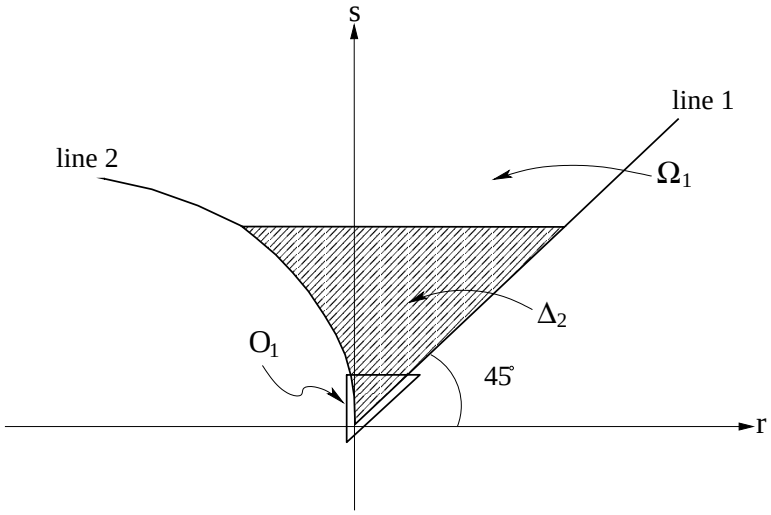



*Figure 16: Domain $\Delta_2$ with coordinate axes $r$ and $s$.*

*to infinity as $r'$ goes to infinity (to keep $||\Phi_1'||_{L^\infty(\mathbb{R})}$ finite) , so we would pick the function $Phi_1'$ such that it describes lines 1 and 2 in the neighborhood of $x = y = 0$ and then smoothly tends to some positive constant.*

*Note: While a curve which is not described by a Lipschitz function in one coordinate system $(r, s)$ may sometimes be described by a Lipschitz function in another coordinate system $(r', s')$ as shown above, sometimes there is no coordinate system such that a curve is described by a Lipschitz function. The cusped corner of $\Delta_1$ is a typical example.*

**Theorem 7.4 Sobolev's inequality** *Let $\Omega$ be an $n$-dimensional domain with a Lipschitz boundary and let $k$ be a positive integer such that*

$$k > n/2 \tag{138}$$

*Then there is a constant $C$ such that for all $u \in H^k(\Omega)$*

$$||u||_{L^\infty(\Omega)} \leq C||u||_{H^k(\Omega)} \tag{139}$$

*Moreover, there is a continuous function $v \in L^\infty(\Omega)$ such that*

$$||u - v||_{L^\infty(\Omega)} = 0 \tag{140}$$

One important consequence of Sobolev's inequality is that in one dimension, any function $u$ of class $H^1$ over a segment $[a, b]$ is bounded and it can be thought of as a continuous function.

However, in two dimensions, we are only guaranteed that an $H^2$ function is bounded over a domain with a Lipschitz boundary.

**Theorem 7.5 Corollary of Sobolev's inequality** *Let $\Omega$ be an $n$-dimensional domain with Lipschitz-continuous boundary, and let $k$ and $m$ be two integers such that $k - m > n/2$. Then*

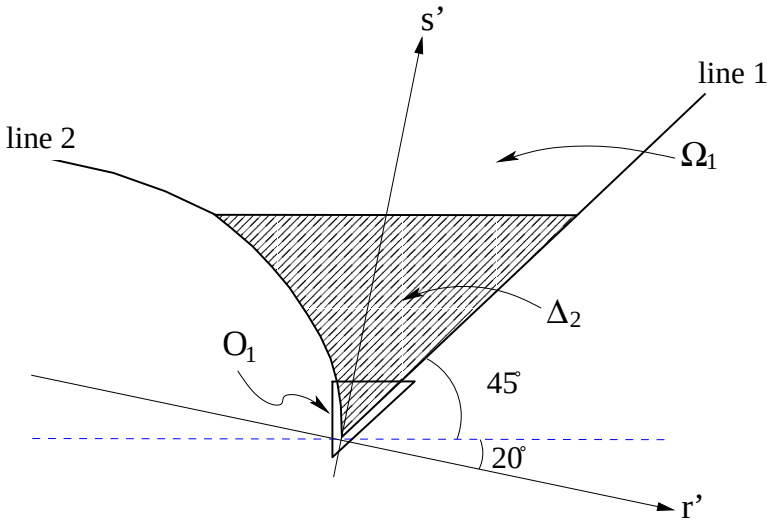



*Figure 17: Domain $\Delta_2$ with coordinate axes $r'$ and $s'$.*

*there is a constant $C$ such that for all $u \in H^k(\Omega)$*

$$\max_{|\alpha| \leq m} ||D^\alpha u||_{L^\infty(\Omega)} \leq C||u||_{H^k(\Omega)} \tag{141}$$

*and there exists a function $v \in C^m(\Omega)$ such that*

$$||u - v||_{L^2(\Omega)} = 0 \tag{142}$$

**Example 7.7** *Consider the following functions defined over $\Omega = ]-1, 1[$:*

$$\begin{cases} f_1(x) = 0 & \text{if} \quad x \neq 0 \\ f_1(x) = 1 & \text{if} \quad x = 0 \end{cases} \tag{143}$$

$$\begin{cases} f_2(x) = 0 & \text{if} \quad x < 0 \\ f_2(x) = 1 & \text{if} \quad 0 \leq x \end{cases} \tag{144}$$

*and*

$$f_3(x) = |x| \tag{145}$$

*The functions $f_1$ and $f_3$ are of class $H^1(\Omega)$.*

*The function $f_2$ is not. It can immediately be seen that there exist a continuous function – the zero function – that is equivalent to $f_1$ for the $L^\infty$ norm. It is also clear that $f_1$ is in fact an $H^k(\Omega)$ function for any $k$, and therefore the corollary tells us that there exists a function of class $C^\infty(\Omega)$ that is equivalent to $f_1$ for the $L^2(\Omega)$ norm. Of course that function is the zero function.*

*The function $f_2$ is of class $L^2(\Omega)$ but it is not of class $H^1(\Omega)$. One quick way to prove this is to consider that if it were of class $H^1(\Omega)$, the corollary to Sobolev's inequality says it would be possible to find a continuous function that is equal to $f_2$ almost everywhere, which is clearly not the case*

*due to the "jump" in $f_2$ at the origin. Consider for instance the functions $h_\epsilon$ defined by*

$$\begin{cases} h_\epsilon(x) = 0 & \text{if } x < -\epsilon \\ h_\epsilon(x) = 1 + \frac{x-\epsilon}{2\epsilon} & \text{if } -\epsilon < x < \epsilon \\ h_\epsilon(x) = 1 & \text{if } x > \epsilon \end{cases} \tag{146}$$

*The function $h_\epsilon$ is continuous and for small values of $\epsilon$ it is very close to $f_2$. However, $f_2$ and $h$ differ on the segment $]-\epsilon;+\epsilon[$ which is of finite measure $2\epsilon$. Therefore the functions $f_2$ and $h_\epsilon$ are not equal almost everywhere.*

*Note that the function $f_1$ cannot be used in mechanical engineering to model a concentrated force applied on a beam or rod at the origin. To model such a loading we would use a Dirac delta distribution. Distributions will be introduced in Section 8. That $f_1$ does not represent a point load can be shown as follows: the virtual work done by a body force $f$ in a displacement field $v$ equals $\int_\Omega f v d\Omega$, where $\Omega$ denotes the domain occupied by the structure considered. We can see that since $f_1$ is zero almost everywhere, so is $f_1 v$ and therefore $\int_\Omega f_1 v d\Omega = 0$. Of course the work done by a point load of magnitude $F$ applied at the origin in the displacement field $v$ equals $Fv(0)$. $f_1$ is therefore not the proper representation of a point load.*

*The function $f_3(\Omega)$ is another example of a function of class $H^1(\Omega)$. Contrary to $f_1$ it is not of class $H^2(\Omega)$. It can be seen that it is equivalent to a continuous function – itself! – but that there does not exist a function of class $C^1(\Omega)$ that is equivalent to $f_3$ for the $L^2(\Omega)$ norm.*

**Example 7.8** *It can be shown that the function*

$$f(x) = ln|\ ln||x||\ | \tag{147}$$

*defined over the domain $\Omega = \{x \in \mathbb{R}^2 \ \ , \ \ ||x|| < 1/2\ \}$ (i.e. the ball of euclidian radius $1/2$ in the plane) is of class $H^1(\Omega)$ but it is clearly not bounded, since it goes to infinity at the origin.*

In two dimensions (or higher, see Sobolev's inequality), there is not necessarily a continuous function that is equivalent in the sense of $||.||_{L^\infty}$ to a function of class $H^1$ and therefore pointwise values of $H^1$ have no real meaning. In one dimension however, pointwise values of $H^1$ functions have a sense so that there cannot be discontinuities in an $H^1$ function, unlike in an $L^2$ function.

# 8 Distributions

**Definition of a distribution** A **distribution** on $\Omega$ is a linear form over $\mathcal{D}(\Omega)$ whose restriction to any $\mathcal{D}(K)$ is continuous ($K$ being a compact subset of $\Omega$) (See [6, 7, 8, 9]). We denote by $T(\Phi)$ the action of the linear form $T$ on the function $\Phi$.

Distributions form a vectorial space (we can add distributions, and we can multiply a distribution by a real number), denoted by $\mathcal{D}'(\Omega)$ and which by definition is the dual space of $\mathcal{D}(\Omega)$ (hence the notation).

We define the derivative $\frac{\partial T}{\partial x_k}$ with respect to the $k$ variable of a distribution $T$ by the functional that satisfies

$$\frac{\partial T}{\partial x_k}(\Phi) = -T(\frac{\partial \Phi}{\partial x_k}) \qquad \forall \Phi \in \mathcal{D} \tag{148}$$

Note that in the equation above, the parentheses are used to indicate to which element of $\mathcal{D}$ the forms $\frac{\partial T}{\partial x_k}$ and $T$ are applied.

**Example 8.1** *Let us consider the application $T$ defined over $\mathcal{D}$ defined by*

$$\Phi \rightarrow T(\Phi) = \int_0^1 (x+1)\Phi(x)dx \tag{149}$$

*For instance, if $\Phi(x) = x$, we have $T(\Phi) = 5/6$. $T$ is a distribution. Its derivative in defined by Equation 148, i.e.*

$$\frac{dT}{dx}(\Phi) = -\int_0^1 (x+1)\frac{d\Phi}{dx}dx \quad \forall \Phi \in \mathcal{D} \tag{150}$$

*Hence (by an immediate integration by parts) its derivative is the application*

$$\frac{dT}{dx}(\Phi) = \Phi(0) - 2\Phi(1) + \int_0^1 \Phi(x)dx \quad \forall \Phi \in \mathcal{D} \tag{151}$$

*Hence for $\Phi(x) = x$, we get $\frac{dT}{dx}(\Phi) = -3/2$. Note that we could also have obtained the same result by plugging in $\Phi(x) = x$ directly in Equation 150*

We can show that the so-defined derivative of a distribution is also a distribution. This means in particular that any distribution is infinitely differentiable.

**Example 8.2** *The functional defined by*

$$\Phi \xrightarrow{T} T(\Phi) = \Phi(0) \tag{152}$$

*is a distribution known as the Dirac distribution. Note that because of its definition, the Dirac distribution can operate on less smooth functions than elements of $\mathcal{D}$ (i.e. in Equation 152, $\Phi$ does not need to be a function of class $C_0^\infty$: indeed the Dirac distribution is in $H^{-1}(\Omega)$, i.e. that $\Phi$ be in $H_0^1(\Omega)$ is sufficient. Like any other distribution, the Dirac delta can be differentiated. The derivative of the Dirac distribution is a distribution of class $H^{-2}(\Omega)$.* □

**Example 8.3** *Let $\Omega \subset \mathbb{R}^n$ be an open set. Then any function $f \in L^2(\Omega)$ can be identified with a distribution on $\Omega$ which is denoted by $T_f$, defined by:*

$$T_f(\Phi) = \int_\Omega f(x)\Phi(x)dx \quad \forall \Phi \in \mathcal{D}(\Omega) \tag{153}$$

*In this case, the action $T_f(\Phi)$ of the distribution $T_f$ on the function $\Phi$ is also sometimes denoted by $< f, \Phi >$.*□

We can use equation 153 to define a distribution corresponding to any $f \in L^1_{\text{loc}}(\Omega)$. However, the reciprocal is not true: it is not true that any distribution can be made to correspond to a function $f \in L^1_{\text{loc}}(\Omega)$. An example of that is the Dirac distribution which is a distribution but has no corresponding $L^1_{\text{loc}}(\Omega)$ function associated with.

Still, even when there does not exist an actual function $f \in L^1_{\text{loc}}(\Omega)$ such that

$$T(\Phi) = \text{⨍}_\Omega f(x)\Phi(x)dx \qquad \forall \Phi \in \mathcal{D} \tag{154}$$

one defines a "generalized function" $f$ which is identified with the distribution $T$ as well as a new integral by the equation:

$$T(\Phi) = \text{⨍}_\Omega f(x)\Phi(x)dx \qquad \forall \Phi \in \mathcal{D}. \tag{155}$$

As an abuse of language, the "generalized function" $f$ is also called the distribution $f$. Note that in fact Equation 154 should be understood as the definition of its entire right hand side by the left-hand side, since the integration sign ⨍ does not have the same meaning as the Lebesgue integration sign $\int$ in general. Also, in general, the generalized function $f$ does not have a meaning by itself: it only has a meaning when inside a ⨍ sign. Of course, in the case where there does exist a function $f \in L^1_{\text{loc}}(\Omega)$ such that $T(\Phi) = T_f(\Phi)$ then the ⨍ and $\int$ integration signs do have the same meaning.

**Example 8.4** *The Dirac distribution is identified with the "generalized function" $\delta$, which is defined by*

$$\forall \Phi \in \mathcal{D}(\Omega) \qquad \Phi(0) = \text{⨍}_\Omega \delta(x)\Phi(x)dx \tag{156}$$

*even though there exists no function $f \in L^1_{loc}(\Omega)$ such that $\int_\Omega f(x)\Phi(x)dx = \Phi(0) \qquad \forall \Phi \in \mathcal{D}(\Omega)$.*
□

**Example 8.5** *In this example, we show how the distributional integral sign is not to be confused with a Lebesgue or Riemann integral sign.*

*As mentioned earlier, point loads are modelled with Dirac delta distributions. For instance, in 1D the work done by a point force of magnitude $F$ at the origin in a displacement field $u$ belonging to $H^1$ is given by*

$$\mathcal{W}(u) = u(0) = \text{⨍}_\Omega F\delta(x)u(x)d\Omega \tag{157}$$

*Let us say that we are interested in calculating the finite element load vector corresponding to a point load of magnitude $F$ applied at the origin on a two node rod element occupying the domain $[-5; +5]$. The formula that is usually used to calculate the consistent nodal forces in a finite element isoparametric element is*

$$F_i = \int_{-5}^{+5} h_i(x) f^B(x)dx \tag{158}$$

*where $h_i$ denote the shape function corresponding to node $i$ and we intentionally omitted to indicate which integral sign (Riemann, Lebesgue, or distributional) is used.*

*If we try to employ Equation 158 to evaluate the consistent nodal load corresponding to our point load, we get*

$$F_i = \int_{-1}^{+1} h_i(r) f^B(r) Det(J) dr \tag{159}$$

*where $Det(J)$ denotes the determinant of the Jacobian of the mapping from $[-5; +5]$ to $[-1; +1]$*

*and is equal to* 5. *We have for our two-node element* $h_1(r) = (1+r)/2$ *and* $h_2(r) = (1-r)/2$. *Hence, continuing down the slippery slope of blindly trying to apply Equation 158, we get*

$$\begin{array}{l} F_1 = \int_{-1}^{+1} \frac{1+r}{2} F\delta(r) 5 dr = 5F/2 \\ F_2 = \int_{-1}^{+1} \frac{1-r}{2} F\delta(r) 5 dr = 5F/2 \end{array} \tag{160}$$

*Obviously this is* **not** *correct: the sum of the consistent nodal forces should be equal to* $F$. *So where did we go wrong? Well, to begin with, we should realize that Equation 158 only holds with the Lebesgue integration sign and can only be used for loadings such that the integrand is actually a function that can be Riemann integrated (and we mentioned earlier that the Dirac* $\delta$ *is a generalized function but not an actual function). So we started going wrong when we used* $f^B = \delta$ *as our loading. As a matter of fact* $\delta$ *is in* $H^{-1}$ *but not in* $L^2$ *and therefore the integrand in Equation 159 is also only in* $H^{-1}$ *and its Lebesgue integral cannot be considered! So how should we handle point loads in our finite element discretization? Well, in general, the correct expression for the work done by a loading in a finite element displacement field* $u$ *for a rod should really read*

$$\mathcal{W}(u) = < u, f^B > \tag{161}$$

*where* $< ., . >$ *denotes the duality pairing between* $H^1$ *and* $H^{-1}$. *This form highlights that* $H^1$ *is the space for displacements over which the problem is well-posed (this is due to the expression of the strain energy for a rod:* $1/2 \int_\Omega (du/dx)^2 dx$ *which can only be evaluated if* $du/dx$ *is in* $L^2$ *i.e. if* $u$ *is in* $H^1$*) and which loadings can be considered, namely those in* $H^{-1}$ *by duality with* $H^1$. *When the loading* $f^B$ *considered is such that the Lebesgue integral of* $f^B$ *times an arbitrary function in* $H^1$ *is well defined, then we can use Equation 158 with the integration sign meaning Lebesgue integration. When it is not, we need to work directly from Equation 161 to establish the consistent nodal loads. This is the case for our point load:*

$$\begin{array}{l} F_1 = < h_1, F\delta > = F h_1(0) = F/2 \\ F_2 = < h_2, F\delta > = F h_2(0) = F/2 \end{array} \tag{162}$$

*Now consider the case where the loading consists in a smooth part and a point load. Then we can decompose the loading into two parts, and then use Lebesgue integration to evaluate the consistent nodal loads corresponding to the smooth part while the consistent nodal loads corresponding to the point load are evaluated as above. The consistent nodal forces for the entire loading is the sum of the consistent nodal forces for each of the two parts.*

*An equivalent way of explaining why our initial calculation of the consistent nodal forces was incorrect consists in saying that Equation 158 holds with the distributional integration sign and that the change of variables* $x \rightarrow r$ *(which introduces the* $Det(J)$ *in the integrand) is not legitimate under the distributional integration sign.*

*We give more details on the rod in tension problem in Example 8.7.*

□

We recall that the definition of the weak derivative $g$ of $f$ is the function that satisfies

$$\int\!\!\!\!\!\!\int_{\Omega} g(x)\ \Phi(x)dx = (-1)^1 \int\!\!\!\!\!\!\int_{\Omega} f(x)\ \Phi^{(1)}(x)dx \quad \forall \Phi \in \mathcal{D}(\Omega) \tag{163}$$

(if such a function exists) and that the definition of the derivative $\frac{\partial T_f}{\partial x}$ of the distribution $T_f$ is

$$\frac{\partial T_f}{\partial x}(\Phi) = -T_f(\frac{\partial \Phi}{\partial x}) = (-1) \int\!\!\!\!\!\!\int_{\Omega} f(x)\Phi^{(1)}(x)dx \quad \forall \Phi \in \mathcal{D}(\Omega). \tag{164}$$

Denoting by $h$ the generalized function associated with the distribution $\frac{\partial T_f}{\partial x}$, we can rewrite Equation 164 in the following form:

$$\int\!\!\!\!\!\!\int_{\Omega} h(x)\Phi(x) = (-1)^{(1)} \int\!\!\!\!\!\!\int_{\Omega} f(x)\Phi^{(1)}(x)dx \quad \forall \Phi \in \mathcal{D}(\Omega) \tag{165}$$

Comparing now Equations 163 and 165, we note the similarity between $g$ and $h$. However, because in general $h$ is a generalized function, not an actual function, and the integral signs in Equations 163 and 165 are different, we can not conclude that $g = h$. However, we can extend the definition of the weak derivative $g$ of a function $f$ by modifying its definition to read:

$$\int\!\!\!\!\!\!\int_{\Omega} g(x)\ \Phi(x)dx = (-1)^1 \int\!\!\!\!\!\!\int_{\Omega} f(x)\ \Phi^{(1)}(x)dx \quad \forall \Phi \in \mathcal{D}(\Omega). \tag{166}$$

This definition allows for the weak derivative to be a generalized function ( also called distribution, as mentioned earlier). This explains why weak derivatives are also called the "derivatives in the sense of distributions". Indeed, in practice, one always works in the context of the theory of distributions, so that whether the derivative of a function is an actual function or only a generalized function is generally not mentioned, and the sign $\int$ is used for all integrals, where $\int\!\!\!\!\!\!\int$ is meant.

**Example 8.6** *Let us consider the Heaviside function*

$$\begin{cases} H(x) = 0 & \text{if } x < 0 \\ H(x) = 1 & \text{if } x \geq 0 \end{cases} \tag{167}$$

*The weak derivative (derivative in the sense of distributions) of this function is the Dirac distribution, as can be checked immediately from the definition of weak derivatives:*

$$\int\!\!\!\!\!\!\int_{\Omega} DH(x)\ \Phi(x)dx = (-1)^1 \int\!\!\!\!\!\!\int_{\Omega} H(x)\ \Phi^{(1)}(x)dx \quad \forall \Phi \in \mathcal{D}(\Omega) \tag{168}$$

*the term on the right being obviously equal to* $\Phi(0)$. □

**Property** Consider a distribution $T \in \mathcal{D}'$, and $\Omega \subset \mathbb{R}^n$. This distribution is an element of $H^{-1}(\Omega)$ if and only if there exist functions $f_0, \cdots, f_n \in L^2(\Omega)$ such that $T = f_0 + \sum_{i=0}^{i=n} \frac{\partial f_i}{\partial x_i}$, and

we then have:

$$\forall \Phi \in H_0^1(\Omega), \qquad < T, \Phi >= \int_\Omega \left( f_o(x)\Phi(x) - \sum_{i=1}^{i=n} f_i(x) \frac{\partial \Phi}{\partial x_i}(x) \right) dx \tag{169}$$

**Application** This result can be used immediately to prove that the Dirac distribution is in $H^{-1}(\Omega)$: take the Heaviside function as $f_0$ and zero for $f_1$.

**Applications of the theory of distributions** Before distributions where introduced physicists employed —without a formal justification that they worked— tricks to analyze certain phenomena governed by differential equations. For instance, they introduced a "Dirac function" denoted by $\delta$ which is 0 everywhere except at the origin where it is infinite, but with an integral equal to 1. This is a self-contradicting definition (if you use either the Riemann or Lebesgue integral) in view of the fact that the Lebesgue integral of a function that is equal to zero outside of a set of measure zero is zero. Let us stress again that there does not exist any actual function that verifies Equation 156. Distributions constitute the rigorous framework in which these tricks can be formally analyzed (Note that another way to go is to employ Laplace transforms, as is commonly done in electricity or control theory. Irregular "functions" tend to have smooth Laplace transforms. For instance the Laplace transform of a Dirac distribution is the constant function equal to one). Distributions are relatively recent, as they date back to Laurent Schwartz's work first published in 1950 [6, 7, 8].

In mechanics, one can employ Dirac distributions to model point forces, the derivatives of Dirac distributions to model concentrated moments, etc. In signal theory, Dirac distributions model impulses. In electromagnetism, they model loads, while derivatives of Dirac distributions model dipoles.

**Example 8.7** *Consider the following bar in tension problem. A bar occupying the region* $x \in \Omega = ]-1, 1[$ *and fixed at both ends is subjected to a concentrated force of magnitude* $F$ *applied at* $x = 0$. *For clarity, we assume that the product of the Young's modulus by the cross-sectional area* $E(x)A(x)$ *is equal to one.*
*The strain energy associated with the solution* $u$ *to this problem is given by*

$$\mathcal{E}(u) = \frac{1}{2} \int_\Omega \left( \frac{du}{dx} \right)^2 dx, \tag{170}$$

*prompting to search the solution in the space* $H_0^1(\Omega)$, *so that the strain energy remains finite and the boundary conditions are satisfied. The external virtual work done by an admissible displacement pattern* $v$ *can be written as the duality pairing*

$$\mathcal{W}_e(v) = \int_\Omega f(x)v(x)dx. \tag{171}$$

*Here* $f \in H^{-1}(\Omega)$ *is the distribution*

$$f(x) = F\delta(x) \tag{172}$$

*(The equation above should be understood in a distributional sense. From now on we will not indicate when equations are meant in a distributional sense) and $v$ is in $H_0^1(\Omega)$. The distribution 172 is chosen so that the virtual work obtained from Equation 171 is equal to its known "physical" value: $Fv(0)$ for any $v$ in $H_0^1(\Omega)$. Note that for Equation 171 to make sense (i.e. for the integral to yield a finite number) the loading must be chosen in $H^{-1}(\Omega)$, which means that this model does not allow moments (which would correspond to a loading distribution in $H^{-2}(\Omega)$ but not in $H^{-1}(\Omega)$) for instance to be applied to the bar.*

*We can determine the element $v_f \in H_0^1(\Omega)$ that satisfies*

$$\forall v \in H_0^1(\Omega), \qquad \int_\Omega f(x)v(x) = (v_f, v)_{H^1(\Omega)} \tag{173}$$

*Such an element exists, as stated by the Riesz representation theorem. The function $v_f$ is obtained by solving the (distributional differential) equation*

$$K_1(v_f) = f \tag{174}$$

*where $K_1 : H^1(\Omega) \to H^{-1}(\Omega)$ operator is given by*

$$K_1(v_f) = v_f - \frac{d^2 v_f}{dx^2} \tag{175}$$

*as can be seen by integrating by parts Equation 173. Note that in the more general case where we want to find the Riesz representation of an element of $H^k(\Omega)$ with $\Omega = ]x_1; x_2[$, one needs to consider the operator $K_k : H^k(\Omega) \to H^{-k}(\Omega)$ given by*

$$K_k(v_f) = \sum_{m=0}^{m=k} (-1)^m \frac{d^{2m} v_f}{dx^{2m}}. \tag{176}$$

*This expression follows by applying Green's formula (i.e. integration by parts) $k$ times for the scalar product $(v_f, v)_{H^k(\Omega)}$. For our problem of a bar in tension, we therefore need to solve*

$$-\frac{d^2 v_f}{dx^2}(x) + v_f(x) = f(x) \tag{177}$$

*where this equation should be understood in a distributional sense, i.e.*

$$\forall v \in H_0^1(\Omega), \qquad \int_\Omega (-\frac{d^2 v_f}{dx^2}(x) + v_f(x))v(x)dx = \int_\Omega f(x)v(x)dx \tag{178}$$

*We can solve Equation 178 in the following way. First we solve Equation 177 separately over $]-1; 0[$ and over $]0; -1[$. This gives rise to four constants, three of which can be expressed in terms of the fourth one by using the boundary conditions at both ends of the bar and the continuity of $v_f$ at $x = 0$. The last constant is obtained by using (for instance) $v(x) = cosh1 - coshx$ (We use the notation cosh for the hyperbolic cosine and sinh for hyperbolic sine) in Equation 178*

*The solution to this differential equation is*

$$\begin{cases} v_f(x) = sinh(1+x)/(2cosh1) & \forall x < 0 \\ v_f(x) = sinh(1-x)/(2cosh1) & \forall x > 0 \end{cases} \tag{179}$$

*and one can then check that, indeed, with this function $v_f$ we have*

$$\int_\Omega f(x)v(x)dx = v(L/2) = (v_f, v)_{H^1(\Omega)} \qquad \forall v \in H_0^1(\Omega). \tag{180}$$

*The interest of obtaining $v_f$ has been stated in Section 3. The analysis of this problem of the bar in tension can be completed in the following way. We note that the total potential energy, which takes into account the internal strain energy and the potential due to the external loads, for a displacement state described by $v \in H_0^1(\Omega)$ is simply*

$$\mathcal{E}_p(v) = \mathcal{E}(v) - \mathcal{W}_e(v) \tag{181}$$

*The solution $u$ corresponds to the minimum of the total potential energy, i.e.*

$$\delta\mathcal{E}_p(u) = 0. \tag{182}$$

*Hence, we have*

$$\forall v \in H_0^1(\Omega), \qquad \int_\Omega \frac{du}{dx}\frac{dv}{dx}dx = \int_\Omega f(x)v(x)dx \tag{183}$$

*which can be stated in the form*

$$A(u, v) = F(v). \tag{184}$$

*Since $A(u,v) = \int_\Omega \frac{du}{dx}\frac{dv}{dx}dx$ and $F(v) = \int_\Omega f(x)v(x)dx$, this is exactly the form required by the Lax-Milgram theorem presented in Section 9, which guarantees the existence and uniqueness of a solution in $H_0^1(\Omega)$.*
*Equation 184 is the form of the problem that would be used to solve this problem numerically by the finite element method. If we want to seek an analytical solution to the problem, we can rewrite Equation 183 in the following form*

$$\forall v \in H_0^1(\Omega), \qquad \int_\Omega -\frac{d^2u}{dx^2}(x)v(x)dx = \int_\Omega f(x)v(x)dx \tag{185}$$

*which is*

$$\frac{d^2u}{dx^2} = -f(x) \tag{186}$$

*understood in a distributional sense. Equation 186 can be solved using the same techniques used to solve Equation 177.*
*This example shows that the familiar results and equations concerning the bar in tension can be obtained rigorously in the framework presented in this write-up. It should be noted that we have not made use of the so-called "singularity function", but that with the help of the Lax-Milgram theorem of Section 9 we have been able to establish the existence and uniqueness of the solution of this simple problem. In the next example, we conduct the same type of analysis for a beam to illustrate that the analysis procedure employed for the bar in tension problem is in fact very general.* □

**Example 8.8** *Consider the following beam bending problem. A beam occupying the region* $x \in \Omega = ]-1,1[$ *and clamped at both ends is subjected to a loading wich involves a distributed force per unit length* $q$ *between* $x = -1$ *and* $x = 1$ *(with* $q \in L^2(\Omega)$*), a concentrated force of magnitude* $F$ *applied at* $x = 0$ *and a concentrated moment of magnitude* $M$ *applied at* $x = 0$*. We assume that the product of the Young's modulus by the moment of inertia of the beam is given by* $E(x)I(x) = 1$*. The strain energy associated with the solution* $u$ *to this problem is given by*

$$\mathcal{E}(u) = \frac{1}{2}\int_\Omega \left(\frac{d^2u}{dx^2}\right)^2 dx, \tag{187}$$

*and so we search the solution in the space* $H_0^2(\Omega)$*, so that the strain energy remains finite and the boundary conditions are satisfied.*
*The external virtual work due to an admissible displacement pattern* $v$ *can be written as the duality pairing*

$$\mathcal{W}_e(v) = \int_\Omega f(x)v(x)dx. \tag{188}$$

*Here* $f \in H^{-2}(\Omega)$ *is the distribution*

$$f = q + F\delta + MD\delta \tag{189}$$

*where by* $D\delta$ *we denote the derivative in the sense of distributions of the Dirac distribution. The distribution 189 is chosen so that the virtual work obtained from Equation 188 is equal to its known "physical" value:* $\int_0^1 q(x)v(x)dx + Fv(0) - Mv'(0)$ *for any* $v$ *in* $H_0^2(\Omega)$*. Similarly to what was done in the case of the bar in tension, we can determine, using Equation 176, the element* $v_f \in H_0^2(]-1,1[)$ *that satisfies* $\forall v \in H^2(\Omega)$, $\int_\Omega f(x)v(x) = (v_f, v)_{H_0^2(\Omega)}$ *(i.e. Riesz representation theorem) by solving*

$$K_2(v_f) = f \tag{190}$$

**More on the Dirac distribution** We say that a sequence of distributions $T_n$ converges strongly to a limit $T$ if for all $\Phi$ in $\mathcal{D}(\Omega)$, the sequence $T_n(\Phi)$ converges to $T(\Phi)$.

Let us consider again the sequence of function defined by

$$\begin{cases} f_n(x) = n & \text{if } x \in [-\frac{1}{2n}, +\frac{1}{2n}] \\ f_n(x) = 0 & \text{otherwise} \end{cases} \tag{191}$$

and an arbitrary smooth function $\Phi$, where by smooth we mean that $\Phi \in \mathcal{D}(\Omega)$. We can develop $\Phi$ in a Taylor sequence in the neighborhood of the origin:

$$\Phi(x) = \Phi(0) + x\Phi'(0) + \frac{x^2}{2}\Phi''(0) + ... \tag{192}$$

Then we have

$$\int_{\mathbb{R}} \Phi(x)f_n(x)dx = \int_{\mathbb{R}} (\Phi(0) + x\Phi'(0) + \frac{x^2}{2}\Phi''(0) + ...)f_n(x)dx = ... \tag{193}$$

$$... = \int\!\!\!\!\!\!-_{-\frac{1}{2n}}^{\frac{1}{2n}} (\Phi(0) + x\Phi'(0) + \frac{x^2}{2}\Phi''(0) + ...)ndx = \Phi(0) + \frac{1}{24n^2}\Phi''(0) + ... \tag{194}$$

Hence we have

$$\lim_{n\to\infty} \int\!\!\!\!\!\!- _{\mathbb{R}} f_n(x)\Phi(x)dx = \Phi(0) \tag{195}$$

However, since the support of $\lim_{n\to\infty} f_n(x)$ is reduced to the origin which is a set of zero measure we have

$$\int\!\!\!\!\!\!- _{\mathbb{R}} (\lim_{n\to\infty} f_n(x))\Phi(x)dx = 0 \tag{196}$$

The Dirac distribution, by definition, satisfies

$$\int\!\!\!\!\!\!- _{\mathbb{R}} \delta(x)\Phi(x)dx = \Phi(0) \tag{197}$$

where once again it must be remembered that the integral sign in Equation 197 does not have the same meaning as the integral sign in equation 196. Comparing Equation 197 and Equation 195, it is seen that the sequence of operators $T_{f_n}$ defined by $\Phi \to \int\!\!\!\!\!\!-_{\mathbb{R}} f_n(x)\Phi(x)dx$ converges strongly to the Dirac distribution.

We mentioned earlier that for any function $f \in L^1_{\text{loc}}$ the corresponding distribution $T_f$ can be identified and that as a result the function $f$ is sometimes called the distribution $f$. In that sense, one can say that the sequence defined by Equation 191 converges strongly to the Dirac distribution or that the $f_n$'s, *considered as distributions*, converge to the Dirac distribution. However, comparing Equation 196 and Equation 197 we see that it is not quite appropriate to say that the sequence of *functions* $f_n$ converges to the Dirac distribution (even though that is often said by abuse of language).

**An approximation theorem** The vector space $\mathcal{D}$ is dense in $\mathcal{D}'$, i.e. any distribution is the limit of a sequence of functions of $\mathcal{D}$.

**Example 8.9** *The Dirac distribution is the limit of the sequence of functions defined by:*

$$f_n(x) = \frac{1}{\sqrt{\pi}} n exp(-n^2x^2) \tag{198}$$

*(Note that other sequences of functions in $\mathcal{D}$ converge to the Dirac distribution.) This means that for any $\Phi \in \mathcal{D}$, we have*

$$\lim_{n\to\infty} \int\!\!\!\!\!\!- _{\Omega} f_n(x)\Phi(x)dx = \delta(\Phi) = \Phi(0) \tag{199}$$

*This is a very useful result. Say for instance we are able to show that the equality*

$$\int\!\!\!\!\!\!-_0^1 EA\frac{du}{dx}\frac{dv}{dx} + f(x)v(x)dx = 0 \tag{200}$$

*holds for all $v \in H^1([0;1])$ that satisfy $v(0) = v(1) = 0$, with $f$ being some function of class $L^2([0;1])$, $u$ is in $H^1([0;1])$ and $E$ and $A$ are two constants. This type of equation can arise in the*

*situation of a rod in tension. We would like to conclude that this implies that* $\frac{d^2u}{dx^2} \equiv -f$ *over* $]0;1[$. *Considering a point* $x_0 \in ]0;1[$, *naturally, we are tempted to use the Dirac* $\delta$ *centered at* $x_0$ *as our* $v$, *but this is not a legal pick since* $v$ *must be in* $H^1([0;1])$ *and the Dirac* $\delta$ *is not in that space. However, we can use the density of* $\mathcal{D}([0;1])$ *in* $\mathcal{D}'([0;1])$ *to argue that there exists a sequence of functions* $v_n \in \mathcal{D}([0;1])$ *that converges to the Dirac* $\delta$. *Since the functions* $v_n$ *are in* $\mathcal{D}([0;1])$ *they are in particular in* $H^1([0;1])$ *and hence they satisfy Equation 200, and also because the functions are in* $\mathcal{D}([0;1])$ *we can use the definition of weak derivatives (which amounts to integration by parts) to obtain*

$$\int_0^1 (-\frac{d^2u}{dx^2} + f(x))v_n(x)dx = 0 \tag{201}$$

*for all* $v_n$. *We now use the fact that we have picked the functions* $v_n$ *such that they converge to the Dirac* $\delta$ *shifted by* $x_0$ *(the generalized function for which is denoted* $\delta_{x_0}$*) to conclude that*

$$\lim_{n\to\infty} \int_0^1 (-\frac{d^2u}{dx^2} + f(x))v_n(x)dx = 0 = \int_0^1 (-\frac{d^2u}{dx^2} + f(x))\delta_{x_0}(x)dx = -\frac{d^2u}{dx^2}(x_0) + f(x_0) \tag{202}$$

*Therefore we conclude that for any* $x \in ]0;1[$ *we have*

$$-\frac{d^2u}{dx^2}(x) + f(x) = 0 \tag{203}$$

*The derivative in Equation 203 is a weak one and this equation has meaning only in the distributional sense in general, but if we pick a loading function* $f$ *such that the solution* $u$ *actually has a second derivative in the sense of calculus then we are assured that Equation 203 also holds in the usual sense.* □

# 9 Do we really need all of this?

Given the non-trivial character of the theory presented here, it is fair to ask whether we can do without it.

Let us consider an arbitrary problem in linear three-dimensional elasticity. The existence and uniqueness of a solution to this problem is assured by the Lax-Milgram theorem

**Definition of coercivity (or ellipticity)** A functional $A(.,.)$ defined over the space $E$ with the norm $||.||_E$ is said to be **coercive** (or **elliptic**) if there exists a strictly positive constant $c$ such that for all $u \in E$

$$A(u,u) \geq c||u||_E^2 \tag{204}$$

**Example 9.1** *The functional over* $\mathbb{R}^2$ *defined by*

$$a(x,y) = xy \tag{205}$$

*is clearly coercive (with constant* $c = 1$*), whereas the functional over* $\mathbb{R}^2$ *defined by*

$$a(x,y) = x^2y^2 \tag{206}$$

*is not coercive: for any constant c, it is possible to find a value of x small enough to make $x^4 - cx^2$ negative.*

**Theorem 9.1** *Lax-Milgram Theorem Consider a continuous, bilinear, coercive functional $A(.,.)$, and a continuous linear functional $F(.)$, both of them operating on a Hilbert space $V$ (with its scalar field being $\mathbb{R}$). Then the problem*

$$\text{Find } u \in V, \qquad A(u,v) = F(v) \qquad \forall v \in V \tag{207}$$

*has a unique solution u, and this solution is continuous with respect to the loading, i.e. there exists a constant $C$ such that*

$$||u||_V \leq C||F||_{V'} \tag{208}$$

**Note 1** It should be noted that the Lax-Milgram is very similar in nature to the Riesz representation theorem. There are some subtle differences though, including the fact the Lax-Milgram theorem does not require the form $A(.,.)$ to be symmetric.

**Note 2** When the form $A(.,.)$ is symmetric, the problem of Theorem 9.1 is equivalent to the minimization problem: Find $u \in V$ such that the functional $\Pi(v)$ defined by

$$\Pi(v) = \frac{1}{2}A(v,v) - F(v) \tag{209}$$

is minimized on $V$. In a mechanical engineering context, $\Pi(.)$ is an energy functional, such as the Hu-Washizu functional or the Hellinger-Reissner functional.

**Proof** Consider the functional $\Pi$ evaluated at $u$ and $u + w$ where $u$ and $w$ are in $V$. We have

$$\begin{array}{l} \Pi(u) = \frac{1}{2}A(u,u) - F(u) \\ \Pi(u+w) = \frac{1}{2}A(u+w,u+w) - F(u+w) \end{array} \tag{210}$$

We can use the bilinearity and symmetry of $A(.,.)$ and the linearity of $F(.)$ to obtain

$$\Pi(u+w) = \frac{1}{2}[A(u,u) + A(w,w) + 2A(u,w)] - [F(u) + F(w)] \tag{211}$$

Hence we have

$$\Pi(u+w) - \Pi(u) = \frac{1}{2}[A(w,w) + 2A(u,w)] - F(w) \tag{212}$$

Rearranging terms we have

$$\Pi(u+w) - \Pi(u) = [A(u,w) - F(w)] + \frac{1}{2}A(w,w) \tag{213}$$

Since $u$ is the solution of the problem in Theorem 9.1, the term in brackets vanishes for all $w \in V$ therefore

$$\Pi(u+w) - \Pi(u) = \frac{1}{2}A(w,w) \tag{214}$$

Because of the ellipticity of $A(.,.)$ the term on the right-hand side is always positive and can vanish only if $||v||_V$ is zero, i.e. only if $v$ is zero. Equation 214 therefore proves that the functional $\Pi$ is minimized over $V$ by the solution of the problem of Theorem 9.1.□

**Application of the Lax-Milgram in solid mechanics** The most famous application of the Lax-Milgram in mechanics is in the proof of the existence and uniqueness of a solution to the linear elasticity equations (its use in the proof of existence and uniqueness solution of various shell and plate models can be found in [3]). Let us first recall these equations:

Consider an open bounded connected [15] domain $\Omega \subset \mathbb{R}^3$, with a boundary denoted by $\partial\Omega$. Let $\partial\Omega_u$ be an open, non-empty subset of $\partial\Omega$ and let $\partial\Omega_f = \partial\Omega - \partial\Omega_u$ (i.e. $\partial\Omega = \partial\Omega_u \cup \partial\Omega_f$ and $\partial\Omega_u \cap \partial\Omega_f = \emptyset$). Let us assume that we have as a volumetric loading field (force per unit volume) $f^B = (f_i^B)_{1\leq i\leq 3} \in L^2(\Omega)^3$ and a surface loading field (force per unit area) $f^S = (f_i^S)_{1\leq i\leq 3}$ where there exist a function $g \in H^1(\Omega)^3$ such that $f^S = g|_{\partial\Omega_f}$. Finally we denote by $\lambda$ and $\mu$ the Lamé constants, related to the Young's modulus $E$, the Poisson ratio $\nu$ and the shear modulus $G$ by the equations

$$\lambda = \frac{E\nu}{(1+\nu)(1-2\nu)} \tag{215}$$

$$\mu = G = \frac{E}{2(1+\nu)}). \tag{216}$$

The traditional equations modelling the infinitesimal three-dimensional linear elastic reaction of a solid made of an isotropic material and occupying at rest the domain $\Omega$ are as follow (the Einstein summation convention is employed, i.e. repeated indices indicate summation):

$$\sigma_{ij,j}(u) = -f_i^B \qquad \text{in } \Omega \text{ for } 1 \leq i \leq 3 \tag{217}$$

$$u = 0 \qquad \text{on } \partial\Omega_u \tag{218}$$

$$\sigma_{ij}(u).n_i = f_i^S \quad \text{on } \partial\Omega_f \text{ for } 1 \leq i \leq 3 \tag{219}$$

where $n$ denotes the outward normal to $\Omega$ and $\sigma_{ij}(u) = \lambda\delta_{ij}u_{k,k} + 2\mu\epsilon_{ij}(u)$ are the stresses associated with the Green-Lagrange strains $\epsilon_{ij}$ defined in Equation 127.

The Lax-Milgram theorem applies with

$$A(u,v) = \lambda \int\!\!\!\!\!\!\int_\Omega div(u)div(v)d\Omega + 2\mu \int\!\!\!\!\!\!\int_\Omega \epsilon_{ij}(u)\epsilon_{ij}(v)d\Omega \tag{220}$$

$$F(v) = \int\!\!\!\!\!\!\int_{\partial\Omega_f} f_i^S v_i d\Omega + \int\!\!\!\!\!\!\int_\Omega f_i^B v_i d\Omega \tag{221}$$

[15] By connected we mean that the domain is "in one piece". Technically, we require that the domain is not the reunion of two non-empty open disjoint sets.

The proof that the bilinear form $A(.,.)$ given by Equation 220 is coercive is rather technical and requires the Korn inequality introduced in section 7.4

Note that the required condition on $f^S$ is needed in the proof to transform the surface integral over $\partial\Omega_f$ into a volume integral. $\square$

It is seen that one of the hypotheses is that the vector space considered is a Hilbert space, i.e. a Banach space with a norm associated with a scalar product that is complete for that norm. We have mentioned in Section 7.4 that if we do not define the Sobolev spaces based on the Lebesgue integral then completeness is not satisfied (this result was shown for $L^1([0,1])$).

Also, let us recall Cea's lemma:

**Cea's lemma** Consider a continuous, bilinear, coercive functional $A(.,.)$, and a continuous linear functional $F(.)$, both of them operating on a Hilbert space $V$ (with its scalar field being $\mathbb{R}$). Consider also $V_h$ a finite-dimensional subspace of $V$. Let us denote by $u_h$ the solution of

$$\text{Find } u_h \in V_h, \qquad A(u_h, v_h) = F(v_h) \quad \forall v_h \in V_h \tag{222}$$

(this solution exists and is unique by application of the Lax-Milgram theorem to $V_h$). Then we have the following bound:

$$||u - u_h|| \leq \inf_{v_h \in V_h} ||u - v_h|| \tag{223}$$

Again, it is seen that this crucial lemma for estimating the error of a finite element analysis requires that the problem be defined over a Hilbert space.

In finite element practice, one only uses shape functions that are smooth outside of a set of zero measure (the inter-element boundaries). These shape functions are Riemann integrable, and their integrals are evaluated approximately by using Gauss quadrature. So, the functions that we use and need to integrate in finite element analysis and continuum mechanics can be integrated with either the Riemann integration or Lebesgue integration and the same results are obtained. However, the Lebesgue integration theory has specific properties that the Riemann integration theory does not have and which make it the right framework to study mechanical engineering and finite element techniques.

## References


[1] J.-M. Bony. *Méthodes mathématiques pour les sciences physiques.* Presses de l'Ecole Polytechnique, 1995.

[2] S.C. Brenner and L.R. Scott. *The mathematical theory of finite element methods (MIT call number:QA379.B745).* Springer-Verlag, 1994.

[3] D. Chapelle and K.J. Bathe. *The finite element analysis of shells.*

[4] L. Debnath and P. Mikusinski. *Introduction to Hilbert spaces with applications (MIT call number:QA322.4.D43).* Academic Press, 1990.

[5] P.R. Halmos. *Measure theory (MIT call number:QA312.H26).* Springer-Verlag, 1974.

[6] L. Schwartz. *Théorie des distributions (MIT call number:QA324.S399).* Hermann, 1961.

[7] L. Schwartz. *Cours d'analyse (MIT call number:QA300.S399).* Presses de l'Ecole Polytechnique, 1965.

[8] L. Schwartz. *Théorie de l'intégration (MIT call number:QA300.S383).* Presses de l'Ecole Polytechnique, 1975.

[9] A.H. Zemanian. *Distribution Theory and transform analysis (MIT call number:QA320.Z53).* McGraw-Hill, 1965.

# Index

absolute convergence in a vector space, 9
application, 2

Banach space, 15
bar in tension, 44, 47
beam bending, 49
bidual, 17
boundedness, 3

Cauchy sequence, 11
Cea's lemma, 54
closed ball, 3
closed interval, 22
closed set, 3
coercivity, 51
compact embedding, 36
compact set, 5
complete space, 12
connected set, 53
consistent nodal forces, 44
continuity, 9
convergence, 4
convergence in a vector space, 9
countable set, 23

density, 14
derivative in the sense of calculus, 9
derivative in the sense of distributions, 31
derivative of a distribution, 42
Dirac distribution, 44
distribution, 42
dominated convergence theorem, 27
dual space, 16
duality pairing, 16, 45

ellipticity, 51
equality almost everywhere (a.e.), 30
equivalence relation, 1
equivalent norms, 9
Euclidian metric, 2

functionals, 52

generalized function, 44

Green's formula, 36

Heaviside function, 46
Hilbert space, 15
homogeneous Sobolev spaces, 35

inner product, 7
isometry, 17

Korn's inequality, 35

Lax-Milgram theorem, 52
Lebesgue integrable function, 25
Lebesgue measure, 22
Lebesgue spaces, 29
Lipschitz boundary, 37
Lipschitz function, 36
Lipschitz norm, 36
Lipschitz-continuity, 36

measurable function, 25
metric, 2
metric space, 2
minimization problem, 52
momotone convergence theorem, 27

norm, 7
norm $L^\infty$, 36
norms $H^p$, 33
norms $L^p$, 29

open ball, 3
open set, 3

pavable set, 22
Poincaré-Friedrichs inequality, 35
pointwise convergence, 10
product of sets, 1
product space norm, 7
product space scalar product, 7

quadratic form, 8

reciprocal image of a set, 25
reflexive relation, 1

reflexive space, 17
Rellich-Kondrachov theorem, 36
Riemann integral, 18
Riesz representation theorem, 16, 48
rod, 44, 47

scalar field, 6
scalar product, 7
scalar product for the space $L^2(\Omega)$, 30
scalar products for the spaces $H^p(\Omega)$, 33
Schwarz's inequality, 8
second dual, 17
semi-norm, 8
set, 1
Sobolev spaces $H^p$, 33
Sobolev's inequality, 38
space $\mathcal{D}(\Omega)$, 29
space of distributions $\mathcal{D}'(\Omega)$, 42
spaces $C^0$, $C^1$, etc, 9
spaces $H^p$, 33
spaces $L^p$, 30
step function, 25
strong convergence, 4
symmetric relation, 2

transitive relation, 2

uniform convergence, 11
uniform norm, 10

vector space, 6
vector subspace, 6

weak convergence, 15
weak derivative, 31