CMMS 11 FOURIER ANALYSIS

Yu. Safarov

Notation

If c_k, $k = 1, 2, \dots$, are real numbers then

$$\limsup c_k = \lim_{m\to\infty} \sup_{k\geqslant m} \{c_k\}\,, \qquad \liminf c_k = \lim_{m\to\infty} \inf_{k\geqslant m} \{c_k\}\,,$$

$\varnothing$ is the empty set;
$\{x_1, x_2, \dots\}$ denotes the set with the elements $x_1, x_2, \dots$;
$\mathbb{N} = 1, 2, 3, \dots$ is the set of positive interges;
$\mathbb{R}$ is the set of real numbers;
$\mathbb{R}_+$ is the set of nonnegative real numbers, that is, A is a subset of B;
∞ is the symbol of infinity;
$\hat{\mathbb{R}}_+$ is the set $\mathbb{R}_+ \cup \{+\infty\}$;
$\hat{\mathbb{R}}$ is the set $\mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$.

If A and B are some sets then
$A \cup B = \{x \,:\, x \in A \text{ or } x \in B\}$ is the union of A and B;
$A \cap B = \{x \,:\, x \in A \text{ and } x \in B\}$ is the intersection of A and B;
$A \subset B$ means that $x \in A$ implies $x \in B$, that is, A is a subset of B;
$B \supset A$ is the same as $A \subset B$;
$B \setminus A = \{x \,:\, x \in B \text{ and } x \notin A\}$;
2^A is the set of all subsets of A.

We shall always fix a set E (usually $E = \mathbb{R}^n$) and deal with subsets B of E and complex functions f defined on E. Then
$B^c = E \setminus B$ is the complement of B in E.
If E is a metric space then
$\overline{B} = \{x \in \mathbb{R}^n \,:\, \operatorname{dist}(x, B) = 0\}$ is the closure (completion) of the set B.
The support of a function f on E is defined as the closure of the set $\{x \in E \,:\, f(x) \neq 0\}$.

For a real function f we define

$$f_+(x) = \max\{f(x), 0\} \quad\text{and}\quad f_-(x) = \max\{-f(x), 0\}.$$

Then f_+ and f_- are non-negative functions such that $f = f_+ - f_-$ and $f_+\, f_- \equiv 0$.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

1. Measures and measure spaces

σ-algebras.

Let E be an arbitrary set, and $\mathcal{A}$ be a set of subsets of E.

Definition 1.1. The set $\mathcal{A}$ is said to be a σ-algebra over E if

(1) $\varnothing \in \mathcal{A}$ and $E \in \mathcal{A}$;
(2) if $B \in \mathcal{A}$ then $B^c \in \mathcal{A}$;
(3) if $B_k \in \mathcal{A}$, $k = 1, \dots,$ then $\cup_{k=1}^{\infty} B_k \in \mathcal{A}$.

Since $\cap_{k=1}^{\infty} B_k = (\cup_{k=1}^{\infty} B_k^c)^c$ and $B_1 \setminus B_2 = B_1 \cap B_2^c$, (1)–(3) imply

(4) if $B_k \in \mathcal{A}$, $k = 0, 1, \dots,$ then $\cap_{k=1}^{\infty} B_k \in \mathcal{A}$;
(5) if $B_1 \in \mathcal{A}$ and $B_2 \in \mathcal{A}$ then $B_1 \setminus B_2 \in \mathcal{A}$.

The conditions (3) and (4) mean that σ-algebra is closed under countable unions and intersections. Since we can take $B_k = \varnothing$ for $k \geqslant N$, a σ-algebra contains also all the finite unions and intersections of its elements.

The set $\mathcal{A}$ which contains only $\varnothing$ and E is a σ-algebra over E. The set $\mathcal{A}$ containing all the subsets of E is also a σ-algebra over E. Of course, most of interesting σ-algebras lie somewhere in between these two extreme examples.

Examples 1.2.

(1) Let $E = \{a, b, c, d\}$. Then the set $\mathcal{A} = \{\varnothing, \{a, b\}, \{c, d\}, E\}$ is a σ-algebra.
(2) Let $E = \{a, b, c\}$ and $\mathcal{A} = \{\varnothing, \{a\}, E\}$. Then $\mathcal{A}$ is not a σ-algebra. Indeed, the set $(\{a\})^c = \{b, c\}$ is not an element of $\mathcal{A}$.
(3) Let $E = \mathbb{N}$ and $\mathcal{A} = N_f \cup N_c$, where N_f is the set of all finite subsets of $\mathbb{N}$ and the set $N_c \subset 2^{\mathbb{N}}$ is defined as follows:

$$X \in N_c \text{ if and only if } X^c = \mathbb{N} \setminus X \in N_f.$$

Then $\mathcal{A}$ is not a σ-algebra over E. The main peculiarity of this example is that the set $\mathcal{A}$ satisfies the conditions (1), (2) of our definition and is closed under finite unions and intersections. However, $\mathcal{A}$ does not satisfy the last condition (3). Indeed, the set $B = \{2, 4, 6, 8, \dots\}$ is the union of elements $\{2k\} \in \mathbb{N}_f$ but $B \notin N_f$ and $B \notin N_c$.

Exercise 1.3. Let $\mathcal{A}_i$, $i \in I$, be some σ-algebras over E where I is some set (not necessarily countable). Prove that $\cap_{i \in I} \mathcal{A}_i$ is a σ-algebra over E.

Let $\mathcal{A}_0$ be a set of subsets of E. Since 2^E is a σ-algebra, it is clear that there always exists at least one σ-algebra containing all the sets from $\mathcal{A}_0$.

Definition 1.4. Denote by $\mathcal{E}(\mathcal{A}_0)$ the intersection of all σ-algebras over E which contain all the sets from $\mathcal{A}_0$. Then $\mathcal{E}(\mathcal{A}_0)$ is a σ-algebra (see Exercise 1.3) which is called the *σ-algebra generated by* $\mathcal{A}_0$.

In other words, $\mathcal{E}(\mathcal{A}_0)$ is the minimal σ-algebra over E which contains all the sets from $\mathcal{A}_0$.

Example 1.5. Let $E = \{a, b, c\}$, $\mathcal{A}_0 = \{\{a\}\}$. Then $\mathcal{E}(\mathcal{A}_0) = \{\varnothing, E, \{a\}, \{b, c\}\}$. Indeed, since the set $\mathcal{A}_1 = \{\varnothing, E, \{a\}, \{b, c\}\}$ is a σ-algebra which contains {a} inclusion $\mathcal{E}(\mathcal{A}_0) \subset \mathcal{A}_1$ holds. Since $\mathcal{E}(\mathcal{A}_0)$ is a σ-algebra and $\{a\} \in \mathcal{E}(\mathcal{A}_0)$ we have $\{\varnothing, E\} \subset \mathcal{A}_0$, $\{b, c\} = (\{a\})^c = E \setminus \{a\}$ and hence $\mathcal{A}_1 \subset \mathcal{A}_0$.

Definition 1.6. Let E be a metric space. The *Borel σ-algebra over E* is the σ-algebra generated by all open sets. We will denote this σ-algebra by $\mathcal{B}(E)$. The sets $B \in \mathcal{B}(E)$ are said to be the *Borel sets*.

Example 1.7.

(1) Obviously, every open or closed set is a Borel set.
(2) Since any countable set $\{x_1, x_2, \dots\}$ coincides with the union of closed sets $\{x_k\}$, the countable sets are also Borel sets.

Definition of a measure space. Complete measure spaces.

The reason for introducing σ-algebras is that they are natural objects on which we can define measures.

Definition 1.8. The map $\mu : \mathcal{A} \to \hat{\mathbb{R}}_+$ is a *measure* if

(1) $\mu(\varnothing) = 0$;
(2) μ is countably additive, that is $\mu\left(\cup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^{\infty} \mu(B_k)$ for all $B_k \in \mathcal{A}$ such that $B_k \cap B_m = \varnothing\,,\ m \neq k$.

If $\mu(E) < \infty$ then μ is said to be a *finite measure*. If $\mu(E) = 1$ then μ is called a *probability measure*. The triple $(E, \mathcal{A}, \mu)$, where E is an arbitrary set, $\mathcal{A}$ is a σ-algebra over E and μ is a measure on $\mathcal{A}$, is called a *measure space*.

Example 1.9. Let $\mathcal{A}$ be the set of all subsets of E.

(1) Let us fix a point $x \in E$ and define the map μ as follows:

$$\mu(B) \;=\; \begin{cases} 1, & \text{if } x \in B, \\ 0, & \text{if } x \notin B. \end{cases}$$

Then μ is a measure. This measure is called the *δ-measure* at the point x and is denoted δ_x .
(2) Let $\mu(B)$ coincides with the number of elements of the set B if B is finite set and $\mu(B) = +\infty$ if B is infinite set. Then μ is a measure. This measure is called the *counting measure*.

Definition 1.10. The measure space $(E, \mathcal{A}, \mu)$ is said to be complete if every subset B_1 of every set $B \in \mathcal{A}$ with $\mu(B) = 0$ also belongs to $\mathcal{A}$ (and then, by Lemma 1.12, $\mu(B_1) = 0$).

Examples 1.11.

(1) If $\mathcal{A} = 2^E$ (as in Example 1.9) then the measure space $(E, \mathcal{A}, \mu)$ is always complete.
(2) Let $E = \{a, b\}$, $\mathcal{A} = \{\varnothing, E\}$, and $\mu(\varnothing) = \mu(E) = 0$. Then the measure space $(E, \mathcal{A}, \mu)$ is not complete.

Lemma 1.12. *Let $(E, \mathcal{A}, \mu)$ be a measure space and let $A, B, B_1, B_2, \dots$ be elements of $\mathcal{A}$. Then the following statements hold true.*

(1) *If $A \subset B$ then $\mu(A) \leqslant \mu(B)$.*
(2) $\mu(\cup_{k=1}^{\infty} B_k) \leqslant \sum_{k=1}^{\infty} \mu(B_k)$.
(3) *If $B_1 \supset B_2 \supset B_3 \supset \dots$, $\cap_{k=1}^{\infty} B_k = \varnothing$, and $\mu(B_1) < \infty$, then $\mu(B_N) \to 0$ as $N \to \infty$.*

Proof.

(1) Since $B = A \cup (B \setminus A)$, $A \cap (B \setminus A) = \varnothing$ and the function μ is nonnegative, we obtain $\mu(B) = \mu(A) + \mu(B \setminus A) \geqslant \mu(A)$.

(2) Since

$$\cup_{k=1}^{\infty} B_k = (B_1) \cup (B_2 \setminus B_1) \cup (B_3 \setminus (B_1 \cup B_2)) \cup \ldots (B_{k+1} \setminus \cup_{m=1}^{m=k} B_m) \ldots$$

and the sets

$$B_1, \quad B_2 \setminus B_1, \quad B_3 \setminus (B_1 \cup B_2), \quad \ldots \quad B_{k+1} \setminus \cup_{m=1}^{m=k} B_m, \quad \ldots$$

are mutually disjoint, from countable additivity of the measure μ it follows that

$$\mu(\cup_{k=1}^{\infty} B_k) = \mu(B_1) + \mu(B_2 \setminus B_1) + \mu(B_3 \setminus (B_1 \cup B_2) + \ldots \\ + \mu(B_{k+1} \setminus \cup_{m=1}^{m=k} B_m) + \ldots$$

Now the required inequality follows from the first statement.

(3) Since $\cap_{k=1}^{\infty} B_k = \varnothing$, we have

$$B_m \;=\; (B_m \setminus B_{m+1}) \cup (B_{m+1} \setminus B_{m+2}) \cup (B_{m+2} \setminus B_{m+3}) \ldots$$

for all $m \in \mathbb{N}$. The sets $B_k \setminus B_{k+1}$ are mutually disjoint, and therefore

$$\mu(B_m) \;=\; \sum_{k=m}^{\infty} \mu\,(B_k \setminus B_{k+1}) \tag{1.1}$$

for all m. In particular, from this equality and the first statement of the theorem it follows that the series $\sum_{k=1}^{\infty} \mu\,(B_k \setminus B_{k+1})$ converges, which implies that (1.1) vanishes as $N \to \infty$. $\square$

2. Lebesgue's measure

Outer measure.

Definition 2.1. Let $(E, \mathcal{A}, \mu)$ be a measure space. The map $\nu : \mathcal{A} \to \hat{\mathbb{R}}_+$ is a outer measure if

(1) $\nu(\varnothing) = 0$;
(2) the inequality $\nu\left(\cup_{k=1}^{\infty} B_k\right) \leqslant \sum_{k=1}^{\infty} \nu(B_k)$ holds for all $B_k \in \mathcal{A}$.

Lemma 1.12 implies that every measure is a outer measure.

Example 2.2. Let $\mathcal{A}$ be the set of all subsets of E. Define the map μ as follows:

$$\mu(B) \;=\; \begin{cases} 1, & \text{if } B \neq \varnothing, \\ 0, & \text{if } B = \varnothing. \end{cases}$$

Then μ is an outer measure. One can show that μ is not a measure if E contains more than one element.

Definition 2.3. If a set $P \subset \mathbb{R}^n$ coincides with the direct product of some open intervals $(a_1, b_1), (a_2, b_2), \dots (a_n, b_n)$ then we say that $P \subset \mathbb{R}^n$ is an open rectangle and denote

$$\mathrm{Vol}_n(P) = (b_1 - a_1)(b_2 - a_2) \dots (b_n - a_n).$$

Definition 2.4. Let $B \subset \mathbb{R}^n$. Denote

$$\mu_{out}(B) = \inf \sum_{k=1}^{\infty} \mathrm{Vol}_n(P_k)\,,$$

where the infimum is taken over all countable families of rectangles $\{P_k\}_{k \in \mathbb{N}}$ such that $B \subset \cup_{k \in \mathbb{N}} P_k$.

Definition of Lebesgue's measure.

Theorem 2.5.

(1) *The map μ_{out} defined on the σ-algebra $2^{\mathbb{R}^n}$ (that is, on the σ-algebra of all subsets of $\mathbb{R}^n$) is an outer measure.*
(2) *For any rectangle P the equality $\mu_{out}(P) = \mathrm{Vol}_n(P)$ holds.*
(3) *Let $\mathcal{B}(\mathbb{R}^n)$ be the Borel σ-algebra, $\mathcal{N}$ the family of sets $B \subset \mathbb{R}^n$ such that $\mu_{out}(B) = 0$, and $\mathcal{L}_n$ be the σ-algebra generated by $\mathcal{N} \cup \mathcal{B}(\mathbb{R}^n)$. Let μ_n be the restriction of μ_{out} to $\mathcal{L}_n$. Then μ_n is a measure and the measure space $(\mathbb{R}^n, \mathcal{L}_n, \mu_n)$ is complete.*

Defintion 2.6. The measure μ_n defined in Theorem 2.5 is called the *Lebesgue measure*. The sets $B \subset \mathcal{B}(\mathbb{R}^n)$ are said to be *Lebesgue measurable*.

Of course, Lebesgue's measure is the most natural measure in $\mathbb{R}^n$. It is invariant with respect to Euclidean transformations (translations and rotations). Under the additional assumption that the measure of the unit cube is equal to 1, it is the only measure (defined on the σ-algebra $\mathcal{L}_n$) which has this property.

The Cantor set.

It is clear that any countable set has Lebesgue measure zero. However, it is not immediately clear that there are uncountable subsets of $\mathbb{R}^1$ whose Lebesgue measure is zero. In this subsection we construct such a set. Namely, we start with the set $B_0 = [0,1]$ and let C_1 be the set obtained by removing the open middle third of C_0;

$$[--------------(\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots..)--------------]$$

$$C_1$$

(i.e., $C_1 = C_0 \setminus (1/3, 2/3)$). Next, let C_2 be the set obtained from C_1 after removing the open middle third of each of the (two) intervals of which C_1 is disjoint union;

$$[----(\cdots\cdots..)----(\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots..)----(\cdots\cdots..)----]$$

$$C_2$$

More generally, given C_k which is a union of 2^k disjoint, closed intervals, let C_{k+1} be the set which one gets by removing the open middle third of each of these closed intervals. Finally, set $C = \cap_{k=0}^{\infty} C_k$.

The set C is called the Cantor set. It is closed, and does not contain any isolated points (i.e., any neighbourhood U of a point $x \in C$ contains also other points from C). There is a one-to-one correspondence between the points from C and the sequences of 0 and 1 (the first k numbers in the sequence indicate the closed interval in C_k which contains x). Therefore C is uncountable.

The measure of $C_0 \setminus C$ is equal to

$$\frac{1}{3} + \frac{1}{3}\cdot\frac{2}{3} + \frac{1}{3}\cdot\left(\frac{2}{3}\right)^2 + \frac{1}{3}\cdot\left(\frac{2}{3}\right)^3 + \ldots = \frac{1}{3}\cdot\frac{1}{1-2/3} = 1\,,$$

so the measure C is equal to zero.

3. Measurable functions

In this section we fix a σ-algebra $\mathcal{A}$ over the set E and say that a set is measurable if it belongs to our σ-algebra $\mathcal{A}$.

Definition 3.1. We say that the real function $f : E \to \hat{\mathbb{R}}$ is *measurable* (or, more precisely, that f is a measurable function with respect to the σ-algebra $\mathcal{A}$) if for any real a the inverse image

$$f^{-1}([-\infty, a)) = \{x \in \mathbb{R}^n : f(x) < a\}$$

is an element of $\mathcal{A}$.

In the case where $E = \mathbb{R}^n$ and $\mathcal{A} = \mathcal{L}_n$ we say that f is Lebesgue measurable, and in the case where E is a metric space and $\mathcal{A} = \mathcal{B}(E)$ we say that f is Borel measurable (or, simply, that f is a Borel function). From inclusion $\mathcal{B}(\mathbb{R}^n) \subset \mathcal{L}_n$ it follows that any Borel function $f : \mathbb{R}^n \to \hat{\mathbb{R}}$ is Lebesgue measurable.

Remark 3.2. A measurable function can be infinite at some points, and such points also form a measurable set. Indeed,

$$f^{-1}(-\infty) = \cap_{k=1}^{\infty} f^{-1}([-\infty, -k)) , \qquad f^{-1}(+\infty) = \cup_{k=1}^{\infty} (f^{-1}([-\infty, k)))^c .$$

Example 3.3. Let E be a metric space and let f be a continuous functon. Then f is a Borel function. Indeed for any $a \in \mathbb{R}^1$ the set $f^{-1}((-\infty, a))$ is open and hence is a Borel set.

Theorem 3.4. *For any function $f : E \to \hat{\mathbb{R}}$ the following two conditions are equivalent:*

(1) *f is measurable*
(2) *the inverse images of all intervals (including degenerate and infinite intervals) are elements of $\mathcal{A}$.*

Simple functions.

From the measure-theoretic standpoint, the most elementary functions are those taking only a finite number of values.

Definition 3.5. The function $\varphi : E \to \mathbb{R}^1$ is said to be simple if it takes a finite number of values a_k and for each a_k the inverse image $\varphi^{-1}(a_k)$ is a measurable set.

One can easily prove the following lemma.

Lemma 3.6.

(1) *A simple function is measurable.*
(2) *Let $\psi_1, \psi_2, \dots \psi_m$ be simple functions and let $F : \mathbb{R}^m \to \mathbb{R}^1$ be a continuous function. Then the function $F(\psi_1, \psi_2, \dots \psi_m)$ is a simple function.*

Lemma 3.6 immediately implies

Corollary 3.7. *Let $\psi_1, \psi_2, \dots \psi_m$ be simple functions and let $a_1, a_2, \dots a_m$ be real numbers. Then $\sum_{k=1}^{m} a_k \psi_k$ is a simple function.*

In other words, Corollary 3.7 says that a linear combination of simple functions is a simple function.

The following lemma implies that every measurable function can be approximated by simple functions.

Lemma 3.8. *If f is measurable then there exists a sequence of simple functions ψ_k such that $\psi_k(x) \to f(x)$ for each x as $k \to \infty$. Moreover, one can choose ψ_k in a such way that for all x and k the estimates*

$$(\psi_k)_+(x) \leqslant (\psi_{k+1})_+(x) \leqslant f_+(x), \qquad (\psi_k)_-(x) \leqslant (\psi_{k+1})_-(x) \leqslant f_-(x)$$

hold. Furthermore, if the function f is bounded then we can choose a sequence ψ_k, $k = 1, 2, 3, \ldots$ which converges to f uniformly on E.

Proof. Since $f = f_+ - f_-$, it is sufficient to prove the lemma assuming that f and ψ_k are non-negative (then the general result is obtained by considering the linear combinations of the simple functions corresponding to f_+ and f_-).

Let f be non-negative. We define

$$\varphi_k(x) \;=\; \max_{m\in\mathbb{Z}_+} \{m/2^k \;:\; m/2^k \leqslant f(x)\} \tag{4.1}$$

where $\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$; in other words

$$\varphi_k(x) = m/2^k \quad \text{if} \quad m/2^k \leqslant f(x) < (m+1)/2^k\,, \quad m = 0, 1, 2, \ldots$$

Obviously, $0 \leqslant \varphi_k(x) \leqslant \varphi_{k+1}(x) \leqslant f(x)$. Since f is measurable, the inverse images $\varphi_k^{-1}(m/2^k) = f^{-1}\left([m/2^k, (m+1)/2^k)\right)$ are measurable sets.

If f is bounded then every function φ_k can take only a finite number of values $0\,,\ 1/2^k\,,\ \ldots\,,\ m_0/2^k$ (where $m_0/2^k < \sup f$). Thus, $\varphi_k^{(R)}$ are simple functions, and

$$0 \;\leqslant\; \varphi_k(x) \;\leqslant\; \varphi_{k+1}(x) \;\leqslant\; f(x)$$

for all x and k. We have $|f(x) - \varphi_k(x)| \leqslant 2^{-k}$, so $\varphi_k \to f$ uniformly as $k \to \infty$.

Let f be unbounded. We define

$$\psi_k(x) = \begin{cases} \varphi_k(x)\,, & \varphi_k(x) \leqslant k\,, \\ k\,, & \varphi_k(x) > k\,. \end{cases}$$

Then ψ_k are simple functions and

$$0 \;\leqslant\; \psi_k(x) \;\leqslant\; \psi_{k+1}(x) \;\leqslant\; f(x)\,, \qquad \forall\, x, k\,.$$

Moreover, $\psi_k(x) \to f(x)$ for each fixed x. Indeed,
if $f(x)$ is finite then $\psi_k(x) = \varphi_k(x)$ for sufficiently large k, so $\psi_k(x) \to f(x)$;
if $f(x)$ is infinite then, obviously, $\psi_k(x) \to +\infty$.
The proof is complete. □

A convergence theorem.

We shall deduce all the basic properties of the measurable functions from the following theorem.

Theorem 3.9. *Let $\varphi_1, \varphi_2, \dots$ be a sequence of measurable functions. Then*

(1) $g_1(x) = \sup_k \{\varphi_k(x)\}$ *and* $g_2(x) = \inf_k \{\varphi_k(x)\}$ *are measurable functions;*
(2) $f_1(x) = \limsup \varphi_k(x)$ *and* $f_2(x) = \liminf \varphi_k(x)$ *are measurable functions;*
(3) *if* $\varphi_k(x) \to f(x)$ *for every* x *as* $k \to \infty$ *then* f *is measurable.*

Proof.
(1) If $a \in \mathbb{R}^1$ then from definitions of supremum and infimum it follows that

$$\begin{aligned} g_1^{-1}(-\infty, a] &= \cap_k \varphi_k^{-1}(-\infty, a] \\ g_2^{-1}[a, +\infty) &= \cap_k \varphi_k^{-1}([a, +\infty). \end{aligned}$$

Now the required statement follows from Theorem 3.4 and the definition of mesurability.

(2) is an easy consequence of (1) and equalities

$$\limsup \varphi_k(x) = \inf_k \sup_{m \geqslant k} \{\varphi_m(x)\}, \qquad \liminf \varphi_k(x) = \sup_k \inf_{m \geqslant k} \{\varphi_m(x)\}.$$

(3) immediately follows from (2). □

Properties of measurable functions.

Lemma 3.8 and Theorem 3.9 imply that a function f is measurable if and only if f is the pointwise limit of a sequence of simple functions. Therefore we can obtain various results concerning the measurable functions by proving them first for the simple functions and then taking the pointwise limit. In particular, we obtain

Theorem 3.10. *Let $F(x_1, \dots, x_p)$ be a continuous function of p real variables. Then for any measurable functions $f_1, \dots, f_p$ the function $F(f_1, \dots, f_p)$ is measurable.*

Proof. From lemma 3.8 it follows that for every m there exists a sequence of simple functions $\psi_{m,k}$ such that $\psi_{m,k}(x) \to f_m(x)$ for each x as $k \to \infty$. Lemma 3.6 implies that for any $k \in \mathbb{N}$ the function

$$F_k(x) = F\left(\varphi_{1,k}(x), \dots, \varphi_{p,k}(x)\right)$$

is simple. Since F is continuous, we have

$$F\left(f_1(x), \dots, f_p(x)\right) = \lim_{k \to \infty} F\left(\varphi_{1,k}(x), \dots, \varphi_{p,k}(x)\right) = \lim_{k \to \infty} F_k(x)$$

for all x. Thus, $F(f_1, \dots, f_p)$ is the pointwise limit of a sequence of simple functions and therefore $F(f_1, \dots, f_p)$ is measurable. □

Theorem 3.10 immediately implies

Corollary 3.11. *If the functions f and g are measurable then $|f|$, $f + g$ and fg are also measurable.*

Complex-valued functions.

A complex-valued function is said to be measurable (simple) if its real and complex parts are measurable (simple) functions. Clearly, all results of this section can be easily extended to the complex-valued functions.

4. Construction of integrals

In this section we fix a measure space $(E, \mathcal{A}, \mu)$.

Definition 4.1. Denote by $\Sigma_0(E, \mathcal{A}, \mu)$ (or simply by Σ_0) the set of all simple functions φ such that $\mu(\{x \in E : \varphi(x) \neq 0\} < \infty$.

Definition 4.2. If $\varphi \in \Sigma_0$ then we define

$$\int_E \varphi \, d\mu \;=\; \sum_{k=1}^{m} a_k \, \mu(\Omega_k)$$

where $a_1, a_2, \ldots, a_m$ are the nonzero values of the function f and $\Omega_k = \varphi^{-1}(a_k) = \{x \in E : \varphi(x) = a_k\}$.

Example 4.3. Let $(E, \mathcal{A}, \mu) = (\mathbb{R}^1, \mathcal{L}_1, \mu_1)$ and

$$\varphi(x) \;=\; \begin{cases} 0, & \text{if } x \in (-\infty, -1), \\ 1, & \text{if } x \in [-1, 1], \\ 3, & \text{if } x \in (1, 2), \\ 0, & \text{if } x \in [2, +\infty). \end{cases}$$

Then the function φ belongs the set Σ_0 and

$$\int_{\mathbb{R}^1} \varphi \, d\mu_1 = \mu_1[-1, 1] + 3\mu_1(1, 2) = 2 + 3 = 5\,.$$

Lemma 4.4. *Let $f \in \Sigma_0$ and $g \in \Sigma_0$. Then for any (real) constants α and β*

$$\int (\alpha f + \beta g) \, d\mu \;=\; \alpha \int f \, d\mu \;+\; \beta \int g \, d\mu\,.$$

Proof. Obviously, $\int \alpha f \, d\mu = \alpha \int f \, d\mu$ for all simple functions f and $\alpha \in \mathbb{R}^1$. Therefore it is sufficient to prove that

$$\int (f + g) \, d\mu \;=\; \int f \, d\mu \;+\; \int g \, d\mu\,. \tag{5.1}$$

Assume that f and g take the values $a_1, \ldots, a_l$ and $b_1, \ldots, b_p$ respectively and denote $f^{-1}(a_k) = \Omega_k$, $g^{-1}(b_m) = \widetilde{\Omega}_m$. Then $f + g$ is a simple function taking the values $a_k + b_m$ on the sets $\Omega_k \cap \widetilde{\Omega}_m$, and by definition

$$\begin{aligned} \int (f + g) \, d\mu \;&=\; \sum_{k,m} (a_k + b_m) \, \mu(\Omega_k \cap \widetilde{\Omega}_m) \\ &= \sum_{k,m} a_k \, \mu(\Omega_k \cap \widetilde{\Omega}_m) \;+\; \sum_{k,m} b_m \, \mu(\Omega_k \cap \widetilde{\Omega}_m)\,. \end{aligned} \tag{5.2}$$

Since the sets $\widetilde{\Omega}_m$ form an exact non-overlapping cover of $\mathbb{R}^n$,

$$\sum_m \mu(\Omega \cap \widetilde{\Omega}_m) = \mu(\Omega)$$

for all Borel sets Ω. Therefore the first term in the right hand side of (5.2) coincides with

$$\sum_k a_k\,\mu(\Omega_k) \;=\; \int f\,d\mu\,.$$

By analogy, the second term in the right hand side of (5.2) is equal to

$$\sum_m a_m\,\mu(\widetilde{\Omega}_m) \;=\; \int g\,d\mu\,.$$

This implies (5.1). $\square$

Definition of integral.

Definition 4.5. Let f be a non-negative measurable function. We define

$$\int_E f\,d\mu \;= \sup \int \varphi\,d\mu\,,$$

where the supremum is taken over all functions $\varphi \in \Sigma_0$ such that $0 \leqslant \varphi \leqslant f$. If E_0 is a measurable subset of E then we define $\int_{E_0} f\,d\mu = \int_E f_0\,d\mu$, where

$$f_0(x) = \begin{cases} f(x), & x \in E_0, \\ 0, & x \notin E_0. \end{cases}$$

Further on in this section we shall always assume that the integrals are taken over E and write $\int f\,d\mu$ instead of $\int_E f\,d\mu$.

One can easily prove that Definition 4.5 does not contradict to Definition 4.2, that is, both definitions give the same result in the case where f is a non-negative function from Σ_0.

Definition 4.6. Let f be a measurable function. We define

(1) if $\int f_+\,d\mu = +\infty$ and $\int f_-\,d\mu < \infty$ then $\int f\,d\mu = +\infty\,$;
(2) if $\int f_+\,d\mu < \infty$ and $\int f_-\,d\mu = +\infty$ then $\int f\,d\mu = -\infty\,$;
(3) if $\int f_+\,d\mu < \infty$ and $\int f_-\,d\mu < \infty$ then $\int f\,d\mu = \int f_+\,d\mu \; - \int f_-\,d\mu$.

In the last case the function f is said to be integrable. The class of integrable functions is denoted by $L_1(E,\mathcal{A},\mu)$ (or simply L_1).

Remark 4.7. *Note that a measurable function f belongs to L_1 if and only if $\int |f|\,d\mu < \infty$.*

Theorem 4.8 (elementary properties of the integral). *Let f and g be integrable functions. Then*

(1) $\int f\,d\mu \;\leqslant \int g\,d\mu$ *whenever* $f \leqslant g$;
(2) $|\int f\,d\mu| \;\leqslant \int |f|\,d\mu$;
(3) $\int af\,d\mu \;= a\int f\,d\mu$ *for every real constant* a;
(4) $\int f\,d\mu \;+ \int g\,d\mu \;= \int f + g\,d\mu$.

Proof. The first three results immediately follow from the definition. The last result is less obvious. It is deduced from Lemma 4.4 with the use of Lemma 3.8 and Theorem 5.2 □

The above definition of integral is due to Lebesgue. Note that Lebesgue's integral of a non-negative measurable function always exists (and may be infinite). But that is not true in the general case. If $\int f_+ \, d\mu = \infty$ and $\int f_- \, d\mu = \infty$ then we cannot say anything about the value of $\int f \, d\mu$.

Example 4.9. Let $f(x) = x^{-1} \sin x$, $x \in \mathbb{R}^1$, and μ be Lebesgue's measure on $\mathbb{R}^1$. Then $\int_1^\infty f_+ \, d\mu = \infty$ and $\int_1^\infty f_- \, d\mu = \infty$, so Lebesgue's integral $\int_1^\infty f \, d\mu$ does not exist. But the Riemann integral is well-defined and finite (this can be easily proved by integrating by parts).

In a sense, Lebesgue's definition of integrals is more restrictive as far as the behaviour at infinity is concerned. On the other hand, it allows us to integrate very non-smooth functions, for which the Riemann integral does not exist. Moreover, it works for an arbitrary measure μ whereas the Riemann integral is defined only for Lebesgue's measure.

Example 4.10. Let μ be the δ-measure at the point x (Example 1.9). Then $\int f \, d\mu = f(x)$.

5. Convergence of integrals

Monotone convergence theorem.

In this section we fix a measure space $(E, \mathcal{A}, \mu)$ and assume that it is complete.

We shall need the following technical lemma.

Lemma 5.1. *Let $\varphi \in \Sigma_0$ be a non-negative simple function and f_k, $k \in 1, 2, 3 \ldots$, be non-negative measurable functions such that*

$$f_k(x) \leqslant f_{k+1}(x)\,, \qquad \lim_{k\to\infty} f_k(x) \geqslant \varphi(x)$$

for all $x \in E$. Then $\lim\limits_{k\to\infty} \int f_k\, d\mu \geqslant \int \varphi\, d\mu$.

Proof. Given $\delta > 0$, denote $E_k = \{x \in E : f_k(x) \geqslant \varphi(x) - \delta\}$. Then, under conditions of the lemma, $E_k \subset E_{k+1}$, and $\cup_k E_k = E$. Therefore $(E \setminus E_k) \supset (E \setminus E_{k+1})$ and $\cap_k (E \setminus E_k) = \varnothing$. Now Lemma 1.12(3) implies that $\mu(E \setminus E_k) \to 0$ as $k \to \infty$.

In view of Theorem 4.8(1) and Lemma 4.4 we have

$$\int f_k\, d\mu \geqslant \int_{E_k} f_k\, d\mu \geqslant \int_{E_k} (\varphi - \delta)\, d\mu$$
$$= \int_E \varphi\, d\mu - \int_{E\setminus E_k} \varphi\, d\mu - \int_{E_k} \delta\, d\mu \geqslant \int_E \varphi\, d\mu - \int_{E\setminus E_k} C\, d\mu - \int_{E_k} \delta\, d\mu$$
$$= \int \varphi\, d\mu - C\,\mu(E\setminus E_k) - \delta\,\mu(E_k)$$

where $C = \max\limits_{x\in E} f(x)$. Since the last two terms in the right hand side can be made arbitrarily small by choosing large k and small δ respectively, the required inequality follows. □

Theorem 5.2 (the monotone convergence theorem). *Let f_m be non-negative measurable functions such that*

$$f_m(x) \leqslant f_{m+1}(x)\,, \qquad \lim_{m\to\infty} f_m(x) = f(x)$$

for all $x \in E$. Then $\lim\limits_{m\to\infty} \int f_m\, d\mu = \int f\, d\mu$.

Proof. The condition $f_m \leqslant f_{m+1}$ implies that $f_k \leqslant f$ and hence, by Theorem 4.8(1),

$$\int f_k\, d\mu \leqslant \int f\, d\mu\,, \qquad \forall k = 1, 2, \ldots$$

On the other hand, in view of Lemma 5.1, we have

$$\int \varphi\, d\mu \leqslant \lim_{m\to\infty} \int f_m\, d\mu.$$

for all $\varphi \in \mathrm{Simp}_0$ such that $0 \leqslant \varphi \leqslant f$. By Definition 4.5 this implies that

$$\lim_{m\to\infty} \int f_m\, d\mu \geqslant \int f\, d\mu.$$

□

Lebesgue's dominated convergence theorem.

Lemma 5.3 (Fatou's lemma). *Let f_k be non-negative measurable functions. Then*

$$\int \liminf f_k \, d\mu \;\leqslant\; \liminf \int f_k \, d\mu \,. \tag{5.1}$$

Proof. Denote $h_m(x) = \inf_{k \geqslant m} \{f_k(x)\}$ Then

$$f_m(x) \geqslant h_m(x), \qquad h_m(x) \leqslant h_{m+1}(x) \tag{5.2}$$

for all $x \in E$ and $m \in \mathbb{N}$, Now (5.2)) and the monotone convergence theorem imply

$$\int \liminf f_k \, d\mu \;=\; \int \lim_{k\to\infty} h_k \, d\mu \;=\; \lim_{k\to\infty} \int h_k \, d\mu$$
$$= \; \liminf \int h_k \, d\mu \;\leqslant\; \liminf \int f_k \, d\mu \,.$$

□

Corollary 5.4. *Let f_k be real measurable functions. Assume that there exists a measurable function g such that $f_k(x) \leqslant g(x)$ and $\int g \, d\mu < \infty$. Then*

$$\int \limsup f_k \, d\mu \;\geqslant\; \limsup \int f_k \, d\mu.$$

Proof. If we apply Fatou's lemma to the functions $(g - f_k)$ then (5.1) turns into the required inequality. □

Theorem 5.5 (Lebesgue's dominated convergence theorem). *Let f_k be measurable functions, and $f_k(x) \to f(x)$ for all x. Assume that there exists a non-negative measurable function g such that $|f_k(x)| \leqslant g(x)$ and $\int g \, d\mu < \infty$. Then*

$$\lim_{k\to\infty} \int |f - f_k| \, d\mu \;=\; 0 \,, \tag{5.5}$$

$$\lim_{k\to\infty} \int f_k \, d\mu \;=\; \int f \, d\mu \,. \tag{5.6}$$

Proof. Since $|f_k(x)| \leqslant g(x)$, we have $|f(x)| \leqslant g(x)$ and $|f(x) - f_k(x)| \leqslant 2g(x)$. By Corollary 5.4

$$\overline{\lim_{k\to\infty}} \int |f - f_k| \, d\mu \;\leqslant\; \int \overline{\lim_{k\to\infty}} |f - f_k| \, d\mu \;=\; \int \lim_{k\to\infty} |f - f_k| \, d\mu \;=\; 0 \,,$$

which implies (5.5). In view of Theorem 4.8(2) and Theorem 4.8(4)

$$\left| \int f \, d\mu - \int f_k \, d\mu \right| \;\leqslant\; \int |f - f_k| \, d\mu$$

for all $k \in \mathbb{N}$. Therefore (5.5) implies (5.6).

Further properties of the integrals.

Using Theorems 5.2 and 5.5 one can prove that many results, which are valid for the integrals of simple functions, remain valid for the integrals of arbitrary integrable functions f. Indeed, in view of Lemma 3.8 any measurable function can be approximated by simple functions, and then one can often take the limit. In particular, $f, g \in L_1$ then

(1) $|\int f\, g\, d\mu| \leqslant (\sup |f|) \int |g|\, d\mu$,
(2) $|\int f\, g\, d\mu| \leqslant \left(\int |f|^2\, d\mu\right)^{1/2} \left(\int |g|^2\, d\mu\right)^{1/2}$,
(3) for all $p \geqslant 1$

$$\left(\int |f+g|^p\, dx\right)^{1/p} \leqslant \left(\int |f|^p\, dx\right)^{1/p} + \left(\int |g|^p\, dx\right)^{1/p} \quad \text{(Minkowski's inequality)}$$

etc.

Complex-valued functions.

If f is a complex-valued function then one defines $\int f\, d\mu = \int \mathrm{Re}\, f\, d\mu + \int \mathrm{Im}\, f\, d\mu$. Using the corresponding results on real-valued functions, one can easily prove similar statements for complex-valued functions.

6. Comparison of measures.

The notion "almost everywhere".

Definition 6.1. We say that an x-dependent statement about quantities on the measure space $(E, \mathcal{A}, \mu)$ holds *almost everywhere* (a.e.) if the set B of x for which the statement fails is an element of $\mathcal{A}$ and $\mu(B) = 0$. If $E_0 \in \mathcal{A}$ then we say that an x-dependent statement holds *almost everywhere on* E_0 if the subset $B \subset E_0$ of x for which the statement fails is an element of $\mathcal{A}$ and $\mu(B) = 0$.

Of course, the notion "almost everywhere" depends on the measure μ.

Example 6.2. Let μ be the δ-measure at the point x_0, and f and g be some measurable functions. Then $f \geqslant g$ a.e. if and only if $f(x_0) \geqslant g(x_0)$.

Example 6.3. Let μ be Lebesgue's measure on $\mathbb{R}^1$, and f and g be measurable functions on $\mathbb{R}^1$ such that $f(x) = g(x)$ for all irrational x. Then $f = g$ a.e.

Theorem 6.4. *Let $(E, \mathcal{A}, \mu)$ be a measure space, $E_0 \in \mathcal{A}$ and f, g be non-negative measurable functions on E_0. Then*

(1) $\int_{E_0} f\, d\mu = 0$ *if and only if* $f(x) = 0$ *a.e. on the set* E_0*;*
(2) $\int_{E_0} f\, d\mu = \int_{E_0} g\, d\mu$ *whenever* $f = g$ *a.e. on the set* E_0

Proof. This Theoremt is an easy consequence of definitions.

Since the integral over a set of measure zero is equal to zero, **all the results from Section 5 remain valid if the corresponding conditions are fulfilled a.e.** In particular, Theorem 5.5 is valid if $f_k \to f$ a.e. and $|f_k| \leqslant g$ a.e.

The examples above show that two functions f and g which are equal a.e. may take different values on a rather rich set. In particular, if μ is Lebesgue's measure on $\mathbb{R}^1$ then $f(x)$ may not concide with $g(x)$ on the set of rational numbers, or even on an uncountable set (for example, on the Cantor set). The typical problem is to prove that for a given f there exists (or does not exist) a smooth function g such that $f = g$ a.e.

Proposition 6.5. *Let μ be Lebesgue's measure and f be a measurable function on $\mathbb{R}^1$. Assume that f has a jump at $x_0 \in \mathbb{R}^1$, that is, the limits $f(x_0+0)$ and $f(x_0-0)$ exist and $f(x_0+0) \neq f(x_0-0)$. Then f cannot coincide a.e. with a function which is continuous at x_0.*

Proof. Let $f = g$ a.e. Then any interval of the form $(x_0 - \varepsilon, x_0)$ or $(x_0, x_0 + \varepsilon)$ contains at least one point x at which $f(x) = g(x)$ (otherwise $f(x) \neq g(x)$ on a set of positive Lebesgue's measure). Therefore there exist sequences $x_k^- \uparrow x_0$ and $x_k^+ \downarrow x_0$ such that $g(x_k^-) = f(x_k^-)$ and $g(x_k^+) = f(x_k^+)$ for all $k = 1, 2, \ldots$ Then

$$\lim_{k\to\infty} g(x_k^-) \;=\; f(x_0 - 0) \;\neq\; f(x_0 + 0) \;=\; \lim_{k\to\infty} g(x_k^+)\,,$$

so the function g is not continuous at x_0. $\square$

Singular and absolutely continuous measures.

Definition 6.6. Let $\mathcal{A}$ be a σ-algebra over E and let μ_1 and μ_2 be measures defined on $\mathcal{A}$. We say that

(1) the measures μ_1 and μ_2 are (mutually) *singular* if there exists a set $\Omega \in \mathcal{A}$ such that $\mu_1(\Omega) = \mu_2(\Omega^c) = 0$.
(2) the measure μ_1 is *absolutely continuous* with respect to the measure μ_2 if $\mu_1(\Omega) = 0$ whenever $\mu_2(\Omega) = 0$.

Example 6.7. Let ν_1 coincides with the Lebesgue measure μ_n and ν_2 coincides with the restriction of the δ-measure at the point x_0 to the σ-algebra $\mathcal{L}_n$ of Lebesgue mesurable sets. Then the measures ν_1 and ν_2 are mutually singular. Indeed, we can take $\Omega = \{x_0\}$.

If two measures are singular then they live on non-overlapping subsets of E and, in fact, have nothing to do with one another. In particular, it can happen that $f > g$ a.e. with respect to μ_1 and $f < g$ a.e. with respect to μ_2.

If μ_1 is absolutely continuous with respect to μ_2 then any statement which holds μ_2-a.e. is also valid μ_1-a.e. Moreover, under some additional restriction the measure μ_1 can be represented as an integral with respect to the measure μ_2 (see Theorem 6.10).

Definition 6.8. We say that the measure space $(E, \mathcal{A}, \mu)$ (or the measure μ) is *σ-finite* if E can be represented as the union of a countable collection of the sets E_k, $k = 1, 2, \ldots$ such that $E_k \in \mathcal{A}$ and $\mu(E_k) < \infty$ for all $k \in \mathbb{N}$.

Example 6.9. The Lebegue measure μ_n is σ-finite. Indeed, we can take $E_k = (-k,\, k)^n$, $k \in \mathbb{N}$.

Theorem 6.10 (Radon–Nikodym theorem). *Let $\mathcal{A}$ be a σ-algebra over the set E and let μ_1 and μ_2 be measures defined on $\mathcal{A}$. Assume that the measure μ_2 is σ-finite. Then μ_1 is absolutely continuous with respect to μ_2 if and only if there exists a measurable non-negative function ρ such that*

$$\mu_1(E_0) \;=\; \int_{E_0} \rho \, d\mu_2 \tag{6.1}$$

for all sets $E_0 \in \mathcal{A}$.

The function ρ from Theorem 6.10 is said to be the density or the Radon–Nikodym derivative of the measure μ_1 with respect to μ_2. The density ρ is "almost uniquely" defined. Indeed, if $\widetilde{\rho}$ is another measurable function satisfying (6.1) then $\int_\Omega \rho\, d\mu_2 = \int_\Omega \widetilde{\rho}\, d\mu_2$ for all Borel sets Ω, which implies that $\widetilde{\rho} = \rho$ a.e. with respect to μ_2.

Remark 6.11. One can prove that the function μ_1 defined by (6.1) is always (even if μ_2 is not σ-finite) a measure which is absolutely continuous with respect to μ_2.

Decomposition of a Borel measure on $\mathbb{R}^1$.

Let μ be a Borel measure on $\mathbb{R}^1$, and

$$N(\lambda) \;=\; \begin{cases} \mu\left([0,\lambda)\right), & \lambda \geqslant 0, \\ -\mu\left([\lambda,0)\right), & \lambda \leqslant 0. \end{cases}$$

Then N is a non-decreasing measurable function on $\mathbb{R}^1$ such that

$$\mu([\lambda_1,\lambda_2)) = N(\lambda_2) - N(\lambda_1), \qquad \forall\,\lambda_1,\lambda_2 \in \mathbb{R}^1 .$$

Since the intervals of the form $[\lambda_1,\lambda_2)$ generate the Borel σ-algebra, the measure μ is uniquely determined by N.

Definition 6.12. The function N is said to be the distribution function of the measure μ.

Definition 6.13. We say that the Borel measure μ is continuous if $\mu(\{x\}) = 0$ for all points x.

By Lemma 1.12 $\lim\limits_{\lambda_1\uparrow\lambda_2} \mu([\lambda_1,\lambda_2)) = 0$ and $\lim\limits_{\lambda_2\downarrow\lambda_1} \mu([\lambda_1,\lambda_2)) = \mu(\{\lambda_1\})$. Therefore the measure μ is continuous if and only if the corresponding distribution function is continuous.

Definition 6.14. We say that the Borel measure μ is discrete if there exists a countable set of points $x_1, x_2, \dots$ such that $\Omega \cap \{x_1, x_2, \dots\} = \varnothing$ implies $\mu(\Omega) = 0$.

A discrete Borel measure can be represented as a sum of the form

$$\mu = \sum_k c_k\,\delta_{x_k}, \tag{6.2}$$

where δ_{x_k} are the delta-measures at the points x_k and $c_k = \mu(\{x_k\})$. The corresponding distribution function is constant on the intervals which do not contain the points x_k and has the jumps c_k at the points x_k.

Clearly, if μ_1 is continuous and μ_2 is discrete then μ_1 and μ_2 are mutually singular (we can take $\Omega = \{x_1, x_2, \dots\}$).

Proposition 6.15. *Any Borel measure μ on $\mathbb{R}^1$ can be uniquely represented as the sum $\mu_d + \mu_c$, where μ_d is a discrete measure and μ_c is a continuous measure.*

Proof. Let Ω be the set of points x for which $\mu(\{x\}) > \varepsilon$, and

$$\Omega = \cup_{\varepsilon>0}\,\Omega_\varepsilon = \{x : \mu(\{x\}) > 0\}.$$

Since the measure of a bounded set is finite, the intersection of Ω_ε with any bounded interval contains only a finite number of points. This implies that the sets Ω_ε are countable. Therefore Ω is also countable. Let $\Omega = \{x_1, x_2, \dots\}$, and μ_d be the discrete measure defined by (6.2). Then $\mu_c = \mu - \mu_d$ is a continuous measure, and $\mu = \mu_d + \mu_c$. The points x_k and the constants c_k are determined uniquely by the measure μ, so μ_d and μ_c are uniquely defined. □

Proposition 6.15 is not very helpful, since we do not have any explicit representation for an arbitrary continuous measure μ_c. It is easy to see that a measure is continuous if it is absolutely continuous with respect to Lebesgue's measure. However, the converse is not true.

Example. Let $N(\lambda)$ be the continuous function on $[0,1]$ such that

$N(\lambda) = 1/2$ as $\lambda \in (1/3, 2/3)$,
$N(\lambda) = 1/4$ as $\lambda \in (1/9, 2/9)$ and $N(\lambda) = 3/4$ as $\lambda \in (7/9, 8/9)$,

etc. Then $N(0) = 0$, $N(1) = 1$, and N is constant on any open interval from $[0,1] \setminus C$, where C is the Cantor set (see Section 2). The corresponding measure μ lives on the Cantor set, and for any fixed point x we have $\mu(\{x\}) = 0$. Therefore μ is neither discrete nor absolutely continuous with respect to Lebesgue's measure.

Definition 6.16. A continuous measure which is not absolutely continuous with respect to Lebesgue's measure is said to be singular continuous.

Theorem 6.17 (Lebesgue decomposition theorem). *Any Borel measure μ on $\mathbb{R}^1$ can be uniquely represented as the sum $\mu_d + \mu_a + \mu_s$, where μ_d, μ_a and μ_s are mutually singular measures such that μ_d is discrete, μ_s is singular continuous and μ_a is absolutely continuous with respect to Lebesgue's measure.*

Of course, with the preceding notation $\mu_c = \mu_a + \mu_s$. By Theorem 6.10 the absolutely continuous component μ_a of the measure μ can be represented as an integral with respect to Lebesgue's measure. In general situation there is no good representation for the singular component μ_s. However, the Borel measures which appear in "real" problems usually do not have the singular component. For such measures Theorem 6.17 allows one to obtain explicit formulae in terms of the densities (corresponding to μ_a) and the δ-measures (corresponding to μ_d).

7. The Lebesgue spaces

Given a measure space $(E, \mathcal{A}, \mu)$, a real number $p \in [1, \infty)$ and a complex-valued measurable function f on E, we define

$$\|f\|_{L_p(E,\mu)} = \left(\int |f|^p \, d\mu \right)^{1/p}.$$

Obviously, as p varies $\|f\|_{L_p(E,\mu)}$ provides different estimates of the size f as it is "seen" by the measure μ. In particular, $\|f\|_{L_p(E,\mu)}$ cannot detect any properties of f which occur on sets having μ-measure zero.

Remark. In applications the most important L_p-norms are those corresponding to $p = 1$ and $p = 2$. Another important object is the L_∞-norm of a measurable function f, which is defined by

$$\|f\|_{L_\infty(E,\mu)} = \inf \{ M \geqslant 0 \, : \, |f| \leqslant M \text{ a.e.} \} .$$

However, we shall consider only finite p.

Definition 7.1. We say that $f \in L_p(E, \mu)$ if f is measurable and $\|f\|_{L_p(E,\mu)} < \infty$. When $E = \mathbb{R}^n$ and μ is the Lebesgue measure, we omit μ and simply write $L_p(\mathbb{R}^n)$.

We shall identify two functions from $L_p(E, \mu)$ which coincide μ-a.e. (which means, in fact, that we assume the elements of $L_p(E, \mu)$ to be the classes of functions). Since $\|f\|_{L_p(E,\mu)} = 0$ implies $f = 0$ a.e., this means that $f = 0$ as an element of $L_p(E, \mu)$ whenever $\|f\|_{L_p(E,\mu)} = 0$. By Minkowski's inequality

$$\|f + g\|_{L_p(E,\mu)} \leqslant \|f\|_{L_p(E,\mu)} + \|g\|_{L_p(E,\mu)} .$$

From here it follows that the metric defined on $\|f\|_{L_p(E,\mu)}$ by

$$\operatorname{dist}(f, g) = \|f - g\|_{L_p(E,\mu)} \tag{7.1}$$

satisfies the triangle inequality. We shall consider $L_p(E, \mu)$ as a metric space with the metric (7.1). In particular, $f_k \to f$ in $L_p(E, \mu)$ if $\|f - f_k\|_{L_p(E,\mu)} \to 0$.

Remark 7.2. Strictly speaking, the elements of L^p are equivalence classes of functions. One can say "the function f belongs to L^p" (Definition 7.1) but bearing in mind that there are many other functions which coincide with f as elements of L^p. In particular, if we modify f on a set of measure zero then the new function coincides with the initial one as an element of the L^p space. Therefore one should be carefull making statements like "the function $f \in L^p(\mathbb{R}^n)$ is equal to zero at the origin" (changing the function f at the origin we do not change it as an element of $L^p(\mathbb{R}^n)$).

By Lebesgue's dominated convergence theorem, if $f_k \to f$ μ-a.e. and there is a function $g \in L_p(E, \mu)$ such that $|f_k(x)| \leqslant g(x)$ then $f_k \to f$ in $L_p(E, \mu)$. The converse is not true; convergence in $L_p(E, \mu)$ does not imply convergence μ-a.e. However, we have the following

Theorem 7.3. *If $\|f - f_k\|_{L_p(E,\mu)} \to 0$ then there is a subsequence f_{k_j} which converges to f μ-a.e. Moreover, if $\{f_k\}$ is a Cauchy sequence in $L_p(E,\mu)$ then there exist a subsequence f_{k_i} and $f \in L_p(E,\mu)$ such that $\|f - f_{k_j}\|_{L_p(E,\mu)} \to 0$ and $f_{k_j} \to f$ μ-a.e.*

Proof. Assume that $\{f_k\}$ is a Cauchy sequence in $L_p(E,\mu)$, that is,

$$\|f_m - f_k\|^p_{L_p(E,\mu)} \;=\; \int |f_m - f_k|^p \, d\mu \;\to\; 0\,, \qquad k, m \to \infty\,.$$

If $f_{k_j} \to \widetilde{f}$ μ-a.e. then by Theorem 3.9 f is measurable and by Fatou's lemma

$$\|f - f_{k_j}\|^p_{L_p(E,\mu)} \;\leqslant\; \liminf_{l\to\infty} \|f_{k_l} - f_{k_j}\|^p_{L_p(E,\mu)} \;\leqslant\; \sup_{l\geqslant j} \|f_{k_l} - f_{k_j}\|^p_{L_p(E,\mu)} \;\to\; 0$$

as $j \to \infty$. This obviously implies $f \in L_p(E,\mu)$. Therefore it is sufficient to prove only that any Cauchy sequence in $L_p(E,\mu)$ contains a subsequence which converges μ-a.e. to some function f.

If $\{f_k\}$ is a Cauchy sequence in $L_p(E,\mu)$ then

$$\lim_{k\to\infty} \sup_{m\geqslant k} \mu\left(\{x : |f_m - f_k| \geqslant \varepsilon\}\right) \;=\; 0$$

for each fixed ε. Let us choose $1 \leqslant k_1 < k_2 < \dots$ in such a way that

$$\sup_{m\geqslant k_j} \mu\left(\{x : |f_m(x) - f_{k_j}(x)| \geqslant 2^{-j-1}\}\right) \;\leqslant\; 2^{-j-1}\,.$$

If $|f_{k_{j+1}}(x) - f_{k_j}(x)| < 2^{-j-1}$ for all $j \geqslant i$ then

$$|f_{k_l}(x) - f_{k_i}(x)| \;\leqslant\; \sum_{j=i}^{l-1} |f_{k_{j+1}}(x) - f_{k_j}(x)| \;<\; \sum_{j=i}^{l-1} 2^{-j-1} \;<\; 2^{-i}\,, \qquad \forall\, l > i\,.$$

Therefore $\sup_{l\geqslant i} |f_{k_l}(x) - f_{k_i}(x)| \geqslant 2^{-i}$ implies $\sup_{j\geqslant i} |f_{k_{j+1}}(x) - f_{k_j}(x)| \geqslant 2^{-j-1}$, and we obtain

$$\begin{aligned}
&\mu\left(\{x : \sup_{l\geqslant i} |f_{k_l}(x) - f_{k_i}(x)| \geqslant 2^{-i}\}\right) \\
&\qquad\leqslant\; \mu\left(\{x : \sup_{j\geqslant i} |f_{k_{j+1}}(x) - f_{k_j}(x)| \geqslant 2^{-j-1}\}\right) \\
&\qquad\leqslant\; \mu\left(\bigcup_{j\geqslant i} \{x : |f_{k_{j+1}}(x) - f_{k_j}(x)| \geqslant 2^{-j-1}\}\right) \\
&\leqslant\; \sum_{j=i}^{\infty} \mu\left(\{x : |f_{k_{j+1}}(x) - f_{k_j}(x)| \geqslant 2^{-j-1}\}\right) \;\leqslant\; \sum_{j=i}^{\infty} 2^{-j-1} \;\leqslant\; 2^{-i}\,. \qquad (7.2)
\end{aligned}$$

Assume that the sequence $f_{k_j}(x)$ is not convergent. Then there exists $\varepsilon > 0$ for which the inequality

$$\sup_{l\geqslant i} |f_{k_l}(x) - f_{k_i}(x)| \geqslant \varepsilon$$

holds with an arbitrarily large i. Therefore the set of points at which $\{f_{k_j}\}$ does not converge is a subset of

$$\Omega_j = \bigcap_{i\geqslant j} \{x : \sup_{l\geqslant i} |f_{k_l}(x) - f_{k_i}(x)| \geqslant 2^{-i}\}$$

for each fixed j. In view of (7.2)

$$\mu(\Omega_j) = \sum_{i=j}^{\infty} \mu\left(\{x : \sup_{l\geqslant i} |f_{k_l}(x) - f_{k_i}(x)| \geqslant 2^{-i}\}\right) \leqslant \sum_{i=j}^{\infty} 2^{-i}.$$

Therefore $\mu(\Omega_j) \to 0$ as $j \to \infty$, which implies that $\{f_{k_j}\}$ converges almost everywhere to some function f. $\square$

Corollary 7.4. *The space $L_p(E,\mu)$ is complete.*

Proof. Given a Cauchy sequence $\{f_k\}$, we choose a subsequence f_{k_j} which converges in $L_p(E,\mu)$ to some function $f \in L_p(E,\mu)$ (Theorem 7.3). Then

$$\|f - f_k\|_{L_p(E,\mu)} \leqslant \|f - f_{k_j}\|_{L_p(E,\mu)} + \|f_{k_j} - f_k\|_{L_p(E,\mu)}$$

for all k and k_j, so

$$\lim_{k\to\infty} \|f - f_k\|_{L_p(E,\mu)} \leqslant \lim_{k_j\to\infty} \|f - f_{k_j}\|_{L_p(E,\mu)} + \lim_{k,k_j\to\infty} \|f_{k_j} - f_k\|_{L_p(E,\mu)} = 0.$$

$\square$

Definition 7.5. We say that a function φ on $\mathbb{R}^1$ is a step function if φ takes only a finite number of values $a_1,\dots,a_m$ and for each $a_k \neq 0$ the inverse image $\varphi^{-1}(a_k)$ is a (possibly degenerate) interval.

Theorem 7.6. *The set of step functions is dense in $L_p(\mathbb{R}^1)$ for all $1 \leqslant p < \infty$.*

The proof of this theorem is based on the following

Lemma 7.7. *Any open set $U \in \mathbb{R}^1$ is the union of a countable collection of mutually disjoint open intervals.*

Proof. If $U \subset \mathbb{R}^1$ is open and $x \in U$, let I_x be the maximal connected component of U containing x. Since U is open, I_x is an open interval. Obviously, for any $x, y \in U$, either $I_x = I_y$ or $I_x \cap I_y = \varnothing$.

We have $U = \cup_x I_x$ where the union is taken over all the rational x. Since the set of rational numbers is countable (and the intervals I_x either coincide or do not intersect), we have obtained the required representation of U. $\square$

Proof of Theorem 7.6. From the definition of Lebesgue's integral it follows that any function $f \in L_p(\mathbb{R}^1)$ is approximated in $L_p(\mathbb{R}^1)$ by compactly supported simple functions. Any compactly supported simple function is a linear combination of the characteristic functions of some measurable sets. Therefore it is sufficient to prove that the characteristic function χ of an arbitrary measurable set $\Omega \in \mathbb{R}^1$ can be approximated by step functions.

By Theorem 2.5 for all $\varepsilon > 0$ there exist open sets $\Omega_\varepsilon \supset \Omega$ such that $\mu_1(\Omega_\varepsilon \setminus \Omega) \leqslant \varepsilon$. Since the characteristic functions χ_ε of the sets Ω_ε converge in $L_p(\mathbb{R}^1)$ to χ as $\varepsilon \to 0$, we can assume without loss of generality that Ω is an open set.

By Lemma 7.7 an open set Ω coincides with the union of a countable collection of open intervals I_k, $k = 1, 2, \dots$ If χ_k is the characteristic function of I_k then the functions $\widetilde{\chi}_j = \sum_{k=1}^{j} \chi_k$ converge to the characteristic function of Ω in $L_p(\mathbb{R}^1)$ as $j \to \infty$. $\square$

Let $C_0^\infty(\mathbb{R}^1)$ be the space of smooth functions with compact supports. Obviously, the characteristic function of a bounded interval I can be approximated in $L_p(\mathbb{R}^1)$ by C_0^∞-functions f_k (for example, we can choose f_k in such a way that $f_k(x) = 1$ for all k and $x \in I$, $0 \leqslant f_k(x) \leqslant 1$ for all x, and $f_k(x) = 0$ if $\operatorname{dist}(x, I) \geqslant k^{-1}$). Therefore Theorem 7.6 implies

Corollary 7.8. *The space $C_0^\infty(\mathbb{R}^1)$ is dense in $L_p(\mathbb{R}^1)$ for all $1 \leqslant p < \infty$.*

8. Distributions

Convergence in $\mathcal{S}(\mathbb{R}^1)$.

We say that the sequence of functions $\varphi_k \in \mathcal{S}(\mathbb{R}^1)$ converges to $\varphi \in \mathcal{S}(\mathbb{R}^1)$ in $\mathcal{S}(\mathbb{R}^1)$ if

$$\sup_x |\, x^l \, \frac{d^m(\varphi - \varphi_k)}{dx^m} \,| \;\to\; 0\,, \qquad k \to \infty\,,$$

for all $l, m = 0, 1, \ldots$. If $\varphi_k \to \varphi$ in $\mathcal{S}(\mathbb{R}^1)$ then for all l and m we have

$$|\, \frac{d^m(\varphi - \varphi_k)}{dx^m} \,| \;\leqslant\; c_{k,l,m}\,(1 + x^2)^{-l}\,,$$

where $c_{k,l,m} \to 0$ as $k \to \infty$. This implies, in particular, that $\varphi_k \to \varphi$ in $L_p(\mathbb{R}^1)$ for all p.

The dual space $\mathcal{S}'(\mathbb{R}^1)$.

A map $f : \mathcal{S}(\mathbb{R}^1) \to \mathbb{C}^1$ is said to be a functional on $\mathcal{S}(\mathbb{R}^1)$. The value of functional f on the function $\varphi \in \mathcal{S}(\mathbb{R}^1)$ is denoted by $\langle f, \varphi \rangle$. We say that the functional f is linear if

$$\langle f, \alpha\varphi + \beta\psi \rangle \;=\; \alpha\,\langle f, \varphi \rangle \;+\; \beta\,\langle f, \psi \rangle\,, \qquad \forall \varphi, \psi \in \mathcal{S}(\mathbb{R}^1)\,, \quad \forall\, \alpha, \beta \in \mathbb{C}^1\,,$$

and f is continuous if $\varphi_k \to \varphi$ implies $\langle f, \varphi_k \rangle \to \langle f, \varphi \rangle$.

Definition 8.1. The linear continuous functionals on $\mathcal{S}(\mathbb{R}^1)$ are said to be the (temperate) distributions.

The distributions form a linear space. If $\varphi, \psi \in \mathcal{S}(\mathbb{R}^1)$ and $\alpha, \beta \in \mathbb{C}^1$, we <u>define</u> the distribution $\alpha f + \beta g$ by

$$\langle \alpha f + \beta g, \varphi \rangle \;=\; \alpha\,\langle f, \varphi \rangle \;+\; \beta\,\langle g, \varphi \rangle\,, \qquad \forall\, \varphi \in \mathcal{S}(\mathbb{R}^1)\,.$$

The linear space of distributions is denoted by $\mathcal{S}'(\mathbb{R}^1)$.

Example 8.2. If $f \in L_p(\mathbb{R}^1)$ with $p \geqslant 1$ then the functional on $\mathcal{S}(\mathbb{R}^1)$ defined by $\langle f, \varphi \rangle = \int f\,\varphi\,dx$ is a distribution. Obviously, if two functions f_1 and f_2 define the same distribution then $f_1 = f_2$ almost everywhere (with respect to the Lebesgue measure). This allows us to identify the functions $f \in L_p(\mathbb{R}^1)$ with distributions. Further on we shall use the same notation f for the function $f \in L_p(\mathbb{R}^1)$ and for the corresponding distribution.

Example 8.3. If μ is a Borel measure on $\mathbb{R}^1$ such that

$$\mu\,(\{x : |x| \leqslant R\}) \leqslant C(R^N + 1)$$

with some positive constants C and N then $\langle \mu, \varphi \rangle = \int \varphi\,d\mu$ is a distribution. One can prove that

$$\int \varphi\,d\mu_1 \;=\; \int \varphi\,d\mu_2\,, \qquad \forall\, \varphi \in \mathcal{S}(\mathbb{R}^1)$$

if and only if $\mu_1 = \mu_2$. Therefore the Borel measures can be also identified with the corresponding distributions.

Example 8.4. Let $x \in \mathbb{R}^1$ be a fixed point. The distribution δ_x defined by

$$\langle \delta_x, \varphi \rangle \;=\; \varphi(x)\,, \qquad \forall\, \varphi \in \mathcal{S}(\mathbb{R}^1)$$

is said to be the δ-function at x. The δ-function is one of the simplest distributions; its value on φ depends only on the value of φ at one fixed point. We have

$$\langle \delta_x, \varphi \rangle \;=\; \int \varphi(x)\, d\delta_x\,, \qquad \forall\, \varphi \in \mathcal{S}(\mathbb{R}^1)\,,$$

we δ_x in the right hand side stands for the δ-measure at x (Example 1.9).

Operations with distributions.

One can do with the distributions almost all the same things as with the functions from $\mathcal{S}(\mathbb{R}^1)$. The basic idea is as follows. Assume that we are going to extend a linear operator T in the space $\mathcal{S}(\mathbb{R}^1)$ to the space $\mathcal{S}'(\mathbb{R}^1)$. First, we take $\psi \in \mathcal{S}(\mathbb{R}^1)$ and write

$$\langle T\psi, \varphi \rangle \;=\; \int T\psi\, \varphi\, dx\,, \qquad \forall\, \varphi \in \mathcal{S}(\mathbb{R}^1)$$

(here we consider the function $T\psi$ as a distribution). Then we try to find a linear operator T' in $\mathcal{S}(\mathbb{R}^1)$ such that

$$\langle T\psi, \varphi \rangle \;=\; \int \psi\, T'\varphi\, dx\,, \qquad \forall\, \varphi \in \mathcal{S}(\mathbb{R}^1)\,.$$

Finally, for $f \in \mathcal{S}(\mathbb{R}^1)$ we <u>define</u> Tf by

$$\langle Tf, \varphi \rangle \;=\; \langle f, T'\varphi \rangle\,, \qquad \forall\, \varphi \in \mathcal{S}(\mathbb{R}^1)\,. \tag{8.1}$$

Obviously, (8.1) defines a linear functional Tf on $\mathcal{S}(\mathbb{R}^1)$. If T' is continuous in $\mathcal{S}(\mathbb{R}^1)$ then $\varphi_k \to \varphi$ implies $\langle Tf, \varphi_k \rangle \to \langle Tf, \varphi \rangle$. In this case the functional Tf is continuous, so $Tf \in \mathcal{S}'(\mathbb{R}^1)$.

Definition 8.5. Let h be an infinitely smooth function on $\mathbb{R}^1$ such that

$$|d^k h/dx^k| \;\leqslant\; c_k\,(1+x^2)^{m_k}\,, \qquad \forall\, k = 0, 1, \ldots, \tag{8.2}$$

with some constants $c_k > 0$ and $m_k > 0$. Then $\varphi \to h\,\varphi$ is a continuous operator in $\mathcal{S}(\mathbb{R}^1)$, and for $f \in \mathcal{S}'(\mathbb{R}^1)$ we define hf by

$$\langle hf, \varphi \rangle \;=\; \langle f, h\,\varphi \rangle\,, \qquad \forall\, \varphi \in \mathcal{S}(\mathbb{R}^1)\,.$$

Definition 8.6. The differentiation is a continuous operator in $\mathcal{S}(\mathbb{R}^1)$, so for $f \in \mathcal{S}'(\mathbb{R}^1)$ we define f' by

$$\langle f', \varphi \rangle \;=\; -\,\langle f, \varphi' \rangle\,, \qquad \forall\, \varphi \in \mathcal{S}(\mathbb{R}^1)\,.$$

In the same manner we can define many other operators in $\mathcal{S}'(\mathbb{R}^1)$, in particular, the operator $f(x) \to g(x) = f\,(y(x))$.

Example 8.7. If h is a continuously differentiable function satisfying (8.2) with $k = 0, 1$ then the derivative of $h \in \mathcal{S}'(\mathbb{R}^1)$ coincides with the usual derivative f'.

Example 8.8. Let

$$f(x) \;=\; \begin{cases} 0, & x < 0, \\ 1, & x \geqslant 0\,. \end{cases}$$

Then for all $\varphi \in \mathcal{S}(\mathbb{R}^1)$ we have

$$-\langle f, \varphi' \rangle \;=\; -\int f\,\varphi'\,dx \;=\; -\int_0^\infty \varphi'\,dx \;=\; \varphi(0)\,.$$

Therefore f' coincides with the δ-function at $x = 0$.

Example 8.9. Let

$$-\infty < a_1 < a_2 < \cdots < a_m < \infty\,.$$

Assume that f is a piecewise continuous function, which is equal to zero as $x < a_1$ or $x > a_m$ and is continuously differentiable on the intervals (a_k, a_{k+1}). Let

$$c_k \;=\; f(a_k + 0) \;-\; f(a_k - 0)$$

be the jumps of f at the points a_k, and let

$$g(x) \;=\; \begin{cases} f'(x), & x \in (a_k, a_{k+1})\,, \qquad k = 1, \dots, m-1, \\ 0, & x < a_1 \text{ or } x > a_m. \end{cases}$$

Integrating by parts we obtain

$$-\langle f, \varphi' \rangle \;=\; -\int f\,\varphi'\,dx \;=\; \int g\,\varphi\,dx \;+\; \sum_{k=1}^{m} c_k\,\varphi(a_k)\,.$$

Therefore $f' = g + \sum_{k=1}^{m} c_k\,\delta_{a_k}$ where δ_{a_k} are the δ-functions at the points a_k.

Example 8.10. Let $f = \delta_x$. Then $\langle f', \varphi \rangle = -\varphi'(x)$. The distribution f' cannot be described in any simpler way. This distribution is called the derivative of the δ-function at x.

Supports of distributions.

Generally speaking, one cannot say what is the value of a distribution at one fixed point. However, we can single out a class of distributions which coincide on an open set.

Definition 8.11. We say that the distribution f vanishes on an open set Ω and write $f|_\Omega = 0$ if $\langle f, \varphi \rangle = 0$ for all $\varphi \in \mathcal{S}(\mathbb{R}^1)$ with $\operatorname{supp}\varphi \subset \Omega$. We say that f coincides with a distribution g on Ω if $(f-g)|_\Omega = 0$.

In particular, we say that the distribution f coincides with a function g on Ω if $(f-g)|_\Omega = 0$ in the sense of distributions.

Definition 8.12. For $f \in \mathcal{S}'(\mathbb{R}^1)$ we define $\operatorname{supp} f = \mathbb{R}^1 \setminus \Omega_f$, where Ω_f is the maximal open set on which f is equal to zero (i.e., Ω_f is the union of all open sets Ω such that $f|_\Omega = 0$).

Example 8.13. The supports of all the derivatives of the δ-function at x coincide with the point x.

The support of a function coincides with the support of the corresponding distribution modulo a set of Lebesgue's measure zero. If h is a function satisfying (8.2) then

$$\operatorname{supp}(hf) \;\subset\; (\operatorname{supp} h) \cap (\operatorname{supp} f)\,, \qquad \forall\, f \in \mathcal{S}'(\mathbb{R}^1)\,.$$

In particular, if $h = 0$ in a neighbourhood of $\operatorname{supp} f$ then $hf = 0$. That is not necessarily true if $h = 0$ only on $\operatorname{supp} f$.

Example 8.14. Let $f = \delta_0'$ be the derivative of the δ-function at zero, and $h(x) = x$. Then $h = 0$ on $\operatorname{supp} f$. However,

$$\langle hf, \varphi \rangle \;=\; -\;(x\varphi)'|_{x=0} \;=\; -\,\varphi(0)\,,$$

so $x\delta_0' = -\delta_0$. In the same manner we obtain that $x^2\delta_0' = -x\delta_0 = 0$.

The set of distributions with compact supports is denoted by $\mathcal{E}'(\mathbb{R}^1)$. We can choose C_0^∞-functions ψ_j such that $\sum_j \psi_j(x) \equiv 1$ (it is called a partition of unity). Then an arbitrary distribution $f \in \mathcal{S}'(\mathbb{R}^1)$ is represented as a sum of distributions $f_j = \psi_j f$ with compact supports.

Theorem 8.15. *If $f \in \mathcal{E}'(\mathbb{R}^1)$ then*

$$f \;=\; \sum_{j=0}^{m} \frac{d^j g_j}{dx^j}\,, \tag{8.3}$$

where g_j are some continuous functions.

Theorem 8.15 implies that any distribution $f \in \mathcal{S}'(\mathbb{R}^1)$ can be represented as a sum of distributions f_j of the form (8.3). However, the number of derivatives m in the representation of f_j may be depending on j and going to infinity as $j \to \infty$.

9. Fourier transform in $\mathcal{S}'(\mathbb{R}^1)$

We shall need the following

Lemma 9.1. *The Fourier transform is continuous in* $\mathcal{S}(\mathbb{R}^1)$.

Proof. Let $\hat{\psi}_k(\xi)$ be the Fourier transforms of $\psi_k \in \mathcal{S}(\mathbb{R}^1)$. We have to prove that $\psi_k \to 0$ in $\mathcal{S}(\mathbb{R}^1)$ implies

$$\sup_\xi | \xi^l \frac{d^m \hat{\psi}_k}{d\xi^m} | \;\to\; 0\,, \qquad k \to \infty \tag{9.1}$$

(with the notation of Section 8 $\psi_k = \varphi - \varphi_k$). Since

$$\xi^l \frac{d^m \hat{f}}{d\xi^m} \;=\; (-i)^{l+m} \left(\widehat{\frac{d^l (x^m f)}{dx^l}} \right), \qquad \forall f \in \mathcal{S}(\mathbb{R}^1)\,,$$

and $f \to x^m \frac{d^l f}{dx^l}$ is a continuous operator in $\mathcal{S}(\mathbb{R}^1)$, it is sufficient to obtain (9.1) only for $l = m = 0$. This follows from the estimates

$$\begin{aligned} \sup_\xi |\hat{\psi}_k(\xi)| \;&=\; (2\pi)^{-1/2} \,| \int e^{-ix\xi}\, \psi_k(x)\, dx \,| \\ &\leqslant\; (2\pi)^{-1/2} \int |\psi_k(x)|\, dx \;\leqslant\; c_k\, (2\pi)^{-1/2} \int (1+x^2)^{-1}\, dx\,, \end{aligned}$$

where $c_k = \sup_x |(1+x^2)\psi_k(x)| \to 0$ as $\psi_k \to 0$ in $\mathcal{S}(\mathbb{R}^1)$. $\square$

If $f \in \mathcal{S}(\mathbb{R}^1)$ then

$$\langle \hat{f}, \varphi \rangle \;=\; \langle f, \hat{\varphi} \rangle\,, \qquad \forall\, \varphi \in \mathcal{S}(\mathbb{R}^1)\,. \tag{9.2}$$

Let now f be a distribution. Then, in view of Lemma 9.1, (9.2) defines a linear continuous functional $\hat{f}$ on $\mathcal{S}(\mathbb{R}^1)$.

Definition 9.2. The Fourier transform of $f \in \mathcal{S}'(\mathbb{R}^1)$ is the distribution $\hat{f}$ defined by (9.2).

Obviously, (9.2) holds if f belongs to $L_p(\mathbb{R}^1)$, $p = 1, 2$, and if the Fourier transform is understood in the L_p sense. Therefore the Fourier transform in $\mathcal{S}'(\mathbb{R}^1)$ coincides with the Fourier transform in $L_p(\mathbb{R}^1)$ if $f \in L_p(\mathbb{R}^1)$.

By analogy, the inverse Fourier transform $\mathcal{F}^{-1} f$ of $f \in \mathcal{S}'(\mathbb{R}^1)$ is defined by

$$\langle \mathcal{F}^{-1} f, \varphi \rangle \;=\; \langle f, \mathcal{F}^{-1} \varphi \rangle\,, \qquad \forall\, \varphi \in \mathcal{S}(\mathbb{R}^1)\,.$$

We have

$$\langle \mathcal{F}^{-1} \hat{f}, \varphi \rangle \;=\; \langle \hat{f}, \mathcal{F}^{-1} \varphi \rangle \;=\; \langle f, \widehat{\mathcal{F}^{-1} \varphi} \rangle \;=\; \langle f, \varphi \rangle\,,$$

for all $\varphi \in \mathcal{S}(\mathbb{R}^1)$, which means that $\mathcal{F}^{-1} \hat{f} = f$. Since $\mathcal{F}^{-1} f(\xi) = \hat{f}(-\xi)$, all the results concerning the Fourier transform in $\mathcal{S}'(\mathbb{R}^1)$ can be easily reformulated for the inverse Fourier transform.

Example 9.3. If $\hat\delta_0$ is the Fourier transform of the δ-function at $x=0$ then

$$\langle \hat\delta_0, \varphi\rangle \;=\; \hat\varphi(0) \;=\; (2\pi)^{-1/2}\int \varphi(x)\,dx\,.$$

Therefore $\hat\delta_0$ is the function identically equal to $(2\pi)^{-1/2}$. In the same way we obtain that $\mathcal{F}^{-1}\delta_0 = \hat\delta_0$. This implies that the direct and inverse Fourier transforms of the function which is identically equal to a, coincide with $a\,(2\pi)^{1/2}\,\delta_0$.

Proposition 9.4. *For all $f \in \mathcal{S}'(\mathbb{R}^1)$*

$$\widehat{f'} \;=\; (i\xi)\,\hat f\,, \qquad \widehat{xf} \;=\; i\,(\hat f)'\,. \tag{9.3}$$

Proof. The equalities (9.3) are valid for all the functions from $\mathcal{S}(\mathbb{R}^1)$. Therefore for all $\varphi \in \mathcal{S}(\mathbb{R}^1)$ we have

$$\langle \widehat{f'}, \varphi\rangle \;=\; \langle f', \hat\varphi\rangle \;=\; -\,\langle f, (\hat\varphi)'\rangle \;=\; i\,\langle f, \widehat{x\varphi}\rangle \;=\; i\,\langle \xi\hat f, \varphi\rangle\,,$$

$$\langle \widehat{xf}, \varphi\rangle \;=\; \langle xf, \hat\varphi\rangle \;=\; \langle f, x\hat\varphi\rangle \;=\; -i\,\langle f, \widehat{\varphi'}\rangle \;=\; i\,\langle (\hat f)', \varphi\rangle\,.$$

This implies (9.3) for all $f \in \mathcal{S}'(\mathbb{R}^1)$. $\square$

From Example 9.3 and Proposition 9.4 it follows that the Fourier transforms of polynomials are linear combinations of the δ-function at $x=0$ and its derivatives.

Definition 9.5. We say that a distribution f belongs to the Sobolev class $H^s(\mathbb{R}^1)$ with $s \in \mathbb{R}^1$ if the Fourier transform $\hat f$ is a function such that $(1+\xi^2)^{s/2}\hat f(\xi) \in L_2(\mathbb{R}^1)$.

The inner product and the norm in $H^s(\mathbb{R}^1)$ is defined for the negative s in the same way as for $s \geqslant 0$. If $s<0$ then, generally speaking, the elements of $H^s(\mathbb{R}^1)$ are not functions but distributions. The value of s indicates how non-smooth the distributions are. By Proposition 9.4 $f \in H^s(\mathbb{R}^1)$ implies $d^kf/dx^k \in H^{s-k}(\mathbb{R}^1)$, so the differentiation makes the distribution more non-smooth. For example, the derivative of a δ-function is a more non-smooth distribution than the δ-function itself.

Applications to differential equations.

Let us consider the differential equation

$$\sum_{k=1}^{m} c_k\,D^k f \;=\; 0\,, \tag{9.4}$$

where $D^k = (-i)^k\,d^k/dx^k$ and c_k are some constants. We assume that f is a distribution and understand the derivatives in the sense of distributions. Then, taking the Fourier transform, we obtain

$$P(\xi)\,\hat f(\xi) \;=\; 0\,, \tag{9.5}$$

where $P(\xi) = \sum_{k=1}^{m} c_k\xi^k$ is the symbol of the corresponding differential operators.

Clearly, if P is not identically equal to zero and $\hat{f}$ is a function then (9.5) implies $\hat{f} = 0$ almost everywhere. Thus, we cannot obtain from (9.5) any non-trivial solution of (9.4) if we assume $\hat{f}$ to be a function. But the equation (9.5) can have non-trivial solutions in the class of distributions. For example, if $P(\xi)$ has a zero of the second order at the point η then (9.5) is fulfilled for $\hat{f} = \alpha\delta_\eta + \beta\delta'_\eta$ with all $\alpha, \beta \in \mathbb{C}^1$. Taking the inverse Fourier transform we obtain non-trivial solutions of (9.4), which appear to be infinitely smooth functions.

Divergent integrals.

We have defined the Fourier transform for any distribution $f \in \mathcal{S}'(\mathbb{R}^1)$. In particular, the distribution f can be a bounded function. This means, in fact, that we have defined the integral

$$\int e^{-ix\xi}\, f(x)\, dx$$

for an arbitrary bounded measurable function f. Of course, the integral does not converge in the classical sense, but it can be understood as a distribution in ξ.

This idea can be generalized in different directions. It often happens that a divergent integral depending on a parameter can be interpreted as a distribution in this parameter. Moreover, the corresponding distribution may well be a function even if the integral does not converge.

Example 9.6. Let f be a bounded measurable function. Let us consider the integral

$$\int e^{itx^2}\, f(x)\, dx\,. \tag{9.6}$$

We formally write

$$\int e^{-itx^2}\, f(x)\, dx \;=\; \int_{-1}^{1} e^{-itx^2}\, f(x)\, dx \;+\; \frac{d}{dt}\left(\int_{|x|>1} ix^{-2}\, e^{-itx^2}\, f(x)\, dx\right) \tag{9.7}$$

and accept the right hand side of (9.7) as the definition of the integral (9.6). The first term in the right hand side of (9.7) is a function of t, the second term is a derivative of a function which we understand as a distribution. One can prove that if f is infinitely differentiable with uniformly bounded derivatives, then the distribution on the right hand side of (9.7) coincides with an infinitely smooth function for $t \neq 0$.