Fourier Analysis

TMA4170

Miłosz Krupski

Ferdinand Hodler, *Lake Thun with symmetrical reflection*, 1905
Musée d'Art et d'Histoire (Geneva)

Based on translations of lecture notes by J. Dziubański, G. Łysik, G. Plebanek and R. Szwarc (originally in Polish), with generous permissions of the authors, and other sources

Contents

CHAPTER 1

Introduction

1. Heat equation.

Consider a metal rod of length l with a given initial distribution of temperature. We want to describe how the heat dissipates over time. Temperature of the rod at point x satisfies

$$\frac{\partial}{\partial t}u(t,x) = \frac{\partial^2}{\partial x^2}u(t,x), \quad t > 0, \quad 0 < x < l. \tag{1}$$

Suppose the initial distribution of heat at time $t = 0$ is given by

$$u(0,x) = f(x), \quad 0 \leqslant x \leqslant l \tag{2}$$

and the ends of the rod have the same temperature

$$u(t,0) = u(t,l) = 0, \quad t \geqslant 0 \tag{3}$$

(for simplicity we set the temperature to 0, which could mean that both ends of the rod are immersed in thawing snow). Notice that in order for conditions (2) and (3) to be consistent we need to assume $f(0) = f(l) = 0$.

We may solve this problem using a method developed by Joseph Fourier in 1822[1]. Suppose the solution has the form

$$u(t,x) = T(t) \cdot X(x).$$

Then, by using equation (1), we obtain

$$T'(t) \cdot X(x) = T(t) \cdot X''(x)$$

and

$$\frac{T'(t)}{T(t)} = \frac{X''(x)}{X(x)}.$$

Notice that the function on left-hand side of the last equation depends only on the variable t, while the function on the right-hand side only depends on x. Hence, **both functions must be constant**, i.e.

$$\frac{T'(t)}{T(t)} = \frac{X''(x)}{X(x)} = \lambda.$$

It follows that

$$X''(x) - \lambda X(x) = 0.$$

[1] J. Fourier, *Théorie analytique de la Chaleur*, Didot, Paris, 1822

This is a second-order ordinary differential equation, which we may solve

$$X(x) = \begin{cases} ae^{\sqrt{\lambda}x} + be^{-\sqrt{\lambda}x} & \text{when } \lambda > 0, \\ a\cos\left(\sqrt{-\lambda}x\right) + b\sin\left(\sqrt{-\lambda}x\right) & \text{when } \lambda < 0, \\ a + bx & \text{when } \lambda = 0. \end{cases}$$

The function $u(t,x)$ satisfies the boundary condition (3), therefore

$$X(0) = X(l) = 0.$$

It follows that

$$a = b = 0, \quad \text{when } \lambda > 0 \text{ or } \lambda = 0 \text{ or } \lambda \notin \left\{ -\frac{\pi^2 k^2}{l^2} : k \in \mathbb{N} \right\} ;$$

and

$$a = 0,\ b \text{ — free}, \quad \text{when } \lambda \in \left\{ -\frac{\pi^2 k^2}{l^2} : k \in \mathbb{N} \right\} .$$

Let $k \in \mathbb{N}$ and $\lambda = -\frac{\pi^2 k^2}{l^2}$. We have

$$X(x) = b_k \sin\left(\frac{\pi k}{l} x\right), \tag{4}$$

where b_k may be any real number, as well as

$$T'(t) = \lambda T(t) = -\frac{\pi^2 k^2}{l^2} \cdot T(t).$$

The last equation is solved by

$$T(t) = \exp\left(-\frac{\pi^2 k^2}{l^2} t\right). \tag{5}$$

If we combine (4) and (5), we get

$$u_k(t,x) = b_k \exp\left(-\frac{\pi^2 k^2}{l^2} t\right) \sin\left(\frac{\pi k}{l} x\right).$$

Because equation (1) is linear and u_k satisfy condition (3) for every $k \in \mathbb{N}$, the following sum

$$u_N(t,x) = \sum_{k=0}^{N} b_k \exp\left(-\frac{\pi^2 k^2}{l^2} t\right) \sin\left(\frac{\pi k}{l} x\right)$$

is also a solution for every $N \in \mathbb{N}$. Notice that this sum **is no longer** of the form $u_N(t,x) = T(t)X(x)$.

We still need to consider the initial condition (2). We have

$$u_k(0,x) = b_k \sin\left(\frac{\pi k}{l} x\right).$$

and

$$u_N(0,x) = \sum_{k=0}^{N} b_k \sin\left(\frac{\pi k}{l} x\right). \tag{6}$$

Thus, if the function f may be represented in the form (6), u_N is a solution we are looking for (here we disregard the question whether it is *the only* solution). However, it is clear that even though we may select the coefficients b_k freely, this is a very limited class of functions.

Instead of considering the finite sum (6), it is therefore tangible to ask the following question: *Given a function f in a certain class, can it be represented as a series, or infinite sum,*

$$f(x) = \sum_{k=0}^{\infty} b_k \sin\left(\frac{\pi k}{l} x\right)?$$

2. Trignonometry and complex functions.

Before proceeding further, we need to recall some basic facts in trignonometry and complex analysis.

In the simplest way, sine and cosine are defined by relations between sides in a triangle (Figure 1).

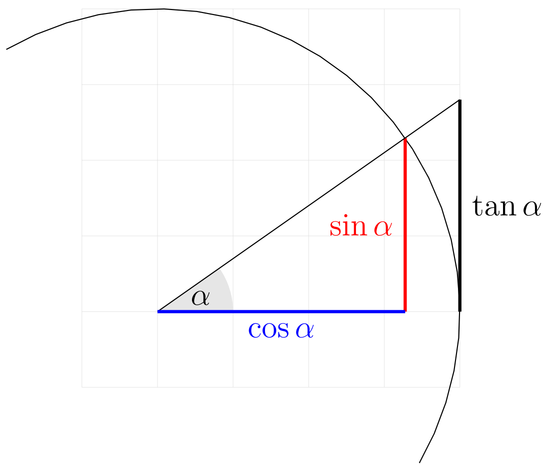

FIGURE 1. Values of $\sin\alpha$, $\cos\alpha$ and $\tan\alpha$ given as lengths of coloured segments in a circle of radius 1.

We also have

$$\frac{d}{dx}\sin x = \cos x, \qquad \frac{d}{dx}\cos x = -\sin x.$$

We may then discover the analytical expressions in the form of power series

$$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n (2n+1)!}{x^{2n+1}}, \qquad \cos x = \sum_{n=0}^{\infty} \frac{(-1)^n (2n)!}{x^{2n}},$$

and use those to define and extend the functions onto the real line, making them 2π-periodic. (Figure 2).

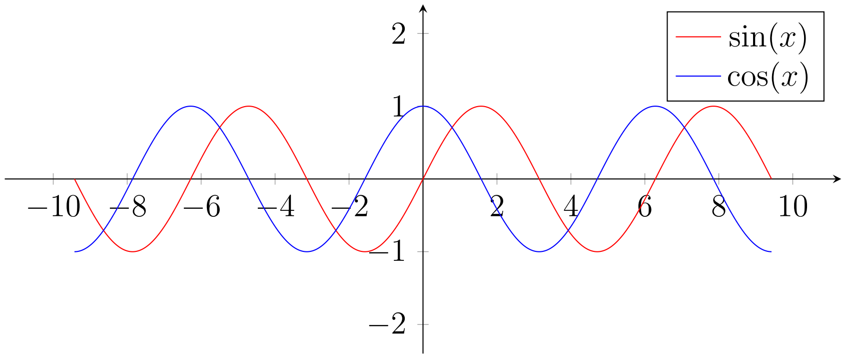

FIGURE 2. Functions $\sin(x)$ and $\cos(x)$ on the real line, x given in radians.

We use sine and cosine to define the other trigonometric functions

$$\tan x = \frac{\sin x}{\cos x}, \qquad \cot x = \frac{\cos x}{\sin x},$$
$$\sec x = \frac{1}{\cos x}, \qquad \csc x = \frac{1}{\sin x},$$

which we call *tangent, cotangent, secant* and *cosecant*, respectively.

Let $z \in \mathbb{C}$. The exponential function is defined by a series

$$e^z = \exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

For $z, w \in \mathbb{C}$ we have

$$e^{z+w} = e^z e^w, \qquad e^{\overline{z}} = \overline{e^z}$$

In particular, $e^{x+iy} = e^x e^{iy}$. If we consider $z = ix$, we have

$$e^{ix} = \sum_{n=0}^{\infty} \frac{(ix)^n}{n!} =$$
$$= \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!} + i \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!} = \cos(x) + i\sin(x).$$

This allows us to write the *Euler formulas*

$$\cos x = \operatorname{Re} e^{ix} = \frac{e^{ix} + e^{-ix}}{2}, \qquad \sin x = \operatorname{Im} e^{ix} = \frac{e^{ix} - e^{-ix}}{2i}.$$

CHAPTER 2

Fourier series

1. Periodic functions.

Definition 1.1. Let $f : \mathbb{R} \to \mathbb{R}$ or $f : \mathbb{R} \to \mathbb{C}$. We say that the function f is p-periodic, $p > 0$ if $f(x + p) = f(x)$ for every $x \in \mathbb{R}$.

It is easy to see that if f is p-periodic, then it is kp-periodic as well for every $k \in \mathbb{N}$.

Definition 1.2. The basic period of a function f is the smallest positive number $p > 0$ such that f is p-periodic, provided such p exists.

Example 1.3. The following function

$$f(x) = \begin{cases} 1 & \text{when } x \in \mathbb{Q}, \\ 0 & \text{when } x \notin \mathbb{Q} \end{cases}$$

does not have a basic period.

In the following we are going to assume that the functions are 2π-periodic. We do not lose any generality, since if a function g is p-periodic, $p > 0$, then $f(x) = g(\frac{px}{2\pi})$ is 2π-periodic and vice-versa.

Notice that 2π-periodic functions may be equivalently treated as functions on a torus (unit circle) $\mathbb{T} = S^1$. Trigonometric functions are basic examples of 2π-periodic functions.

Definition 1.4. The following infinite set of functions

$$\{\sin nx, \cos nx\}_{n \in \mathbb{N}_0} = \{1, \sin x, \cos x, \sin 2x, \cos 2x, \ldots\}$$

is called the trigonometric system.

Lemma 1.5. *The trigonometric system is an orthogonal system, i.e. if $f_j, f_k \in \{\sin nx, \cos nx\}_{n \in \mathbb{N}_0}$, then*

$$\int_0^{2\pi} f_j(x) f_k(x) = c_{ik}\delta_{ik},$$

where $c_{jk} > 0$ are positive numbers.

Lemma 1.6. *The set $\left\{(2\pi)^{-1/2} e^{inx}\right\}_{n \in \mathbb{Z}}$ is an orthonormal system, i.e.*

$$\int_0^{2\pi} \frac{1}{\sqrt{2\pi}} e^{inx}(x) \frac{1}{\sqrt{2\pi}} e^{imx}(x) = \delta_{nm}.$$

Corollary 1.7. *Both systems are linearly independent.*

2. Trigonometric polynomials.

In many applications (recall the example of the heat equation), we are interested in considering only real-valued functions. However, it is often more convenient to study general complex-valued functions.

Definition 2.1. A 2π-periodic function $W : \mathbb{R} \to \mathbb{C}$ which may be expressed in the form

$$W(x) = a_0 + \sum_{k=1}^{N} \big(a_k \cos(kx) + b_k \sin(kx)\big) \tag{7}$$

for (finite sets of) complex coefficients $\{a_k\}$, $\{b_k\}$, is called a trigonometric polynomial.

Notice that the function given by expression (6) is an example of a trigonometric polynomial. A more concise formula to define trigonometric polynomials is the following

$$W(x) = \sum_{k=-N}^{N} c_k e^{ikx}, \qquad \{c_k\} \subset \mathbb{C}. \tag{8}$$

Lemmas 1.5, 1.6 provide us with a method of calculating the coefficients of a trigonometric polynomial. Namely, given a trigonometric polynomial $W(x)$, we have

$$W(x) = \sum_{n\in\mathbb{Z}} c_n e^{inx},$$

where

$$c_n = \frac{1}{2\pi} \int_0^{2\pi} W(x) e^{-inx}\, dx.$$

Notice that only a finite number of coefficients c_n are non-zero.

3. Stone-Weierstrass theorem.

Let K be a compact space, e.g a bounded subset of $\mathbb{R}^d$ or $\mathbb{C}^d$. By $C(K)$ and $C_{\mathbb{R}}(K)$ we denote the spaces of complex-valued and real-valued continuous functions on K. We consider the following metric (norm)

$$d_\infty(f,g) = \|f-g\|_\infty = \sup_{x\in K} |f(x)-g(x)|,$$

which describes the topology of uniform convergence of functions in $C(K)$ and $C_{\mathbb{R}}(K)$.

Definition 3.1. We say that a set $\mathcal{A} \subset C_{\mathbb{R}}(K)$ (or $\mathcal{A} \subset C(K)$) is an **algebra** if $f, g \in \mathcal{A}$ implies $fg \in \mathcal{A}$, $f+g \in \mathcal{A}$ and $cf \in \mathcal{A}$ for every $c \in \mathbb{R}$ (or every $c \in \mathbb{C}$).

Remark 3.2. *Notice that $\|fg\|_\infty \leqslant \|f\|_\infty \|g\|_\infty$, i.e. the $\|\cdot\|_\infty$ norm is **submultiplicative**.*

Examples 3.3. The following sets are algebras

(1) Polynomials $\mathbf{P}$ in $C_{\mathbb{R}}[0,1]$ (or any $C_{\mathbb{R}}[a,b]$).
(2) $\mathcal{A} = \{f \in C_{\mathbb{R}}([0,1]) : f(\frac{1}{2}) = 0\}$.

Lemma 3.4. *If $\mathcal{A} \subset C_{\mathbb{R}}(K)$ is an algebra then $\operatorname{cl}\mathcal{A}$ (uniform limits of functions from $\mathcal{A}$) is also an algebra.*

Proof. If $f, g \in \operatorname{cl}\mathcal{A}$ then we may find sequences $f_n, g_n \in \mathcal{A}$ such that $\|f_n - f\|_\infty \xrightarrow{n} 0$ and $\|g_n - g\|_\infty \xrightarrow{n} 0$. We have

$$\begin{aligned}\|f_n g_n - fg\|_\infty &= \big\|(f_n - f)(g_n - g) + f(g_n - g) + g(f_n - f)\big\|_\infty \\ &\leqslant \|f_n - f\|_\infty \|g_n - g\|_\infty + \|f\|_\infty \|g_n - g\|_\infty + \|g\|_\infty \|f_n - f\|_\infty.\end{aligned}$$

Thus $\|f_n g_n - fg\|_\infty \xrightarrow{n} 0$, which means that $fg \in \operatorname{cl}\mathcal{A}$. In a similar way we show that $f + g \in \operatorname{cl}\mathcal{A}$ and $cf \in \operatorname{cl}\mathcal{A}$ for every $c \in \mathbb{R}$. □

Definition 3.5. We say that an algebra $\mathcal{A} \subset C_{\mathbb{R}}(K)$ (or $C(K)$) **separates points** if for every pair of points $x_1, x_2 \in K$ there exists a function $f \in \mathcal{A}$ such that $f(x_1) \neq f(x_2)$.

Definition 3.6. We say that an algebra $\mathcal{A} \subset C_{\mathbb{R}}(K)$ (or $C(K)$) **does not vanish** in K if for every point $x \in K$ there exists a function $f \in \mathcal{A}$ such that $f(x) \neq 0$.

Example 3.7. Algebra of polynomials $\mathbf{P} \subset C[a,b]$ does not vanish, because $x \mapsto 1 \in \mathbf{P}$. Algebra $\mathbf{P}$ separates points, because the function $x \mapsto x$ is injective (one-to-one).

Lemma 3.8. *If $\mathcal{A} \subset C_{\mathbb{R}}(K)$ is an algebra which separates points and does not vanish in K, then for every pair of points $x_1, x_2 \in K$ and numbers $a_1, a_2 \in \mathbb{R}$, we may find a function $f \in \mathcal{A}$ satisfying $f(x_1) = a_1, f(x_2) = a_2$.*

Proof. There exist functions $h_1, h_2, g \in \mathcal{A}$ such that

$$h_1(x_1) \neq 0, \qquad h_2(x_2) \neq 0, \qquad g(x_1) \neq g(x_2).$$

Let us define functions

$$\begin{aligned}u(x) &= g(x)h_1(x) - g(x_2)h_1(x),\\ v(x) &= g(x)h_2(x) - g(x_1)h_2(x).\end{aligned}$$

Then $u, v \in \mathcal{A}$ and

$$u(x_1) \neq 0, \qquad u(x_2) = 0, \qquad v(x_1) = 0, \qquad v(x_2) \neq 0.$$

Let us notice that the function

$$f(x) = a_1 \frac{u(x)}{u(x_1)} + a_2 \frac{v(x)}{v(x_1)}$$

satisfies the lemma. □

THEOREM 3.9 (Stone-Weierstrass). *Let $\mathcal{A} \subset C_{\mathbb{R}}(K)$ be an algebra which separates points and does not vanish in K. Then $\mathcal{A}$ is dense in $C_{\mathbb{R}}(K)$.*

REMARK 3.10. *The following statements are equivalent*

(1) *$\mathcal{A}$ is dense in $C_{\mathbb{R}}(K)$;*
(2) $\operatorname{cl}\mathcal{A} = C_{\mathbb{R}}(K)$.
(3) *for every continuous function $f \in C_{\mathbb{R}}(K)$ there exists a sequence $f_n \in \mathcal{A}$ such that $f_n \rightrightarrows^n f$ (uniformly) in K;*

Theorem 3.9 is a consequence of the following three lemmas. From now on we always assume that $\mathcal{A} \subset C_{\mathbb{R}}(K)$ is an algebra which separates points and does not vanish in K.

LEMMA 3.11. *If $f \in \operatorname{cl}\mathcal{A}$ then $|f| \in \operatorname{cl}\mathcal{A}$.*

PROOF. We may assume that $f \neq 0$. Consider

$$g = \frac{1}{2}\frac{f}{\|f\|_\infty}.$$

Then $\|g\|_\infty = \frac{1}{2}$. Therefore $|g(x)| \leqslant \frac{1}{2} < 1$ for every $x \in K$ and we know that $g \in \operatorname{cl}\mathcal{A}$. It is enough to show that $|g| \in \operatorname{cl}\mathcal{A}$. It follows from the Weierstrass theorem[1] that there exists a sequence of polynomials $p_n(y)$ such that

$$p_n(y) \rightrightarrows^n |y|, \qquad -1 \leqslant y \leqslant 1.$$

Then

$$\sup_K \big|p_n(g(x)) - |g(x)|\big| \leqslant \sup_{|y|\leqslant 1} \big|p_n(y) - |y|\big| \xrightarrow{n} 0,$$

thus

$$p_n(g(x)) \rightrightarrows^n |g(x)|, \qquad x \in K.$$

Because every polynomial $p_n(y) = \sum_{n=0}^N c_n y^n$ is constructed by only using operations "allowed" in an algebra, we know from Lemma 3.4 that $p_n(g(x)) \in \operatorname{cl}\mathcal{A}$. Therefore $|g(x)| \in \operatorname{cl}\mathcal{A}$. □

LEMMA 3.12. *If $f, g \in \operatorname{cl}\mathcal{A}$ then $\min\{f,g\}, \max\{f,g\} \in \operatorname{cl}\mathcal{A}$.*

PROOF. It follows from Lemma 3.11 and the subsequent relations

$$\min\{f,g\} = \frac{f+g-|f-g|}{2}, \ \max\{f,g\} = \frac{f+g+|f-g|}{2}. \ \square$$

REMARK 3.13. *As an immediate consequence we obtain that if $f_1, f_2, \dots, f_n \in \operatorname{cl}\mathcal{A}$, then $\min\{f_1, f_2, \dots, f_n\}$ and $\max\{f_1, f_2, \dots, f_n\}$ belong to $\operatorname{cl}\mathcal{A}$.*

[1] The sequence of polynomials which approximates $y \mapsto |y|$ may be constructed explicitly. Thus we do not have to rely on the Weierstrass theorem and its original formulation is indeed a special case of the Stone-Weierstrass theorem;

LEMMA 3.14. *Let $f \in C_{\mathbb{R}}(K)$ and $x \in K$. For an arbitrary $\varepsilon > 0$ there exists a function $g_x \in \operatorname{cl}\mathcal{A}$ such that*

$$g_x(x) = f(x)$$
$$g_x(t) > f(t) - \varepsilon, \quad \textit{when } t \in K.$$

PROOF. It follows from Lemma 3.8 that for every $y \in K$ there exists a function $h_y \in \operatorname{cl}\mathcal{A}$ such that

$$h_y(x) = f(x), \qquad h_y(y) = f(y).$$

Because h_y is continuous, there exists a neighbourhood U_y of the point y such that $h_y(t) > f(t) - \varepsilon$, for every $t \in U_y$. The neighbourhoods U_y, $y \in K$ cover the set K. By the property of compactness of K, we may thus choose a finite subcover

$$K = U_{y_1} \cup U_{y_2} \cup \ldots \cup U_{y_n}.$$

Let us define the following function

$$g_x(t) = \max\{h_{y_1}(t), h_{y_2}(t), \ldots, h_{y_n}(t)\}.$$

We use the remark following Lemma (3.12) to show that $g_x \in \operatorname{cl}\mathcal{A}$. We also know that $g_x(x) = f(x)$. Finally, for every $t \in K$, we have $t \in U_{y_j}$ for at least one index $j = 1, 2, \ldots, n$ and so

$$g_x(t) \geqslant h_{y_j}(t) > f(t) - \varepsilon. \qquad \square$$

LEMMA 3.15. *Let $f \in C_{\mathbb{R}}(K)$ and $\varepsilon > 0$. Then there exists a function $h \in \operatorname{cl}\mathcal{A}$ such that*

$$|h(x) - f(x)| < \varepsilon \qquad \textit{for every } x \in K.$$

PROOF. It follows from Lemma 3.14 that for every point $x \in K$ there exists a function $g_x \in \mathcal{A}$ such that

$$g_x(x) = f(x) \quad \text{and} \quad g_x(t) > f(t) - \varepsilon \quad \text{for every } t \in K.$$

Because g_x is continuous, there exists a neighbourhood V_x of the point x such that

$$g_x(t) < f(t) + \varepsilon \qquad \text{for every } t \in V_x.$$

The neighbourhoods V_x cover the set K. We select a finite subcover

$$K = V_{x_1} \cup V_{x_2} \cup \ldots \cup V_{x_n}$$

and define

$$h(t) = \min\{g_{x_1}(t), g_{x_2}(t), \ldots, g_{x_n}(t)\}.$$

It follows from Lemma 3.12 that $h \in \operatorname{cl}\mathcal{A}$. Because $g_{x_j}(t) > f(t) - \varepsilon$ for every $j = 1, 2, \ldots, n$, we also have $h(t) > f(t) - \varepsilon$. For every $t \in K$ we have $t \in V_{x_j}$ for some index $j = 1, 2, \ldots, n$ and so

$$h(t) \leqslant g_{x_j}(t) < f(t) + \varepsilon.$$

As a result

$$|h(t) - f(t)| < \varepsilon, \quad \text{for every } t \in K. \qquad \square$$

We now discuss the complex case.

DEFINITION 3.16. An algebra $\mathcal{A} \subset C(K)$ is called **self-adjoint** if $f \in \mathcal{A}$ implies $\overline{f} \in \mathcal{A}$.

THEOREM 3.17 (Stone-Weierstrass, complex variant). *If $\mathcal{A}$ is a self-adjoint algebra in $C(K)$ which separates points and does not vanish, then $\mathcal{A}$ is dense in $C(K)$.*

PROOF. Let $\mathcal{A}_{\mathbb{R}} = \{f \in \mathcal{A} : f = \overline{f}\} \subset C_{\mathbb{R}}(K)$ and notice that $\mathcal{A}_{\mathbb{R}}$ is an algebra in $C_{\mathbb{R}}(K)$. Let us verify that $\mathcal{A}_{\mathbb{R}}$ satisfies the hypothesis of Theorem 3.9. We know that for $x_1 \neq x_2 \in K$ there exists a function $f \in \mathcal{A}$ such that $f(x_1) \neq f(x_2)$. Therefore

$$\operatorname{Re} f(x_1) \neq \operatorname{Re} f(x_2) \quad \text{or} \quad \operatorname{Im} f(x_1) \neq \operatorname{Im} f(x_2).$$

But $\operatorname{Re} f, \operatorname{Im} f \in \mathcal{A}_{\mathbb{R}}$, because algebra $\mathcal{A}$ is self-adjoint and

$$\operatorname{Re} f = \frac{f + \overline{f}}{2}, \qquad \operatorname{Im} f = \frac{f - \overline{f}}{2i}.$$

Hence $\mathcal{A}_{\mathbb{R}}$ separates points. For every $x \in K$ there exists $f \in \mathcal{A}$ such that $f(x) \neq 0$. Therefore

$$\operatorname{Re} f(x) \neq 0 \quad \text{or} \quad \operatorname{Im} f(x) \neq 0.$$

Hence $\mathcal{A}_{\mathbb{R}}$ does not vanish. By Theorem 3.9 the algebra $\mathcal{A}_{\mathbb{R}}$ is dense in $C_{\mathbb{R}}(K)$.

Let $f \in C(K)$. Then $f = \operatorname{Re} f + i \operatorname{Im} f$ and both functions $\operatorname{Re} f$ and $\operatorname{Im} f$ may be uniformly approximated by functions in $\mathcal{A}_{\mathbb{R}}$. Therefore f may be approximated by functions in $\mathcal{A}$. □

4. Fourier series.

In the same way as in the case of trigonometric series, with an integrable, 2π-periodic function f we may associate a function series, which we denote

$$f \sim \sum_{n \in \mathbb{Z}} c_n e^{inx}, \tag{9}$$

where

$$c_n = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-inx}\, dx. \tag{10}$$

The numbers c_n are called the Fourier coefficients and denoted

$$c_n = \widehat{f}(n).$$

By convergence of a (Fourier) series (9) we mean the convergence of the sequence of its partial sums

$$S_N f(x) = \sum_{|n| \leqslant N} \widehat{f}(n) e^{inx}.$$

We already know that in the case when f is a trigonometric polynomials, we have

$$f(x) = \sum_{n \in \mathbb{Z}} c_n e^{inx} = \lim_{N \to \infty} S_N f(x) \qquad \text{fo every } x \in \mathbb{R}^d.$$

Our aim now is twofold. First, we want to find as many as possible functions for which the series (9) converges, and converges to the function f. Second, when the series diverges, we want to describe other modes of convergence such that we can still say that $f(x) \equiv \sum_{n \in \mathbb{Z}} c_n e^{inx}$.

Notice that the Stone-Weierstrass theorem is not sufficient for convergence, even when the function is continuous: we may find trigonometric polynomials arbitrarily close (in the uniform sense) to any f, but they need not match the partial sums of the series (9) (think of the relation between the Taylor expansion and analytical functions on one hand, and the Weierstrass theorem about polynomials on the other). In fact, it turns out that there are continuous functions with divergent Fourier series.

5. Dirichlet kernel.

Let us denote the n-th partial sum of the Fourier series of a function $f : \mathbb{T} \to \mathbb{C}$ by $S_n f$

$$\begin{aligned} S_n f(x) = \sum_{|k| \leqslant n} \widehat{f}(k) e^{ikx} &= \sum_{|k| \leqslant n} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-ikt} e^{ikx} \, dt \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \sum_{|k| \leqslant n} e^{ik(x-t)} \, dt. \end{aligned}$$

If we define

$$D_n(x) = \sum_{|k| \leqslant n} e^{ikx} \tag{11}$$

then we obtain

$$S_n f(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) D_n(x-t) \, dt.$$

Definition 5.1. For functions u, v on $\mathbb{T}$ the following operation (in the alebraic sense)

$$(u * v)(x) = \int_{-\pi}^{\pi} u(t) v(x-t) \, dt$$

is called the **convolution** (we omit the question of when it is well-defined, and how can the underlying group be described).

Hence

$$S_n f(x) = \frac{1}{2\pi} f * D_n(x).$$

Let us calculate the (finite) sum (11) which defines $D_n(t)$

$$D_n(t) = \sum_{k=-n}^{n} e^{ikt} = e^{-int} \sum_{k=0}^{2n} e^{ikt} = e^{-int} \frac{e^{i(2n+1)t} - 1}{e^{it} - 1} = \frac{e^{i(n+1)t} - e^{-int}}{e^{it} - 1}. \tag{12}$$

By multiplying the numerator and the denominator by $e^{-it/2}$ and using the Euler formulas, we obtain

$$D_n(t) = \frac{\sin\left((n+\frac{1}{2})t\right)}{\sin\frac{t}{2}}.$$

Notice that D_n is an even function. By using the de l'Hôpital rule we get

$$\lim_{t\to 0} \frac{\sin\left((n+\frac{1}{2})t\right)}{\sin\frac{t}{2}} = \lim_{t\to 0} \frac{(2n+1)\cos\left((n+\frac{1}{2})t\right)}{\cos\frac{t}{2}} = 2n+1,$$

thus D_n may be extended at 0 to a continuous function by putting $D_n(0) = 2n+1$. Moreover,

$$\int_{-\pi}^{\pi} D_n(t)\,dt = 2\pi.$$

By using the Euler formulas again directly in formula (11), we obtain yet another representation

$$D_n(x) = \sum_{|k|\leqslant n} e^{ikx} = e^0 + \sum_{k=1}^{n} \left(e^{ikx} + e^{-ikx}\right) = 1 + 2\sum_{k=1}^{n} \cos(kx), \tag{13}$$

Definition 5.2. The (continuous) function

$$D_n(t) = \begin{cases} \dfrac{\sin\left((n+\frac{1}{2})t\right)}{\sin\frac{t}{2}} & \text{when } t \in [-\pi,\pi]\backslash\{0\} \\ 2n+1 & \text{when } t = 0, \end{cases}$$

is called the **Dirichlet kernel**.

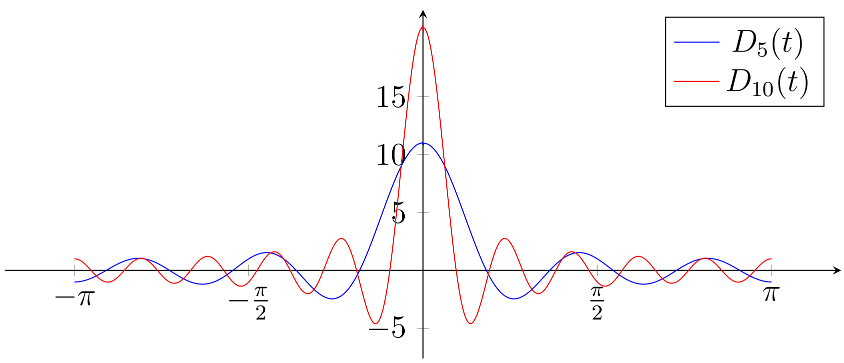

Figure 1. The Dirichlet kernel.

6. Hilbert spaces

Let $\mathcal{H}$ be a linear space over the field $\mathbb{C}$ (or $\mathbb{R}$). Recall that a function $\langle\,\cdot\,,\cdot\,\rangle : \mathcal{H}\times\mathcal{H}\to\mathbb{C}$ (or $\mathbb{R}$) which is

- (conjugate) symmetric

  $\langle x, y\rangle = \overline{\langle y, x\rangle};$

- linear (sesquilinear)

  $\langle ax + by, z\rangle = a\langle x, z\rangle + b\langle y, z\rangle;$

- and positive-definite

  $\langle x, x\rangle > 0, \quad x \in \mathcal{H}\backslash\{0\}.$

is called an inner product on $\mathcal{H}$.

Definition 6.1. Let $\mathcal{H}$ be a linear space over the field $\mathbb{C}$, equipped with an inner product $\langle\,\cdot\,,\cdot\,\rangle$ (we say that $\big(\mathcal{H},\ \langle\,\cdot\,,\cdot\,\rangle\big)$ is an **inner product space**).

An inner product defines the norm (and hence the metric) on $\mathcal{H}$

$$\|x\| = \sqrt{\langle x, x\rangle}, \quad d(x, y) = \|x - y\|.$$

We have the following properties

- polarization formula

  $$\langle x, y\rangle = \frac{1}{4}\sum_{k=0}^{3}\|x + i^k y\|^2 i^k;$$

- parallelogram identity

  $$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + \|y\|^2;$$

- law of cosines

  $$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\,\mathrm{Re}\,\langle x, y\rangle;$$

- Cauchy-Schwarz-Bunyakovsky inequality

  $$|\langle x, y\rangle| \leqslant \|x\|\|y\|.$$

Definition 6.2. We say that the vectors $e_j \in \mathcal{H}$ form an **orthonormal system** if

$$\langle e_j, e_k\rangle = \delta_{jk}, \qquad \|e_j\|^2 = \langle e_j, e_j\rangle = 1,$$

where

$$\delta_{jk} = \begin{cases} 0, & \text{when } j \neq k, \\ 1, & \text{when } j = k. \end{cases}$$

Example 6.3. Consider the space $\mathcal{H} = C(\mathbb{T})$ and let

$$\langle f, g\rangle = \int_0^{2\pi} f(x)g(x)\,dx.$$

Then $\langle\,\cdot\,,\cdot\,\rangle$ is an inner product on $\mathcal{H}$. Let

$$e_n = \frac{1}{\sqrt{2\pi}}e^{inx}$$

then $\{e_n\}$ is an orthonormal system in $\big(\mathcal{H}, \langle\,\cdot\,,\cdot\,\rangle\big)$.

Definition 6.4. An inner product space which is complete is called a **Hilbert space**. A linearly dense orthonormal system is called a **Hilbert basis**, or simply a basis (*not to be confused with a linear basis!*).

Definition 6.5. We denote by $L^2(\mathbb{T})$ the completion of $C(\mathbb{T})$ with respect to the inner product

$$\int_0^{2\pi} f(x)g(x)\,dx.$$

(recall that a continuous function on a closed interval is integrable in both Riemann and Lebesgue sense).

6.1. Bessel inequality.

Theorem 6.6. *Let $\{e_n\}$ be an orthonormal system in an inner product space $\big(\mathcal{H}, \langle\,\cdot\,,\cdot\,\rangle\big)$. For a fixed x let*

$$s_n = \sum_{k=1}^{n} \langle x, e_k\rangle\, e_k$$

and for an arbitrary sequence $\{a_k\}$ let

$$t_n = \sum_{k=1}^{n} a_k e_k.$$

Then

$$\|x - s_n\|^2 \leqslant \|x - t_n\|^2.$$

Moreover, the equality holds only when $a_k = \langle x, e_k\rangle$.

Proof. Let $c_m = \langle x, e_m\rangle$. We have

$$\langle x, t_n\rangle = \sum_{m=1}^{n} c_m \overline{a_m}$$

and

$$\|t_n\|^2 = \sum_{m=1}^{n} |a_m|^2.$$

By orthonormality we get

$$\|x - t_n\|^2 = \|x\|^2 + \|t_n\|^2 - 2\,\mathrm{Re}\,\langle x, t_n\rangle$$
$$= \|x\|^2 + \sum_{m=1}^{n} \left(|a_m|^2 - 2\,\mathrm{Re}\, c_m \overline{a_m}\right)$$
$$= \|x\|^2 - \sum_{m=1}^{n} |c_m|^2 + \sum_{m=1}^{n} |a_m - c_m|^2.$$

The last expression attains its minimum when $a_m = c_m$. By substituting $a_m = c_m$, we obtain the result. □

Corollary 6.7 (Bessel inequality). *If $\{e_n\}$ is an orthonormal system in $(\mathcal{H}, \langle\,\cdot\,,\cdot\,\rangle)$, then for every $x \in \mathcal{H}$ we have*

$$\sum_{k=1}^{n} \left|\langle x, e_k\rangle\right|^2 \leqslant \|x\|^2.$$

Proof. Using notation from the previous theorem, we have

$$\|x\|^2 - \sum_{m=1}^{n} |c_m|^2 = \|x - s_n\|^2 \geqslant 0.$$ □

Corollary 6.8 (Parseval identity). *If an orthonormal system e_j is linearly dense, i.e. the linear combinations of e_j constitute a dense set in $\mathcal{H}$, then*

$$\|x\|^2 = \sum_{k=1}^{\infty} \left|\langle x, e_k\rangle\right|^2.$$

Proof. Let $\varepsilon > 0$ and $\{a_k\}$ be such that $\|x - \sum_{k=1}^{N} a_k e_k\| < \varepsilon$. Denote $c_k = \langle x, e_k\rangle$. Then

$$\left\|x - \sum_{k=1}^{N} c_k e_k\right\| \leqslant \left\|x - \sum_{k=1}^{N} a_k e_k\right\| < \varepsilon.$$

Hence

$$0 \leqslant \left\|x - \sum_{k=1}^{N} c_k e_k\right\| = \|x\|^2 - \sum_{k=1}^{N} |c_k|^2 < \varepsilon.$$ □

Theorem 6.9 (Parseval identity for Fourier series.). *If $f \in L^2(\mathbb{T})$ then we have*

$$\lim_{N\to\infty} \int_{-\pi}^{\pi} \left|f(x) - S_N f(x)\right|^2 dx = 0,$$
$$\int_{-\pi}^{\pi} |f(x)|^2\, dx = 2\pi \sum_{n\in\mathbb{Z}} \left|\widehat{f}(n)\right|^2,$$

where $\widehat{f}(n)$ are given by (10) *and $S_N f = \frac{1}{2\pi} D_N * f$ is the N-th partial sum of the Fourier series of function f.*

Corollary 6.10 (Riemann-Lebesgue Lemma). *If f is an integrable function on $\mathbb{T}$ then* $\lim_{|n|\to\infty} \widehat{f}(n) = 0$.

Proof.
For those who do not know the Lebesgue integral. Let $f \in C(\mathbb{T})$ (or in fact, $f \in L^2(\mathbb{T})$). Then

$$\sum_n |\widehat{f}(n)|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(t)|^2 \, dt.$$

This proves the lemma, because the series is convergent.

For those who do know the Lebesgue integral. Let $f \in L^1(\mathbb{T})$. Take $\varepsilon > 0$ and let $g \in C(\mathbb{T})$ be such that

$$\int_{-\pi}^{\pi} |f - g| \, dx \leqslant \varepsilon.$$

Then

$$\left|\widehat{g}(n)\right| < \varepsilon \qquad \text{for } |n| > N_\varepsilon.$$

Moreover,

$$\begin{aligned}\left|\widehat{f}(n) - \widehat{g}(n)\right| &= \frac{1}{2\pi}\left| \int_{-\pi}^{\pi} \big(f(x) - g(x)\big) e^{-inx} \, dx \right| \\ &\leqslant \frac{1}{2\pi} \int_{-\pi}^{\pi} |f - g| \, dx < \frac{1}{2\pi}\varepsilon,\end{aligned}$$

which proves the lemma. □

7. Pointwise divergence of Fourier series

In Theorem 6.9 we showed convergence of the series of partial sums of a Fourier series $S_N f$ in the norm of the space $L^2(\mathbb{T})$, but we don't know whether the series converges pointwise. It turns out that for a *typical* continuous function, it is not the case.

Definition 7.1. A subset S of a metric space X is **nowhere dense** if the closure $\operatorname{cl} S$ has an empty interior. In other words, for every open ball B in X we have $B \setminus \operatorname{cl} S \neq \varnothing$.

Examples 7.2. A finite subset of the real line is nowhere dense. A countable set $\mathbb{Z} \subset \mathbb{R}$ is nowhere dense. But a countable union of nowhere dense sets may not be nowhere dense, e.g. $\mathbb{Q} \subset \mathbb{R}$ is dense. The Cantor set $C \subset [0, 1]$ is nowhere dense even though it is uncountable.

Definition 7.3. A set S is a **first category** set (or **meagre**) if S is a countable sum of nowhere dense sets in X.

Example 7.4. $\mathbb{Q}$ is a first category set, which is itself not nowhere dense.

REMARK 7.5. *A countable sum of first category sets is a first category set.*

Recall that a bounded linear operator $T : X \to Y$ between normed spaces X and Y is continuous

$$\|Tx_n - Tx\| = \|T(x_n - x)\| \leqslant C\|x_n - x\| \to 0, \quad \text{if } x_n \to x.$$

We define the operator norm as

$$\|T\| = \sup_{x \in X, x \neq 0} \frac{\|Tx\|_Y}{\|x\|_X}.$$

THEOREM 7.6 (Banach-Steinhaus). *Let X and Y be normed spaces. If $\mathcal{F}$ is a family of bounded linear operators from X to Y then either the set of numbers $\{\|T\| : T \in \mathcal{F}\}$ is bounded or*

$$\{x \in X : \sup_{T \in \mathcal{F}} \|Tx\| < \infty\}$$

is a first category set in X.

REMARK 7.7. *If the set $\{\|T\| : T \in \mathcal{F}\}$ is bounded, i.e. there exists a number $c > 0$ such that $\|T\| \leqslant c$ for all $T \in \mathcal{F}$, then*

$$\|Tx\| \leqslant \|T\|\|x\| \leqslant c\|x\|, \quad x \in X, \quad T \in \mathcal{F}.$$

Hence $\{x \in X : \sup_{T \in \mathcal{F}} \|Tx\| < \infty\} = X$.

PROOF. Assume that

$$A = \{x \in X : \sup_{T \in \mathcal{F}} \|Tx\| < \infty\}$$

is not a first category set. Let

$$A_n = \{x \in X : \sup_{T \in \mathcal{F}} \|Tx\| \leqslant n\}.$$

Then

$$A = \bigcup_{n=1}^{\infty} A_n.$$

The sets A_n are closed, because if $x_k \in A_n$ and $x_k \to x$, then

$$\|Tx\| = \lim_k \|Tx_k\| \leqslant n.$$

Therefore for some n the set A_n contains a ball

$$A_n \supset B_r(x_0) = \{x \in X : \|x - x_0\| \leqslant r\}.$$

Let $\|y\| \leqslant r$. Then $y + x_0 \in B_r(x_0) \subset A_n$. Therefore

$$\|Ty\| = \|T(y + x_0) - Tx_0\| \leqslant \|T(y + x_0)\| + \|Tx_0\| \leqslant n + n = 2n.$$

For every $x \neq 0$ the element $y = r\frac{x}{\|x\|}$ satisfies $\|y\| = r$, thus

$$\|Tx\| = \frac{\|Ty\|}{r}\|x\| \leqslant \frac{2n}{r}\|x\|,$$

which means that

$$\|T\| \leqslant \frac{2n}{r}, \quad T \in \mathcal{F}. \qquad \square$$

Corollary 7.8 (duBois Reymond). *There exists a continuous function $f \in C(\mathbb{T})$ such that $S_n f$ diverges at a point.*

Proof. Notice that $f \mapsto T_n f = 2\pi S_n f(0) = D_n * f(0)$ is a family of bounded linear operators, $T_n : C(\mathbb{T}) \to \mathbb{C}$. We are going to show that the set $\{\|T_n\| : n \in \mathbb{N}\}$ is unbounded.

Let ϕ_n be a sequence of continuous functions such that $|\phi_n| \leqslant 1$, each of which is equal to $\operatorname{sgn} D_n$ except for small neighbourhoods of its zeroes. Suppose those neighbourhoods are of the length $(2n)^{-2}$ and we know that D_n has exactly $2n$ zeroes. Denote the sum of those intervals by I_n. Then we have $|I_n| = \frac{1}{2n}$ and

$$\int_{I_n} |D_n(x)|\, dx + \left| \int_{I_n} \phi_n(x) D_n(x)\, dx \right| \leqslant 2|I_n|(2n+1) < 3.$$

Therefore

$$\begin{aligned}
\|T_n\| &\geqslant \|T_n\phi_n\| = \left| \int_{\mathbb{T}} \phi_n(x) D_n(x)\, dx \right| \\
&\geqslant \int_{I_n^c} |D_n(x)|\, dx - \left| \int_{I_n} \phi_n(x) D_n(x)\, dx \right| \\
&= \int_{\mathbb{T}} |D_n(x)|\, dx - \int_{I_n} |D_n(x)|\, dx - \left| \int_{I_n} \phi_n(x) D_n(x)\, dx \right| \\
&\qquad > \frac{1}{\pi} \sum_{k=1}^{n} \frac{1}{k} - 3.
\end{aligned}$$

This means that the set of norms $\|T_n\|$ is unbounded. By the Banach-Steinhaus theorem there exists a function f such that $S_n f(0)$ is not convergent. In fact, the set of such functions is **residual** (a complement of a first category set), i.e. we are extremly lucky if we find a function for which the series converges *at every point.* $\square$

Lemma 7.9. *We have*

$$\int_{-\pi}^{\pi} |D_n| \geqslant \frac{1}{\pi} \sum_{k=1}^{n} \frac{1}{k}$$

Proof. Notice that $\sin x \leqslant x$ for $x \geqslant 0$. We have

$$\begin{aligned}\int_{-\pi}^{\pi} |D_n(x)|\,dx &= 2\int_0^{\pi} \frac{|\sin((n+\frac{1}{2})x|}{\sin\frac{x}{2}}\,dx \\ &\geqslant 2\int_0^{\pi} \frac{|\sin((n+\frac{1}{2})x|}{\frac{x}{2}}\,dx \geqslant \int_0^{\pi(n+\frac{1}{2})} \frac{|\sin x|}{x}\,dx \\ &\geqslant \sum_{k=1}^{n} \int_{\pi(k-1)}^{\pi k} \frac{|\sin x|}{x}\,dx \geqslant \sum_{k=1}^{n} \frac{1}{\pi k}\int_0^1 |\sin x|\,dx.\end{aligned}$$

Hence $\int_{-\pi}^{\pi} |D_n(x)|\,dx \geqslant \frac{1}{\pi}\sum_{k=1}^{n}\frac{1}{k}$. □

8. Pointwise convergence of Fourier series

Notice that the definition of the coefficients of the Fourier series is *non-local*, i.e. by changing a function at any point, the exact values of its Fourier coefficients may change far away from this point.

Nevertheless, we have the following result, which says that a small modification is not going to affect *convergence* of the series at a distance.

Theorem 8.1 (Riemann Localization Principle). *If $f \in C(\mathbb{T})$ is zero in a neighbourhood of x, then* $\lim_{n\to\infty} S_n f(x) = 0$.

Because of linearity, this formulation it is equivalent to saying that if two functions agree in a neighbourhood of x, then their Fourier series behave in the same way at x.

Proof. Suppose that $f(t) = 0$ on $(x-\delta, x+\delta)$. Then

$$S_n f(x) = \int_{\delta\leqslant|t|\leqslant\pi} f(x-t)\frac{\sin((n+\frac{1}{2})t)}{\sin(\frac{t}{2})}\,dt.$$

Let

$$g(t) = \frac{f(x-t)}{2i\sin(\frac{t}{2})}\mathbb{1}_{\delta\leqslant|t|\leqslant\pi}(t).$$

Then $g(t) \in L^2(\mathbb{T})$ and because of the Euler formulas we have

$$\begin{aligned}&f(x-t)\mathbb{1}_{\delta\leqslant|t|\leqslant\pi}(t)\frac{\sin((n+\frac{1}{2})t)}{\sin(\frac{t}{2})} \\ &\qquad = 2ig(t)\sin\left((n+\tfrac{1}{2})t\right) = 2ig(t)\frac{e^{i(n+\frac{1}{2})t} - e^{-i(n+\frac{1}{2})t}}{2i} \\ &\qquad = g(t)e^{it/2}e^{int} - g(t)e^{-it/2}e^{-int}.\end{aligned}$$

Thus for $g_1(t) = g(t)e^{it/2}$ and $g_2(t) = g(t)e^{-it/2}$ we obtain

$$\begin{aligned}S_n f(x) &= \int_{-\pi}^{\pi} \left(g(t)e^{it/2}e^{int} - g(t)e^{-it/2}e^{-int}\right)dt \\ &= \widehat{g_1}(n) + \widehat{g_2}(-n)\end{aligned}$$

By the Riemann-Lebesgue lemma we have $\widehat{g_1}(n) \to 0$ and $\widehat{g_2}(-n) \to 0$ and we conclude that

$$\lim_{n\to\infty} S_n f(x) = 0. \qquad \square$$

DEFINITION 8.2. We say that f is a functon of bounded variation on an interval $[a, b]$ if it is the difference of two bounded monotone functions[2].

We also introduce the following notation for right and left limits of a function at a point

$$f(x^+) = \lim_{y\to x^+} f(y) \quad f(x^-) = \lim_{y\to x^-} f(y).$$

The distinction is of course only important if the function is not continuous at x.

LEMMA 8.3 (Second mean value theorem for integrals). *If h is positive monotonically increasing function on $[a, b]$ and ϕ is integrable on $[a, b]$, then there exists $c \in [a, b)$ such that*

$$\int_a^b h(x)\phi(x)\,dx = h(b^-)\int_c^b \phi(x)\,dx$$

PROOF. It is left as an exercise. $\square$

THEOREM 8.4 (Jordan Criterion). *If f is a function of bounded variation in a neighbourhood of x, then*

$$\lim_{n\to\infty} S_n f(x) = \frac{1}{2}\big(f(x^+) + f(x^-)\big)$$

PROOF. We may assume that f is monotone in a neighbourhood of x. Since

$$\begin{aligned} S_n f(x) = \int_{-\pi}^{\pi} f(t)D_n(x-t)\,dt &= \int_{-\pi}^{\pi} f(x-t)D_n(t)\,dt \\ &= \int_0^{\pi} \big(f(x-t) + f(x+t)\big)D_n(t)\,dt, \end{aligned}$$

it suffices to show that for every monotone g we have

$$\lim_{n\to\infty} \int_0^{\pi} g(t)D_n(t)\,dt = \frac{1}{2}g(0^+).$$

We may also assume that $g(0^+) = 0$ and that g is increasing to the right of 0. Now we need to show that the sequence of integrals converges to 0.

Given $\varepsilon > 0$, choose $\delta > 0$ such that $g(t) < \varepsilon$ if $0 < t < \delta$. Then

$$\int_0^{\pi} g(t)D_n(t)\,dt = \int_0^{\delta} g(t)D_n(t)\,dt + \int_{\delta}^{\pi} g(t)D_n(t)\,dt.$$

[2] The functions of bounded variation are usually defined in a different way, but it requires prior introduction of unnecessary (here, now) formalism.

By the Riemann Localization Principle, the second integral tends to 0. To estimate the first integral, we use the second mean value theorem for integrals. For some y, $0 < y < \delta$ we have

$$\int_0^\delta g(t)D_n(t)\,dt = g(\delta^-)\int_y^\delta D_n(t)\,dt.$$

Furthermore,

$$\begin{aligned}
&\left|\int_y^\delta D_n(t)\,dt\right| \\
&\leqslant \left|\int_y^\delta \sin((n+\tfrac{1}{2})t)\left(\frac{1}{\sin\frac{t}{2}} - \frac{2}{t}\right)dt\right| + \left|\int_y^\delta \frac{\sin((n+\frac{1}{2})t)}{\frac{t}{2}}\,dt\right| \\
&\qquad\leqslant \left|\int_y^\delta \left|\frac{1}{\sin\frac{t}{2}} - \frac{2}{t}\right|dt + 2\sup_{M>0}\left|\int_0^M \frac{\sin t}{t}\,dt\right| < C\right.
\end{aligned}$$

Hence $\int_0^\delta g(t)D_n(t)\,dt < C\varepsilon$. □

Now we prove another criterion. The two are incomparable, i.e. there are examples of functions which satisfy hypotheses of one but not the other, both ways. Other, more general criteria are also known.

Recall that

$$\int_{-\pi}^{\pi} D_n(t)\,dt = \int_{-\pi}^{\pi}\sum_{|k|\leqslant n} e^{ikt}\,dt = \int_{-\pi}^{\pi} e^{i0t}\,dt = 2\pi.$$

THEOREM 8.5 (Dini Criterion). *Let $f \in C(\mathbb{T})$ and $x \in \mathbb{T}$. If there exists $\delta > 0$ such that*

$$\int_{|t|<\delta}\left|\frac{f(x-t)-f(x)}{t}\right|dt < \infty,$$

then $\lim_{n\to\infty} S_nf(x) = f(x)$.

PROOF. Since the integral of D_n equals 2π

$$\begin{aligned}
S_nf(x) - f(x) &= \int_{-\pi}^{\pi}\big(f(x-t)-f(x)\big)\frac{\sin((n+\frac{1}{2})t)}{\sin\frac{t}{2}}\,dt \\
&= \int_{|t|<\delta}\ldots\,dt + \int_{\delta\leqslant|t|\leqslant\pi}\ldots\,dt
\end{aligned}$$

By the Riemann-Lebesgue lemma both of these integrals tend to 0. The second – if we use the Riemann Localization Principle, the first – since we assume the function

$$\frac{f(x-t)-f(x)}{t}\mathbb{1}_{|t|<\delta}(t)$$

to be integrable. Indeed, we have

$$\int_{|t|<\delta}\frac{|f(x-t)-f(x)|}{|\sin\frac{t}{2}|}\,dt = \int_{|t|<\delta}\big|f(x-t)-f(x)\big|\frac{2}{|t|}\cdot\left|\frac{\frac{t}{2}}{\sin\frac{t}{2}}\right|dt$$

and $t \mapsto \frac{t}{\sin t}$ is a function decreasing to 1 for both $t \to 0^+$ and $t \to 0^-$. More particularly,

$$\frac{t}{\sin t} \leqslant \frac{\delta}{\sin \delta} \quad \text{for } 0 < t \leqslant \delta < \pi.$$

This means that

$$\frac{f(x-t) - f(x)}{\sin \frac{t}{2}}$$

is integrable and we use the Euler formulas to argue like in the proof of the Riemann Localization Principle to show that

$$\int_0^\delta \frac{f(x-t) - f(x)}{\sin \frac{t}{2}} \sin((n + \tfrac{1}{2})t)\, dt$$

converges to 0 as $n \to \infty$. □

9. Cesàro means and Fejér kernel

. Consider a numerical series

$$c_0 + c_1 + c_2 + \ldots = \sum_{k=0}^{\infty} c_k$$

and let $S_n = \sum_{k=0}^{n} c_k$ be its partial sums. The series is (conditionally) convergent if the sequence S_n converges. Otherwise, it is divergent.

Notice that the series

$$\sum_{k=0}^{\infty} (-1)^k$$

is divergent. However, the partial sums form the sequence $1, 0, 1, 0, \ldots$ and one may "intuitively" say, that the "limit" of these numbers is equal to $\frac{1}{2}$.

Let us try to give it a precise meaning. Consider the arithmetic mean of the partial sums

$$\sigma_N = \frac{S_0 + S_1 + \ldots + S_{N-1}}{N}.$$

If the series σ_N converges, then we say that the series $\sum c_n$ is **summable in the sense of Cesàro** (it does not make it *convergent*!).

Lemma 9.1. *Let c_n be a numerical sequence and let*

$$S_n = \sum_{k=0}^{n} c_k, \qquad \sigma_N = \frac{S_0 + S_1 + \ldots + S_{N-1}}{N}.$$

If $\lim_{n\to\infty} S_n = a$ *then* $\lim_{n\to\infty} \sigma_n = a$

We leave the proof as an exercise.

DEFINITION 9.2. Consider the arithmetic mean of the Dirichlet kernels

$$K_n(x) = \frac{1}{n+1}\sum_{k=0}^{n} D_k(x).$$

K_n is called the **Fejér kernel**.

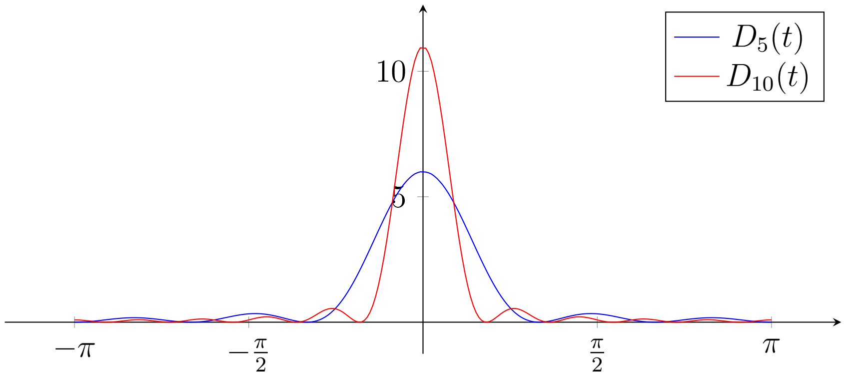

FIGURE 2. The Fejér kernel.

THEOREM 9.3. *The Fejér kernel may also be defined explicitly as*

$$K_n(x) = \frac{1}{n+1}\frac{1-\cos\big((n+1)x\big)}{1-\cos x}$$

for $x \neq 0$ and $K_n(0) = n+1$.

PROOF. We have from identity (12) that

$$(e^{ix}-1)D_n(x) = e^{i(n+1)x} - e^{-inx}.$$

Then notice the following identities

$$\begin{aligned}(e^{ix}-1)(e^{-ix}-1) &= 1 - e^{ix} - e^{-ix} + 1 = 2 - 2\cos x,\\ \frac{e^{-ix}-1}{e^{ix}-1} &= -e^{-ix}.\end{aligned}$$

Hence

$$\begin{aligned}(n+1)(2-2\cos x)K_n(x) &= \sum_{k=1}^{n}\big(e^{i(k+1)x} - e^{-ikx}\big)(e^{-ix}-1)\\ &= \left(\frac{e^{ix}\big(e^{i(n+1)x}-1\big)}{e^{ix}-1} - \frac{e^{-i(n+1)x}-1}{e^{-ix}-1}\right)(e^{-ix}-1)\\ &= e^{ix}\big(e^{i(n+1)x}-1\big)(-e^{-ix}) + 1 - e^{-i(n+1)x}\\ &= 1 - e^{i(n+1)x} + 1 - e^{-i(n+1)x}\\ &= 2 - 2\cos\big((n+1)x\big)\end{aligned}$$

and finally

$$K_n(x) = \frac{1}{n+1}\frac{2-2\cos\big((n+1)x\big)}{2-2\cos x}.$$

We also have

$$K_n(0) = \frac{1}{n+1}\sum_{k=0}^{n}(2n+1) = \frac{(n+1)^2}{n+1} = n+1. \qquad \square$$

COROLLARY 9.4. *We have $K_n(x) \geqslant 0$ and $\int_{-\pi}^{\pi} K_n(x)\,dx = 2\pi$.*

DEFINITION 9.5. An approximate identity on $\mathbb{T}$ is a family of integrable functions $\{k_n\}$ with the following three properties:

(1) There exists a constant $c > 0$ such that $\int_{-\pi}^{\pi} |k_n(x)|\,dx \leqslant c$ for all $n \in \mathbb{N}$.
(2) $\int_{-\pi}^{\pi} k_n(x)\,dx = 1$ for all $n \in \mathbb{N}$.
(3) For any neighbourhood $\delta > 0$ we have $\int_{|x|>\delta} |k_n(x)|\,dx \to 0$ as $n \to \infty$.

Notice the subtle difference between the first two properties.

EXAMPLE 9.6. Let $k(x)$ be a continuous function supported within $(-\pi,\pi) \subset \mathbb{R}$ with integral one. Let

$$k_n(x) = nk(nx).$$

Then $k_n(x)$ is an approximate identity on $\mathbb{T}$. Here we consider k_n as a restriction of a function defined on $\mathbb{R}$ to the interval $[-\pi,\pi]$, but we know that $k_n(-\pi) = k_n(\pi) = 0$, so $k_n \in C_{\text{per}}([-\pi,\pi]) \equiv C(\mathbb{T})$.

The last property follows from the fact that

$$\int_{|x|\geqslant n\delta} |k(x)|\,dx = 0 \quad \text{for } n\delta > \pi.$$

LEMMA 9.7. *The Dirichlet kernels multiplied by $\frac{1}{2\pi}$ are* ***not*** *an approximate identity.*

PROOF. While we have $\frac{1}{2\pi}\int_{-\pi}^{\pi} D_n(x)\,dx = 1$, we also know from Lemma 7.9 that $\int_{-\pi}^{\pi} |D_n(x)|\,dx \to \infty$ as $n \to \infty$ (note: the third property in the definition of approximate identities also fails for the Dirichlet kernels). $\square$

LEMMA 9.8. *The Fejér kernels multiplied by $\frac{1}{2\pi}$ are an approximate identity.*

PROOF. We have $K_n(x) \geqslant 0$, hence $|K_n(x)| = K_n(x)$ and

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(x)\,dx = 1.$$

Notice that

$$K_n(x) = \frac{1}{n+1}\frac{1-\cos\big((n+1)x\big)}{1-\cos x} \leqslant \frac{1}{n+1}\frac{2}{1-\cos x}$$

and for $x \in [-\pi, -\delta] \cup [\delta, \pi]$ we have

$$1 - \cos(x) \geqslant 1 - \cos(\delta).$$

Thus for every $\varepsilon > 0$, we may find n large enough that

$$K_n(x) = \frac{1}{n+1} \frac{1 - \cos\big((n+1)x\big)}{1 - \cos x} \leqslant \frac{1}{n+1} \frac{2}{1 - \cos\delta} \leqslant \varepsilon$$

i.e. for every $\delta > 0$ the sequence of functions K_n converges to 0 uniformly on $[-\pi, -\delta] \cup [\delta, \pi]$. Therefore

$$\int_{|x|>\delta} K_n(x)\, dx \to 0 \quad \text{as } n \to \infty. \qquad \square$$

REMARK 9.9. *Notice that* $1 - \cos x = 2\sin(\frac{x}{2})^2$, *thus the Fejér kernel may also be expressed as*

$$K_n(x) = \frac{1}{n+1} \left(\frac{\sin\big((n+1)\frac{x}{2}\big)}{\sin\frac{x}{2}} \right)^2.$$

THEOREM 9.10. *Let* k_n *be an approximate identity on* $\mathbb{T}$. *If* $f \in C(\mathbb{T})$ *then* $\|k_n * f - f\|_\infty \to 0$ *as* $n \to \infty$.

PROOF. Let $c \geqslant \int_{\mathbb{T}} |k_n(x)|\, dx$. Since f is continuous and $\mathbb{T}$ is compact, we may find $\delta > 0$ such that

$$|f(x-h) - f(x)| < \frac{\varepsilon}{2c} \quad \text{for } |h| < \delta \text{ and every } x \in \mathbb{T}$$

and then find $N_0 > 0$ such that for $n > N_0$ we have

$$\int_{\delta \leqslant |y| \leqslant \pi} |k_n(y)|\, dy < \frac{\varepsilon}{4\|f\|_\infty}.$$

Using these estimates we conclude that

$$\begin{aligned}
\sup_{x\in\mathbb{T}} \big|(k_n * f)(x) - f(x)\big| &= \sup_{x\in\mathbb{T}} \left| \int_{\mathbb{T}} k_n(y) f(x-y)\, dy - f(x) \right| \\
&\leqslant \sup_{x\in\mathbb{T}} \int_{\mathbb{T}} |k_n(y)| \big|f(x-y) - f(x)\big|\, dy \\
&= \sup_{x\in\mathbb{T}} \left(\int_{-\delta}^{\delta} |k_n(y)| \big|f(x-y) - f(x)\big|\, dy \right. \\
&\qquad + \left. \int_{\delta \leqslant |y| \leqslant \pi} |k_n(y)| \big|f(x-y) - f(x)\big|\, dy \right) \\
&\leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,
\end{aligned}$$

which shows that $k_n * f$ converges uniformly to f on $\mathbb{T}$ as $n \to \infty$. $\square$

COROLLARY 9.11. *If* f *is a continuous* 2π*-periodic function then* $\frac{1}{2\pi} f * K_n$ *converges to* f *uniformly.*

We conclude this section with the following observations.

- For $f \in C(\mathbb{T})$, the partial sums of its Fourier series $S_n f$ are examples of trigonometric polynomials.
- The partial sums may be represented by convolutions with Dirichlet kernels $S_n f = D_n * f$.
- The Fejér kernels K_n are the Cesàro means of Dirchlet kernels.
- The Cesàro means $\sigma_n f$ of the Fourier series of the function f are the convolutions with the Fejér kernels $\sigma_n f = K_n * f$.
- The convolutions $K_n * f$ are trigonometric polynomials.
- $K_n * f \rightrightarrows^n f$.

Therefore we now have a constructive method of approximating (uniformly) continuous functions by trigonometric polynomials. Previously, we had to rely on the Stone-Weierstrass theorem, which only says that such an approximation exists (but also covers other algebras, which will be useful when we discuss wavelets).

Moreover, notice that in Theorem 9.10 we could only require the function f to be continuous on some compact set K around an arbitrary point t_0 and then prove that

$$\sup_{x \in K} |k_n * f(x) - f(x)| \to 0 \quad \text{as } n \to \infty.$$

Then we would obtain the following result.

COROLLARY 9.12. *If f is a bounded integrable function which is continuous at t_0 then*

$$\lim_{n\to\infty} \frac{1}{2\pi} f * K_n(t_0) = f(t_0).$$

This means that the Fourier series $\sum_n \widehat{f}(n) e^{int}$ of the function f (bounded) is summable in the sense of Cesàro to the values of f at its points of continuity.

10. Gibbs phenomenon

If a sequence of continuous functions converges uniformly, then its limit is also a continuous function. Thus if a function f is discontinuous at a point x, then its Fourier series cannot be uniformly convergent to f in the neighbourhood of this point. It turns out that the "tip" of the largest "wave" of the Fourier series near the discontinuity point converges to a value which differs from either $f(x^+)$ or $f(x^-)$ by about 9% of the size of the "jump" $|f(x^+) - f(x^-)|$. The same can be said about the smaller "waves", with smaller differences.

This phenomenon is called after Josiah Willard Gibbs who described it in 1899 (it was also observed earlier by Henry Wilbraham in 1848 and studied in more detail by Bôcher in 1906).

The effects of this behaviour have significant impact in applications, because it implies that near a point of discontinuity the function cannot

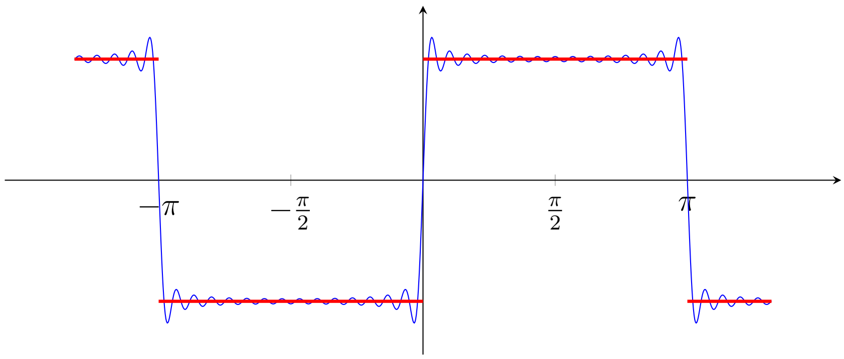

FIGURE 3. Gibbs phenomenon observed for a function $f(x) = \pm\frac{1}{2}$; notice the decreasing sequence of waves near the point of discontinuity.

be well approximated by the partial sums of its Fourier series, no matter how "long" of a sum we consider. Notice that in the digital world, all functions are essentially step functions. Fortunately, the Gibbs phenomenon does not occur when using the Cesàro method of summation (Figure 4) and other methods are also available as a remedy.

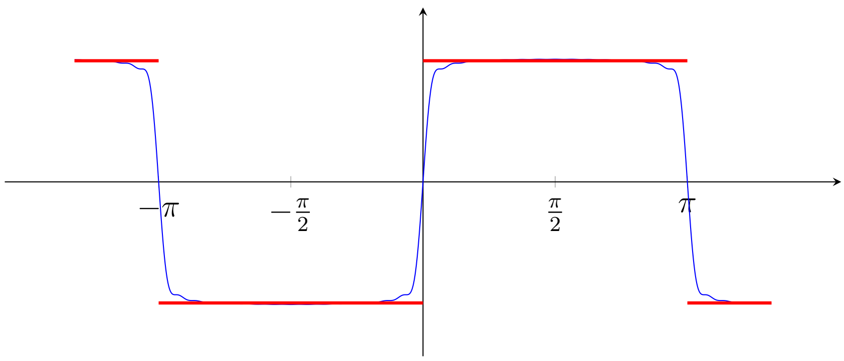

FIGURE 4. Gibbs phenomenon is eliminated by Cesàro summation.

Let us describe this effect on the example shown in Figure 3

$$f(x) = \begin{cases} -\frac{1}{2}, & \text{when } x \in [-\pi, 0), \\ \frac{1}{2}, & \text{when } x \in [0, \pi), \end{cases}$$

whose Fourier series is given by the following series of sines

$$\frac{2}{\pi}\sum_{n=0}^{\infty}\frac{\sin(2n+1)x}{2n+1}.$$

THEOREM 10.1. *If*

$$S_n f(x) = \frac{2}{\pi}\sum_{k=0}^{n}\frac{\sin(2k+1)x}{(2k+1)}.$$

then

$$\lim_{n\to\infty} \max\Big\{S_n f(x) - \frac{1}{2} : 0 \leqslant x \leqslant \frac{\pi}{n}\Big\}$$
$$= \frac{1}{\pi}\int_0^\pi \frac{\sin u}{u}\,du - \frac{1}{2} \approx 0,089.$$

PROOF. Using the Euler formulas in a way similar to the calculation of the Dirichlet kernel in expression (12) we have

$$\begin{aligned}\pi\frac{d}{dx}\big(S_{n-1}f\big)(x) &= \sum_{k=0}^{n-1} 2\cos(2k+1)x = \sum_{k=-n}^{n-1} e^{i(2k+1)x}\\ &= e^{-i(2nx)}\sum_{k=0}^{2n-1} e^{ix}e^{(2ix)k} = e^{-i(2nx)}e^{ix}\cdot\frac{1-e^{2i(2nx)}}{1-e^{2ix}}\\ &= \frac{e^{-i(2nx)}-e^{i(2nx)}}{e^{-ix}-e^{ix}} = \frac{\sin(2nx)}{\sin x}.\end{aligned}$$

Therefore

$$\begin{aligned}\pi S_{n-1}(x) &= \int_0^x \frac{\sin(2ny)}{\sin y}\,dy = \int_0^x \frac{\sin(2ny)}{y}\,dy + W_n(x)\\ &= \int_0^{2nx} \frac{\sin y}{y}\,dy + W_n(x),\end{aligned}$$

where

$$W_n(x) = \int_0^x \Big(\frac{1}{\sin y} - \frac{1}{y}\Big)\sin(2ny)\,dy.$$

Thanks to the de l'Hôpital rule we obtain

$$\begin{aligned}\lim_{y\to 0}\left(\frac{1}{\sin y} - \frac{1}{y}\right) &= \lim_{y\to 0}\frac{y-\sin y}{y\sin y} \overset{(H)}{=} \lim_{y\to 0}\frac{1-\cos y}{\sin y + y\cos y}\\ &\overset{(H)}{=} \lim_{y\to 0}\frac{\sin y}{2\cos y - y\sin y} = 0.\end{aligned}$$

Therefore for an arbitrary $\varepsilon > 0$ we may find a $\delta > 0$ such that for every $x \in (0,\delta)$ and every $n \in \mathbb{N}$ we have

$$W_n(x) \leqslant \int_0^x \Big|\frac{1}{\sin y} - \frac{1}{y}\Big|\,dy < \varepsilon.$$

Hence, for $x \in (0,\delta)$,

$$\Big|S_{n-1}f(x) - \frac{1}{2}\Big| \leqslant \frac{1}{\pi}\int_0^{2nx}\frac{\sin y}{y}\,dy - \frac{1}{2} + \varepsilon.$$

Notice that the function

$$I(x) = \int_0^x \frac{\sin y}{y}\,dy$$

attains its maximum at $x = \pi$. Therefore for every n such that $n > \pi/\delta$ we obtain

$$\max_{0 \leqslant nx \leqslant \pi} \left| S_{n-1}(x) - \frac{1}{2} \right| \leqslant \max_{0 \leqslant nx \leqslant \pi} \frac{1}{\pi} \int_0^{2nx} \frac{\sin y}{y}\, dy - \frac{1}{2} + \varepsilon$$
$$\leqslant \frac{1}{\pi} \int_0^{\pi} \frac{\sin y}{y}\, dy - \frac{1}{2} + \varepsilon \approx 0,089 + \varepsilon. \quad \square$$

11. Curves on the plane and isoperimetric inequality.

Let us try to find a curve on the complex plane which can be laid with piece of twine of a given length, say 2π meters and encloses the largest possible area. We are going to assume that the curve has no sharp edges (is C^1) and does not intersect itself.

Let $\gamma(s) = (x(s), y(s))$ be the parametrization of the curve. Without loss of generality we may assume that $s \in [0, 2\pi]$ and $|\gamma'(s)| = 1$ (we lay the twine while walking at a constant speed, placing each $\frac{s}{100}$-th centimeter of thread at the point $(x(s), y(s))$.

THEOREM 11.1. *Let A be the area bounded by the curve satisfying our assumptions. Then $A \leqslant \pi$. The equality holds if and only if γ is a circle of radius* 1.

PROOF. We know that $x'(s)^2 + y'(s)^2 = 1$. It follows from the Green theorem for path integrals that

$$A = \frac{1}{2} \left| \int_\gamma (x\, dy - y\, dx) \right| = \frac{1}{2} \left| \int_0^{2\pi} x(s) y'(s) - y(s) x'(s)\, ds \right|.$$

Let $x(s) \sim \sum a_n e^{ins}$, $y(s) \sim \sum b_n e^{ins}$ be Fourier series of x and y. Then

$$x'(s) \sim \sum a_n in e^{ins}, \qquad y'(s) \sim \sum b_n in e^{ins}$$

are the Fourier series of their derivatives. From the Parseval identity we have

$$A = \pi \left| \sum_n n(a_n \overline{b_n} - b_n \overline{a_n}) \right|, \qquad \sum_n |n|^2 \Big(|a_n|^2 + |b_n|^2 \Big) = 1.$$

Notice that

$$(14) \qquad |a_n \overline{b_n} - b_n \overline{a_n}| \leqslant 2|a_n||b_n| \leqslant |a_n|^2 + |b_n|^2.$$

Because $|n| \leqslant |n|^2$, we have

$$A \leqslant \pi \sum_n |n|^2 \Big(|a_n|^2 + |b_n|^2 \Big) = \pi.$$

If $A = \pi$ then, because $|n| < |n|^2$ when $|n| > 1$, we have

$$x(s) = a_{-1} e^{-is} + a_0 + a_1 e^{is},$$
$$y(s) = b_{-1} e^{-is} + b_0 + b_1 e^{is}.$$

The functions $x(s)$, $y(s)$ are real. Therefore $a_{-1} = \overline{a_1}$ and $b_{-1} = \overline{b_1}$, which implies $(|a_1|^2 + |b_1|^2) = \frac{1}{2}$. Moreover, inequalities (14) must reduce to equalities. This gives us $|a_1|^2 = |b_1|^2 = \frac{1}{4}$. Hence

$$a_1 = \frac{e^{i\alpha}}{2}, \qquad b_1 = \frac{e^{i\beta}}{2}.$$

From $|a_n\overline{b_n} - b_n\overline{a_n}| = \frac{1}{2}$ it follows that $|\sin(\alpha - \beta)| = 1$, that is $\alpha - \beta = k\pi + \pi/2$. Finally,

$$x(s) = a_0 + \cos(\alpha + s), \qquad y(s) = b_0 \pm \sin(\alpha + s).$$

Let $z_0 = a_0 + ib_0$. Notice that the curve γ is described by the mapping $\gamma(s) = z_0 + e^{i(\alpha \pm s)}$, which is exactly like tracing the unit circle centered at the point z_0, starting at the phase α and going either "left" or "right". $\square$

We now ask kind of an opposite question: given a trigonometric polynomial, or a convergent Fourier series of a continuous function on $\mathbb{T}$, how to draw the curve it descibes on the complex plane?

Recall that

$$c_n = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-int}\, dt,$$

hence

$$c_0 = \frac{1}{2\pi} \int_0^{2\pi} f(t)\, dt,$$

is the mean value of f. If the curve was made of a uniform cord, c_0 would be its centre of mass (arguably, the best approximation of a curve by a single point). Then

$$c_1 = r_1 e^{i\theta_1}, \qquad c_{-1} = r_{-1} e^{i\theta_{-1}}$$

is a pair of vectors and so on for each $|n|$. We are going to treat t as an angle changing from 0 to $2\pi \equiv 0$ in 1 second at a constant speed. At $t = 0$ we have

$$f(0) = \sum_{n \in \mathbb{Z}} c_n,$$

It is simply a sum (possibly infinite) of vectors pointing at the spot on the curve, where we started its parametrization. Then t turns by a small angle as we trace the curve. Let us look at the individual components of the sum

$$f(t) = \sum_{n \in \mathbb{Z}} c_n e^{int}.$$

We may notice that for each $n \in \mathbb{Z}$, the mapping $t \mapsto r_n e^{i(\theta_n + nt)}$ describes *in a unique way* the movement with frequency $|n|$ hertz along a circle of radius r_n centered at 0, which starts at the phase θ_n and travels in one or the other direction.

Thus after time t has passed, on the one hand, we moved along the curve, and on the other, we need to add all the vectors, each of which has turned at its own speed, in its own direction along its own circle. Hence the curve is an image of epicicles of epicicles of epicicles... all the way down. In essence, this is what the geo-centric Ptolemaic view of the Solar system aimed to model. It is not wrong, in this sense, but horrendously complicated.

A beautiful visualization of this behaviour and further explanations may be found on an excellent YouTube channel 3blue1brown:

- https://www.youtube.com/watch?v=r6sGWTCMz2k
- https://www.youtube.com/watch?v=-qgreAUpPwM

12. Temperature of the Earth[3]

Consider the yearly fluctuation of temperature at a given point on Earth and assume it is a periodic function of time. Then the temperature $u(t,x)$ at time $t \geqslant 0$ and depth $x \geqslant 0$ below that point is also periodic in t for every x. It is natural to assume that $\|u\|_\infty \leqslant \|f\|_\infty < \infty$. Let us adjust units of time and space such that the length of the year is 2π and the temperature conductivity of the soil equals $\frac{1}{2}$ (in reality the latter is about $2 \cdot 10^{-3} \frac{\mathrm{cm}^2}{\mathrm{s}}$). Then we have

$$\frac{\partial}{\partial t} u(t,x) = \frac{1}{2} \frac{\partial^2}{\partial x^2} u(t,x)$$

and

$$u(t,x) = \sum_{n \in \mathbb{Z}} c_n(x) e^{int},$$

where

$$c_n(x) = \frac{1}{2\pi} \int_0^{2\pi} u(t,x) e^{-int}\, dt.$$

We thus also have

$$\begin{aligned} \frac{\partial^2}{\partial x^2} c_n(x) &= \frac{1}{2\pi} \int_0^{2\pi} \frac{\partial^2}{\partial x^2} u(t,x) e^{-int}\, dt \\ &= \frac{1}{2\pi} \int_0^{2\pi} 2 \frac{\partial}{\partial t} u(t,x) e^{-int}\, dt \quad = 2in c_n(x) \\ &\qquad = \left(\sqrt{|n|}(1 \pm i)\right)^2 c_n(x), \end{aligned} \tag{15}$$

where the sign is $+$ for $n > 0$ and $-$ for $n < 0$. Moreover,

$$c_n(0) = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{int}\, dt = \widehat{f}(n).$$

A general solution to equation $y'' = ay$ is

$$y(x) = \alpha e^{\sqrt{a}x} + \beta e^{-\sqrt{a}x},$$

[3]Adapted from Dym & McKean and Sommerfeld

but we assume $|c_n(x)| \leqslant \|f\|_\infty$, hence by solving (15) we obtain

$$c_n(x) = \widehat{f}(n)e^{-(\sqrt{|n|}(1\pm i)x)}.$$

In this way we arrive at the solution

$$u(t,x) = \sum_{n\in\mathbb{Z}} \widehat{f}(n)e^{-\sqrt{|n|}x}e^{int\mp i\sqrt{|n|}x}.$$

This means that the temperature at depth x, at the level of n-th "component" e^{int}, is damped by the factor $\exp(-\sqrt{|n|}x)$ and is shifted in time by $\sqrt{|n|}x$.

Let the annual surface temperature be given by the sine function

$$f(t) = \frac{e^{it} - e^{-it}}{2i}.$$

Then $\widehat{f}(1) = \widehat{f}(-1) = \frac{1}{2i}$ and

$$u(t,x) = \frac{e^{-x}e^{i(t-x)} + e^{-x}e^{-i(t-x)}}{2i} = e^{-x}\sin(t-x).$$

At the depth $x = \pi$ the function is damped by $e^{-\pi} \approx \frac{1}{25}$ and completely out of phase with the seasons – warmest in winter and coolest in summer. If we repeat this calculation with true units, we can discover that this depth is about 4 meters, which is therefore the correct choice for a root cellar.

The same calculation may be performed for daily fluctuations of temperature. In this case we may discover that the same phenomenon occurs already at the depth of 20 cm.

CHAPTER 3

Fourier transform

1. Schwartz class

DEFINITION 1.1. The space

$$\mathcal{S}(\mathbb{R}) = \Big\{ f \in C^\infty(\mathbb{R}) : \lim_{|x|\to\infty} |x|^N \big|\tfrac{d^n}{dx^n} f(x)\big| < \infty$$
$$\text{for every pair } n, N \in \mathbb{N} \Big\}$$

is called the **Schwartz class** or the **space of rapidly decaying functions**. Let

$$p_{n,N}(f) = \sup_{x\in\mathbb{R}} |x|^N \big|\tfrac{d^n}{dx^n} f(x)\big|.$$

We say that f_k converges to f in $\mathcal{S}(\mathbb{R})$ if $p_{n,N}(f_k - f) \xrightarrow{k} 0$ for every pair n, N.

In simple words, the space $\mathcal{S}(\mathbb{R})$ contains those smooth functions which together with all their derivatives decay to 0 stronger than any polynomial grows to ∞.

The space $\mathcal{S}(\mathbb{R})$ is linear and completely metrizable, but it is not normed (it is a so-called Fréchet space).

REMARK 1.2. *We may characterize the space $\mathcal{S}(\mathbb{R})$ in another, equivalent way. The function f belongs to $\mathcal{S}(\mathbb{R})$ if and only if for every pair $n, N \in \mathbb{N}$ there exists a constant $C_{n,N}$ such that*

$$\big|\big(\tfrac{d^n}{dx^n} f\big)(x)\big| \leqslant C_{n,N}\big(1 + |x|\big)^{-N}.$$

PROPOSITION 1.3.

- *If $f \in \mathcal{S}(\mathbb{R})$, then $f'(x), xf(x), f(x-h)$ also belong to $\mathcal{S}(\mathbb{R})$;*
- *the class $\mathcal{S}(\mathbb{R})$ is an algebra;*
- *$C_c^\infty(\mathbb{R}^d) \subset \mathcal{S}(\mathbb{R})$;*
- *the function $x \mapsto e^{-x^2}$ belongs to $\mathcal{S}(\mathbb{R})$.*

DEFINITION 1.4. For $f \in \mathcal{S}(\mathbb{R})$ we define

$$\widehat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x\xi}\, dx.$$

We call $\widehat{f}$ the Fourier transform of f.

LEMMA 1.5. *If $f \in \mathcal{S}(\mathbb{R})$ and $n \in \mathbb{N}$ then $\widehat{\tfrac{d^n}{dx^n} f}(\xi) = (2\pi i\xi)^n \widehat{f}(\xi)$ and $\tfrac{d^n}{d\xi^n}\widehat{f} = \big((-2\pi i x)^n f(x)\big)\widehat{\ }$.*

PROOF. We have

$$\int_{-R}^{R} f'(x)e^{-2\pi ix\xi}\,dx$$
$$= \big(f(x)e^{-2\pi ix\xi}\big)\Big|_{x=-R}^{x=R} - 2\pi i\xi \int_{-R}^{R} f(x)e^{2\pi ix\xi}\,dx$$

and by taking the limit with $R \to \infty$ we obtain the result for the first derivative. Now the general result follows from repeating the same argument.

Consider

$$\frac{\widehat{f}(\xi+h) - \widehat{f}(\xi)}{h} - \big(-2\pi i\widehat{xf(x)}\big)(\xi)$$
$$= \int_{-\infty}^{\infty} f(x)e^{-2\pi ix\xi}\left(\frac{e^{-2\pi ixh} - 1}{h} + 2\pi ix\right) dx$$

Notice that

$$\sup_{h\in\mathbb{R}} \left|\frac{e^{-2\pi ixh} - 1}{h}\right| \leqslant C(1 + |x|)$$

and so, because f is a rapidly decaying function, we can find R large enough that

$$\int_{|x|>R} f(x)e^{-2\pi ix\xi}\left(\frac{e^{-2\pi ixh} - 1}{h} + 2\pi ix\right) dx \leqslant \varepsilon$$

independently of h. On the other hand, we may find h small enough that

$$\left|\frac{e^{-2\pi ixh} - 1}{h} + 2\pi ix\right| \leqslant \varepsilon$$

for all $|x| < R$. □

COROLLARY 1.6. *If $f \in \mathcal{S}(\mathbb{R})$ then $\widehat{f} \in \mathcal{S}(\mathbb{R})$.*

PROPOSITION 1.7. *Let $f(x) = e^{-\pi x^2}$ on $\mathbb{R}$. Then $\widehat{f}(\xi) = f(\xi)$.*

PROOF. Notice that

$$\widehat{f}(0) = \int_{-\infty}^{\infty} e^{-\pi x^2}\,dx = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{x^2}\,dx = 1.$$

Moreover,

$$\widehat{f}'(\xi) = \int_{-\infty}^{\infty} -2\pi ixe^{-\pi x^2}e^{-2\pi ix\xi}\,dx$$
$$= i\int_{-\infty}^{\infty} \big(\tfrac{d}{dx}e^{-\pi x^2}\big)e^{-2\pi ix\xi}\,dx = -2\pi\xi\widehat{f}(\xi).$$

Let $g(\xi) = e^{\pi\xi^2}\widehat{f}(\xi)$. Then we have

$$g'(\xi) = e^{\pi\xi^2}\widehat{f}'(\xi) + 2\pi\xi e^{-\pi\xi^2}\widehat{f}(\xi) = 0,$$

which means that g is a constant function and $g(\xi) = g(0)$. But $g(0) = \widehat{f}(0) = 1$ and thus $\widehat{f}(\xi) = e^{-\pi\xi^2}$. $\square$

While $e^{-\pi x^2}$ is the most important example of a **fixed point** of the Fourier transform, it is not the only one. Others include (every fourth of) the Hermite polynomials as well as the hyperbolic secant $\operatorname{sech} x = \frac{2}{e^x + e^{-x}}$.

Definition 1.8. We define the operation of convolution on $\mathcal{S}(\mathbb{R})$ by

$$f * g = \int_{\mathbb{R}} f(y)g(x-y)\,dy.$$

Proposition 1.9. *Let $f, g \in \mathcal{S}(\mathbb{R})$, $a \in \mathbb{C}$, $y \in \mathbb{R}$, $n \in \mathbb{N}$ and $t > 0$. Denote $\tau_y f(x) = f(x+y)$ and $\widetilde{f}(x) = f(-x)$. We have*

(1) $\widehat{f+g} = \widehat{f} + \widehat{g}$;
(2) $\widehat{af} = a\widehat{f}$;
(3) $\widehat{\widetilde{f}} = \widetilde{\widehat{f}}$;
(4) $\widehat{\overline{f}} = \overline{\widetilde{\widehat{f}}}$;
(5) $\widehat{\tau_y f}(\xi) = e^{-2\pi i y\xi}\widehat{f}(\xi)$;
(6) $f * g \in \mathcal{S}(\mathbb{R})$ *and* $\widehat{f * g} = \widehat{f}\widehat{g}$

Definition 1.10. An approximate identity on $\mathbb{R}$ is a family of integrable functions $\{k_t\}$ with the following three properties:

(1) There exists a constant $c > 0$ such that $\int_{-\infty}^{\infty} |k_t(x)|\,dx \leqslant c$ for all $t > 0$.
(2) $\int_{-\infty}^{\infty} k_t(x)\,dx = 1$ for all $t > 0$.
(3) For any neighbourhood $\delta > 0$ we have $\int_{|x|>\delta} |k_t(x)|\,dx \to 0$ as $t \to 0$.

Lemma 1.11. *The family*

$$h_t(x) = \frac{1}{\sqrt{t}} e^{-\frac{\pi x^2}{t}},$$

called the heat kernel, is an approximate identity.

Definition 1.12. For $f \in \mathcal{S}(\mathbb{R})$ we define the **inverse Fourier transform**

$$\check{f}(x) = \widehat{f}(-x) = \int_{\mathbb{R}^d} f(\xi)e^{2\pi i x\xi}\,d\xi.$$

Theorem 1.13 (Parseval-Plancharel identity). *Let $f, g \in \mathcal{S}(\mathbb{R})$. We have*

$$\int_{\mathbb{R}} f(x)\widehat{g}(x)\,dx = \int_{\mathbb{R}} \widehat{f}(x)g(x)\,dx$$

COROLLARY 1.14. *If $f \in \mathcal{S}(\mathbb{R})$ then*

$$(\widehat{f})^{\vee} = f = (\check{f})^{\wedge}$$

PROOF. Let $G_t(x) = e^{-\pi t x^2}$ and notice that $h_t = \widehat{G_t}$. We have

$$\begin{aligned} f(x) &= \lim_{t\to 0} \int_{-\infty}^{\infty} f(x-\xi) h_t(\xi)\, d\xi = \lim_{t\to 0} \int_{-\infty}^{\infty} f(x-\xi)\widehat{G_t}(\xi)\, d\xi \\ &= \lim_{t\to 0} \int_{-\infty}^{\infty} \widehat{f}(\xi) e^{2\pi i x \xi} e^{-\pi t \xi^2}\, d\xi = \int_{-\infty}^{\infty} \widehat{f}(\xi) e^{2\pi i x\xi}\, d\xi \end{aligned}$$

□

COROLLARY 1.15. *Let $f, g \in \mathcal{S}(\mathbb{R})$. We have*

(1)

$$\int_{\mathbb{R}} f(x)\overline{g(x)}\, dx = \int_{\mathbb{R}} \widehat{f}(\xi)\overline{\widehat{g}(\xi)}\, d\xi$$

(2)

$$\|\widehat{f}\|_{L^2(\mathbb{R})} = \|f\|_{L^2(\mathbb{R})} = \|\check{f}\|_{L^2(\mathbb{R})}$$

(3)

$$\int_{\mathbb{R}} f(x)g(x)\, dx = \int_{\mathbb{R}} \widehat{f}(x)\check{g}(x)\, dx$$

LEMMA 1.16. *If $f_k, f \in \mathcal{S}(\mathbb{R})$ and $f_k \to f$ in $\mathcal{S}(\mathbb{R})$, then $\widehat{f_k} \to \widehat{f}$ in $\mathcal{S}(\mathbb{R})$*

COROLLARY 1.17. *The Fourier transform is a homeomophism from $\mathcal{S}(\mathbb{R})$ onto itself.*

In the next theorem we prove that a function and its image under the Fourier transform cannot be simultaneously localized. It is also the mathematical argument behind the Heisenberg uncertainty principle in quantum mechanics.

We measure "localization" by the variance of a probability density, here given by $|\psi|^2$ and $|\widehat{\psi}|^2$. In quantum mechanics, those correspond to wave functions.

THEOREM 1.18 (Uncertainty principle). *Let $\psi \in \mathcal{S}(\mathbb{R})$ be such that $\|\psi\|_{L^2(\mathbb{R})} = 1$. Then*

$$\left(\int_{\mathbb{R}} x^2 |\psi(x)|^2\, dx\right)\left(\int_{\mathbb{R}} \xi^2 |\widehat{\psi}(\xi)|^2\, d\xi\right) \geqslant \frac{1}{16\pi^2}.$$

PROOF. Integrating by parts we obtain

$$\begin{aligned} 1 = \int_{\mathbb{R}} |\psi(x)|^2\, dx &= -\int_{\mathbb{R}} x \tfrac{d}{dx} |\psi(x)|^2\, dx \\ &= -\int_{\mathbb{R}} \Big(x\psi'(x)\overline{\psi(x)} + x\overline{\psi'(x)}\psi(x)\Big)\, dx. \end{aligned}$$

Applying the Cauchy-Schwarz inequality we get

$$\begin{aligned} 1 &\leqslant 2\int_{\mathbb{R}} |x||\psi(x)||\psi'(x)|\,dx \\ &\leqslant 2\Big(\int_{\mathbb{R}} x^2|\psi(x)|^2\,dx\Big)^{1/2}\Big(\int_{\mathbb{R}} |\psi'(x)|^2\,dx\Big)^{1/2}. \end{aligned}$$

Notice that

$$\int_{\mathbb{R}} |\psi'(x)|^2\,dx = 4\pi^2\int_{\mathbb{R}} \xi^2|\widehat{\psi}(\xi)|^2\,d\xi. \qquad \square$$

REMARK 1.19. *The equality in above theorem holds only when*

$$\psi(x) = Ae^{-Bx^2}, \quad B > 0, \quad A^2 = \sqrt{2B/\pi}.$$

REMARK 1.20. *The inequality*

$$\Big(\int_{\mathbb{R}} (x-x_0)^2|\psi(x)|^2\,dx\Big)\Big(\int_{\mathbb{R}} (\xi-\xi_0)^2|\widehat{\psi}(\xi)|^2\,d\xi\Big) \geqslant \frac{1}{16\pi^2}$$

is also true for every $x_0, \xi_0 \in \mathbb{R}$.

2. Elements of measure theory

Consider a non-empty space X. By $\mathcal{P}(X)$ we denote the **power set** of X, i.e. the family of all subsets of X.

DEFINITION 2.1. We say that a family $\mathcal{R} \subseteq \mathcal{P}(X)$ is a **ring** (of subsets of X) if

(1) $\varnothing \in \mathcal{R}$;
(2) if $A, B \in \mathcal{R}$ then $A \cup B, A\backslash B \in \mathcal{R}$.

A family $\mathcal{R}$ is a **field** (or alternatively an **algebra**) if it is a ring and $X \in \mathcal{R}$.

In other words, a field is a family closed under a finite number of operations on sets like taking unions, intersections, complements or differences.

EXAMPLE 2.2. $\{\varnothing\}$ is a ring. If X is an infinite space and $\mathcal{R}$ is a family of all finite subsets of X, then $\mathcal{R}$ is a ring, but it is not a field.

LEMMA 2.3. *Let $\mathcal{R}$ be the family of subsets $A \subset \mathbb{R}$, which may be represented as*

$$A = \bigcup_{k=1}^{n} [a_k, b_k), \tag{16}$$

for some $n \in \mathbb{N}$ and $a_k, b_k \in \mathbb{R}$. Then $\mathcal{R}$ is a ring of subsets of $\mathbb{R}$. Moreover, every $A in \mathcal{R}$ may be represented in a form of (16)*, where $[a_k, b_k)$ are pairwise disjoint.*

PROOF. We have $\varnothing = [0,0) \in \mathcal{R}$; by the very structure of formula (16) it follows that the family $\mathcal{R}$ is closed under finite unions.

Notice that any set of the form $[a,b)\backslash[c,d)$ can only be:

- empty
- an interval $[x,y)$,
- when $a < c < d < b$, a set $[a,c) \cup [d,b) \in \mathcal{R}$.

Now we may use induction to show that $[a,b)\backslash A \in \mathcal{R}$ for every set A given by formula (16). Then it follows that $\mathcal{R}$ is closed under taking a difference of two sets. □

DEFINITION 2.4. We say that a family $\mathcal{A} \subseteq \mathcal{P}(X)$ is a σ-field or a σ-algebra (of subsets of X) if

(1) $\varnothing \in \mathcal{A}$;
(2) if $A \in \mathcal{A}$ then $X\backslash A \in \mathcal{A}$;
(3) if $A_1, A_2, \ldots \in \mathcal{A}$ then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

In other words, a σ-field is a family closed under *countable* operations on sets. Notice that a σ-field is also a ring and a field.

EXAMPLE 2.5. $\{\varnothing, X\}$, $\mathcal{P}(X)$ are σ-fields.

The notion of a σ-field is rather an abstract one and it may be difficult to determine whether a given set belongs to a given σ-field. Most often, however, we decribe such families by well-understood *generating* sets.

PROPOSITION 2.6. *If $\mathcal{A}_\alpha$ are σ-fields, then $\bigcap_\alpha \mathcal{A}_\alpha$ is a σ-field.*

Let $\mathcal{F}$ be any family of sets. Because every σ-field is a subset of $\mathcal{P}(X)$ and because of the above proposition, by considering *all* σ-fields that contain $\mathcal{F}$ and then taking their intersection, we are left with the *smallest* σ-field $\mathcal{A}$ such that $\mathcal{F} \subset \mathcal{A}$. We say that $\mathcal{A}$ is **generated** by the family $\mathcal{F}$ and denote $\mathcal{A} = \sigma(\mathcal{F})$.

DEFINITION 2.7. The σ-field generated by the family of open subsets of X is called the **Borel σ-field** and we denote it by $\mathrm{Bor}(\mathbb{R}^d)$. Elements of $\mathrm{Bor}(\mathbb{R}^d)$ are called **Borel sets**.

LEMMA 2.8. *The family $\mathcal{F}$ of intervals $[p,q)$, where $p,q \in \mathbb{Q}$ generates* $\mathrm{Bor}(\mathbb{R})$.

PROOF. Notice that $[p,q) = \bigcap_{n=1}^{\infty}(p - \frac{1}{n}, q)$, thus $[p,q)$ is a countable intersection of open sets, and hence belongs to $\mathrm{Bor}(\mathbb{R})$. Thus $\mathcal{F} \subseteq \mathrm{Bor}(\mathbb{R})$ and $\sigma(\mathcal{F}) \subseteq \mathrm{Bor}(\mathbb{R})$.

For every $a < b$ we may find sequences of rational numbers p_n, q_n such that $(a,b) = \bigcap_{n=1}^{\infty}[p_n, q_n)$ and thus $(a,b) \in \sigma(\mathcal{F})$

Recall that every open set $U \subset \mathbb{R}$ may be represented as $U = \bigcup(r_n, s_n)$, where $r_n, s_n \in \mathbb{Q}$ ($\mathbb{R}$ is a second-countable space).

It follows that $\mathrm{Bor}(\mathbb{R}) \subseteq \sigma(\mathcal{F})$. □

It is not trivial to show that there exists subsets of $\mathbb{R}$ which are not Borel sets. It is even more difficult to construct such sets.

2.1. Set functions. A function defined on a given family of sets is called a **set function** (it is only a descriptive name, it is still a function in a regular sense). Here we only consider positive extended real-valued set functions, i.e.

$$\mu : \mathcal{F} \to [0, +\infty], \qquad \mathcal{F} \subseteq \mathcal{P}(X).$$

Definition 2.9. We say a set function is additive if for every $E, F \in \mathcal{F}$ such that $E \cap F = \varnothing$ and $E \cup F \in \mathcal{F}$ we have

$$\mu(E \cup F) = \mu(E) + \mu(F)$$

Notice that if only there exists a set E such that $\mu(E) < \infty$ and $\varnothing \in \mathcal{F}$, then $\mu(E) = \mu(E \cup \varnothing) = \mu(E) + \mu(\varnothing)$ and $\mu(\varnothing) = 0$. This means that either $\mu(\varnothing) = 0$ or $\mu(E) = \infty$ for every $E \in \mathcal{F}$.

From now on we will only consider set functions on rings (or σ-fields, which are also rings).

Proposition 2.10. *Let μ be an additive set function on a ring $\mathcal{R}$ and $E, F, E_i \in \mathcal{R}$. We have*

- *if $E \subset F$ then $\mu(E) \leqslant \mu(F)$;*
- *if $E \subset F$ and $\mu(E) < \infty$ then $\mu(F \backslash E) = \mu(F) - \mu(E)$;*
- *if E_i are pairwise disjoint then $\mu\Big(\bigcup_{i=1}^n E_i\Big) = \sum_{i=1}^n \mu(E_i)$.*

Definition 2.11. We say a set function μ on a ring $\mathcal{R}$ is countably additive if for every pairwise disjoint sequence $E_i \in \mathcal{R}$ such that $\bigcup_{i=1}^\infty E_i \in \mathcal{R}$ we have

$$\mu\Big(\bigcup_{n=1}^\infty E_n\Big) = \sum_{n=1}^\infty \mu(E_n).$$

Notice that since the function is extended real-valued, the expression above is meaningful both in the case when the series on the right-and side is convergent and when it diverges to $+\infty$.

Proposition 2.12. *Let μ be a countably additive set function on a ring $\mathcal{R}$. For every sequence $E_n \in \mathcal{R}$ such that $\bigcup_{n=1}^\infty E_n \in \mathcal{R}$ we have*

$$\mu\Big(\bigcup_{i=1}^\infty E_i\Big) \leqslant \sum_{i=1}^\infty \mu(E_i).$$

Proof. Let $A_1 = E_1$ and $A_n = E_n \backslash \bigcup_{i<n} E_i$ for $n > 1$. The sets A_n are pairwise disjoint, $A_n \subseteq E_n$ and $\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty E_n$. Therefore $\mu(A_n) \leqslant \mu(E_n)$ for every n and because of countable additivity

$$\mu\Big(\bigcup_{n=1}^\infty E_n\Big) = \mu\Big(\bigcup_{n=1}^\infty A_n\Big) = \sum_{n=1}^\infty \mu(A_n) \leqslant \sum_{n=1}^\infty \mu(E_n). \qquad \square$$

DEFINITION 2.13. A countably additive set function μ defined on a σ-field and such that $\mu(\varnothing) = 0$ is called a **measure**.

EXAMPLES 2.14.

- Let $f : X \to [0, \infty]$ and let $\mathcal{R}$ be the ring of finite subsets of X. Then
$$\mu(\{x_1, \ldots, x_n\}) = \sum_{k=1}^{n} f(x_k)$$
is an additive set function such that $\mu(\varnothing) = 0$.
- Choose $x_0 \in X$ and let μ be defined on $\mathcal{P}(X)$ by $\mu(A) = 1$ if $x_0 \in A$, $\mu(A) = 0$ if $x_0 \notin A$. Then μ is a measure.

DEFINITION 2.15. A **measure space** is a triplet (X, Σ, μ), where $\Sigma \subseteq \mathcal{P}(X)$ is a σ-field and $\mu : \Sigma \to [0, \infty]$ is a measure. We say a measurable space is

- **finite** if $\mu(X) < \infty$ (or **probabilistic** if $\mu(X) = 1$);
- **σ-finite** if there exist sets $E_n \in \Sigma$ such that $\bigcup_n E_n = X$ and $\mu(E_n) < \infty$ for every n;
- **complete** if for every pair of sets E, F such that $F \subseteq E$ when $E \in \Sigma$ and $\mu(E) = 0$ then $F \in \Sigma$ (and necessarily $\mu(F) = 0$).

2.2. Outer measures.

DEFINITION 2.16. Let μ be a countably additive set function on a ring $\mathcal{R}$. For every $E \subseteq X$ we define the **outer measure** $\mu^* : \mathcal{P}(X) \to [0, \infty]$ by
$$\mu^*(E) = \inf \Big\{ \sum_n \mu(R_n) \ : R_n \in \mathcal{R},\ E \subseteq \bigcup_n R_n \Big\}$$
(we assume that $\inf \varnothing = \infty$).

Throughout the rest of this section we keep the notation for X, $\mathcal{R}$, μ, μ^* etc. to denote relevant objects without change once they are introduced.

PROPOSITION 2.17. *The outer measure μ^* has the following properties*

- $\mu^*(\varnothing) = 0$;
- *if $E \subseteq F \subseteq X$ then $\mu^*(E) \leqslant \mu^*(F)$;*
- *if $E_n \subseteq X$ then $\mu^*\big(\bigcup_n E_n\big) \leqslant \sum_n \mu^*(E_n)$.*

PROOF. The first two statements are easy to prove and the third is obvious if $\mu^*(E_n) = \infty$ for at least single n. Suppose $\mu^*(E_n) \leqslant \infty$ for every n and fix $\varepsilon > 0$. Then there exist $R^n_k \in \mathcal{R}$ such that
$$E_n \subseteq \bigcup_k R^n_k \quad \text{and} \quad \sum_k \mu(R^n_k) \leqslant \mu^*(E_n) + \frac{\varepsilon}{2^n}.$$

Then

$$\bigcup_n E_n \subseteq \bigcup_{n.k} R_k^n$$

and

$$\mu^*\Big(\bigcup_n E_n\Big) \leqslant \sum_n \Big(\mu^*(E_n) + \frac{\varepsilon}{2^n}\Big) = \sum_n \mu^*(E_n) + \varepsilon. \qquad \square$$

Given those properties, we say that an outer measure is *monotone* and *countably subadditive*. In general, it is not countably additive on $\mathcal{P}(X)$, but we are going to prove that it is countably additive on $\sigma(\mathcal{R})$ (and hence is a measure on $\sigma(\mathcal{R})$).

Definition 2.18. We say that a set $E \subseteq X$ is μ^*-measurable if it satisfies the following **Carathéodory condition**

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c) \quad \text{for every } A \subseteq X.$$

By $\mathrm{Meas}(\mu^*)$ we denote the family of all μ^*-measurable sets.

Theorem 2.19. *The family* $\mathrm{Meas}(\mu^*)$ *is a σ-field and μ^* restricted to* $\mathrm{Meas}(\mu^*)$ *is a measure.*

Proof. We have $\varnothing \in \mathrm{Meas}(\mu^*)$ and if $E \in \mathrm{Meas}(\mu^*)$ then $E^c \in \mathrm{Meas}(\mu^*)$. Let $E, F \in \mathrm{Meas}(\mu^*)$ and $A \subseteq X$. Then

$$\begin{aligned} \mu^*(A) &= \mu^*(A \cap E) + \mu^*(A \cap E^c) \\ &= \mu^*(A \cap E \cap F) + \mu^*(A \cap E \cap F^c) + \mu^*(A \cap E^c) \\ &\qquad \geqslant \mu^*(A \cap E \cap F) + \mu^*(A \cap (E \cap F)^c), \end{aligned}$$

because $(A \cap E \cap F^c) \cup (A \cap E^c) \supseteq A \cap (E^c \cup F^c) = A \cap (E \cap F)^c$ and μ^* is subadditive. The converse inequality is always true, thus $E \cap F \in \mathrm{Meas}(\mu^*)$. This means that $\mathrm{Meas}(\mu^*)$ is a field.

Let $E, F \in \mathrm{Meas}(\mu^*)$ be a pair of disjoint sets. Then for $A \subset X$ we have

$$\begin{aligned} \mu^*\big(A \cap (E \cup F)\big) &= \mu^*\big(A \cap (E \cup F) \cap E\big) + \mu^*\big(A \cap (E \cup F) \cap E^c\big) \\ &= \mu^*(A \cap E) + \mu^*(A \cap F). \end{aligned}$$

Let $E_1, E_2 \dots$ be a sequence of disjoint sets in $\mathrm{Meas}(\mu^*)$. By induction we have

$$\mu^*\left(A \cap \Big(\bigcup_{k=1}^n E_n\Big)\right) = \sum_{k=1}^n \mu^*(A \cap E_k).$$

Notice that by taking $A = X$ we obtain that μ^* is an additive set function on $\mathrm{Meas}(\mu^*)$.

Let $E = \bigcup E_n$. Because $\mathrm{Meas}(\mu^*)$ is a field and μ^* is monotone,

$$\mu^*(A) = \mu^*\left(A \cap \left(\bigcup_{k=1}^{n} E_n\right)\right) + \mu^*\left(A \cap \left(\bigcup_{k=1}^{n} E_n\right)^c\right)$$
$$\geqslant \sum_{k=1}^{n} \mu^*(A \cap E_k) + \mu^*(A \cap E^c).$$

Because μ^* is countably subadditive, we have

$$\mu^*(A) \geqslant \sum_{k=1}^{\infty} \mu^*(A \cap E_k) + \mu^*(A \cap E^c) \geqslant \mu^*(A \cap E) + \mu^*(A \cap E^c)$$

and because we always have the converse inequality,

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c).$$

Therefore $E \in \mathrm{Meas}(\mu^*)$, which means that $\mathrm{Meas}(\mu^*)$ is a σ-field. Morover, by taking $A = E$ we get

$$\mu^*(E) = \sum_{k=1}^{\infty} \mu^*(E \cap E_k) + \mu^*(E \cap E^c) = \sum_{k=1}^{\infty} \mu^*(E_k),$$

which means that $\mu^*(E)$ is a measure on $\mathrm{Meas}(\mu^*)$. □

THEOREM 2.20. *We have $\sigma(\mathcal{R}) \subset \mathrm{Meas}(\mu^*)$ and $\mu(R) = \mu^*(R)$ for every $R \in \mathcal{R}$.*

PROOF. Let $R \in \mathcal{R}$ and $A \subseteq X$. If $\mu^*(A) = \infty$, then we have $\mu^*(A) \geqslant \mu^*(A \cap R) + \mu^*(A \cap R^c)$. If $\mu^*(A) < \infty$, then for every $\varepsilon > 0$ there exists a sequence of pairwise disjoint sets $R_n \in \mathcal{R}$ such that $A \subseteq \bigcup_n R_n$ and $\sum_n \mu(R_n) \leqslant \mu^*(A) + \varepsilon$. Then by monotonicity of μ^* and additivity of μ we have

$$\mu^*(A \cap R) + \mu^*(A \cap R^c) \leqslant \sum_n \Big(\mu(R_n \cap R) + \mu(R_n \cap R^c)\Big)$$
$$= \sum_n \mu(R_n) \leqslant \mu^*(A) + \varepsilon,$$

thus $\mu^*(A) \geqslant \mu^*(A \cap R) + \mu^*(A \cap R^c)$ and $R \in \mathrm{Meas}(\mu^*)$. By definition we have $\mu^*(R) \leqslant \mu(R)$. If $R \subseteq \bigcup_n R_n$ for pairwise disjoint $R_n \in \mathcal{R}$ then, because μ is countably additive,

$$\mu(R) = \mu^*(R \cap \bigcup_n R_n) = \sum_n \mu(R \cap R_n) \leqslant \mu^*(R)$$ □

THEOREM 2.21. *Suppose X is σ-finite with respect to μ, i.e. there exist X_n, $\mu(X_n) < \infty$ such that $X = \bigcup_{n=1}^{\infty} X_n$. For every $E \in \mathrm{Meas}(\mu^*)$ there exist $A, B \in \sigma(\mathcal{R})$ such that $A \subseteq E \subseteq B$ and $\mu^*(B \backslash A) = 0$.*

PROOF. First suppose that $X \in \mathcal{R}$ and $\mu(X) < \infty$. Then for $E \in \mathrm{Meas}(\mu^*)$ and $k \in \mathbb{N}$ there exist $R_n^k \in \mathcal{R}$ such that

$$E \subseteq \bigcup_n R_n^k, \quad \mu^*(E) + \frac{1}{k} \geqslant \sum_n \mu(R_n^k).$$

Let

$$B = \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} R_n^k.$$

Then $B \in \sigma(\mathcal{R})$, $E \subseteq B$ and for every $k \in \mathbb{N}$

$$\mu^*(B) \leqslant \sum_{n=1}^{\infty} \mu(R_n^k) \leqslant \mu(E) + \frac{1}{k},$$

hence $\mu^*(E) = \mu^*(B)$. In the same way we may find $C \in \sigma(\mathcal{R})$ such that $E^c \subseteq C$ and $\mu^*(E^c) = \mu^*(C)$. We have $\mu(X) = \mu^*(E) + \mu^*(E^c)$, thus if we let $A = C^c$ then

$$\mu^*(B) = \mu^*(E) = \mu(X) - \mu^*(E^c) = \mu(X) - \mu^*(C) = \mu^*(A).$$

This means that $\mu^*(B \backslash A) = 0$, because μ^* is additive on $\mathrm{Meas}(\mu^*)$ and $\sigma(\mathcal{R}) \subset \mathrm{Meas}(\mu^*)$

In the general case we consider a sequence X_n such that $\mu(X_n) < \infty$ and $X = \bigcup_n X_n$ and a sequence of rings

$$\mathcal{R}_n = \{R \in \mathcal{R} : R \subseteq X_n\}.$$

For $E \subseteq X$ we define $E_n = X \cap X_n$ and using the first part of the proof we may find $A_n, B_n \in \mathcal{R}_n$ such that $A_n \subseteq E_n \subseteq B_n \subseteq X_n$ and $\mu^*(B_n \backslash A_n) = 0$.

Then it suffices to take $A = \bigcup_n A_n$ and $B = \bigcup_n B_n$. □

SUMMARY. The construction of a measure (on a σ-field) from a countably additive set function on a ring is an important one, but it is not essential to remember all the fine details. Below is the summary of the properties that follow from the construction.

THEOREM 2.22. *Let*

- X *be a non-empty space,*
- $\mathcal{R}$ *be a ring of subsets of* X *(Def. 2.1)*
- $\mu : \mathcal{R} \to [0, +\infty]$ *be a countably additive set function (Def. 2.11) such that* $\mu(\varnothing) = 0$
- μ^* *be the outer measure induced by* μ *(Def. 2.16)*
- $\mathrm{Meas}(\mu^*)$ *be the family of* μ^**-measurable sets (Def. 2.18)*

Assume that there exist sets $X_n \in \mathcal{R}$ *such that* $\mu(X_n) < \infty$ *and* $\bigcup_{n=1}^{\infty} = X$. *If*

$$\overline{\sigma(\mathcal{R})} = \{A \cup B : A \in \sigma(\mathcal{R}),\ \mu^*(B) = 0\}$$

then

(1) *The family* $\overline{\sigma(\mathcal{R})}$ *is a* σ*-field and* $\overline{\sigma(\mathcal{R})} = \mathrm{Meas}(\mu^*)$.
(2) *The outer measure* μ^* *is countably additive on* $\overline{\sigma(\mathcal{R})}$, *i.e. it is a measure on* $\overline{\sigma(\mathcal{R})}$.
(3) *We have* $\mu(R) = \mu^*(R)$ *for every* $R \in \mathcal{R}$*; in other words,* μ^* *is an extension of* μ *onto the* σ*-field* $\overline{\sigma(\mathcal{R})}$.
(4) *For every* $E \in \overline{\sigma(\mathcal{R})}$ *and* $\varepsilon > 0$ *there exists* $B = \bigcup_{n=1}^{\infty} R_n$, $R_n \in R$, *such that* $\mu^*(B \backslash E) < \varepsilon$.
(5) *For every* $E \in \overline{\sigma(\mathcal{R})}$ *there exist* $A, B \in \sigma(\mathcal{R})$, *such that* $A \subseteq E \subseteq B$ *and* $\mu^*(B \backslash A) = 0$.

3. Lebesgue measure

Let $\mathcal{R}$ be the ring of subsets $A \subset \mathbb{R}$, which may be represented as

$$A = \bigcup_{n=1}^{N} [a_n, b_n),$$

for some $N \in \mathbb{N}$ and $a_n, b_n \in \mathbb{R}$. Recall that we may assume that the intervals $[a_n, b_n)$ are disjoint.

PROPOSITION 3.1. *Let* $[a_n, b_n)$ *be a finite or infinite family of disjoint intervals. If* $\bigcup_n [a_n, b_n) \subseteq [a, b)$ *then* $\sum_{n=1}^{\infty} (b_n - a_n) \leqslant b - a$.

PROOF. If the family is finite, we obtain the result by induction. If the family is infinite, we notice that for every $N \in \mathbb{N}$ we have

$$\sum_{n=1}^{N} (b_n - a_n) \leqslant b - a$$

(because the result holds for finite families). Therefore the series is convergent and the result follows (the sequence of partial sums is non-decreasing and bounded). □

PROPOSITION 3.2. *Let* $[a_n, b_n)$ *be finite or infinite family of disjoint intervals. If* $[a, b) \subseteq \bigcup_{n=1}^{\infty} [a_n, b_n)$ *then* $b - a \leqslant \sum_{n=1}^{\infty} (b_n - a_n)$.

PROOF. If the family is finite, we prove the result by induction.

For an infinite family the situation is a bit more complicated then before. For a fixed $\varepsilon > 0$ we consider a closed (and hence compact) interval $[a, b - \varepsilon]$. Then $[a, b - \varepsilon) \subseteq \bigcup_{n=1}^{\infty} (a_n - \frac{1}{n}, b_n)$ and because of compactness we may choose a finite sub-covering

$$[a, b - \varepsilon) \subseteq [a, b - \varepsilon] \subseteq \bigcup_{n=1}^{N} \left(a_n - \tfrac{\varepsilon}{2^n}, b_n\right)$$

Then we may use the finite case to obtain

$$b - a - \varepsilon \leqslant \sum_{n=1}^{N} \left(b_n - a_n - \tfrac{\varepsilon}{2^n}\right) \leqslant \sum_{n=1}^{\infty} (b_n - a_n) + \varepsilon$$

and the result follows because ε may be arbitarily small. $\square$

LEMMA 3.3. *Let $\lambda : \mathcal{R} \to [0, \infty]$ be defined by*

$$\lambda\Big(\bigcup_{n=1}^{N}[a_n, b_n)\Big) = \sum_{n=1}^{N} b_n - a_n,$$

where $[a_n, b_n)$ are disjoint intervals. The function λ is well-defined (i.e. it doesn't depend on the choice of representation of any given set).

PROOF. Let

$$R = \bigcup_{n=1}^{N}[a_n, b_n) = \bigcup_{k=1}^{K}[c_k, d_k),$$

where $[a_n, b_n)$ are pairwise disjoint and so are $[c_k, d_k)$. Consider all possible intersections, $I_{n,k} = [a_n, b_n) \cap [c_k, d_k)$. Then every $I_{n,k}$ is either empty or it is an interval. Moreover, for every $n \in \mathbb{N}$ we have $[a_n, b_n) = \bigcup_{k=1}^{K} I_{n,k}$ and by the two previous propositions we know that

$$b_n - a_n = \sum_{k=1}^{K} \lambda(I_{n,k}).$$

We can make a similar observation for intervals $[c_k, d_k)$. Finally, we have

$$\sum_{n=1}^{N}(b_n - a_n) = \sum_{n=1}^{N}\sum_{k=1}^{K} \lambda(I_{n,k}) = \sum_{k=1}^{K}(d_k - c_k) \qquad \square$$

LEMMA 3.4. *The function $\lambda : \mathcal{R} \to [0, \infty]$ is a countably additive set function.*

PROOF. If $[a, b) = \bigcup_{n=1}^{\infty} R_n$ and $R_n = \bigcup_{k=1}^{K_n}[a_k^n, b_k^n)$ then by the two propositions we proved before

$$\lambda\big([a, b)\big) = b - a = \sum_{n=1}^{\infty}\sum_{k=1}^{K_n}(b_k^n - a_k^n) = \sum_{n=1}^{\infty} \lambda(R_n)$$

The general case follows by induction. $\square$

THEOREM 3.5. *There exists a unique measure $\widetilde{\lambda}$ on* Bor($\mathbb{R}$) *such that $\widetilde{\lambda}\big([a, b)\big) = b - a$.*

PROOF. Because λ is a countably additive set function on a ring and $\mathbb{R} = \bigcup_{n=1}^{\infty}[-n, n)$ we can prove existence of $\widetilde{\lambda}$ by consdering the extension of λ to Meas(λ^*) and then the restriction of λ^* to Bor($\mathbb{R}$) $= \sigma(\mathcal{R})$.

We are not going to prove uniqueness (but it is not very difficult). $\square$

DEFINITION 3.6. We call $\widetilde{\lambda} = \lambda$ the **Lebesgue measure**.

DEFINITION 3.7. A measure defined on the σ-field $\mathrm{Bor}(\mathbb{R})$ is called a **Borel measure**.

THEOREM 3.8. *If C is the Cantor ternary set then $\lambda(C) = 0$.*

PROOF. Consider the sequence of sets

$$A_1 = (1/3, 2/3),\ A_2 = (1/9, 2/9) \cup (7/9, 8/9), \ldots$$

that we consecutively "cut out" in the construction of the Cantor set. They are pairwise disjoint and each A_n is itself a finite sum of pairwise disjoint open intervals. We have $\lambda(A_n) = 2^{n-1}\frac{1}{3^n}$. Then

$$\lambda(C) = \lambda\Big([0,1]\backslash \bigcup_{n=1}^{\infty} A_n\Big)$$
$$= 1 - \sum_{n=1}^{\infty} \lambda(A_n) = 1 - \frac{1}{2}\sum_{n=1}^{\infty}(2/3)^n = 0.$$

□

THEOREM 3.9. *The σ-fields $\mathrm{Bor}(\mathbb{R})$ and $\mathrm{Meas}(\lambda^*)$ are not equal. The measure space $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}), \lambda)$ is not complete. The measure space $(\mathbb{R}, \mathrm{Meas}(\lambda^*), \lambda)$ is complete.*

PROOF. It can be shown that the cardinality of $\mathrm{Bor}(\mathbb{R})$ is equal to $\mathfrak{c}$. But the Cantor set has measure zero and therefore all of its subsets are λ^*-measurable. The cardinality of $\mathcal{P}(C)$ is $2^{\mathfrak{c}} > \mathfrak{c}$, hence $\mathrm{Bor}(\mathbb{R}) \neq \mathrm{Meas}(\lambda^*)$ and $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}), \lambda)$ is not complete.

The completeness of $(\mathbb{R}, \mathrm{Meas}(\lambda^*), \lambda)$ follows from the construction of $\mathrm{Meas}(\lambda^*)$. □

THEOREM 3.10. *For every $B \in \mathrm{Bor}(\mathbb{R})$ and $x \in \mathbb{R}$ we have $x + B \in \mathrm{Bor}(\mathbb{R})$ and $\lambda(x + B) = \lambda(B)$.*

PROOF. Let $\mathcal{A}$ be the family of those $B \in \mathrm{Bor}(\mathbb{R})$, for which all translations $x + B$ are Borel sets. Then $\mathcal{A}$ certainly contains all open intervals (a, b). On the other hand $\mathcal{A}$ is σ-field (if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$ and if $A_n \in \mathcal{A}$ then $\bigcup_n A_n \in \mathcal{A}$), hence $\mathcal{A} = \mathrm{Bor}(\mathbb{R})$.

For a fixed x consider a set function μ on $\mathrm{Bor}(\mathbb{R})$, given by $\mu(A) = \lambda(x + A)$ Then μ is a measure and for every $a < b$ we have $\mu([a, b)) = \lambda([x + b, x + b)) = b - a = \lambda([a, b))$. It follows that $\mu(R) = \lambda(R)$ on the ring $\mathcal{R}$ of finite unions of intervals, and hence $\mu(B) = \lambda(B)$ for every $B \in \mathrm{Bor}(\mathbb{R})$ because of uniqueness of the extension. □

REMARK 3.11. *In fact all measures on $\mathbb{R}$ with the property that $\mu(B) = \mu(x + B)$ are given by $c\lambda$ for some $c \geqslant 0$.*

COROLLARY 3.12. *Consider the restriction of λ to the interval $[-\pi, \pi] \cong \mathbb{T}$. For every $B \in \mathrm{Bor}(\mathbb{T})$ and $t \in [-\pi, \pi]$ we have $\lambda(e^{it}B) = \lambda(B)$.*

4. Measurable functions

For every function $f : X \to Y$ and all sets $A \subseteq X$ and $B \subseteq Y$, we define the image of A by

$$f[A] = \{f(x) \in Y : x \in A\},$$

and the pre-image of B

$$f^{-1}[B] = \{x \in X : f(x) \in B\}.$$

Pre-image preserves all set operations, for example

$$f^{-1}\left[\bigcap_n B_n\right] = \bigcap_n f^{-1}[B_n],$$

for every sequence of sets $B_n \subseteq Y$.

Definition 4.1. We say a function $f : X \to Y$ is continuous if the pre-image $f^{-1}[V]$ of every open set $V \subset Y$ is open in X.

Remark 4.2. *Let $f : \mathbb{R} \to \mathbb{R}$ be continuous and $V \subseteq \mathbb{R}$ be an open set. If $x_0 \in f^{-1}[V]$ then $y_0 = f(x_0) \in V$, Becasue V is open, for some $\varepsilon > 0$ we have $(y_0 - \varepsilon, y_0 + \varepsilon) \subseteq V$. Because of the Cauchy definition of continuity of f at x_0, we may find $\delta > 0$, such that $(x_0 - \delta, x_0 + \delta) \subseteq f^{-1}[V]$, which is equivalent to saying that $f^{-1}[V]$ is open.*

Consider a fixed measure space (X, Σ, μ).

Definition 4.3. We say that a function $f : X \to \mathbb{R}$ is Σ–measurable (or simply measurable) if $f^{-1}[B] \in \Sigma$ for every set $B \in \mathrm{Bor}(\mathbb{R})$ (notice that this definiton is independent of μ).

Lemma 4.4. *Let $\mathcal{G} \subseteq \mathrm{Bor}(\mathbb{R})$ be a family of sets, such that $\sigma(\mathcal{G}) = \mathrm{Bor}(\mathbb{R})$, Then a function $f : X \to \mathbb{R}$ is measurable if and only if $f^{-1}[G] \in \Sigma$ for every $G \in \mathcal{G}$.*

Proof. Consider a family $\mathcal{A}$ which consists of those $B \in \mathrm{Bor}(\mathbb{R})$, for which $f^{-1}[B] \in \Sigma$. Then $\mathcal{A}$ is a σ-field: if $A_n \in \mathcal{A}$ and $A = \bigcup_n A_n$ then $f^{-1}[A_n] \in \Sigma$ for every n and $f^{-1}[A] = \bigcup_n f^{-1}[A_n] \in \Sigma$. If $A \in \mathcal{A}$ then also $A^c \in \mathcal{A}$, because

$$f^{-1}[A^c] = \left(f^{-1}[A]\right)^c \in \Sigma.$$

Because $\mathcal{A}$ is a σ-field, then from $\mathcal{G} \subseteq \mathcal{A}$ it follows that $\mathrm{Bor}(\mathbb{R}) = \sigma(\mathcal{G}) \subseteq \mathcal{A}$, thus $\mathcal{A} = \mathrm{Bor}(\mathbb{R})$, which proves that the condition is sufficient. It is clear that it is necessary as well. □

Corollary 4.5. *Each of the following implies that $f : X \to \mathbb{R}$ is measurable:*

- $\{x : f(x) < t\} \in \Sigma$ *for every* $t \in \mathbb{R}$;
- $\{x : f(x) \leqslant t\} \in \Sigma$ *for every* $t \in \mathbb{R}$;
- $\{x : f(x) > t\} \in \Sigma$ *for every* $t \in \mathbb{R}$;
- $\{x : f(x) \geqslant t\} \in \Sigma$ *for every* $t \in \mathbb{R}$.

Proof. Let $\mathcal{G}$ be the family of half-lines $(-\infty, t)$ for $t \in \mathbb{R}$. Then $f^{-1}[G] \in \Sigma$ for $G \in \mathcal{G}$ thus f is measurable, because G generates $\mathrm{Bor}(\mathbb{R})$. □

Corollary 4.6. *If a function $f : \mathbb{R} \to \mathbb{R}$ is continuous then it is measurable with respect to* $\mathrm{Bor}(\mathbb{R})$.

Example 4.7. For every $A \in \Sigma$ a function $\mathbb{1}_A : X \to \mathbb{R}$, where $\mathbb{1}_A(x) = 1$ for $x \in A$ and $\mathbb{1}_A(x) = 0$ for $x \notin A$ is called an indicator (or characteristic) function of a set A. Such a function is measurable, because $\mathbb{1}_A^{-1}[U]$ is an element of the family $\{\varnothing, A, A^c, X\} \subseteq \Sigma$.

For every $B \in \mathrm{Bor}(\mathbb{R})$ the function $\mathbb{1}_B$ is thus a Borel function. Notice that $\mathbb{1}_{\mathbb{Q}}$ is not continuous at any point of the real line, which shows that measurability is a much more general property.

Lemma 4.8. *If a function $f : X \to \mathbb{R}$ is Σ-measurable and a function $g : \mathbb{R} \to \mathbb{R}$ is continuous then a function $g \circ f : X \to \mathbb{R}$ is Σ-measurable.*

Proof. . For every open set $U \subseteq \mathbb{R}$, the set $g^{-1}[U]$ is open because g is continuous; thus

$$(g \circ f)^{-1}[U] = f^{-1}[g^{-1}[U]] \in \Sigma.$$ □

Corollary 4.9. *If a function $f : X \to \mathbb{R}$ is Σ-measurable then*

$$cf, \quad f^2, \quad |f|$$

are also Σ-measurable.

Lemma 4.10. *If functions $f, g : X \to \mathbb{R}$ are Σ-measurable then the function $f + g$ is Σ-measurable.*

Proof. . It suffices to show that for $h = f + g$ and $t \in \mathbb{R}$ we have $h^{-1}[(-\infty, t)] \in \Sigma$. But

$$\{x : f(x) + g(x) < t\} = \bigcup_{\substack{p+q<t,\\ p,q\in\mathbb{Q}}} \{x : f(x) < p\} \cap \{x : g(x) < q\},$$

which can be easily verified, because $\mathbb{Q}$ in dense in $\mathbb{R}$. Notice that the union in the expression above is countable and hence belongs to Σ. □

Corollary 4.11. *If functions $f, g : X \to \mathbb{R}$ are Σ-measurable then functions fg, $\max(f, g)$, $\min(f, g)$ are also measurable.*

Proof.

$$fg = \frac{(f+g)^2 - f^2 - g^2}{2},$$

$$\max(f, g) = \frac{|f-g| + f + g}{2},$$

$$\min(f, g) = \frac{-|f-g| + f + g}{2}.$$

□

It is convenient to consider functions $f : X \to \mathbb{R} \cup \{-\infty, \infty\}$. Then it is natural to assume that Σ-measurability of f in addition means that the sets $f^{-1}(\{-\infty\})$ and $f^{-1}(\{\infty\})$ belong to Σ.

Under such a convention, for any sequence of measurable functions $f_n : X \to \mathbb{R}$, we may define for example $\sup_n f_n$, without the need to assume that the set $\{f_n(x) : n \in \mathbb{N}\}$ is bounded for every $x \in X$. Similarly, we may consider a function $f(x) = \limsup_n f_n(x)$.

LEMMA 4.12. *If functions $f_n : X \to \mathbb{R}$ are Σ-measurable then functions $\liminf_n f_n$, $\limsup_n f_n$, $\inf_n f_n$, $\sup_n f_n$ are also measurable.*

PROOF. We are going to show that the function $f = \limsup_n f_n$ is measurable. It follows from the identities

$$\{x : f(x) = \infty\} = \bigcap_k \bigcap_m \bigcup_{n \geqslant m} \{x : f_n(x) > k\},$$
$$\{x : f(x) \leqslant t\} = \bigcap_k \bigcup_m \bigcap_{n \geqslant m} \{x : f_n(x) < t + 1/k\},$$

and a similar formula for $-\infty$. The second identity follows from the fact that $f(x) \leqslant t$ if and only if for every k almost every element (i.e. all of them except a finite number) of the sequence $f_n(x)$ satisfies $f_n(x) < t + 1/k$. $\square$

COROLLARY 4.13. *A pointwise limit of a conergent sequence of measurable functions is measurable.*

Intitively, every countable operation involving measurable functions leads to a measurable function. For example every function $\mathbb{R} \to \mathbb{R}$ expressed by a formula, which contains countable quantifiers is a Borel function.

LEMMA 4.14. *Every Σ-measurable function $f : X \to \mathbb{R}$ may be expressed as a difference of two measurable and non-negative functions $f = f^+ - f^-$.*

PROOF. Let $f^+ = \max(f, 0)$, $f^- = -\min(f, 0)$ $\square$

4.1. Simple functions.

DEFINITION 4.15. A function $f : X \to \mathbb{R}$ is called **simple** if its range $f[X]$ is a finite set.

An indicator function $\mathbb{1}_A$ for every set $A \subseteq X$ is simple. In fact, all simple functions are finite linear combinations of indicator functions.

LEMMA 4.16. *A function $f : X \to \mathbb{R}$ is simple if and only if*

$$f = \sum_{n=1}^{N} a_n \mathbb{1}_{A_n}$$

for some $a_n \in \mathbb{R}$ and $A_n \subseteq X$. A simple function is Σ-measurable if and only if it is a linear combination of indicator functions of sets in Σ.

PROOF. . If $f[X] = \{a_1, \ldots, a_N\}$ then by taking $A_n = f^{-1}[a_n]$ we get $f = \sum_{n \leqslant N} a_n \mathbb{1}_{A_n}$.

Reversely, for a function of the form $f = \sum_{n \leqslant N} a_n \mathbb{1}_{A_n}$, its range is contained in a finite set consisting of 0 and all numbers which are finite sums of elements of the set $\{a_1, \ldots, a_N\}$. The second statements follows from these remarks. □

THEOREM 4.17. *Let $f : X \to \mathbb{R}$ be a non-negative, Σ-measurable function. Then there exists a sequence of measurable simple functions $s_n : X \to \mathbb{R}$, such that $0 \leqslant s_1(x) \leqslant s_2(x) \leqslant \ldots$, and $\lim_n s_n(x) = f(x)$, for every $x \in X$. Moreover, if the function f is bounded then the sequence s_n may be chosen such that it converges to f uniformly.*

PROOF. Fix n and for every $1 \leqslant k \leqslant n2^n$ let

$$A_{n,k} = \left\{ x : \frac{k-1}{2^n} \leqslant f(x) < \frac{k}{2^n} \right\}.$$

Then $A_{n,k} \in \Sigma$ because the function f is measurable. Let s_n be defined by

$$s_n(x) = \frac{k-1}{2^n} \quad \text{for } x \in A_{n,k}, \quad s_n(x) = n \quad \text{for } f(x) > n.$$

Simple functions s_n defined in this way are measurable and non-negative. If $x \in A_{n,k}$ for some k then $s_n(x) = \frac{k-1}{2^n}$, while

$$s_{n+1}(x) = \frac{k-1}{2^n} \quad \text{or} \quad s_{n+1}(x) = \frac{2k-1}{2^{n+1}},$$

i.e. $s_n(x) \leqslant s_{n+1}(x)$.

For a fixed x and $n > f(x)$ we have $f(x) \geqslant s_n(x) \geqslant f(x) - 1/2^n$, which shows that $\lim_n s_n(x) = f(x)$. If f is bounded then for $n > f[X]$ we have $0 \leqslant f(x) - s_n \leqslant 1/2^n$ uniformly in $x \in X$. □

5. Almost everywhere

THEOREM 5.1. *An additive set function μ on a ring $\mathcal{R}$ is countably additive if and only if it is continuous from below, i.e. for every $A \in \mathcal{R}$ and a sequence $A_n \in \mathcal{R}$ such that $A_n \uparrow A$, we have $\lim_n \mu(A_n) = \mu(A)$.*

PROOF. First, let μ be countably additive. For an increasing sequence of sets $A_n \uparrow A$ let $B_1 = A_1$ and $B_n = A_n \backslash A_{n-1}$ for $n > 1$. Then $A = \bigcup_n B_n$, and B_n are pairwise disjoint. Thus

$$\mu(A) = \mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n)$$

$$= \lim_N \sum_{n=1}^{N} \mu(B_n) = \lim_N \mu(A_N).$$

Now suppose μ is continuous from below and consider pairwise disjoint A_n and $A = \bigcup_n A_n \in \mathcal{R}$. Let $S_N = \bigcup_{n=1}^N A_n$. Then $S_N \uparrow A$ and we have

$$\mu(A) = \lim_N \mu(S_N) = \lim_N \big(\mu(A_1) + \ldots + \mu(A_N)\big) = \sum_n \mu(A_n),$$

i.e. the countable additivity. □

THEOREM 5.2. *For an additive set function μ on a ring $\mathcal{R}$, which only attains finite values, the following conditions are equivalent (below always $A_n, A \in \mathcal{R}$)*

(1) *μ is countably additive;*
(2) *μ is continuous from above, i.e.* $\lim_n \mu(A_n) = \mu(A)$ *if $A_n \downarrow A$;*
(3) *μ is continuous from above on the set $\varnothing$, i.e.* $\lim_n \mu(A_n) = 0$ *if $A_n \downarrow \varnothing$.*

PROOF. (1) $\Rightarrow$ (2) Let $B_n = A_1 \backslash A_n$; then $B_n \uparrow (A_1 \backslash A)$ thus because of the previous theorem

$$\lim_n (A_1 \backslash A_n) = \lim_n \mu(B_n) = \mu(A_1 \backslash A) = \mu(A_1) - \mu(A),$$

which implies $\lim_n \mu(A_n) = \mu(A)$.

(2) $\Rightarrow$ (3) is obvious; we just take $A = \varnothing$.

(3) $\Rightarrow$ (1) Consider pairwise disjoint sets A_n and $A = \bigcup_{n=1}^\infty A_n$. Let $S_n = \bigcup_{n=1}^N A_n$. Then $S_n \uparrow A$ and

$$\mu(A) = \mu(A_1) + \ldots + \mu(A_n) + \mu(A \backslash S_n).$$

Because $\lim_n \mu(A \backslash S_n) = 0$, this implies that the series converges to $\mu(A)$. □

DEFINITION 5.3. For a fixed measure space (X, Σ, μ) and measurable functions $f, g : X \to \mathbb{R}$ we say that $f = g$ **μ-almost everywhere** if

$$\mu\big(\{x : f(x) \neq g(x)\}\big) = 0.$$

Being equal μ-almost everywhere is an equivalence relation. We often write *a.e.* instead of *almost everwhere*.

EXAMPLE 5.4. Identifying functions which are equal almost everywhere has to be done with caution. We have $\mathbb{1}_{\mathbb{Q}} = 0$ λ-almost everywhere, but $\mathbb{1}_{\mathbb{Q}}$ is not continuous at any point, while 0 is. Moreover, $\mathbb{1}_{\mathbb{Q}} \neq 0$ δ_{x_0}-almost everywhere, where δ_{x_0} is the Dirac measure, when $x_0 \in \mathbb{Q}$.

DEFINITION 5.5. A sequence of measurable functions $f_n : X \to \mathbb{R}$ converges $\mu - almosteverywhere$ to a function f if there exists a set $E \in \Sigma$ such that $\mu(E) = 0$ and $\lim_n f_n(x) = f(x)$ for every $x \in X \backslash E$.

EXAMPLE 5.6. Let $X = [0, 1]$ and $f_n(x) = x^n$. Then $f_n \to 0$ λ-almost everywhere, but $f_n \to 1$ δ_1-almost everywhere.

THEOREM 5.7. *For every λ-measurable function f there exists a Borel function g such that $f = g$ λ-almost everywhere.*

We already know that every measurable function is a pointwise limit of a sequence of simple functions and every bounded measurable function is a uniform limit of simple functions.

The sequence $f_n : [0,1] \to \mathbb{R}$, $f_n(x) = x^n$ converges to 0 pointwise, but does not converge uniformly. Notice, however, that for every $\varepsilon > 0$, the sequence f_n does converge uniformly to 0 on the interval $[0, 1-\varepsilon]$. It can be said that removing a small set improves convergence properties of the sequence.

THEOREM 5.8 (Egorov). *If (X, Σ, μ) is a finite measure space and $f_n : X \to \mathbb{R}$ is a sequence of measurable functions converging almost everywhere to f, then for every $\varepsilon > 0$ there exists $A \in \Sigma$ such that $\mu(A) \leqslant \varepsilon$ and f_n converges uniformly to f on the set $X \backslash A$.*

PROOF. Assume that $f(x) = lim_n f_n(x)$ for every $x \in X$. In the general case, we can simply remove the "offending" set of measure zero. For every $m, n \in \mathbb{N}$ consider sets

$$E(m,n) = \bigcap_{k=n}^{\infty} \{x : |f_k(x) - f(x)| < 1/m\}.$$

Then for every m we have $E(m,1) \subseteq E(m,2) \subseteq \ldots$ and

$$\bigcup_n E(m,n) = X,$$

which follows from the fact that $f_k(x) \to f(x)$, which means that for every x there exists k such that $|f_k(x) - f(x)| < 1/m$. Let us fix $\varepsilon > 0$. Because $E(m,n) \uparrow X$, we have $X \backslash E(m,n) \downarrow \varnothing$ and because a finite measure is continuous from above on the empty set, for every m there exists n_m such that

$$\mu\big(X \backslash E(m, n_m)\big) < \varepsilon / 2^m.$$

Then by putting

$$A = \bigcup_m \big(X \;\; E(m, n_m)\big),$$

we obtain

$$\mu(A) \leqslant \sum_m \mu\big(X \;\; E(m, n_m)\big) \leqslant \sum_m \varepsilon / 2^m = \varepsilon.$$

Moreover, $|f_n(x) - f(x)| < 1/m$ for $n > n_m$ and $x \notin A$, which implies uniform convergence of f_n on $X \backslash A$. $\square$

REMARK 5.9. *The assumption $\mu(X) < \infty$ in the Egorov theorem is important. The sequence $f_n(x) = x/n$ on the real line converges pointwise to 0, but it is not uniformly convergent on any unbounded subset of the line.*

The Egorov theorem spurs the following definition.

Definition 5.10. We say that a sequence of measurable functions is **almost uniformly convergent** if form every $\varepsilon > 0$ the sequence f_n converges uniformly on the complement of a set of measure smaller than ε.

6. Lebesgue integral

6.1. Integration of simple functions. Consider a fixed measure space (X, Σ, μ). Our goal is to define the integral, i.e. a linear operator assiging numerical values to functions, which, for a non-negative function, measures the "area under the graph".

Because of this, it is clear how the integral should be defnied for simple functions.

Definition 6.1. If $f = \sum_{n \leqslant N} a_n \mathbb{1}_{A_n}$ for $A_n \in \Sigma$ then we define

$$\int_X f \, d\mu = \sum_{n \leqslant N} a_n \, \mu(A_n),$$

if only the expression on the right-hand side is meaningful (including $\pm\infty$). We say that the function f is **integrable** if $\int_X f \, d\mu$ has a finite value.

Remark 6.2. *For the symbols ∞ and $-\infty$, we assume $x + \infty = \infty$, $x - \infty = -\infty$ for $x \in \mathbb{R}$ as well as $0 \cdot \infty = 0 \cdot (-\infty) = 0$. Expression $\infty - \infty$ has no numerical sense nor value.*

Remark 6.3. *Let $f = 2\mathbb{1}_{[0,1]} + c\mathbb{1}_{[3,\infty]}$. Then*

$$\int_{\mathbb{R}} f \, d\lambda = \begin{cases} 2, & \text{for } c = 0 \\ \infty, & \text{for } c > 0 \\ -\infty & \text{for } c < 0. \end{cases}$$

For the function $g = \mathbb{1}_{[-\infty,0)} - \mathbb{1}_{[1,\infty)}$ the expression $\int_{\mathbb{R}} g \, d\lambda$ has no numerical sense.

Lemma 6.4. *The integral of a simple function is well-defined, i.e. if $f = \sum_{n \leqslant N} a_n \mathbb{1}_{A_n} = \sum_{k \leqslant K} b_k \mathbb{1}_{B_k}$ then*

$$\sum_{n \leqslant N} a_n \, \mu(A_n) = \sum_{k \leqslant K} b_k \, \mu(B_k).$$

Apart from the integral over the entire space X, we may consider the integral over any set $A \in \Sigma$, which we simply define by

$$\int_A f \, d\mu = \int_X f \cdot \mathbb{1}_A \, d\mu.$$

Theorem 6.5. *For a simple measurable function h and simple integrable functions f and g we have the following*

(1) $\int_X (af + bg)\, d\mu = a\int_X f\, d\mu + b\int_X g\, d\mu$;
(2) *if* $h = 0$ *almost everywhere then* $\int_X h\, d\mu = 0$;
(3) *if* $f \leqslant g$ *almost everywhere then* $\int_X f\, d\mu \leqslant \int_X g\, d\mu$;
(4) $|\int_X (f + g)\, d\mu| \leqslant \int_X |f|\, d\mu + \int_X |g|\, d\mu$;
(5) *if* $a \leqslant f \leqslant b$ *almost everywhere then* $a\mu(X) \leqslant \int_X f\, d\mu \leqslant b\mu(X)$;
(6) *for* $A, B \in \Sigma$, *if* $A \cap B = \varnothing$ *then* $\int_{A\cup B} f\, d\mu = \int_A f\, d\mu + \int_B f\, d\mu$.

6.2. Integration of measurables functions. We still assume we work in a fixed, σ-finte measure space (X, Σ, μ), and all functions we discuss are assumed to be Σ-measurable.

First we define the integral of a measurable non-negative function $f : X \to \mathbb{R}$. Notice that if s is a non-negative simple function such that $0 \leqslant s \leqslant f$ and $s = \sum_{n\leqslant N} a_n \mathbb{1}_{A_n}$, where A_n are pairwise disjoint and $a_n \geqslant 0$ then the condition means that, geometrically speaking, the rectangles $A_n \times [0, a_n]$ fit under the graph of the function f and hence we should have $\int_X f\, d\mu \geqslant \int_X s\, d\mu$ (see Figure 1).

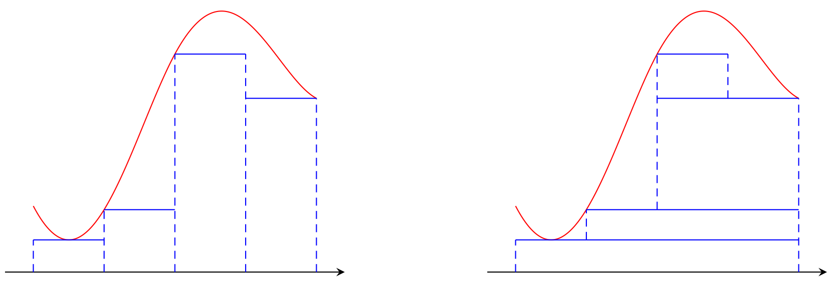

FIGURE 1. Riemann's (Darboux's) idea for approximating the integral (left) compared with Lebesgue's (right).

DEFINITION 6.6. For a non-negative measurable function f we define

$$\int_X f\, d\mu = \sup\left\{ \int_X s\, d\mu : s \text{ is simple and } 0 \leqslant s \leqslant f \right\},$$

The function f is called **integrable**, if the integral $\int_X f\, d\mu$ is finite.

Notice that in fact the integral of a non-negative function f may be defined as the supremum of value $\int_X s\, d\mu$, taken only for simple integrable functions. The following theorem often serves as a more useful definition.

THEOREM 6.7. *If f is a non-negative measurable function, and s_n is a sequence of simple functions, such that $s_1 \leqslant s_2 \leqslant \ldots$ and $\lim_n s_n = f$ almost everywhere then*

$$\int_X f\, d\mu = \lim_n \int_X s_n\, d\mu.$$

Proof. Because the sequence of integrals $\int_X s_n \, d\mu$ is non-decreasing the limit $\lim_n \int_X s_n \, d\mu$, proper or improper, always exists. Thanks to the definition of the integral we have the inequality

$$\int_X f \, d\mu \geqslant \lim_n \int_X s_n \, d\mu.$$

Consider a simple function $g = \sum_{n \leqslant N} a_n \mathbb{1}_{A_n}$ where A_n are pairwise disjoint sets of finite measure and which satisfies $0 \leqslant g \leqslant f$. Then $X_0 = \bigcup_{n \leqslant N} A_n$ has finite measure. Let $M = \max_n a_n$ (the values $\mu(X_0)$ and M are fixed). It follows from the Egorov theorem that s_n converges to f almost uniformly on the set X_0. For a fixed $\varepsilon > 0$ there exists $A \subseteq X_0$ such that $\mu(A) < \varepsilon/M$ and the covergence on $X_0 \backslash A$ is uniform. This means that for large n we have the inequality

$$g(x) - s_n(x) \leqslant f(x) - s_n(x) < \varepsilon/\mu(X_0), \quad \text{for } x \in X_0 \backslash A$$

and hence

$$\begin{aligned} \int_X g \, d\mu = \int_{X_0} g \, d\mu &= \int_{X_0 \backslash A} g \, d\mu + \int_A g \, d\mu \\ &\leqslant \int_{X_0 \backslash A} \big(s_n + \varepsilon/\mu(X_0)\big) \, d\mu + M\mu(A) \leqslant \int_{X_0} s_n \, d\mu + \varepsilon + \varepsilon, \end{aligned}$$

which proves that $\lim_n \int_X s_n \, d\mu \geqslant \int_X g \, d\mu$.

Suppose that we can find a simple function $g = a\mathbb{1}_A$ such that $0 \leqslant g \leqslant f$ and $\mu(A) = \infty$. Then $\int f \, d\mu = \infty$. On the other hand, we may consider $B_N = \bigcup_{n=1}^N A \cap X_n$, where $X = \bigcup X_n$ and $\mu(X_n) < \infty$. Thanks to the result of the previous paragraph, we know that

$$\lim_n \int_X s_n \, d\mu \geqslant \lim_n \int_{B_N} s_n \, d\mu \geqslant \int_{B_N} g \, d\mu = a\mu(B_N).$$

But $\lim_{N \to \infty} \mu(B_N) = \mu(A) = \infty$, hence $\lim_n \int_X s_n = \infty$

This allows us to conclude that

$$\lim_n \int_X s_n \, d\mu \geqslant \sup \left\{ \int_X g \, d\mu : g \text{ is simple and } 0 \leqslant g \leqslant f \right\}$$

and $\lim_n \int_X s_n \, d\mu = \int f \, d\mu$. □

Finally, for a general measurable function we define the integral with the help of the decomposition we described in Lemma 4.14.

Definition 6.8. We say that a measurable function $f : X \to \mathbb{R}$ is integrable if $\int_X |f| \, d\mu < \infty$; in such case, the integral of f is defined by

$$\int_X f \, d\mu = \int_X f^+ \, d\mu - \int_X f^- \, d\mu,$$

where $f = f^+ - f^-$ and $f^+ = \max(f, 0)$, $f^- = -\min(f, 0)$.

Notice that the function f is integrable if and only if the functions f^+ and f^- are integrable. Of course, in case that $\int_X f^+\,d\mu = \infty$ and $\int_X f^-\,d\mu < \infty$ it is natural to assume that $\int_X f\,d\mu = \infty$ etc. We may also notice that for an integrable function f and $A \in \Sigma$ we have $\int_A f\,d\mu = \int_X f \cdot \mathbb{1}_A\,d\mu$.

Now we can easily extend the fundamental properties of the integral to measurable functions.

THEOREM 6.9. *For integrable functions f, g and a measurable h we have the following*

(1) $\int_X (f+g)\,d\mu = \int_X f\,d\mu + \int_X g\,d\mu$;
(2) *if* $f \leqslant g$ *then* $\int_X f\,d\mu \leqslant \int_X g\,d\mu$;
(3) *if* $a \leqslant f \leqslant b$ *then* $a\mu(X) \leqslant \int_X f\,d\mu \leqslant b\mu(X)$;
(4) *if* $h = 0$ *almost everywhere then* $\int_X h\,d\mu = 0$;
(5) *if* $\int_X h\,d\mu = 0$ *and* $h \geqslant 0$ *almost everywhere then* $h = 0$ *almost everywhere;*
(6) $|\int_X (f+g)\,d\mu| \leqslant \int_X |f|\,d\mu + \int_X |g|\,d\mu$;
(7) *for* $A, B \in \Sigma$, *if* $A \cap B = \varnothing$ *then* $\int_{A\cup B} f\,d\mu = \int_A f\,d\mu + \int_B f\,d\mu$.

REMARK 6.10. *Property (3) is still valid even if either or both integrals only have numerical sense (they may be equal to* $\pm\infty$*).*

6.3. Limit theorems. One of the main advantages of the Lebesgue integral over the Riemann integral is the availability of limit theorems, which allow us to calculate or estimate integrals of possibly complicated functions with mimimal effort.

THEOREM 6.11 (Monotone convergence theorem). *If f_n is a sequence of non-negative functions and $f_1 \leqslant f_2 \leqslant \ldots$ almost everywhere then the limit function $f = \lim_n f_n$ satisfies $\int_X f\,d\mu = \lim_n \int_X f_n\,d\mu$.*

The proof is simply an adaptation of the proof of Theorem 6.7. Notice that we do not assume functions f_n to be integrable. The limit function f is well defined almost everywhere if we allow it to attain infinite values.

THEOREM 6.12 (Fatou Lemma). *If f_n is a sequence of non-negative functions then*

$$\int_X \liminf_n f_n\,d\mu \leqslant \liminf_n \int_X f_n\,d\mu.$$

PROOF. By denoting

$$g_n = \inf_{k\geqslant n} f_k, \quad f = \liminf_n f_n,$$

we obtain $g_n \leqslant f_n$, $0 \leqslant g_1 \leqslant g_2 \leqslant \ldots$ and $\lim_n g_n = f$. Hence from the monotone convergence theorem we get

$$\int_X f_n\,d\mu \geqslant \int_X g_n\,d\mu \to \int_X f\,d\mu,$$

and then the result follows immediately. □

EXAMPLE 6.13. If $f_n = \mathbb{1}_{[n,n+1]}$, then $\liminf_n f_n = 0$, while gdy $\int_{\mathbb{R}} f_n \, d\lambda = 1$ for every n. This simple example shows that the Fatou lemma indeed requires an inequality. It is also an easy way to remember, in which direction is the inequality pointing.

THEOREM 6.14 (Lebesgue dominated convergence theorem). *Let f_n and g be such measurable functions that for every n the inequality $|f_n| \leqslant g$ is satisfied almost everywhere and $\int_X g \, d\mu < \infty$. If $f = \lim_n f_n$ almost everywhere then*

$$\lim_n \int_X |f_n - f| \, d\mu = 0 \quad \textit{and} \quad \int_X f \, d\mu = \lim_n \int_X f_n \, d\mu.$$

PROOF. Let $h_n = |f_n - f|$ and $h = 2g$. Then $h_n \to 0$ almost everywhere and $0 \leqslant h_n \leqslant h$. Thus by applying the Fatou lemma to the sequence $h - h_n$, we obtain

$$\begin{aligned} \int_X h \, d\mu = \int_X \liminf_n (h - h_n) \, d\mu &\leqslant \liminf_n \int_X (h - h_n) \, d\mu \\ &= \int_X h \, d\mu - \limsup_n \int_X h_n \, d\mu. \end{aligned}$$

This gives us $\limsup_n \int_X h_n \, d\mu = 0$, because $\int_X h \, d\mu < \infty$. Thus we have shown that $\int_X |f_n - f| \, d\mu \to 0$. Because

$$\int_X f_n \, d\mu - \int_X f \, d\mu \leqslant \int_X |f_n - f| \, d\mu,$$

the second relation follws from the first. □

REMARK 6.15. *Let $X = [0,1]$ and $f_n = n\mathbb{1}_{[0,1/n]}$. Then we have $f_n \to 0$ λ-almost everywhere, but $\int_{[0,1]} f_n \, d\lambda = 1$. The assumption of "dominated convergence", appearing in (the very name of) Theorem 6.14 is therefore important.*

COROLLARY 6.16. *Let $\mu(X) < \infty$ and let functions f_n be uniformly bounded. If $f = \lim_n f_n$ almost everywhere then $\int_X f \, d\mu = \lim_n \int_X f_n \, d\mu$.*

THEOREM 6.17. *If f is a measurable and non-negative function on a measure space (X, Σ, μ) then the set function $\nu : \Sigma \to [0, \infty]$ given for every $A \in \Sigma$ by $\nu(A) = \int_A f \, d\mu$ is a measure on Σ.*

PROOF. By the properties of the integral, we know that ν is an additive set function on Σ. If $A_n \uparrow A$ for some sets $A_n, A \in \Sigma$ then $\mathbb{1}_{A_n}$ is a non-decreasing sequence of functions converging to $\mathbb{1}_A$, while $f\mathbb{1}_{A_n} \to f\mathbb{1}_A$. By the monotone convergence theorem we thus have

$$\nu(A) = \int_A f \, d\mu = \int_X f\mathbb{1}_A \, d\mu = \lim_n \int_X f\mathbb{1}_{A_n} \, d\mu = \lim_n \nu(A_n)$$

Hence ν is continuous from below and thus countably additive (it is a measure). $\square$

7. L^p spaces

DEFINITION 7.1. We say that a sequence of measurable functions $f_n : X \to \mathbb{R}$ converges in measure to a function f if for every $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \mu\big(\{x : |f_n(x) - f(x)| \geqslant \varepsilon\}\big) = 0.$$

In such case we denote $f_n \xrightarrow{\mu} f$.

PROPOSITION 7.2. *A sequence which converges almost uniformly, converges in measure.*

PROOF. If functions f_n converge to f almost uniformly, then for every $\varepsilon > 0$ there exists a set A such that $\mu(A) < \varepsilon$ and $|f_n(x) - f(x)| < \varepsilon$ for large enough n and all $x \notin A$. Thus $\{x : |f_n(x) - f(x)| \geqslant \varepsilon\} \subseteq A$ and $\mu\big(\{x : |f_n(x) - f(x)| \geqslant \varepsilon\}\big) \leqslant \mu(A) < \varepsilon$. $\square$

REMARK 7.3. *Let $f_n : [0,1] \to \mathbb{R}$ denote the sequence*

$$\mathbb{1}_{[0,1]},\ \mathbb{1}_{[0,1/2]},\ \mathbb{1}_{[1/2,1]},\ \mathbb{1}_{[0,1/4]},\ \mathbb{1}_{[1/4,1/2]}, \ldots$$

We can check that f_n converges to 0 in Lebesgue measure, but

$$\liminf_n f_n(x) = 0, \quad \limsup_n f_n(x) = 1 \quad \textit{for every } x \in [0,1],$$

so the sequence doesn't converge almost uniformly.

LEMMA 7.4 (Chebyshev inequality). *If f is a measurable function then for every $\varepsilon > 0$*

$$\varepsilon \cdot \mu\big(\{x : |f(x)| \geqslant \varepsilon\}\big) \leqslant \int_X |f| \, d\mu.$$

PROOF. Let $A_\varepsilon = \{x : |f(x)| \geqslant \varepsilon\}$. Then $|f| \, \mathbb{1}_{A_\varepsilon} \geqslant \varepsilon \, \mathbb{1}_{A_\varepsilon}$ and $\int_X |f| \, d\mu \geqslant \int_{A_\varepsilon} |f| \, d\mu \geqslant \varepsilon\mu(A_\varepsilon)$ $\square$

THEOREM 7.5 (Riesz). *Let (X, Σ, μ) be a finite measure space and let $f_n : X \to \mathbb{R}$ be a sequence of measurable functions satisfying the Cauchy condition in measure, i.e.*

$$\lim_{n,m\to\infty} \mu\big(\{x : |f_n(x) - f_m(x)| \geqslant \varepsilon\}\big) = 0$$

for every $\varepsilon > 0$. Then

- *there exists a subsequence $n(k) \in \mathbb{N}$, such that the sequence of functions $f_{n(k)}$ is convergent almost everywhere;*
- *the sequence f_n converges in measure to some function f.*

Proof. Notice that the Cauchy condition we assumed implies that for every k there exists $n(k)$, such that for any $n, m \geqslant n(k)$ we have

$$\mu\big(\{x : |f_n(x) - f_m(x)| \geqslant 1/2^k\}\big) \leqslant 1/2^k,$$

and in addition we can take $n(1) < n(2) < \ldots$. Let

$$E_k = \big\{x : |f_{n(k)}(x) - f_{n(k+1)}(x)| \geqslant 1/2^k\big\}, \quad A_k = \bigcup_{n \geqslant k} E_n.$$

Then $\mu(A_k) \leqslant 1/2^{k-1}$ and hence the set $A = \bigcap_k A_k$ has measure zero. If $x \notin A$ then for every k such that $x \notin A_k$ and every $i \geqslant k$ we have

$$|f_{n(i)} - f_{n(i+1)}| \leqslant 1/2^i.$$

It follows from the triangle inequality that for $j > i \geqslant k$ we have

$$|f_{n(i)} - f_{n(j)}| \leqslant 1/2^{i-1}.$$

This means that for $x \notin A$ the numerical sequence $f_{n(i)}(x)$ satisfies the Cauchy condition and hence converges to a number, which we (unsurprisingly) denote as $f(x)$. In this way we obtain that $f_{n(k)}$ converges almost everywhere to the fuction f and this proves the first part of the theorem.

In order to verify the second part it suffices to notice that $f_n \xrightarrow{\mu} f$, which follows from

$$\begin{aligned}&\big\{x : |f_n(x) - f(x)| \geqslant \varepsilon\big\}\\ &\subseteq \big\{x : |f_n(x) - f_{n(k)}(x)| \geqslant \frac{\varepsilon}{2}\big\} \cup \big\{x : |f_{n(k)}(x) - f(x)| \geqslant \frac{\varepsilon}{2}\big\},\end{aligned}$$

and the Cauchy condition for the convergence in measure. □

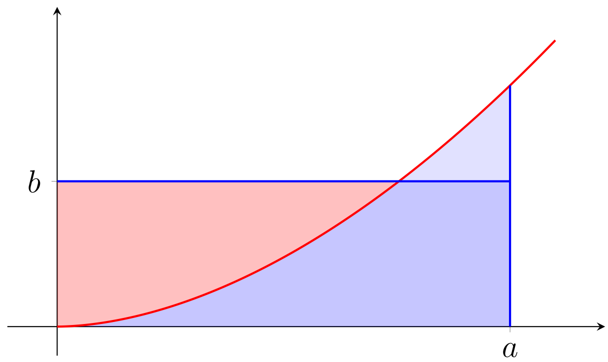

Figure 2. Young inequality: the rectangle $[0, a] \times [0, b]$ is covered by the blue and red areas, but there is an excess of blue, hence the inequality.

Lemma 7.6 (Young inequality for products). *For any positive numbers a, b, p, q, if $1/p + 1/q = 1$ then*

$$ab \leqslant \frac{a^p}{p} + \frac{b^q}{q}.$$

PROOF. Consider the function $f(t) = t^{p-1}$ on the interval $[0, a]$. We assume $p > 1$ therefore f has the inverse function $g(s) = s^{1/(p-1)}$. Note that the areas under the graphs of $f : [0, a] \to \mathbb{R}$ and $g : [0, b] \to \mathbb{R}$ cover the rectangle $[0, a] \times [0, b]$ (see Figure 2).

Thus

$$ab \leqslant \int_0^a t^{p-1}\,dt + \int_0^b s^{1/(p-1)}\,ds = \frac{t^p}{p}\bigg|_0^a + \frac{s^q}{q}\bigg|_0^b = \frac{a^p}{p} + \frac{b^q}{q},$$

because $1 + 1/(p-1) = p/(p-1) = q$. □

DEFINITION 7.7. For every measurable function (integrable or not) $f : X \to \mathbb{R}$ and $p \geqslant 1$ the expression

$$\|f\|_p = \left(\int_X |f|^p\,d\mu\right)^{1/p}$$

is called the p-th integral norm of the function f.

THEOREM 7.8 (Hölder inequality). *For every pair of functions f, g and numbers $p, q > 0$ such that $1/p + 1/q = 1$ we have the following inequality*

$$\|fg\|_1 = \int_X |f \cdot g|\,d\mu \leqslant \|f\|_p \cdot \|g\|_q.$$

PROOF. The inequality is obviously true if one of the norms on the right-hand side is infinite. Otherwise, for a given $x \in X$ we substitute

$$a = \frac{|f(x)|}{\|f\|_p}, \quad b = \frac{|g(x)|}{\|g\|_q}$$

into the inequality in the previous lemma in order to obtain (for every $x \in X$)

$$\frac{|f(x) \cdot g(x)|}{\|f\|_p \cdot \|g\|_q} \leqslant \frac{1}{p} \cdot \frac{|f(x)|^p}{\|f\|_p^p} + \frac{1}{q} \cdot \frac{|g(x)|^q}{\|g\|_q^q}.$$

By integrating the last inequality we get

$$\int_X |fg|\,d\mu\|f\|_p \cdot \|g\|q \leqslant 1p + 1q = 1.$$ □

THEOREM 7.9 (Minkowski inequality). *For every pair of functions f, g and a number $p \geqslant 1$, we have the following inequality*

$$\|f + g\|_p \leqslant \|f\|_p + \|g\|_p.$$

PROOF. The inequality is satisfied for $p = 1$. For $p > 1$ we may find a number q satisfying the condition $1/p + 1/q = 1$. Notice that

$(p-1)q = p$ and $p/q = p-1$. We use the Hölder inequality to get

$$\begin{aligned}
\|f+g\|_p^p &= \int_X |f+g|^p \, d\mu \\
&\leqslant \int_X |f| \cdot |f+g|^{p-1} \, d\mu + \int_X |g| \cdot |f+g|^{p-1} \, d\mu \\
&\leqslant \|f\|_p \Big(\int_X |f+g|^{(p-1)q} \, d\mu \Big)^{\frac{1}{q}} + \|g\|_p \Big(\int_X |f+g|^{(p-1)q} \, d\mu \Big)^{\frac{1}{q}} \\
&= \big(\|f\|_p + \|g\|_p\big) \cdot \Big(\int_X |f+g|^p \, d\mu \Big)^{\frac{1}{q}} = \big(\|f\|_p + \|g\|_p\big) \cdot \|f+g\|_p^{p/q}.
\end{aligned}$$

We now divide both sides by $\|f+g\|_p^{p/q}$ and we get result.

Note that in order for this proof to be entirely correct, we need to verify that $\|f\|_p, \|g\|_p < \infty$ implies $\|f+g\|_p < \infty$. $\square$

7.1. Banach spaces. Recall that a norm on a linear space X is a function $\|\cdot\| : X \to \mathbb{C}$ (or $X \to \mathbb{R}$) such that

(1) $\|x\| \geqslant 0$ for every $x \in \mathbb{R}^d$ and $\|x\| = 0$ if and only if $x = 0$;
(2) (triangle inequality) $\|x+y\| \leqslant \|x\| + \|y\|$ for every $x, y \in X$;
(3) (homogeneity) $\|ax\| = |a|\|x\|$ for every $x \in X$ and every $a \in \mathbb{C}$ (or $a \in \mathbb{R}$).

Definition 7.10. A normed space $(X, \|\cdot\|)$ is called a Banach space if the metric induced by the norm is complete, i.e. for every sequence $x_n \in X$ satisfying the Cauchy condition

$$\lim_{n,k\to\infty} \|x_n - x_k\| = 0,$$

there exists $x \in X$ such that $\|x_n - x\| \to 0$ (x is the limit of the sequece).

The p-th norm function $\|\cdot\|_p$ is in fact a norm: Minkowski inequality is the triangle inequality for $\|\cdot\|_p$ and homogeneity follows directly from the properties of the integral.

The only problem is with the first axiom, since $\|f\|_p = 0$ is only equivalent to saying that $f = 0$ almost everywhere.

Definition 7.11. For a given measure space (X, Σ, μ), by $L^p(\mu)$ we denote the space of all measurable functions $f : X \to \mathbb{R}$ for which $\|f\|_p < \infty$. Elements of $L^p(\mu)$ which are equal almost everywhere are identified as classes of abstraction.

In this way $L^p(\mu)$ equipped with the p-th integral norm becomes a normed space, but formally speaking it consists not of functions, but classes of abstaction (of functions). Most often. however, we may still refer to the elements of $L^p(\mu)$ as functions without any confusion.

It is nonetheless important not to forget about this distinction. For example, if f is a measurable function and $[f]$ is its class of abstraction such that $f \sim [f] \in L^p(\lambda)$, then for a chosen point $x \in \mathbb{R}$ the value

$f(x)$ is "undefined", since a single point has Lebesgue measure zero. In fact, $[f]$ contains functions attaining all possible values at x.

Notice that if $f_n \to f$ almost everywhere, then the same is true for every representative of the respective classes of abstraction, while it is not true for the actual pointwise convergence (everywhere without "almost").

In different contexts, $L^p(\mu)$ may also be denoted by $L^p(X, \Sigma, \mu)$ or as $L^p(X)$. For example, we usually write $L^p(\mathbb{R})$ or $L^p(\mathbb{T})$ to refer to spaces defined using the Lebesgue measure on $\mathbb{R}$ or $\mathbb{T}$.

THEOREM 7.12. *Spaces $L^p(\mu)$ with norms $\|\cdot\|_p$ are Banach spaces for $p \geqslant 1$.*

PROOF. Consider $p = 1$ and let $f_n \in L^1(\mu)$ be a Cauchy sequence in the norm $\|\cdot\|_1$, that is

$$\lim_{n,k\to\infty} \int_X |f_n - f_k| \, d\mu = 0.$$

Then for $\varepsilon > 0$ it follows from the Chebyshev inequality that

$$\int_X |f_n - f_k| \, d\mu \geqslant \varepsilon \cdot \mu\left(\{x : |f_n(x) - f_k(x)| \geqslant \varepsilon\}\right),$$

which means that f_n is a Cauchy sequence in measure.

It follows from the Riesz theorem that there exists an increasing sequence $n_k \in \mathbb{N}$ and a function f such that $f_{n_k} \to f$ almost everywhere. On the other hand, the Fatou lemma gives us

$$\int_X |f| \, d\mu \leqslant \liminf_k \int_X |f_{n_k}| \, d\mu < \infty,$$

because the Cauchy condition implies that the sequence of integrals $\int_X |f_n| \, d\mu$ is bounded.

Using the Fatou lemma once again we obtain

$$\begin{aligned}\int_X |f - f_{n_k}| \, d\mu &= \int_X \liminf_j |f_{n_j} - f_{n_k}| \, d\mu \\ &\leqslant \liminf_j \int_X |f_{n_j} - f_{n_k}| \, d\mu \leqslant \varepsilon,\end{aligned}$$

for k large enough. Finally, because

$$\int_X |f - f_n| \, d\mu \leqslant \int_X |f - f_{n_k}| \, d\mu + \int_X |f_{n_k} - f_n| \, d\mu,$$

we obtain that f is in fact the limit of the sequence f_n in the space $L^1(\mu)$. The proof for $p > 1$ is a rather mechanical modification of this argument. $\square$

REMARK 7.13. *Consider measure spaces (X, Σ, μ) and (Y, Θ, ν). We may then define*

$$\Sigma \otimes \Theta = \sigma\big(\{A \times B : \; A \in \Sigma, \; B \in \Theta\}\big),$$

which is a σ-field of subsets of $X \times Y$. Similarly, we may define

$$(\mu \otimes \nu)(A \times B) = \mu(A) \cdot \nu(B),$$

and show that $\mu \otimes \nu$ extends to a measure on $(X \times Y, \Sigma \otimes \Theta)$. It can also be shown that

$$\mathrm{Bor}(\mathbb{R} \times \mathbb{R}) = \mathrm{Bor}(\mathbb{R}) \otimes \mathrm{Bor}(\mathbb{R}).$$

This allows us to easily consider spaces of functions of complex values. For a measure space (X, Σ, μ) and a function $f : X \to \mathbb{C}$ we say that f is measurable if $f^{-1}[B] \in \Sigma$ for every Borel set $B \subseteq \mathbb{C}$. Here $\mathbb{C}$ may be identified with $\mathbb{R} \times \mathbb{R}$ and so $\mathrm{Bor}(\mathbb{C}) = \mathrm{Bor}(\mathbb{R}) \otimes \mathrm{Bor}(\mathbb{R})$.

We may express such a function as $f = \mathrm{Re}\, f + i\, \mathrm{Im}\, f$, where $\mathrm{Re}\, f$ and $\mathrm{Im}\, f$ are real-valued functions. Then f is measurable if and only if $\mathrm{Re}\, f$ and $\mathrm{Im}\, f$ are measurable.

Hence if f is measurable then its modulus $|f| = \sqrt{(\mathrm{Re}\, f)^2 + (\mathrm{Im}\, f)^2}$ is measurable as well. The function f is integrable when $\int_X |f|\, d\mu < \infty$, while

$$\int_X f\, d\mu = \int_X \mathrm{Re}\, f\, d\mu + i \int_X \mathrm{Im}\, f\, d\mu$$

defines the integral. The basic properties of the integral remain valid.