

Chapter 2

Fourier Series

2.1 Approximation theorems

2.1.1 The Weierstrass approximation theorem

Definition 2.1 Let $[a, b]$ be an interval. We denote by $\Pi_n(a, b)$ the vector space of polynomials of degree less or equal n on $[a, b]$, and by

$$\Pi(a, b) = \bigcup_{n=0}^{\infty} \Pi_n(a, b)$$

the vector space of all polynomials on $[a, b]$.

Comment: The space of polynomials is more than just a vector space; the product of every two polynomials is also a polynomial, which means that $\Pi(a, b)$ is closed under pointwise multiplication. A vector space that is closed under a multiplication operation is called an **algebra**.

The following celebrated theorem states that every continuous function can be approximated uniformly by polynomials:

Theorem 2.2 (Weierstrass) Let $[a, b]$ be an interval. The set of polynomials $\Pi(a, b)$ is a dense subset of $C([a, b])$ with respect to the maximum norm. That is,

for every $f \in C([a, b])$ there exists a sequence of polynomial $P_n \in \Pi(a, b)$, such that

$$\lim_{n \rightarrow \infty} \|P_n - f\|_\infty = 0.$$

Proof: Step 1: Reduction. It suffices to prove this theorem for $[a, b] = [0, 1]$. Indeed, given $f \in C([a, b])$ we may define $\tilde{f} \in C([0, 1])$ by

$$\tilde{f}(x) = f(a + x(b - a)).$$

Suppose that a sequence $\tilde{P}_n \in \Pi(0, 1)$ converges uniformly to \tilde{f} . Then, $P_n : [a, b] \rightarrow \mathbb{R}$ defined by

$$P_n(x) = \tilde{P}_n((x - a)/(b - a))$$

are polynomials converging uniformly to f , as

$$\sup_{x \in [a, b]} |P_n(x) - f(x)| = \sup_{x \in [0, 1]} |\tilde{P}_n(x) - \tilde{f}(x)| = \sup_{x \in [0, 1]} |\tilde{P}_n(x) - \tilde{f}(x)|.$$

Step 2: The approximating sequence. Given $f \in C([0, 1])$, we define a sequence of polynomials $B_n f \in \Pi_n(0, 1)$, known as the **Bernstein polynomials** of f ,

$$B_n f(x) = \sum_{k=1}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1 - x)^{n-k}.$$

We view $B_n : f \mapsto B_n f$ as a mapping $C([0, 1]) \rightarrow \Pi_n(0, 1)$.

Step 3: Properties of the mapping B_n . The mapping B_n satisfies the following properties:

(a) Linearity:

$$B_n(\alpha f + \beta g) = \alpha B_n f + \beta B_n g.$$

(b) Positivity: if $f \geq 0$ then $B_n f \geq 0$.

(c) For $f = 1$, using the binomial formula,

$$B_n 1(x) = \sum_{k=0}^n \binom{n}{k} x^k (1 - x)^{n-k} = (x + 1 - x)^n = 1,$$

i.e., $B_n 1 = 1$ for every $n \in \mathbb{N}$.

(d) For $f = \text{Id}$, using again the binomial formula,

$$\begin{aligned}
 B_n \text{Id}(x) &= \sum_{k=0}^n \binom{n}{k} \frac{k}{n} x^k (1-x)^{n-k} \\
 &= \sum_{k=1}^n \binom{n-1}{k-1} x^k (1-x)^{n-k} \\
 &= x \sum_{k=0}^{n-1} \binom{n-1}{k} x^k (1-x)^{n-1-k} \\
 &= x,
 \end{aligned}$$

so that $B_n \text{Id} = \text{Id}$ for every $n \in \mathbb{N}$.

(e) For $f = \text{Id}^2$,

$$B_n \text{Id}^2(x) = \sum_{k=0}^n \binom{n}{k} \frac{k^2}{n^2} x^k (1-x)^{n-k} = \frac{n-1}{n} x^2 + \frac{1}{n} x,$$

so that $\|B_n \text{Id}^2 - \text{Id}^2\|_\infty \rightarrow 0$.

To summarize,

$$\|B_n f - f\|_\infty \rightarrow 0 \quad \text{for } f = 1, \text{Id}, \text{Id}^2.$$

Step 4: The properties of B_n imply uniform convergence. The next theorem shows that the properties of B_n ensure that for every $f \in C([0, 1])$, the sequence of polynomials $B_n f$ converges uniformly to f . ■

Theorem 2.3 (Bohman-Korovkin) Let $B_n : C([0, 1]) \rightarrow C([0, 1])$ be a sequence of operators that are linear, positive, and satisfy

$$\lim_{n \rightarrow \infty} \|B_n f - f\|_\infty = 0 \quad \text{for } f = 1, \text{Id}, \text{Id}^2.$$

Then $\|B_n f - f\|_\infty \rightarrow 0$ for all $f \in C([0, 1])$.

Proof: By linearity and positivity, if $f \geq g$, then

$$B_n f - B_n g = B_n(f - g) \geq 0,$$

i.e., $B_n f \geq B_n g$. In particular, since $\pm f \leq |f|$ it follows that $\pm B_n f \leq B_n |f|$, hence

$$|B_n f| \leq B_n |f|. \quad (2.1)$$

Let $f \in C([0, 1])$ be given as well as $\varepsilon > 0$. Since f is continuous on a bounded domain, it is uniformly continuous: there exists a $\delta > 0$ such that for every x, y such that $|x - y| < \delta$, $|f(x) - f(y)| < \varepsilon$. On the other hand, if $|x - y| > \delta$ then $|f(x) - f(y)| \leq 2\|f\|_\infty < 2\|f\|_\infty(x - y)^2/\delta^2$. In either case, there exists a constant C_ε such that

$$|f(x) - f(y)| < C_\varepsilon(x - y)^2 + \varepsilon.$$

View now this inequality as an inequality between functions of y with x being a parameter, namely

$$|f(x) - f| \leq C_\varepsilon(x^2 - 2x\text{Id} + \text{Id}^2) + \varepsilon.$$

By (2.1) and by positivity,

$$\begin{aligned} |f(x) - B_n f| &\leq |f(x) - f(x) B_n 1| + |f(x) B_n 1 - B_n f| \\ &\leq |f(x)| |1 - B_n 1| + B_n |f(x) - f| \\ &\leq \|f\|_\infty |1 - B_n 1| + C_\varepsilon(x^2 B_n 1 - 2x B_n \text{Id} + B_n \text{Id}^2) + \varepsilon B_n 1 \\ &\leq \|f\|_\infty |1 - B_n 1| + C_\varepsilon(x^2 - 2x\text{Id} + \text{Id}^2) \\ &\quad + C_\varepsilon x^2 |1 - B_n 1| + 2x C_\varepsilon |\text{Id} - B_n \text{Id}| + C_\varepsilon |\text{Id}^2 - B_n \text{Id}^2| \\ &\quad + \varepsilon + \varepsilon |1 - B_n 1| \\ &\leq \|f\|_\infty \|1 - B_n 1\|_\infty + C_\varepsilon(x^2 - 2x\text{Id} + \text{Id}^2) \\ &\quad + C_\varepsilon x^2 \|1 - B_n 1\|_\infty + 2x C_\varepsilon \|\text{Id} - B_n \text{Id}\|_\infty + C_\varepsilon \|\text{Id}^2 - B_n \text{Id}^2\|_\infty \\ &\quad + \varepsilon + \varepsilon \|1 - B_n 1\|_\infty. \end{aligned}$$

In particular, this should hold at x , where $x^2 - 2x\text{Id} + \text{Id}^2 = 0$, hence

$$\begin{aligned} |f(x) - B_n f(x)| &\leq \|f\|_\infty \|1 - B_n 1\|_\infty + \varepsilon + \varepsilon \|1 - B_n 1\|_\infty \\ &\quad + C_\varepsilon \|1 - B_n 1\|_\infty + 2C_\varepsilon \|\text{Id} - B_n \text{Id}\|_\infty + C_\varepsilon \|\text{Id}^2 - B_n \text{Id}^2\|_\infty, \end{aligned}$$

where we also used the fact that $x \in [0, 1]$. Taking the maximum over x and letting $n \rightarrow \infty$,

$$\limsup_{n \rightarrow \infty} \|f(x) - B_n f(x)\|_\infty \leq \varepsilon.$$

Since this holds for every $\varepsilon > 0$, the limit is zero. ■

Comment: Given $f \in C([a, b])$, one could choose $n + 1$ equi-distributed points x_i on $[a, b]$ and construct an **interpolation polynomial** of degree up to n through the points $(x_i, f(x_i))$. It turns out that in many cases this will not yield a sequence of approximating polynomials. There is something very special about the Bernstein polynomials, which makes them always work.

2.1.2 The Stone-Weierstrass theorem

Definition 2.4 Let K be a set. A collection \mathcal{A} of functions on K is called **point-separating** (מפרידה בין נקודות) if for every $x, y \in K$ there exists a function $f \in \mathcal{A}$, such that

$$f(x) \neq f(y).$$

Theorem 2.5 (Stone-Weierstraß) Let \mathcal{A} be a point-separating algebra of continuous real-valued functions on a compact set $K \subset \mathbb{R}^n$, where \mathcal{A} includes the constant function $f = 1$. Then \mathcal{A} is dense in $C(K)$ with respect to the maximum norm. That is,

$$\forall f \in C(K) \quad \forall \varepsilon > 0 \quad \exists g \in \mathcal{A} : \|f - g\|_\infty < \varepsilon.$$

Comment: The Weierstrass approximation theorem is a special instance of the Stone-Weierstrass theorem, as the polynomials are an algebra of real-valued function that is point-separating and includes the constants.

Proof: Step 1: We may assume that \mathcal{A} is closed. We need to prove that $\bar{\mathcal{A}} = C(K)$. The closure of \mathcal{A} is a vector subspace of $C(K)$; moreover, if $f, g \in \bar{\mathcal{A}}$, then there exist $f_n, g_n \in \mathcal{A}$, such that $f_n \rightarrow f$ and $g_n \rightarrow g$. It is easy to see that $\mathcal{A} \ni f_n g_n \rightarrow fg$, implying that $\bar{\mathcal{A}}$ is also an algebra of continuous functions. Thus, it suffices to show that if \mathcal{A} is a *closed* point-separating algebra of continuous real-valued functions including the constants, then $\mathcal{A} = C(K)$. We will henceforth assume that \mathcal{A} is closed.

Step 2: If $f \in \mathcal{A}$ then $|f| \in \mathcal{A}$. Let $f \in \mathcal{A}$ be given. By the Weierstraß approximation theorem,

$$\forall \varepsilon > 0 \quad \exists P \in \Pi[-\|f\|_\infty, \|f\|_\infty] : \max_{|t| \leq \|f\|_\infty} |t - P(t)| < \varepsilon.$$

Since \mathcal{A} is an algebra, $P \circ f$ is also in \mathcal{A} , and

$$\max_{x \in K} \|f(x) - P(f(x))\| \leq \max_{|t| \leq \|f\|_\infty} \|t - P(t)\| < \varepsilon,$$

i.e., $|f| \in \mathcal{A}$, where we used the fact that the latter is closed.

Step 3: If $f, g \in \mathcal{A}$ then $\min(f, g), \max(f, g) \in \mathcal{A}$. This follows from the fact that

$$\min(f, g) = \frac{1}{2}(f + g) - \frac{1}{2}|f - g| \quad \text{and} \quad \max(f, g) = \frac{1}{2}(f + g) + \frac{1}{2}|f - g|.$$

Step 4: We prove that

$$\forall F \in C(K) \quad \forall x \in K \quad \forall \varepsilon > 0 \quad \exists f \in \mathcal{A} : f(x) = F(x) \quad \text{and} \quad f < F + \varepsilon.$$

Since \mathcal{A} separates between points and $1 \in \mathcal{A}$, there exists for every $x, y \in K$ and every $\alpha, \beta \in \mathbb{R}$ a function $g \in \mathcal{A}$ such that $g(x) = \alpha$ and $g(y) = \beta$ (very easy to show). Thus, given F , x , and ε , then

$$\forall y \in K \quad \exists g_y \in \mathcal{A} : g_y(x) = F(x) \quad \text{and} \quad g_y(y) = F(y).$$

Since both F and g_y are continuous, there exists an open neighborhood U_y of y such that

$$g_y|_{U_y} < F|_{U_y} + \varepsilon.$$

The collection $\{U_y\}$ is an open covering of K , and since K is compact, there exists a finite open sub-covering $\{U_{y_i}\}_{i=1}^n$. The function

$$f = \min(g_{y_1}, g_{y_2}, \dots, g_{y_n})$$

satisfies the required properties.

Step 5: Finalization. Let $F \in C(K)$ and $\varepsilon > 0$ be given. We have seen that

$$\forall x \in K \quad \exists f_x \in \mathcal{A} : f_x(x) = F(x) \quad \text{and} \quad f_x < F + \varepsilon.$$

Relying again on continuity, there exists an open neighborhood V_x of x , such that

$$f_x|_{V_x} > F|_{V_x} - \varepsilon.$$

By the compactness of K , K can be covered by a finite number of $\{V_{x_i}\}_{i=1}^n$. The function

$$f = \max(f_{x_1}, f_{x_2}, \dots, f_{x_n})$$

satisfies,

$$\|f - F\|_\infty < \varepsilon,$$

which completes the proof. ■

Corollary 2.6 Let

$$K = \{(\cos x, \sin x) : 0 \leq x < 2\pi\} \subset \mathbb{R}^2$$

be the unit circle. Then, the **trigonometric polynomials** (פולינומים טריגונומטריים),

$$\mathcal{T} = \text{Span}\{1, \sin x, \sin 2x, \dots, \cos x, \cos 2x, \dots\}$$

are dense in $C(K)$. That is, every continuous function f on the circle is a uniform limit of trigonometric polynomials.

Proof: The trigonometric polynomials form an algebra of continuous functions on K (products of trigonometric polynomials are trigonometric polynomials by elementary trigonometric identities), which includes the constants and separates point (if $0 \leq x < y < 2\pi$, then either $\sin x \neq \sin y$ or $\cos x \neq \cos y$). ■

Putting this slightly differently:

Corollary 2.7 Every continuous function $f \in C([0, 2\pi])$ satisfying $f(0) = f(2\pi)$ is a uniform limit of trigonometric polynomials.

2.2 Inner-product spaces

2.2.1 Basic definitions

Definition 2.8 Let V be a real vector space. A function $(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}$ is called an **inner-product** (מכפלה פנימית) if it satisfies the following conditions:

(a) **Symmetry:** For every $x, y \in V$,

$$(x, y) = (y, x).$$

(b) **Bilinearity:** For every $x, y, z \in V$ and $a \in \mathbb{R}$,

$$(ax, y) = a(x, y) \quad \text{and} \quad (x + y, z) = (x, z) + (y, z).$$

(c) **Positivity:** For every $x \in V$,

$$(x, x) \geq 0,$$

with equality if and only if $x = 0$.

Example: \mathbb{R}^n with the Euclidean inner-product,

$$(x, y) = \sum_{i=1}^n x_i y_i.$$

▲ ▲ ▲

Example: The vector space of n -by- n real matrices with

$$(A, B) = \text{Tr}(A^T B).$$

▲ ▲ ▲

Example: The vector space of infinite sequences

$$\ell^2 = \{(x_n)_{n=1}^{\infty} : \sum_{n=1}^{\infty} |x_n|^2 < \infty\},$$

with

$$(x, y) = \sum_{n=1}^{\infty} x_n y_n.$$

▲ ▲ ▲

Definition 2.9 Let $(V, (\cdot, \cdot))$ be an inner-product space. Then we denote

$$\|x\| = (x, x)^{1/2}. \quad (2.2)$$

As the notation suggests, this defines a norm (hence a metric) on V . To prove it, we need the following key inequality:

Proposition 2.10 (Cauchy-Schwarz inequality) Let $(V, (\cdot, \cdot))$ be an inner-product space. For every $x, y \in V$,

$$|(x, y)| \leq \|x\| \|y\|.$$

Proof: By the positivity of the inner-product, for $x, y \neq 0$,

$$\begin{aligned} 0 &\leq \left(\frac{x}{\|x\|} - \frac{(x, y)}{\|x\| \|y\|} \frac{y}{\|y\|}, \frac{x}{\|x\|} - \frac{(x, y)}{\|x\| \|y\|} \frac{y}{\|y\|} \right) \\ &= \frac{\|x\|^2}{\|x\|^2} - 2 \frac{|(x, y)|^2}{\|x\|^2 \|y\|^2} + \frac{|(x, y)|^2}{\|x\|^2 \|y\|^2} \frac{\|y\|^2}{\|y\|^2}, \end{aligned}$$

from which follows that

$$|(x, y)| \leq \|x\| \|y\|.$$

■

Proposition 2.11 Let $(V, (\cdot, \cdot))$ be an inner-product space. Then, $\|\cdot\|$ as defined in (2.2) is a norm on V .

Proof: Positivity and homogeneity are immediate. The triangle inequality follows from the Cauchy-Schwarz inequality, as

$$\begin{aligned} \|x + y\|^2 &= (x + y, x + y) \\ &= \|x\|^2 + 2(x, y) + \|y\|^2 \\ &\leq \|x\|^2 + 2|(x, y)| + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\| \|y\| + \|y\|^2 \\ &= (\|x\| + \|y\|)^2, \end{aligned}$$

hence

$$\|x + y\| \leq \|x\| + \|y\|.$$

■

Comment: As you will notice, in an inner-product space $\|x\|^2$ is a much more convenient object than $\|x\|$.

Proposition 2.12 Let $(V, (\cdot, \cdot))$ be an inner-product space. Then, the inner-product is continuous with respect to the metric induced by the norm (2.2).


Proof: Let $x_n \rightarrow x$ and $y_n \rightarrow y$. Then,

$$(x_n, y_n) = (x, y) + (x_n - x, y) + (x_n, y_n - y),$$

and by the Cauchy-Schwarz inequality,

$$|(x_n, y_n) - (x, y)| \leq \|x_n - x\| \|y\| + \|x_n\| \|y_n - y\|.$$

Since the norm in a normed space is continuous, $\|x_n\| \rightarrow \|x\|$, and by limit arithmetics the right-hand side tends to zero as $n \rightarrow \infty$. ■

 **Exercise 2.1** Prove that a norm $\|\cdot\|$ on a vector space V is induced by an inner-product if and only if it satisfies the **parallelogram law** (זהות המקבילית),

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2),$$

in which case the inner-product is given by the **polarization identity**

$$(x, y) = \frac{1}{2} (\|x + y\|^2 - \|x\|^2 - \|y\|^2).$$

2.2.2 Orthonormal systems

Definition 2.13 Let $(V, (\cdot, \cdot))$ be an inner-product space. Two vectors x, y are called **orthogonal** (ניצבים), denoted $x \perp y$, if

$$(x, y) = 0.$$

A set of vectors $\{x_i : i \in I\}$ is called an **orthogonal system** (מערכת אורתונגלית) if $x_i \perp x_j$ for all $i \neq j$. It is called an **orthonormal system** (מערכת אורתונורמלית) if in addition $\|x_i\| = 1$ for all $i \in I$. (Note that the index set I needs not be countable.)

Proposition 2.14 (Pythagoras law) Let $\{x_1, \dots, x_n\}$ be an orthogonal system, then

$$\|x_1 + \dots + x_n\|^2 = \|x_1\|^2 + \dots + \|x_n\|^2.$$

Proof: For a pair of orthogonal vectors x, y ,

$$\|x + y\|^2 = (x + y, x + y) = \|x\|^2 + 2(x, y) + \|y\|^2 = \|x\|^2 + \|y\|^2.$$

The general case follows by induction. ■

Proposition 2.15 Let $\{x_1, \dots, x_n\}$ be a finite orthonormal system. Then, for every set of real numbers (a_1, \dots, a_n) ,

$$\left\| \sum_{j=1}^n a_j x_j \right\| = \left(\sum_{j=1}^n |a_j|^2 \right)^{1/2}.$$

Proof: This is immediate as by linearity and orthogonality,

$$\left\| \sum_{j=1}^n a_j x_j \right\|^2 = \left(\sum_{i=1}^n a_i x_i, \sum_{j=1}^n a_j x_j \right) = \sum_{i,j=1}^n a_i a_j (x_i, x_j) = \sum_{j=1}^n |a_j|^2.$$

■

Proposition 2.16 Let $\{x_i : i \in I\}$ be an orthogonal system. Then, it is linearly independent. That is, for every finite subset of indexes $J \subset I$, the vectors $\{x_j : j \in J\}$ are linearly independent.

Proof: Suppose that for some finite $J \subset I$ and a set of real numbers $\{a_j : j \in J\}$,

$$\sum_{j \in J} a_j x_j = 0.$$

Taking an inner-product with x_k , $k \in J$, we obtain

$$\sum_{j \in J} a_j (x_j, x_k) = a_k \|x_k\|^2 = 0,$$

i.e., $a_k = 0$.

■

Theorem 2.17 Let V be an inner-product space. Let $W \subset V$ be a finite-dimensional vector subspace spanned by an orthonormal system $\{x_i : i = 1, 2, \dots, d\}$. Define the function $\pi_W : V \rightarrow V$,

$$\pi_W(y) = \sum_{i=1}^d (y, x_i) x_i.$$

Then,

- (a) π_W is a continuous linear map.
 (b) $\text{Image}(\pi_W) \subset W$.
 (c) For every $y \in V$, $y - \pi_W(y) \perp W$.
 (d) For every $y \in V$, $\pi_W(y)$ is the element in W that is closest to y , i.e., for every $z \in W$,

$$\|z - y\| \geq \|\pi_W(y) - y\|$$

with equality if and only if $z = \pi_W(y)$.

- (e) For every $z \in W$, $\pi_W(z) = z$. In particular, $\text{Image}(\pi_W) = W$ and $\pi_W \circ \pi_W = \pi_W$.
 (f) For every $y \in V$,

$$\|\pi_W(y)\| \leq \|y\|,$$

with equality if and only if $y \in W$.

Proof:

- (a) Linearity and continuity follow from the properties of the inner-product.
 (b) By definition $\pi_W(y)$ is a linear combination of elements in the vector subspace W .
 (c) For every $y \in V$ and $k = 1, \dots, d$,

$$(y - \pi_W(y), x_k) = (y, x_k) - \sum_{i=1}^d (y, x_i)(x_i, x_k) = 0.$$

That is, $y - \pi_W(y)$ is orthogonal to a basis in W , hence to every vector in W .

- (d) Let $y \in V$ and $z \in W$. Then, since $y - \pi_W(y) \perp z - \pi_W(y)$, it follows from Pythagoras' Law that

$$\begin{aligned} \|z - y\|^2 &= \|(z - \pi_W(y)) - (y - \pi_W(y))\|^2 \\ &= \|(z - \pi_W(y))\|^2 + \|(y - \pi_W(y))\|^2 \\ &\geq \|(y - \pi_W(y))\|^2, \end{aligned}$$

with equality if and only if $z = \pi_W(y)$.

- (e) Let $z \in W$. Then z is the element in W closest to z , hence $z = \pi_W(z)$. The surjectivity and idempotence of π_W follow at once.

(f) Using the previous items and Pythagoras law,

$$\|y\|^2 = \|(y - \pi_W(y)) + \pi_W(y)\|^2 = \|y - \pi_W(y)\|^2 + \|\pi_W(y)\|^2 \geq \|\pi_W(y)\|^2,$$

with equality if and only if $y = \pi_W(y)$, i.e., $y \in W$.

■

Comment: It follows from Item (d) that π_W does not depend on the orthonormal system spanning W , but only on the subspace W . The map π_W is called the **orthogonal projection** (הטלה ניצבת) from V onto W .

Proposition 2.18 (Bessel inequality) Let $\{x_n : n \in \mathbb{N}\}$ be a countable orthonormal system in an inner-product space V . Then, for every $y \in V$,

$$\sum_{n=1}^{\infty} |(y, x_n)|^2 \leq \|y\|^2.$$

Proof: For every N define

$$V_N = \text{Span}\{x_1, \dots, x_N\},$$

and let

$$\pi_N(y) = \sum_{n=1}^N (y, x_n) x_n$$

be the corresponding orthogonal projection. By Theorem 2.17(f),

$$\begin{aligned} \|y\|^2 &\geq \|\pi_N(y)\|^2 = \left\| \sum_{n=1}^N (y, x_n) x_n, \sum_{m=1}^N (y, x_m) x_m \right\|^2 \\ &= \sum_{n,m=1}^N (y, x_n)(y, x_m)(x_n, x_m) \\ &= \sum_{n=1}^N |(y, x_n)|^2. \end{aligned}$$

Letting $N \rightarrow \infty$ we obtain the desired result.

■

Corollary 2.19 Let $\{x_n : n \in \mathbb{N}\}$ be a countable orthonormal system in an inner-product space V . For every $y \in V$,

$$\lim_{n \rightarrow \infty} (y, x_n) = 0.$$

Proof: $|(y, x_n)|^2$ is an element in a converging series. ■

Definition 2.20 An orthonormal system $\{x_n : n \in \mathbb{N}\}$ is called **complete** (מערכת שלמה) if its span is dense in V , where

$$\text{Span}\{x_n : n \in \mathbb{N}\} = \left\{ \sum_{j \in J} a_j x_j : a_j \in \mathbb{R}, |J| < \infty \right\}.$$

Theorem 2.21 Let $\{x_n : n \in \mathbb{N}\}$ be an orthonormal system in an inner-product space V . The following statements are equivalent:

- (a) It is complete.
- (b) For every $y \in V$,

$$y = \sum_{n=1}^{\infty} (y, x_n) x_n,$$

which is called the **expansion** (פיתוח) of y according to the orthonormal system $\{x_n\}$. Note that this condition means that

$$\lim_{N \rightarrow \infty} \left\| y - \sum_{n=1}^N (y, x_n) x_n \right\| = 0.$$

- (c) For every $y \in V$,

$$\|y\|^2 = \sum_{n=1}^{\infty} |(y, x_n)|^2,$$

an identity known as the **Parseval identity**.

Proof: Suppose that Condition (a) holds and let $y \in V$. Denote by

$$V_N = \text{Span}\{x_1, \dots, x_N\}$$

and by

$$\pi_N(y) = \sum_{n=1}^N (y, x_n) x_n$$

the corresponding orthogonal projection. Since the orthonormal system is complete, there exists given $\varepsilon > 0$ an $N \in \mathbb{N}$ and an element $y_N \in V_N$ such that

$$\|y_N - y\| < \varepsilon.$$

Since $\pi_N(y)$ is the element of V_N that is closest to y , we have

$$\left\| \sum_{n=1}^N (y, x_n) x_n - y \right\| < \varepsilon.$$

Moreover, if $N' > N$, then $V_N \subset V_{N'}$, hence $\pi_{N'}(y)$ may be an even better approximation to y than $\pi_N(y)$, i.e., for every $N' > N$,

$$\left\| \sum_{n=1}^{N'} (y, x_n) x_n - y \right\| < \varepsilon.$$

That is,

$$\forall \varepsilon > 0 \quad \exists N : \forall N' > N \quad \left\| \sum_{n=1}^{N'} (y, x_n) x_n - y \right\| < \varepsilon,$$

proving that

$$\lim_{N \rightarrow \infty} \sum_{n=1}^N (y, x_n) x_n = y.$$

Suppose next that Condition (b) holds. Condition (a) follows from the fact that every $y \in V$ is a limit of linear combinations of x_n 's.

Suppose once again that Condition (b) holds. By the continuity of the norm,

$$\|y\|^2 = \left\| \sum_{n=1}^{\infty} (y, x_n) x_n \right\|^2 = \lim_{N \rightarrow \infty} \left\| \sum_{n=1}^N (y, x_n) x_n \right\|^2 = \lim_{N \rightarrow \infty} \sum_{n=1}^N |(y, x_n)|^2 = \sum_{n=1}^{\infty} |(y, x_n)|^2.$$

Finally, suppose that Condition (c) holds. For every N , by the orthogonality $y - \pi_N(y) \perp V_N$,

$$\begin{aligned} \|y\|^2 &= \|(y - \pi_N(y)) + \pi_N(y)\|^2 \\ &= \|y - \pi_N(y)\|^2 + \|\pi_N(y)\|^2 \\ &= \|y - \pi_N(y)\|^2 + \sum_{n=1}^N |(y, x_n)|^2. \end{aligned}$$

Letting $N \rightarrow \infty$, we obtain that

$$\lim_{N \rightarrow \infty} \|y - \pi_N(y)\|^2 = 0.$$

—23h(2019)—

■

Corollary 2.22 Let $\{x_n : n \in \mathbb{N}\}$ be a complete orthonormal system. If

$$y = \sum_{n=1}^{\infty} a_n x_n,$$

then $a_n = (y, x_n)$. That is, there exists a unique representation of y as a sum of x_n 's.

Proof: By the continuity of the inner-product, for every $k \in \mathbb{N}$,

$$(y, x_k) = \left(\sum_{n=1}^{\infty} a_n x_n, x_k \right) = \sum_{n=1}^{\infty} a_n (x_n, x_k) = a_k.$$

■

Comment: Strictly speaking, an orthonormal system is *not* a basis for V unless V has finite-dimension. Yet, it is very similar to a basis in the sense that every vector is a limit of a sequence of unique linear combinations.

Comment: Consider the mapping

$$y \mapsto \{(y, x_n) : n \in \mathbb{N}\},$$

taking an element in V into a sequence of real numbers; this is clearly a linear map. Since

$$\|y\|^2 = \sum_{n=1}^{\infty} |(y, x_n)|^2,$$


this map is a distance-preserving map V to $\ell^2(\mathbb{R})$. But is it also onto? If it were, the inverse map would be

$$\{a_n\} \mapsto \sum_{n=1}^{\infty} a_n x_n.$$

The question is whether the limit exists for every $(a_n) \in \ell^2(\mathbb{R})$. Note that the partial sums form a Cauchy sequence as


$$\left\| \sum_{n=1}^N a_n x_n - \sum_{n=1}^M a_n x_n \right\|^2 = \sum_{n=M+1}^N a_n^2.$$

It follows that an inverse map exists if and only if V is complete. That is, metrically-speaking, every complete inner-product space assuming a complete countable orthonormal system is equivalent to $\ell^2(\mathbb{R})$.

 **Exercise 2.2** Show that if an orthonormal system $\{x_n\}$ is complete, then the **generalized Parseval identity**,

$$(y, z) = \sum_{n=1}^{\infty} (y, x_n)(x_n, z)$$

holds.

 **Exercise 2.3** Let $\{x_n\}$ be an orthonormal system in an inner-product space V . Show that if the system is complete, then there is no element $y \in V$ which is orthogonal to all the x_n 's.

2.3 The space $H(a, b)$

Definition 2.23 Let $a < b \in \mathbb{R}$. We denote by $H(a, b)$ the real vector space of functions $[a, b] \rightarrow \mathbb{R}$ that are bounded and Riemann-integrable.

Comment: We are making a compromise here; Riemann integrability is a problematic notion which eventually led to the definition of a stronger concept—**Lebesgue integrability**. For example, Riemann integrability does not pass to limits. As an example, consider an enumeration $\{q_n\}$ of the rationals and the sequence of functions

$$f_n(x) = \begin{cases} 1 & x = q_k \text{ for some } k \leq n \\ 0 & \text{otherwise.} \end{cases}$$

Then each f_n is Riemann integrable, but its pointwise limit as $n \rightarrow \infty$ is the Dirichlet function, which is not integrable. Lebesgue integrability, however is only taught in 3rd year in the framework of **measure theory** (תורת המידה), so we will make do with Riemann integrability.

Definition 2.24 For $f, g \in H(a, b)$ we define

$$(f, g) = \int_a^b f(x)g(x) dx. \quad (2.3)$$

The immediate question is whether (2.3) is an inner-product on $H(a, b)$. Symmetry and bilinearity follows from the properties of the integral. For positivity,

$$(f, f) = \int_a^b f^2(x) dx$$

is non-negative, however may be zero even if f is not identically zero. To overcome this difficulty, we proceed as follows:

Definition 2.25 A set $A \subset \mathbb{R}$ is said to have **measure zero** (מידה אפס) if for every $\varepsilon > 0$ there exists a countable number of open segments (a_n, b_n) , such that

$$A \subset \bigcup_{n=1}^{\infty} (a_n, b_n) \quad \text{and} \quad \sum_{n=1}^{\infty} (b_n - a_n) < \varepsilon.$$

Example: Every singleton $\{x\}$ has measure zero in \mathbb{R} . This can be proved by taking for example, given $\varepsilon > 0$,

$$a_n = x - \frac{\varepsilon}{2^n} \quad \text{and} \quad b_n = x + \frac{\varepsilon}{2^n}.$$

▲ ▲ ▲

Lemma 2.26 A countable union of sets of measure zero is a set of measure zero.

Proof: Suppose that a sequence of sets A_n all have measure zero. Given $\varepsilon > 0$, there exists for every n a sequence of open segments $(a_n^k, b_n^k)_{k=1}^{\infty}$, such that

$$A_n \subset \bigcup_{k=1}^{\infty} (a_n^k, b_n^k) \quad \text{and} \quad \sum_{k=1}^{\infty} (b_n^k - a_n^k) < \frac{\varepsilon}{2^n}.$$

Thus,

$$\bigcup_{n=1}^{\infty} A_n \subset \bigcup_{n=1}^{\infty} \bigcup_{k=1}^{\infty} (a_n^k, b_n^k) \quad \text{and} \quad \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} (b_n^k - a_n^k) < \varepsilon.$$

■

Example: The rationals \mathbb{Q} have measure zero in \mathbb{R} . ▲ ▲ ▲

Definition 2.27 Two functions $f, g : [a, b] \rightarrow \mathbb{R}$ are said to be **equal almost everywhere** (שווה כמעט בכל מקום) if the set

$$\{x \in [a, b] : f(x) \neq g(x)\}$$

has measure zero.

Lemma 2.28 The property of two functions being equal almost everywhere is an equivalence relation.

Proof: This is immediate. ■

Proposition 2.29 Let $f \in H(a, b)$. Then, $(f, f) = 0$ if and only if $f = 0$ almost everywhere.

Proof: Suppose that

$$\int_a^b f^2(x) dx = 0,$$

and that it were not true that f is non-zero at most on a set of measure zero. Denote

$$A = \{x : f(x) \neq 0\},$$

and note that

$$A = \bigcup_{n=1}^{\infty} A_n \quad \text{where} \quad A_n = \{x : f^2(x) > 1/n\}.$$

By Lemma 2.26, there exists an n for which A_n does not have measure zero. By definition, this means that there exists an $\varepsilon > 0$, such that for any partition $[c_k, d_k]_{k=1}^N$ of the segment $[a, b]$,

$$\sum_{k \in J} (b_k - a_k) \geq \varepsilon \quad \text{where} \quad J = \{k : A_n \cap [a_k, b_k] \neq \emptyset\}.$$

For all $k \in J$, take $\xi_k \in [a_k, b_k) \cap A_n$ and for all $k \notin J$ take any $\xi_k \in [a_k, b_k)$. Then,

$$\sum_{k=1}^N f^2(\xi_k)(b_k - a_k) \geq \sum_{k \in J} f^2(\xi_k)(b_k - a_k) \leq \frac{1}{n} \sum_{k \in J} (b_k - a_k) \geq \frac{\varepsilon}{n}.$$

The left-hand side is a Riemann sum; this inequality holds for any partition of $[a, b]$, it follows from the definition of the Riemann integral that

$$\int_a^b f^2(x) dx \geq \frac{\varepsilon}{n},$$

which is a contradiction.

Conversely, if $f = 0$ almost everywhere, it cannot be non-zero on any segment. This means that every partition of $[a, b]$ has a vanishing Riemann sum for f^2 , and since f^2 is Riemann-integrable, its integral is zero. ■

Define now $\tilde{H}(a, b)$ to be the set of equivalence classes of functions in $H(a, b)$ differing on sets of measure zero; we denote the equivalence class of $f \in H(a, b)$ by $[f]$. We endow $\tilde{H}(a, b)$ with a vector space structure by defining

$$a[f] + b[g] = [af + bg].$$

For $[f], [g] \in \tilde{H}(a, b)$, we set

$$([f], [g]) = (f, g). \quad (2.4)$$

One has to check that this is well-defined, i.e., that if $[f] = [f_1]$ and $[g] = [g_1]$ then

$$(f, g) = (f_1, g_1),$$

which is indeed the case, as

$$|(f_1, g_1) - (f, g)| = |(f_1 - f, g) + f_1(g_1 - g)| \leq \|f_1 - f\| \|g\| + \|f_1\| \|g_1 - g\| = 0.$$

Proposition 2.30 The product (2.4) is an inner-product on $\tilde{H}(a, b)$.

Proof: Symmetry follows from

$$([f], [g]) = (f, g) = (g, f) = ([g], [f]).$$

Linearity from

$$\begin{aligned} (a[f] + b[g], [h]) &= (af + bg, [h]) \\ &= (af + bg, h) \\ &= a(f, h) + b(g, h) \\ &= a([f], [h]) + b([g], [h]). \end{aligned}$$

Finally, by Proposition 2.29,

$$([f], [f]) = 0,$$

implies that $f = 0$ almost-everywhere, i.e., $[f] = [0]$. ■

Comment: Note that it may well be that $f \in H(a, b)$, $g = f$ almost everywhere however $g \notin H(a, b)$. For example the Dirichlet function on $[0, 1]$ equals zero almost everywhere, but it is not Riemann integrable. This is the price for working with Riemann integrability.

Henceforth, we will omit the tilde in $\tilde{H}(a, b)$ and refer to functions rather than equivalence classes of functions, but remember that whenever we refer to a function, we really refer to its equivalence class.

Definition 2.31 Let $f_n, f \in H(a, b)$. We say that f_n converges to f **in the mean** (בממוצע) if

$$\lim_{n \rightarrow \infty} \|f_n - f\| = 0,$$

i.e., if it converges to f in the norm induced by the inner-product.

Thus, we now have three different notions of convergence in $H(a, b)$:

- (a) Pointwise convergence.
- (b) Uniform convergence.
- (c) Convergence in the mean.

We already know that uniform convergence implies pointwise convergence, but that the converse is not true.

Proposition 2.32 Uniform convergence implies convergence in the mean.

Proof: Suppose that $f_n \rightarrow f$ uniformly. Then,

$$\|f_n - f\|^2 = \int_a^b (f_n(x) - f(x))^2 dx \leq (b - a) \sup_{x \in [a, b]} (f_n(x) - f(x))^2 \rightarrow 0.$$

■

Proposition 2.33 Pointwise convergence does not imply convergence in the mean.

Proof: Take for example the sequence of functions $f_n \in H(0, 1)$,

$$f_n(x) = \begin{cases} \sqrt{n} \sin nx & 0 \leq x \leq \pi/n \\ 0 & \text{otherwise.} \end{cases}$$

Then, f_n convergence to zero pointwise, however

$$\|f_n - 0\|^2 = \int_0^{\pi/n} n \sin^2 nx dx = \int_0^{\pi} \sin^2 t dt = \frac{\pi}{2},$$

i.e., f_n does not convergence to zero in the mean. ■

Proposition 2.34 Convergence in the mean does not imply pointwise convergence.

Proof: Take for example the sequence of functions $f_n \in H(0, 1)$,

$$\begin{aligned} f_1(x) &= \chi_{[0,1]} \\ f_2(x) &= \chi_{[0,1/2]} & f_3(x) &= \chi_{[1/2,1]} \\ f_4(x) &= \chi_{[0,1/4]} & f_5(x) &= \chi_{[1/4,1/2]} & f_6(x) &= \chi_{[1/2,3/4]} & f_7(x) &= \chi_{[3/4,1]} \\ & & & \text{etc} \end{aligned}$$

Then, f_n converges to zero in the mean however not pointwise, since for every x , $f_n(x) \neq 0$ infinitely-often. ■

Definition 2.35 We denote by $C_{per}(a, b)$ the set of continuous functions on $[a, b]$ satisfying $f(a) = f(b)$. (Note that this is a vector subspace of $H(a, b)$.)

Theorem 2.36 $C_{per}(a, b)$ is dense in $H(a, b)$. That is, for every $f \in H(a, b)$ there exists a sequence $f_n \in C_{per}(a, b)$ such that $f_n \rightarrow f$ in the mean.

Proof: It suffices to prove the theorem for the subset of $H(a, b)$ consisting of functions f for which $\|f\|_\infty \leq 1/2$. Indeed, let $f \in H(a, b)$, then there exists a sequence of functions $f_n \in C(a, b)$ such that

$$\lim_{n \rightarrow \infty} \left\| f_n - \frac{f}{2\|f\|_\infty} \right\| = 0,$$

hence $2\|f\|_\infty f_n \in C(a, b)$ converges to f in the mean.

So let $f \in H(a, b)$ satisfy $\|f\|_\infty \leq 1/2$ and let $\varepsilon > 0$. We need to prove that there exists a $g \in C_{\text{per}}(a, b)$ satisfying $\|f - g\| < \varepsilon$. Let $P = \{t_0, t_1, \dots, t_n\}$ be a partition of (a, b) and denote by

$$S(f, P) = \sum_{i=1}^n (t_i - t_{i-1}) \inf_{t_{i-1} \leq x < t_i} f(x)$$

the lower Darboux sum of f with respect to the partition P . By definition of the Riemann integral, there exists a partition P such that

$$0 \leq \int_a^b f(x) dx - S(f, P) < \frac{\varepsilon^2}{4}.$$

Define next the discontinuous function

$$h(x) = \inf_{t_{i-1} \leq x < t_i} f(x) \quad \text{for } x \in [t_{i-1}, t_i),$$

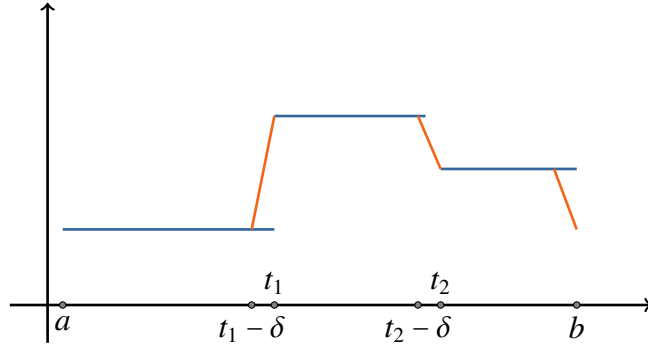
and $h(b) = f(b)$. By the definition of h , $\|h\|_\infty \leq 1/2$ hence $|f(x) - h(x)| \leq 1$, hence

$$\|f - h\|^2 = \int_a^b (f(x) - h(x))^2 dx \leq \int_a^b (f(x) - h(x)) dx = \int_a^b f(x) dx - S(f, P) < \frac{\varepsilon^2}{4}.$$

We have thus found a “step function” h satisfying $\|f - h\| < \varepsilon/2$. It remains to find a function $g \in C_{\text{per}}(a, b)$ satisfying $\|g - h\| < \varepsilon/2$. Take

$$\delta < \frac{\varepsilon^2}{4n} \quad \text{and} \quad \delta < (t_i - t_{i-1}), \quad i = 1, \dots, n.$$

Consider the following illustration where the blue segments are the graph of h and the orange segments are the graph of g where it differs from h :



Then, $g \in C_{\text{per}}(a, b)$, and

$$\|g - h\|^2 \leq n\delta < \frac{\varepsilon^2}{4},$$

which completes the proof. ■

Comment: As a side result, we have also proved that the step functions are dense in $H(a, b)$.

2.4 Fourier series

In this section we will consider the inner-product space $H(0, 2\pi)$.

Definition 2.37 The trigonometric system in $H(0, 2\pi)$ is the family of functions

$$\mathcal{T} = \{\psi_n\}_{n=0}^{\infty} \cup \{\varphi_n\}_{n=1}^{\infty},$$

where

$$\begin{aligned} \psi_0(x) &= \frac{1}{\sqrt{2\pi}} \\ \psi_n(x) &= \frac{1}{\sqrt{\pi}} \cos nx, \quad n = 1, 2, \dots \\ \varphi_n(x) &= \frac{1}{\sqrt{\pi}} \sin nx, \quad n = 1, 2, \dots \end{aligned}$$

(Note that these functions all belong to $C_{\text{per}}(0, 2\pi)$.)

Proposition 2.38 The trigonometric system forms an orthonormal system in $H(0, 2\pi)$.

Proof: This is verified by direct integration. For example,

$$(\varphi_m, \varphi_n) = \frac{1}{\pi} \int_0^{2\pi} \sin mx \sin nx \, dx = \delta_{m,n}.$$

■

Proposition 2.39 The trigonometric system forms a complete orthonormal system in $H(0, 2\pi)$.

Proof: This is an immediate consequence of the Stone-Weierstrass theorem, whereby the span of the trigonometric system is dense in $C_{\text{per}}(0, 2\pi)$ with respect to the maximum norm, and $C_{\text{per}}(0, 2\pi)$ is dense in $H(0, 2\pi)$ with respect to the inner-product norm. Thus,

$$\overline{\text{Span } \mathcal{T}} = C_{\text{per}}(0, 2\pi),$$

and

$$\overline{C_{\text{per}}(0, 2\pi)} = H(0, 2\pi).$$

This may seem confusing as it seems to insinuate that $C_{\text{per}}(0, 2\pi)$ is closed, hence equals $H(0, 2\pi)$ (which is of course not true). One has to remember, however that both equalities refer to different metrics. The first is with respect to the maximum norm, whereas the second is with respect to the inner-product norm. Since for $f \in H(0, 2\pi)$,

$$\|f\|^2 = \int_0^{2\pi} f^2(x) \, dx \leq 2\pi \|f\|_{\infty},$$

(or as we've already seen, uniform convergence implies convergence in the mean), it follows that

$$\overline{\text{Span } \mathcal{T}} \supseteq C_{\text{per}}(0, 2\pi),$$

where now the closure is with respect to the inner-product norm. Thus,

$$\overline{\text{Span } \mathcal{T}} \supseteq \overline{C_{\text{per}}(0, 2\pi)} = H(0, 2\pi),$$

which concludes the proof. ■

Corollary 2.40 For every $f \in H(0, 2\pi)$,

$$f = \sum_{n=0}^{\infty} (f, \psi_n) \psi_n + \sum_{n=1}^{\infty} (f, \varphi_n) \varphi_n. \quad (2.5)$$

The right-hand side in (2.5) is called the **Fourier series** (טור פורייה) of f . The scalars

$$\begin{aligned} a_0^f &= \frac{1}{\sqrt{2\pi}} (f, \psi_0) = \frac{1}{2\pi} \int_0^{2\pi} f(x) dx \\ a_n^f &= \frac{1}{\sqrt{\pi}} (f, \psi_n) = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos nx dx \\ b_n^f &= \frac{1}{\sqrt{\pi}} (f, \varphi_n) = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin nx dx \end{aligned}$$

are called the **Fourier coefficients** (מקדמי פורייה) of f . Thus, for every $f \in H(0, 2\pi)$,

$$f(x) = \sum_{n=0}^{\infty} a_n^f \cos nx + \sum_{n=1}^{\infty} b_n^f \sin nx.$$

We will also denote by

$$S_n f(x) = \sum_{k=0}^n a_k^f \cos kx + \sum_{k=1}^n b_k^f \sin kx$$

the partial trigonometric sum. Hence

$$f = \lim_{n \rightarrow \infty} S_n f,$$

but note that the convergence is in the mean, and not necessarily uniform, and neither pointwise. That is, (2.5) cannot be interpreted as a pointwise equation, but only as an equality in $H(0, 2\pi)$, where functions may differ on a set of measure zero.

Corollary 2.41 For every $f, g \in H(0, 2\pi)$,

$$(f, g) = 2\pi a_0^f a_0^g + \pi \sum_{n=1}^{\infty} a_n^f a_n^g + \pi \sum_{n=1}^{\infty} b_n^f b_n^g,$$

and in particular

$$\|f\|^2 = 2\pi|a_0^f|^2 + \pi \sum_{n=1}^{\infty} |a_n^f|^2 + \pi \sum_{n=1}^{\infty} |b_n^f|^2.$$

Proof: By the generalized Parseval's identity,

$$(f, g) = \sum_{n=0}^{\infty} (f, \psi_n)(g, \psi_n) + \sum_{n=1}^{\infty} (f, \varphi_n)(g, \varphi_n),$$

and it remains to substitute the definitions of a_n^f and b_n^f . ■

Example: Consider the function $f(x) = x$. Then,

$$\begin{aligned} 2\pi a_0^f &= \int_0^{2\pi} x \, dx = 2\pi^2 \\ \pi a_n^f &= \int_0^{2\pi} x \cos nx \, dx = \frac{1}{n} x \sin nx \Big|_0^{2\pi} - \frac{1}{n} \int_0^{2\pi} \sin nx \, dx = 0 \\ \pi b_n^f &= \int_0^{2\pi} x \sin nx \, dx = -\frac{1}{n} x \cos nx \Big|_0^{2\pi} + \frac{1}{n} \int_0^{2\pi} \cos nx \, dx = -\frac{2\pi}{n}. \end{aligned}$$

Thus,

$$x = \pi - \sum_{n=1}^{\infty} \frac{2}{n} \sin nx.$$

Note that for any truncation the right-hand side is continuous periodic whereas the limit is not periodic; clearly, the convergence cannot be uniform. Moreover, for $x = 0$ it looks utterly wrong, as the left-hand side equals zero and the right-hand side equals π . This equality holds however as a limit in $H(0, 2\pi)$, hence almost everywhere.

Parseval's identity yields

$$\int_0^{2\pi} x^2 \, dx = \frac{8\pi^3}{3} = 2\pi^3 + \pi \sum_{n=1}^{\infty} \frac{4}{n^2},$$

which reduces to

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

This is a very well-known formula, which is very difficult to obtain by other methods. The calculation of the sum of inverse square was known as the **Basel problem**. It was first posed by Pietro Mengoli in 1650 and solved by Leonhard Euler in 1734 (not using Fourier series, which had not yet been invented). ▲ ▲ ▲

We proceed to investigate under what conditions the Fourier series converges pointwise and/or uniformly:

Proposition 2.42 Let $f \in H(a, b)$ and let $\theta \in \mathbb{R}$. Then,

$$\lim_{n \rightarrow \infty} \int_a^b f(x) \sin((n + \theta)x) dx = 0,$$

and

$$\lim_{n \rightarrow \infty} \int_a^b f(x) \cos((n + \theta)x) dx = 0,$$

Proof: Suppose first that $[a, b] = [0, 2\pi]$. Then,

$$\int_a^b f(x) \sin((n + \theta)x) dx = \int_0^{2\pi} f(x) (\sin nx \cos \theta x + \cos nx \sin \theta x) dx.$$

Defining $f_1(x) = f(x) \cos \theta x$ and $f_2(x) = f(x) \sin \theta x$, we have that

$$\int_a^b f(x) \sin((n + \theta)x) dx = 2\pi(b_n^{f_1} + a_n^{f_2}),$$

which converges to zero as $n \rightarrow \infty$. We proceed similarly for the cos term.

Next, suppose that $[a, b] \subset [0, 2\pi]$, define

$$g(x) = \begin{cases} f(x) & x \in [a, b] \\ 0 & \text{otherwise.} \end{cases}$$

Then, $g \in H(0, 2\pi)$,

$$\int_a^b f(x) \sin((n + \theta)x) dx = \int_0^{2\pi} g(x) \sin((n + \theta)x) dx,$$

which tends to zero as $n \rightarrow \infty$.

Consider then the case where $[a, b] = [2\pi k, 2\pi\ell]$. Then,

$$\begin{aligned} \int_{2\pi k}^{2\pi\ell} f(x) \sin((n + \theta)x) dx &= \sum_{j=k}^{\ell} \int_{2\pi j}^{2\pi(j+1)} f(x) \sin((n + \theta)x) dx \\ &= \sum_{j=k}^{\ell} \int_0^{2\pi} f(x - 2\pi j) \sin((n + \theta)(x - 2\pi j)) dx, \end{aligned}$$

and the limit as $n \rightarrow \infty$ vanishes by the $[0, 2\pi]$ case. Finally, any closed interval can be encapsulated in an interval $[2\pi k, 2\pi\ell]$ as above. ■

Proposition 2.43 Let $f \in H(0, 2\pi)$.

- (a) A necessary condition for its Fourier series to converge pointwise is that $f(0) = f(2\pi)$.
- (b) A necessary condition for its Fourier series to converge uniformly is that f be continuous (and also that $f(0) = f(2\pi)$).

Proof:

- (a) The functions $S_n f$ satisfy $S_n f(0) = S_n f(2\pi)$. Hence, if $S_n f$ converges to f pointwise, it must hold that $f(0) = f(2\pi)$.
- (b) The functions $S_n f$ are continuous, hence if they converge uniformly, their limit must be continuous.

■

—27h(2019)—

Proposition 2.44 Let $f \in H(0, 2\pi)$ be continuous, piecewise continuously-differentiable and satisfies $f(0) = f(2\pi)$. Then, $f' \in H(0, 2\pi)$ and its Fourier series is obtained by term-by-term differentiation of the Fourier series of f , namely,

$$f' = - \sum_{k=0}^{\infty} k a_k^f \sin kx + \sum_{k=1}^{\infty} k b_k^f \cos kx,$$

that is

$$a_n^{f'} = n b_n^f \quad \text{and} \quad b_n^{f'} = -n a_n^f.$$

Proof: Clearly, $f' \in H(0, 2\pi)$ as a piecewise-continuous function. Integrating by parts,

$$\begin{aligned} a_n^f &= \frac{1}{\pi} \int_0^{2\pi} f(x) \cos nx \, dx \\ &= \frac{1}{\pi n} f(x) \sin nx \Big|_0^{2\pi} - \frac{1}{\pi n} \int_0^{2\pi} f'(x) \sin nx \, dx \\ &= -\frac{1}{n} b_n^{f'}. \end{aligned}$$

We proceed similarly with b_n^f . ■

Proposition 2.45 Let $f \in H(0, 2\pi)$ be continuous, piecewise continuously-differentiable and satisfies $f(0) = f(2\pi)$. Then,

$$\sum_{n=1}^{\infty} |a_n^f| < \infty \quad \text{and} \quad \sum_{n=1}^{\infty} |b_n^f| < \infty.$$

Moreover, the Fourier series of f converges absolutely and uniformly. Moreover,

Proof: From the previous proposition, the Cauchy-Schwarz inequality and the Parseval identity for f' ,

$$\begin{aligned} \sum_{n=1}^{\infty} |a_n^f| &= \sum_{n=1}^{\infty} \frac{|b_n^{f'}|}{n} \\ &\leq \left(\sum_{n=1}^{\infty} \frac{1}{n^2} \right)^{1/2} \left(\sum_{n=1}^{\infty} |b_n^{f'}|^2 \right)^{1/2} \\ &\leq \frac{\pi}{\sqrt{6}} \frac{1}{\sqrt{\pi}} \|f'\| < \infty, \end{aligned}$$

and similarly for b_n^f . Then, the series

$$S_n f(x) = \sum_{k=0}^n a_k^f \cos kx + \sum_{k=1}^n b_k^f \sin kx$$

satisfies the Weierstrass M-test, hence converging absolutely and uniformly to some limit g . Since uniform convergence implies convergence in the mean, it follows that the Fourier series of f converges in the mean both to f and to g , hence f and g are equal almost everywhere. However g is continuous as a uniform limit of continuous functions and f is continuous as well, hence $f = g$. ■

Lemma 2.46 Let $f \in H(0, 2\pi)$. Then, its partial Fourier sums can be represented as

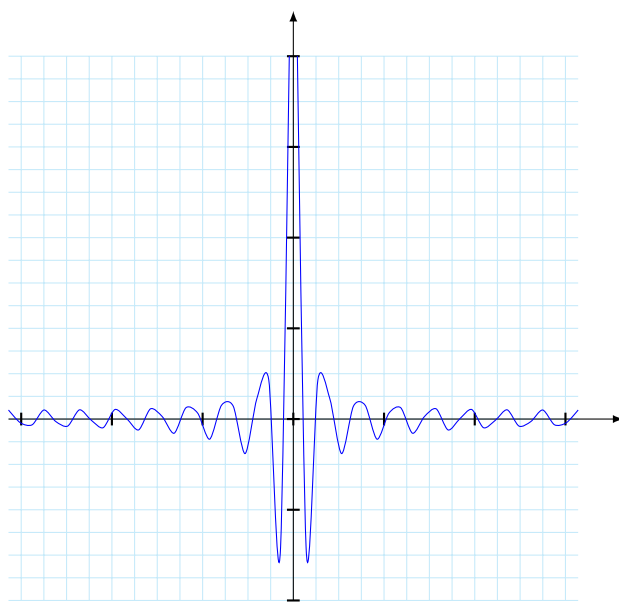
$$S_n f = \frac{1}{2\pi} \int_0^{2\pi} f(t) D_n(x-t) dt,$$

where

$$D_n(x) = \frac{\sin(n + \frac{1}{2})x}{\sin \frac{1}{2}x}$$

is called the **Dirichlet kernel** (גרעין דיריכלה).

The following figure shows $D_{32}(x)$:



Proof: This follows from straightforward algebra:

$$\begin{aligned}
 S_n f(x) &= a_0^f + \sum_{k=1}^n a_k^f \cos kx + \sum_{k=1}^n b_k^f \sin kx \\
 &= \frac{1}{2\pi} \int_0^{2\pi} f(t) dt + \frac{1}{\pi} \int_0^{2\pi} \left(\sum_{k=1}^n f(t) \cos kt \cos kx \right) dt \\
 &\quad + \frac{1}{\pi} \int_0^{2\pi} \left(\sum_{k=1}^n f(t) \sin kt \sin kx \right) dt \\
 &= \frac{1}{2\pi} \int_0^{2\pi} f(t) \left(1 + 2 \sum_{k=1}^n \cos k(x-t) \right) dt.
 \end{aligned}$$

It remains to prove that

$$1 + 2 \sum_{k=1}^n \cos kx = D_n(x),$$

which can be done using induction. Alternatively, using the fact that

$$\cos x = \frac{e^{ix} + e^{-ix}}{2} \quad \text{and} \quad \sin x = \frac{e^{ix} - e^{-ix}}{2i}$$

$$\begin{aligned}
 1 + 2 \sum_{k=1}^n \cos kx &= \sum_{k=-n}^n e^{ikx} = e^{-inx} \frac{e^{i(2n+1)x} - 1}{e^{ix} - 1} \\
 &= \frac{e^{i(n+1)x} - e^{-inx}}{e^{ix} - 1} = \frac{e^{\frac{1}{2}ix} e^{i(n+\frac{1}{2})x} - e^{-i(n+\frac{1}{2})x}}{e^{\frac{1}{2}ix} (e^{\frac{1}{2}ix} - e^{-\frac{1}{2}ix})} \\
 &= \frac{\sin(n + \frac{1}{2})x}{\sin \frac{1}{2}x}.
 \end{aligned}$$

■

Lemma 2.47 The Dirichlet kernel satisfies the following properties:

- (a) $D_n(x) = D_n(-x)$,
- (b) For every n ,

$$\frac{1}{2\pi} \int_0^{2\pi} D_n(x) dx = 1.$$

(c) Let $x \in \mathbb{R}$ and let $[a, b]$ be an interval not containing points of the form $x + 2\pi k$, then for every $f \in H(a, b)$,

$$\lim_{n \rightarrow \infty} \int_a^b f(t) D_n(x - t) dt = 0.$$

Proof:

(a) Obvious.

(b) Set $f = 1 \in H(0, 2\pi)$. Then,

$$a_0^f = \frac{1}{2\pi} \int_0^{2\pi} f(x) dx = 1 \quad \text{and} \quad a_n^f = b_n^f = 0 \quad \text{for } n \geq 1.$$

It follows that for every $n \geq 1$,

$$S_n f(x) = 1 = \frac{1}{2\pi} \int_0^{2\pi} D_n(x - t) dt.$$

Substituting $x = 0$ we recover the desired result.

(c) We have

$$\int_a^b f(t) D_n(x - t) dt = \int_a^b \frac{f(t)}{\sin\left(\frac{x-t}{2}\right)} \sin\left((n + \frac{1}{2})(x - t)\right) dt.$$

Since the interval $[a, b]$ does not contain points of the form $x + 2\pi k$, the denominator is bounded away from zero, hence the first term of the integrand is bounded and integrable. It follows from Proposition 2.42 that the limit as $n \rightarrow \infty$ vanishes.

■

Theorem 2.48 (Localization principle) Let $f, g \in H(0, 2\pi)$ and suppose that $f(x) = g(x)$ in a neighborhood of a point $x_0 \in (0, 2\pi)$. Then,

$$\lim_{n \rightarrow \infty} (S_n f(x_0) - S_n g(x_0)) = 0.$$

In particular, if the Fourier series of f converges at x_0 , so does the Fourier series of g , and to the same limit. (This means that the convergence of the Fourier series of a function at a point only depends on its value in an arbitrarily small neighborhood of that point, hence the term “localization”.)

Comment: The Fourier series of f at 0 only depends on its values in a neighborhood of 0 and 2π .

Proof: Suppose that $f(x) = g(x)$ in $(x_0 - \delta, x_0 + \delta)$. Then

$$\begin{aligned} S_n f(x_0) - S_n g(x_0) &= \frac{1}{2\pi} \int_0^{2\pi} (f(t) - g(t)) D_n(x_0 - t) dt \\ &= \frac{1}{2\pi} \int_0^{x_0 - \delta} (f(t) - g(t)) D_n(x_0 - t) dt \\ &\quad + \frac{1}{2\pi} \int_{x_0 + \delta}^{2\pi} (f(t) - g(t)) D_n(x_0 - t) dt, \end{aligned}$$

where we used the fact that $f = g$ in the rest of the interval. Since t in the integrands does not assume values of the form $x_0 + 2\pi k$, it follows from Lemma 2.47(c) that the limits as $n \rightarrow \infty$ vanish. ■

2.5 Fejér sums

The basic question remains: under what conditions does the Fourier series of a function (if it exists) converge pointwise to the function? The question was asked already by Fourier himself in the beginning of the 19th century. Dirichlet proved that the Fourier series of continuously differentiable functions converges everywhere. It turns out, however, that continuity is not enough for pointwise convergence everywhere. For example, the series

$$f(x) = \sum_{n=1}^{\infty} \frac{1}{n^2} \sin\left((2^{n^3} + 1) \frac{x}{2}\right)$$

converges uniformly to a continuous function, but its Fourier series diverges at certain points.

Rather than looking at the partial sums $S_n f$, we can consider their running average:

$$\sigma_n f = \frac{1}{n+1} \sum_{k=0}^n S_k f.$$

Such sums are called after Lipót Fejér. It turns out that Fejér sums are much better behaved than the Fourier partial sums. Note that it is trivial that if $S_n f(x)$ converges as $n \rightarrow \infty$, then so does $\sigma_n f(x)$, and to the same limit.

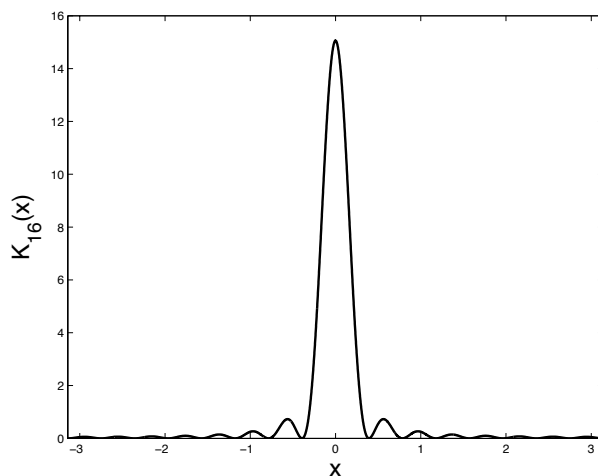
Consider the Fejér sums in explicit form:

$$\begin{aligned} \sigma_n f(x) &= \frac{1}{n+1} \sum_{k=0}^n \frac{1}{2\pi} \int_0^{2\pi} f(t) D_k(x-t) dt \\ &= \frac{1}{2\pi} \int_0^{2\pi} f(t) \left(\frac{1}{n+1} \sum_{k=0}^n D_k(x-t) \right) dy \\ &\equiv \frac{1}{2\pi} \int_0^{2\pi} f(t) K_n(x-t) dy, \end{aligned}$$

where,

$$K_n(x) = \frac{1}{n+1} \frac{\sin^2 \frac{1}{2} nx}{\sin^2 \frac{1}{2} x},$$

which can be proved by induction. The function $K_n(x)$ is known as the **Fejér kernel** (גרעין פייר).



The Fejér kernel is also normalized, as found by setting $f \equiv 1$:

$$1 = \sigma_n f(x) = \frac{1}{2\pi} \int_0^{2\pi} K_n(y) dy.$$

It differs a lot from the Dirichlet kernel in that it is non-negative.

Like the Dirichlet kernel it is “centered” at zero. That is, for every $\varepsilon > 0$,

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_{\varepsilon}^{2\pi-\varepsilon} K_n(x) dx &= \lim_{n \rightarrow \infty} \frac{1}{n+1} \int_{\varepsilon}^{2\pi-\varepsilon} \frac{\sin^2 \frac{1}{2}nx}{\sin^2 \frac{1}{2}x} dx \\ &\leq \lim_{n \rightarrow \infty} \frac{1}{n+1} \int_0^{2\pi} \frac{dx}{\sin^2 \frac{1}{2}\delta} = 0. \end{aligned}$$

Theorem 2.49 (Fejér) For every $f \in C([0, 2\pi])$ such that $f(0) = f(2\pi)$:

$$\lim_{n \rightarrow \infty} \|\sigma_n f - f\|_{\infty} = 0.$$

Proof: Since continuous functions on compact intervals are uniformly continuous,

$$\forall \varepsilon > 0 \quad \exists \delta > 0 : \forall |x - y| < \delta \quad |f(x) - f(y)| < \varepsilon.$$

Using the normalization of the Fejér kernel:

$$\begin{aligned} |\sigma_n f(x) - f(x)| &= \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} (f(t) - f(x)) K_n(x-t) dt \right| \\ &\leq \left| \frac{1}{2\pi} \int_{|x-t| \geq \delta} (f(t) - f(x)) K_n(x-t) dt \right| + \left| \frac{1}{2\pi} \int_{|x-t| < \delta} (f(t) - f(x)) K_n(x-t) dt \right| \\ &\leq 2\|f\|_{\infty} \frac{1}{2\pi} \int_{|x| \geq \delta} K_n(y) dy + \frac{\varepsilon}{2\pi} \int_0^{2\pi} K_n(y) dy, \end{aligned}$$

and we used the fact that the Fejér kernel is non-negative. For sufficiently large n the right hand side can be made smaller than a constant (independent of x) times ε , which completes the proof. ■

2.6 Applications of Fourier series

The Fourier series is named after Jean-Baptiste Joseph Fourier (1768–1830). Fourier introduced the series for the purpose of solving the heat equation in a metal plate, publishing his initial results in his 1807. Fourier's research established that an arbitrary *continuous* function can be represented by a trigonometric series. Early ideas of decomposing a periodic function into the sum of simple oscillating functions date back to the 3rd century BC, when ancient astronomers proposed an empiric model of planetary motions.

TA material 2.1 The **heat equation** (משוואת החום) describes the evolution over time of a temperature field. In its simplest formulation, let $T(x, t)$ denote the temperature at a point x and time t , where $x \in [0, 2\pi]$ and $t \geq 0$. Suppose that the initial temperature $T(x, 0)$ is given. Then, $T(x, t)$ satisfies the **partial differential equation**,

$$\frac{\partial T}{\partial t}(x, t) = k \frac{\partial^2 T}{\partial x^2}(x, t),$$

where $k > 0$ is a constant. In addition, one has to specify **boundary conditions** (תנאי שפה). For example, one may dip the ends of the rod in ice, so that for every t ,

$$T(0, t) = T(2\pi, t) = 0.$$

This is the problem that Fourier wanted to solve. He notices that every function of the form

$$T(x, t) = \sin nx e^{-kn^2 t}$$

is a solution to the heat equation satisfying the boundary conditions. Since the heat equation is linear, every linear combination of the form

$$T(x, t) = \sum_{n=1}^{\infty} a_n \sin nx e^{-kn^2 t}$$

is also a solution to the heat equation, satisfying the boundary conditions.