5 Basic Principles of Linear Analysis

Mathematical objects studied in linear analysis are linear transformations between vector spaces endowed with proper concepts of limit. Linear analysis, therefore, provides suitable language and framework for modeling linear phenomena, and, moreover, often suggests feasible methods for solving the corresponding problems. This is most clearly seen in the case of linear algebra when the vector spaces concerned are finite-dimensional.

This chapter is devoted to the most basic principles of linear analysis. Emphasis will be placed on the case when vector spaces are normed vector spaces, although weaker concepts of limit other than in terms of norm will occasionally be considered in view of subsequent applications.

The first basic principles are those arising from the Baire category theorem, and those from separation of sets by hyperplanes. These principles will be presented first, because they are fundamental in many constructs of linear analysis.

In the latter part of the chapter, considerable weight is laid on geometric aspects of linear analysis, with the introduction of Hilbert spaces. The main ingredients are the Riesz representation of continuous linear functionals on Hilbert spaces and Fourier expansion of elements of a Hilbert space with respect to an orthonormal basis.

Recall that vector spaces considered in our discourse are either over the complex field $\mathbb{C}$ or over the real field $\mathbb{R}$; when specification is desirable, they are called complex vector spaces or real vector spaces, according to whether they are over the complex or the real field. As usual, the smallest vector subspace containing a subset S of a vector space is called the vector space **spanned** by S and is denoted by $\langle S\rangle$.

5.1 The Baire category theorem

The Baire category theorem reveals a deep property of complete metric spaces; it is not usually applied directly, but through its consequences, such as the principle of uniform boundedness and the open mapping theorem. We shall present in this section the Baire category theorem and the principle of uniform boundedness; while the open mapping theorem and some of its consequences will be treated in Section 5.2.

Let M be a metric space. A subset S of M is said to be **nowhere dense** in M if the closure $\bar{S}$ of S contains no nonempty open balls of M. A subset A of M is said to be **of the first category** if A is a countable union of nowhere dense subsets of M. Otherwise A is said to be **of the second category**.

Theorem 5.1.1 (Baire category theorem) *A complete metric space M is of the second category.*

Proof It is required to show that if M is a union $\bigcup_{n=1}^{\infty} S_n$ of closed sets, then one of the S_n contains a nonempty open ball. Suppose the contrary, then each S_n^c has a nonempty intersection with every open ball. Thus if B_0 is an open ball with radius 1, $S_1^c \cap B_0$ contains an open ball $B_1 = B_{r_1}(x_1)$ as well as the closed ball $C_1 = C_{r_1}(x_1)$ with $r_1 < \frac{1}{2}$. Then $S_2^c \cap B_1$ contains an open ball $B_2 = B_{r_2}(x_2)$ and the closed ball $C_2 = C_{r_2}(x_2)$ with $r_2 < \frac{1}{2^2}$. Proceed in this way; a sequence of open balls $\{B_k\}$, $B_k = B_{r_k}(x_k)$, is obtained such that the closed ball $C_{k+1} := C_{r_{k+1}}(x_{k+1}) \subset S_{k+1}^c \cap B_k$ and $0 < r_k < 2^{-k}$, $k = 1, 2, \ldots$. Since $\{C_k\}$ is decreasing, $\{x_k, x_{k+1}, \ldots\} \subset C_k$, the sequence $\{x_k\}$ is a Cauchy sequence, hence $x_k \to x$ in M. But for each k, $x \in C_k \subset B_{k-1}$, or $x \in \bigcap_{k=1}^{\infty} B_k$, hence, $x \in \bigcap_{k=1}^{\infty} S_k^c = (\bigcup_{k=1}^{\infty} S_k)^c = \emptyset$, which is absurd. ■

Theorem 5.1.2 (Principle of uniform boundedness) *Let $\{f_\alpha\}$ be a family of continuous nonnegative functions defined on a Banach space X with the following properties:*

(1) *$f_\alpha(x+y) \le f_\alpha(x) + f_\alpha(y)$ for x, y in X and for each α;*
(2) *$f_\alpha(\lambda x) = |\lambda| f_\alpha(x)$, for $\lambda \in \mathbb{C}$ or $\mathbb{R}$ (depending on whether X is a complex or a real space), $x \in X$ and for each α; and*
(3) *$\sup_\alpha f_\alpha(x) < \infty$ for each $x \in X$.*

Then there is $N > 0$, such that

$$\sup_\alpha f_\alpha(x) \le N\|x\|$$

for all $x \in X$.

Proof For each $n \in \mathbb{N}$, let

$$S_n = \{x \in X : f_\alpha(x) \le n \; \forall \alpha\} = \bigcap_\alpha \{x \in X : f_\alpha(x) \le n\}.$$

Each S_n is closed and from (3), $X = \bigcup_n S_n$. By Theorem 5.1.1, for some n_0, S_{n_0} contains a ball $B = C_r(x_0)$, or

$$\sup_{\alpha; x \in B} f_\alpha(x) \le n_0.$$

Now, there is $N > 0$ such that

$$f_\alpha(x) \leq N$$

for all α if $\|x\| = 1$. To see this, for $x \in X$ with $\|x\| = 1$ and any α,

$$\begin{aligned} f_\alpha(x) = \frac{1}{r} f_\alpha(rx) &\leq \frac{1}{r}\{f_\alpha(rx + x_0) + f_\alpha(-x_0)\} \\ &\leq \frac{1}{r}\left\{n_0 + \sup_\alpha f_\alpha(-x_0)\right\} =: N. \end{aligned}$$

Now, for any $x \neq 0$ and any α,

$$f_\alpha(x) = \|x\| f_\alpha\left(\frac{x}{\|x\|}\right) \leq N\|x\|. \qquad \blacksquare$$

Actually, the principle of uniform boundedness is usually referred to the following special case of Theorem 5.1.2.

Theorem 5.1.3 *Let $\{T_\alpha\} \subset L(X, Y)$, where X is a Banach space and Y a n.v.s. Then $\sup_\alpha \|T_\alpha\| < \infty$ if and only if $\sup_\alpha \|T_\alpha x\| < +\infty$ for each $x \in X$.*

Proof That $\sup_\alpha \|T_\alpha\| < \infty$ implies that $\sup_\alpha \|T_\alpha x\| < \infty$ for all $x \in X$ is obvious; to show the other direction of implication, let $f_\alpha(x) = \|T_\alpha x\|$ and apply Theorem 5.1.2. ■

Theorem 5.1.4 (Banach–Steinhaus) *Let $\{T_n\} \subset L(X, Y)$, where X is a Banach space and Y a n.v.s. Suppose that $Tx = \lim_{n\to\infty} T_n x$ exists for each $x \in X$. Then $T \in L(X, Y)$ and $\|T\| \leq \liminf_{n\to\infty} \|T_n\| \leq \sup_n \|T_n\| < \infty$.*

Proof T is obviously a linear operator from X into Y. Since $\lim_{n\to\infty} T_n x$ exists, it follows that $\sup_n \|T_n x\| < \infty$ and hence $\sup_n \|T_n\| < \infty$, by Theorem 5.1.3. Now,

$$\begin{aligned} \|T\| &= \sup_{\|x\|=1} \|Tx\| = \sup_{\|x\|=1} \left\|\lim_{n\to\infty} T_n x\right\| \\ &= \sup_{\|x\|=1} \left(\lim_{n\to\infty} \|T_n x\|\right) \leq \sup_{\|x\|=1} \left(\liminf_{n\to\infty} \|T_n\| \cdot \|x\|\right) \\ &= \liminf_{n\to\infty} \|T_n\| \leq \sup_n \|T_n\| < \infty. \end{aligned} \qquad \blacksquare$$

Exercise 5.1.1 Let $\{T_n\} \subset L(X, Y)$, where both X and Y are Banach spaces. A necessary and sufficient condition for $\lim_{n\to\infty} T_n x$ to exist for each $x \in X$ is:

$$\begin{cases} (1) & \lim_{n\to\infty} T_n x \text{ exists for } x \text{ in a dense subset of } X; \\ (2) & \{\|T_n\|\} \text{ is bounded.} \end{cases}$$

Theorem 5.1.5 (C. Neumann) *Suppose that T is a bounded linear operator from a Banach space X into itself with $\|T\| < 1$. Then $(1-T)^{-1}$ exists, $(1-T)^{-1} \in L(X)$, and $(1-T)^{-1}x = \lim_{n\to\infty} \sum_{k=0}^{n} T^k x = \sum_{k=0}^{\infty} T^k x$.*

Proof For each $x \in X$, let $x_n = \sum_{k=0}^{n} T^k x$. Since for $n > m$,

$$\|x_n - x_m\| = \left\| \sum_{k=m+1}^{n} T^k x \right\| \leq \left(\sum_{k=m+1}^{n} \|T\|^k \right) \|x\|,$$

$\{x_n\}$ is a Cauchy sequence in X. Let $Sx = \lim_{n\to\infty} x_n = \lim_{n\to\infty}(\sum_{k=0}^{n} T^k x)$. By Theorem 5.1.4, S is a bounded linear operator. Now,

$$\begin{aligned}(1-T)Sx &= (1-T)\left(\lim_{n\to\infty} \sum_{k=0}^{n} T^k x \right) = \lim_{n\to\infty} \left((1-T) \sum_{k=0}^{n} T^k x \right) \\ &= \lim_{n\to\infty} (x - T^{n+1}x) = x,\end{aligned}$$

because $\|T^{n+1}x\| \leq \|T\|^{n+1}\|x\| \to 0$, implying that $T^{n+1}x \to 0$; similarly, $S(1-T)x = x$ for $x \in X$. Hence $S = (1-T)^{-1}$. ∎

Exercise 5.1.2 Suppose that $T \in L(X)$, $T \neq 0$, where X is a Banach space. Show that for $\lambda \in \mathbb{C}$ with $|\lambda| < \|T\|^{-1}$ the operator $I - \lambda T$ is bijective. Expand $(I - \lambda T)^{-1}$ in terms of λ and T and their powers.

We now apply the Baire category theorem to show the existence of continuous functions on the finite closed interval $[a, b]$ which are nowhere differentiable on $[a, b]$.

Fix a finite closed interval $[a, b]$ and let $I = [a, c]$, where $b < c < \infty$.

Lemma 5.1.1 *Suppose that $f \in C(I)$ and let $\varepsilon > 0$ and $L > 0$ be given. Then there is a continuous and piece-wise linear function g on I such that $\max_{x\in I} |g(x) - f(x)| \leq \varepsilon$, and the absolute value of the slope of each line segment of the graph of g is greater than L.*

Proof Let $\delta > 0$ be chosen so that $|f(x) - f(y)| < \frac{\varepsilon}{4}$ if $|x - y| < \delta$. Consider a partition $a = x_0 < x_1 < \cdots < x_{k-1} < x_k = c$ of I, with $|x_j - x_{j-1}| < \delta$ for $j = 1, \ldots, k$, and let $P_0 = (x_0, f(x_0))$, $P_1 = (x_1, f(x_1) + \frac{3}{4}\varepsilon), \ldots, P_j = (x_j, f(x_j) + (-1)^{j-1}\frac{3}{4}\varepsilon), \ldots, P_k = (x_k, f(x_k))$. Let g be the piece-wise linear function whose graph consists of the line segments $[P_0, P_1]$, $[P_1, P_2], \ldots, [P_{k-1}, P_k]$. Then g is continuous and $\max_{x\in I} |g(x) - f(x)| \leq \varepsilon$. If we choose δ small enough, then the absolute value of the slope of each $[P_{j-1}, P_j]$, $j = 1, \ldots, k$, is greater than L. ∎

Theorem 5.1.6 *There is a continuous function on $[a, b]$ which is nowhere differentiable on $[a, b]$.*

Proof Let $I = [a, c]$, $b < c < \infty$. It is sufficient to show that there is $f \in C(I)$ such that f is not differentiable at every point of $[a, b]$; actually, should f be differentiable from the left at b, the function $f + g$ is differentiable nowhere on $[a, b]$ if g is a continuous function on $[a, b]$ which is differentiable on $[a, b)$, but not differentiable from the left at b. As usual, we endow $C(I)$ with sup-norm, then $C(I)$ is a complete metric space. Consider the set S of functions f in $C(I)$ such that for some $\xi \in [a, b]$, the set $\left\{\frac{f(\xi+h)-f(\xi)}{h} : 0 < h \le c - b\right\}$ is bounded. Clearly, S contains all functions in $C(I)$ which are differentiable somewhere on $[a, b]$. For $n \in \mathbb{N}$, let

$$S_n = \left\{f \in S : \sup_{0<h\le c-b}\left|\frac{f(\xi + h) - f(\xi)}{h}\right| \le n \text{ for some } \xi \in [a, b]\right\}.$$

Observe that $S = \bigcup_n S_n$. We claim first that each S_n is closed. Let $\{f_k\}$ be a sequence in S_n which converges to f in $C(I)$. To claim that S_n is closed is to show that $f \in S_n$. For each k, there is $\xi_k \in [a, b]$ such that

$$\sup_{0<h\le c-b}\left|\frac{f_k(\xi_k + h) - f_k(\xi_k)}{h}\right| \le n.$$

Since $[a, b]$ is compact, $\{\xi_k\}$ has a subsequence which converges to $\xi \in [a, b]$. If necessary, replace $\{f_k\}$ by a subsequence of itself; we may assume that $\{\xi_k\}$ converges to ξ. For $0 < h \le c - b$ and $\varepsilon > 0$, there is $N = N(h, \varepsilon) \in \mathbb{N}$ such that $k > N$ implies $\sup_{x\in I} |f_k(x) - f(x)| < \frac{\varepsilon h}{4}$. Since f is uniformly continuous on I and $\xi_k \to \xi$, there is $N_1 > N$ such that $|f(\xi_k) - f(\xi)| < \frac{\varepsilon h}{4}$ and $|f(\xi + h) - f(\xi_k + h)| < \frac{\varepsilon h}{4}$ whenever $k > N_1$. Thus, for $k > N_1$, we have

$$\begin{aligned}\left|\frac{f(\xi + h) - f(\xi)}{h}\right| &\le \frac{1}{h}\{|f_k(\xi_k + h) - f_k(\xi_k)| + |f(\xi_k) - f(\xi)| + |f_k(\xi_k) - f(\xi_k)| \\ &\quad + |f(\xi_k + h) - f_k(\xi_k + h)| + |f(\xi + h) - f(\xi_k + h)|\} \\ &< \left|\frac{f_k(\xi_k + h) - f_k(\xi_k)}{h}\right| + \varepsilon \le n + \varepsilon;\end{aligned}$$

hence, $\sup_{0<h\le c-b}\left|\frac{f(\xi+h)-f(\xi)}{h}\right| \le n + \varepsilon$ for $\varepsilon > 0$, consequently,

$$\sup_{0<h\le c-b}\left|\frac{f(\xi + h) - f(\xi)}{h}\right| \le n$$

and $f \in S_n$. This shows that S_n is closed for $n \in \mathbb{N}$.

Next we claim that each S_n is nowhere dense in $C(I)$. For this, it is sufficient to show that $C(I)\backslash S_n$ is dense in $C(I)$. Consider $f \in C(I)$ and $\varepsilon > 0$; we claim that there is $g \in C(I)\backslash S_n$ such that $\sup_{x\in I} |g(x) - f(x)| \le \varepsilon$. Let g be the continuous and piecewise linear function in Lemma 5.1.1 corresponding to ε and $L = n$, then, $g \in C(I)\backslash S_n$ and $\sup_{x\in I} |g(x) - f(x)| \le \varepsilon$. Hence, $C(I)\backslash S_n$ is dense in $C(I)$, and therefore S_n is nowhere dense in $C(I)$. Since $S = \bigcup_n S_n$ and each S_n is closed and nowhere dense in $C(I)$, S is of the first category. By Theorem 5.1.1, $C(I)$ is of the second category and therefore there is $f \in C(I)\backslash S$. Since S contains all functions which are somewhere differentiable on $[a, b]$, f is nowhere differentiable on $[a, b]$. ∎

An interesting application of Theorem 5.1.3 is considered in Exercise 5.9.1, to show the existence of a continuous periodic function whose Fourier series diverges at a point.

5.2 The open mapping theorem

Theorem 5.2.1 (Banach open mapping theorem) *Suppose that T is a bounded linear map from a Banach space X onto a Banach space Y. Then T maps open sets into open sets.*

Proof Since $T(G + x_0) = TG + Tx_0$, it suffices to prove that if G is a neighborhood of 0 in X, then TG contains an open ball centered at 0 in Y.

Step 1. A weaker claim will be shown first. Here is the claim: Let B^X be an open ball in X centered at 0, then there is an open ball B^Y in Y centered at 0 such that $B^Y \subset \overline{TB^X}$. For the proof, the open ball in X centered at x with radius r will be denoted by $B_r^X(x)$; the connotation of $B_r^Y(y)$ as an open ball in Y is similarly defined. Let $B^X = B_r^X(0)$ and $U = B_{\frac{r}{2}}^X(0)$. Then, $X = \bigcup_{n=1}^\infty (nU)$, and $Y = TX = \bigcup_{n=1}^\infty nTU$. The Baire category theorem implies that there is n_0 such that $\overline{n_0 TU} = n_0 \overline{TU}$ contains an open ball in Y and hence $\overline{TU}$ contains an open ball, say $B_\sigma^Y(\hat{y})$. Since $\hat{y} \in \overline{TU}$, there is $x_0 \in U$ such that $y_0 = Tx_0 \in B_{\frac{\sigma}{2}}^Y(\hat{y})$, and therefore $B_{\frac{\sigma}{2}}^Y(y_0) \subset B_\sigma^Y(\hat{y}) \subset \overline{TU}$. Now put $B^Y = B_{\frac{\sigma}{2}}^Y(0)$, then,

$$B^Y = B_{\frac{\sigma}{2}}^Y(y_0) - y_0 \subset \overline{TU} - Tx_0 \subset \overline{TU - Tx_0} \subset \overline{T(U+U)} \subset \overline{TB^X},$$

as is claimed.

Step 2. Let G be any open set containing 0 in X and let $B_r^X(0) \subset G$. Put $B_0^X = B_{\frac{r}{2}}^X(0)$. By Step 1, there is a ball $B_0^Y = B_\sigma^Y(0)$ in Y such that $B_0^Y \subset \overline{TB_0^X}$. It will be shown that $TB_r^X(0) \supset B_0^Y$. For this purpose, let $B_i^X = B_{\varepsilon_i}^X(0)$, $\varepsilon_i = \frac{r}{2^{i+1}}$, $i = 1, 2, \ldots$. By Step 1, there is a sequence $B_i^Y = B_{\eta_i}^Y(0)$ of balls in Y such that $\eta_i \to 0$ and $B_i^Y \subset \overline{TB_i^X}$. For $y \in B_0^Y$, there is $x_0 \in B_0^X$ such that $\|y - Tx_0\| < \eta_1$; then there is $x_1 \in B_1^X$ such that $\|y - Tx_0 - Tx_1\| < \eta_2$. Proceeding in this way, we find a sequence $\{x_i\}$ such that $x_i \in B_i^X$ and

$$\left\| y - \sum_{i=0}^n Tx_i \right\| = \left\| y - T\left(\sum_{i=0}^n x_i\right) \right\| < \eta_n,$$

$n = 1, 2, \ldots$. Now, $\| \sum_{i=m}^{m+l} x_i \| \le \sum_{i=m}^{m+l} \varepsilon_i \to 0$ uniformly in l as $m \to \infty$, which implies that $\{\sum_{i=0}^n x_i\}$ is a Cauchy sequence. Set $x = \lim_{n\to\infty} \sum_{i=0}^n x_i$, then

$$Tx = \lim_{n\to\infty} T\left(\sum_{i=0}^n x_i\right) = \lim_{n\to\infty} \sum_{i=0}^n Tx_j = y.$$

But $\|x\| \le \sum_{i=0}^\infty \|x_i\| < \sum_{i=0}^\infty \frac{r}{2^{i+1}} = r$, i.e. $x \in B_r^X(0)$, hence $y \in TB_r^X(0)$. ∎

Corollary 5.2.1 *If T is an injective continuous linear map from a Banach space onto a Banach space, then T^{-1} is a bounded linear map.*

Exercise 5.2.1 As a complement to Theorem 5.2.1, show that if l is a nonzero linear functional on a n.v.s. not necessarily continuous, then l maps open sets into open sets.

Exercise 5.2.2 Let X be a n.v.s. and F a closed vector subspace of X. For $x \in X$, let $[x] = x + F$.

(i) Show that $[x] = [y]$ if and only if $y \in [x]$.

(ii) Define $[x] + [y] = [x + y]$, $\lambda[x] = [\lambda x]$ (λ scalar). Show that both operations are well defined and $X/F := \{[x] : x \in X\}$ becomes a vector space under these operations.

(iii) For $[x] \in X/F$, define $\|[x]\| = \inf_{y\in[x]} \|y\|$. Show that $\|[x]\|$ is well defined and that it defines a norm on X/F.

(iv) Define $\tau : X \mapsto X/F$ by $\tau(x) = [x]$. Show that τ is a linear open mapping from X onto X/F. The map τ is called the **canonical map** from X onto X/F.

5.3 The closed graph theorem

For n.v.s.'s X and Y over the same field, a n.v.s. $X \oplus Y$, called the direct sum of X and Y, is constructed as follows. Let $X \oplus Y = \{[x, y] : x \in X,\ y \in Y\}$, on which vector space operations are defined by

$$[x_1, y_1] + [x_2, y_2] = [x_1 + x_2, y_1 + y_2]; \quad \alpha[x, y] = [\alpha x, \alpha y],$$

and a norm is defined by

$$\|[x, y]\| = \{\|x\|^2 + \|y\|^2\}^{\frac{1}{2}}.$$

This norm is so chosen, that when X and Y are inner product spaces (to be introduced later in Section 5.6), so is $X \oplus Y$.

That $X \oplus Y$ is a n.v.s. is a direct consequence of its definition. Observe that when both X and Y are Banach spaces, so is $X \oplus Y$.

Henceforth, by a **linear operator** T from a vector space X into a vector space Y, we shall mean that the domain of T, denoted $D(T)$, is a vector subspace of X, not necessarily the whole space X. Now, if both X and Y are n.v.s.'s over the same field, and if T is a linear operator from X into Y, T is called a **closed operator** if its graph $G(T) := \{[x, Tx] : x \in D(T)\}$ is closed in $X \oplus Y$; i.e. if $\{x_n\} \subset D(T)$ with $\lim_{n\to\infty} x_n = x$ and $\lim_{n\to\infty} Tx_n = y$, then $x \in D(T)$ and $Tx = y$. If T is a linear operator from X into Y and the closure of $G(T)$ in $X \oplus Y$ is the graph of a linear operator, then T is called closable.

Example 5.3.1 Let $X = Y = C[0,1]$, $D(T) = \{f \in X : f' \in X\}$ and $Tf = f'$ for $f \in D(T)$. Then T is not bounded on $D(T)$, but T is a closed operator. That T is not bounded on $D(T)$ follows from

$$\|Tf_n\| = n\|f_{n-1}\|, \quad n = 1, 2, \dots,$$

where $f_n(t) = t^n$, $t \in [0,1]$. That T is closed is left as an exercise.

Exercise 5.3.1 Show that the linear operator T in Example 5.3.1 is closed.

Remark For a linear operator, its domain of definition has to be specified. For example, the differential operator T in Example 5.3.1 has to be considered as a different operator if its domain of definition $D(T)$ is changed to $D(T) = \{f \in X : f'' \in X\}$. Note that when defined on the new domain of definition, T is not closed, but closable.

Proposition 5.3.1 *If X and Y are n.v.s.'s, then a linear operator from X into Y is closable if and only if*

$$\{x_n\} \subset D(T), \quad \lim_{n\to\infty} x_n = 0, \quad \textit{and} \quad \lim_{n\to\infty} Tx_n = y, \quad \textit{then}\; y = 0. \tag{5.1}$$

Proof That (5.1) is necessary for T to be closable is obvious. To show that (5.1) is sufficient for T to be closable, let $[x, y] \in \overline{G(T)}$, i.e. there is $[x_n, Tx_n] \in G(T)$ such that $[x_n, Tx_n] \to [x, y]$. Define $Sx = y$. Because of (5.1), one verifies easily that S is well defined i.e. if $[x_n, Tx_n] \to [x, y]$ and $[x_n', Tx_n'] \to [x, y']$, then $y' = y$. Clearly, S is linear and $G(S) = \overline{G(T)}$. ∎

Example 5.3.2 Let Ω be an open set in $\mathbb{R}^n$, $C_\alpha \in C^k(\Omega)$ for $\alpha = (\alpha_1, \dots, \alpha_n)$, with $|\alpha| = \alpha_1 + \cdots + \alpha_n \le k$. Define $D(A) = \{f \in L^2(\Omega) \cap C^k(\Omega) : Af \in L^2(\Omega)\}$, where $A = \sum_{|\alpha|\le k} C_\alpha \partial^\alpha$. Then A is a closable linear operator from $L^2(\Omega)$ into $L^2(\Omega)$. If $\{f_j\} \subset D(A)$, $f_j \to 0$ in $L^2(\Omega)$, and $Af_j \to g$ in $L^2(\Omega)$, then for any $\varphi \in C_c^\infty(\Omega)$,

$$\begin{aligned}
\int_\Omega g(x)\varphi(x)dx &= \lim_{j\to\infty} \int_\Omega (Af_j)\varphi d\lambda^n \\
&= \lim_{j\to\infty} \int_\Omega \left(\sum_{|\alpha|\le k} C_\alpha(x)\partial^\alpha f_j(x) \right) \varphi(x)dx \\
&= \lim_{j\to\infty} \int_\Omega \sum_{|\alpha|\le k} (-1)^{|\alpha|}\partial^\alpha (C_\alpha(x)\varphi(x)) f_j(x)dx \\
&= \lim_{j\to\infty} \int_\Omega [A'\varphi](x) f_j(x)dx = 0,
\end{aligned}$$

which implies that $g = 0$. By Proposition 5.3.1, A is closable. Note that in the sequence of equalities above, the Fubini theorem and integration by parts have been used.

Exercise 5.3.2 Show that if T is a 1-1 closed operator, then T^{-1} is also closed.

Theorem 5.3.1 (Closed graph theorem) *A closed operator T with $D(T) = X$, a Banach space, and range in a Banach space Y, is bounded.*

Proof $G(T)$ is a closed subspace of $X \oplus Y$, and is therefore a Banach space. The linear operator $U : G(T) \mapsto X$ defined by

$$U[x, Tx] = x, \quad x \in X$$

is clearly one-to-one and continuous. Since $U(G(T)) = X$, by Corollary 5.2.1, U^{-1} is a continuous linear map from X onto $G(T)$, thus $T = VU^{-1}$ is continuous, where $V[x, Tx] = Tx$ is a continuous linear map from $G(T)$ to Y. ■

The following exercise is a comment on Theorem 5.3.1.

Exercise 5.3.3 Let X be the space of all sequences $(a_k)_{k\in\mathbb{N}}$ of real numbers such that $a_k \neq 0$ only for finitely many k's. X is a vector space under the usual way of defining addition and multiplication by scalars. For (a_k) in X, let $\|(a_k)\| = \max_k |a_k|$; then X is a n.v.s. Define $T : X \to X$ by $T(a_k) = (ka_k)$. Show that X is a closed operator on X, but is not bounded.

5.4 Separation principles

Consider a real vector space X; a subset E of X is said to be **convex** if $\alpha x + \beta y \in E$ whenever x and y are in E and α, β are nonnegative numbers with $\alpha + \beta = 1$. E is called a **convex cone** if it is convex and $\gamma E \subset E$ for all $\gamma > 0$. For a set $S \subset X$, there is a smallest convex set containing S. The smallest convex set containing S is called the **convex hull** of S and is usually denoted by Conv S, while the smallest convex cone containing S will be denoted by Con S. For $x \neq y$ in X, Conv$\{x, y\}$ is usually denoted by $[x, y]$ and is called the **line segment** with endpoints x and y, while, for $x \neq 0$ in X, Con$\{x\}$ is called the **half line** through x. In $\mathbb{R}^k$, the convex set $\Delta^{k-1} := \{x = (x_1, \ldots, x_k) : x_j \geq 0,\ j = 1, \ldots, k,\ \sum_{j=1}^k x_j = 1\}$ is called the **standard** $(k-1)$-**simplex**. Elements in X of the form $\sum_{j=1}^k \alpha_j x_j$ (k varies from element to element), where $x_1, \ldots, x_k$ are in X and $\alpha = (\alpha_1, \ldots, \alpha_k) \in \Delta^{k-1}$, are called **convex combinations** of $x_1, \ldots, x_k$; if $x_1, \ldots, x_k$ are in $S \subset X$, they are called convex combinations of elements in S.

For convenience, the fact that a real-valued function f assumes values $\geq \alpha$ on a set A will be expressed by $f(A) \geq \alpha$; the meaning of each of the expressions $f(A) > \alpha$, $f(A) \leq \alpha$, and $f(A) < \alpha$ is parallelly given.

Exercise 5.4.1 Let $S \subset X$. Prove the following statements:

(i) Conv S is the set of all convex combinations of elements in S.

(ii) Con $S = \{\sum_{j=1}^k \gamma_j x_j : k \in \mathbb{N},\ x_1, \ldots, x_k \in S,\ \gamma_j > 0,\ j = 1, \ldots, k\}$.

(iii) S is a convex cone if and only if $S + S \subset S$ and $\gamma S \subset S$ for all $\gamma > 0$.

A set $E \subset X$ is said to be **linearly open** if for any $x \in E$ and $y \in X$, $x + ty \in E$ if $|t|$ is small enough. Clearly, open sets in a n.v.s. X are linearly open. Note that if a linearly open convex cone contains the origin 0, then $E = X$.

Exercise 5.4.2 Show that a convex set $E \subset \mathbb{R}^n$ is linearly open if and only if E is open.

Exercise 5.4.3 Suppose that E is a convex cone in X, and S a convex set in X.

(i) Show that if $E \cap S = \emptyset$, then $E \cap (\text{Con } S) = \emptyset$.

(ii) If S is also a convex cone, then $E + S$ and $E - S$ are convex cones and they are linearly open if one of E and S is linearly open.

Theorem 5.4.1 *If E is a nonempty linearly open convex cone not containing* 0*, then there is a hyperplane H such that $E \cap H = \emptyset$.*

Proof Denote by $\mathcal{F}$ the family of all vector subspaces F of X such that $F \cap E = \emptyset$. $\mathcal{F}$ is not empty, because $\{0\} \in \mathcal{F}$. Order $\mathcal{F}$ by set-inclusion i.e. $F_1 \leq F_2$ if $F_1 \subset F_2$ for F_1 and F_2 in $\mathcal{F}$. If $\mathcal{T}$ is a chain in $\mathcal{F}$, then $\bigcup_{F \in \mathcal{T}} F$ is in $\mathcal{F}$ and is an upper bound of $\mathcal{T}$. By Zorn's lemma (cf. Section 3.12), $\mathcal{F}$ has a maximal element H.

Let $D = H + E$; by Exercise 5.4.3, D is a linearly open convex cone. We claim that $X = D \cup H \cup (-D)$ is a disjoint union. It is obvious that $D \cap H = \emptyset$, and hence $(-D) \cap H = \emptyset$. If $h \in D \cap (-D)$, then both h and $-h$ are in D and consequently $h + (-h) = 0$ is in D, contradicting the fact that $D \cap H = \emptyset$. Thus $D \cup H \cup (-D)$ is a disjoint union. It remains to show that $X = D \cup H \cup (-D)$. Let $x \in X$, but $x \notin H$. Then $H + \langle x \rangle$ meets E, because H is a maximal element of $\mathcal{F}$. Then there is $h \in H$ and $\lambda \in \mathbb{R}$, $\lambda \neq 0$ such that $h + \lambda x \in E$; as a consequence $\lambda x \in H + E = D$, and then $x \in D$ or $(-D)$ depending on $\lambda > 0$ or $\lambda < 0$. This shows that $X = D \cup H \cup (-D)$.

It will be shown presently that H is a hyperplane. This amounts to showing that if $x \in X$, but $x \notin H$, then $H + \langle x \rangle = X$. Fix such an x and let $y \in X$, $y \notin H$. One has to show that $y \in H + \langle x \rangle$ to conclude the proof. For this purpose, one may assume that $x \in D$ and $y \in (-D)$. Since $[x, y]$ is connected (see Theorem 1.9.1) and $X = D \cup H \cup (-D)$,

$$[x, y] \cap \{D \cup (-D)\} \subsetneq [x, y] = [x, y] \cap \{D \cup H \cup (-D)\},$$

therefore there is $h \in H \cap [x, y]$. Since $h \in [x, y]$ there are $\alpha \geq 0$, $\beta \geq 0$ with $\alpha + \beta = 1$, such that $h = \alpha x + \beta y$. Now $h \in H$ implies that $h \notin D$, which forces β to be > 0 and hence $y = \frac{1}{\beta} h - \frac{\alpha}{\beta} x \in H + \langle x \rangle$. The proof of the theorem is complete. ■

A basic principle on separation of sets by linear functional is the following consequence of Theorem 5.4.1.

Corollary 5.4.1 *Suppose that E is a nonempty linearly open convex cone in X, and C is a nonempty convex set in X such that $C \cap E = \emptyset$, then there is $\ell \in X'$ such that $\ell(C) \geq 0$ and $\ell(E) < 0$.*

Proof Put $D = E - \text{Con } C$. D is a linearly open convex cone and $0 \notin D$, because E and $\text{Con } C$ are disjoint, by Exercise 5.4.3. By Theorem 5.4.1, there is a hyperplane H in

X such that $H \cap D = \emptyset$. Choose $\ell \in X'$ with $\ker \ell = H$ and $\ell(D) < 0$. Now, for $x \in \operatorname{Con} C$, $y \in E$, and $\gamma > 0$

$$\begin{cases} \ell(y) < \gamma \ell(x); \\ \gamma \ell(y) < \ell(x). \end{cases}$$

Let $\gamma \searrow 0$; it follows that $\ell(y) \leq 0$ for $y \in E$ and $\ell(x) \geq 0$ for $x \in \operatorname{Con} C$. In particular, $\ell(C) \geq 0$.

It remains to show that $\ell(y) < 0$ for $y \in E$. Choose $x_0 \in X$ with $\ell(x_0) > 0$, then $y + tx_0 \in E$ if $|t|$ is small enough, because E is linearly open. Since $y + tx_0 \in E$, $\ell(y + tx_0) \leq 0$, and hence $\ell(y) \leq -t\ell(x_0) < 0$ if $t > 0$ is small, as is to be shown. ■

Note that in the proof of Corollary 5.4.1 we have used the well-known fact in linear algebra that a vector subspace of X is a hyperplane in X if and only if it is the kernel of a nonzero linear functional on X.

A real-valued function φ defined on a convex set S in X is called a **convex** function if $\varphi(\alpha x + \beta y) \leq \alpha\varphi(x) + \beta\varphi(y)$ for any x, y in S and any convex pair (α, β). If φ is convex, then $\varphi(\sum_{j=1}^k \alpha_j x_j) \leq \sum_{j=1}^k \alpha_j \varphi(x_j)$ for any convex combination $\sum_{j=1}^k \alpha_j x_j$ of elements of S, as is easily seen by induction on k.

Consider now a convex function φ defined on an open interval I of $\mathbb{R}$. For $a < b < c$ in I, from $b = \frac{c-b}{c-a}a + \frac{b-a}{c-a}c$ it follows that $\varphi(b) \leq \frac{c-b}{c-a}\varphi(a) + \frac{b-a}{c-a}\varphi(c) = \varphi(a) - \frac{b-a}{c-a}\{\varphi(c) - \varphi(a)\}$, or

$$\frac{\varphi(b) - \varphi(a)}{b - a} \leq \frac{\varphi(c) - \varphi(a)}{c - a};$$

similarly,

$$\frac{\varphi(c) - \varphi(a)}{c - a} \leq \frac{\varphi(c) - \varphi(b)}{c - b}.$$

From the sequence of inequalities,

$$\frac{\varphi(b) - \varphi(a)}{b - a} \leq \frac{\varphi(c) - \varphi(a)}{c - a} \leq \frac{\varphi(c) - \varphi(b)}{c - b},$$

one infers that if $x \neq y$ are in I, the quotient $\frac{\varphi(y)-\varphi(x)}{y-x}$ is bounded for y near x and is an increasing function of y. Thus, both $\varphi'_-(x) := \lim_{y\to x-} \frac{\varphi(y)-\varphi(x)}{y-x}$ and $\varphi'_+(x) = lim_{y\to x+} \frac{\varphi(y)-\varphi(x)}{y-x}$ exist and are finite; furthermore, $\varphi'_-(x) \leq \varphi'_+(x)$ and $\varphi'_+(x) \leq \varphi'_-(y)$ if $x < y$ are in I. The last inequality follows from $\varphi'_+(x) \leq \frac{\varphi(z)-\varphi(x)}{z-x} \leq \frac{\varphi(y)-\varphi(z)}{y-z}$ for z strictly between x and y, by letting $z \to y$. Since the left and right derivatives of φ exist and are finite at each point of I, φ is continuous on I. Now, for $x < y$ in I, the inequalities $\varphi'_-(x) \leq \varphi'_+(x) \leq \varphi'_-(y) \leq \varphi'_+(y)$ imply that both φ'_- and φ'_+ are monotone increasing functions on I. Next, for $x < y < z$ in I, one verifies that $\varphi'_+(x) \leq \varphi'_+(y) \leq \frac{\varphi(z)-\varphi(y)}{z-y}$, from which $\varphi'_+(x) \leq \varphi'_+(x+) \leq \frac{\varphi(z)-\varphi(x)}{z-x}$ follows when $y \to x$ (note that φ is continuous); then one concludes that $\varphi'_+(x) = \varphi'_+(x+)$, by letting $z \to x$. Thus φ'_+ is a right-continuous function; similarly, one can verify that φ'_- is a left-continuous function. The following proposition has been proved.

Proposition 5.4.1 *Suppose that φ is a convex function defined on an open interval I in $\mathbb{R}$. The following statements hold:*

(i) *The left derivative $\varphi'_-(x)$ and the right derivative $\varphi'_+(x)$ exist and are finite at each point x of I; and for $x < y$ in I, $\varphi'_-(x) \leq \varphi'_+(x) \leq \varphi'_-(y)$.*

(ii) *Both φ'_- and φ'_+ are monotone increasing.*

(iii) *φ'_- is left-continuous and φ'_+ is right-continuous.*

(iv) *For $x \in I$ and $m \in [\varphi'_-(x), \varphi'_+(x)]$, $\varphi(y) \geq \varphi(x) + m(y-x)$ for all $y \in I$.*

Exercise 5.4.4 Show that if φ is a convex function on a vector space X, then, for any $t \in \mathbb{R}$, the set $\{\varphi \leq t\}$ is convex and the set $\{\varphi < t\}$ is convex and linearly open.

A real-valued function q on a real vector space X is called a **sublinear functional** on X if

(1) $q(x+y) \leq q(x) + q(y)$, x, y in X;

(2) $q(\lambda x) = \lambda q(x)$, $x \in X$, $\lambda > 0$.

Note that a sublinear functional is necessarily convex.

Exercise 5.4.5 Suppose that q is a sublinear functional on X, and put $Q = \{q < 0\}$. Show that Q is a linearly open convex cone. Also show that $q(0) = 0$ and $-q(-x) \leq q(x)$ for $x \in X$.

Exercise 5.4.6 Suppose that q is the sublinear functional on $\mathbb{R}^n$ defined by $q(x) = \max_{1 \leq j \leq n} x_j$ if $x = (x_1, \ldots, x_n)$. Show that a linear functional on $\mathbb{R}^n$ satisfies $l \leq q$ if and only if there is $\alpha \in \Delta^{n-1}$ such that $l(x) = \sum_{j=1}^n \alpha_j x_j$.

Lemma 5.4.1 *Suppose that q is a sublinear functional on X with $Q = \{q < 0\} \neq \emptyset$. Let τ be a map from a set T into X. Then there is $\ell \in X'$, $\ell \neq 0$, with $\ell \leq q$ such that $\ell(\tau(T)) \geq 0$ if and only if $q(\mathrm{Con}\,\tau(T)) \geq 0$.*

Proof Suppose $q(\mathrm{Con}\,\tau(T)) \geq 0$. Then $(\mathrm{Con}\,\tau(T)) \cap Q = \emptyset$. By Corollary 5.4.1, there is $\hat{\ell} \in X'$, $\hat{\ell} \neq 0$, such that $\hat{\ell}(\mathrm{Con}\,\tau(T)) \geq 0$ and $\hat{\ell}(Q) < 0$. It will be shown presently that there is $\sigma > 0$ such that $\sigma\hat{\ell} \leq q$.

Define a map f from X into $\mathbb{R}^2$ by

$$f(x) = (q(x), -\hat{\ell}(x)), \quad x \in X,$$

and let C be the convex hull of $f(X)$; then $C \cap \mathring{\mathbb{R}}^2_- = \emptyset$, where $\mathring{\mathbb{R}}^2_- = \{(r_1, r_2) \in \mathbb{R}^2 : r_1 < 0, r_2 < 0\}$. Actually, if $v \in C$, there are $x_1, \ldots, x_k$ in X and

$\alpha = (\alpha_1, \ldots, \alpha_k) \in \Delta^{k-1}$ such that $v = (\sum_{j=1}^k \alpha_j q(x_j), -\hat{\ell}(\sum_{j=1}^k \alpha_j x_j))$; if $\sum_{j=1}^k \alpha_j q(x_j) < 0$, then $q(\sum_{j=1}^k \alpha_j x_j) \leq \sum_{j=1}^k \alpha_j q(x_j) < 0$, implying that $\sum_{j=1}^k \alpha_j x_j \in Q$ and hence $-\hat{\ell}(\sum_{j=1}^k \alpha_j x_j) > 0$; thus $v \notin \mathring{\mathbb{R}}_-^2$. By Corollary 5.4.1, there is (α_1, α_2) in $\mathbb{R}^2$ with $\alpha_1^2 + \alpha_2^2 > 0$ such that

$$\begin{cases} \alpha_1 r_1 + \alpha_2 r_2 < 0 & \text{for } (r_1, r_2) \in \mathring{\mathbb{R}}_-^2; \\ \alpha_1 q(x) - \alpha_2 \hat{\ell}(x) \geq 0 & \text{for } x \in X. \end{cases} \tag{5.2}$$

The first inequality in (5.2) shows that $\alpha_1 \geq 0, \alpha_2 \geq 0$, while the second inequality shows that $\alpha_1 > 0$ and $\alpha_2 > 0$, in that $Q \neq \emptyset$. Then, $\sigma \hat{\ell}(x) \leq q(x)$ for $x \in X$ by taking $\sigma = \alpha_1^{-1}\alpha_2$. Then $\ell := \sigma \hat{\ell}$ satisfies $\ell \leq q$, $\ell \neq 0$, and $\ell(x) \geq 0$ for $x \in \tau(T)$. The other direction of the Lemma is obvious. ■

Remark

(i) Since q is sublinear, the condition $q(\mathrm{Con}\,\tau(T)) \geq 0$ in Lemma 5.4.1 is equivalent to $q(\mathrm{Conv}\,\tau(T)) \geq 0$;

(ii) since $Q \neq \emptyset$, $\ell \neq 0$ is a consequence of $\ell \leq q$;

(iii) when $Q = \emptyset$, Lemma 5.4.1 also holds if we do not require that $\ell \neq 0$, because in this case $q(\mathrm{Con}\,\tau(T)) \geq 0$ always holds and ℓ is simply taken to be the zero functional.

It follows from the preceding remarks that Lemma 5.4.1 can be generalized to the following theorem.

Theorem 5.4.2 *Suppose that q is a sublinear functional on a real vector space X and τ a map from a set T into X. Then there is $\ell \in X'$ with $\ell \leq q$ such that $\ell(\tau(T)) \geq 0$ if and only if $q(\mathrm{Con}\,\tau(T)) \geq 0$.*

An immediate consequence of Theorem 5.4.2 is the following historically interesting result of Banach.

Corollary 5.4.2 (Banach) *If q is a sublinear functional on X, then there is $\ell \in X'$ such that $\ell \leq q$ on X.*

Proof In Theorem 4.5.2, take $\tau(t)$ to be the zero element of X for each $t \in T$. ■

If, for a real vector space X and a sublinear functional q on X, we let $X'(q)$ be the set of all those $\ell \in X'$ such that $\ell \leq q$, then $X'(q)$ is obviously convex, and is nonempty, by Corollary 5.4.2.

From Theorem 5.4.2, there follow two important consequences.

Theorem 5.4.3 (Hahn–Banach) *Let q be a sublinear functional on a real vector space X and suppose that Y is a vector subspace of X and $\ell \in Y'(q)$. Then there is $\hat{\ell} \in X'(q)$ such that $\ell(y) = \hat{\ell}(y)$ for $y \in Y$.*

Proof Define a sublinear functional $\hat{q}$ on $X \oplus Y$ by

$$\hat{q}(x,y) = q(x) + \ell(y), \quad x \in X,\ y \in Y,$$

and a map $\hat{\tau}$ from Y into $X \oplus Y$ by

$$\hat{\tau}(y) = (y, -y), \quad y \in Y.$$

Since $\hat{\tau}$ is linear, $\operatorname{Conv}\hat{\tau}(Y) = \hat{\tau}(Y)$. Now let $v \in \operatorname{Conv}\hat{\tau}(Y) = \hat{\tau}(Y)$. Then, $v = (y, -y)$ for some $y \in Y$ and $\hat{q}(v) = q(y) + \ell(-y) \geq 0$; this means that $\hat{q}(\operatorname{Conv}\hat{\tau}(Y)) \geq 0$. By Theorem 5.4.2, there is $(\hat{\ell}, \ell_Y) \in (X \oplus Y)'$ with $(\hat{\ell}, \ell_Y) \leq \hat{q}$ such that $(\hat{\ell}, \ell_Y)(y, -y) \geq 0$ for all $y \in$ Y, where $\hat{\ell} \in X'$ and $\ell_Y \in Y'$. But $(\hat{\ell}, \ell_Y) \leq \hat{q}$ if and only if $\hat{\ell} \leq q$ on X and $\ell_Y \leq \ell$ on Y. Now, $\ell_Y \leq \ell$ implies that $\ell_Y = \ell$ and $(\hat{\ell}, \ell_Y)(y, -y) = \hat{\ell}(y) - \ell_Y(y) = \hat{\ell}(y) - \ell(y) \geq 0$ for $y \in Y$ forces $\hat{\ell}(y) = \ell(y)$ for $y \in Y$. ■

Theorem 5.4.4 (Mazur–Orlicz) *Let q be a sublinear functional on a real vector space X and τ a map from a set T into X. Suppose that θ is a map from T into $\mathbb{R}$. Then there is $\ell \in X'(q)$ such that $\theta(t) \leq \ell(\tau(t))$ for all $t \in T$ if and only if for every positive integer n,*

$$\sum_{j=1}^{n} \alpha_j \theta(t_j) \leq q\left(\sum_{j=1}^{n} \alpha_j \tau(t_j)\right) \tag{5.3}$$

for all $t_1, \ldots, t_n$ in T and $\alpha = (\alpha_1, \ldots, \alpha_n) \in \Delta^{n-1}$.

Proof Consider $\hat{X} = X \oplus \mathbb{R}$. Define $\hat{q} : \hat{X} \mapsto \mathbb{R}$ by

$$\hat{q}(x, \lambda) = q(x) + \lambda, \quad x \in X,\ \lambda \in \mathbb{R},$$

then $\hat{q}$ is a sublinear functional on $\hat{X}$. Let now

$$\hat{\tau}(t) = (\tau(t), -\theta(t)), \quad t \in T.$$

Suppose that (5.3) holds, then for $n \in \mathbb{N}$, $t_1, \ldots, t_n$ in T and $\alpha = (\alpha_1, \ldots, \alpha_n) \in \Delta^{n-1}$,

$$\begin{aligned}\hat{q}\left(\sum_{j=1}^{n} \alpha_j \tau(t_j), -\sum_{j=1}^{n} \alpha_j \theta(t_j)\right) &= \hat{q}\left(\sum_{j=1}^{n} \alpha_j \hat{\tau}(t_j)\right) \\ &= q\left(\sum_{j=1}^{n} \alpha_j \tau(t_j)\right) - \sum_{j=1}^{n} \alpha_j \theta(t_j) \geq 0,\end{aligned}$$

or $\hat{q}(\operatorname{Conv}\hat{\tau}(T)) \geq 0$. By Theorem 5.4.2, there is $\hat{\ell} \in \hat{X}'$ with $\hat{\ell} \leq \hat{q}$ on $\hat{X}$ such that $\hat{\ell}(\hat{\tau}(T)) \geq 0$. But $\hat{\ell} = (\ell, \alpha)$, $\ell \in X'$, $\alpha \in \mathbb{R}$, and $\hat{\ell}(x, \lambda) = \ell(x) + \alpha\lambda$ for $x \in X$ and $\lambda \in \mathbb{R}$. Observe then that $\hat{\ell} \leq \hat{q}$ on $\hat{X}$ means that $\ell \leq q$ on X and $\alpha = 1$; hence, $\hat{\ell}(\hat{\tau}(t)) \geq 0$ for $t \in T$ implies that $\theta(t) \leq \ell(\tau(t))$ for $t \in T$. On the other hand, if there is $\ell \in X'$ with $\ell \leq q$ and $\ell(\tau(t)) \geq \theta(t)$ for $t \in T$, then (5.3) obviously holds. ■

Corollary 5.4.3 *Let X, q, and τ be as in Theorem 5.4.4, then*

$$\max_{\ell\in X'(q)} \inf \ell(\tau(T)) = \inf q(\operatorname{Conv}\tau(T)).$$

Proof Observe firstly that $\inf \ell(\tau(T)) \leq \inf q(\operatorname{Conv}\tau(T))$ holds for any $\ell \in X'(q)$, hence $\sup_{\ell\in X'(q)} \inf \ell(\tau(T)) \leq \inf q(\operatorname{Conv}\tau(T))$, and it remains to show that there is $\ell \in X'(q)$ such that $\inf \ell(\tau(T)) = \inf q(\operatorname{Conv}\tau(T))$. In the case where $\inf q(\operatorname{Conv}\tau(T)) = -\infty$, just take any $\ell \in X'(q)$ (recall that $X'(q) \neq \emptyset$, by Corollary 5.4.2). If $\inf q(\operatorname{Conv}\tau(T)) = \beta > -\infty$, let a function θ on T be defined by $\theta(t) = \beta$ for all $t \in T$. Then (5.3) holds trivially and we may apply Theorem 5.4.4 to find $\ell \in X'(q)$ such that $\beta \leq \ell(\tau(t))$ for all $t \in T$, i.e.

$$\inf \ell(\tau(T)) \geq \beta = \inf q(\operatorname{Conv}\tau(T)).$$

But, as we observed at the beginning of the proof, $\inf \ell(\tau(T)) \leq \inf q(\operatorname{Conv}\tau(T))$, therefore $\inf \ell(\tau(T)) = \inf q(\operatorname{Conv}\tau(T))$ and the proof is complete. ■

Exercise 5.4.7 Show that if C is a convex set in a real n.v.s. X, such that $\inf_{x\in C} \|x\| = \sigma > 0$, then there is $l \in X^*$ with $\|l\| = 1$ such that $l(x) \geq \sigma$ for all $x \in C$. (Hint: apply Corollary 5.4.3.)

The conclusion of Corollary 5.4.3 is a general form of J. von Neumann's minimax theorem in game theory, as illustrated in Exercise 5.4.8.

Exercise 5.4.8 (von Neumann minimax theorem) Suppose that (a_{ij}), $1 \leq i \leq m$, $1 \leq j \leq n$, is a given $m \times n$-matrix with real entries. For each $j = 1, \ldots, n$, define a function f_j on Δ^{m-1} by

$$f_j(\alpha) = \sum_{i=1}^{m} a_{ij}\alpha_i, \quad \alpha = (\alpha_1, \ldots, \alpha_n) \in \Delta^{m-1},$$

and define a quadratic form A on $\Delta^{m-1} \times \Delta^{n-1}$ by

$$A(\alpha, \beta) = \sum_{j=1}^{n} \beta_j f_j(\alpha) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\alpha_i\beta_j.$$

Now consider the sublinear functional q on $\mathbb{R}^n$, defined by $q(x) = \max_{1\leq j\leq n} x_j$ for $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, and let the map τ from Δ^{m-1} to $\mathbb{R}^n$ be defined by $\tau(\alpha) = (f_1(\alpha), \ldots, f_n(\alpha))$. Use Corollary 5.4.3 and the assertion of Exercise 5.4.6 to show the following minimax equality of von Neumann:

$$\min_{\alpha\in\Delta^{m-1}} \max_{\beta\in\Delta^{n-1}} A(\alpha, \beta) = \max_{\beta\in\Delta^{n-1}} \min_{\alpha\in\Delta^{m-1}} A(\alpha, \beta).$$

Exercise 5.4.9 Let q be a sublinear functional on a real vector space X and put $Q = \{x \in X : q(x) < 0\}$. Suppose that S is a convex cone in X such that $Q \cap S = \emptyset$, and define $\hat{q}$ on X by $\hat{q}(x) = \inf_{y\in S} q(x + y)$.

(i) Show that $\hat{q}$ is a sublinear functional on X and $\hat{q} \leq q$.

(ii) Show that if $\ell \in X'(\hat{q})$, then $\ell(x) \geq 0$ for $x \in S$.

Exercise 5.4.10 Show that Theorem 5.4.2 is a consequence of Corollary 5.4.2. (Hint: apply Corollary 5.4.2 with q replaced by $\hat{q}$, as defined in Exercise 5.4.9, with $S = \mathrm{Con}(\tau(T))$.)

Exercise 5.4.11 Show that Theorem 5.4.2, Theorem 5.4.3, and Theorem 5.4.4 are equivalent to each other. (Hint: Corollary 5.4.2 is a special case of the Hahn–Banach theorem.)

Exercise 5.4.12 Let Q be a proper linearly open convex cone in a real vector space X. Fix $x_0 \in Q$.

(i) Show that the family $L = \{\ell \in X' : \ell < 0 \text{ on } Q \text{ and } \ell(x_0) = -1\}$ is nonempty and that for $x \in X$, $\sup_{\ell \in L} \ell(x)$ is finite. (Hint: for $x \in X$ there is $\sigma > 0$ such that $x_0 + \sigma x \in Q$, from which assert that $\ell(x) \leq \frac{1}{\sigma}$ for $\ell \in L$.)

(ii) Put $q(x) = \sup_{\ell \in L} \ell(x)$ for $x \in X$. Show that q is a sublinear functional on X and that $Q = \{x \in X : q(x) < 0\}$.

In this final part of the section our discussion is restricted to real normed vector spaces; and our concern is the separation of convex sets by closed affine hyperplane. By an **affine hyperplane** we mean a translation of a hyperplane in a vector space, i.e. an affine hyperplane in a vector space X is a set of the form $x + H$, where $x \in X$ and H is a hyperplane in X. We recall from elementary linear algebra that a vector subspace of a vector space X is a hyperplane if and only if it is the kernel of a nonzero linear functional on X. Note that if ℓ_1, ℓ_2 are nonzero linear functionals on X, then $\ker \ell_1 = \ker \ell_2$ if and only if $\ell_1 = \alpha\ell_2$ for some nonzero scalar α. Thus an affine hyperplane in X is a set of the form $\{x \in X : \ell(x) = \alpha\}$ for some $\ell \in X'$ $(\ell \neq 0)$ and some scalar α. If X is a normed vector space, then, since the closure of a vector subspace of X is a vector subspace of X, every hyperplane in X is either closed or dense in X. Observe that a hyperplane $H = \ker \ell$, $\ell \in X'$, in a normed vector space X is closed if and only if $\ell \in X^*$, and hence a closed affine hyperplane in X is of the form $\{x \in X : \ell(x) = \alpha\}$ for some $\ell \in X^*$ $(\ell \neq 0)$ and some scalar α.

We fix now a real n.v.s. X. Nonempty sets A and B in X are said to be separated strictly by a closed affine hyperplane if there are $\ell \in X^*$ and $\alpha \in \mathbb{R}$ such that $\ell(x) < \alpha$ for $x \in A$ and $\ell(y) > \alpha$ for $y \in B$; while they are separated strictly in the strong sense if there are $\ell \in X^*, \alpha \in \mathbb{R}$, and $\varepsilon > 0$ such that $\ell(x) \leq \alpha - \varepsilon$ for $x \in A$ and $\ell(y) \geq \alpha + \varepsilon$ for $y \in B$. Note that $\ell \in X^*$ in the above definition is necessarily nonzero, and $\{x \in X : \ell(x) = \alpha\}$ is the closed affine hyperplane in question. A closed set of the form $\{x \in X : \ell(x) \leq \alpha\}$, where $\ell \in X^*$ and $\alpha \in \mathbb{R}$, is called a closed half-space in X.

Lemma 5.4.2 *Let G be a nonempty open convex set in X not containing* 0. *Then there is $\ell \in X^*$ such that $\ell(x) < 0$ for $x \in G$.*

Proof Put $E = \bigcup_{\lambda > 0} \lambda G$. Clearly E is a nonempty open convex cone not containing 0, and we infer from Corollary 5.4.1 by taking $C = \{0\}$ that there is $\ell \in X'$ such that $\ell(x) < 0$ for $x \in E$ (and hence for $x \in G$). Since G is disjoint with the hyperplane $H := \ker \ell$, H cannot be dense in X and therefore is closed. Consequently $\ell \in X^*$. ■

Theorem 5.4.5 *Any two nonempty disjoint open convex sets A and B in X can be separated strictly by a closed affine hyperplane.*

Proof Let $G = A - B$. G is a nonempty open convex set in X not containing 0; we infer then from Lemma 5.4.2 that there is $\ell \in X^*$ such that $\ell(x - y) < 0$ for $x \in A$ and $y \in B$, and hence $\ell(A)$ is bounded above and $\ell(B)$ is bounded below. Observe that $\ell(A)$ and $\ell(B)$ are open intervals. Let $a = \sup \ell(A)$ and $b = \inf \ell(B)$; then $a \leq b$. Choose $\alpha \in [a, b]$, then $f(x) < \alpha$ for $x \in A$ and $\ell(y) > \alpha$ for $y \in B$. Thus A and B are separated strictly by the closed affine hyperplane $\{x \in X : \ell(x) = \alpha\}$. ■

Theorem 5.4.6 *Suppose that A and B are disjoint closed convex sets in X, one of which is compact. Then there is a closed affine hyperplane which separates A and B strictly in the strong sense.*

Proof We may assume that B is compact and let $G = X\backslash A$. Then G is an open set containing B. For $x \in B$, choose $r_x > 0$ such that $x + B_{r_x}(0) \subset G$. The family $\{x + B_{\frac{1}{2}r_x}(0)\}_{x\in B}$ is an open covering of B, hence there are $x_1, \ldots, x_k$ in B such that $B \subset \bigcup_{j=1}^k \{x_j + B_{\frac{1}{2}r_{x_j}}(0)\}$. Let $r = \min_{1\leq j\leq k} \frac{1}{2} r_{x_j} > 0$, then $B + B_r(0) \subset G$. Therefore $\{B + B_r(0)\} \cap A = \emptyset$, and consequently

$$\left\{B + B_{\frac{1}{2}r}(0)\right\} \cap \left\{A + B_{\frac{1}{2}r}(0)\right\} = \emptyset.$$

We infer then from Theorem 5.4.5 that there are $\ell \in X^*$ and $\alpha \in \mathbb{R}$, such that

$$\ell(x + z) < \alpha,\ x \in A,\ z \in B_{\frac{1}{2}r}(0);$$
$$\ell(y + z) > \alpha,\ y \in B,\ z \in B_{\frac{1}{2}r}(0).$$

Now put $\varepsilon = \sup\{|\ell(z)| : z \in B_{\frac{1}{2}r}(0)\}$. Then, by choosing sequences $\{z'_k\}$ and $\{z''_k\}$ in $B_{\frac{1}{2}r}(0)$ such that $\ell(z'_k) \to \varepsilon$ and $\ell(z''_k) \to -\varepsilon$, we conclude from $\ell(x) < \alpha - \ell(z'_k)$ for $x \in A$ by letting $k \to \infty$ that $\ell(x) \leq \alpha - \varepsilon$; and conclude from $\ell(y) \geq \alpha - \ell(z''_k)$ for $y \in B$ by letting $k \to \infty$ that $\ell(y) \geq \alpha + \varepsilon$. ■

Exercise 5.4.13 Show that a set K in a real n.v.s. X is closed convex if and only if K is the intersection of a family of closed half-spaces in X.

Remark Since a complex vector space is also a real vector space, sublinear functionals are also defined on complex vector spaces. This fact is often used without being noted explicitly.

5.5 Complex form of Hahn–Banach theorem

Let X be a vector space. A **semi-norm** on X is a sublinear functional q on X such that $q(\alpha x) = |\alpha| q(x)$ for $x \in X$ and for scalar α (cf. Remark at the end of Section 5.4). Note that a semi-norm is nonnegative, because if $q(x) < 0$ for some x, then $0 = q(0) = q(x + (-x)) \leq q(x) + q(-x) = 2q(x) < 0$, which is absurd.

Theorem 5.5.1 *Let X be a vector space and q a semi-norm on X. Suppose that ℓ is a linear functional on a vector subspace Y of X such that $|\ell| \leq q$ on Y, then there is $\hat{\ell} \in X'$ with $|\hat{\ell}| \leq q$ on X such that $\hat{\ell}(y) = \ell(y)$ for $y \in Y$.*

Proof If X is a real vector space, then the theorem is a consequence of Theorem 5.4.3, as is easily verified. So we assume that X is a complex vector space. Write

$$\ell(y) = \ell_1(y) + i\ell_2(y), \quad y \in Y,$$

where $\ell_1(y) = \operatorname{Re} \ell(y)$ and $\ell_2(y) = \operatorname{Im} \ell(g)$. Then ℓ_1 and ℓ_2 are real linear functionals on Y. Since $i\ell(y) = \ell(iy)$, it follows that $\ell_2(y) = -\ell_1(iy)$, i.e.

$$\ell(y) = \ell_1(y) - i\ell_1(iy), \quad y \in Y.$$

Obviously, $|\ell_1| \leq q$ on Y. Hence there is a real linear functional $\hat{\ell}_1$ on X extending ℓ_1 such that $|\hat{\ell}_1(x)| \leq q(x)$ for $x \in X$.

Define $\hat{\ell}$ on X by

$$\hat{\ell}(x) = \hat{\ell}_1(x) - i\hat{\ell}_1(ix), \quad x \in X.$$

One can see that $\hat{\ell}$ is a linear functional on X and $\hat{\ell}$ extends ℓ. It remains only to show that $|\hat{\ell}(x)| \leq q(x)$ for $x \in X$. For any $x \in X$, there is $\beta \in \mathbb{C}$ with $|\beta| = 1$ such that $|\hat{\ell}(x)| = \beta\hat{\ell}(x)$, then,

$$|\hat{\ell}(x)| = \beta\hat{\ell}(x) = \hat{\ell}(\beta x) = \hat{\ell}_1(\beta x) - i\hat{\ell}_1(i\beta x)$$

$$= \hat{\ell}_1(\beta x) \leq q(\beta x) = |\beta| q(x) = q(x).$$

■

Some relevant consequences of Theorem 5.5.1 are now considered.

Corollary 5.5.1 *Let X be a normed vector space, then for any $x_0 \in X$, there is $\ell \in X^*$, with $\|\ell\| = 1$ such that $\ell(x_0) = \|x_0\|$.*

Proof Suppose first that $x_0 \neq 0$, and let $Y = \langle\{x_0\}\rangle$ be the vector subspace of X spanned by $\{x_0\}$. Define a linear functional ℓ_1 on Y by

$$\ell_1(\alpha x_0) = \alpha\|x_0\|,$$

then $|\ell_1(\alpha x_0)| = \|\alpha x_0\|$, implying $\|\ell_1\|_{Y^*} = 1$. By Theorem 5.5.1 with q being the norm on X, there is $\ell \in X'$ extending ℓ_1 such that $|\ell(x)| \leq \|x\|$. Then, $\ell(x_0) = \ell_1(x_0) = \|x_0\|$ and $\|\ell\| = 1$.

Now if $x_0 = 0$, simply take ℓ to be any $\ell \in X'$ with $\|\ell\| = 1$ (note that the first part of the proof shows that there is $\ell \in X'$ with $\|\ell\| = 1$). ■

Corollary 5.5.2 *Let X be any normed vector space. Then for any x and y in X, $x \neq y$, there is $\ell \in X^*$ such that $\ell(x) \neq \ell(y)$. i.e. X^* **separates points** of X.*

Proof Let $x_0 = x - y$. By Corollary 5.5.1, there is $\ell \in X^*$ with $\|\ell\| = 1$ such that $\ell(x_0) = \|x_0\| = \|x - y\|$. But,

$$|\ell(x) - \ell(y)| = |\ell(x - y)| = |\ell(x_0)| = \|x_0\| > 0.$$ ■

Exercise 5.5.1 Show that if $x_0 \in X$ and $x_0 \neq 0$, then there is $\ell \in X^*$ with $\|\ell\| = \|x_0\|$ and $\ell(x_0) = \|x_0\|^2$.

Exercise 5.5.2 Let $X = L^1[0,1]$ and $Y = C[0,1]$. Choose $x_0 \in (0,1)$ and let $\ell(f) = f(x_0)$ for $f \in Y$. Is it possible to extend ℓ to a bounded linear functional on X?

For a normed vector space X, define a function $\langle\cdot,\cdot\rangle$ on $X \times X^*$ by

$$\langle x, x^*\rangle = x^*(x), \quad (x, x^*) \in X \times X^*,$$

$\langle\cdot,\cdot\rangle$ is called the natural pairing between X and X^*.

For $x \in X$, let $j(x) \in X^{**} := (X^*)^*$ be defined by

$$\langle x^*, j(x)\rangle = \langle x, x^*\rangle, \quad x^* \in X^*.$$

The mapping j is a linear map from X into X^{**}, and since X^* separates points of X, it is one-to-one; furthermore it is an **isometry** in the sense that $\|j(x)\| = \|x\|$ for all $x \in X$.

Theorem 5.5.2 *The mapping j is a linear isometry from X into X^{**}.*

Proof It is left only to show that $\|j(x)\| = \|x\|$, where $\|j(x)\|$ is the norm of $j(x)$ in X^{**}. From

$$\begin{aligned} \|j(x)\| &= \sup_{\substack{x^* \in X^* \\ \|x^*\|=1}} |\langle x^*, j(x)\rangle| = \sup_{\substack{x^* \in X^* \\ \|x^*\|=1}} |\langle x, x^*\rangle| \\ &\leq \sup_{\substack{x^* \in X^* \\ \|x^*\|=1}} \|x\|\,\|x^*\| = \|x\|, \end{aligned}$$

it follows that $\|j(x)\| \leq \|x\|$. On the other hand, by Corollary 5.5.1, there is $x^* \in X^*$ with $\|x^*\| = 1$ such that $\langle x, x^*\rangle = \|x\|$, hence $\|j(x)\| \geq \|x\|$. Thus, $\|j(x)\| = \|x\|$. ■

Because of Theorem 5.5.2 we shall consider X as embedded in X^{**} as a normed vector subspace through the mapping j. If $X = X^{**}$, then X is called a **reflexive** space. A reflexive normed vector space is necessarily a Banach space. In general, the closure of X in X^{**} is a Banach space, which is called the **completion** of X. Note that if x is in the completion of a n.v.s. X, then there is a Cauchy sequence $\{x_n\}$ in $X \subset X^{**}$ such that $x_n \to x$ in X^{**}.

Example 5.5.1 Let $X = L^\infty[-1,1]$ and $Y = C[-1,1]$, and let $\delta \in Y^*$ be defined by $\delta(f) = f(0)$ for $f \in Y$. Since δ is a bounded linear functional with norm 1 on Y, it can be extended to be a bounded linear functional on X with the same norm by the Hahn–Banach theorem; we also denote the extended functional by δ, i.e., $\delta \in L^\infty[-1,1]^*$. It will be shown in Chapter 6 that $L^1[-1,1]^* = L^\infty[-1,1]$, in the sense that for $\ell \in L^1[-1,1]^*$ there is $h \in L^\infty[-1,1]$ such that $\ell(f) = \int_{[-1,1]} fh d\lambda$ for all $f \in L^1[-1,1]$. We know from this fact that $\delta \in L^1[-1,1]^{**}$. But there is no $h \in L^1[-1,1]$ such that $\delta(f) = \int_{[-1,1]} fh d\lambda = f(0)$ for $f \in C[-1,1]$; this means that $\delta \notin L^1[-1,1]$, i.e., $L^1[-1,1] \subsetneq L^1[-1,1]^{**}$.

Exercise 5.5.3 Suppose that Y is a vector subspace of a n.v.s. X such that $\overline{Y} \neq X$, and let $Y^\perp = \{x^* \in X^* : \langle y, x^* \rangle = 0 \text{ for all } y \in Y\}$.

(i) For $x \in X \backslash \overline{Y}$, show that there is $x^* \in Y^\perp$ such that $\|x^*\| = 1$ and $\langle x, x^* \rangle = \inf_{y \in Y} \|x - y\|$. (Hint: define $l \in (\langle \{x\} \rangle + Y)^*$ by $l(\alpha x + y) = \alpha \inf_{y \in Y} \|x - y\|$ for scalar α and $y \in Y$, then extend l to be defined on X by the Hahn–Banach theorem.)

(ii) For $x \in X$, show that

$$\inf_{y \in Y} \|x - y\| = \max_{\substack{x^* \in Y^\perp \\ \|x^*\| \leq 1}} |\langle x, x^* \rangle| = \max_{\substack{x^* \in Y^\perp \\ \|x^*\| = 1}} |\langle x, x^* \rangle|.$$

Exercise 5.5.4 Let F be a closed vector subspace in a real n.v.s. X and let τ be the canonical map from X onto X/F.

(i) Suppose now that C is an open convex set with $C \cap F = \emptyset$. Show that $\tau(C)$ is an open convex set in X/F, not containing $[0]$.

(ii) Suppose that Y is a vector subspace of X and C an open convex set in X, such that $C \cap Y = \emptyset$; show that there is a closed hyperplane H such that $H \supset Y$ and $H \cap C = \emptyset$. (Hint: use Theorem 5.4.1 in $X/\overline{Y}$ and note that a hyperplane in a n.v.s. X is either closed or dense in X.)

5.6 Hilbert space

Let E be a vector space. For definiteness, it will be assumed that E is a complex space throughout this section. The case of E being a real vector space can be treated similarly.

E is called an **inner product space** if there is a map $(\cdot,\cdot) : E \times E \to \mathbb{C}$ satisfying the following conditions:

(i) $(x, x) \geq 0 \; \forall x \in E$, and $(x, x) = 0$ if and only if $x = 0$;

(ii) $(\cdot, x)$ is linear on E for each $x \in E$; and

(iii) $(x, y) = \overline{(y, x)}$ for all x, y in E (for $z \in \mathbb{C}$, $\bar{z}$ is the conjugate of z).

The map $(\cdot, \cdot)$ is called an **inner product** on E. We always consider a vector subspace F of an inner product space E as an inner product space, with the inner product inherited from that on E, i.e. the inner product on F is the restriction to $F \times F$ of that on E. Note that when E is a real vector space, condition (iii) is replaced by $(x, y) = (y, x)$. If E is an inner product space, put $\|x\| = (x, x)^{1/2}$ for $x \in E$.

Theorem 5.6.1 *If E is an inner product space, then for x, y in E, the following hold:*

(a) $\|x - y\|^2 + \|x + y\|^2 = 2(\|x\|^2 + \|y\|^2)$ (***Parallelogram identity***);

(b) $|(x, y)| \leq \|x\| \cdot \|y\|$ (***Schwarz inequality***); *and*

(c) $\|x + y\| \leq \|x\| + \|y\|$ (***Triangle inequality***).

Proof For x and y in E,

$$\|x - y\|^2 = (x - y, x - y) = \|x\|^2 - 2\,\mathrm{Re}(x, y) + \|y\|^2; \tag{5.4}$$

$$\|x + y\|^2 = (x + y, x + y) = \|x\|^2 + 2\,\mathrm{Re}(x, y) + \|y\|^2. \tag{5.5}$$

(a) follows by adding (5.4) and (5.5).

To show (b), it is sufficient to show that $|(x, y)| \leq 1$ whenever $\|x\| = \|y\| = 1$. Now if $\|x\| = \|y\| = 1$, $|\mathrm{Re}(x, y)| \leq 1$ follows from (5.4) or (5.5) according to whether $\mathrm{Re}(x, y) \geq 0$ or $R(x, y) < 0$, because the far left sides of (5.4) and (5.5) are both greater than or equal to zero. If $\theta \in \mathbb{C}$ with $|\theta| = 1$ is chosen so that $(x, \theta y) = |(x, y)|$, then

$$|(x, y)| = (x, \theta y) = \mathrm{Re}(x, \theta y) \leq 1,$$

this concludes (b). Finally,

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

and thus,

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

From Theorem 5.6.1 (c), E is a normed vector space if the norm $\|x\|$ of x in E is defined by $\|x\| = (x, x)^{1/2}$. For an inner product space, the norm so defined is called the **norm associated with its inner product**. Unless stated otherwise, for an inner product space the norm associated with its inner product is always chosen as its norm.

An inner product space E is called a **Hilbert space** if it is complete when considered as a normed vector space. Obviously, a closed vector subspace of a Hilbert space is a Hilbert space.

The most important class of Hilbert spaces is the class of all $L^2(\Omega, \Sigma, \mu)$ with inner product (f, g), defined by $\int_\Omega f\bar{g}d\mu$ for f, g in $L^2(\Omega, \Sigma, \mu)$. The norm associated with this inner product is the L^2-norm. The space $\mathbb{C}^n$ with inner product $(z, w) = \sum_{k=1}^n z_k\bar{w}_k$ for $z = (z_1, \ldots, z_n)$ and $w = (w_1, \ldots, w_n)$ is a particular case; the norm associated with this inner product is the norm introduced for $\mathbb{C}^n$ in Section 1.4, hence $\mathbb{C}^n$ with this inner product is called the n-dimensional unitary space. Correspondingly, the Euclidean norm of $\mathbb{R}^n$ is associated with the inner product $(x, y) = \sum_{k=1}^n x_k y_k$ for $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$.

Suppose that E is a finite-dimensional vector space of dimension n and let $b_1, \ldots, b_n$ form a basis of E. For $x = \sum_{j=1}^n x_j b_j$, $y = \sum_{j=1}^n y_j b_j$ in E, where the x_j's and y_j's are scalars, define $(x, y) = \sum_{j=1}^n x_j\bar{y}_j$. E is clearly a Hilbert space with inner product so defined. Then it follows from Proposition 1.7.2 that every finite-dimensional inner product space is a Hilbert space.

An example of infinite-dimensional Hilbert space is the real space $\ell^2(\mathbb{Z})$ considered in Section 1.6 whose norm is associated with the inner product $(x, y) = \sum_{k\in\mathbb{Z}} x_k y_k$ for $x = (x_k)$ and $y = (y_k)$. We shall also use $\ell^2(\mathbb{Z})$ to denote the complex Hilbert space of all those complex sequences $(z_k)_{k\in\mathbb{Z}}$ such that $\sum_{k\in\mathbb{Z}} |z_k|^2 < \infty$, and with inner product $(z, w) := \sum_{k\in\mathbb{Z}} z_k\bar{w}_k$ for $z = (z_k)$ and $w = (w_k)$. Whether $\ell^2(\mathbb{Z})$ is a complex or real space will either be stated explicitly or occasioned by context.

As inner product on an inner product space is a generalization of the scalar product for vectors in three-dimensional Euclidean space in which two nonzero vectors are perpendicular to each other if and only if their scalar product is zero. Therefore, two elements x and y in an inner product space E are said to be **orthogonal** if $(x, y) = 0$, and, for a nonempty subset A of E, call the set $A^\perp := \{x \in E : (x, y) = 0\ \forall y \in A\}$, the **orthogonal complement** of A in E. Obviously, $A^\perp$ is a closed vector subspace of E.

Exercise 5.6.1 Let M be a vector subspace of an inner product space E; show that $M \cap M^\perp = \{0\}$. Also show that if an element x of E can be expressed as the sum $x = y + z$ of an element y in M and an element z in $M^\perp$, then such an expression is unique.

Theorem 5.6.2 (Orthogonal projection theorem) *Suppose that E is a Hilbert space and M a closed vector subspace of E. Then for any $x \in E$, there is a unique element $y \in M$ such that*

$$\|x - y\| = \min_{z\in M} \|x - z\|. \tag{5.6}$$

Furthermore, y is characterized by

$$x - y \in M^\perp. \tag{5.7}$$

Proof Let $\alpha = \inf_{z\in M} \|x - z\|$. There is a sequence $\{y_n\}$ in M such that

$$\alpha^2 \le \|x - y_n\|^2 \le \alpha^2 + \frac{1}{n}, \quad n = 1, 2, \ldots.$$

By parallelogram identity,

$$\|(y_n - x) - (y_m - x)\|^2 + \|(y_n - x) + (y_m - x)\|^2$$
$$= 2(\|y_n - x\|^2 + \|y_m - x\|^2) \leq 4\alpha^2 + \frac{2}{n} + \frac{2}{m},$$

or

$$\|y_n - y_m\|^2 \leq 4\alpha^2 + \frac{2}{n} + \frac{2}{m} - 4\left\|\frac{y_n + y_m}{2} - x\right\|^2 \leq 2\left(\frac{1}{n} + \frac{1}{m}\right),$$

from which it follows that $\{y_n\}$ is a Cauchy sequence in M. Since M is complete, there is $y \in M$ such that $\lim_{n\to\infty} \|y_n - y\| = 0$. Then,

$$\|x - y\|^2 = \lim_{n\to\infty} \|x - y_n\|^2 = \alpha^2,$$

i.e.

$$\|x - y\| = \alpha = \inf_{z\in M} \|x - z\| = \min_{z\in M} \|x - z\|.$$

We have shown that there is $y \in M$ such that

$$\|x - y\| = \min_{z\in M} \|x - z\|.$$

Now let y be any element of M which satisfies (5.6); then for $z \in M$ and $t \in \mathbb{R}$, we have

$$\|x - y - tz\|^2 = \|x - y\|^2 - 2\,\mathrm{Re}(x - y, z)t + t^2\|z\|^2,$$

or

$$0 \leq \|x - y - tz\|^2 - \|x - y\|^2 \leq -2\,\mathrm{Re}(x - y, z)t + t^2\|z\|^2.$$

Then for $t > 0$,

$$0 \leq -2\,\mathrm{Re}(x - y, z) + t\|z\|^2,$$

and hence,

$$\mathrm{Re}(x - y, z) \leq 0,$$

by letting $t \searrow 0$; while for $t < 0$,

$$0 \geq -2\,\mathrm{Re}(x - y, z) + t\|z\|^2$$

holds, and by letting $t \nearrow 0$, we have

$$\operatorname{Re}(x - y, z) \geq 0.$$

Hence,

$$\operatorname{Re}(x - y, z) = 0. \tag{5.8}$$

If we replace z in (5.8) by iz, then $\operatorname{Im}(x - y, z) = 0$. Thus $(x - y, z) = 0$, i.e. y satisfies (5.7). Suppose now that (5.6) holds for $y = y'$ and y'' in M, then $(x - y', y' - y'') = 0 = (x - y'', y' - y'') = 0$ by (5.7), and consequently,

$$(y' - y'', y' - y'') = (x - y'' + y' - x, y' - y'') = 0,$$

which implies that $\|y' - y''\| = 0$ or $y' = y''$. Hence, there is unique $y \in M$ that satisfies (5.6).

Finally, suppose $y \in M$ satisfies (5.7), then for $z \in M$,

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

or y satisfies (5.6). ■

The map that associates each $x \in X$ with the unique element y in M which satisfies (5.6) (or (5.7)) is called the **orthogonal projection** from X onto M. This map will be denoted by P_M.

Corollary 5.6.1 *Suppose that M is a closed vector subspace of a Hilbert space E; then every $x \in E$ can be expressed uniquely as $x = y + z$, where $y \in M$ and $z \in M^\perp$. In other words, $E = M \oplus M^\perp$.*

Proof For $x \in E$, let $y = P_M x$. Then $x - y \in M^\perp$, by (5.7), hence $x = y + (x - y) \equiv y + z$, where $y \in M$ and $z \in M^\perp$. The uniqueness of such an expression follows from Exercise 5.6.1. ■

Exercise 5.6.2 Let M be a closed vector subspace of a Hilbert space E.

(i) Show that P_M is linear and that the following properties hold:
(a) $P_M x = x$ if and only if $x \in M$; (b) $P_M^2 = P_M$; and (c) $\|P_M x\| \leq \|x\|$ for all $x \in E$.

(ii) Show that $1 - P_M = P_{M^\perp}$.

(iii) Show that $\|x\|^2 = \|P_M x\|^2 + \|P_{M^\perp} x\|^2$ for $x \in E$ (**Pythagoras relation**).

Theorem 5.6.3 (Riesz representation theorem) *If E is a Hilbert space, and $x^* \in E^*$, then there is a unique $y_0 \in E$ such that*

$$\langle x, x^* \rangle = (x, y_0), \quad x \in E.$$

Furthermore,

$$\|x^*\| = \|y_0\|,$$

and the map $x^* \to y_0$ *is conjugate linear (an operator* T *from a vector space into a vector space is conjugate linear if* $T(\alpha x + \beta y) = \bar{\alpha}Tx + \bar{\beta}Ty$ *for all* x, y *in* $D(T)$ *and all scalars* α *and* β*).*

Proof If $x^* = 0$, take $y_0 = 0$. Suppose now that $x^* \neq 0$ and let $M = \ker x^* := \{x \in E : \langle x, x^*\rangle = 0\}$. M is clearly a closed vector subspace of E. Since $x^* \neq 0$, there is $x_0 \in M^\perp$ such that $\langle x_0, x^*\rangle = 1$. Now let $x \in E$ and put $\lambda = \langle x, x^*\rangle$. By Corollary 5.6.1, $x = y + z$, where $y \in M$ and $z \in M^\perp$, hence, $\lambda = \langle x, x^*\rangle = \langle z, x^*\rangle = \langle \lambda x_0, x^*\rangle$, or $\langle z - \lambda x_0, x^*\rangle = 0$, which means that $z - \lambda x_0 \in M$. But $z - \lambda x_0$ is also in $M^\perp$, consequently $z = \lambda x_0$, by Exercise 5.6.1. Now, from $x = y + \lambda x_0$ we have $(x, x_0) = (y + \lambda x_0, x_0) = \lambda\|x_0\|^2 = \langle x, x^*\rangle\|x_0\|^2$. If we take $y_0 = \frac{x_0}{\|x_0\|^2}$, then $(x, y_0) = \langle x, x^*\rangle$ for $x \in E$. Suppose that $y_0' \in E$ also satisfies $\langle x, x^*\rangle = (x, y_0')$ for all $x \in E$, then $(y_0' - y_0, x) = 0$ for all x in E; in particular, $(y_0' - y_0, y_0' - y_0) = 0$ or $\|y_0' - y_0\| = 0$, implying $y_0' = y_0$. Hence, there is unique $y_0 \in E$ satisfying $\langle x, x^*\rangle = (x, y_0)$ for all x in E. From $\langle x, x^*\rangle = (x, y_0)$ it follows readily that $\|x^*\| \leq \|y_0\|$; but $\|y_0\|^2 = (y_0, y_0) = |\langle y_0, x^*\rangle| \leq \|y_0\| \cdot \|x^*\|$, hence, $\|y_0\| \leq \|x^*\|$. Thus $\|y_0\| = \|x^*\|$. That $x^* \to y_0$ is conjugate linear is obvious. ■

Exercise 5.6.3

(i) Denote by R the map $x^* \mapsto y_0$ in Theorem 5.6.3. Show that E^* is a Hilbert space with inner product $(\cdot,\cdot)_*$, defined by $(x^*, y^*)_* = (Ry^*, Rx^*)$ for x^*, y^* in E^*.

(ii) Show that Hilbert spaces are reflexive.

Example 5.6.1 Define on $C[0,1]$ an inner product by

$$(f, g) = \int_0^1 f(t)\overline{g(t)}dt, \quad f, g \in C[0,1].$$

We claim that $C[0,1]$ is not complete with the norm associated with this inner product. We denote this inner product space by $\hat{C}[0,1]$ in this example. Let f be the indicator function of $[\frac{1}{2}, 1]$ on $[0,1]$ and for each integer $n > 2$, let f_n be a continuous function such that $0 \leq f_n \leq 1$ and coincides with f on $[0, \frac{1}{2} - \frac{1}{n}] \cup [\frac{1}{2}, 1]$. Then $f_n \to f$ in $L^2[0,1]$, i.e., $\|f_n - f\|_2 \to 0$. Let g be any function in $\hat{C}[0,1]$, then $\|f_n - g\|_2 \geq \|f - g\|_2 - \|f_n - f\|_2$ and hence $\liminf_{n\to\infty} \|f_n - g\|_2 \geq \|f - g\|_2 > 0$. Thus $\{f_n\}$, which is a Cauchy sequence in $\hat{C}[0,1]$, does not converge in $\hat{C}[0,1]$.

The Riesz representation theorem for linear functionals on Hilbert spaces might lead to far reaching results, even when the spaces concerned are finite dimensional. We illustrate this fact by proving an interesting result of A.P. Calderón and A. Zygmund about Friederich mollifiers. Recall that from a real-valued C^∞ function φ on $\mathbb{R}^n$ with compact support in the unit closed ball $C_1(0)$ and with $\int \varphi d\lambda^n = 1$, one can construct a family

$\{J_\varepsilon\}_{\varepsilon>0}$ of operators on $L_{\text{loc}}(\mathbb{R}^n)$ in the following way (cf. Section 4.9). For $\varepsilon > 0$, let $\varphi_\varepsilon(x) = \varepsilon^{-n}(\frac{x}{\varepsilon})$ for $x \in \mathbb{R}^n$, then $\operatorname{supp}\varphi_\varepsilon \subset C_\varepsilon(0)$ and $\int \varphi_\varepsilon d\lambda^n = 1$. If $f \in L_{\text{loc}}(\mathbb{R}^n)$, define a function $J_\varepsilon f$ by

$$J_\varepsilon f(x) = \int_{\mathbb{R}^n} f(y)\varphi_\varepsilon(x-y)d\lambda^n(y), \quad x \in \mathbb{R}^n.$$

The family $\{J_\varepsilon\}_{\varepsilon>0}$ depends on φ and is called a Friederich mollifier.

Theorem 5.6.4 (Calderón–Zygmund) *For each $k \in \mathbb{N}$, there is a Friederichs mollifier $\{J_\varepsilon\}_{\varepsilon>0}$ such that $J_\varepsilon p = p$ for every polynomial p of degree $\leq k$ defined on $\mathbb{R}^n$.*

Proof Let E be the space of all real polynomials p of degree $\leq k$ on $\mathbb{R}^n$. E is a real vector space of finite dimension. Choose a nonnegative and nonzero C^∞ function η on $\mathbb{R}^n$ with $\operatorname{supp}\eta \subset C_1(0)$ and define an inner product $(\cdot,\cdot)$ on E by $(p,q) = \int_{\mathbb{R}^n} pq\eta d\lambda^n$ for p, q in E. Since $\dim E < \infty$, E is a Hilbert space. Let l be a linear functional on E defined by

$$l(p) = p(0), \quad p \in E.$$

Since $\dim E < \infty$, every linear functional on E is bounded. By Theorem 5.6.3, there is $q_0 \in E$ such that

$$p(0) = (p, q_0) = \int_{\mathbb{R}^n} pq_0\eta d\lambda^n.$$

If we choose p to be the constant polynomial 1 in the above equality, we have $\int_{\mathbb{R}^n} q_0\eta d\lambda^n = 1$. Let $\varphi = q_0\eta$ and $\{J_\varepsilon\}_{\varepsilon>0}$ the corresponding Friederich mollifier. Now for $p \in E$ and $x \in \mathbb{R}^n$,

$$\begin{aligned} J_\varepsilon p(x) &= \int_{\mathbb{R}^n} p(y)\varphi_\varepsilon(x-y)d\lambda^n(y) = \varepsilon^{-n}\int_{\mathbb{R}^n} p(y)\varphi\left(\frac{x-y}{\varepsilon}\right)d\lambda^n(y) \\ &= \int_{\mathbb{R}^n} p(x-\varepsilon y)\varphi(y)d\lambda^n(y) = \widehat{p_x}(0) = p(x), \end{aligned}$$

where $\widehat{p_x}(y) = p(x - \varepsilon y)$. ∎

Another remarkable application of the Riesz representation theorem will be presented in Section 5.7.

5.7 Lebesgue–Nikodym theorem

We consider in this section an interesting application of the Riesz representation theorem to measure theory.

Let (Ω, Σ) be a measurable space, and suppose that μ and ν are finite measures on Σ. The following theorem asserts that ν can be decomposed in a certain way relative to μ.

Theorem 5.7.1 (Lebesgue–Nikodym theorem) *Let (Ω, Σ) be a measurable space, and μ, ν finite measures on Σ. Then there is a unique $h \in L^1(\Omega, \Sigma, \mu)$ and a μ-null set N, such that*

$$\nu(A) = \int_A h d\mu + \nu(A \cap N), \ \ A \in \Sigma. \tag{5.9}$$

Proof Let $\rho = \mu + \nu$; then ρ is a finite measure on Σ. Consider the real Hilbert space $L^2(\Omega, \Sigma, \rho)$ and consider the linear functional ℓ on $L^2(\Omega, \Sigma, \rho)$, defined by

$$\ell(f) = \int f d\nu.$$

Since

$$\begin{aligned} |\ell(f)| &\leq \left(\int |f|^2 d\nu\right)^{1/2} \left(\int 1 d\nu\right)^{1/2} \leq \nu(\Omega)^{1/2} \left[\int |f|^2 d\rho\right]^{1/2} \\ &= \nu(\Omega)^{1/2} \|f\|_{L^2(\rho)}, \end{aligned}$$

ℓ is a bounded linear functional on $L^2(\Omega, \Sigma, \rho)$. By the Riesz representation theorem there is unique $g \in L^2(\Omega, \Sigma, \rho)$, such that

$$\int f d\nu = \int f g d\rho = \int f g d\mu + \int f g d\nu$$

for all $f \in L^2(\Omega, \Sigma, \rho)$, or

$$\int f(1-g) d\nu = \int f g d\mu \tag{5.10}$$

for all $f \in L^2(\Omega, \Sigma, \rho)$.

We claim first that there is a μ-null set N such that $0 \leq g(x) < 1$ for $x \in \Omega \backslash N$. Let $A_1 = \{x \in \Omega : g(x) < 0\}$ and $A_2 = \{x \in \Omega : g(x) \geq 1\}$. If we let $f = I_{A_1}$ in (5.10), then $0 \leq \nu(A_1) \leq \int_{A_1} (1-g) d\nu = \int_{A_1} g d\mu$, from which it follows that $\mu(A_1) = 0$. Next choose $f = I_{A_2}$ in (5.10); we have $0 \geq \int_{A_2} (1-g) d\nu = \int_{A_2} g d\mu \geq \mu(A_2)$. This implies that $\mu(A_2) = 0$. Put $N = A_1 \cup A_2$, then $\mu(N) = 0$ and $0 \leq g(x) < 1$ for $x \in \Omega \backslash N$.

We show next that (5.10) holds for every nonnegative measurable function f which vanishes on N. Suppose that f is such a function; for each positive integer n, let $f_n = f \wedge n$, i.e. $f_n(x) = f(x)$ if $f(x) \leq n$, otherwise $f_n(x) = n$. Since $1 - g > 0$ and $g \geq 0$ on $\Omega \backslash N$, $0 \leq f_n(1-g) \nearrow f(1-g)$, and $0 \leq f_n g \nearrow fg$, then from the monotone convergence theorem and the fact that (5.10) holds for each f_n, it follows that

$$\int f(1-g) d\nu = \lim_{n \to \infty} \int f_n (1-g) d\nu = \lim_{n \to \infty} \int f_n g d\mu = \int f g d\mu.$$

This shows that (5.10) holds for every such function. For $A \in \Sigma$, let $B = A \cap (\Omega \backslash N)$; then (5.10) holds for the function $f := I_B(1-g)^{-1}$ and we have $\int I_B d\nu = \int I_B \frac{g}{1-g} d\mu = \int_A I_{\Omega\backslash N} \cdot \frac{g}{1-g} d\mu$, or

$$\nu(A \cap (\Omega\backslash N)) = \int_A h d\mu$$

if we put $h = I_{\Omega\backslash N}\frac{g}{1-g}$. Note that $h \geq 0$, and, since $\int_\Omega h d\mu = \nu(\Omega\backslash N) < \infty$, $h \in L^1(\Omega, \Sigma, \mu)$. Now,

$$\nu(A) = \nu(A \cap (\Omega\backslash N)) + \nu(A \cap N) = \int_A h d\mu + \nu(A \cap N),$$

hence (5.9) holds. Now suppose that there is $h' \in L^1(\Omega, \Sigma, \mu)$ and μ-null set N', such that

$$\nu(A) = \int_A h' d\mu + \nu(A \cap N'), \quad A \in \Sigma;$$

if we put $\hat{N} = N \cup N'$, then $\int_{A\cap\hat{N}^c} h d\mu = \int_{A\cap\hat{N}^c} h' d\mu$ for all $A \in \Sigma$, and consequently $h = h'$ μ-a.e. on $\Omega\backslash\hat{N}$; but $\hat{N}$ being a μ-null set implies that $h = h'$ μ-a.e. on Ω. Thus h is unique. ■

Exercise 5.7.1 Show that Theorem 5.7.1 holds if both μ and ν are σ-finite. But in this case h may not be μ-integrable; however it is μ-integrable if ν is finite.

Measure ν is said to be **μ-absolutely continuous** on Σ, if $A \in \Sigma$ and $\mu(A) = 0$ results in $\nu(A) = 0$; while ν is **μ-singular** on Σ, if there is a μ-null set N such that $\nu(A) = (A \cap N)$ for all $A \in \Sigma$. Note that if we use μ^* and ν^* to denote the outer measures on Ω, constructed respectively from μ and ν on Σ by Method I, then the definitions given here for μ-absolute continuity and μ-singularity for ν as measure on Σ are the same as μ^*-absolute continuity and μ^*-singularity for ν^*, introduced in Section 4.6.

Corollary 5.7.1 (Radon–Nikodym) *If μ and ν are σ-finite measures on Σ and ν is μ-absolutely continuous, then there is a unique nonnegative measurable function h on Ω such that*

$$\nu(A) = \int_A h d\mu, \quad A \in \Sigma.$$

Proof We know that Theorem 5.7.1 also holds true if μ and ν are σ-finite (cf. Exercise 5.7.1). We may then apply (5.9). Since $\mu(A \cap N) = 0$ implies that $\nu(A \cap N) = 0$ for all $A \in \Sigma$ by the μ-absolute continuity of ν, the corollary follows. ■

Remark The function h in Corollary 5.7.1 is called the **Radon–Nikodym derivative** of ν w.r.t. μ, and the conclusion of the corollary is usually referred to as the Radon–Nikodym theorem and is expressed by $d\nu = h d\mu$ or $h = \frac{d\nu}{d\mu}$.

5.8 Orthonormal families and separability

Hilbert spaces considered in this section are assumed to be of infinite dimension. The finite-dimensional case can be treated similarly, but in a simpler fashion.

A family $\{e_\alpha\}_{\alpha\in I}$ of elements in a Hilbert space E is said to be **orthonormal** if $(e_\alpha, e_\beta) = \delta_{\alpha\beta} := \begin{cases} 0 & \text{if } \alpha \neq \beta \\ 1 & \text{if } \alpha = \beta \end{cases}$. It is clear that an orthonormal family is linearly independent.

Consider first a finite orthonormal family $\{e_j\}_{j=1}^n$ and let $E_n = \langle\{e_1, \dots, e_n\}\rangle$. Then E_n is a closed vector subspace of E, by Corollary 1.7.1.

Lemma 5.8.1 *Let P_n denote the orthogonal projection from E onto E_n; then $P_n x = \sum_{j=1}^n (x, e_j)e_j$ for $x \in E$.*

Proof It is clear that $P_n x = \sum_{j=1}^n (P_n x, e_j)e_j$. For each $j = 1, \dots, n$, we have $(x - P_n x, e_j) = 0$, by (5.7), hence $(P_n x, e_j) = (x, e_j)$. ■

Exercise 5.8.1 Suppose that $\{e_\alpha\}_{\alpha\in I}$ is an orthonormal family in a Hilbert space E. Show that for any $x \in E$, $\{|(x, e_\alpha)|^2\}_{\alpha\in I}$ is summable and $\sum_{\alpha\in I} |(x, e_\alpha)|^2 \le \|x\|^2$.

Now let $\{e_k\}_{k=1}^\infty$ be an orthonormal family in E. For each $n \in \mathbb{N}$, put $E_n = \langle\{e_1, \dots, e_n\}\rangle$ and let E_∞ be the closure of $\langle\{e_k\}_{k=1}^\infty\rangle$, i.e. E_∞ is the smallest closed vector subspace containing $\{e_k\}_{k=1}^\infty$.

Theorem 5.8.1 *For $x \in E_\infty$, we have*

(i) $x = \sum_{k=1}^\infty (x, e_k)e_k$, *i.e.* $\lim_{n\to\infty} \|x - \sum_{k=1}^n (x, e_k)e_k\| = 0$.

(ii) $\|x\|^2 = \sum_{k=1}^\infty |(x, e_k)|^2$.

Proof

(i): Given that $\varepsilon > 0$, there is $y \in \langle\{e_k\}_{k=1}^\infty\rangle$ such that $\|x - y\|^2 < \varepsilon$. Now, $y = \sum_{k=1}^m \alpha_k e_k$, $\alpha_k \in \mathbb{C}$, $k = 1, \dots, m$, hence, $y \in E_m \subset E_n$ for $n \ge m$. Thus if $n \ge m$, we have

$$\|x - P_n x\|^2 \le \|x - y\|^2 < \varepsilon,$$

or, by Lemma 5.8.1,

$$\left\| x - \sum_{k=1}^n (x, e_k)e_k \right\|^2 < \varepsilon$$

if $n \ge m$. This proves (i).

(ii): From (i),

$$\|x\|^2 = \lim_{n\to\infty} \left\| \sum_{k=1}^n (x, e_k)e_k \right\|^2.$$

But,

$$\left\|\sum_{k=1}^{n}(x,e_k)e_k\right\|^2 = \left(\sum_{j=1}^{n}(x,e_j)e_j, \sum_{k=1}^{n}(x,e_k)e_k\right)$$
$$= \sum_{j,k=1}^{n}(x,e_j)\overline{(x,e_k)}(e_j,e_k) = \sum_{k=1}^{n}|(x,e_k)|^2,$$

hence $\|x\|^2 = \sum_{k=1}^{\infty}|(x,e_k)|^2$. ■

Corollary 5.8.1 (Bessel inequality) *For $x \in E$, $\sum_{k=1}^{\infty}|(x,e_k)|^2 \le \|x\|^2$, and the equality holds if and only if $x \in E_\infty$.*

Proof Let P be the orthogonal projection from E onto E_∞, then $\|x\|^2 = \|Px\|^2 + \|x - Px\|^2$, by Exercise 5.6.1. Hence $\|Px\|^2 \le \|x\|^2$. But by Theorem 5.8.1,

$$\|Px\|^2 = \sum_{k=1}^{\infty}|(Px,e_k)|^2 = \sum_{k=1}^{\infty}|(x,e_k)|^2,$$

because $(x - Px, e_k) = 0$ for each k by (5.7). Hence,

$$\|x\|^2 = \|x - Px\|^2 + \sum_{k=1}^{\infty}|(x,e_k)|^2,$$

from which it follows that $\sum_{k=1}^{\infty}|(x,e_k)|^2 \le \|x\|^2$, and that equality holds if and only if $x = Px$ or $x \in E_\infty$. ■

Exercise 5.8.2

(i) Show that for x, y in E_∞ we have

$$(x,y) = \sum_{k=1}^{\infty}(x,e_k)\overline{(y,e_k)}.$$

(ii) Show that $E = E_\infty$ if and only if $\|x\|^2 = \sum_{k=1}^{\infty}|(x,e_k)|^2$ for all $x \in E$.

(iii) Show that $E = E_\infty$ if and only if

$$x = \sum_{k=1}^{\infty}(x,e_k)e_k$$

for all $x \in E$.

Theorem 5.8.2 (Riesz–Fischer) *Let $\{e_k\}_{k\in\mathbb{N}}$ be an orthonormal family in E and $\{\alpha_k\}_{k\in\mathbb{N}}$ a sequence of scalars, then there is $x \in E$ such that $x = \sum_{k=1}^{\infty}\alpha_k e_k$ if and only if $\sum_k |\alpha_k|^2 < \infty$.*

Proof Suppose that $\sum_{k=1}^{\infty} |\alpha_k|^2 < \infty$. For each $n \in \mathbb{N}$ let $x_n = \sum_{k=1}^{n} \alpha_k e_k$. We claim that $\{x_n\}$ is a Cauchy sequence in E. Actually, for $n > m$ in $\mathbb{N}$,

$$\|x_n - x_m\|^2 = \left(\sum_{k=m+1}^{n} \alpha_k e_k, \sum_{j=m+1}^{n} \alpha_j e_j \right) = \sum_{k,j=m+1}^{n} \alpha_k \bar{e}_j (e_k, e_j)$$
$$= \sum_{k=m+1}^{n} |\alpha_k|^2 \to 0$$

as $n > m \to \infty$, so $\{x_n\}$ is a Cauchy sequence, and there is $x \in E$ such that $x = \lim_{n\to\infty} x_n$, or $x = \lim_{n\to\infty} \sum_{k=1}^{n} \alpha_k e_k = \sum_{k=1}^{\infty} \alpha_k e_k$.

Next, suppose that $x = \sum_{k=1}^{\infty} \alpha_k e_k$. This means that $x = \lim_{n\to\infty} \sum_{k=1}^{n} \alpha_k e_k$; but each $\sum_{k=1}^{n} \alpha_k e_k$ is in E_n, and hence $x \in E_\infty$. Now for each $j \in \mathbb{N}$,

$$(x, e_j) = \lim_{n\to\infty} \left(\sum_{k=1}^{n} \alpha_k e_k, e_j \right) = \alpha_j;$$

consequently,

$$\sum_{j=1}^{\infty} |\alpha_j|^2 = \sum_{j=1}^{\infty} |(x, e_j)|^2 = \|x\|^2 < \infty,$$

by Theorem 5.8.1 (ii). ■

An orthonormal family $\{e_k\}_{k=1}^{\infty}$ is called an **orthonormal basis** for E if

$$x = \sum_{k=1}^{\infty} (x, e_k) e_k$$

for all $x \in E$.

Theorem 5.8.3 *Let $\{e_k\}_{k=1}^{\infty}$ be an orthonormal family in a Hilbert space E and define E_∞ as before.*

(i) *$\{e_k\}_{k=1}^{\infty}$ is an orthonormal basis for E if and only if $E = E_\infty$.*

(ii) *$\{e_k\}_{k=1}^{\infty}$ is an orthonormal basis for E if and only if for $x \in E$, $x = 0$ whenever $(x, e_k) = 0$ for all k.*

Proof It is clear that (i) follows from Theorem 5.8.1 (i), and the fact that if $\{e_k\}_{k=1}^{\infty}$ is an orthonormal basis, then $E = E_\infty$. For the proof of (ii), in view of (i) one need only observe that for $x \in E$, $(x - Px, e_k) = 0$ for all k, where P is the orthonormal projection from E onto E_∞. ■

Exercise 5.8.3 Show that an orthonormal family $\{e_k\}_{k\in\mathbb{N}}$ in E is an orthonormal basis for E if and only if $\|x\|^2 = \sum_{k=1}^{\infty} |(x, e_k)|^2$ for all $x \in E$.

Example 5.8.1 (Hermite polynomials and Hermite functions) For nonnegative integer n and $x \in \mathbb{R}$, let

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2};$$

then $H_n(x)$ is a polynomial in x of degree n with the coefficient of x^n being 2^n. The polynomials $H_n(x)$ are called **Hermite polynomials** and the functions $\psi_n(x) = e^{-\frac{x^2}{2}} H_n(x)$ are called **Hermite functions**. We have, for nonnegative integers m and n,

$$\begin{aligned}\int_{-\infty}^{\infty} \psi_n(x)\psi_m(x)dx &= \int_{-\infty}^{\infty} e^{-x^2} H_n(x)H_m(x)dx \\ &= \int_{-\infty}^{\infty} H_m(x)(-1)^n \frac{d^n}{dx^n} e^{-x^2} dx,\end{aligned}$$

from which we conclude by repeated integration by parts that

$$\begin{aligned}\int_{-\infty}^{\infty} \psi_n(x)\psi_m(x)dx &= \int_{-\infty}^{\infty} e^{-x^2} \frac{d^n}{dx^n} H_m(x)dx \\ &= \begin{cases} 0 & \text{if } m < n; \\ 2^n n! \sqrt{\pi} & \text{if } m = n. \end{cases}\end{aligned}$$

Thus $\{\psi_0, \psi_1, \psi_2, \ldots\}$ is an orthogonal family in $L^2(\mathbb{R})$. If we define **the normalized Hermite functions** $\mathcal{E}_n$ by

$$\mathcal{E}_n(x) = (2^n n! \sqrt{\pi})^{-\frac{1}{2}} \psi_n(x),$$

then $\{\mathcal{E}_0, \mathcal{E}_1, \mathcal{E}_2, \ldots\}$ is an orthonormal family in $L^2(\mathbb{R})$. Observe that $\mathcal{E}_n(x) = e^{-\frac{x^2}{2}} h_n(x)$, where $h_n(x) = (2^n n! \sqrt{\pi})^{-\frac{1}{2}} H_n(x)$; the polynomials $h_0(x)$, $h_1(x)$, $h_2(x)$, ... are called **normalized Hermite polynomials**. Observe that since $h_n(x)$ is a polynomial of degree n, each monomial x^n is a linear combination of $h_0(x), \ldots, h_n(x)$. Let us now put $w(x) = e^{-x^2}$ and denote by $L^2_w(\mathbb{R})$ the space $L^2(\mathbb{R}, \mathcal{L}, \mu)$, where $\mu(A) = \int_A w d\lambda = \int_A e^{-x^2} dx$ for $A \in \mathcal{L}$. The space $L^2_w(\mathbb{R})$ is called the **weighted $\mathbf{L^2}$ space** on $\mathbb{R}$ with weight w. Then, Hermite polynomials form an orthogonal family in $L^2_w(\mathbb{R})$ and normalized Hermite polynomials form an orthonormal family in $L^2_w(\mathbb{R})$. We shall see in Chapter 7 that $\{\mathcal{E}_0, \mathcal{E}_1, \mathcal{E}_2, \ldots\}$ is an orthonormal basis for $L^2(\mathbb{R})$, or equivalently, $\{h_0, h_1, h_2, \ldots\}$ is an orthonormal basis for $L^2_w(\mathbb{R})$ (cf. Corollary 7.1.1).

A procedure, the **Gram–Schmidt process**, for orthonormalizing a given countable linearly independent family $\{u_k\}$ in E is now introduced. Let $e_1 = \frac{u_1}{\|u_1\|}$. Suppose now that $e_1, \ldots, e_n$ have been defined so that they form an orthonormal family and $\langle\{e_1, \ldots, e_n\}\rangle = \langle\{u_1, \ldots, u_n\}\rangle$; put $E_n = \langle\{e_1, \ldots, e_n\}\rangle$ and let z_n be the image of u_{n+1} in E_n under the orthogonal projection from E onto E_n. Since u_{n+1} is not in $\langle\{u_1, \ldots, u_n\}\rangle$, it is not in E_n and hence $u_{n+1} - z_n \neq 0$. Define $e_{n+1} = \frac{u_{n+1} - z_n}{\|u_{n+1} - z_n\|}$, then $\|e_{n+1}\| = 1$ and $e_{n+1} \in E_n^{\perp}$.

Thus $e_1,\dots,e_{n+1}$ form an orthonormal family; it is readily seen that $\langle\{u_1,\dots,u_{n+1}\}\rangle = \langle\{e_1,\dots,e_{n+1}\}\rangle$. We have therefore defined, by induction, an orthonormal family $\{e_k\}$ from $\{u_k\}$ such that $\langle\{e_1,\dots,e_n\}\rangle = \langle\{u_1,\dots,u_n\}\rangle$ for all $n\in\mathbb{N}$.

Theorem 5.8.4 *A Hilbert space E has an orthonormal basis if and only if E is separable.*

Proof If E has an orthonormal basis $\{e_k\}$, then the countable set $\bigcup_{n=1}^{\infty}\{\sum_{j=1}^{n}\alpha_j e_j : \alpha_j\in\gamma,\ j=1,\dots,n\}$ is dense in E; hence E is separable. We have denoted by γ the countable set of rational complex numbers.

If now E is separable, say $\{x_n\}_{n=1}^{\infty}$ is dense in E. We may assume that $x_1\neq 0$. By an obvious selection procedure, we can select a linearly independent subsequence $\{x_{n_k}\}$ of $\{x_n\}$ such that $\langle\{x_{n_k}\}\rangle = \langle\{x_n\}\rangle$. Put $x_{n_k}=y_k$. Let $\{e_k\}$ be the orthonormal family obtained from $\{y_k\}$ by the Gram–Schimdt procedure, then $\{e_k\}$ is an orthonormal family such that $\langle\{e_k\}\rangle = \langle\{y_k\}\rangle = \langle\{x_k\}\rangle$. Consequently the closure of $\langle\{e_k\}\rangle$ is E. Then $\{e_k\}_{k=1}^{\infty}$ is an orthonormal basis of E. ■

5.9 The space $L^2[-\pi,\pi]$

Historically, the most well-known orthonormal family is $\{\frac{1}{\sqrt{2\pi}}e^{ikt}\}_{k\in\mathbb{Z}}$ in $L^2[-\pi,\pi]$. It was introduced by **J. Fourier** in his study of heat conduction by means of expansion of functions as trigonometric series, and is usually referred to as the **Fourier basis**. Here $L^2[-\pi,\pi]$ stands for $L^2([-\pi,\pi],\mathcal{L}|[-\pi,\pi],\lambda)$.

For $f\in L^1[-\pi,\pi]$, the function $\hat{f}$ defined on $\mathbb{Z}$ by

$$\hat{f}(k) = \frac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi} f(t)e^{-ikt}\,dt$$

is called the **Fourier transform** of f, and $\hat{f}(k)$'s, $k\in\mathbb{Z}$, are called **Fourier coefficients** of f. If we put $e_k(t)=\frac{1}{\sqrt{2\pi}}e^{ikt}$, then for $f\in L^2[-\pi,\pi]$, $\hat{f}(k)=(f,e_k)$, $k\in\mathbb{Z}$, where $\int_{-\pi}^{\pi} f(t)\overline{g(t)}\,dt \equiv (f,g)$ is the inner product for L^2-spaces. It is easily verified that $(e_k,e_j)=\delta_{kj}$, hence $\{e_k\}$ is indeed an orthonormal family in $L^2[-\pi,\pi]$.

We shall show in this section that $\{e_k\}_{k\in\mathbb{Z}}$ is an orthonormal basis for $L^2[-\pi,\pi]$.

Let $f\in L^1[-\pi,\pi]$ and n be a nonnegative integer; define the Fourier n-th partial sum $S_n(f,t)$ of f by

$$\begin{aligned} S_n(f,t) &= \sum_{k=-n}^{n}\hat{f}(k)e_k(t) = \sum_{k=-n}^{n}\left(\int_{-\pi}^{\pi} f(s)\frac{e^{-iks}}{\sqrt{2\pi}}ds\right)\frac{1}{\sqrt{2\pi}}e^{ikt} \\ &= \frac{1}{2\pi}\sum_{k=-n}^{n}\int_{-\pi}^{\pi} f(s)e^{ik(t-s)}\,ds. \end{aligned}$$

We derive firstly an integral representation for $S_n(f,t)$. Define

$$D_n(t) := \frac{1}{2\pi}\left[1+2\sum_{k=1}^{n}\cos kt\right],$$

then,

$$\begin{aligned}\sin\frac{1}{2}tD_n(t) &= \frac{1}{2\pi}\left[\sin\frac{1}{2}t + 2\sum_{k=1}^{n}\sin\frac{1}{2}t\cos kt\right]\\ &= \frac{1}{2\pi}\left[\sin\frac{1}{2}t + \sum_{k=1}^{n}\left\{\sin\left(k+\frac{1}{2}\right)t - \sin\left(k-\frac{1}{2}\right)t\right\}\right]\\ &= \frac{1}{2\pi}\sin\left(n+\frac{1}{2}\right)t,\end{aligned}$$

hence if t is not an even multiple of π, we have

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

Now,

$$\begin{aligned}S_n(f,t) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(s)\sum_{k=-n}^{n} e^{ik(t-s)}ds\\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(s)\left\{1 + 2\sum_{k=1}^{n}\cos k(t-s)\right\}ds;\end{aligned}$$

thus,

$$S_n(f,t) = \int_{-\pi}^{\pi} f(s)D_n(t-s)ds. \tag{5.11}$$

The functions D_n, $n = 0, 1, 2, \ldots$ are called **Dirichlet kernels**.

It is a common practice to extend a function on $(a, b]$ to be a periodic function on $\mathbb{R}$ with period $(b-a)$; we follow this practice by regarding f as defined on $(-\pi, \pi]$ and extend it periodically to $\mathbb{R}$ with period 2π; then,

$$\begin{aligned}S_n(f,t) &= \int_{-\pi}^{\pi} f(s)D_n(t-s)ds = \int_{-\pi-t}^{\pi-t} f(t+s)D_n(-s)ds\\ &= \int_{-\pi-t}^{\pi-t} f(t+s)D_n(s)ds = \int_{-\pi}^{\pi} f(t+s)D_n(s)ds,\end{aligned}$$

where the last equality follows from the fact that the function $s \mapsto f(t+s)D_n(s)$ is of period 2π (cf. Exercise 4.3.3). Thus (5.11) can be put in the form

$$S_n(f,t) = \int_{-\pi}^{\pi} f(t+s)D_n(s)ds. \tag{5.11$'$}$$

Exercise 5.9.1 Let $X = \{f \in C[-\pi, \pi] : f(-\pi) = f(\pi)\}$; X is a Banach space with sup-norm. For $n = 0, 1, 2, \ldots$ define $\ell_n(f) = S_n(f, 0)$ for $f \in X$.

(i) Show that $\ell_n \in X^*$, $n = 0, 1, 2, \dots$ and

$$\|\ell_n\| = \int_{-\pi}^{\pi} |D_n(t)|dt;$$

(ii) Show that $\lim_{n\to\infty} \|\ell_n\| = \infty$;

(iii) Show that there is $f \in X$ such that

$$\limsup_{n\to\infty} |S_n(f,0)| = \infty.$$

(Hint: cf. Theorem 5.1.3.)

In general, $S_n(f,t)$ is not well behaved as $n \to \infty$, so it is expedient to consider the Cesàro mean of the sequence: for $n = 0, 1, 2, \dots$; let

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

Using (5.11) we have

$$\sigma_n(f,t) = \frac{1}{n+1}\int_{-\pi}^{\pi} f(s)\sum_{k=0}^{n} D_k(t-s)ds = \int_{-\pi}^{\pi} f(s)F_n(t-s)ds, \tag{5.12}$$

where $F_n(t) = \frac{1}{n+1}\sum_{k=0}^{n} D_k(t)$. Since

$$\begin{aligned}
\sin^2\frac{1}{2}tF_n(t) &= \frac{1}{2\pi(n+1)}\sum_{k=0}^{n}\sin(k+\frac{1}{2})t\sin\frac{1}{2}t \\
&= \frac{1}{2\pi(n+1)}\frac{1}{2}\sum_{k=0}^{n}\{\cos kt - \cos(k+1)t\} \\
&= \frac{1}{2\pi(n+1)}\cdot\frac{1}{2}\{1-\cos(n+1)t\} \\
&= \frac{1}{2\pi(n+1)}\sin^2\frac{n+1}{2}t,
\end{aligned}$$

we have

$$F_n(t) = \frac{1}{2\pi(n+1)}\left(\frac{\sin\frac{n+1}{2}t}{\sin\frac{1}{2}t}\right)^2$$

if t is not an even multiple of π. $F_n(t)$, $n = 0, 1, 2, \dots$, are called the **Féjer kernels**. Take $f = 1$ in (5.11) and (5.12), we have

$$\int_{-\pi}^{\pi} D_n(t-s)ds = \int_{-\pi}^{\pi} F_n(t-s)ds = 1, \quad t \in [-\pi,\pi]. \tag{5.13}$$

Theorem 5.9.1 (Féjer) *Suppose that f is continuous on $[-\pi,\pi]$ and $f(-\pi) = f(\pi)$. Then $\sigma_n(f,t) \to f(t)$ uniformly for $t \in [-\pi,\pi]$ when $n \to \infty$.*

Proof From (5.13),

$$\begin{aligned}|\sigma_n(f,t)-f(t)| &= \left|\int_{-\pi}^{\pi}\{f(s)-f(t)\}F_n(t-s)ds\right| \\ &\le \int_{-\pi}^{\pi}|f(s)-f(t)|F_n(t-s)ds.\end{aligned}$$

Since f is continuous on $[-\pi,\pi]$ and $f(-\pi)=f(\pi)$, for any given $\varepsilon>0$, there is $\delta>0$, such that when either $|s-t|<\delta$ or $|s-t|>2\pi-\delta$, we have $|f(s)-f(t)|\le\frac{\varepsilon}{2}$. It is obvious from the form of the function $F_n(s-t)$ that there is $N\in\mathbb{N}$ such that when $n\ge N$,

$$\sup_{\delta\le|t-s|\le2\pi-\delta} F_n(t-s)\le\frac{\varepsilon}{8\pi M}, \tag{5.14}$$

where $M=\sup_{t\in[-\pi,\pi]}|f(t)|$. For $n\ge N$, by (5.14) and the choice of δ,

$$\begin{aligned}|\sigma_n(f,t)-f(t)| &\le \int_{\substack{|t-s|<\delta\\ \text{or } |t-s|>2\pi-\delta}}|f(s)-f(t)|F_n(t-s)ds \\ &\quad+\int_{\delta\le|t-s|\le2\pi-\delta}|f(s)-f(t)|F_n(t-s)ds \\ &\le\frac{\varepsilon}{2}\int_{-\pi}^{\pi}F_n(t-s)ds+2M\cdot\frac{\varepsilon}{8\pi M}\cdot2\pi=\varepsilon;\end{aligned}$$

this shows that $\sigma_n(f,t)\to f(t)$ uniformly for $t\in[-\pi,\pi]$ when $n\to\infty$, because our choice of N is independent of t. ■

Since each $\sigma_n(f,t)$ is a linear combination of $\{\frac{1}{\sqrt{2\pi}}e_k\}_{|k|\le n}$, it follows from the Féjer theorem that $\langle\{\frac{1}{\sqrt{2\pi}}e_k\}_{k\in\mathbb{Z}}\rangle$ is dense in the space of all continuous functions f on $[-\pi,\pi]$ with $f(-\pi)=f(\pi)$ w.r.t. the L^2-norm in $L^2[-\pi,\pi]$. But the latter space contains $C_c(-\pi,\pi)$ which is dense in $L^2[-\pi,\pi]$. As a consequence, the closure of $\langle\{\frac{1}{\sqrt{2\pi}}e_k\}_{k\in\mathbb{Z}}\rangle$ in $L^2[-\pi,\pi]$ is $L^2[-\pi,\pi]$. Thus we have established the following theorem.

Theorem 5.9.2 $\{\frac{1}{\sqrt{2\pi}}e_k\}_{k\in\mathbb{Z}}$, *where* $e_k(t)=e^{ikt}$ *is an orthonormal basis for* $L^2[-\pi,\pi]$.

Because $e^{ikt}=\cos kt+i\sin kt$, it follows from direct computation that

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

Hence, if we put

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos nt dt, \quad n = 0, 1, 2, \dots;$$
$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin nt dt, \quad n = 1, 2, 3, \dots, \tag{5.15}$$

then,

$$S_n(f, x) = \frac{1}{2}a_0 + \sum_{k=1}^{n}\{a_k \cos kx + b_k \sin kx\}. \tag{5.16}$$

This is the traditional form of **Fourier partial sums**; the numbers $a_0, a_1, b_1, a_2, b_2, \dots$ defined by (5.15) are called the **Fourier trigonometric coefficients** of the function f and are expressed symbolically by

$$f(x) \sim \frac{1}{2}a_0 + \sum_{n=1}^{\infty}\{a_n \cos nx + b_n \sin nx\};$$

the series $\frac{1}{2}a_0 + \sum_{n=1}^{\infty}\{a_n \cos nx + b_n \sin nx\}$ is usually referred to as the **Fourier trigonometric series** of f. Whether or not $f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty}\{a_n \cos nx + b_n \sin nx\}$ for $x \in [-\pi, \pi]$ is a well-known problem in analysis, which leads to discovery of many tools in real analysis, including the introduction of Lebesgue measure and Lebesgue integration. Since $\{\frac{1}{\sqrt{2\pi}}e_k\}_{k\in\mathbb{Z}}$ is an orthonormal basis for $L^2[a, b]$ if $b - a = 2\pi$, our discussion so far also holds on any interval of length 2π; in particular, Fourier trigonometric coefficients for integrable functions on such an interval are defined similarly.

Exercise 5.9.2 Consider $L^2[0, 2\pi]$.

(i) Show that $\{\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\cos x, \frac{1}{\sqrt{\pi}}\sin x, \frac{1}{\sqrt{\pi}}\cos 2x, \frac{1}{\sqrt{\pi}}\sin 2x, \dots\}$ is an orthonormal basis for $L^2[0, 2\pi]$.

(ii) For f, g in $L^2[0, 2\pi]$, suppose that

$$f(x) \sim \frac{1}{2}a_0 + \sum_{n=1}^{\infty}\{a_n \cos nx + b_n \sin nx\},$$
$$g(x) \sim \frac{1}{2}c_0 + \sum_{n=1}^{\infty}\{c_n \cos nx + d_n \sin nx\}.$$

Show that

$$\frac{1}{\pi}\int_0^{2\pi} f\bar{g} d\lambda = \frac{1}{2}a_0\bar{c}_0 + \sum_{n=1}^{\infty}\{a_n\bar{c}_n + b_n\bar{d}_n\}.$$

(iii) Suppose that $f \in L^2[0, 2\pi]$ and $a_n = b_n = 0$ for $n \geq k$ for some k. Show that $f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{k-1}\{a_n \cos nx + b_n \sin nx\}$ for a.e. $x \in [0, 2\pi]$.

(iv) Suppose that f is AC on $[0, 2\pi]$ with $f' \in L^2[0, 2\pi]$ and satisfies $f(0) = f(2\pi)$. Show that

$$\frac{1}{\pi}\int_0^{2\pi} |f'|^2 d\lambda = \sum_{n=1}^{\infty} n^2(|a_n|^2 + |b_n|^2),$$

where a_n and b_n are as defined in (ii).

(v) Let f be as in (iv). Show that $\sum_{n=1}^{\infty}(|a_n| + |b_n|) < \infty$ and infer that the Fourier trigonometric series of f converges uniformly to f on $[0, 2\pi]$.

To give a flavor of orthonormal basis in infinite-dimensional spaces, we now prove a classical isoperimetric inequality, following A. Hurwicz.

Theorem 5.9.3 (Isoperimetric inequality) *For any piece-wise C^1 simple closed plane curve with given length L, the following inequality holds:*

$$A \le \frac{L^2}{4\pi},$$

where A is the area of the region enclosed by the curve; and equality holds when and only when the curve is a circle.

Proof Let C be such a curve and choose a parametric representation, $x = x(s), y = y(s)$, $0 \le s \le L$, with arc length as the parameter so that, when s goes from 0 to L, the curve C is traced counter clockwise. Choose the new parameter $t = 2\pi s/L$ and let

$$x(t) \sim \frac{1}{2}a_0 + \sum_{n=1}^{\infty}\{a_n \cos nt + b_n \sin nt\},$$
$$y(t) \sim \frac{1}{2}c_0 + \sum_{n=1}^{\infty}\{c_n \cos nt + d_n \sin nt\};$$

then, using the results in Exercise 5.9.2, we have

$$\frac{dx}{dt} \sim \sum_{n=1}^{\infty}\{nb_n \cos nt - na_n \sin nt\},$$
$$\frac{dy}{dt} \sim \sum_{n=1}^{\infty}\{nd_n \cos nt - nc_n \sin nt\};$$

and

$$\frac{1}{\pi}\int_0^{2\pi}\left\{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2\right\} dt = \sum_{n=1}^{\infty} n^2(a_n^2 + b_n^2 + c_n^2 + d_n^2),$$
$$\frac{1}{\pi}\int_0^{2\pi} x\frac{dy}{dt} dt = \sum_{n=1}^{\infty} n(a_n d_n - b_n c_n).$$

Since $(\frac{dx}{dt})^2 + (\frac{dy}{dt})^2 = (\frac{L}{2\pi})^2\{(\frac{dx}{ds})^2 + (\frac{dy}{ds})^2\} = (\frac{L}{2\pi})^2$ and $A = \int_0^{2\pi} x\frac{dy}{dt}dt$, we have

$$\begin{aligned}\frac{L^2}{4\pi} - A &= \frac{\pi}{2}\sum_{n=1}^{\infty}\{n^2(a_n^2 + b_n^2 + c_n^2 + d_n^2) - 2n(a_n d_n - b_n c_n)\}\\ &= \frac{\pi}{2}\sum_{n=1}^{\infty}\{(na_n - d_n)^2 + (nb_n + c_n)^2 + (n^2-1)(c_n^2 + d_n^2)\} \geq 0.\end{aligned}$$

Hence $A \leq \frac{L^2}{4\pi}$. Now, $\sum_{n=1}^{\infty}\{(na_n - d_n)^2 + (nb_n + c_n)^2 + (n^2-1)(c_n^2 + d_n^2)\} = 0$ if and only if $a_1 = d_1$, $b_1 = -c_1$, and $a_n = b_n = c_n = d_n = 0$ for $n \geq 2$; it follows that $\frac{L^2}{4\pi} = A$ if and only if

$$x = \frac{1}{2}a_0 + a_1\cos t + b_1 \sin t, \quad y = \frac{1}{2}c_0 - b_1\cos t + a_1\sin t,$$

or C is a circle. ■

Theorem 5.9.4 (Weierstrass approximation theorem) *Any continuous function on a finite closed interval $[a,b]$ can be approximated uniformly by polynomials in the interval.*

Proof We may assume without loss of generality that $[a,b] = [-\pi,\pi]$. Since any continuous function f on $[-\pi,\pi]$ can be expressed as

$$f(x) = f(-\pi) + \frac{\{f(\pi) - f(-\pi)\}}{2\pi}(x+\pi) + g(x),$$

where $g(-\pi) = g(\pi) = 0$, it is sufficient to prove the theorem for continuous functions f on $[-\pi,\pi]$ satisfying $f(-\pi) = f(\pi)$. For such a function f, $\sigma_n(f,x) \to f(x)$ uniformly for $x \in [-\pi,\pi]$, by Theorem 5.9.1. Now, $\sigma_n(f,x)$ is a finite linear combination of trigonometric functions $\cos x, \sin x, \cos 2x, \sin 2x, \ldots$; hence, each $\sigma_n(f,x)$ can be approximated uniformly by polynomials on $[-\pi,\pi]$ by Taylor's theorem. Thus, given $\varepsilon > 0$, there is n_0 such that $\sup_{x\in[-\pi,\pi]}|f(x) - \sigma_{n_0}(f,x)| \leq \frac{\varepsilon}{2}$; then let $p(x)$ be a Taylor polynomial of $\sigma_{n_0}(f,x)$ such that $\sup_{x\in[-\pi,\pi]}|\sigma_{n_0}(f,x) - p(x)| \leq \frac{\varepsilon}{2}$; therefore, $\sup_{x\in[-\pi,\pi]}|f(x) - p(x)| \leq \varepsilon$. ■

Exercise 5.9.3 Let $f_n(x) = x^n$, $n = 0,1,2,\ldots$. Show that the Gram–Schmidt process applied to the family $\{f_0, f_1, f_2, \ldots\}$ in $L^2[a,b]$ yields an orthonormal basis for $L^2[a,b]$ $(-\infty < a < b < \infty)$. When $a = -1, b = 1$, denote the orthonormal basis so obtained by $\{\pi_0, \pi_1, \pi_2, \ldots\}$. Show that π_n is a polynomial of degree n, $n = 0,1,2,\ldots$ and find π_0, π_1, and π_2.

Exercise 5.9.4 For $n = 0,1,2,\ldots$, let P_n be the polynomial defined by $P_n(x) = \frac{1}{2^n n!}\frac{d^n(x^2-1)^n}{dx^n}$; P_0, P_1, P_2, ... are called **Legendre polynomials**. Show that $\{P_0, P_1, P_2, \ldots\}$ is an orthogonal family in $L^2[-1,1]$ and $\int_{-1}^1 x^k P_n(x)dx = 0$ for $n \geq 1$ and $0 \leq k < n$.

Exercise 5.9.5 Let $\{\pi_0, \pi_1, \pi_2, \ldots\}$ and $\{P_0, P_1, P_2, \ldots\}$ be as in Exercises 5.9.3 and 5.9.4. Show that for $n = 0, 1, 2, \ldots$, there is a positive constant α_n such that $\pi_n = \alpha_n P_n$.

We digress now from the main theme of this section to discuss briefly the pointwise convergence of Fourier trigonometric series. For this we first prove the Riemann–Lebesgue lemma.

Lemma 5.9.1 (Riemann–Lebesgue) *If f is an integrable function on a finite interval $[a, b]$, then*

$$\lim_{l \mapsto \infty} \int_a^b f(t) \sin lt dt = 0.$$

Proof If J is an interval with endpoints $c < d$ in $[a, b]$, then $\int_J \sin lt dt = -\frac{1}{l}\{\cos ld - \cos lc\} \to 0$ as $l \to \infty$; consequently, the lemma holds if f is a step function. In general, given $\varepsilon > 0$, there is a step function g on $[a, b]$ such that $\int_a^b |f(t) - g(t)| dt < \frac{\varepsilon}{2}$, and therefore,

$$\begin{aligned} \left| \int_a^b f(t) \sin lt dt \right| &\le \left| \int_a^b \{f(t) - g(t)\} \sin lt dt \right| + \left| \int_a^b g(t) \sin lt dt \right| \\ &< \frac{\varepsilon}{2} + \left| \int_a^b g(t) \sin lt dt \right| < \varepsilon, \end{aligned}$$

if l is sufficiently large, because the lemma holds for the step function g. ■

Theorem 5.9.5 (Dini test) *Suppose that f is an integrable function on $(-\pi, \pi)$ and is extended to $\mathbb{R}$ periodically. Let $t_0 \in [-\pi, \pi]$, then,*

$$\lim_{n \to \infty} S_n(f, t_0) = f(t_0)$$

if $s \mapsto \frac{1}{s}\{f(t_0 + s) - f(t_0)\}$ is integrable in a neighborhood of 0.

Proof If $s \mapsto \frac{1}{s}\{f(t_0 + s) - f(t_0)\}$ is integrable in a neighborhood of 0, then the function g defined by

$$g(s) = \frac{1}{2\pi} \frac{f(t_0 + s) - f(t_0)}{\sin \frac{1}{2} s} = \frac{1}{2\pi} \frac{s}{\sin \frac{1}{2} s} \frac{f(t_0 + s) - f(t_0)}{s}, \quad s \in [-\pi, \pi],$$

is integrable on $[-\pi, \pi]$. Now, from (5.11)′ we have,

$$\begin{aligned} S_n(f, t_0) - f(t_0) &= \int_{-\pi}^{\pi} \{f(t_0 + s) - f(t_0)\} D_n(s) ds \\ &= \int_{-\pi}^{\pi} g(s) \sin \left(n + \frac{1}{2} \right) s ds \to 0 \end{aligned}$$

as $n \to \infty$, by the Riemann–Lebesgue lemma. ■

Exercise 5.9.6 Let f be an even function on $[-\pi,\pi]$ defined on $[0,\pi]$ by $f(s) = 1 - \frac{s}{\pi}$. Show that the Fourier trigonometric series of f converges uniformly to f on $[-\pi,\pi]$. In particular, verify that $\sum_{k=0}^{\infty} \frac{1}{(2k+1)^2} = \frac{\pi^2}{8}$.

Exercise 5.9.7 Suppose that f is a periodic function of period 2π on $\mathbb{R}$ and is integrable on $[-\pi,\pi]$. Show that if $f = 0$ on a neighborhood of t_0, then $S_n(f,t) \to 0$ uniformly on a neighborhood of t_0.

Exercise 5.9.8 Suppose that f is integrable on $[-\pi,\pi]$ and $f(t_0+), f(t_0-)$ exist at $t_0 \in [-\pi,\pi]$. Show that

$$\lim_{n\to\infty} S_n(f,t_0) = \frac{1}{2}\{f(t_0-) + f(t_0+)\}$$

if $\int_{-\varepsilon}^{0} \left|\frac{f(t_0+s)-f(t_0-)}{s}\right| ds < \infty$, and $\int_0^{\varepsilon} \left|\frac{f(t_0+s)-f(t_0+)}{s}\right| ds < \infty$ for some $\varepsilon > 0$. (Hint: $\int_{-\pi}^{0} D_n(s)ds = \int_0^{\pi} D_n(s)ds = \frac{1}{2}$.)

Exercise 5.9.9 Let f be a periodic function with period π on $\mathbb{R}$, and $f(s) = s$ for $0 \le s < \pi$. Find the Fourier trigonometric series for f and evaluate $\sum_{n=1}^{\infty} a_n$, where

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(s)\cos nsds, \qquad n = 0, 1, 2, \ldots.$$

Lemma 5.9.2 *There is $c > 0$ s.t. $\left|\int_{\delta}^{\eta} D_n(s)ds\right| \le c$ for all $n \in \mathbb{N}$ and $0 \le \delta < \eta \le \pi$.*

Proof Let $n \in \mathbb{N}$ and $0 \le \delta < \eta \le \pi$. It will be clear from the following argument that we may assume $\delta < \frac{2}{2n+1} < \eta$; then,

$$0 \le \frac{\sin(n+\frac{1}{2})s}{\sin\frac{1}{2}s} = \frac{\frac{1}{2}s}{\sin\frac{1}{2}s} \cdot \frac{\sin(n+\frac{1}{2})s}{\frac{1}{2}s} < 1 \cdot (2n+1)$$

for $0 < s < \frac{2}{2n+1}$, and hence,

$$\int_{\delta}^{\frac{2}{2n+1}} \frac{\sin(n+\frac{1}{2})s}{\sin\frac{1}{2}s} < (2n+1)\cdot\frac{2}{2n+1} = 2.$$

Thus,

$$\left|\int_{\delta}^{\eta} D_n(s)ds\right| \le \frac{1}{2\pi}\left\{2 + \left|\int_{\frac{2}{2n+1}}^{\eta} \frac{\sin(n+\frac{1}{2})s}{\sin\frac{1}{2}s}ds\right|\right\}.$$

But by the second mean-value theorem (actually, Lemma 4.5.2), there is $\frac{2}{2n+1} \leq \eta' \leq \eta$ such that

$$\begin{aligned}\left|\int_{\frac{2}{2n+1}}^{\eta} \frac{\sin(n+\frac{1}{2})s}{\sin\frac{1}{2}s}ds\right| &= \left|\frac{1}{\sin(\frac{1}{2n+1})}\int_{\frac{2}{2n+1}}^{\eta'} \sin\left(n+\frac{1}{2}\right)sds\right| \\ &= \frac{1}{\sin(\frac{1}{2n+1})}\left|\frac{1}{n+\frac{1}{2}}\left\{\cos 1 - \cos\left(n+\frac{1}{2}\right)\eta'\right\}\right| \\ &\leq \frac{1}{(2n+1)\sin(\frac{1}{2n+1})} \\ &= \left\{(2n+1)\left[\frac{1}{2n+1} - \frac{1}{3!}\left(\frac{1}{2n+1}\right)^3 + \cdots\right]\right\}^{-1} \\ &\leq \left\{1 - \frac{1}{3!}\left(\frac{1}{2n+1}\right)^2\right\}^{-1} = \frac{54}{53},\end{aligned}$$

and consequently,

$$\left|\int_{\delta}^{n} D_n(s)ds\right| \leq \frac{1}{2\pi}\left(2 + \frac{54}{53}\right).$$

Thus we may take it that $c = \frac{1}{2\pi}\left(2 + \frac{54}{53}\right)$. ■

Theorem 5.9.6 (Dirichlet–Jordan) *Let f be a BV function on $[-\pi, \pi]$; then* $\frac{1}{2}a_0 + \sum_{n=1}^{\infty}\{a_n \cos nt + b_n \sin nt\} = \lim_{n\to\infty} S_n(f,t) = \frac{1}{2}\{f(t-) + f(t+)\}$.

Proof Since f is the difference of two monotone increasing functions, we may assume without loss of generality that f is monotone increasing, and consider f as defined on $(-\pi, \pi]$ and then extend f to $\mathbb{R}$ as a periodic function with period 2π. Now fix $t \in [-\pi, \pi]$. Given that $\varepsilon > 0$, there is $\delta > 0$ such that $f(t+s) - f(t+) < \frac{\varepsilon}{2c}$ for $0 < s \leq \delta$, where c is the constant in Lemma 5.9.2. We choose δ small enough so that $f(t+s)$ is monotone increasing in s on $[0, \delta]$, if $f(t+0)$ is understood to be $f(t+)$. Then, $\int_0^{\delta}\{f(t+s) - f(t+)\}D_n(s)ds = \{f(t+\delta) - f(t+)\}\int_{\delta'}^{\delta} D_n(s)ds$ for some $\delta' \in [0, \delta]$ by the second-mean value theorem, and hence

$$\left|\int_0^{\delta}\{f(t+s) - f(t+)\}D_n(s)ds\right| < \frac{\varepsilon}{2c}\cdot c = \frac{\varepsilon}{2}.$$

Now,

$$\left|\int_0^{\pi} f(t+s)D_n(s)ds - \frac{1}{2}f(t+)\right| = \left|\int_0^{\pi}\{f(t+s) - f(t+)\}D_n(s)ds\right\}$$

$$\leq \left|\int_0^{\delta}\{f(t+s) - f(t+)\}D_n(s)ds\right| + \left|\int_{\delta}^{\pi}\{f(t+s) - f(t+)\}D_n(s)ds\right|$$

$$< \frac{\varepsilon}{2} + \left|\frac{1}{2\pi}\int_{\delta}^{\pi}\frac{f(t+s) - f(t+)}{\sin\frac{1}{2}s}\cdot\sin\left(n+\frac{1}{2}\right)sds\right| < \varepsilon$$

if n is sufficiently large, by the Riemann–Lebesgue lemma, because the function $s \mapsto \frac{f(t+s)-f(t+)}{\sin\frac{1}{2}s}$ is integrable on $[\delta, \pi]$. Thus $\lim_{n\to\infty}\int_0^{\pi} f(t+s)D_n(s)ds = \frac{1}{2}f(t+)$. Similarly, $\lim_{n\to\infty}\int_{-\pi}^{0} f(t+s)D_n(s)ds = \frac{1}{2}f(t-)$. Consequently, $\lim_{n\to\infty}\int_{-\pi}^{\pi} f(\mathrm{t+s})D_n(s)ds = \lim_{n\to\infty} S_n(f,t) = \frac{1}{2}\{f(t-) + f(t+)\}$. ■

5.10 Weak convergence

The concept of limit for sequences in a metric space is defined in Section 1.4 in terms of the metric of the space. When normed vector spaces are concerned, there is a weaker form of concept of limit for sequences, towards the introduction of which we now turn.

Suppose that X is a n.v.s. and $\{x_k\}$ a sequence in X. If $x \in X$ satisfies $\langle x, x^*\rangle = \lim_{k\to\infty}\langle x_k, x^*\rangle$ for every $x^* \in X^*$, x is called a **weak limit** of the sequence $\{x_k\}$; since X^* separates points of X, if x is a weak limit of $\{x_k\}$, it is the only weak limit of $\{x_k\}$, and hence is the weak limit of $\{x_k\}$ and is denoted by $w\text{-}\lim_{k\to\infty} x_k$. We often write $x_k \rightharpoonup x$ to indicate that $x = w\text{-}\lim_{k\to\infty} x_k$. To distinguish between weak limit and limit defined in terms of the norm of X, the latter is called the limit in norm and we employ notation $x = \lim_{k\to\infty} x_k$ or $x_k \to x$ to mean that x is the limit of $\{x_k\}$ in norm. If the weak (norm) limit of a sequence exists, the sequence is said to be **weakly convergent** (**convergent in norm**) or is said to **converge weakly** (in **norm**). Clearly, in a Hilbert space E, $x = w\text{ - }\lim_{k\to\infty} x_k$ if and only if $(x, y) = \lim_{x\to\infty}(x_k, y)$ for all $y \in E$, and $x_k \to x$ implies that $x_k \rightharpoonup x$.

Proposition 5.10.1 *A weakly convergent sequence in a n.v.s. X is bounded.*

Proof Let $\{x_k\}$ be a weakly convergent sequence in X. For $k \in \mathbb{N}$, let l_k be the bounded linear functional on X^*, defined by $l_k(x^*) = \langle x_k, x^*\rangle$ for $x^* \in X^*$. Note that X^* is a Banach space and by Theorem 5.5.2, $\|l_k\| = \|x_k\|$ for $k \in \mathbb{N}$. Let $x = w\text{-}\lim_{k\to\infty} x_k$, then since $\lim_{k\to\infty}|l_k(x^*)| = |\langle x, x^*\rangle|$, $\sup_k |l_k(x^*)| < \infty$ for each $x^* \in X^*$. By the principle of uniform boundedness (Theorem 5.1.3), $\sup_k \|l_k\| = \sup_k \|x_k\| < \infty$. ■

Remark Proposition 5.10.1 is actually contained in Theorem 5.1.4.

Exercise 5.10.1 Show that a bounded sequence $\{x_k\}$ converges to x weakly in a n.v.s. X if and only if there is $S \subset X^*$ such that $\langle S \rangle$ is dense in X^* and $\langle x, x^* \rangle = \lim_{k\to\infty} \langle x_k, x^* \rangle$ for $x^* \in S$.

Exercise 5.10.2 Show that a sequence $\{x_n\}$ in a finite-dimensional n.v.s. X converges weakly if and only if it converges in norm.

Theorem 5.10.1 *Every bounded sequence $\{x_k\}$ in a Hilbert space E has a subsequence which converges weakly in E.*

Proof Let F be the closure of $\langle \{x_k\} \rangle$ in E, then F is a Hilbert space with inner product inherited from E. Put $\sup_k \|x_k\| = M < \infty$. We show first that $\{x_k\}$ has a subsequence which converges weakly in F.

Since $\{(x_k, x_1)\}_k$ is a bounded sequence in $\mathbb{C}$, there is a subsequence $\{x_k^{(1)}\}$ of $\{x_k\}$ such that $\lim_{k\to\infty}(x_k^{(1)}, x_1)$ exists. Suppose now that sequences $\{x_k^{(1)}\}, \ldots, \{x_k^{(n)}\}$ have been chosen so that each of them except the first is a subsequence of the preceding one and $\lim_{k\to\infty}(x_k^{(n)}, x_j)$ exists for $j = 1, \ldots, n$. Since $\{(x_k^{(n)}, x_{n+1})\}$ is bounded, there is a subsequence $\{x_k^{(n+1)}\}$ of $\{x_k^{(n)}\}$ such that $\lim_{k\to\infty}(x_k^{(n+1)}, x_{n+1})$ exists. Clearly, $\lim_{k\to\infty}(x_k^{(n+1)}, x_j)$ exists for $j = 1, \ldots, n$, because $\{x_k^{(n+1)}\}$ is a subsequence of $\{x_k^{(n)}\}$. We have therefore obtained a sequence $\{x_k^{(1)}\}, \{x_k^{(2)}\}, \ldots, \{x_k^{(n)}\}, \ldots$ of subsequences of $\{x_k\}$ such that $\{x_k^{(n+1)}\}$ is a subsequence of $\{x_k^{(n)}\}$ for each $n \in \mathbb{N}$ and where $\lim_{k\to\infty}(x_k^{(n)}, x_j)$ exists for $j = 1, \ldots, n$. Now, $\{x_k^{(k)}\}$ is a subsequence of $\{x_k\}$ and $\lim_{k\to\infty}(x_k^{(k)}, x_j)$ exists for each $j \in \mathbb{N}$. For convenience, put $y_k = x_k^{(k)}$ for $k \in \mathbb{N}$, then $\lim_{k\to\infty}(y_k, z)$ exists for $z \in \langle \{x_k\} \rangle$. Let $l(z) = \overline{\lim_{k\to\infty}(y_k, z)}$, then l is a linear functional on $\langle \{x_k\} \rangle$; obviously, $|l(z)| \le M\|z\|$ for $z \in \langle \{x_k\} \rangle$, hence l is bounded on $\langle \{x_k\} \rangle$, and can be extended uniquely to be a bounded linear functional on F, still denoted by l. By the Riesz representation theorem, there is unique $x \in F$ such that $l(u) = (u, x)$ for $u \in F$; in particular, for $z \in \langle \{x_k\} \rangle$, $(z, x) = \overline{\lim_{k\to\infty}(y_k, z)}$ i.e. $(x, z) = \lim_{k\to\infty}(y_k, z)$. Since $\langle \{x_k\} \rangle$ is dense in F, $y_k \rightharpoonup x$ in F, by Exercise 5.10.1.

We claim now that $y_k \rightharpoonup x$ in E. Let $u \in E$, then $u = z + v$, where $z \in F$ and $v \in F^\perp$, by Corollary 5.6.1. Thus,

$$(x, u) = (x, z + v) = (x, z) = \lim_{k\to\infty}(y_k, z) = \lim_{k\to\infty}(y_k, z + v) = \lim_{k\to\infty}(y_k, u),$$

and hence $y_k \rightharpoonup x$ in E. ∎

Exercise 5.10.3 Suppose that $\{e_k\}$ is an orthonormal sequence in a Hilbert space E. Show that $e_k \rightharpoonup 0$, but 0 is not a limit of $\{e_k\}$ in norm. (Hint: for $x \in E$, $\sum_{k=1}^{\infty} |(x, e_k)|^2 \le \|x\|^2$.)

Exercise 5.10.4 (Cf. Example 2.7.2) Show that if $1 < p < \infty$, then $f_n \rightharpoonup f$ in $l^p(\Omega)$ if and only if $\sup_n \|f_n\|_p < \infty$ and $f_n(\omega) \to f(\omega)$ for all $\omega \in \Omega$.

Exercise 5.10.5 Suppose that X is a reflexive Banach space and $\{x_n\}$ is a bounded sequence in X. Assume that X^* is separable and let $\{x_1^*, x_2^*, \ldots\}$ be a countable dense set in X^*. Show that $\{x_n\}$ has a subsequence which converges weakly by the following steps.

(i) Show that $\{x_n\}$ has a subsequence $\{y_n\}$ such that $\lim_{n\to\infty}\langle y_n, x_k^*\rangle$ exists and is finite for all $k \in \mathbb{N}$.

(ii) Show that $\lim_{n\to\infty}\langle y_n, x^*\rangle$ exists and is finite for all $x^* \in X^*$.

(iii) Put $l(x^*) = \lim_{n\to\infty}\langle y_n, x^*\rangle$. Show that $l \in X^{**}$, and there is $x \in X$ such that $l(x^*) = \langle x, x^*\rangle$ for all $x^* \in X^*$.

Theorem 5.10.2 (Banach–Saks) *If $\{x_k\}$ is a bounded sequence in a Hilbert space E, then it has a subsequence $\{y_k\}$ such that $\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} y_k$ in norm exists.*

Proof There is a subsequence $\{z_k\}$ of $\{x_k\}$ and $x \in E$ such that $z_k \rightharpoonup x$, by Theorem 5.10.1. Let $\hat{z}_k = z_k - x$, then $\hat{z}_k \rightharpoonup 0$. Choose inductively a subsequence $\{\hat{y}_k\}$ of $\{\hat{z}_k\}$ so that

$$|(\hat{y}_1, \hat{y}_{n+1})| \leq \frac{1}{n}, \ \ldots, \ |(\hat{y}_n, \hat{y}_{n+1})| \leq \frac{1}{n}$$

for all $n \in \mathbb{N}$. Then,

$$\begin{aligned}
\left\| n^{-1}\sum_{k=1}^{n}\hat{y}_k \right\|^2 &= n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\hat{y}_i, \hat{y}_j) \\
&= n^{-2}\left\{\sum_{i=1}^{n}(\hat{y}_i, \hat{y}_i) + 2\sum_{1\leq i<j\leq n}\operatorname{Re}(\hat{y}_i, \hat{y}_j)\right\} \\
&\leq n^{-2}\left\{nC + 2\sum_{j=2}^{n}\sum_{i=1}^{j-1}|(\hat{y}_i, \hat{y}_j)|\right\} \\
&\leq n^{-2}\{nC + 2(n-1)\} < n^{-1}\{C+2\},
\end{aligned}$$

where $C = \sup_n\{\|\hat{y}_n\|^2\} \leq \sup_n(\|x_n\| + \|x\|)^2 < \infty$. Thus $\lim_{n\to\infty} n^{-1}\sum_{k=1}^{n}\hat{y}_k = 0$. We complete the proof by letting $y_k = \hat{y}_k + x$. ■

Theorems 5.10.1 and 5.10.2 have already shown the relevance of weak convergence, in that in terms of weak convergence, bounded sets in a Hilbert space reveal a certain compactness property. We shall now apply Theorem 5.10.1 to prove a mean ergodic theorem of F. Riesz which shows that bounded linear operators from a Hilbert space into itself of a certain kind have eigenvalue 1 whose eigenspace can be explicitly described.

In the following, we fix a bounded linear operator T from a Hilbert space E into itself, having the property that $\|T^n\| \leq \alpha < \infty$ for all $n \in \mathbb{N}$ for some $\alpha > 0$. Let $T_1 = T$ and $T_n = \frac{1}{n}\{T + T^2 + \cdots + T^n\}$ for $n \geq 2$, and for $x \in E$, put $x_n = T_n x$ for $n \in \mathbb{N}$.

Lemma 5.10.1 *If $x \in \overline{(1-T)E}$, then* $\lim_{n\to\infty}\|x_n\| = 0$.

Proof If $x \in (1 - T)E$, i.e. $x = y - Ty$ for some y in E, then

$$x_n = (y - Ty)_n = \frac{1}{n}\{T(y - Ty) + \cdots + T^n(y - Ty)\} = \frac{1}{n}\{Ty - T^{n+1}y\},$$

and hence, $\|x_n\| \leq \frac{2\alpha}{n}\|y\|$, from which $\|x_n\| \to 0$ follows. Now suppose that $x \in \overline{(1 - T)E}$. Given $\varepsilon > 0$, there is $z \in (1 - T)E$ such that $\|x - z\| < \frac{\varepsilon}{2\alpha}$. It is clear that $\|(x - z)_n\| \leq \alpha\|x - z\| < \frac{\varepsilon}{2}$. Since $\|z_n\| \to 0$, by the first part of the proof, there is $n_0 \in \mathbb{N}$ such that $\|z_n\| < \frac{\varepsilon}{2}$ whenever $n \geq n_0$, hence, $\|x_n\| = \|z_n + (x - z)_n\| \leq \|z_n\| + \|(x - z)_n\| < \|z_n\| + \frac{\varepsilon}{2} < \varepsilon$ whenever $n \geq n_0$. Thus $\|x_n\| \to 0$. ■

Lemma 5.10.2 *If x_∞ is the weak limit of a subsequence of $\{x_n\}$, then x_∞ is a fixed point of T, i.e. $Tx_\infty = x_\infty$.*

Proof Let $\{x_{n_k}\}$ be a subsequence of $\{x_n\}$ such that $w\text{-}\lim_{k\to\infty} x_{n_k} = x_\infty$. Since $TT_nx - T_nx = T_nTx - T_nx = T_n(Tx - x) = (Tx - x)_n$, $\|TT_nx - T_nx\| \to 0$, by Lemma 5.10.1 and hence $TT_{n_k}x \rightharpoonup x_\infty$. But, since for each $y \in E$, $(Tz, y) = (z, \hat{y})$ for some $\hat{y} \in E$ and for all $z \in E$ by the Riesz representation theorem, we have $(TT_{n_k}x, y) = (x_{n_k}, \hat{y}) \to (x_\infty, \hat{y}) = (Tx_\infty, y)$ and consequently, $TT_{n_k}x \rightharpoonup Tx_\infty$. We infer from this last fact and the fact that $TT_{n_k}x \rightharpoonup x_\infty$, that $T_{x_\infty} = x_\infty$. ■

To prepare for the statement of the mean ergodic theorem of Riesz, we shall say that a sequence $\{T_n\} \subset L(X, Y)$ **converges strongly** to $T \in L(X, Y)$ if $\lim_{n\to\infty} T_nx = Tx$ for all $x \in X$, where X and Y are n.v.s.'s over the same scalar field $\mathbb{C}$ or $\mathbb{R}$. To distinguish this mode of convergence, if $\lim_{n\to\infty} \|T_n - T\| = 0$, we say that T_n **converges in operator norm** to T.

Theorem 5.10.3 (Mean ergodic theorem of Riesz) *T_n converges strongly in $L(E)$ to a linear operator T_∞ with the property that $TT_\infty = T_\infty$.*

Proof For $x \in E$, $\|x_n\| = \|\frac{1}{n}\{Tx + \cdots + T^nx\}\| \leq \alpha\|x\|$, hence $\{x_n\}$ is a bounded sequence in E. $\{x_n\}$ has a subsequence $\{x_{n_k}\}$ which converges weakly to x_∞ in E. We know from Lemma 5.10.2 that $Tx_\infty = x_\infty$, and hence $(x_\infty)_n = x_\infty$. We claim that $\lim_{n\to\infty} x_n = x_\infty$, i.e. x_n converges strongly to x_∞. Now, $x_n = (x_\infty + \{x - x_\infty\})_n = (x_\infty)_n + (x - x_\infty)_n = x_\infty + (x - x_\infty)_n$, thus $\|x_n - x_\infty\| = \|(x - x_\infty)_n\|$; to verify the claim it is sufficient to show that $x - x_\infty \in \overline{(1 - T)E}$, by Lemma 5.10.1. To see this, let Y be the orthogonal complement of $(1 - T)E$ in E and observe that $x - x_{k_n} = \frac{1}{n_k}\{(x - Tx) + \cdots + (x - T^{n_k}x)\}$ is in $(1 - T)E$, because $(x - T^mx) = (1 - T)(1 + T + \cdots + T^{m-1})x \in (1 - T)E$ for each $m \in \mathbb{N}$; then for $y \in Y$, we have $(x - x_{n_k}, y) = 0$, which implies that

$$(x - x_\infty, y) = \lim_{k\to\infty} (x - x_{n_k}, y) = 0,$$

i.e. $x - x_\infty \in Y^\perp = \overline{(1 - T)E}$. Thus we have shown that $\|x_n - x_\infty\| \to 0$. This last fact shows in particular that all weakly convergent subsequences of $\{x_n\}$ converge weakly to the same element x_∞. Let $x_\infty = T_\infty x$, then T_∞ is a linear operator from E

into E and is the strong limit of $\{T_n\}$, i.e. $T_\infty x = \lim_{n\to\infty} T_n x$. That T_∞ is a bounded linear operator follows from the Banach–Steinhaus theorem (Theorem 5.1.4). From Lemma 5.10.2, $Tx_\infty = x_\infty$ and consequently, $TT_\infty x = T_\infty x$, or $TT_\infty = T_\infty$. ■

Corollary 5.10.1 $TT_\infty = T_\infty = T_\infty T = T_\infty^2$.

Proof From $TT_\infty = T_\infty$, it follows that $T^n T_\infty = T_\infty$ and $T_n T_\infty = T_\infty$ for all $n \in \mathbb{N}$; by letting $n \to \infty$ in the last equality, we obtain $T_\infty^2 = T_\infty$. To see that $T_\infty T = T_\infty$, note first that $T_n T - T_n = \frac{1}{n}(T^{n+1} - T)$ and hence $\|T_n Tx - T_n x\| \leq \frac{2\alpha}{n}\|x\|$ for all $x \in E$; thus $T_\infty T = T_\infty$ follows. ■

Exercise 5.10.6 Show that 1 is an eigenvalue of T and $T_\infty E$ is the eigenspace of T belonging to the eigenvalue 1.

The well-known ergodic theorem of J. von Neumann is a consequence of Theorem 5.10.3, as we shall now show.

Let (Ω, Σ, p) be a probability space. A bijective map $T : \Omega \to \Omega$ is called a **flow** on (Ω, Σ, p) if T is measurable and measure preserving.

Theorem 5.10.4 (von Neumann mean ergodic theorem) *Suppose that T is a flow on a probability space (Ω, Σ, p). Define a linear operator $\widehat{T}$ from $L^2(\Omega, \Sigma, p)$ to itself by*

$$(\widehat{T}f)(\omega) = f \circ T(\omega),\ \omega \in \Omega,\ f \in L^2(\Omega, \Sigma, p);$$

and let $\widehat{T}_n = \frac{1}{n}\{\widehat{T} + \cdots + \widehat{T}^n\}$. Then for $f \in L^2(\Omega, \Sigma, p)$, $\widehat{T}_n f \to f^*$ in $L^2(\Omega, \Sigma, p)$. Furthermore, $\widehat{T}f^* = f^*$ i.e. $f^*(T\omega) = f^*(\omega)$ for a.e. $\omega \in \Omega$.

Proof Since T is a flow on (Ω, Σ, p), $\|\widehat{T}f\| = \|f\|$ for all $f \in L^2(\Omega, \Sigma, p)$. Hence $\|\widehat{T}\| = 1$ and $\|\widehat{T}^n\| \leq \|\widehat{T}\|^n = 1$ for all $n \in \mathbb{N}$. The theorem follows from Theorem 5.10.3. ■

A flow T on (Ω, Σ, p) is called an **ergodic** flow if for each $f \in L^2(\Omega, \Sigma, p)$, the element f^* in the conclusion of Theorem 5.10.4 is constant a.e. on Ω.

Corollary 5.10.2 *Suppose that T is an ergodic flow on (Ω, Σ, p) and $\widehat{T}$, $\widehat{T}_n$, $n \in \mathbb{N}$, are defined as in Theorem 5.10.4. Then for $f \in L^2(\Omega, \Sigma, p)$, $\widehat{T}_n f \to \int_\Omega f dp$ in $L^2(\Omega, \Sigma, p)$.*

Proof For $f \in L^2(\Omega, \Sigma, p)$, let f^* be as in Theorem 5.10.4. Since $\widehat{T}_n f \to f^*$ in $L^2(\Omega, \Sigma, p)$, $\widehat{T}_n f \to f^*$ in $L^1(\Omega, \Sigma, p)$ and, a fortiori, $\lim_{n\to\infty} \int_\Omega \widehat{T}_n f dp = \int_\Omega f^* dp$; but $\int_\Omega \widehat{T} f dp = \int_\Omega \widehat{T}^2 f dp = \cdots = \int_\Omega \widehat{T}_n f dp = \cdots = \int_\Omega f dp$, from the fact that T is measure preserving, hence $\int_\Omega f^* dp = \int_\Omega f dp$. Now that $f^* =$ constant a.e. implies $f^* = \int_\Omega f dp$ a.e. ■