Functional Analysis Notes

泛函分析笔记

何憾[1]

February 28, 2021

[1]访问我的个人主页, 可查看并下载其他笔记.

Contents

0.1 Basic Linear Algebra

Let X and Y be linear spaces over the same scalar field $\mathbb{F}$. A mapping T defined on a linear subspace $D = D(T)$ of X and taking values in Y is said to be *linear*, if

$$T(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 T x_1 + \alpha_2 T x_2, \text{ for all } x_1, x_2 \in D \text{ and all } \alpha, \beta \in \mathbb{F}.$$

T is also called a *linear operator* or *linear transformation* on $D(T) \subset X$ into Y. If T is taking values in the scalar field, i.e., $Y = \mathbb{F}$, we say T is a *linear functional* on $D(T)$. We say that $D(T)$ is the *domain* of T and sometimes is denoted by $\operatorname{dom}(T)$. Moreover, the *range* the of T, denoted by $R(T)$ or $\operatorname{ran}(T)$, is given by

$$R(T) \coloneqq T(X) = \{y \in Y : y = Tx \text{ for some } x \in D(T)\}.$$

The *null space* or the *kernel* of T, denoted by $N(T)$ or $\ker(T)$, is given by

$$N(T) \coloneqq T^{-1}\{0\} = \{x \in D(T) : Tx = 0\}.$$

Clearly, $N(T)$ is a linear subspace of X and $R(T)$ is a linear subspace of Y.

If a linear operator T gives a one-to-one map of $D(T)$ onto $R(T)$, then the inverse map T^{-1} gives a linear operator on $R(T)$ onto $D(T)$:

$$T^{-1}Tx = x \text{ for } x \in D(T)\ ;\ TT^{-1}y = y \text{ for } y \in R(T).$$

T^{-1} is the *inverse operator* or, in short, the *inverse* of T. Clearly, T admits the inverse T^{-1} if and only if $N(T) = \{0\}$; i.e., $Tx = 0$ implies $x = 0$.

Let T_1 and T_2 be linear operators with domains $D(T_1)$ and $D(T_2)$ both contained in X, and ranges $R(T_1)$ and $R(T_2)$ both contained in Y. Then

$$T_1 = T_2 \text{ iff } D(T_1) = D(T_2) \text{ and } T_1 x = T_2 x \text{ for all } x \in D(T_1).$$

If $D(T_1) \subset D(T_2)$ and $T_1 x = T_2 x$ for all $x \in D(T_1)$, then T_2 is called an *extension* of T_1, and T_1 a *restriction* of T_2; we shall then write

$$T_1 \subset T_2 \text{ or } T_2|_{D(T_1)} = T_1.$$

Exercise 0.1. Let X, Y be two vector space, and S, T are linear operators with domain in X and range in Y ao that $S \subset T$. If S is surjective and T is injective, then $S = T$.

Let X and Y be linear space over $\mathbb{F}$. Denote by $\mathcal{L}(X, Y)$ all the linear operator defined on X and taking values in Y. For $T, S \in \mathcal{L}(X, Y)$ and $\alpha \in \mathbb{F}$, define the operations of addition and scalar multiplication as follows:

$$(T + S)(x) = Tx + Sx, \text{ for each } x \in X$$
$$(\alpha T)(x) = \alpha Tx, \text{ for each } x \in X.$$

Then $\mathcal{L}(X, Y)$ becomes a linear space over $\mathbb{F}$.

Exercise 0.2. Let X be a vector space, and $S, T \in \mathcal{L}(X)$. Suppose that $ST = TS$ and that ST is bijective. Then also S and T are bijective.

Ordinary algebraic operations with operators whose domain is not the hull space must be handled with care; the domains have to be watched. Let $(T, (D(T)),\ (S, (D(S)),\ (R, D(R))$ be linear operators whose domain and range lie in X. Here are the natural definitions for the domains of sums and products:

$$D(S + T) = D(S) \cap D(T)$$
$$D(ST) = \{x \in D(T) : Tx \in D(S)\}$$

Then the usual associative laws hold:

$$(R + S) + T = R + (S + T)\ ,\ (RS)T = R(ST)\,.$$

Scalar multiplication is defined as follows: If $\alpha = 0$, then $D(\alpha T) = H$ and $\alpha T = 0$. If $\alpha \neq 0$, then $D(\alpha T) = D(T)$ and $(\alpha T)x = \alpha(Tx)$ for $x \in D(T)$.

Lemma 0.1. *Let X be a vector space over $\mathbb{F}$. Let $\ell, \ell_1, \cdots, \ell_n$ be linear functionals on X. Then $\ell \in \operatorname{span}\{\ell_1, \cdots, \ell_n\}$ if and only if*

$$\bigcap_{k=1}^{n} \ker(\ell_k) \subset \ker(\ell)\,. \tag{1}$$

Proof. We only prove the sufficiency. Define $T : X \to \mathbb{F}^n$ by

$$Tx = (\ell_1 x, \cdots, \ell_n x) \ .$$

If $Tx = Tx'$, then $\cap_k \ker(\ell_k) \subset \ker(\ell)$ implies that $\ell x = \ell x'$. Thus we can define a linear functional f on $\mathrm{ran}(T)$ by letting

$$f(Tx) = \ell x \quad \text{for each } Tx \in \mathrm{ran}(T) \, .$$

We can extend f to a linear functional on $\mathbb{F}^n$. This means that there exist scalars $\alpha_i, i = 1, 2 \cdots, n$ such that

$$f\left(u_1, \ldots, u_n\right) = \alpha_1 u_1 + \cdots + \alpha_n u_n \, .$$

Thus

$$\ell x = f(Tx) = f\left(\ell_1 x, \ldots, \ell_n x\right) = \sum_{i=1}^{n} \alpha_i \ell_i x \, . \qquad \square$$

Corollary 0.2. *Let X be a vector space over $\mathbb{F}$. Let f, g be linear functionals on X. Then there exists a scalar $c \in \mathbb{F}$ so that $g = cf$ if and only if* $\ker(f) \subset \ker(g)$.

Chapter 1

Normed Linear Spaces

1.1 Fundamentals

Definition 1.1. A *seminorm* on a linear space X is a nonnegative real-valued function $p : X \to [0, \infty)$ which satisfies the following properties.

(a) Absolute homogeneity: $p(\lambda x) = |\lambda| p(x)$ for all $\lambda \in \mathbb{F}$ and $x \in X$.

(b) Trigonometric inequality: $p(x + y) \leq p(x) + p(y)$ for all $x, y \in X$.

It follows from (a) that $p(0) = 0$. A *norm* is a seminorm p satisfying:

(c) Positive definiteness: $p(x) = 0$ if and only if $x = 0$.

Usually a norm is denoted by $\|\cdot\|$, A *normed linear space* is a pair $(X, \|\cdot\|)$, where X is a linear space and $\|\cdot\|$ a norm on X. If $(X, \|\cdot\|)$ is a normed linear space, then

$$d(x, y) \coloneqq \|x - y\| \text{ for all } x, y \in X$$

defines a metric on X. Such a metric d is said to be induced or generated by the norm $\|\cdot\|$. Thus, every normed linear space is a metric space, and unless otherwise specified, we shall henceforth regard any normed linear space as a metric space with respect to the metric induced by its norm.

Remark 1.1. It's natural to ask when a metric is induced by a norm or when a linear space X equipped with a metric d become a normed linear space? It's not hard to prove the following proposition, which shows that a metric d on linear space X is induced by a norm if and only if d is *translation-invariant* and *absolutely homogeneous*.

Exercise 1.1. If d is a metric on a linear space X satisfying for all $x, y, z \in X$ and $\lambda \in \mathbb{F}$,

(a) $d(x, y) = d(x + z, y + z)$ (translation invariance)

(b) $d(\lambda x, \lambda y) = |\lambda| d(x, y)$ (absolute homogeneity)

then $\|x\| \coloneqq d(x, 0)$ defines a norm on X and d is induced by this norm.

In the following proposition, we collect some elementary but fundamental facts baout normed linear spaces.

Exercise 1.2. $(X, \|\cdot\|)$ is a normed linear space over $\mathbb{F}$. Let $(x_n)_{n\geq 1}, (y_n)_{n\geq 1}$ be sequences in X and $(\lambda_n)_{n\geq 1}$ be a sequence in $\mathbb{F}$.

(a) Norm is a continuous funcrion, i.e., $x_n \to x$ (with respect to metric d) implies $\|x_n\| \to \|x\|$.

(b) Vector addition and scalar multiplicationis are continuous, i.e., $x_n \to x$ and $y_n \to y$ implies $x_n + y_n \to x + y$. is continuous, and $\lambda_n \to \lambda$ and $x_n \to x$ implies $\lambda_n x_n \to \lambda x$.

Definition 1.2. A normed linear space is complete with respect to the metric induced by the norm is called a *Banach space*.

There are two types of properties of a Banach space: those that are topological and those that are metric. The metric properties depend on the precise norm (such as uniformly convex space); the topological ones depend only on the equivalence class of norms.

Definition 1.3. Let $(x_n)_{n=1}^{\infty}$ be a sequence in $(X, \|\cdot\|)$. If the sequence $(s_n)_{n=1}^{\infty}$ of partial sums, where $s_n = \sum_{k=1}^{n} x_k$ for each $n \in \mathbb{N}$, converges to s, then we say the series $\sum_{k=1}^{\infty} x_k$ *converges* and its *sum* is s. In this case we write $\sum_{k=1}^{\infty} x_k = s$. The series $\sum_{k=1}^{\infty} x_k$ is said to be *absolutely convergent* if $\sum_{k=1}^{\infty} \|x_k\| < \infty$.

Theorem 1.1. $(X, \|\cdot\|)$ *is a Banach space if and only if every absolutely convergent series in* X *is convergent.*

Proof. Let X be a Banach space and suppose that $\sum_{j=1}^{\infty} \|x_j\| < \infty$. For any $\epsilon > 0$ and $n \in \mathbb{N}$, let $s_n = \sum_{j=1}^{n} x_j$. Let $K = K_\epsilon$ be a positive integer such that $\sum_{j=K+1}^{\infty} \|x_j\| < \epsilon$. Then, for all $m > n > K$, we have

$$\|s_m - s_n\| = \left\| \sum_{n+1}^{m} x_j \right\| \le \sum_{n+1}^{\infty} \|x_j\| \le \sum_{K+1}^{\infty} \|x_j\| < \epsilon\,.$$

Hence the sequence (s_n) of partial sums forms a Cauchy sequence in X. since X is complete, the sequence (s_n) converges to some element $s \in X$. That is, the series $\sum_{j=1}^{\infty} x_j$ converges.

Conversely, assume that $(X, \|\cdot\|)$ is a normed linear space in which every absolutely convergent series converges. Let (x_n) be a Cauchy sequence in X. Then there is an sequence (x_{n_k}) such that for each $k \in \mathbb{N}$, $\left\|x_{n_{k+1}} - x_{n_k}\right\| < 2^{-k}$. Then $\sum_{k=1}^{\infty} \left\|x_{n_{k+1}} - x_{n_k}\right\| < \infty$. By our assumption, the series $\sum_{k=1}^{\infty}(x_{n_{k+1}} - x_{n_k})$ is convergent to some $s \in X$. It follows that as $j \to \infty$

$$s_j = \sum_{k=1}^{j} (x_{n_k+1} - x_{n_k}) = x_{n_j+1} - x_{n_1} \to s\,.$$

Thus, the subsequence (x_{n_k}) of (x_n) converges in X. But if a Cauchy sequence has a convergent subsequence, then then the sequence itself also converges to the same limit as the subsequence. Hence X is complete. $\square$

1.1.1 Examples

We give some examples of normed linear spaces and Banach spaces. We always denote by $\mathbb{N}$ all the positive integers, and by $\mathbb{N}_0$ all the non-negative integers.

Example 1.1. Let $n \in \mathbb{N}$, for each $x = (x_1, x_2, \cdots, x_n) \in \mathbb{F}^n$, define

$$\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}, \text{ for } 1 \leq p < \infty$$
$$\|x\|_\infty := \max_{1 \leq i \leq n} |x_i|$$

Then $(\mathbb{F}^n, \|\cdot\|_p)$ and $(\mathbb{F}^n, \|\cdot\|_\infty)$ are Banach spaces.

Example 1.2. Let $M_n(\mathbb{F})$ be the linear space of all $n \times n$ matrices over $\mathbb{F}$. For each $A = (a_{ij}) \in M_n(\mathbb{F})$, define

$$\|A\|_{tr} := \left(\sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|^2\right)^{\frac{1}{2}}$$
$$\|A\|_\infty := \max_{1 \leq i,j \leq n} |a_{ij}|$$

Then $(M_n(\mathbb{F}), \|\cdot\|_{tr})$ and $(M_n(\mathbb{F}), \|\cdot\|_\infty)$ are normed linear spaces.

Example 1.3. For each $x = (x_i)_{i \in \mathbb{N}} \in \ell^p = \ell^p(\mathbb{N}, \mathbb{F})$, where $1 \leq p < \infty$, define

$$\|x\|_p := \left(\sum_{i \in \mathbb{N}} |x_i|^p\right)^{\frac{1}{p}}$$

Then $(\ell^p, \|\cdot\|_p)$ is a normed linear space over $\mathbb{F}$.

Example 1.4. Let ℓ_0 be the set of all sequences $x = (x_i)_{i \in \mathbb{N}}$ of real or complex numbers such that $x_i = 0$ for all but finitely many incluses i. Let c be the set of all convergent sequences $x = (x_i)_{i \in \mathbb{N}}$ of real or complex numbers. Let c_0 is the set of all convergent sequences $x = (x_i)_{i \in \mathbb{N}}$ of real or complex numbers which converge to 0. Suppose that $X = \ell_\infty, \ell_0, c$ or c_0. For each For each $x = (x_i)_{i\in\mathbb{N}} \in X$, define

$$\|x\|_\infty \coloneqq \max_{i\in\mathbb{N}} |x_i|$$

Then ℓ_∞, c, c_0 is Banach space, but ℓ_0 is an incomplete subspace of ℓ_∞.

Example 1.5. Let X be any Hausdorff space [1] and let $C_b(X, \mathbb{F})$ be all bounded continuous functions $f : X \to \mathbb{F}$. Define

$$\|f\|_\infty \coloneqq \sup_{x\in X} |f(x)| .$$

Then $C_b(X, \mathbb{F})$ is a Banach space over $\mathbb{F}$. Particularly, when X is compact, let $C(X, \mathbb{F})$ be the set of all continuous $\mathbb{F}$-valued functions, then $C(X, \mathbb{F}) = C_b(X, \mathbb{F})$.

Example 1.6. Let X be a locally compact Hausdorff (LCH) space and $C_0(X, \mathbb{F})$ be all continuous functions $f : X \to \mathbb{F}$ such that that vanish at infinity: For any $\epsilon > 0, \{|f| \geq \epsilon\}$ is compact in X. Then for each f, Define

$$\|f\|_\infty \coloneqq \sup_{x\in X} |f(x)| .$$

then $C_0(X, \mathbb{F})$ is a closed subspace of $C_b(X, \mathbb{F})$ and hence is a Banach space. If X is compact, $C_0(X, \mathbb{F}) = C_b(X, \mathbb{F}) = C(X, \mathbb{F})$.

To prove this, observe that $C_0(X, \mathbb{F})$ is a linear subspace in $C_b(X, \mathbb{F})$. It will only be shown that $C_0(X, \mathbb{F})$ is closed in $C_b(X, \mathbb{F})$. Let $\{f_n\} \subset C_0(X, \mathbb{F})$ and suppose $f_n \to f$ in $C_b(X, \mathbb{F})$. Given $\epsilon > 0$, there is an integer n such that

$$\|f_n - f\| < \frac{\epsilon}{2}$$

Thus,

$$\{|f| \geq \epsilon\} \subset \left\{|f_n| \geq \frac{\epsilon}{2}\right\}$$

so that $f \in C_0(X, \mathbb{F})$.

[1] All topological spaces in this notes are assumed to be Hausdorff unless the contrazy is specified

Remark 1.2. Let X be LCH space and let $X_\infty = X \cup \{\infty\}$ be the one-point compactification of X. Then one can show that $\{f \in C(X_\infty, \mathbb{F}) : f(\infty) = 0\}$, with the norm it inherits as a subspace of $C(X_\infty, \mathbb{F})$, is *isometrically isomorphic* to $C_0(X, \mathbb{F})$. Paricularly,

- $C_0(\mathbb{R}, \mathbb{F})$ = all of the $\mathbb{F}$-valued continuous functions f such that $|f(x)| \to 0$ as $|x| \to \infty$.

- Let $\mathbb{N}$ equipped with discrete topology, then $C_0(\mathbb{N}, \mathbb{F}) = c_0$.

Exercise 1.3. Let X be LCH and define $C_c(X, \mathbb{F})$ to be the continuous functions $f : X \to \mathbb{F}$ such that $\operatorname{supp}(f)$ (the closure of $\{f \neq 0\}$) is compact. Show that $C_c(X, \mathbb{F})$ is dense in $C_0(X, \mathbb{F})$.

The next example is usually proved in courses on integration and no proof is given here.

Example 1.7. For $1 \leq p \leq \infty$, $L^p(X, \mathcal{F}, \mu; \mathbb{F})$ is Banach space.

Example 1.8. Let I be a set and $1 \leq p < \infty$. Define $\ell^p = \ell^p(I, \mathbb{F})$ to be the set of all functions $f : I \to \mathbb{F}$ such that $\sum_{\alpha \in \Lambda} |f(i)|^p < \infty$, and define

$$\|f\|_p = \left(\sum_{\alpha \in \Lambda} |f(i)|^p \right)^{1/p}.$$

Then $\ell^p(I)$ is a Banach space. If $I = \mathbb{N}$, then $\ell^p(\mathbb{N})$ is often denoted as ℓ^p. In fact, let 2^I is all the subset of I and $\#$ be counting measure on $(I, 2^I)$. We have $\ell^p(I, \mathbb{F}) = L^p(I, 2^I, \#; \mathbb{F})$.

Example 1.9. Let $n \geq 1$ and let $C^{(n)}[0,1] = C^{(n)}([0,1], \mathbb{R})$ be the collection of functions $f : [0,1] \to \mathbb{R}$ such that f has n continuous derivatives. Define

$$\|f\| = \sum_{k=0}^{n} \sup_{x \in [0,1]} |f^{(k)}(x)|.$$

Then $C^{(n)}([0,1], \mathbb{F})$ is a Banach space.

Example 1.10 (Sobolev space)**.** Let $1 \leq p < \infty$ and $n \in \mathbb{N}$ and let $W_p^n[0,1] = W_p^n([0,1], \mathbb{R})$ be all the functions $f : [0,1] \to \mathbb{R}$ such that f has $n-1$ continuous derivatives, $f^{(n-1)}$ is absolutely continuous, and $f^{(n)} \in L^p[0,1]$. For f in $W_p^n[0,1]$, define

$$\|f\| = \sum_{k=0}^{n} \left[\int_0^1 \left| f^{(k)}(x) \right|^p dx \right]^{1/p}$$

Then $W_p^n[0,1]$ is a Banach space.

1.1.2 Separability and Schauder Bases

A subset S of a normed linear space $(X, \|\cdot\|)$ is said to be *dense* in X if $\overline{S} = X$. If $(X, \|\cdot\|)$ contains a countable dense subset S, we say $(X, \|\cdot\|)$ is *separable*.

Lemma 1.2. *X is separable if and only if it contains a countable set S such that $\overline{\operatorname{span}}(S) = X$.*

Proof. It suffices to show that if there is a countable subset S of X so that $\overline{\operatorname{span}}(S) = X$, then X is separable. To this end, let $\mathbb{K}$ be a countable dense subset of $\mathbb{F}$ and define

$$A = \left\{ \sum_{i=1}^{n} \lambda_i x_i : n \geq 1, \lambda_i \in \mathbb{K}, x_i \in S \right\}.$$

Clearly A is countable dense subset of $\operatorname{span}(S)$. Hence A is a countable dense subset of X. □

Example 1.11. $\mathbb{R}, \mathbb{C}$ are separable, ℓ^p is separable, ℓ_∞ is not separable. To see the last one, consider

$$\{(a_n) : a_n \in \{0,1\} \text{ for all } n \in \mathbb{N}\}.$$

Definition 1.4. A sequence $(e_n)_{n\in\mathbb{N}}$ in a separable Banach space is called a *Schauder basis* if, for any $x \in X$, there is a *unique* sequence $(\alpha_n(x))$ of scalars such that

$$x = \sum_{n=1}^{\infty} \alpha_n(x) e_n \, .$$

It is clear from this definition that if $(e_n)_{n\in\mathbb{N}}$ is a Schauder basis, then $\overline{\operatorname{span}}\{e_n : n \in \mathbb{N}\} = X$ Uniqueness of the expansion clearly implies that the set $\{e_n : n \in \mathbb{N}\}$ is linear independent.

Example 1.12.

- The sequence $(e_n)_{n\in\mathbb{N}}$ where $e_n = (\delta_{nm})_{m\in\mathbb{N}}$ is a Schauder basis for ℓ^p.
- The sequence $(e_n)_{n\in\mathbb{N}}$ where $e_n = (\delta_{nm})_{m\in\mathbb{N}}$ is a Schauder basis for c_0.
- The sequence $\{e\} \cup (e_n)_{n\in\mathbb{N}}$ where $e = (1, 1, 1, \cdots)$ is a Schauder basis for c.
- ℓ^∞ has no Schauder basis.

Remark 1.3. In 1937, Per Enflo constructed a separeble Banach space with no Schauder basis.

1.1.3 The External Direct Sum of Banach Spaces

Now for the product or (external) direct sum of normed spaces. Here there is a difficulty because, unlike Hilbert space, there is no canonical way to proceed. Suppose $\{X_\alpha\}_{\alpha\in\Lambda}$ is a collection of normed vector spaces. Then $\Pi_{\alpha\in\Lambda} X_\alpha$ is a vector space if the linear operations are defined coordinate-wise. The idea is to put a norm on a linear subspace of this product.

Let $\|\cdot\|$ denote the norm on each X_α. For $1 \le p < \infty$, define

$$\oplus_p X_\alpha \coloneqq \left\{ x \in \prod_\alpha X_\alpha : \|x\| \coloneqq \left[\sum_\alpha \|x_\alpha\|^p \right]^{1/p} < \infty \right\} .$$

and

$$\oplus_\infty X_\alpha \coloneqq \left\{ x \in \prod_\alpha X_\alpha : \|x\| \coloneqq \sup_\alpha \|x_\alpha\| < \infty \right\}$$

If $\{X_1, X_2, \ldots\}$ is a sequence of normed spaces, define

$$\oplus_0 X_n \equiv \left\{ x \in \prod_{n=1}^{\infty} X_n : \|x_n\| \to 0 \right\}$$

give $\oplus_0 X_n$ the norm it has as a subspace of $\oplus_\infty X_n$.

Proposition 1.3. *Let $\{X_\alpha\}_{\alpha\in\Lambda}$ be a collection of normed spaces and $p \in [1, \infty]$. Denote $X = \oplus_p X_\alpha$, then*

(a) X is a normed space, and X is a Banach space if and only if each X_α is a Banach space.

(b) The canonical projection $P_\alpha : X \to X_\alpha$ is a continuous open linear map with $\|P_\alpha x\| \le \|x\|$ for each x in X.

Proof. Observe that for $x = (x_\alpha)$ and $y = (y_\alpha)$ in X, by Minkovski inequality,

$$\begin{aligned} \|x+y\| &= \left(\sum_\alpha \|x_\alpha + y_\alpha\|^p \right)^{1/p} \le \left(\sum_\alpha (\|x_\alpha\| + \|y_\alpha\|)^p \right)^{1/p} \\ &\le \left(\sum_\alpha \|x_\alpha\|^p \right)^{1/p} + \left(\sum_\alpha \|y_\alpha\|^p \right)^{1/p} = \|x\| + \|y\| \,. \end{aligned}$$

It's now easy to see that X is a normed vector space. Obviously, if X is complete then each X_α is. On the other hand, suppose each X_α is complete and $\{x^{(n)}\}$ a Cauchy sequence in X. Then for every $\epsilon > 0$, there exists a positive integer $N = N_\epsilon > 0$ so that

$$\|x^{(n)} - x^{(m)}\| = \left(\sum_\alpha \|x_\alpha^{(n)} - x_\alpha^{(m)}\|^p \right)^{1/p} \le \epsilon \,.$$

Thus for each $\alpha \in \Lambda$, $\{x_\alpha^{(n)}\}$ is a Cauchy sequence in X_α. Suppose $x_\alpha^{(n)} \to x_\alpha$ in X_α. Given any fixed finite subset S of Λ, for $n \geq N_\epsilon$, we have

$$\left(\sum_{\alpha \in S} \|x_\alpha^{(n)} - x_\alpha\|^p \right)^{1/p} = \lim_{m \to \infty} \left(\sum_{\alpha \in S} \|x_\alpha^{(n)} - x_\alpha^{(m)}\|^p \right)^{1/p} \leq \epsilon \, .$$

Taking the limit, we get

$$\left(\sum_{\alpha \in \Lambda} \|x_\alpha^{(n)} - x_\alpha\|^p \right)^{1/p} \leq \epsilon \, , \qquad \text{for } n \geq N_\epsilon \, .$$

Thus $x = (x_\alpha) \in X$ and $x^{(n)} \to x$. So X is a Banach space. Part (b) is trivial, so we omit the proof. $\square$

Remark 1.4. A similar result holds for $\oplus_0 X_n$, but the formulation and proof of this is the same as before.

1.2 Bounded Linear Operators

In this section, we always assume X and Y are two vector space over the same field $\mathbb{F}$.

Definition 1.5. A linear operator A on $D(A) \subset X$ into Y is said to be *bounded* if there exists a constant $M > 0$ such that

$$\|Ax\| \leq M\|x\|,^2 \text{ for all } x \in D(A).$$

Obviously, A is bounded if and only if A maps a bounded set into a bounded set. Moreover, boundedness of linear operator in the sense of Definition 1.5 is equivalent to the continuity. The proof is easy so we omit it.

[2] It should be emphasised that the norm on the left side is in Y and that on the right side is in X .

Proposition 1.4. *Let A be a linear operator on $D(A)$ into Y. Then the following statements are equivalent: (i) A is bounded. (ii) A is continuous on $D(A)$. (iii) A is continuous at some point x_0 in $D(A)$.*

Definition 1.6. Let A be a linear operator on $D(A)$ into Y. The *operator norm* (or simply norm) of A, denoted by $\|A\|$, is defined as

$$\|A\| := \inf\{M : \|Tx\| \le M\|x\|,\ \forall\, x \in D(A)\}\,.$$

As we can see, the definition of $\|A\|$ can be change as the following equations:

$$\|A\| = \sup_{x\neq 0} \frac{\|Tx\|}{\|x\|} = \sup\{\|Tx\| : \|x\| = 1\} = \sup\{\|Tx\| : \|x\| \le 1\}\,.$$

We shall denote by $\mathcal{B}(X, Y)$ the set of all bounded linear operators defined on X taking values in Y. We shall write $\mathcal{B}(X)$ for $\mathcal{B}(X, X)$. It's easy to check that the function $\|\cdot\|$ defined above is a norm on $\mathcal{B}(X, Y)$, so $\mathcal{B}(X, Y)$ become a normed linear space. So, it's natural to ask: When it become a Banach space?

Theorem 1.5. *$\mathcal{B}(X, Y)$ is a Banach space if and only if Y is a Banach space.*

Proof. Sufficiency. Let $\{T_n\}$ be a Cauchy sequence in $\mathcal{B}(X, Y)$. Then for any $x \in X$, $\{T_n x\}$ is Cauchy sequence in Y. So we can define $T : X \to Y$ by letting

$$Tx := \lim_{n\to\infty} T_n x \quad \text{ for } x \in X\,.$$

Evidently, $T \in \mathcal{B}(X, Y)$. To show that $T_n \to T$ in $\mathcal{B}(X, Y)$, note that for all $x \le 1$,

$$\|T_n x - Tx\| = \lim_{m\to\infty} \|T_n x - T_m x\| \le \liminf_{m\to\infty} \|T_n - T_m\|\,.$$

Since $\{T_n\}$ is a Cauchy sequence, we can see that $T_n \to T$.

Necessity. Let $\{y_n\}$ be a Cauchy sequence in Y. Let ℓ be a continuous non-zero linear functional on X. Define $T_n : X \to Y$ by letting

$$T_n x \coloneqq \ell(x) y_n \quad \text{ for } x \in X\,.$$

Claerly $T_n \in \mathcal{B}(X, Y)$, and $\|T_n - T_m\| \le \|\ell\| \, \|y_n - y_m\|$, so $\{T_n\}$ is a Cauchy sequence, and there exists $T \in \mathcal{B}(X, Y)$ so that $T_n \to T$ in $\mathcal{B}(X, Y)$. Choose $x_0 \in X$ so that $\ell(x_0) = 1$, then

$$T_n x_0 = y_n \to T x_0\,.$$

So Y is complete. □

Remark 1.5. Let X be a normed linear space and Y be a Banach space, Denote $\tilde{X}$ be the completion of X. Define $\rho : \mathcal{B}(\tilde{X}, Y) \to \mathcal{B}(X, Y)$ by $\rho(A) = A|_X$, then ρ is an isometric isomorphism.

Let $T : X \to Y$ and $S : Y \to Z$. We define the composition of T and S as the map $ST : X \to Z$ defined by

$$(ST)(x) = (S \circ T)(x) = S(T(x)) \text{ for all } x \in X.$$

Exercise 1.4. Let X, Y be normed vector spaces. Let $T \in \mathcal{B}(X, Y), S \in \mathcal{B}(Y, Z)$. Then $ST \in \mathcal{B}(X, Z)$ and $\|ST\| \le \|S\|\|T\|$.

1.2.1 Examples

We give some examples of normed linear spaces to end the section.

Example 1.13. Let $X = \mathbb{F}^n$ with the norm $\|\cdot\|_\infty$ and $A = (\alpha_{ij}) \in M_n(\mathbb{F})$. For $x = (x_1, x_2 \cdots, x_n)' \in \mathbb{F}^n$, define $T : X \to X$ by

$$Ax = A(x_1, x_2 \cdots, x_n)' = \left(\sum_{j=1}^n \alpha_{1j} x_j, \sum_{j=1}^n \alpha_{2j} x_j \cdots, \sum_{j=1}^n \alpha_{nj} x_j \right)'.$$

Then $A \in \mathcal{B}(X)$ and $\|A\| = \sup\limits_{1 \le i \le n} \sum\limits_{j=1}^n |\alpha_{ij}|$.

Example 1.14. Define an operator $L : \ell^2 \to \ell^2$ by

$$Lx = L(x_1, x_2, x_3, \cdots) = (x_2, x_3, \cdots)$$

Then $L \in \mathcal{B}(\ell^2)$ and $\|L\| = 1$. The operator L is called the *shift operator*.

Example 1.15. Let $X = \mathcal{P}[0,1]$, the set of polynomials on the interval $[0,1]$ with the uniform norm $\|\cdot\|_\infty$. For each $x \in X$, define $A : X \to X$ by

$$Tx = x'(t), \text{ for all } t \in [0,1]$$

Then A is a linear operator but NOT bounded.

Example 1.16 (Multiplication Operator)**.** Let $(X, \mathcal{F}, \mu)$ be a σ-finite measure space and $\phi \in L^\infty(\mu)$. Let $1 \le p \le \infty$. Define $M_\phi : L^p(\mu) \to L^p(\mu)$, by

$$M_\phi f = \phi f \text{ for all } f \in L^p(\mu)\,.$$

Then $M_\phi \in \mathcal{B}(L^p(\mu))$ and $\|M_\phi\| = \|\phi\|_\infty$. The operator M_ϕ is called a *multiplication operator*. The function ϕ is it's *symbol*. To see this, it's easy to see that if $f \in L^p(\mu)$, then

$$\int |\phi f|^p d\mu \le \|\phi\|_\infty^p \int |f|^p d\mu.$$

Thus, $M_\phi \in \mathcal{B}(L^p(\mu))$ and $\|M_\phi\| \le \|\phi\|_\infty$. On the other hand (assume that $\|\phi\|_\infty > 0$), for any $\epsilon > 0$, the σ-finiteness of the measure space implies that there is a set B in $\mathcal{F}$, such that $0 < \mu(B) < \infty$ and

$$|\phi(x)| \ge \|\phi\|_\infty - \epsilon \quad \text{ for } x \in B\,.$$

Let $f = \mu(B)^{-1/p}\chi_B$, then $f \in L^p(\mu)$ and $\|f\|_p = 1$. So

$$\|M_\phi\|^p \ge \|\phi f\|_p^p = (\mu(B))^{-1} \int_B |\phi|^p d\mu \ge (\|\phi\|_\infty - \epsilon)^p\,.$$

Letting $\epsilon \to 0$, we get that $\|M_\phi\| \ge \|\phi\|_\infty$. Thus $\|M_\phi\| = \|\phi\|_\infty$.

Remark 1.6. We should note that if the measure space $(X, \mathcal{F}, \mu)$ is not σ-finite, then the conclusion is not necessarily valid. Indeed, let $\mathcal{F}$ be the Borel subsets of $[0,1]$ and define μ on $\mathcal{F}$ by

$$\mu(A) = \begin{cases} \lambda(A), & 0 \notin A. \\ \infty, & 0 \in A. \end{cases}$$

where λ is the Lebesgue measure. This measure μ has an infinite *atom* at 0 and therefore, is not σ-finite. Let $\phi = \chi_{\{0\}}$. Then $\phi \in L^\infty(\mu)$ and $\|\phi\|_\infty = 1$. We claim that $M_\phi = 0$, so $\|M_\phi\| = 0 < 1 = \|\phi\|_\infty$.

To see this, note that for any $f \in L^p(\mu)$, $f(0) = 0$, then $M_\phi f = f(0)\chi_{\{0\}} = 0$.

Example 1.17 (Integral Operator I)**.** Let $(X, \mathcal{F}, \mu)$ be a measure space. Suppose $k \in L^2(X \times X, \mathcal{F} \times \mathcal{F}, \mu \times \mu)$. Define $K : L^2(\mu) \to L^2(\mu)$ by

$$(Kf)(x) = \int k(x,y) f(y) \mu(dy)\,, \text{ for } x \in X\,.$$

Then K is a bounded linear operator (in fact K is compact, see Example 7.6) with

$$\|K\| \leq \|k\|_{L^2}\,.$$

K is called an *integral operator* and the function k is called its *kernel*. (There exists other conditions on the kernel implying that K is bounded.) To see this, it must be shown that $Kf \in L^2(\mu)$, but it will follow from the argument that demonstrates the boundedness of K. For $f \in L^2(\mu)$, using the Cauchy-Schwartz inequality, we have

$$\begin{aligned} \|Kf\|^2 &= \int \left| \int k(x,y) f(y) \mu(dy) \right|^2 \mu(dx) \\ &\leq \int \left[\int |k(x,y)|^2 \mu(dy) \int |f(y)|^2 \mu(dy) \right] \mu(dx) = \|k\|_{L^2}^2 \|f\|^2\,. \end{aligned}$$

A particular example of an integral operator is the Volterra operator defined below.

Example 1.18. Let $k : [0,1] \times [0,1] \to \mathbb{R}$ be the characteristic function of $\{(x,y) : y < x\}$. The corresponding operator $V : L^2(0,1) \to L^2(0,1)$ defined by $Vf(x) = \int_0^1 k(x,y)f(y)dy$ is called the *Volterna operator*. Note that

$$Vf(x) = \int_0^x f(y)dy$$

Example 1.19 (Integral Operator II)**.** Let $(X, \mathcal{F}, \mu)$ be a measure space and suppose $k : X \times X \to \mathbb{F}$ is an $\mathcal{F} \times \mathcal{F}$-measurable function for which there are constants C_1 and C_2 such that

$$\int |k(x,y)|\mu(dy) \le C_1\,,\ \mu\text{-a.e. } x\,.$$
$$\int |k(x,y)|\mu(dx) \le C_2\,,\ \mu\text{-a.e. } y\,.$$

Define $K : L^p(\mu) \to L^p(\mu)$ $(1 < p < \infty)$ by

$$(Kf)(x) = \int f(y)k(x,y)\mu(dy)\,,$$

then K is a bounded linear operator and

$$\|K\| \le C_1^{1/q} C_2^{1/p},\ \text{where } \frac{1}{p} + \frac{1}{q} = 1\,.$$

This operator is also called an *integral operator* and the function k is called its *kernel*. (There exists other conditions on the kernel imply that K is bounded.) To show this, for $f \in L^p(\mu)$,

$$\begin{aligned}
|Kf(x)| &\le \int |k(x,y)||f(y)|\mu(dy) = \int |k(x,y)|^{1/q}|k(x,y)|^{1/p}|f(y)|\mu(dy) \\
&\le \left[\int |k(x,y)|\mu(dy)\right]^{1/q} \left[\int |k(x,y)||f(y)|^p\mu(dy)\right]^{1/p} \\
&\le C_1^{1/q} \left[\int |k(x,y)||f(y)|^p\mu(dy)\right]^{1/p}
\end{aligned}$$

Hence

$$\begin{aligned}
\int |Kf(x)|^p\mu(dx) &\le C_1^{p/q} \iint |k(x,y)||f(y)|^p\mu(dy)\mu(dx) \\
&= C_1^{p/q} \int |f(y)|^p\mu(dy) \int |k(x,y)|\mu(dx) \\
&\le C_1^{p/q} C_2\|f\|^p
\end{aligned}$$

Thus $\|K\| \le C_1^{1/q} C_2^{1/p}$.

1.3 Finite Dimensional Space

In this section we will show two surprising results. Firstly, there norm on finite dimensional space is unique in the sense of equivalence. Secondly, the completeness of the unit ball implies the dimension of space is finite. Now let us make the "equivalence" clear.

Definition 1.7. If $\|\cdot\|_1$ and $\|\cdot\|_2$ are two norms on X, they are said to be *equivalent* if they define the same topology on X.

We will give a much direct condition to judge equivalence.

Lemma 1.6. *Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be two norms on X, then they are equivalent iff there are positive constants C_1 and C_2 such that*

$$C_1\|x\|_1 \le \|x\|_2 \le C_2\|x\|_1 \text{ for all } x \in X\,.$$

Proof. Note that the two topology coincides iff, for any sequence (x_n) and point x in X,

$$\|x_n - x\|_1 \to 0 \Leftrightarrow \|x_n - x\|_2 \to 0\,. \qquad \square$$

Suppose that X is a finite-dimensional linear space over $\mathbb{F}$. Let $\{e_1, \cdots, e_n\}$ be a basis of X. Then we can define a "natural" norm on X. For $x = \sum_{j=1}^n \alpha_j e_j$, let

$$|x| \coloneqq \sum_{j=1}^n |\alpha_j|\,.$$

It's easy to check that $|\cdot|$ is a norm on X. A surprisingly result told us, this norm is the unique norm on X in the sense of equivalence.

Theorem 1.7. *All the norms on X are equivalent.*

Proof. Let $\|\cdot\|$ be a norm on X. We shall show that $|\cdot|$ is equivalent to $\|\cdot\|$. Clearly, $S = \left\{(\alpha_1, \cdots, \alpha_n) \in \mathbb{F}^n : \sum_{j=1}^n |\alpha_j| = 1\right\}$ is compact subset of $\mathbb{F}^n$. We define $f : S \to \mathbb{R}$ by

$$f(\alpha_1, \alpha_2, \cdots, \alpha_n) \coloneqq \left\| \sum_{j=1}^n \alpha_j e_j \right\| .$$

Clearly, f is continuous on S. Since S is compact, f attains its minimum and maximum on S, which is positive. Thus there exists a sonstant C_1, C_2 so that for all x

$$C_1 |x| \le \|x\| \le C_2 |x| .$$

Hence we complete the proof. $\square$

By the theorem above, we can find that

Corollary 1.8. *Every finite-dimensional normed linear space $(X, \|\cdot\|)$ is Banach space.*

Corollary 1.9. *In a finite-dimensional normed linear space $(X, \|\cdot\|)$, a subset $K \subset X$ is compact if and only if it is closed and bounded.*

Corollary 1.10. *Let X and Y be normed linear spaces and $\dim(X) < \infty$. Then $\mathcal{B}(X, Y) = \mathcal{L}(X, Y)$.*

Next, we shall give a topological characterization of the algebraic concept of finite dimensionality. The following lemma is needed, and we will need it when discussing compact operator.

Lemma 1.11 (Riesz's Lemma)**.** *Let M be a poper closed subspace of $(X, \|\cdot\|)$. Then for each $\epsilon \in (0,1)$, there is an element $y \in X$, depending on ϵ, such that*

$$\|y\| = 1 \quad \text{and} \quad d(y, M) > 1 - \epsilon .$$

The element y is thus "nearly orthogonal" to M.

Proof. Choose $x \in M^c$ and denote $d = d(x, M)$. $d > 0$ since M is closed. Then for any given $\epsilon > 0$, there is a $m \in M$ such that

$$d \leq \|x - m\| < (1 + \epsilon)d\,.$$

Then for any $u \in M$,

$$\left\| \frac{x - m}{\|x - m\|} - u \right\| \geq \frac{d}{\|x - m\|} > \frac{d}{(1 + \epsilon)d} > 1 - \epsilon\,.$$

Let $y = \frac{x-m}{\|x-m\|}$, we have completed the proof. $\square$

Second proof. By Corollary 1.34, one can find a non-trivial linear functional ℓ on X which vanishes on M with $\|\ell\| = 1$. By definition of the operator norm $\|\ell\|$ of ℓ, one can find a unit vector x such that $|\ell(x)| \geq 1 - \epsilon$. The claim follows from that $d(x, M) = \|\widetilde{x}\| \geq |\widetilde{\ell}\widetilde{x}| = |\ell(x)| \geq 1 - \epsilon$. $\square$

We now give a topological characterization of the algebraic concept of finite dimensionality.

Theorem 1.12. *$(X, \|\cdot\|)$ is finite-dimensional if and only if its closed unit ball $B_X = \{x : \|x\| \leq 1\}$ is compact.*

Proof. Assume that the closed unit ball $B_X = \{x \in X : \|x\| \leq 1\}$ is compact. Then B_X is totally bounded. Hence there is a finite $\frac{1}{2}$-net

$$\{x_1, x_2, \ldots, x_n\} \subset B_X\,.$$

Let $M = \operatorname{span}\{x_1, x_2, \ldots, x_n\}$. Then M is a finite-dimensional linear subspace of X and hence closed.

If M is a proper subspace of X, then, by Riesz's lemma, there is an element $x_0 \in B_X$ such that $d\,(x_0, M) > \frac{1}{2}$. In particular, $\|x_0 - x_k\| > \frac{1}{2}$ for all $k = 1, 2, \ldots, n$. However this contradicts the fact that $\{x_1, x_2, \ldots, x_n\}$ is a $\frac{1}{2}$-net in B_X. Hence $M = X$ and, consequently, X is finite-dimensional. $\square$

To give an example of a relatively compact infinite subset of a Banach space of infinite dimension, we shall give the

Theorem 1.13 (Ascoli-Arzelá)**.** *Let (S, d) be a compact metric space, and $C(S)$ the Banach space of $\mathbb{F}$-valued continuous functions on S endowed with supremum norm. Then a subset $F \subset C(S)$ is relatively compact in $C(S)$ if the following two conditions are satisfied:*

(a) *F is (uniformly) bounded, i.e.,* $\sup_{x\in F} \|x\| < \infty$*;*

(b) *F is equi-continuous, i.e.,*

$$\lim_{\delta\downarrow 0} \sup_{x\in F} \{\left|x\left(s'\right) - x\left(s''\right)\right| : d\left(s', s''\right) \le \delta\} = 0\,.$$

1.4 Quotient Space

Assume that M is a subspace of X. We learned the quotinent space X/M in the course of linear algebra. We denoe by $\widetilde{x}$ the element in X/M, i.e., for $x \in X$,

$$\widetilde{x} \coloneqq x + M = \{x + m : m \in M\}\,.$$

Next, we shall make X/M become a normed vector space.

Theorem 1.14. *Let M be a closed linear subspace of a normed linear space X over $\mathbb{F}$ .Then the quotient space X/M is a normed linear space with respect to the quotient norm defined by*

$$\|\widetilde{x}\| \coloneqq \inf_{y\in\widetilde{x}} \|y\| = d(x, M)\ ,\ \ \textit{where } \widetilde{x} \in X/M. \tag{1.1}$$

Proof. Observe that

$$\|\widetilde{x}\| = \inf_{y\in\widetilde{x}} \|y\| = \inf_{m\in M} \|x + m\| = \inf_{m\in M} \|x - m\| = d(x, M)$$

Then it's easy to check the three conditions which makes $\|\cdot\|$ be a norm. In fact, since M is closed, we have

$$\|\widetilde{x}\| = 0 \Leftrightarrow d(x, M) = 0 \Leftrightarrow x \in M \Leftrightarrow \widetilde{x} = \widetilde{0}\,.$$

Since the metric d is absolutely homogeneous,

$$\|\lambda\widetilde{x}\| = d(\lambda x, M) = |\lambda| d(x, M) = |\lambda|\|\widetilde{x}\| \,.$$

By the trigonometric inequality and translation invaruance of d,

$$\|\widetilde{x} + \widetilde{y}\| = d(x + y, M) \leq d(x, M) + d(y, M) = \|\widetilde{x}\| + \|\widetilde{y}\|.$$

So the desired result follows. □

Remark 1.7. From the proof above, we should note that if M is a subspace but not closed, then (1.1) is only a seminorm.

Let M be a closed subspace of the normed linear space X. The mapping $Q = Q_M : X \to X/M$ defined by

$$Q_M(x) \coloneqq \widetilde{x} \quad \text{for all } x \in X \,,$$

is called the *quotient map* (or *natural map*) of X onto X/M.

Theorem 1.15. *M be a proper closed subspace of normed linear space X, Q is the quotinent map. Then*

(a) Q is continuous linear operator and $\|Q\| = 1$.

(b) Q is an open surjective map.

(c) The topology induced by quotient norm on X/M coincides with the quotient topology on X/M. In other words, a subset U of X/M is open if and only if $Q^{-1}(U)$ is open.

Proof. (a). Clearly Q is linear. By the definition of quotinent norm, $\|Q\| \leq 1$. By Lemma 1.11, $\|Q\| = 1$.

(b). Clearly Q is surjective. Fix a open subset G, and for any $\widetilde{x} \in Q(G)$, without loss of generality, we can asuume $x \in G$. Then there exists $\delta > 0$ so that $B_X(x, \delta) \subset G$. We shall show that $B_{X/M}(\widetilde{x}, \delta) \subset Q(G)$, and thus

$Q(G)$ is open. To this end, for any $\widetilde{y} \in B_{X/M}(\widetilde{x}, \delta)$, there exists $m \in M$ so that $\|y - m - x\| < \delta$. So $y - m \in B_X(x, \delta)$ and $\widetilde{y} = Q(y - m) \in Q(G)$.

(c) If U is open, $Q^{-1}(U)$ is open since Q is continuous. If $Q^{-1}(U)$ is open, then $U = Q(Q^{-1}(U))$ is open since Q is a open mapping. $\square$

Remark 1.8. In fact, for topology space X and Y, if $f : X \to Y$ is a continuous open mapping, then the topology on Y is the quotinent topology (with respect to f).

Corollary 1.16. *Let X be a normed vector space. Let M be a closed subspace and N be a finite dimensional subspace. Then $M + N$ is closed.*

Proof. Let Q_M be the quotient map. Observing that

$$M + N = Q_M^{-1}(Q_M(N)),$$

$Q_M(N)$ is a finite-dimensional subspace of X/M, and Q_M is continuous, the desired result follows. $\square$

Example 1.20. If $A : X \to Y$ is a linear operator between vector spaces X and Y, then it descends to a injective linear operator from the quotient space $X/N(A)$ to Y. That is,

$$\tilde{A} : X/N(A) \to Y\,;\, \tilde{x} \to Ax$$

It's easy to see that $\tilde{A}$ is well-defined and is an injective linear operater. Moreover, if X and Y are normed vector space and A is continuous, then so is $\tilde{A}$. Indeed,

$$\|\tilde{A}\| = \|A\|\,. \tag{1.2}$$

On the one hand, $\|\tilde{A}\tilde{x}\| = \|Ax\| \leq \|A\|\|x\|$ for all $x \in \tilde{x}$, and hence $\|\tilde{A}\| \leq \|A\|$. On the other hand, observe that $A = \tilde{A}Q$, where $Q : X \to X/N(A)$ is the quotinent map with $\|Q\| \leq 1$, we get $\|A\| \leq \|\tilde{A}\|\|Q\|$. Then (1.2) follows.

Theorem 1.17. *Let X be a Banach space. Let M be a subspace. Then the quotient space X/M, equipped with the quotient norm, is a Banach space.*

Proof. Let $(\widetilde{x_n})_{n\geq 1}$ be a sequence in X/M such that $\sum_{j=1}^{\infty} \|\widetilde{x_j}\| < \infty$. For each $j \in \mathbb{N}$, choose an element $m_j \in M$ such that

$$\|x_j - m_j\| \leq \|\widetilde{x_j}\| + 1/2^j.$$

It now follows that $\sum_{j=1}^{\infty} \|x_j - m_j\| < \infty$. Since X is a Banach space, the series $\sum_{j=1}^{\infty} (x_j - m_j)$ converges to some element $z \in X$. Since the quotient mapping is contious, the series $\sum_{j=1}^{\infty} \widetilde{x_j}$ converges to $\widetilde{z}$. Hence, every absolutely convergent series in X/M is convergent, and so X/M is complete. $\square$

1.5 Hahn-Banach Theorems

The Hahn-Banach Theorem is one of the most important results in mathematics. It is used so often it is rightly considered as a cornerstone of functional analysis. It is one of those theorems that when it or one of its immediate consequences is used, it is used without quotation or reference and we should to realize that it is being invoked.

1.5.1 Extension Theorems

Let X be a vector space over $\mathbb{F}$. A *positive homogeneous, subadditive* functional is a function $p : X \to \mathbb{R}$ satisfying the following properties.

(a) Positive homogeneity: $p(\alpha x) = \alpha p(x)$ for x in X and $\alpha \geq 0$.

(b) Subadditivity: $p(x + y) \leq p(x) + p(y)$ for all x, y in X.

Trivially, every seminorm is a positive homogeneous, subadditive functional, but not conversely. It should be emphasized that a positive homogeneous, subadditive functional is allowed to assume negative values and that (b) in the definition only holds for $\alpha \geq 0$.

First of all, we deal with the real vector spaces.

Theorem 1.18. *Let X be a vector space over $\mathbb{R}$. Let p be a positive homogeneous, subaddititive functional on X. Let Y be a subspace in X, and ℓ be a linear functional on Y. If ℓ satisfies*

$$\ell(y) \leq p(y), \text{ for all } y \in Y,$$

then ℓ can be extended to X as a linear functional satisfying

$$\ell(x) \leq p(x), \text{ for all } x \in X.$$

Remark 1.9. Note that the substance of the theorem isn't that the extension exists but that an extension can be found that remains *dominated* by p.

Just to find an extension, let $\{e_i\}$ be a Hamel basis for Y and let $\{\varepsilon_j\}$ be vectors in X such that $\{e_i\} \cup \{\varepsilon_j\}$ is a Hamel basis for X. Now define $L : X \to \mathbb{R}$ by

$$L\Bigg(\sum_i \alpha_i e_i + \sum_j \beta_j \varepsilon_j\Bigg) = \sum_i \alpha_i \ell\left(e_i\right) = \ell\Bigg(\sum_i \alpha_i e_i\Bigg).$$

This extends ℓ. If $\{\gamma_j\}$ is any collection of real numbers, then

$$L\Bigg(\sum_i \alpha_i e_i + \sum_j \beta_j \varepsilon_j\Bigg) = \ell\Bigg(\sum_i \alpha_i e_i\Bigg) + \sum_j \beta_j \gamma_j$$

is also an extension of ℓ. Moreover, any extension of ℓ has this form. The difficulty is that we must find one still dominated by p.

Proof. Without loss of generality, we suppose that Y is not all of X.

Step 1. There is some z in X that is not in Y. Denote by $Z = Y \oplus \operatorname{span}\{z\}$. Our aim is to extend ℓ as a linear functional L on Z such that L is dominated by p. Let's see what L must look like, if L exists. Put $\alpha_0 = L\left(z\right)$.

(a) Given $t > 0$ and $y \in Y$, we have $L(y + tz) = t\alpha_0 + \ell(y) \leq p(y + tz)$. Hence

$$\alpha_0 \leq \frac{p(y + tz)}{t} - \frac{\ell(y)}{t} = p\left(\frac{y}{t} + z\right) - \ell\left(\frac{y}{t}\right)$$

for every y in Y. Since $y/t \in Y$, this gives that

$$\alpha_0 \leq p(y + z) - \ell(y) \tag{1.3}$$

for all y in Y. On the other hand, if α_0 satisfies (1.3), then by reversing the preceding argument, it follows that $t\alpha_0 + \ell(y) \leq p(y + tz)$ whenever $t \geq 0$.

(b) Given $t > 0$ and $y' \in Y$, we have $L(y' - tz) = \ell(y') - t\alpha_0 \leq p(y' - tz)$. For the sme reason, this is equivalent to

$$\alpha_0 \geq \ell\left(y'\right) - p\left(y' - z\right) . \tag{1.4}$$

Combining (1.3) and (1.4), we see that that α_0 can be chosen satisfying them both simultaneously. Such an α_0 exists iff for all pairs $y, y' \in Y$,

$$\ell\left(y'\right) - p\left(y' - z\right) \leq p(y + z) - \ell(y) \tag{1.5}$$

Using the linearity of ℓ and subadditivity of p we have

$$\ell\left(y + y'\right) \leq p(y + y') \leq p(y + z) + p\left(y' - z\right) .$$

So pick α_0 satisfying $\sup_{y' \in Y} \ell(y') - p(-z + y') \leq \alpha_0 \leq \inf_{y \in Y} p(z + y) - \ell(y)$, and define

$$L(y + tz) = \ell(y) + t\alpha_0 \,, \text{ for } t \in \mathbb{R} .$$

we get a extension of ℓ on Z dominated by p.

Step 2. Consider all extensions of ℓ to linear spaces Z containing Y and dominated by p. We order these extensions by defining

$$(Z, \ell) \leq \left(Z', \ell'\right)$$

to mean that Z' contains Z, and that ℓ' agrees with ℓ on Z.

Let $\{Z_v, \ell_v\}$ be a totally ordered collection of extensions of ℓ. Then we can define $\widehat{\ell}$ on the union $\widehat{Z} = \cup_v Z_v$ as being ℓ_v on Z_v. Clearly, $\widehat{\ell}$ on $\widehat{Z}$ dominated by p, and $(Z_v, \ell_v) \leq (Z, \widehat{\ell})$ for all v. This shows that every totally ordered collection of extensions of ℓ has an upper bound. So the hypothesis of Zorn's lemma is satisfied, and we conclude that there exists a maximal extension. But according to the foregoing, a maximal extension must be to the whole space X. $\square$

To extend the result to complex vector spaces, we need the following lemma.

Lemma 1.19. *Let X be a vector space over $\mathbb{C}$.*

(a) *If $\ell : X \to \mathbb{C}$ is $\mathbb{C}$-linear, let $\ell_1 = \operatorname{Re} \ell$, then ℓ_1 is a $\mathbb{R}$-linear functional, and we have*

$$\ell(x) = \ell_1(x) - i\ell_1(ix)\,, \text{ for all } x \in X\,.$$

(b) *If $\ell_1 : X \to \mathbb{R}$ is an $\mathbb{R}$-linear functional, then $\ell(x) = \ell_1(x) - i\ell_1(ix)$ is a $\mathbb{C}$-linear functional. Moreover,*

$$\ell = \operatorname{Re} \tilde{\ell}\,.$$

(c) *Let p be a seminorm on X and ℓ and ℓ_1 are as in (a) or (b), then*

$$|\ell(x)| \leq p(x) \text{ for all } x \Leftrightarrow \ell_1(x) \leq p(x) \text{ for all } x\,.$$

Proof. The proofs of (a) and (b) are left as an exercise. To prove (c), suppose $|\ell(x)| \leq p(x)$. Then

$$\ell_1(x) = \operatorname{Re} \tilde{\ell}(x) \leq |\tilde{\ell}(x)| \leq p(x)\,.$$

Now assume that $\ell_1(x) \leq p(x)$. for all x. Let $\ell(x) = e^{i\theta}|\ell(x)|$. Hence

$$|\ell(x)| = \ell\left(e^{-i\theta}x\right) = \operatorname{Re} \ell\left(e^{-i\theta}x\right) = \ell_1\left(e^{-i\theta}x\right) \leq p\left(e^{-i\theta}x\right) = p(x)\,. \quad \square$$

Theorem 1.20 (Hahn-Banach Theorem I)**.** *Let X be a vector space over $\mathbb{F}$, let Y be a subspace, let p be a seminorm, and ℓ is a linear functional on Y. If ℓ is dominated by p,*

$$|\ell(y)| \leq p(y), \text{ for all } y \in Y,$$

then ℓ can be extended to X as a linear functional, still dominated by p:

$$|\ell(x)| \leq p(x), \text{ for all } x \in X.$$

Proof. There are two case for $\mathbb{F}$. When $\mathbb{F}$ is $\mathbb{R}$, then $\ell(y) \leq |\ell(y)| \leq p(y)$ for y in Y. By Theorem 1.18 , ℓ can be extended to X such that $\ell(x) \leq p(x)$ for all $x \in X$. Hence $-\ell(x) = \ell(-x) \leq p(-x) = p(x)$, That is, $|\ell(x)| \leq p(x)$.

When $\mathbb{F} = \mathbb{C}$. Let $\ell_1 = \operatorname{Re} \ell$. By Lemma 1.19, $\ell_1 \leq p$. From the proof above, ℓ_1 can be extended on X as $\mathbb{R}$-linear functional such that $|\ell_1| \leq p$. Let

$$\ell(x) = \ell_1(x) - i\ell_1(ix)$$

for all x in X, then ℓ is a extension. By Lemma 1.19, $|\ell| \leq p$. $\square$

1.5.2 Geometric Hahn-Banach Theorems

In spite (or perhaps because) of its nonconstructive proof, the Hahn-Banach theorem has plenty of very concrete applications. One of the most important is to separation theorems concerning convex sets; these are sometimes called *geometric Hahn-Banach theorems*.

Minkowski Functionals. Suppose X is a vector space over $\mathbb{F}$. Let K be a subset of X.

(a) K is called *convex* if, for any $t \in [0, 1]$, $tK + (1 - t)K \subset K$.

(b) K is called *absorbing at* $x \in K$, if for any $y \in X$, there exists an $\epsilon > 0$, depending on y, such that

$$x + ty \in K \quad \text{for all real } t, |t| < \epsilon.$$

If K contains 0 and absorbs at 0, we say K is *absorbing* for short.

(c) K is called *balanced*, if for any $\lambda \in \mathbb{F}$ and $|\lambda| \le 1$, we have $\lambda K \subset K$.

Example 1.21. Let p be real-valued functional on X.

(a) If p is a positive homogeneous, subadditive functional on X, then $\{x : p(x) < 1\}$ is a absorbing convex subset of X.

(b) If p is a seminorm on X, then $\{x : p(x) < 1\}$ is a balanced absorbing convex subset of X.

The following observation is quite useful. If p is a seminorm, letting $K = \{x : p(x) < 1\}$, there holds

$$p(x) = \inf\left\{t > 0 : \frac{x}{t} \in K\right\} .$$

If we let K be any balanced absorbing convex set, then we can see that the avobe p satisfies that $p(x) < \infty$ for all x. Hence, we introduce the definition:

Definition 1.8. Let K be a balanced absorbing convex set, the *Minkowski functional of* K is defined by

$$p_K(x) = \inf\left\{t > 0 : \frac{x}{t} \in K\right\} , \text{ for } x \in X . \tag{1.6}$$

Lemma 1.21. *Let K be a balanced absorbing convex in X, then p_K is a seminorm.*

Proof. We only show that for all x, y in X,

$$p_K(x + y) \le p_K(x) + p_K(y) .$$

For any $a, b > 0$, such that $\frac{x}{a}, \frac{y}{b} \in K$, note that

$$\frac{x + y}{a + b} = \frac{a}{a + b}\,\frac{x}{a} + \frac{b}{a + b}\,\frac{y}{b} \in K .$$

So $p_K(x + y) \le a + b$. Letting $a \downarrow p_K(x)$ and $b \downarrow p_K(y)$, we get the desired result. $\square$

Lemma 1.22. *For any balanced absorbing convex set K,*

(a) $\{x : p_K(x) < 1\} \subset K \subset \{x : p_K(x) \le 1\}$.

(b) $p_K(x) < 1$ *if and only if K is absorbing at x.*

Proof. We only show (b). When K is absorbing at x, clearly $p_K(x) < 1$. Suppose now that $p_K(x) < 1$. Given $y \in X$, note that

$$p_K(x + ty) \le p_K(x) + |t| p_K(y) \,.$$

So there exists some $\epsilon > 0$, depending on y, so that when $|t| < \epsilon$,

$$p_K(x + ty) \le p_K(x) + |t| p_K(y) \le p_K(x) + \epsilon\, p_K(y) < 1 \,.$$

By (a), $x + ty \in K$. ☐

Note that in (1.6), if we let K be any absorbing convex set, give up the condition that K is balanced, $p_K(x) < \infty$ for all $x \in X$ still holds. Thuswe extends the Minkovski functional to all the absorbing convex set: Let K be absorbing and convex. The *Minkowski functional of K* is defined by

$$p_K(x) = \inf\left\{t > 0 : \frac{x}{t} \in K\right\} \,.$$

As the preceding theorems, we have

Lemma 1.23. *K is an absorbing convex set, p_K is the Minkowski functional of K. Then*

(a) p_K *is a positive homogeneous, subadditive functional on X.*

(b) $\{x : p_K(x) < 1\} \subset K \subset \{x : p_K(x) \le 1\}$.

(c) $p_K(x) < 1$ *if and only if K is absorbing at x.*

Hyperplane and Linear Functionals. We turn now to the notion of a hyperplane. Let X be a vector space over $\mathbb{F}$, a subspace M is called a *hyperplane* in X if it has codimension 1, in other words,

$$\dim(X/M) = 1\,.$$

An *affine hyperplane* in X is a hyperplane shifted from the origin by a vector, i.e., $x_0 + M$ is the affine hyperplane for some $x_0 \in X$.

It's easy to find that is a closed connection between hyperplanes and linear functionals.

- Let ℓ be a *nonzero* linear functional, then $\ker\ell$ is a hyperplane. In fact, there is an isomorhism between $X/\ker\ell$ and $\mathbb{F}$ naturally induced by ℓ.
- Let M be a hyperplane, and $Q : X \to X/M$ be the quotinent map and let $T : X/M \to \mathbb{F}$ be an isomorphism. Then $\ell \coloneqq T \circ Q$ is a linear functional with kernel $\ker\ell = M$.

This is summarized as follows :

Proposition 1.24. *let X be a vector space, $M \subset X$, then M is a hyperplane if and only if there is a nonzero linear functional ℓ such that $M = \ker\ell$. Consequently, M is an affine hyperplane if and only if there is non-zero linear functional ℓ and some scalar c such that $M = \{\ell = c\}$.*

Hyperplane Seperation Theorem. There is a great advantage inherent in a *geometric discussion* of *real* vector space, X. Since, if ℓ is a nonzero linear functional, then any affine hyperplane $\{\ell = c\}$ "disconnects" the space: all points of X belong to one, and only one, of the following three sets:

$$\{x : \ell(x) < c\}, \quad \{x : \ell(x) = c\}, \quad \{x : \ell(x) > c\}\,.$$

The sets where $\{\ell < c\}$, or$\{\ell > c\}$ are called *open halfspaces.* The sets where $\{\ell \geq c\}$, or $\{\ell \leq c\}$ are called *closed halfspaces.*

However, When X is a complex vector space, and ℓ is nonzero linear functional, then $X \backslash \{l = c\}$ is "connected". However, we can regard any complex vector space as a real vector space, in this case, we say ℓ is "linear" means ℓ is $\mathbb{R}$ -linear, not $\mathbb{C}$-linear. Then the resluts in real vector spaces can be applied in vector spaces over both $\mathbb{R}$ and $\mathbb{C}$ after the slight modification.

We say two subsets A and B of *real* vectors space X are said to be *strictly separated* if they are contained in disjoint open half-spaces; they are *separated* if they are contained in two closed half-spaces whose intersection is a affine hyperplane.

Theorem 1.25 (Hyperplane Separation Theorem)**.** *Let X be a real vector space. Let $K \subset X$ be convex and absorbing at each of it's points. Then, any point $y \notin K$ can be separated from K by a hyperplane $\{x : \ell(x) = c\}$. In other words, there is a linear functional ℓ, depending on y, such that*

$$\ell(x) < c = \ell(y) \quad \textit{for all } x \textit{ in } K\,.$$

Proof. Without loss of generality, we assume that $0 \in K$. So K is absorbing. Denote by p_K the Minkovski functional of K. We define ℓ on $\operatorname{span}\{y\}$ by

$$\ell(ay) = a\,, \text{ for all } a \in \mathbb{R}\,.$$

We claim that for all such ay,

$$\ell(ay) \leq p_K(ay)\,.$$

This is obvious for $a \leq 0$, for then $\ell(ay) = a \leq 0$ while $p_K \geq 0$. If $a > 0$, since $y \notin K$, $p_K(y) \geq 1$. So, $p_K(ay) \geq a = \ell(ay)$ for $a > 0$.

Having shown that ℓ, as defined on the above one-dimensional subspace, is dominated by p_K, we conclude from the Hahn-Banach theorem that ℓ can be so extended to all of X. We deduce from this and Lemma 1.23, since K is absorbing at each of it's points, it follows from that $p_K(x) < 1$ for every x in K, thus

$$\ell(x) \leq p_K(x) < 1 = l(y)\,. \qquad \square$$

Corollary 1.26. *Let K denote a absorbing convex set. For any y not in K there is a nonzero linear functional ℓ that satisfies*

$$\ell(x) \leq \ell(y) \;\; \textit{for all } \; x \; \textit{in } K.$$

Theorem 1.27 (Extended Hyperplane Separation)**.** *X is real vector space. A and B disjoint convex subsets of X. A is absorbing at some $a_0 \in A$. Then A and B can be separated by a hyperplane $\{x : \ell(x) = c\}$. That is, there is a nonzero linear functional ℓ, and a number c, such that*

$$\ell(a) \leq c \leq \ell(b) \quad \textit{for all}\, a \in A,\, b \in B\,.$$

Proof. Let $G = A - B = \{a - b : a \in A, b \in B\}$; it is easy to verify that G is convex and absorbsing at each of it's points (do it!). Moreover, $0 \notin G$, because $A \cap B = \emptyset$. By hyperplane separation theorem, there is a linear functional ℓ on X such that

$$\ell(a - b) < 0 = \ell(0)\,, \text{for any}\, a \in A, b \in B$$

Thus

$$\sup\{\ell(a) : a \in A\} \leq c \leq \inf\{\ell(b) : b \in B\}\,. \qquad \square$$

1.6 Continuous Linear Functionals

We have pointed that, a subspace in X is a hyperplane if and only if it is the kernel of a non-zero linear functional. Two linear functionals have the same kernel if and only if one is a non-zero multiple of the other.

Hyperplanes in a normed space fall into one of two categories.

Proposition 1.28. *Let X be a normed linear space. Let M be a hyperplane in X. Then either M is closed or M is dense.*

Proof. Note that $\dim(X/\overline{M}) \leq \dim(X/M)$, then the desired result follows. $\square$

Example 1.22. To get an example of a dense hyperplane, consider normed linear space c_0. Denote e_n as the element of c_0 such that $e_n(k) = 0$ if $k \neq n$ and $e_n(n) = 1$. Let $x_0(n) = 1/n$ for all n, so $x_0 \in c_0$ and $\{x_0, e_1, e_2, \ldots\}$ is a linearly independent set in c_0. Let $\mathscr{B} =$ a Hamel basis in c_0 which contains $\{x_0, e_1, e_2, \ldots\}$. Put $\mathscr{B} = \{x_0, e_1, e_2, \ldots\} \cup \{b_i : i \in I\}$. Define $f : c_0 \to \mathbb{F}$ by

$$f\left(\alpha_0 x_0 + \sum_n \alpha_n e_n + \sum_i \beta_i b_i\right) = \alpha_0.$$

(Remember that in the preceding expression at most a finite number of the α_n and β_i are not zero). Since $e_n \in \ker f$ for all $n \geq 1$, $\ker f$ is dense but clearly $\ker f \neq c_0$.

Theorem 1.29. *Let X be a normed space over $\mathbb{F}$ and $f : X \to \mathbb{F}$ is a linear functional, then f is continuous if and only if $\ker f$ is closed.*

First proof. If f is continuous, then $\ker f$ must be closed. Assume now that $\ker f$ is closed and let $Q : X \to X/\ker f$ be the natural map. Let $T : X/\ker f \to \mathbb{F}$ be an isomorphism. Note that both Q, T are continuous. Thus, if $g = T \circ Q : X \to \mathbb{F}$, then g is continuous and $\ker f = \ker g$. Hence $f = \alpha g$ for some α in $\mathbb{F}$ and so f is continuous. $\square$

Second proof. Suppose $\ker f$ is closed and f is not continuous. Then there exists $\{x_n\}$ and $\epsilon > 0$, so that $\|x_n\| \to 0$ but $|f(x_n)| \geq \epsilon$ for all n. Observe that

$$\frac{x_n}{f(x_n)} - \frac{x_1}{f(x_1)} \in \ker f \quad \text{for each } n \geq 1\,.$$

Since $\ker f$ is closed and $\frac{x_n}{f(x_n)} \to 0$, we deduce that $\frac{x_1}{f(x_1)} \in \ker f$, which is absurd. Since $f(\frac{x_1}{f(x_1)}) = 1$. $\square$

We will denote by X^* all the continuous linear functional on the norm linear space X. In other words, $X^* := \mathcal{B}(X, \mathbb{F})$. Then it follows from Theorem 1.5 that X^* is a Banach space endowed with the operator norm.

Theorem 1.30 (Hahn-Banach Extension Theorem in Normed Space)**.** *Let X be a normed vector space. Let Y is a subspace in X. Let $f : Y \to \mathbb{F}$ is a bounded linear functional, then there is an $\bar{f}$ in X^* such that $\bar{f}|_Y = f$ and $\|\bar{f}\| = \|f\|$.*

Proof. Use Theorem 1.20 with $p(\cdot) = \|f\|\|\cdot\|$. □

Note that Y don't need to be closed. In fact, without using Hahn-Banach theorem we can prove there is a bounded linear functional $\bar{f}$ on $\overline{Y}$ such that $\bar{f}|_Y = f$ and $\|\bar{f}\| = \|f\|$.

Corollary 1.31. *If X is a normed space, $\{x_1, x_2, \ldots, x_d\}$ is a linearly independent subset of X, and $\alpha_1, \alpha_2, \ldots, \alpha_d$ are arbitrary scalars in $\mathbb{F}$, then there is an f in X^* such that $f(x_j) = \alpha_j$ for $1 \le j \le d$.*

Proof. Let $Y =$ the linear span of $x_1, \ldots, x_d$ and define $g : Y \to \mathbb{F}$ by

$$g\left(\sum_j \beta_j x_j\right) = \sum_j \beta_j \alpha_j \,.$$

So g is linear. since Y is finite dimensional, g is continuous. Let f be a continuous extension of g to X. □

Corollary 1.32. *Let X be a normed linear space. For any $x \in X$,*

$$\|x\| = \sup_{x^* \in X^*, \|x^*\| \le 1} |\langle x, x^* \rangle| \,. \tag{1.7}$$

Moreover, this supremum is attained.

Proof. Obviously,

$$\sup_{x^* \in X^*, \|x^*\| \le 1} |\langle x, x^* \rangle| \le \|x\| \,.$$

On the other hand, define $f : \mathrm{span}\{x\} \to \mathbb{F}$ by

$$f(\beta x) = \beta \|x\| \,, \text{ for all } \beta \in \mathbb{F} \,.$$

Then f is bounded and $\|f\| = 1$. By Hahn-Banach extension theorem, there is a x^* in X^* such that $\|x^*\| = 1$ and $x^*(x) = f(x) = \|x\|$. □

Corollary 1.33. *X, Y are normed space. $T \in \mathcal{B}(X, Y)$, then*

$$\|T\| = \sup_{\|x\| \leq 1, \|y^*\| \leq 1} |\langle Tx, y^* \rangle| . \tag{1.8}$$

Corollary 1.34. *Let X be a normed linear space. Let M be a proper subspace of X. For each $x_0 \notin \overline{M}$, there exists f in X^*, satisfying that (i) $\|f\| = 1$; (ii) $f(x_0) = d(x_0, M)$; and (iii) $f(x) = 0$ for all x in M.*

Proof. Let $Y = \operatorname{span}\{x_0\} \oplus M$. Define a functional $f : Y \to \mathbb{F}$ by

$$f(\alpha x_0 + m) = \alpha d(x_0, M) \qquad \text{for } \alpha \in \mathbb{F},\ m \in M .$$

Then it's easy to check that f is a continuous linear function on Y. We show that $\|f\| = 1$. To this end, observe that for $\alpha \in \mathbb{F}$ and $m \in M$,

$$|f(\alpha x_0 + m)| = |\alpha| d(x_0, M) = \inf_{m' \in M} \|\alpha x_0 + m'\| .$$

On the one hand, $|f(\alpha x_0 + m)| \leq \|\alpha x_0 + m\|$ implies $\|f\| \leq 1$. On the other hand, running m though M

$$\inf_{m' \in M} \|\alpha x_0 + m'\| \leq \|f\| \inf_{m \in M} \|\alpha x_0 + m\|$$

which implies $\|f\| \leq 1$. So by Hahn-Banach extension theorem, we can extend f to X as a continuous linear functional with the same norm. □

Chapter 2

Hilbert Space

A Hilbert space is the abstraction of the finite-dimensional Euclidean spaces of geometry. Its properties are very regular and contain few surprises, though the presence of an infinity of dimensions guarantees a certain amount of surprise. Historically, it was the properties of Hilbert spaces that guided mathematicians when they began to generalize.

2.1 Fundamentals

Definition 2.1. Let X be a vector space over $\mathbb{F}$. An *inner product* (or *scalar product*) on X is a scalar-valued function $\langle \cdot\,,\cdot \rangle : X \times X \to \mathbb{F}$ such that for all $x, y, z \in X$ and forall $\alpha, \beta \in \mathbb{F}$, we have

(a) $\langle x, x \rangle \geq 0$, and $\langle x, x \rangle = 0$ iff $x = 0$. (Positive definiteness)

(b) $\langle x, y \rangle = \overline{\langle y, x \rangle}$. (Hermitian property)

(c) $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$. (Linearity)

In other words, inner product is a positive definite sesquilinear form. If we change (a) to (a$'$): $\langle x, x \rangle \geq 0$, $\langle \cdot\,,\cdot \rangle$ is called a *semi-inner production.* X is called *inner production space*, or *semi-inner production space*, respectively.

Theorem 2.1 (Cauchy-Bunyakowsky-Schwarz Inequality)**.** *Let* $(X, \langle \cdot, \cdot \rangle)$ *be an semi-inner product space over* $\mathbb{F}$. *Then, for all* $x, y \in X$.

$$|\langle x, y\rangle|^2 \leq \langle x, x\rangle \langle y, y\rangle \,,$$

Moreover, equality occurs if and only if there are scalars α, β, *both not* 0, *such that* $\langle \alpha x + \beta y, \alpha x + \beta y\rangle = 0$.

Proof. We only show this inequality in the case of $\mathbb{F} = \mathbb{C}$. Note that, for any $\lambda \in \mathbb{C}$ and $x, y \neq 0$, we have

$$\langle x + \lambda y, x + \lambda y\rangle = \langle y, y\rangle |\lambda|^2 + 2Re\{\langle x, y\rangle \overline{\lambda}\} + \langle x, x\rangle \geq 0\,.$$

One can show that $f(z) = |\alpha z|^2 - 2Re\{\beta z\}$, where $\alpha, \beta \in \mathbb{C}, \alpha \neq 0$, achieve its minimum $-\frac{|\beta|^2}{|\alpha|^2}$ if and only if $z = \frac{\overline{\beta}}{|\alpha|^2}$. Then let $\lambda = -\frac{\langle x,y\rangle}{\langle y,y\rangle}$, we get

$$-\frac{|\langle x, y\rangle|^2}{\langle y, y\rangle} + \langle x, x\rangle \geq 0\,.$$

Moreover, the equality orrurs if and only if α, β, both not 0, such that $\langle \alpha x + \beta y, \alpha x + \beta y\rangle = 0$. □

Let $(X, \langle \cdot\,, \cdot\rangle)$ be an inner product (semi-inner product) space over $\mathbb{F}$. For each $x \in X$, define

$$\|x\| \coloneqq \langle x, x\rangle^{\frac{1}{2}} \tag{2.1}$$

Then $\| \cdot \|$ is a norm, called the norm induced by the inner product. Using this notation, the CBS Inequality now becomes

$$|\langle x, y\rangle| \leq \|x\|\|y\|\,. \tag{2.2}$$

From this, we can see that the inner product is continuous: If $x_n \to x$, $y_n \to y$ with respect to the norm, then $\langle x_n, y_n\rangle$ tends to $\langle x, y\rangle$.

A natural question arises: Is every normed linear space an inner product space? If the answer is NO, how then does one recognise among all normed linear spaces those that are inner product spaces in disguise, i.e., those whose norms are induced by an inner product?

Proposition 2.2 (Polarization Identity). *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space over $\mathbb{F}$. Given any $x, y \in X$,*

- *if* $\mathbb{F} = \mathbb{R}$, $\langle x, y\rangle = \dfrac{\|x+y\|^2 - \|x-y\|^2}{4}$.
- *if* $\mathbb{F} = \mathbb{C}$, $\langle x, y\rangle = \dfrac{\|x+y\|^2 - \|x-y\|^2}{4} + i\left(\dfrac{\|x+iy\|^2 - \|x-iy\|^2}{4}\right)$.

Proof. We only show the polarization identity in the case of $\mathbb{F} = \mathbb{C}$. Since

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

We get

$$\mathrm{Re}\langle x, y\rangle = \frac{\|x+y\|^2 - \|x-y\|^2}{4} .$$

Note that

$$\mathrm{Im}\langle x, y\rangle = \mathrm{Re}\langle x, iy\rangle ,$$

the desired result follows. □

Proposition 2.3 (Parallelogram Identity). *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space over a field $\mathbb{F}$. then for all $x, y \in X$,*

$$\|x-y\|^2 + \|x+y\|^2 = 2\left(\|x\|^2 + \|y\|^2\right) . \tag{2.3}$$

Moreover, a normed linear space X over $\mathbb{F}$ is an inner product space if and only if the parallelogram identity holds for all $x, y \in X$.

Proof. If $(X, \langle\cdot,\cdot\rangle)$ is an inner product, it's trivial that the parallelogram identity holds.

Step 1. If X is normed space over $\mathbb{R}$ and the parallelogram identity holds, define

$$\langle x, y\rangle = \frac{\|x+y\|^2 - \|x-y\|^2}{4} .$$

Then we can check $\langle\cdot,\cdot\rangle$ is an inner product. Clearly $\langle\cdot,\cdot\rangle$ is positive definite and symmetric. We firstly show that

$$\langle x+z, y\rangle = \langle x, y\rangle + \langle z, y\rangle .$$

This follows from parallelogram identity. By induction, for all $n \in \mathbb{N}$, we have $\langle nx, y\rangle = n\langle x, y\rangle$. Then it follows that for $r \in \mathbb{Q}$, $\langle rx, y\rangle = r\langle x, y\rangle$. Since $\langle\cdot,\cdot\rangle$ is continuous, for any $\lambda \in \mathbb{R}$, we have

$$\langle \lambda x, y\rangle = \lambda\langle x, y\rangle\,.$$

Step 2. If X is normed space over $\mathbb{C}$ and the parallelogram identity holds, define

$$\langle x, y\rangle = \frac{\|x+y\|^2 - \|x-y\|^2}{4} + i\,\frac{\|x+iy\|^2 - \|x-iy\|^2}{4}\,.$$

Clearly $\langle\cdot,\cdot\rangle$ is positive definite and Hermitian. We have showed that for given y, $\langle\cdot, y\rangle$ is $\mathbb{R}$-linear in step 1. Observe that

$$\langle ix, y\rangle = i\langle x, y\rangle\,,$$

we have that $\mathrm{Re}\langle\cdot, y\rangle$ is $\mathbb{C}$-linear for given y. So the desired result follows. □

Remark 2.1. From Proposition 2.3 , we know if every two-dimensional linear subspace of normed linear space X is an inner product space, then X is an inner product space.

The mathematical concept of a Hilbert space, named after David Hilbert, generalizes the notion of Euclidean space. Hilbert spaces, as the following definition states, are inner product spaces which in addition are required to be complete,a property that stipulates the existence of enough limits in the space to allow the techniques of calculus to be used.

Definition 2.2. $(H, \langle\cdot,\cdot\rangle)$ is inner product space. If H is complete, with respect to the norm induced by $\langle\cdot,\cdot\rangle$, then we say that H is a *Hilbert space.*

Remark 2.2. Given a linear space with a inner product, it can be completed with respect to the norm derived from the inner product. It follows from the C-B-S inequality that the inner product is a continuous function of its factors; therefore it can be extended to the completed space. Thus the completion is a Hilbert space.

Example 2.1. Fix a positive integer n. Let $X = \mathbb{F}^n$. For $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ in X, define

$$\langle x, y\rangle = \sum_{\alpha=1}^{n} x_i\overline{y_i}\,.$$

The space $\mathbb{R}^n$ (resp. $\mathbb{C}^n$) with this inner product is called the Euclidean space (resp. unitary space) and is a (trivial) Hilbert space.

Example 2.2. Let $M_n(\mathbb{C})$, the linear space of all $n \times n$ complex matrices. For any $A \in M_n(\mathbb{C})$ let $\mathrm{tr}(A) = \sum_{\alpha=1}^{n} (A)_{ii}$ be the trace of A. For $A, B \in M_n(\mathbb{C})$, define

$$\langle A, B\rangle = \mathrm{tr}\,(B^* A)$$

where B^* denotes conjugate transpose of matrix B. Then $(M_n(\mathbb{C}), \langle\cdot,\cdot\rangle)$ is a Hilbert space.

Example 2.3. Let $X = \ell_0$, the linear space of finitely non-zero sequences of real or complex numbers. For $x = (x_1, x_2, \ldots)$ and $y = (y_1, y_2, \ldots)$ in X, define

$$\langle x, y\rangle = \sum_{\alpha=1}^{\infty} x_i\overline{y_i}\,.$$

since this is essentially a finite sum, $\langle\cdot\,,\cdot\rangle$ is well-defined. ℓ^0 is an incomplete inner product space.

Example 2.4. Let $X = \ell^2$, the space of all sequences $x = (x_1, x_2, \ldots)$ of real or complex numbers with $\sum_{\alpha=1}^{\infty} |x_i|^2 < \infty$. For $x = (x_1, x_2, \ldots)$ and $y = (y_1, y_2, \ldots)$ in X, define

$$\langle x, y\rangle = \sum_{\alpha=1}^{\infty} x_i\overline{y_i}\,.$$

ℓ^2 is a Hilbert space. Moreover, ℓ^2 is the completion of ℓ_0 in previous example.

Example 2.5. Let I be any noempty set and let $\ell^2(I)$ denote the set of all functions $x : I \to \mathbb{F}$ such that $\sum_{\alpha \in I} |x(i)|^2 < \infty$. For x and y in $\ell^2(I)$ define

$$\langle x, y \rangle = \sum_{\alpha \in I} x(i)\overline{y(i)}.$$

For any noempty set I, $\ell^2(I)$ is a Hilbert space. One can find that $\ell^2(\mathbb{N})$ is exactly ℓ^2.

Example 2.6. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Denote by $L^2(\Omega, \mathcal{F}, \mu)$ all $\mathbb{F}$-valued square integrable functions If f and $g \in L^2$, then Hölder's inequality implies $f\overline{g} \in L^1$. Definite

$$\langle f, g \rangle = \int f\overline{g}\, \mathrm{d}\mu\,,$$

then this defines an inner product on L^2. Then L^2 becomes a Hilbert space.

Example 2.7. Let $X = C[a, b]$, the space of all continuous $\mathbb{F}$-valued functions on $[a, b]$. For $x, y \in X$, define

$$\langle x, y \rangle = \int_a^b x(t)\overline{y(t)}dt$$

$C[a, b]$ is an incomplete inner product space. Evidently, $L^2[a, b]$ is the Completion of the $C[a, b]$.

The (External) Direct Sum of Hilbert Spaces Suppose H and K are Hilbert spaces. We want to define $H \oplus K$ so that it becomes a Hilbert space. This is not a difficult assignment.

Proposition 2.4. *If H and K are Hilbert spaces, the external direct sum $H \oplus K$ endowed with the inner product*

$$\langle (h_1, k_1), (h_2, k_2) \rangle := \langle h_1, k_1 \rangle + \langle h_2, k_2 \rangle$$

for every $(h_1, k_1), (h_2, k_2) \in H \oplus K$ is a Hilbert space.

Now what happens if we want to define $H_1 \oplus H_2 \oplus \cdots$ for a sequence of Hilbert spaces $H_1, H_2, \ldots$? There is a problem about the completeness of this infinite (external) direct sum, but this can be overcome as follows.

Proposition 2.5. *Let $\{H_\alpha\}_{\alpha\in\Lambda}$ be a collection of Hilbert spaces. Let*

$$H = \left\{ (h_\alpha) : h_\alpha \in H_\alpha \ \textit{and} \ \sum_\alpha \|h_\alpha\|^2 < \infty \right\} .$$

For $h = (h_\alpha)$ and $g = (g_\alpha)$ in H, define

$$\langle h, g\rangle = \sum_\alpha \langle h_\alpha, g_\alpha\rangle \tag{2.4}$$

Then $\langle , \rangle$ is an inner product on H and the norm relative to this inner product is $\|h\| = \left[\sum \|h_\alpha\|^2\right]^{1/2}$. With this inner product, H is a Hilbert space.

Proof. If $h = (h_\alpha)$ and $g = (g_\alpha) \in H$, then the CBS inequality implies

$$\sum_\alpha |\langle h_\alpha, g_\alpha\rangle| \le \sum_\alpha \|h_\alpha\| \, \|g_\alpha\| \le \left(\sum_\alpha \|h_\alpha\|^2\right)^{1/2} \left(\sum_\alpha \|g_\alpha\|^2\right)^{1/2} < \infty.$$

Hence the series in (2.4) converges absolutely. Trivially, we check that $\langle , \rangle$ is an inner product. H is a Hilbert space follows form Proposition 1.3. □

The space H is called the (external) direc sum of $\{H_\alpha\}_{\alpha\in\Lambda}$ and is denotd by $\oplus_\alpha H_\alpha$.

2.2 Orthogonality

Definition 2.3. Two elements x and y in an inner product space $(X, \langle\cdot,\cdot\rangle)$ are said to be *orthogonal*, denoted by $x \perp y$, if

$$\langle x, y\rangle = 0 .$$

A subset $\Omega \subset X$ is called *orthogonal* if it consists of non-zero pairwise orthogonal elements.

Pythagorean theorem still holds in this case. In other words, if

$$\{x_1, x_2, \ldots, x_n\}$$

is an orthogonal set , then

$$\left\| \sum_{i=1}^{n} x_i \right\|^2 = \sum_{i=1}^{n} \|x_i\|^2 .$$

Definition 2.4. Let S be a subset of X. If $\langle x, s\rangle = 0$ for all $s \in S$, then we say x is *orthogonal to* S and write $x \perp S$. We shall denote by

$$S^\perp = \{x \in X : \langle x, s\rangle = 0\,, \forall\, s \in S\}$$

the set of all elements orthogonal to S. The set $S^\perp$ is called the *orthogonal complement* of S.

It's easy to see that for any $S \subset X$, $S^\perp$ is a closed linear subspace of X with $S \subset S^{\perp\perp} := \left(S^\perp\right)^\perp$, and

$$S^\perp = (\mathrm{span} S)^\perp = (\overline{\mathrm{span}} S)^\perp \tag{2.5}$$

2.2.1 Best Approximation

Theorem 2.6 (Existence of the Unique Best Approximation)**.** *Let* $(X, \langle \cdot\,, \cdot\rangle)$ *be a inner product space. Let* K *be a nonempty complete convex subset of* X*. Then for each* $x \in X$ *has a unique best approximation in* K*, i.e. there is a unique point* $k_0 \in K$ *satisfying*

$$\|x - k_0\| = d(x, K) := \inf_{k \in K} \|x - k\| .$$

Proof. Without loss of generality, assume $x = 0 \notin K$. Then there exists a sequence $\{k_n\}_{n \geq 1}$ in K such that

$$\|k_n\| \downarrow d(0, K) .$$

By parallelogram identity

$$\left\|\frac{k_n - k_m}{2}\right\|^2 + \left\|\frac{k_n + k_m}{2}\right\|^2 = \frac{\|k_n\|^2 + \|k_m\|^2}{2}.$$

Thus $\{k_n\}$ is a Cauchy sequence. Since K is complete, there exists $k_0 \in K$ such taht $k_n \to k_0$, then $\|k_0\| = d(0, K)$.

Using parallelogram identity again, we will get the uniqueness of k_0. $\square$

Remark 2.3. The proof above is using the *uniformly convexness* of the norm induced by the inner product. In fact, in any uniformly convex normed linear space, Theorem 2.6 holds.

Theorem 2.7 (Characterization of the Unique Best Approximation)**.** *Let K be a nonempty complete convex subset of a inner product space $(X, \langle \cdot\,, \cdot \rangle)$. Assume $x \in X \backslash K$ and $k_0 \in K$, then k_0 is the best approximation to x from K if and only if*

$$\mathrm{Re}\,\langle x - k_0, k - k_0 \rangle \leq 0, \textit{ for all } k \in K. \tag{2.6}$$

Proof. Take any $k \in K$ and fix it.

$$\begin{aligned} \|x - k\|^2 &= \|x - k_0 - (k - k_0)\|^2 \\ &= \|x - k_0\|^2 + \|k - k_0\|^2 - 2\mathrm{Re}\,\langle x - k_0, k - k_0 \rangle . \end{aligned}$$

So if (2.6) holds , we get $\|x-k\|^2 \geq \|x-k_0\|^2$. so k_0 is a best approximation. The uniqueness is guaranteed by Theorem 2.6.

If $\|x - k\|^2 \geq \|x - k_0\|^2$ for all $k \in K$, we have

$$2\mathrm{Re}\,\langle x - k_0, k - k_0 \rangle \leq \|k - k_0\|^2 .$$

For any $\lambda \in (0, 1)$ and $k' \in K$, let $k = \lambda k' + (1 - \lambda)k_0$, then we have

$$2\mathrm{Re}\,\langle x - k_0, k' - k_0 \rangle \leq \lambda \|k' - k_0\|^2 .$$

So we let λ tends to zero , we get (2.6). $\square$

2.2.2 Orthogonal Decomposition

Theorem 2.8. *H is Hilbert space, M is a closed subspace. Given $x \in H$, then $m \in M$ is the unique best approximation to x from M if and only if*

$$x - m \perp M.$$

Proof. By Theorem 2.7, $m \in M$ is the unique best approximation to x from M if and only if

$$\operatorname{Re}\left\langle x - m, m' \right\rangle \leq 0, \text{ for all } m' \in M\,.$$

Let $m'' = -m'$, we can see that

$$\operatorname{Re}\left\langle x - m, m' \right\rangle = 0, \text{ for all } m' \in M\,.$$

If H is a real Hilbert space, then we get $x - m \perp M$ directly. If If H is a complex Hilbert space, note that

$$\operatorname{Im}\left\langle x - m, m' \right\rangle = \operatorname{Re}\left\langle x - m, im' \right\rangle = 0, \text{ for all } m' \in M\,,$$

the desired result follows. □

Theorem 2.8 says that if M is a closed linear subspace of a Hilbert space H, then $P_M(x)$ is the best approximation to x from M if and only if $x - P_M(x) \perp M$. That is, the unique best approximation is obtained by "dropping the perpendicular from x onto M". Therefore, the map

$$P_M : H \to M; x \mapsto P_M(x) \tag{2.7}$$

is also called the (*orthogonal*) *projection* of H onto M. And we get the following important theorem :

Theorem 2.9 (Orthogonal Decomposition)**.** *H is Hilbert space, M is a closed subspace. Then*

$$H = M \oplus M^{\perp}\,. \tag{2.8}$$

That is, each $x \in H$ can be uniquely decomposed in the form $x = y + z$ with $y \in M$ and $z \in M^{\perp}$.

Corollary 2.10. *Let S be a nonempty subset of a Hilbert space H. Then*

(a) $(S^{\perp})^{\perp} = \overline{\operatorname{span}}(S)$.

(b) $S^{\perp} = \{0\}$ *if and only if* $\overline{\operatorname{span}}(S) = H$.

2.3 Orthonormal Basis

In this section, we always assume that $(X, \langle \cdot, \cdot \rangle)$ is an inner product space over the field $\mathbb{F}$.

Definition 2.5. Let $\mathcal{E}$ be a subset of X. We say that $\mathcal{E} = \{e_\alpha\}_{\alpha \in \Lambda}$ is *orthonormal* if $\mathcal{E}$ is orthogonal and $\|e_\alpha\| = 1$ for all $\alpha \in \Lambda$.

For any $x \in X$, the numbers $\langle x, e_\alpha \rangle$ are called the α^{th} *Fourier coefficients* of x with respect to $\mathcal{E}$; and the formal series $\sum_{\alpha \in \Lambda} \langle x, e_\alpha \rangle e_\alpha$ are called the *Fourier series* of x with respect to $\mathcal{E}$.

Lemma 2.11 (Gram-Schmidt Orthonormalisation Procedure)**.** *Let $\{x_k\}_{k \geq 1}$ be a linearly independent set in X. There exists an orthonormal set $\{e_k\}_{k \geq 1}$ in X such that for all $n \in \mathbb{N}$,*

$$\operatorname{span}\{x_1, x_2, \dots, x_n\} = \operatorname{span}\{e_1, e_2, \dots, e_n\} .$$

Proof. Set $e_1 = \frac{x_1}{\|x_1\|}$. Then $\operatorname{span}\{x_1\} = \operatorname{span}\{e_1\}$. Then we define e_k by induction. Now assume $\{e_1, \cdots, e_k\}$ is orthonormal, and

$$\operatorname{span}\{e_1, e_2, \dots, e_k\} = \operatorname{span}\{x_1, x_2, \dots, x_k\} .$$

Let $\widehat{x}_{k+1}$ be the projection of x_{k+1} on $\operatorname{span}\{e_1, e_2, \dots, e_k\}$, and

$$y_{k+1} = x_{k+1} - \widehat{x}_{k+1} \ , \ e_{k+1} = \frac{y_{k+1}}{\|y_{k+1}\|} .$$

Obviously, $\{e_1, \cdots, e_{k+1}\}$ is orthonormal, and

$$\operatorname{span}\{e_1, e_2, \dots, e_{k+1}\} = \operatorname{span}\{x_1, x_2, \dots, x_{k+1}\} . \qquad \square$$

Theorem 2.12 (Riesz-Fischer Theorem for Fourier Series). *Let $\mathcal{E} = \{e_\alpha\}_{\alpha\in\Lambda}$ be an orthonormal family in X. Let $\{\lambda_\alpha\}_{\alpha\in\Lambda}$ be a family of scalars.*

(a) If $\sum_{\alpha\in\Lambda} \lambda_\alpha e_\alpha$ converges, then $\sum_{\alpha\in\Lambda} |\lambda_\alpha|^2$ converges, and

$$\Big\| \sum_{\alpha\in\Lambda} \lambda_\alpha e_\alpha \Big\|^2 = \sum_{\alpha\in\Lambda} |\lambda_\alpha|^2 .$$

(b) If X is a Hilbert space, the convergence of the two series $\sum_{\alpha\in\Lambda} |\lambda_\alpha|^2$ and $\sum_{\alpha\in\Lambda} \lambda_\alpha e_\alpha$ are equivalent.

Proof. (a). Suppose $\sum_{\alpha\in\Lambda} \lambda_\alpha e_\alpha$ converges to x. By the continuity of norm, for any $\epsilon > 0$, there exists a finite subset S of Λ, depending on ϵ, so that for any finite subset T containing S, we have

$$\Big\| \sum_{i\in T} \lambda_\alpha e_\alpha \Big\|^2 = \sum_{i\in T} |\lambda_\alpha|^2 \le \|x\|^2 + \epsilon .$$

Therefore, $\sum_{\alpha\in\Lambda} |\lambda_\alpha|^2$ converges, and clearly

$$\sum_{\alpha\in\Lambda} |\lambda_\alpha|^2 = \|x\|^2 .$$

(b). First of all, suppose that Λ is countable. Without loss of generality, let $\Lambda = \mathbb{N}$. Then if $\sum_{n=1}^\infty |\lambda_n|^2$ converges, since X is complete, $\sum_{n=1}^\infty \lambda_n e_n$ absolutely converges. Thus $\sum_{n\in\mathbb{N}} \lambda_n e_n$ converges (as a net). If Λ is uncountable, since

$$\sum_{\alpha\in\Lambda} |\lambda_\alpha|^2 < \infty ,$$

there exists a countable subset Λ_0 of Λ so that $\alpha \in \Lambda_0$ if and only if $\lambda_\alpha \neq 0$. So $\sum_{\alpha\in\Lambda} \lambda_\alpha e_\alpha$ converges, by the first step. $\square$

The following lemma is easy to prove so the proof is omitted. Nonetheless, it is very important.

Lemma 2.13. *Let $\{e_1, \cdots, e_n\}$ be orthonormal. Let $M = \operatorname{span}\{e_1, \cdots, e_n\}$. (Recall that M is complete because it is finite dimensional.) Let $x \in H$.*

(a) $\hat{x} = \sum_{k=1}^{n} \langle x, e_k \rangle e_k$ *is the projection of* x *on* M*, i.e.,* $\hat{x} = P_M x$*; and*

$$\|\hat{x}\|^2 = \sum_{k=1}^{n} |\langle x, e_k \rangle|^2 = \|x\|^2 - \|x - \hat{x}\|^2 \leq \|x\|^2.$$

(b) *For an* $\lambda_k \in \mathbb{F}$*,* $k = 1, 2, \cdots, n$*, we have*

$$\|x - \sum_{k=1}^{n} \langle x, e_k \rangle e_k\| \leq \|x - \sum_{k=1}^{n} \lambda_k e_k\| .$$

Theorem 2.14 (Bessel Inequality)**.** *Let* $\mathcal{E} = \{e_\alpha\}_{\alpha \in \Lambda}$ *be an orthonormal family in* X*. Then, for every* $x \in X$*, the series* $\sum_{\alpha \in \Lambda} |\langle x, e_\alpha \rangle|^2$ *converges, and*

$$\sum_{\alpha \in \Lambda} |\langle x, e_\alpha \rangle|^2 \leq \|x\|^2 .$$

Proof. The theorem follows directly from Lemma 2.13 and Proposition A.11. □

Theorem 2.15 (Fourier Series)**.** *Suppose* $\mathcal{E} = \{e_\alpha\}_{\alpha \in \Lambda}$ *is an orthonormal family in* X*. Take* $x \in H$*. Then the following statements are equivalent.*

(a) *The Fourier series of* x *with respect to* $\mathcal{E}$ *converges to* x*, in other words,* $x = \sum_{\alpha \in \Lambda} \langle x, e_\alpha \rangle e_\alpha$*.*

(b) *There holds the Parseval equality:* $\|x\|^2 = \sum_{\alpha \in \Lambda} |\langle x, e_\alpha \rangle|^2$*.*

(c) *The closed linear subspace generated by* $\mathcal{E}$ *contains* x*, i.e.,* $x \in \overline{\mathrm{span}}(\mathcal{E})$*.*

Proof. Step 1. We shall show that (a) and (b) are equivalent.

(a) implies (b) follows directly from the continuity of the norm. To prove the converse, note that by Lemma 2.13, for any finite subset S of Λ,

$$\|x - \sum_{\alpha \in S} \langle x, e_\alpha \rangle e_\alpha\|^2 = \|x\|^2 - \sum_{\alpha \in S} |\langle x, e_\alpha \rangle|^2 .$$

Since $\sum_{\alpha\in\Lambda} |\langle x, e_\alpha\rangle|^2$ converges to $\|x\|^2$, So $\sum_{\alpha\in\Lambda}\langle x, e_\alpha\rangle e_\alpha$ converges and the limit is x, as desired.

Step 2. We shall show that (a) and (c) are equivalent.

Obviously, (a) implies (c). So we have only to prove (c) implies (a) or (b). To this end, take any $x \in \overline{\mathrm{span}}(\mathcal{E})$. Then for each $\epsilon > 0$, there exists $e_{\alpha_1}, \cdots, e_{\alpha_n} \in \mathcal{E}$ and $\lambda_1, \cdots, \lambda_n \in \mathbb{F}$ such that

$$\|x - \sum_{k=1}^{n} \lambda_k e_{\alpha_k}\| \le \epsilon.$$

By Lemma 2.13, we have

$$\|x - \sum_{k=1}^{n} \langle x, e_{\alpha_k}\rangle e_{\alpha_k}\| \le \|x - \sum_{k=1}^{n} \lambda_k e_{\alpha_k}\| \le \epsilon\,.$$

So

$$\|x - \sum_{k=1}^{n} \langle x, e_{\alpha_k}\rangle e_{\alpha_k}\| = \|x\|^2 - \sum_{k=1}^{n} |\langle x, e_{\alpha_k}\rangle|^2 \le \epsilon\,,$$

and hence

$$\|x\|^2 \le \sum_{k=1}^{n} |\langle x, e_{\alpha_k}\rangle|^2 + \epsilon \le \sum_{\alpha\in\Lambda} |\langle x, e_\alpha\rangle|^2 + \epsilon\,.$$

Let $\epsilon \downarrow 0$, then combining the Besell inequality we get $\|x\|^2 = \sum_{\alpha\in\Lambda} |\langle x, e_\alpha\rangle|^2$.

□

Remark 2.4. There exists the projection of x on $\overline{\mathrm{span}}(\mathcal{E})$ if and only if the series $\sum_{\alpha\in\Lambda}\langle x, e_\alpha\rangle e_\alpha$ converges, when the projection $\hat{x}$ exists, we have

$$\hat{x} = \sum_{\alpha\in\Lambda} \langle x, e_\alpha\rangle e_\alpha\,. \tag{2.9}$$

In fact, if the series converges, then

$$\langle x - \sum_{\alpha\in\Lambda} \langle x, e_\alpha\rangle e_\alpha, e_\beta\rangle = 0 \text{ for all } \beta \in \Lambda\,,$$

and hence $(x - \sum_{\alpha\in\Lambda}\langle x, e_\alpha\rangle e_\alpha) \perp \overline{\text{span}}(\mathcal{E})$. By Theorem 2.8 the projection $\hat{x}$ exists and (2.9) holds. If the projection $\hat{x}$ exists, then by Theorem 2.15

$$\sum_{\alpha\in\Lambda}\langle \hat{x}, e_\alpha\rangle e_\alpha$$

converges to $\hat{x}$. Observe that $\langle \hat{x}, e_\alpha\rangle = \langle x, e_\alpha\rangle$ for all α, so (2.9) holds.

An orthonormal family $\mathcal{E} = \{e_\alpha\}_{\alpha\in\Lambda}$ in X is called *complete*, if

$$x = \sum_{\alpha\in\Lambda}\langle x, e_\alpha\rangle e_\alpha \quad \text{for every } x \in X\,.$$

By Theorem 2.15 that an orthonormal family $\mathcal{E} = \{e_\alpha\}_{\alpha\in\Lambda}$ in X is complete is equivalent to the follows propositions. (a) $\overline{\text{span}}(\mathcal{E}) = H$. (b) For each $x \in H$, the Parseval equality holds. (c) For every $x, y \in H$,

$$\langle x, y\rangle = \sum_{\alpha\in\Lambda}\langle x, e_\alpha\rangle\overline{\langle y, e_\alpha\rangle}\,.$$

An orthonormal family $\mathcal{E} = \{e_\alpha\}_{\alpha\in\Lambda}$ in X is called *total*,

$$\mathcal{E}^\perp = \{0\}\,.$$

Obviously, a completely orthonormal family must be total, and if X is a Hilbert space, a totally orthonormal set must be complete, which is also called a *Hilbert basis* of the Hilbert space. But in the general case, the converse doesn't hold well. For a counterexample, see the page 127 of 《实变函数论与泛函分析 *(下册)*》 by Daoxing Xia.

2.4 Isometric Isomorphism

Every mathematical theory has its concept of “isomorphism”. In topology there is homeomorphism and homotopy equivalence, algebra calls them isomorphisms. The basic idea is to define a map which preserves the basic structure of the spaces in the category. In this section, We introduce the isometric isomorphism between Hilbert spaces.

Let H and H' be inner product spaces over the same field $\mathbb{F}$. An *isometric isomorphism* between H and H' is a linear bijective isometry. In this case H and H' are said to be *isometricly isomorphic*.

It is easy to see that if $U : H \to H'$ is an isometric isomorphism, then so is $U^{-1} : H' \to H$. Similar such arguments show that the concept of "isometrically isomorphic" is an equivalence relation on inner product spaces.

Proposition 2.16. *Let $U : H \to H'$ be a linear transform . Then U is an isometry if anf only if U protects the inner products; i.e.,*

$$\langle Ux, Uy\rangle = \langle x, y\rangle\,,\ \textit{for all } x, y \in H\,.$$

Proof. If U protects the inner product, then clearly U protects the norm. On the other hand, if $\|Ux\| = \|x\|$ for all x, by polarization identity, we have

$$\begin{aligned}\operatorname{Re}\langle Ux, Uy\rangle &= \frac{\|Ux+Uy\|^2 - \|Ux-Uy\|^2}{4} \\ &= \frac{\|x+y\|^2 - \|x-y\|^2}{4} = \operatorname{Re}\langle x, y\rangle\,.\end{aligned}$$

Thus $\langle Ux, Uy\rangle = \langle x, y\rangle$ if H and H' are real Hilbert spaces. If H and H' are real Hilbert spaces, then

$$\operatorname{Im}\langle Ux, Uy\rangle = \operatorname{Re}\langle Ux, iUy\rangle = \operatorname{Re}\langle Ux, Uiy\rangle = \operatorname{Re}\langle x, iy\rangle = \operatorname{Im}\langle x, y\rangle\,.$$

Thus $\langle Ux, Uy\rangle = \langle x, y\rangle$. □

In linear algebra, we have learned that every two finite dimensional linear space are isomorphic if and only if they have the same dimension. Therefore, it's nature to do the same thing : give a proper definition about "definition" about Hilbert space. The Hilbert basis occurs to us mind easily.

Definition 2.6. The *dimension* of Hilbert space H is the cardinality of a Hilbert basis, and is denoted by $\dim H$.

Is the definition well-defined ?

Lemma 2.17. *Any two Hilbert bases of Hilbert space H have the same cardinality.*

Proof. Let $\{\varepsilon_i\}_{i\in I}$ and $\{e_j\}_{j\in J}$ be two Hilbert basis for H. For any $j \in J$, e_j has a Fourier expansion

$$e_j = \sum_{i\in I} \langle \varepsilon_i, e_j\rangle \varepsilon_i$$

Let $I_j = \{i \in I : \langle \varepsilon_i, e_j\rangle \neq 0\}$, then I_j is countable, and we have $\cup_{j\in J} I_j = I$. Therefore

$$|I| = \left| \bigcup_{j\in J} I_j \right| \leq \aleph_0 |J| = |J| \,.$$

For the same reason, $|J| \leq |I|$, so $|I| = |J|$. $\square$

Theorem 2.18. *H is an infinite dimensional Hilbert space. Then* $\dim H$ *is* $\aleph_0$ *if and only if H is separable.*

Proof. If $\dim H$ is $\aleph_0$, by Lemma 1.2 we have H is separable. If H is separable, then there is a countable dense subset $\{x_n\}_{n\geq 1}$. Without loss of generality, we assume it is linearly independent. Using Gram-Schmidt orthonormalisation procedure we get a countable Hilbert basis. $\square$

Now, we can give the isomorphism theorem in Hilbert space which likes to finite linear space .

Theorem 2.19. *Two Hilbert spaces are isomorphic if and only if they have the same dimension. Particularly, all separable infinite dimensional Hilbert spaces are isomorphic to ℓ^2.*

Proof. Let $\{e_i\}_{i\in I}$ be an orthonormal basis for H. Define $T : H \to \ell^2(I)$ by

$$Tx = (\langle x, e_i\rangle)_{i\in I} \text{ for each } x \in H.$$

It follows from Bessel's Inequality that the right hand side is in $\ell^2(I)$. We must show that T is a surjective linear isometry. Clearly, T is linear. By Theorem 2.12, T is a surjection. By Theorem 2.15, T isometry. $\square$

Example 2.8. If for each $k \in \mathbb{Z}$,

$$e_k(t) := \frac{1}{\sqrt{2\pi}}\, \mathrm{e}^{ikt}, \quad t \in [0, 2\pi]\,,$$

then $\{e_k : k \in \mathbb{Z}\}$ is a basis for $L^2([0,2\pi], \mathbb{C})$. In fact, by using trigonometric polynomials to uniformly approach continuous function, we can prove

$$\overline{\operatorname{span}}\{e_k\} = L^2([0,2\pi], \mathbb{C})\,.$$

Then for any $f \in L^2([0,2\pi], \mathbb{C})$, let

$$\widehat{f}(k) := \langle f, e_k \rangle = (2\pi)^{-1/2} \int_0^{2\pi} f(t) e^{-ikt} dt$$

is called the *kth Fourier coefficient of f*, k in $\mathbb{Z}$, and we have

$$f = \sum_{k=-\infty}^{\infty} \widehat{f}(k) e_k\,,$$

where this infinite series converges to f in the metric defined by the norm of $L^2([0,2\pi], \mathbb{C})$. This is the classic Fourier series.

For f in $L^2([0,2\pi], \mathbb{C})$, the function $\widehat{f} : \mathbb{Z} \to \mathbb{C}$ is called the Fourier transform of f; the map $U : L^2([0,2\pi], \mathbb{C}) \to l^2(\mathbb{Z})$ defined by $Uf = \widehat{f}$ is the Fourier transform. As we can see,

The Fourier transform is a linear isometry from $L^2([0,2\pi], \mathbb{C})$ onto $\ell^2(\mathbb{Z}, \mathbb{C})$.

2.5 Riesz's Representation Theorem

In a inner product space, we can introduce the notion of *orthogonality* of two vectors. Thanks to this fact, a Hilbert space may be identified with its dual space, i.e., the space of bounded linear functionals. This result is the representation theorem of F.Riesz, and the whole theory of Hilbert spaces is founded on this theorem.

Let $(X, \langle \cdot, \cdot \rangle)$ be an inner product space over a field $\mathbb{F}$. Choose and fix $y \in X$. Define a map $f_y : X \to \mathbb{F}$ by $f_y(x) = \langle x, y \rangle$. We claim that f_y is a bounded linear functional on X. In fact, for any $x \in X$, $|f_y(x)| = |\langle x, y\rangle| \leq \|x\|\|y\|$ (by the CBS Inequality). That is, f_y is bounded and $\|f_y\| \leq \|y\|$. Since

$$f_y(y) = \langle y, y\rangle = \|y\|^2 \Rightarrow \frac{|f_y(y)|}{\|y\|} = \|y\|$$

we have that $\|f_y\| = \|y\|$. The above observation simply says that each element y in an inner product space $(X, \langle\cdot,\cdot\rangle)$ determines a bounded linear functional on X. The following theorem asserts that if $X = H$ is a Hilbert space then the converse of this statement is true. That is, every bounded linear functional on a Hilbert space H is, in fact, determined by some element $y \in H$.

Theorem 2.20 (Riesz's Representation Theorem)**.** *Let H be a Hilbert space over $\mathbb{F}$. If $f : H \to \mathbb{F}$ is a bounded linear functional on H (i.e., $f \in H^*$), then there exists a uniquely determined vector $y = y_f \in H$ such that*

$$f(x) = \langle x, y\rangle \quad \textit{for all } x \in H \quad \textit{and } \|f\| = \|y\|\,.$$

Proof. The uniqueness of y is clear, since $\langle x, y-y'\rangle = 0$ for all $x \in X$ implies $y = y'$. To prove its existence, consider the null space $N = \ker(f)$. If $f = 0$ then take $y = 0$. Then assume that $f \neq 0$, and hence N is a closed proper subspace of H. By Theorem 2.9, there exists $z \in N^{\perp}$ and $z \neq 0$. Without loss of generality, let $\|z\| = 1$. Then define $f_z : H \to \mathbb{F}$ by $f_z(x) = \langle x, z\rangle$. We have shown that f_z is a bounded linear functional. Observe that

$$\ker(f) = N \subset \ker(f_z)\,,$$

then we have that there exists a constant $c \in \mathbb{F}$ so that $f_z = cf$. Since $f_z(z) = \langle z, z\rangle = \|z\|^2 = 1$, we get that $c = 1/f(z)$. Thus

$$f(x) = \frac{1}{f(z)}\langle x, z\rangle \quad \text{for all } x \in H\,.$$

Taking $y = \dfrac{1}{\bar{f(z)}}z$, the desired result follows. □

Remark 2.5. The conclusion of Riesz's representation theorem may fail if $(X, \langle\cdot,\cdot\rangle)$ is an incomplete inner product space. It's easy to give a counterexample. In fact, let H be the completion of X. Take any $z \in H\backslash X$ and define $f : X \to \mathbb{F}$ by $f(x) = \langle x, z\rangle$ for all x. Obviously, there does not exist an element $y \in X$ so that $\langle x, y\rangle = \langle x, z\rangle$ for all $x \in H$. If exists, since X is dense in H, we have $y = z$, which contradicts to $z \notin X$.

Let H be a Hilbert space. Riesz's representation theorem gives a norm-preserving, one-to-one correspondence $f \leftrightarrow y_f$ between H^* and H. By this correspondence, H^* may be identified with H as an abstract set but it is not allowed to identify, by this correspondence, H^* with H as linear spaces, since the correspondence $f \leftrightarrow y_f$ is *conjugate linear* :

$$(\alpha_1 f_1 + \alpha_2 f_2) \leftrightarrow (\bar{\alpha}_1 y_{f_1} + \bar{\alpha}_2 y_{f_3})$$

where α_1, α_2 are scalars. Indeed, the dual space H^* of H constitutes also a Hilbert space. Firstly, the operator norm on H^* satisfies the *parallelogram identity*. Secondly, in the case of complex field, for every $f, g \in H^*$,

$$\begin{aligned}
\langle f, g\rangle &= \frac{\|f+g\|^2 - \|f-g\|^2}{4} + i\left(\frac{\|f+ig\|^2 - \|f-ig\|^2}{4}\right) \\
&= \frac{\|y_f+y_g\|^2 - \|y_f-y_g\|^2}{4} + i\left(\frac{\|y_f-iy_g\|^2 - \|y_f+iy_g\|^2}{4}\right) \\
&= \frac{\|y_g+y_f\|^2 - \|y_g-y_f\|^2}{4} + i\left(\frac{\|y_g+iy_f\|^2 - \|y_g-iy_f\|^2}{4}\right) \\
&= \langle y_g, y_f\rangle = f(y_g)\,.
\end{aligned}$$

Moreover, any continuous linear functional T on the Hilbert space H^* is thus identified with a uniquely determined element y_T of H as follows:

$$T(f) = f(y_T) \quad \text{for all} \quad f \in H^*\,.$$

This fact will be referred to as the *reflexivity* of Hilbert spaces.

Chapter 3

Topological Vector Space and Locally Convex Space

There still are many important spaces carrying natural topologies that cannot be induced by norms. Here are some examples:

(a) $C(\Omega)$, the space of all continuous complex functions on some unbounded open set Ω in a euclidean space $\mathbb{R}^n$.

(b) $H(\Omega)$, the space of all holomorphic functions in some open set Ω in the complex plane.

(c) C_K^∞, the space of all infinitely differentiable complex functions on $\mathbb{R}^n$ that vanish outside some fixed compact set K with nonempty interior.

So we need a generalization of the concept of a Banach space to describe these spaces, that is topological vector space. As a special case for topological vector space, the locally convex spaces are encountered repeatedly when discussing weak topologies on a Banach space, sets of operators on Hilbert space, or the theory of distributions. We will only skim the surface of this theory, but it will treat locally convex spaces in sufficient detail as to enable us to understand the use of these spaces in the three areas of analysis just mentioned.

3.1 Elementary properties

A vector space equipped with a Hausdorff topology such that the linear structure and the topological structure are "vitally connected", is called a topological vector space. Here is a more precise way of stating the definition:

Definition 3.1. A *topological vector space* (*TVS*) is a vector space X together with a topology τ such that

(a) (X, τ) is Hausdorff space.

(b) the vector space operations are continuous with respect to τ.

and such topology τ is called a *vector topology* on X.

Remark 3.1. In many texts, (a) is omitted from the definition of a topological vector space. Since (a) is satisfied in almost every application, and since most theorems of interest require (a) in their hypotheses, it seems best to in clude it in the axioms. Later, we willsee that under condition (b), (X, τ) is regular space. So (a) can be reduced that X is T_1 space.

3.1.1 Invariance of the Local Base

Let X be a topological vector space. Associate to each $a \in X$ and to each scalar $\lambda \neq 0$ the *translation operator* T_a and the *multiplication operator* M_λ, by the formulas

$$T_a(x) = a + x, \quad M_\lambda(x) = \lambda x \quad (x \in X)$$

The following simple proposition is very important:

Proposition 3.1. *T_a and M_λ are homeomorphisms of X onto X.*

Proof. The vector space axioms alone imply that T_a and M_λ are one-to-one, that they map X onto X, and that their inverses are T_{-a} and $M_{1/\lambda}$, respectively. The assumed continuity of the vector space operations implies that all these mappings are continuous. Hence each of them is a homeomorphism. $\square$

As a consequence of this proposition, every vector topology τ is *translation-invariant* (or simply invariant, for brevity): A set $U \in \tau$ is open if and only if each of its translates $a + U$ is open. Thus τ is completely determined by any local base. In the vector space context, the term *local base* will always mean a local base at 0. A local base of a topological vector space X is thus a collection $\mathcal{B}$ of open neighborhoods of 0 such that every neighborhood of 0 contains a member of $\mathcal{B}$. The open sets of X are then precisely those that are unions of translates of members of $\mathcal{B}$.

We check the definition that addition is continuous, which means that the mapping

$$(x, y) \mapsto x + y$$

of the Cartesian product $X \times X$ into X is continuous: if $x_i \in X$ for $i = 1, 2$, and if U is a neighborhood of 0 there should exist open neighborhoods V_i of 0 such that $(x_1 + V_1) + (x_2 + V_2) \subset x_1 + x_2 + V$. Thus let $V = V_1 \cap V_2$, we have

$$V + V \subset U\,.$$

Similarly, the assumption that scalar multiplication ts continuous means that the mapping

$$(\lambda, x) \mapsto \lambda x$$

of $\mathbb{F} \times X$ into X is continuous. Thus for any U is a neighborhood of 0, then for some $\epsilon > 0$ and some open neighborhood V of 0, we have $\lambda V \subset U$ whenever $|\lambda| \leq \epsilon$. If we let

$$W = \bigcup_{|\lambda| \leq \epsilon} \lambda V\,,$$

then W is a balanced open neighborhood of 0 contained in U. Thus we have

Proposition 3.2. *Every topology vector space has a balanced local base.*

3.1.2 Separation Properties

Theorem 3.3. *X is a topological vector space, K is compact, C is closed, and $K \cap C = \emptyset$. Then there is a neighborhood V of 0 such that*

$$(K + V) \cap (C + V) = \emptyset\,.$$

Proof. For any $x \in K$, there is a balanced open neighborhood V_x of 0 such that $(x + V_x + V_x) \cap C = \emptyset$. Since K is compact, there are finitely many points $x_1, \ldots, x_n$ in K such that

$$K \subset (x_1 + V_{x_1}) \cup \cdots \cup (x_n + V_{x_n})$$

Put $V = V_{x_1} \cap \cdots \cap V_{x_n}$. Then

$$K + V \subset \bigcup_{i=1}^{n} (x_i + V_{x_i} + V) \subset \bigcup_{i=1}^{n} (x_i + V_{x_i} + V_{x_i})$$

and no term in this last union intersects $C+V$. This completes the proof. □

Remark 3.2. In the proof we have not uses the assumption that X is Hausdorff space. Let K be any single point, then we know X is regular. Thus X is T_1 imples X is Hausdorff, see Remark 3.1.

Some other simple properties is followed. We omit the proof

Proposition 3.4. *Let X be a TVS, $A, B \subset X$.*

(a) $\overline{A} = \cap(A + V)$, *where V runs through all neighborhoods of* 0.

(b) $\overline{A} + \overline{B} \subset \overline{A + B}$

(c) *If Y is a subspace of X, so is $\overline{Y}$.*

We can also discuss whether the convexity, balance still holds under the topology operations.

Proposition 3.5. *Let X be a TVS, K is subset of X.*

(a) *If K is convex, then so are $\overline{K}$ and K°.*

(b) *If B is balanced, so is $\overline{B}$. In addition, $0 \in B^\circ$ then B° is balanced.*

3.1.3 Types of Topological Vector Spaces

Let (X, τ) be a TVS. A subset B of X is called *bounded* if for any neighborhood U of 0, there exist some $\epsilon > 0$, depending on U, such that $|\lambda|B \subset U$ for all $|\lambda| \le \epsilon$. For example, every compact subset of X is bounded. Note that any subset of a bounded set is also bounded.

X is a topological vector space, with topology τ. We say

(a) X is *locally convex* if there is a local base $\mathcal{B}$ whose members are *convex*.

(b) X is *locally bounded* if 0 has a *bounded* neighborhood.

(c) X is *locally compact* if 0 has a neighborhood whose closure is compact.

(d) X is *metrizable* if τ is compatible with some metric d.

(e) X is an *F-space* if τ is induced by a complete *translation invariant* metric.

(f) X is a *Fréchet space* if X is a locally convex F-space.

(g) X is *normable* if a norm exists on X such that the metric induced by the norm is compatible with τ.

The terminology of (e) and (f) is not universally agreed upon: In some texts, local convexity is omitted from the definition of a Frechet space, whereas others use F-space to describe what we have called Frechet space.

Relations. Here is a list of some relations between these properties of a topological vector space X.

(a) If X is locally bounded, then X has a countable local base.

(b) X is metrizable *iff* X has a countable local base.

(c) X is normable *iff* X is locally convex and locally bounded.

(d) X has finite dimension *iff* X is locally compact.

(e) If a locally bounded space X has the Heine-Borel property, i.e., every closed bounded subset of X is compact, then X has finite dimension.

The spaces $H(\Omega)$ and C_B^∞ mentioned before are infinite-dimensional Fréchet spaces with the Heine-Borel property, they are therefore not locally bounded, hence not normable; they also show that the converse of (a) is false. On the other hand, there exist locally bounded F-spaces that are not locally convex : L^p space when $p \in (0,1)$.

3.2 Locally Convex Spaces

Firstly, we give an example of locally convex space. Let X be a vector space. Since that for any seminorm p on X and for $\epsilon > 0$,

$$V(p,\epsilon) \coloneqq \{x \in X : p(x) < \epsilon\}$$

is a balanced absorbing convex set. If we use these sets to induce a vector topology on X, it must be locally convex.

Example 3.1. X be a vector space. Let $\mathcal{P}$ be a family of seminorms on X. Let τ be the weakest topology on X satisfying that p is continuous for each $p \in \mathcal{P}$. What do the basic open sets in this topology look like? To answer this equation, we are going to find a base of τ. For any $x_0 \in X$, let

$$V(x_0;\Phi,\epsilon) \coloneqq \{x \in X : p(x - x_0) < \epsilon \text{ for all } p \in \Phi\}\,,$$

where Φ is a finite subset of $\mathcal{P}$ and $\epsilon > 0$. Clearly, $V(x_0;\Phi,\epsilon) \in \tau$. Moreover, τ is the weakest topology containing $\{V(x_0;\Phi,\epsilon)\}$. To see this, it suffices to show that every p is continuous relative to the topology generated by $\{V(x_0;\Phi,\epsilon)\}$. Take any $x_0 \in X$ and $\epsilon > 0$, then

$$\{x : |p(x) - p(x_0)| < \epsilon\} = \bigcup_{x:|p(x)-p(x_0)|<\epsilon} V(x;p,\epsilon - |p(x) - p(x_0)|)\,.$$

Thus τ is generated by $\{V(x_0; \Phi, \epsilon)\}$.

Recall that a collection $\mathfrak{B}$ of subsets of a set X is a base for a topology on X if and only if

- $X = \cup\{B : B \in \mathfrak{B}\}$; i.e., each $x \in X$ belongs to some $B \in \mathfrak{B}$, and
- if $x \in B_1 \cap B_2$ for some B_1 and B_2 in $\mathfrak{B}$, then there is a $B_3 \in \mathfrak{B}$ such that $x \in B_3 \subset B_1 \cap B_2$.

Then, it's obvious to show that indeed

$$\mathfrak{B} = \{V(x; \Phi; \epsilon) : x \in X,\ \Phi \subset \mathcal{P} \text{ is finite}, \epsilon > 0\}$$

is a base for the topology τ. We write $V(0, \Phi, \epsilon)$ as $V(\Phi, \epsilon)$ for brevity. Then one can see that

$$\mathcal{B} = \{V(\Phi, \epsilon) :\ \Phi \subset \mathcal{P} \text{ is finite, } \epsilon > 0\}$$

is a balanced convex open neighborhood base of 0. Besides, it's easy to find that,

$$V(x_0; \Phi, \epsilon) = x_0 + V(\Phi, \epsilon)\,.$$

We now show that addition and the scalar multiplication are continuous relative to τ. In fact, for any $V(\Phi, \epsilon)$, $V(\Phi, \epsilon/2) + V(\Phi, \epsilon/2) \subset V(\Phi, \epsilon)$, so addition is continuous. Pick $\alpha_0 \in \mathbb{F}$ and $x_0 \in X$, for any $\alpha_0 x_0 + V(\Phi, \epsilon)$, let $|\alpha - \alpha_0| \leq \delta_1$ and $x \in x_0 + V(\Phi, \delta_2)$, then for all $p \in \Phi$,

$$\begin{aligned} p(\alpha x - \alpha_0 x_0) &\leq |\alpha - \alpha_0| p(x_0) + \alpha p(x - x_0) \\ &< \delta_1 p(x_0) + (|\alpha_0| + |\delta_1|)\delta_2 < \epsilon\,. \end{aligned}$$

Thus the scalars multiplication is continuous.

Now we try to make (X, τ) be a TVS, then since 0 has a (balanced) convex local base, it is a LCS. To this end, it suffices to make (X, τ) be a Hausdorff space. We need to assume that $\mathcal{P}$ is *separating*:

$$\bigcap_{p \in \mathcal{P}} \{x : p(x) = 0\} = \{0\}\,. \tag{3.1}$$

In fact, suppose that $x \neq y$. Then there is a p in $\mathcal{P}$ such that $p(x-y) \neq 0$, pick an $\epsilon > 0$ such that $p(x-y) > 2\epsilon$. Then $x + V(p,\epsilon)$ and $y + V(p,\epsilon)$ are disjoint neighborhoods of x and y, respectively. Conversely, it's easy to check that this condition is necessary.

Thus when $\mathcal{P}$ is separating, (X,τ) is a LCS. Since τ is induced by the separating family $\mathcal{P}$ of seminorms, there are two interesting consequences.

(i) $B \subset X$ is bounded if and only if $\{p(x) : x \in B\}$ is bounded for any $p \in \mathcal{P}$.

(ii) A net $\{x_i\}$ in X is convergent to some x_0 if and only if $p(x_i) \to p(x_0)$ for all $p \in \mathcal{P}$.

If $\mathcal{P}$ is a family of seminorms of X that makes X into a LCS, it is often convenient to enlarge $\mathcal{P}$ by assuming that $\mathcal{P}$ is closed under the formation of finite sums and supremums of "bounded families". Sometimes it is convenient to assume that $\mathcal{P}$ consists of all continuous seminorms. In either case the resulting topology on X remains unchanged.

Seminorms and Local Convexity. We shall show that any LCS has the same structure as in Example 3.1.

Proposition 3.6. *Let X be a TVS and let p be a seminorm on X. The following statements are equivalent. (a) p is continuous. (b) $0 \in \operatorname{int}V(p,1)$. (c) p is continuous at 0.*

Proof. (a) implies (b) is obvious.

To show (b) implies (c), since $0 \in \operatorname{int}\{x : p(x) < 1\}$, there exists some neighborhood U of 0 contained in $V(p,1)$. Note that $p(0) = 0$, for any $\epsilon > 0$, each $x \in \epsilon U \subset \epsilon V(p,1) = V(p,\epsilon)$, so p is continuous at 0.

(c) implies (a) is obvious. □

Corollary 3.7. *Let X be a TVS and let p be a seminorm. Then p is continuous iff there is a continuous seminorm q such that $p \leq q$.*

Theorem 3.8. *Every locally convex space has a balanced convex local base.*

Proof. Suppose U is a convex neighborhood of 0. Note that U is convex and absorbing, then $rU \subset U$ for any $0 \le r \le 1$. A balanced convex open neighborhood of 0 contained in U must contained in

$$V = \bigcap_{|\alpha|=1} \alpha U .$$

Firstly, being an intersection of convex sets, V is convex. Secondly, there is a balanced neighborhood W of 0 contained in U, then $\alpha W = W \subset \alpha U$ for $|\alpha| = 1$, thus $W \subset V$, Thus, the interior V° of V is a convex neighborhood of 0.

Besides, we can see that V is balanced: choose r and β so that $0 \le r \le 1, |\beta| = 1$.

$$r\beta V = \bigcap_{|\alpha|=1} r\beta\alpha U = \bigcap_{|\alpha|=1} r\alpha U \subset V$$

Then V° is balanced since $0 \in V^\circ$. Thus, V° is balanced convex open neighborhood of 0 contained in U, so LCS has a balanced convex local base. □

Theorem 3.9. *Suppose $\mathcal{B}$ is a convex balanced local base in a LCS (X, τ). Associate to every $U \in \mathcal{B}$ its Minkowski functional p_U. Then*

$$\mathcal{P} = \{p_U : U \in \mathcal{B}\}$$

is a separating family of continuous seminorms on X, which induced τ.

Proof. Firstly, we show that $\mathcal{P}$ is separating. If $p_U(x) = 0$ for all $U \in \mathcal{B}$, $x \in U$ for all U. Since τ is Hausdorff, $x = 0$.

Secondly, let τ_1 is the topology induced by p, we show $\tau_1 = \tau$. Since U is open, U is absorbing at each of it's point, then $V(p_U, 1) = U$ by Lemma 1.23, then p_U is continuous for all $U \in \mathcal{B}$. Thus $\tau_1 \subset \tau$. Note that τ is determined by it's local base $\mathcal{B}$, and for any $U \in \mathcal{B}$, $U = V(p_U, 1) \in \tau_1$, so $\tau \subset \tau_1$. □

3.2.1 Metrizable Locally Convex Spaces

We recall that a topology τ on a set X is said to be metrizable if there is a metric d on X which is compatible with τ. In that case, the balls with radius $1/n$ centered at x form a local base at x. This gives a necessary condition for metrizability which, for topological vector spaces, turns out to be also sufficient.

Which LCS's are metrizable? That is, which have a topology which is defined by a metric? Which LCS's have a topology that is defined by a norm? Both are interesting questions and both answers could be useful.

If $\mathcal{P}$ is a family of seminorms on topological vector X, say that $\mathcal{P}$ determines the topology on X if the topology of X is the same as the topology induced by $\mathcal{P}$.

Lemma 3.10. *Let (X, τ) be a LCS , $\mathcal{P} = \{p_n : n \in \mathbb{N}\}$ is countable separating family of seminorms which induces τ. For x and y in X, define*

$$d(x, y) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{p_n(x-y)}{1 + p_n(x-y)} .$$

Then d is a translation invariant metric on X, and the topology induced by d coincides with τ. Moreover, $\{B(0, \frac{1}{n}) : n \geq 1\}$ is a balanced (may NOT convex) local base τ.

Proof. It's easy to check that d is a translation invariant metric on X. We only show that the topology induced by the metric d, denoted by τ_d, coincides with τ. Firstly, in order that $\tau_d \subset \tau$, it suffices to show that for every fixed y, the (open) ball centered at y with radius $r > 0$ is in τ, i.e.,

$$B(y, r) \coloneqq \{x : d(x, y) < r\} \in \tau .$$

Observe that for fixed $y \in X$, the series

$$\sum_{n=1}^{\infty} \frac{\min\{p_n(x-y), 1\}}{2^n} \text{ converges to } d(x, y) \text{ uniformly in } x .$$

Since each p_n is continuous relative to τ, the map $x \mapsto d(x,y)$ is continuous relative to τ for fixed y. Thus $B(y,r) \in \tau$.

On the other hand, in order that $\tau \subset \tau_d$, it suffices to show that for each fixed n, p_n is continuous relative to τ_d. For any gien $\epsilon \in (0, 1/2^n)$, note that

$$p_n(x) < \epsilon \quad \text{for all} \quad d(x,0) < \frac{1}{2^n}\frac{\epsilon}{1+\epsilon}\,.$$

Thus p_n is continuous at 0 relative to τ_d. By Proposition 3.6, p_n is continuous.

Finally, we show that $B(0,1)$ is balanced and convex. For $x \in B(0,1)$ and $|\lambda| \leq 1$, since $x \mapsto \frac{x}{1+x}$ is increasing on $[0,\infty)$, we have

$$d(\lambda x, 0) = \sum_{n=1}^{\infty} \frac{1}{2^n}\frac{|\lambda| p_n(x)}{1+|\lambda| p_n(x)} \leq \sum_{n=1}^{\infty} \frac{1}{2^n}\frac{p_n(x)}{1+p_n(x)} = d(x,0) < 1\,.$$

So $\lambda x \in B(0,1)$. Thus $B(0,1)$ is balanced. $\square$

Remark 3.3. We should emphasize that the open ball relative to such a d may not be convex. Indeed, for every $x, y \in B(0,1)$ and $\lambda \in (0,1)$, we want

$$\begin{aligned} d(\lambda x + (1-\lambda)y, 0) &= \sum_{n=1}^{\infty} \frac{1}{2^n}\frac{p_n(\lambda x + (1-\lambda)y)}{1+p_n(\lambda x + (1-\lambda)y)} \\ &\leq \sum_{n=1}^{\infty} \frac{1}{2^n}\left[\frac{\lambda p_n(x)}{1+p_n(x)} + \frac{(1-\lambda)p_n(y)}{1+p_n(y)}\right] \\ &= \lambda d(x,0) + (1-\lambda)d(y,0)\,. \end{aligned}$$

In order that the argument valid, we need the inequality

$$\frac{p_n(\lambda x + (1-\lambda)y)}{1+p_n(\lambda x + (1-\lambda)y)} \leq \frac{\lambda p_n(x)}{1+p_n(x)} + \frac{(1-\lambda)p_n(y)}{1+p_n(y)}\,.$$

But take $y = 0$, the inequality is in the opposite direction.

However, there does exist a translation invariant metric ρ on X compatible with the topology τ so that all open balls relative to ρ centered at 0 are balanced and convex. See 1.24 Theorem in *Functional Analysis* by W.Rudin.

If $\mathcal{P}$ is a family of seminorms on LCS X, say that $\mathcal{P}$ determines the topology on X if the topology of X is the same as the topology induced by $\mathcal{P}$.

Theorem 3.11 (Metrizable LCS)**.** *Let (X,τ) be a locally convex space. Then the followings are equivalent.*

(a) *(X,τ) is measurable.*

(b) *(X,τ) is first-countable. In other words, (X,τ) has a countable local base.*

(c) *τ is determined by a countable separating family of seminorms.*

Proof. $(a) \Rightarrow (b)$ is obvious. $(c) \Rightarrow (a)$ was proved in Lemma 3.10. We only need to show that $(b) \Rightarrow (c)$. To this end, assume that $\{U_n\}$ is a local base of τ. Given $n \geq 1$, because (X,τ) is locally convex, there are continuous seminorms $\Phi_n = \{q_{n,1}, \ldots, q_{n,k}\}$ and positive real number ϵ_n such that $V(\Phi_n, \epsilon_n) \subset U_n$. Let

$$p_n = \frac{1}{\epsilon_n}(q_{n,1} + \cdots + q_{n,k})\,.$$

Then $x \in U_n$ whenever $p_n(x) < 1$, i.e.,

$$V(p_n, 1) \subset U_n\,. \tag{3.2}$$

We claim that $\{p_n\}$ determines the topology τ. Denote by τ' the vector topology induced by $\{p_n\}$. Since p_n is continuous with respect to τ for each n, there must be $\tau' \subset \tau$. To show the other direction, note that (3.2) implies that $\{V(p_n, 1)\}$ is a local base of τ. Therefore, $\tau \subset \tau'$. We now complete the proof. $\square$

Remark 3.4. In fact, every first-countable topological vector space is metrizable. Specifically, if (X,τ) is a topological vector space with a countable local base, then there exists a translation invariant mertic ρ on X compatible with the topology τ so that all open balls relative to ρ centered at 0 are balanced. See 1.24 Theorem in *Functional Analysis* by W.Rudin.

3.2.2 Normable Locally Convex Spaces

Recall that if X is a TVS and $B \subset X$, then B is *bounded* if for every open set U containing 0, there is an $\epsilon > 0$ such that $\lambda B \subset U$ for all $|\lambda| \leq \epsilon$. When X is metrizable, there is a possibility of misunderstanding, since another very familiar notion of boundedness exists in metric spaces:

If d is a metric on a set X, a set $E \subset X$ is said to be d-bounded if there is a number $M < \infty$ such that $d(x, y) \leq M$ for all x and y in E.

If X is a topological vector space with a compatible metric d, the bounded sets and the d-bounded ones need NOT be the same, even if d is invariant. For instance, if d is a metric such as the one constructed in Lemma 3.10, then X itself is d-bounded (with $M = 1$) but, as we shall see presently, X cannot be bounded, unless $X = \{0\}$. If X is a normed space and d is the metric induced by the norm, then the two notions of boundedness coincide: B is bounded lerative to the topology if and onley if $\sup_{b \in B} \|b\| < \infty$; but if d is replaced by $d_1 = d/(1 + d)$ (an invariant metric which induces the same topology they do not.

Theorem 3.12 (Normable LCS)**.** *Let (X, τ) be a LCS. Then (X, τ) is normable if and only if X locally bounded, i.e., 0 has a bounded open neighborhood.*

Proof. We have shown that the open unit ball in a normed space is bounded. So assume that X is a LCS that 0 has a bounded open neighborhood U. It must be shown that there is a norm on X that defines the same topology. By local convexity, there is a continuous seminorm p so that $V(p, 1) \subset U$. It will be shown that p is a norm and defines the topology on X.

To see that p is a norm, we need to show that $p(x) = 0$ implies $x = 0$. For every neighborhood W of 0, there exists $\epsilon > 0$ so that $\epsilon U \subset W$. Observing that

$$x \in V(p, \epsilon) = \epsilon V(p, 1) \subset \epsilon U \subset W, \tag{3.3}$$

since W is arbitary and X is a Hausdorff space, we have $x = 0$. Thus p is a norm. Denote by τ' the topology induced by the norm p. Because p is continuous on (X, τ), $\tau' \subset \tau$. On the orther hand, it follows from (3.3) that $\{V(p, \epsilon)\}$ is a local base of τ. Thus $\tau \subset \tau'$. □

3.3 Linear Mappings

Here are some properties of linear mappings $T : X \to Y$ whose proofs are so easy that we omit them; it is assumed that $A \subset X$ and $B \subset Y$:

(a) If A is a subspace (or a convex set, or a balanced set) the same is true of $T(A)$.

(b) If B is a subspace (or a convex set, or a balanced set) the same is true of of $T^{-1}(B)$.

The following proposition is obviously .

Proposition 3.13. *Let X and Y be topological vector spaces. If $T : X \to Y$ is linear and continuous at* 0, *then T is continuous.*

Now we turn to discuss continuous linear functionals on TVS, namely (X, τ), and we denote by $(X, \tau)^*$ (or X^* for short) all the continuous linear functional on (X, τ).

Theorem 3.14. *Let X be a TVS. Let f be a nonzero linear functional on X. Then each of the following three properties implies the other three:*

(a) f is continuous.

(b) The null space $\ker f$ *is closed.*

(c) *f is locally bounded in the following sense: there exists $M > 0$ and a neighborhood V of 0 such that $|f(x)| \leq M$ for all $x \in V$.*

Proof. (a) implies (b) is obviously.

To show (b) implies (c), since $\ker f$ is proper closed subspace, pick any $x \notin \ker f$. Then there exists a balanced neighborhood V of 0, suth that

$$(x + V) \cap \ker f = \emptyset .$$

Since $f(V)$ is balanced subset of $\mathbb{F}$, and $-f(x) \notin f(V)$, $f(V)$ must be bounded.

(c) implies (a) : if (c) holds, then $|f(x)| < M$ for all x in V and for some $M < \infty$. Then For any $\epsilon > 0$, let

$$W = \frac{\epsilon}{M} V ,$$

then $|f(x)| < \epsilon$ for every x in W, hence f is continuous at the origin, and then f is continuous. □

Theorem 3.15. *Let X be a LCS, $\mathcal{P}$ is a separating family of seminorms that defines the topology on X, and f a linear functional on X. Then f is continuous if and only if there are $p_1, \ldots, p_n$ in $\mathcal{P}$ and positive scalars $c_1, \ldots, c_n$ such that*

$$|f(x)| \leq \sum_{k=1}^{n} c_k p_k(x) , \quad \textit{for all } x \in X.$$

Proof. The sufficiency is trivial. We show the necessity. If f is continuous, then there exists $\Phi = \{p_1, \cdots, p_n\} \subset \mathcal{P}$ and $\epsilon > 0$ such that

$$V(\Phi, \epsilon) \subset \{x : |f(x)| < 1\} .$$

Let

$$q(x) := \sum_{k=1}^{n} \frac{1}{\epsilon} p_k(x) , \quad \text{for all } x \in X.$$

Obviously, q is a (continuous) seminorm, and

$$V(q,1) \subset \{x : |f(x)| < 1\} .$$

Thus for any $x \in X$, and $\delta > 0$,

$$q\left(\frac{x}{q(x)+\delta}\right) < 1 \Rightarrow \frac{|f(x)|}{q(x)+\delta} < 1 \Rightarrow |f(x)| \leq q(x) . \qquad \square$$

Using the same argument, we can show that

Theorem 3.16. *Let X, Y be two LCS. Suppose $\{p_\alpha\}$, $\{p_\beta\}$ determine the vector topology on X, Y, respectively. Let $T : X \to Y$ be a linear operator. Then T is continuous if and only if for every β, there are $\alpha_1, \ldots, \alpha_n$ and positive scalars $c_1, \ldots, c_n$ such that*

$$|p_\beta(Tx)| \leq \sum_{k=1}^{n} c_k p_{\alpha_k}(x) , \quad \textit{for all } x \in X.$$

3.4 Finite-Dimensional Spaces

We have shown that on finite-dimensional vector space X, all the norm topologys are the same one. It's natural to ask is it to for vector topologys?

Lemma 3.17. *X is TVS, and $f : \mathbb{F}^n \to X$ is linear, then f is continuous.*

Proof. Let $\{e_1, \ldots, e_n\}$ be the standard basis of $\mathbb{F}^n$. Put $u_k = f(e_k)$, for $k = 1, \ldots, n$ Then

$$f(z) = z_1 u_1 + \cdots + z_n u_n \text{ for every } z = (z_1, \ldots, z_n) \in \mathbb{F}^n.$$

Every z_k is a continuous function of z. The continuity of f is therefore an immediate consequence of the fact that addition and scalar multiplication are continuous in X. $\square$

Theorem 3.18. *X is an n-dimensional TVS. Then every isomorphism of $\mathbb{F}^n$ onto X is a linear homeomorphism.*

Proof. Let S be the sphere which bounds the open unit ball B of $\mathbb{F}^n$, i.e.,

$$S = \{\lambda \in \mathbb{F}^n : \Sigma \left|\lambda_i\right|^2 = 1, \}$$

Suppose $f : \mathbb{F}^n \to X$ is an isomorphism, That is, f is a linear bijection. Since f is continuous, $f(S)$ is compact. Note that $0 \notin f(S)$, then there is a balanced neighborhood V of 0 in X which does not intersect $f(S)$, i.e.,

$$V \cap f(S) = \emptyset\,.$$

The set

$$E = f^{-1}(V)$$

is therefore disjoint from S. Since f is linear, V is balanced, hence E is balanced. Note that $0 \in E$, thus $E \subset B$. This implies that the linear map f^{-1} takes V into B. This implies that f^{-1} is locally bounded, by Theorem 3.14 (c) we have f^{-1} is continuous. Thus f is a homeomorphism. □

Corollary 3.19. *X is a finite-dimensional vector space and τ_1, τ_2 are two vector topologys on X, then $\tau_1 = \tau_2$.*

Proof. Let f be a linear isomorphism between X and $\mathbb{F}^n$, then f is linear homeomorphism betwwen $\mathbb{F}^n$, (X, τ_1) and (X, τ_2), so $\tau_1 = \tau_2$. □

Next, we shall give a topological characterization of the algebraic concept of finite dimensionality, as a generalization of Theorem 1.12.

Theorem 3.20. *TVS has finite dimension iff it is locally compact.*

Proof. The origin of X has a neighborhood V whose closure is compact. V is bounded, and the sets $2^{-n}V(n = 1, 2, 3, \ldots)$ form a local base for X. The compactness of $\overline{V}$ shows that there exist $x_1, \ldots, x_m$ in X such that

$$\overline{V} \subset \left(x_1 + \frac{1}{2}V\right) \cup \cdots \cup \left(x_m + \frac{1}{2}V\right)$$

Let Y be the vector space spanned by $x_1, \ldots, x_m$. Then $\dim Y \le m$. Thus Y is a closed subspace of X. Since $V \subset Y + \frac{1}{2}V$ and since $\lambda Y = Y$ for every scalar $\lambda \neq 0$, it follows that

$$\frac{1}{2}V \subset Y + \frac{1}{4}V$$

If we continue in this way, we see that

$$V \subset \bigcap_{n=1}^{\infty} \left(Y + 2^{-n}V\right)$$

Since $\{2^{-n}V\}$ is a local base, we have $V \subset \overline{Y}$. But $\overline{Y} = Y$. Thus $V \subset Y$, which implies that $kV \subset Y$ for $k \in \mathbb{N}_+$. But $X = \cup_{k=1}^{\infty} kV$, thus $Y = X$.. $\square$

Corollary 3.21. *If X is a locally bounded topological vector space with the Heine-Borel property: every closed and bounded subset of X is compact. Then X has finite dimension.*

3.5 Quotient spaces

Let N be a subspace of a vector space X. For every $x \in X$, let $Q(x)$ (sometimes write $\tilde{x}$ or $[x]$) be the coset of N that contains x, thus

$$Q(x) = x + N.$$

These cosets are the elements of a vector space X/N, called the quotient space of X modulo N, in which addition and scalar multiplication are defined by

$$Q(x) + Q(y) = Q(x + y), \quad \alpha Q(x) = Q(\alpha x)$$

Since N is a vector space, the operations are well defined.

The origin of X/N is $Q(0) = N$. Q is a linear mapping of X onto X/N with N as its null space. Q is often called the *quotient map* or the *natural map* of X onto X/N.

Suppose now that τ is a vector topology on X and that M is a subspace of X. Let τ_M be the *quotient topology* on X/M. That is

$$\tau_M = \{U \in X/M : Q^{-1}(U) \in \tau\}\,. \tag{3.4}$$

To guarantee that τ_M is T_1, $Q^{-1}(Q(x)) = x + M$ must be closed in τ, which is equivalent to that M is *closed*. Henceforth in this section we always assume M is a closed subspace of X.

Theorem 3.22. *Let M be a closed subspace of a TVS (X, τ). Let τ_M is quotient topology on X/M. Then the following propositions hold.*

(a) The quotient map $Q : X \to X/M$ is linear continuous open mapping.

(b) τ_M is a vector topology on X/M.

(c) If $\mathcal{B}$ is a local base for τ, then $Q(\mathcal{B})$ is a local base for $(X/M, \tau_M)$.

Proof. To show (a), note that the continuity of Q follows directly from the definition of τ_M. Mext, suppose $V \in \tau$. Since

$$Q^{-1}(Q(V)) = M + V$$

and $M + V \in \tau$, it follows that $Q(V) \in \tau_M$. Thus Q is an open mapping.

To show (b), if now W is a neighborhood of 0 in X/M, there is a neighborhood V of 0 in X such that

$$V + V \subset Q^{-1}(W)$$

Hence $Q(V) + Q(V) \subset W$. Since Q is open, $Q(V)$ is a neighborhood of 0 in X/M. Addition is therefore continuous in X/M. The continuity of scalar multiplication in X/M is proved in the same manner. This establishes (b).

It is clear that (a) implies (c). $\square$

Corollary 3.23. *Suppose M and F are subspaces of a topological vector space X, M is closed, and F has finite dimension. Then $M + F$ is closed.*

Proof. Let Q be the quotient map of X onto X/M, and give X/M its quotient topology. Then $Q(F)$ is a finite-dimensional subspace of X/M. Since X/M is TVS, $Q(F)$ is subspace of X/M, so is closed in X/M. Since $M + F = Q^{-1}(Q(F))$ and Q is continuous, we conclude that $M + F$ is closed. $\square$

Corollary 3.24. *Each of the following properties of X is inherited by X/N: local convexity, local bounded ness, metrizability, normability.*

Proof. Since all of these properties are determined by the local base, and Q do not change these properties. $\square$

Other Definitions. If X is a LCS and $\mathcal{P}$ is the separating family of seminorms on X, induced the topology. For any seminorm p on X, define $\tilde{p}$ on X/M by

$$\tilde{p}(\tilde{x}) = \inf\{p(x+m) : m \in M\}\,.$$

then $\tilde{p}$ is a seminorm on X/M. The family $\tilde{\mathcal{P}} \coloneqq \{\tilde{p} : p \in \mathcal{P}\}$ is a separating family of seminorms on X/M, and it follows from Theorem 3.22 (c) that $\tilde{\mathcal{P}}$ induces the quotient topology on X/M.

Suppose next that d is an invariant metric on X, compatible with τ. Define ρ by

$$\tilde{\rho}(\tilde{x}, \tilde{y}) = d(x-y, M) = \inf\{d(x-y, m) : m \in M\}\,,$$

is interpreted as the distance from $x - y$ to M. We omit the verifications that are now needed to show that $\tilde{\rho}$ is well defined and that it is an invariant metric on X/N. Since

$$Q(\{x : d(x,0) < r\}) = \{\tilde{x} : \tilde{\rho}(\tilde{x}, 0) < r\}\,,$$

it follows from Theorem 3.22 (c) that $\tilde{\rho}$ is compatible with τ_M.

If X is normed, this definition of ρ specializes to yield what is usually called the quotient norm of X/M :

$$\|\tilde{x}\| = \inf\{\|x+m\| : m \in M\}\,.$$

3.6 Hahn-Banach theorems

3.6.1 Separation theorems

Theorem 3.25. *X is TVS, A and B are two disjoint, nonempty, convex subsets. If A is open, there exist $\ell \in X^*$ and $c \in R$ such that*

$$\operatorname{Re}\ell x < c \leq \operatorname{Re}\ell y \tag{3.5}$$

for every $x \in A$ and for every $y \in B$.

Proof. It is enough to prove this for real scalars.

Fix $a_0 \in A, b_0 \in B$,and put $x_0 = b_0 - a_0$, put $C = A - B + x_0$. Then C is a convex open neighborhood of 0 in X. Denote by p the Minkowski functional of C. Since $x_0 \notin C$, $p(x_0) \geq 1$.

Just like in Theorem 1.25, we define ℓ on $\operatorname{span}\{x_0\}$ by

$$\ell(ax_0) = a\,,\ \text{for all}\ \ a \in \mathbb{R}\,.$$

Then for all such ax_0

$$\ell(ax_0) \leq p_K(ax_0)\,.$$

So by Theorem 1.18, ℓ can be extened as a linear functional on X dominated by p. Note that $p(x) \leq 1$ for all $x \in C$, then

(i) ℓ is continuous, since ℓ is locally bounded (Theorem 3.14 (c)). indeed,

$$|\ell(x)| < 1\,,\quad \text{for all } x \in C \cap (-C).$$

(ii) $\ell(a) < \ell(b)$ for all $a \in A, b \in B$, since

$$\ell a - \ell b + 1 = \ell\,(a - b + x_0) \leq p\,(a - b + x_0) < 1$$

It follows that $\ell(A)$ and $\ell(B)$ are disjoint convex subsets of $\mathbb{R}$, with $\ell(A)$ to the left of $\ell(B)$.

The key is that: *every nonconstant linear functional on X is an open mapping.* Since $\ell(A) = \tilde{\ell}(Q(A))$, where $Q : X \to X/\mathrm{ker} f$ and $\tilde{\ell}(\tilde{x}) = \ell(x)$, Q is open mapping, $\tilde{\ell}$ is linear homeomorphism between $X/\mathrm{ker}\ell$ and $\mathbb{R}$, by Theorem 3.18. (See Figure 3.1) Thus $\ell(A)$ is an open convex set, c be the right end point of $\ell(A)$ to get the conclusion of (3.5). $\square$

$$\begin{tikzcd}
X \arrow[r, "Q"] \arrow[dr, "l"'] & \frac{X}{\ker l} \arrow[d, "\tilde{l}"] \\
& \mathbb{R}
\end{tikzcd}$$

Figure 3.1: ℓ is open mapping

Theorem 3.26. *X is LCS, A, B are two disjoint, nonempty, convex subsets. If A is compact, and B is closed, then there exist $\ell \in X^*$ and $c \in \mathbb{R}$ such that*

$$\operatorname{Re} \ell x < c < \operatorname{Re} \ell y \tag{3.6}$$

for every $x \in A$ and for every $y \in B$.

Proof. It is enough to prove this for real scalars.

By Theorem 3.3, and X is locally convex, there is a convex open neighborhood V of 0 such that

$$(A + V) \cap B = \emptyset.$$

With $A+V$ in place of A, Theorem 3.25 shows that there exists $\ell \in X^*$ such that $\ell(A+V)$ and $\ell(B)$ are disjoint convex subsets of $\mathbb{R}$, with $\ell(A+V)$ open (See Figure 3.1) and to the left of $\ell(B)$. Note that $\ell(A)$ is compact, we obtain the conclusion. $\square$

Corollary 3.27. *X is a LCS, then X^* separates points on X.*

Remark 3.5. In Theorem 3.26,

(a) the hypothesis that X is locally convex does not appear in the preceding results, since the existence of an open convex subset of X is assumed. In this theorem such a set must be manufactured. Without the hypothesis of local convexity it may be that the only open convex sets are the whole space itself and the empty set, see Example 3.2 ;

(b) the fact that one of the two closed convex sets in the preceding theorem is assumed to be compact is necessary. In fact, if $X = \mathbb{R}^2$, $A = \{(x, y) \in \mathbb{R}^2 : y \leq 0\}$, and $B = \left\{(x, y) \in \mathbb{R}^2 : y \geq x^{-1} > 0\right\}$, then A and B are disjoint closed convex subsets of $\mathbb{R}^2$ that cannot be strictly separated.

Example 3.2. For $0 < p < 1$, let $L^p(0,1)$ be the collection of equivalence classes of measurable functions $\ell : (0,1) \to \mathbb{R}$ such that

$$((\ell))_p = \int_0^1 |\ell(x)|^p dx < \infty$$

It will be shown that $d(f,g) = ((f-g))_p$ is a metric on $L^p(0,1)$ and that with this metric $L^p(0,1)$ is a F-space. It will also be shown, however, that $L^p(0,1)$ has only one nonempty open convex set, namely itself. So $L^p(0,1)$ $0 < p < 1$, is most emphatically not locally convex.

Theorem 3.28. *X is a LCS, M is a closed linear subspace of X, and $x_0 \notin M$, then there is a nonzero continuous linear functional $\ell \in X^*$ such that (a) $\ell(x_0) = 1$, (b) $\ell(y) = 0$ for all y in M.*

Proof. By Theorem 3.26, there is nonzero $\ell \in X^*$ such that $\ell(x_0)$ and $\ell(M)$ are disjoint. Hence, $\ell(M)$ must be a *proper subspace* of the scalar field. This forces

$$\ell(M) = \{0\} \text{ and } \ell(x_0) \neq 0 .$$

The desired functional is obtained by dividing ℓ by $\ell(x_0)$. □

There is an another useful corollary of the separation theorem.

Theorem 3.29. *X is a LCS, B is a balanced, closed, convex subset of X. Then, for any $x_0 \notin B$, there exists $\ell \in X^*$ such that*

$$|\ell x| \leq 1 < \ell x_0 , \quad \textit{for all } x \in B .$$

Proof. By Theorem 3.26, there is nonzero $\ell_1 \in X^*$ such that

$$\operatorname{Re} \ell_1 x < c < \operatorname{Re} \ell_1 x_0 , \quad \text{for all } x \in B .$$

Since B is balanced, so is $\ell_1(B)$. Hence

$$|\ell_1 x| < c < |\ell_1 x_0| , \quad \text{for all } x \in B .$$

Let $\ell_1 x_0 = |\ell_1 x_0| e^{i\theta}$, then $\ell = c^{-1} e^{-i\theta} \ell_1$ has the desired properties. □

3.6.2 Geometric interpretations

The geometric consequences of the Hahn-Banach Theorem are achieved by interpreting that theorem in light of the correspondence between linear functionals and hyperplanes and between sublinear functionals and open convex neighborhoods of the origin.

As bofore, there is a great advantage inherent in a geometric discussion of real TVS's. Namely, if $\ell : X \to \mathbb{R}$ is a nonzero continuous $\mathbb{R}$-linear functional, then the closed hyperplane $\ker\ell$ disconnects the space. Indeed, $X\backslash\ker\ell$ has two connected components: $\{x : \ell(x) > 0\}$ and $\{x : \ell(x) < 0\}$. But If X is a complex TVS and $\ell : X \to \mathbb{C}$ is a nonzero continuous linear function, then $X\backslash \ker\ell$ is connected.

Theorem 3.25 and Theorem 3.26 can be rewrited as

(a) X is a real TVS and A and B are disjoint convex sets with A open, then A and B are separated.

(b) X is a real LCS and A and B are two disjoint closed convex subsets. If A is compact, then A and B are strictly separated.

Proposition 3.30. *If X is a real LCS, A is a subset of X.*

(a) $\overline{\text{co}}(A)$ *is the intersection of the closed half-spaces containing A.*

(b) $\overline{\text{span}}(A)$ *is the intersection of the closed hyperplanes containing A.*

3.6.3 An extension theorem

Theorem 3.31. *Let X be a LCS. ℓ is a continuous linear functional on a subspace Y, then there ℓ can be extended on X as a continuous linear functional.*

Proof. Assume, without loss of generality, that ℓ is not identically 0. Put

$$N = \{y \in Y : \ell(y) = 0\}$$

and pick $y_\text{o} \in Y$ such that $\ell(y_\text{o}) = 1$. Since ℓ is continuous on Y, $y_\text{o} \notin N = \text{cl}_Y(N) = \text{cl}_X(N) \cap Y$. So $y_\text{o} \notin \text{cl}_X(N)$. Then there exists a $\Lambda \in X^*$ such that $\Lambda y_\text{o} = 1$ and $\Lambda = 0$ on $\text{cl}_X(N)$. For any $y \in Y$, note that $y - \ell(y)y_\text{o} \in N$, since $\ell(y_\text{o}) = 1$. Hence

$$\Lambda y - \ell(y) = \Lambda y - \ell(y)\Lambda y_\text{o} = \Lambda\left(y - \ell(y)y_\text{o}\right) = 0\,.$$

Thus $\Lambda = \ell$ on Y. $\square$

3.7 Weak topologies on LCS

Let X be a LCS. Denote by X^* the space of continuous linear functionals on X. Obviously, X^* has a natural vector-space structure. It is convenient and, more importantly, helpful to introduce the notation, because of a certain symmetry,

$$\langle x, x^* \rangle$$

to stand for $x^*(x)$, for x in X and x^* in X^*.

Definition 3.2. (X, τ) is a LCS, the *weak topology* on X, denoted by τ_w or $\sigma(X, X^*)$, is the topology defined by the separating family of seminorms $\{p_{x^*} : x^* \in X^*\}$, where

$$p_{x^*}(x) = |\langle x, x^* \rangle| \text{ , for all } x \in X \,. \tag{3.7}$$

On the other hand, The *weak* topology* on X^*, denoted by τ_{w^*} or $\sigma(X^*, X)$, is the topology defined by the separating family of seminorms $\{p_x : x \in X\}$, where

$$p_x(x^*) = |\langle x, x^* \rangle| \text{ , for all } x^* \in X^* \,. \tag{3.8}$$

In fact, the weak topology on X is the weakest one with respect to which each element in X^* is continuous. Also, regarding each $x \in X$ as a linear functional on X^*, denoted as Jx and given by

$$Jx : x^* \mapsto \langle x, x^* \rangle \text{ , for all } x^* \in X^* \,, \tag{3.9}$$

the weak* topology on X^* is the weakest one with respect to which $J(X)$ is continuous.

Obviously, (X, τ_w) and (X^*, τ_{w^*}) both are LCS, by Example 3.1, and

(a) (X, τ_w) has a local base, namely,

$$\left\{ V(\Phi, \epsilon) := \{x : |\langle x, x^* \rangle| < \epsilon, x^* \in \Phi\} : \Phi \subset X^* \text{ is finite, } \epsilon > 0 \right\}, \tag{3.10}$$

and (X^*, τ_{w^*}) has a local base, namely,

$$\left\{ V(\Phi, \epsilon) \coloneqq \{x^* : |\langle x, x^* \rangle| < \epsilon, x \in \Phi\} : \Phi \subset X \text{ is finite}, \epsilon > 0 \right\}. \tag{3.11}$$

(b) A net $\{x_i\}$ in X is convergent to some $x \in X$ in τ_w if and only if

$$\langle x_i, x^* \rangle \to \langle x, x^* \rangle \ , \text{ for all } x^* \in X^* . \tag{3.12}$$

A net $\{x_i^*\}$ in X^* is convergent to some $x^* \in X$ in τ_{w^*} if and only if

$$\langle x, x_i^* \rangle \to \langle x, x^* \rangle \ , \text{ for all } x \in X \tag{3.13}$$

(c) Let $(Y, \mathcal{T})$ be a topological space. Then $f : (Y, \mathcal{T}) \to (X, \tau_w)$ is continuous if and only if

$$y \mapsto \langle f(y), x^* \rangle; Y \to \mathbb{R}$$

is continuous for every $x^* \in x^*$. $g : (Y, \mathcal{T}) \to (X^*, \tau_{w^*})$ is continuous if and only if

$$y \mapsto \langle x, f(y) \rangle; Y \to \mathbb{R}$$

is continuous for every $x \in X$.

(d) $M \subset X$ is bounded if and only if

$$\sup_{x \in M} |\langle x, x^* \rangle| < \infty \, , \text{ for all } x^* \in X^* . \tag{3.14}$$

$N \subset X^*$ is bounded if and only if

$$\sup_{x^* \in N} |\langle x, x^* \rangle| < \infty \, , \text{ for all } x \in X . \tag{3.15}$$

3.7.1 Duality

Theorem 3.32. *Let (X, τ) be a LCS, then $(X, \tau_w)^* = X^*$.*

Proof. $X^* \subset (X, \tau_w)^*$, since the weak topology on X is the weakest one with respect to which X^* is continuous. On the other hand, since $\tau_w \subset \tau$, thus $(X, \tau_w)^* \subset X^*$. $\square$

Theorem 3.33. *Let (X, τ) be a LCS, then $(X^*, \tau_{w^*})^* = J(X)$.*

Proof. $J(X) \subset (X^*, \tau_{w^*})^*$, since the weak* topology on X^* is the weakest one with respect to which $J(X)$ is continuous.

On the other hand, for any $f \in (X^*, \tau_{w^*})^*$, there exists $x_1, \dots, x_n$ in X and positive scalars $\alpha_1, \dots, \alpha_n$, by Theorem 3.15, such that

$$|f(x^*)| \leq \sum_{k=1}^{n} \alpha_k |\langle x_k, x^* \rangle|, \quad \text{for all } x^* \in X^*.$$

Thus

$$\ker f \subset \bigcap_{k=1}^{n} \ker(Jx_k).$$

By Lemma 0.1, $f \in \text{span}\{Jx_1, \cdots, Jx_n\}$, of course $f \in J(X)$. $\square$

Remark 3.6. Since $J(X) = (X^*, \tau_{w^*})^*$, we can define a weak* topology on X by regarding $J(X)$ as X (Indeed, if $\mathcal{T}$ is the weak* topology on $J(X)$, then $J^{-1}(\mathcal{T})$ is the topology on X induced by J). Then one can show that this topology is exactly τ_w !

3.7.2 Weak closure

Theorem 3.34. *Let (X, τ) be a LCS. Let K be a nonempty convex subset of X. Then the weak closure of K is equal to its original closure, i.e.,*

$$\overline{K} = \overline{K}^{\tau_w}. \tag{3.16}$$

Proof. $\overline{K}^{\tau_w}$ is weakly closed, hence originally closed, so that

$$\overline{K} \subset \overline{K}^{\tau_w}.$$

To obtain the opposite inclusion, if there exists $x_0 \in X$, and

$$x_0 \in \overline{K}^{\tau_w} \backslash \overline{K}$$

By Hahn-Banach separation theorem, there exist $x^* \in X^*$, $\gamma \in \mathbb{R}$ and $\epsilon > 0$ such that, for every $x \in K$,

$$\mathrm{Re}\langle x_0, x^* \rangle < \gamma < \gamma + \epsilon < \mathrm{Re}\langle x, x^* \rangle \, .$$

Thus $|\langle x - x_0, x^* \rangle| \geq \epsilon$ for all $x \in K$, i.e.,

$$K \cap V(x_0; x^*, \epsilon) = \emptyset \, .$$

Thus $x_0 \notin \overline{K}^{\tau_w}$, this is a contradiction. □

Corollary 3.35. *Let (X, τ) be a LCS. Let K be a nonempty convex subset of X. Then K is closed if and only if K is weakly closed.*

Corollary 3.36. *Let (X, τ) be a metrizable locally convex space. If $\{x_n\}$ is a sequence in X that converges weakly to some $x \in X$, then there is a sequence $\{y_m\}$ in X such that*

(a) each y_m is a convex combination of finitely many x_n, and

(b) $y_m \to x$ originally.

Proof. Let H be the convex hull of $\{x_n\}$, let K be the weak closure of H. Clearly, $x \in K$. By Theorem 3.34, x is also in the original closure of H. Since the original topology of X is assumed to be metrizable, it follows that there is a sequence $\{y_m\}$ in H that converges originally to x. □

To get a feeling for what is involved in Corollary 3.36, consider the following example.

Example 3.3. Let X be a compact Hausdorff space, for example, the unit interval on the real line. Assume that f and f_n are continuous scalar (real or complex) functions on X such that

(a) $f_n(x) \to f(x)$ for every $x \in X$;

(b) $|f_n(x)| \leq 1$ for all n and x.

Theorem 3.36 asserts that there are convex combinations of the f_n that converge uniformly to f. To see this, let $C(X)$ be the Banach space of all continuous scalar functions on X, normed by the supremum. Then norm convergence is the same as uniform convergence on X. If μ is any scalar Borel measure on X, with finite total variation, Lebesgue's dominated convergence theorem implies that

$$\int f_n d\mu \to \int f d\mu.$$

Hence $f_n \to f$ weakly, by the Riesz representation theorem which identifies $C(X)^*$ with the space of all regular scalar Borel measures on X.

3.7.3 Annihilators

Annihilator is in an analogue of the orthogonal complement in Hilbert space.

Definition 3.3. Let (X, τ) be a LCS. Let X^* be the dual space of X.

- For a subspace of X, namely M, the *annihilator of* M is defined by

$$M^{\perp} = \{x^* \in X^* : \langle x, x^* \rangle = 0 \text{ for all } x \in M\} . \tag{3.17}$$

- For a subspace of X^*, namely N, the *annihilator of* N is defined by

$$^{\perp}N = \{x \in X : \langle x, x^* \rangle = 0 \text{ for all } x^* \in N\} . \tag{3.18}$$

In other words, $M^{\perp}$ consists of all continuous linear functionals on X vanishing on M, and $^{\perp}N$ is the subset of X on which every member of N vanishes.

It is clear that $M^\perp$ is a weak-closed subspace of X, and ${}^\perp N$ is a weak*-closed subspace of X^*. The following theorem describes the duality between these two types of annihilators.

Theorem 3.37. *Under the preceding hypotheses,*

$$ {}^\perp\left(M^\perp\right) = \overline{M}^{\tau_w} = \overline{M}\,, \ \left({}^\perp N\right)^\perp = \overline{N}^{\tau_{w^*}}\,. \tag{3.19}$$

Proof. Firstly, ${}^\perp\left(M^\perp\right)$ is weak-closed, so

$$\overline{M}^{\tau_w} \subset {}^\perp\left(M^\perp\right)\,.$$

If $x \notin \overline{M}^{\tau_w}$, by Theorem 3.28, there exists nonzero continuous linear functional $x^* \in X^*$ vanishing on M and $\langle x, x^*\rangle = 1$. Thus

$$x^* \in M^\perp \ \text{ and, } \ x \notin {}^\perp(M^\perp)\,.$$

So the equality is established.

Similarly, $\left({}^\perp N\right)^\perp$ is weak*-closed, so

$$\overline{N}^{\tau_{w^*}} \subset \left({}^\perp N\right)^\perp\,.$$

If $x^* \notin \overline{N}^{\tau_{w^*}}$, by Theorem 3.28 implies the existence of an $x \in {}^\perp N$ such that $\langle x, x^*\rangle = 1$. Thus $x^* \left({}^+N\right)^2$, Thus

$$x \in^\perp N \ \text{ and, } \ x^* \notin \left({}^\perp N\right)^\perp\,.$$

Now we get the desired equality. $\square$

3.8 The Krein-Milman Theorem

3.8.1 Compact Convex Sets

Let X be a vector space and $E \subset X$, the *convex hull* of E will be denoted by $\mathrm{co}(E)$. Recall that $\mathrm{co}(E)$ is the intersection of all convex subsets of X which contain E. Equivalently, $\mathrm{co}(E)$ is the set of all finite convex combinations of members of E. If X is a topological vector space and $E \subset X$, the *closed convex hull* of E, written $\overline{\mathrm{co}}(E)$, is the closure of $\mathrm{co}(E)$.

We now turn to the question: What can one say about the convex hull $\mathrm{co}(K)$ of a compact set K? Even in a Hilbert space, $\mathrm{co}(K)$ need not be closed, and there are situations in which $\overline{\mathrm{co}}(K)$, is not compact. In Fréchet spaces the latter pathology does not occur (Theorem 3.39). The proof of this will depend on the fact that a subset of a complete metric space is compact if and only if it is closed and totally bounded.

Lemma 3.38. *If $A_1, \ldots, A_n$ are compact convex sets in a topological vector space X, then $\mathrm{co}\,(A_1 \cup \cdots \cup A_n)$ is compact.*

Proof. Let S be the simplex in $\mathbb{R}^n$ consisting of all $s = (s_1, \ldots, s_n)$ with $s_i \geq 0$, $s_1 + \cdots + s_n = 1$. Put $A = A_1 \times \cdots \times A_n$. Define $f : S \times A \to X$ by

$$f(s, a) = s_1 a_1 + \cdots + s_n a_n$$

and put $K = f(S \times A)$. Clearly, $S \times A$ is compact and f is continuous, so K is compact with $K \subset \mathrm{co}\,(A_1 \cup \cdots \cup A_n)$. We will see that this inclusion is actually an equality because each A_i is convex. To this end, we show that K is convex, then since $A_i \subset K$, we have $K \supset \mathrm{co}\,(A_1 \cup \cdots \cup A_n)$.

If (s, a) and (t, b) are in $S \times A$ and if $\alpha \geq 0, \beta \geq 0, \alpha + \beta = 1$ then

$$\alpha f(s, a) + \beta f(t, b) = f(u, c)$$

where $u = \alpha s + \beta t \in S$ and $c \in A$, because

$$c_i = \frac{\alpha s_i a_i + \beta t_i b_i}{\alpha s_i + \beta t_i} \in A_i \quad (1 \leq i \leq n)\,.$$

We complete the proof. $\square$

Theorem 3.39. *Let X be a Frechét space and $K \subset X$ is compact, then $\overline{\mathrm{co}}(K)$ is compact.*

Proof. It suffices to show that $\mathrm{co}(K)$ is totally bounded. For any given $\epsilon > 0$, there exists $\{k_1, \cdots, k_n\} \subset K$ so that

$$K \subset \{k_1, \cdots, k_n\} + B(0, \epsilon) .$$

Since X is locally convex, we suppose that $B(0, \epsilon)$ is convex. Thus

$$\mathrm{co}(K) \subset \mathrm{co}\{k_1, \cdots, k_n\} + B(0, \epsilon) .$$

By Lemma 3.38, $\mathrm{co}\{k_1, \cdots, k_n\}$ is compact, so there exists $\{k_1', \cdots, k_m'\} \subset \mathrm{co}(K)$ with

$$\mathrm{co}\{k_1, \cdots, k_n\} \subset \{k_1', \cdots, k_m'\} + B(0, \epsilon) .$$

Therefore,

$$\mathrm{co}(K) \subset \{k_1', \cdots, k_m'\} + B(0, 2\epsilon) ,$$

and the desired result follows. □

The following theorem is a deep result and we will not give the proof. It can be found in Chapter V 13.4 in *A Course in Functional Analysis* by J.Conway.

Theorem 3.40 (Krein)**.** *The closed convex hull of a weakly compact set K in a Banach space X is again weakly compact.*

We give an example that there exists a compact set in a Hilbert space whose convex hull is not closed.

Example 3.4. Consider

$$u_n = (\underbrace{0, \ldots, 0}_{n-1}, 1/n, 0, \ldots) \text{ for } n \geq 1 ,$$

and $K = \{u_n\} \cup \{0\}$ a compact subset of $\ell^2(\mathbb{N})$ since $u_n \to 0$. The convex hull of K is given by

$$\mathrm{co}(K) = \left\{ \sum_{n=1}^{k} a_n u_n : a_n \geq 0 , \sum_{n=1}^{k} a_n \leq 1 \right\} .$$

So also $\sum_{n=1}^{k} 2^{-n} u_n$ lies in $\mathrm{co}(K)$ for each k. But this sequence converges to $\sum_{n=1}^{\infty} 2^{-n} u_n$ which does not lie in it. Thus $\mathrm{co}(K)$ is not closed.

Proposition 3.41. *If $E \subset \mathbb{R}^n$ and $x \in \mathrm{co}(E)$, then x lies in the convex hull of some subset of E which contains at most $n+1$ points (depending on x).*

Proof. It is enough to show that if $k > n$ and $x = \sum_{i=1}^{k+1} t_i x_i$ is convex combination of some $k+1$ vectors $x_i \in \mathbb{R}^n$, then x is actually a convex combination of some k of these vectors. Assume, with no loss of generality, that $t_i > 0$ for $1 \le i \le k+1$ The null space of the linear map (8)

$$(a_1, \ldots, a_{k+1}) \to \left(\sum_{1}^{k+1} a_i x_i, \sum_{1}^{k+1} a_i \right)$$

which sends R^{k+1} into $R^n \times R$, has positive dimension, since $k > n$ Hence there exists $(a_1, \ldots, a_{k+1})$, with some $a_i \neq 0$, so that $\sum a_i x_i = 0$ and $\sum a_i = 0$. since $t_i > 0$ for all i, there is a constant λ such that $|\lambda a_i| \le t_i$ for all i and $\lambda a_j = t_j$ for at least one j. Setting $c_i = t_i - \lambda a_i$ we conclude that $x = \sum c_i x_i$ and that at least one c_j is 0; note also that $\sum c_i = \sum t_i = 1$ and that $c_i \ge 0$ for all i. $\square$

Corollary 3.42. *If K is a compact set in $\mathbb{R}^n$, then $\mathrm{co}(K)$ is compact.*

Proof. Let S be the simplex in $\mathbb{R}^{n+1}$ consisting of all $t = (t_1, \ldots, t_{n+1})$ with $t_i \ge 0$ and $\sum t_i = 1$. Let K be compact, $K \subset \mathbb{R}^n$. By the proposition that follows, $x \in \mathrm{co}(K)$ if and only if

$$x = t_1 x_1 + \cdots + t_{n+1} x_{n+1}$$

for some $t \in S$ and $x_i \in K$. In other words, $\mathrm{co}(K)$ is the image of $S \times K^{n+1}$ under the continuous mapping

$$(t, x_1, \ldots, x_{n+1}) \to t_1 x_1 + \cdots + t_{n+1} x_{n+1} \,.$$

Hence co(K) is compact. $\square$

3.8.2 The Krein-Milman Theorem

Let K be a convex subset of a linear space X. A nonempty set $S \subset K$ is called an *extreme set* of K if no point of S is an internal point of any line interval whose end points are in K, except when both end points are in S. Analytically, the condition can be expressed as follows: For any $z \in S$, if there exists $x, y \in K$ and $t \in [0, 1]$ with

$$z = (1 - t)x + ty\,,$$

then it must be $x, y \in S$. Clearly, if $\{S_i : i \in I\}$ is a family of extreme sets of K, then the intersection $\cap_{i \in I} S_i$ is also an extreme set of K. The *extreme points* of K are the extreme sets that consist of just one point. The set of all extreme points will be denoted by $E(K)$.

Example 3.5. K is the interval $0 \le x \le 1$; the two endpoints are extreme points.

Example 3.6. K is the closed disk

$$x^2 + y^2 \le 1\,.$$

Every point on the circle $x^2 + y^2 = 1$ is an extreme point. The open disk

$$x^2 + y^2 < 1$$

has no extreme points.

Example 3.7. K a polyhedron, including faces. Its extreme subsets are its faces, edges, vertices, and of course K itself.

Lemma 3.43. *Let K be a convex set, E an extreme subset of K, and F an extreme subset of E. Then F is an extreme subset of K.*

Proof. If $x, y \in K$, $t \in (0,1)$ so that $tx + (1-t)y \in F \subset E$, since E is an extreme subset of K, we have x, y in E. Since since F is an extreme subset of E, $tx + (1-t)y \in F$, we have x, y in F. Thus F is an extreme subset of K. □

Lemma 3.44. *Let X, Y be two linear spaces. Let $T : X \to Y$ be a REAL linear. Let K be a convex subset of Y, E an extreme subset of K. Then $T^{-1}E$ is either empty or an extreme subset of $T^{-1}K$.*

Proof. If $T^{-1}E$ is not empty, then so is $T^{-1}K$. As we all know, $T^{-1}K$ is convex. If $x_1 = T^{-1}y_1, x_2 = T^{-1}y_2$ in $T^{-1}E$, $t \in (0,1)$ satisfy that $tx_1 + (1-t)x_2 \in T^{-1}K$, then $ty_1 + (1-t)y_2 \in E$. Since E is an extreme subset of K, y_1, y_2 are in K and hence x_1, x_2 are in $T^{-1}K$. □

Taking Y to be one dimensional, we get

Corollary 3.45. *Denote by H a convex subset of a linear space X, ℓ a REAL linear map of X into $\mathbb{R}$, $H_{\min}$ and $H_{\max}$ the subsets of H, where ℓ achieves its minimum and maximum, respectively.*

Assertion: When nonempty, $H_{\min}$ and $H_{\max}$ are extreme subsets of H.

The following two theorems show that under certain conditions $E(K)$ is quite a large set.

Theorem 3.46 (The Krein-Milman Theorem)**.** *Let X be a locally convex space. Let K be a nonempty compact convex subset of X. Then*

(i) K has at least one extreme point, i.e., $E(K) \neq \emptyset$; and

(ii) K the closed convex hull of the set of its extreme points, i.e.,

$$K = \overline{\mathrm{co}}(E(K)) \,.$$

Proof of part (i). Consider the collection

$$\{E : E \text{ is a nonempty closed extreme subsets of } K\} \,.$$

This collection is nonempty, for it contains K itself. Partially order this collection by inclusion. We claim that every totally ordered subcollection $\{E_j\}$ has a lower bound. That lower bound is the intersection $\cap_j E_j$. To see this, we have to show that $\cap_j E_j$ is nonempty, closed, and extreme.

We claim that every finite subset of the totally ordered collection $\{E_j\}$ has a nonempty intersection. This is because in being totally ordered by inclusion, the intersection of a finite subset of the collection $\{E_j\}$ is the smallest member of that subset, and hence $\cap_j E_j$ is nonempty as a consequence of the compactness of K. Being the intersection of closed sets, $\cap_j E_j$ is closed. $\cap_j E_j$ is an extreme subset of K, since the nonempty intersection of extreme subsets of a convex set K is itself an extreme subset of K.

We conclude from Zorn's lemma that K has a closed extreme subset E that is minimal with respect to inclusion. We claim that such an E consists of a single point. To see this, suppose, on the contrary, that E contains two distinct points. According to Theorem 3.26, there exists a continuous linear functional ℓ that separates these two points. Since E is compact, and $\mathrm{Re}\ell$ continuous and not constant on E, $\mathrm{Re}\ell$ achieves its maximum on some *proper* subset M of E.

Since ℓ is continuous and E is closed, M is closed. By Corollary 3.45 Since M is a real linear functional $\mathrm{Re}\ell$ assumes its maximum on a convex set E is an extreme subset of E. It is easy to show further that if E is an extreme subset of K, and M an extreme subset of E, then M is an extreme subset of K. Since E is a minimal extreme subset of K, and M an extreme subset smaller than E, we have a contradiction, into which we got by assuming that E contains more than one point. We conclude therefore that a minimal E consists of a single point. This single point is an extreme point of K. This completes the proof of part (i), and gives a little more:

(i') Every closed, extreme subset of K contains an extreme point. $\square$

Proof of part (ii). To show that every point of K belongs to the closure of $\mathrm{co}(E(K))$ is the same as showing that a point z that does not belong to $\overline{\mathrm{co}}(E(K))$ does not belong to K. Clearly, the closure of $\mathrm{co}(E(K))$ is compact and convex. So, if z does not belong to the closure, then according to Theorem 3.26 there is a continuous linear functional ℓ such that

$$\mathrm{Re}\,\ell(y) < c < \mathrm{Re}\,\ell(z) \quad \text{for all } y \text{ in } \mathrm{co}(E(K))\,.$$

Since K is compact and ℓ continuous, $\mathrm{Re}\ell$ achieves its maximum over K on some closed subset M of K, and M is an extreme subset of K. According to part (i') noted above, M contains some extreme point p of K. Since p belongs to $E(K)$, and so to $\mathrm{co}(E(K))$, it follows that $\ell(p) < c$. since by construction $\mathrm{Re}\ell(p) = \max_{x\in K} \mathrm{Re}\ell(x)$, $\mathrm{Re}\ell(x) \le \mathrm{Re}\ell(p) < c$ for all x in K. Since by $\mathrm{Re}\ell(z) > c$, this proves that z does not belong to K. $\square$

3.8.3 Applications: The Stone-Weierstrass Theorem

The Weierstrass Approximation Theorem shows that the continuous real- valued fuctions on a compact interval can be uniformly approximated by polynomials. In other words, the polynomials are uniformly dense in $C([a,b],\mathbb{R})$ with respect to the sup-norm.

Theorem (Weierstrass Approximation Theorem I)**.** *Let $f \in C([a,b],\mathbb{R})$. Then there is a sequence of polynomials $p_n(x)$ that converges uniformly to $f(x)$ on $[a,b]$.*

Recall that real-valued functions f defined on a subset of $\mathbb{R}$ is called a *trignometric polynomial* if f is finite sums of the form

$$f(x) = a_0 + \sum_{n=1}^{N} \left[a_n \cos(nx) + b_n \sin(nx)\right] \,.$$

Theorem (Weierstrass Approximation Theorem II)**.** *The set of all trigonometric polynomials are uniformly dense in $C_{per}\,([0,2\pi],\mathbb{R})$.*

It will be seen that the Weierstrass Approximation Theorem is in fact a special case of the more general Stone-Weierstrass Theorem, proved by Stone in 1937, who realized that very few of the properties of the polynomials were essential to the theorem. Although this proof is not constructive and relies on more machinery than that of Bernstein, it is much more efficient and has the added power of generality.

Let X be a locally compact Hausdorff space, $C_0(X) = C_0(X, \mathbb{R})$ the set of all real-valued continuous functions on X vanishing at infinity, that is, a continuous function f is in $C_0(X)$ if, for every $\epsilon > 0$, there exists a compact set $K \subset X$ such that $|f| < \epsilon$ on K^c. Clearly, $C_0(X)$ is a Banach algebra with the supremum norm. Let $\mathcal{A}$ be a subalgebra of $C_0(X)$, that is, $\mathcal{A}$ is a linear subspace of $C_0(X)$ and the product of two functions in $\mathcal{A}$ belongs to $\mathcal{A}$. The subalgebra $\mathcal{A}$ is said wo

- *separates points* of X, if given any pair of points p and q with $p \neq q$, there is a function f in $\mathcal{A}$ such that $f(p) \neq f(q)$;
- *vanish nowhere*, if not all of the elements of $\mathcal{A}$ simultaneously vanish at a point; i.e., for every p in X, there is some f in $\mathcal{A}$ such that $f(p) \neq 0$.

The theorem generalizes as follows.

Theorem 3.47 (Stone-Weierstrass Theorem)**.** *Let X be a locally compact Hausdorff space. If $\mathcal{A}$ is a subalgebra of $C_0(X)$ which separates points of $C_0(X)$ and vanishes nowhere, then $\mathcal{A}$ is dense in $C_0(X)$.*

We present Louis de Branges's elegant proof, based on the Krein-Milman theorem, of Stone's generalization of the Weierstrass theorem.

Proof. According to the spanning criterion, $\mathcal{A}$ is dense in $C_0(X)$ if the only bounded linear functional ℓ on $C_0(X)$ that is zero on $\mathcal{A}$ is the zero functional. According to the Riesz-Kakutani representation theorem (see Example 5.6), the bounded linear functionals on $C_0(X)$ are of the form

$$\ell(f) = \int_X f d\nu \,,$$

ν a signed Radon measure of finite total variation $\|\nu\| = \int d|\nu|$. So what we have to show is that, if $\int_X f d\nu = 0$ for all $f \in \mathcal{A}$, then $\nu = 0$.

Suppose not; denote by U the set of signed Radon measures of finite total mass is ≤ 1 that annihilate all functions in $\mathcal{A}$. This is a convex set, and according to Theorem 5.15, compact in the weak-star topology. So according to the Krein-Milman theorem, if U contained a nonzero measure, it would contain a nonzero extreme point; call it μ. Since μ is extreme, it's easy to see that $\|\mu\| = 1$.

Lemma. *Let g be a function in $\mathcal{A}$ whose values lie between* 0 *and* 1*, i.e.,* $0 < g(p) < 1$ *for all $p \in X$. Then there exists a constant C so that $g = C$ μ-a.e..*

Proof of the Lemma. Since $\mathcal{A}$ is an algebra, if f belong to $\mathcal{A}$, so does gf. Since μ annihilates every function in $\mathcal{A}$, $\int (fg) d\mu = 0$. It follows that the measure $g d\mu$ also annihilates every function in $\mathcal{A}$. Let g be a function in $\mathcal{A}$ whose values lie between 0 and 1, i.e., $0 < g(p) < 1$ for all p in X. Denote

$$a = \|g.\mu\| = \int g d|\mu| \;,\; b = \|(1-g).\mu\| = \int (1-g) d|\mu| \,.$$

Clearly a and b are positive. Add them, $a + b = \int |d\mu| = 1$. The identity

$$\mu = a\,\frac{g.\mu}{a} + b\,\frac{(1-g).\mu}{b}$$

represents μ as a nontrivial convex combination of $g.\mu/a$ and $(1-g).\mu/b$, both points in U. Since μ is an extreme point, μ must be equal to $g.\mu/a$. It follows from the uniqueness of R-N derivative that $g = a$ μ-a.e.. □

Define the support of the measure μ to be the set of points p that have the property that $|\mu|(N) > 0$ for any open set N containing p.

Claim: The support of μ consists of a single point.

To see this, suppose that both p and q $(p \neq q)$ belong to the support μ. Since the functions in $\mathcal{A}$ separate points of X, there is a function h in $\mathcal{A}$,

$h(p) \neq h(q)$. Adding a large enough constant to h and dividing it by another large constant, we obtain a function g whose values lie between 0 and 1, and $g(p) \neq g(q)$. Since g is continuous, this contradicts our previous lemma.

A Radon measure $|\mu|$ whose support consists of a single point p, and $|\mu|(X) = 1$, is a unit point mass at p. This is because of the inner regularity on open sets of a Radon measure:

$$|\mu|(G) = \sup\{|\mu|(K) : K \subset G : K \text{is compact}\}\,, \text{ for each } G \text{ is open }\ .$$

In fact, take any open set G with $p \in G$ and any compact subset K of G. Since the support of μ is $\{p\}$, every $x \in K$ has a openneighboorhood N_x with $|\mu|(N_x) = 0$. Since K is compact and $\{N_x : x \in K\}$ is a open cover of K, $K \subset \cup_j N_{x_j}$ for finitely many $\{x_j\}$. Then we deduce that $|\mu|(K) = 0$. Since K is arbitary, $|\mu|(G) = 0$. Thus $|\mu| = \delta_p$ on the faimly of open sets which is a π system generating the Boel algebra. Thus $|\mu| = \delta_p$, $\mu = \delta_p$ or $\mu = -\delta_p$. Therefore

$$\int f d\mu = f(p) \text{ or } -f(p) = 0\,, \text{ for all } f \in \mathcal{A}\,.$$

Since, by hypothesis $\mathcal{A}$ vanishes nowhere, this a contradiction. We now complete the proof. $\square$

We can use the Stone-Weierstrass theorem to prove the complex Stone-Weierstrass theorem. The complex version, however, requires additional assumptions.

Theorem 3.48 (Complex Stone-Weierstrass Theorem)**.** *Let $\mathcal{A}$ be a (complex) sub-algebra of $C_0(X, \mathbb{C})$ which is closed under complex conjugation, i.e., if $f \in \mathcal{A}$, then $\bar{f} \in \mathcal{A}$,. If $\mathcal{A}$ separates points of X and vanishes nowhere, then $\mathcal{A}$ is dense in $C_0(X, \mathbb{C})$.*

Proof. Let $f \in \mathcal{A}$. Then, $\mathrm{Re}(f) = \frac{1}{2}(f + \bar{f}) \in \mathcal{A}$ and $\mathrm{Im}(f) = \frac{1}{2i}(f - \bar{f}) \in \mathcal{A}$. Denote by $E_{\mathbb{R}}$ the sub-algebra of $\mathcal{A}$ containing all real-valued functions.

Note that $E_{\mathbb{R}}$ separates points in X and vanishes nowhere because $\mathcal{A}$ does. By the Stone-Weierstrass Theorem for real valued functions, $E_{\mathbb{R}}$ is dense in $C_0(X, \mathbb{R})$ since $\mathcal{A} = \{f + ig \mid f, g \in E_{\mathbb{R}}\}$, we see that $\mathcal{A}$ is indeed dense in $C_0(X, \mathbb{C}) = C_0(X, \mathbb{R}) + iC_0(X, \mathbb{R})$. $\square$

3.9 Vector-valued Integration

Sometimes it is desirable to be able to integrate functions f that are defined on some measure space $(\Omega, \mathcal{F}, \mu)$ and whose values lie in some topological vector space X. The first problem is to associate with these data a vector in X that deserves to be called

$$\int_\Omega f d\mu$$

i.e., which has at least some of the properties that integrals usually have. For instance, the equation

$$\Lambda\left(\int_\Omega f d\mu\right) = \int_\Omega (\Lambda f) d\mu$$

ought to hold for every $\Lambda \in X^*$, because it does hold for sums, and because integrals are (or ought to be) limits of sums in some sense or other. In fact, our definition will be based on this single requirement.

Many other approaches to vector-valued integration have been studied in great detail; in some of these, the integrals are defined more directly as limits of sums (see Exercise 3.2).

Definition 3.4. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, let X be a topological vector space on which X^* separates points. Let f be a function from Ω into X such that the scalar functions Λf are integrable with respect to μ, for every $\Lambda \in X^*$. If there exists a vector $y \in X$ such that

$$\Lambda y = \int_\Omega (\Lambda f) d\mu \quad \text{for every } \Lambda \in X^*,$$

then we define

$$\int_\Omega f d\mu := y.$$

Remark 3.7. It is clear that there is at most one such y, because X^* separates points on X. Thus there is no uniqueness problem.

Existence will be proved only in the rather special case (sufficient for many applications) in which Ω is compact, f is continuous and X is a Fréchet space. In that case, $f(\Omega)$ is compact, and the closed convex hull of $f(\Omega)$ should be compact by Theorem 3.39, which is why we assume X is a Fréchet space.

Recall that a Borel measure on a compact Hausdorff space Ω is a measure defined on the σ-algebra of all Borel sets in Ω this is the smallest σ-algebra that contains all open subsets of Ω. A probability measure is a positive measure of total mass 1.

Theorem 3.49 (Existence of Vector-Valued Integrations)**.** *Let X be a Fréchet space. Let μ be a Borel probability measure on a compact Hausdorff space Ω. If $f : \Omega \to X$ is continuous, then the integral*

$$y = \int_\Omega f d\mu$$

exists in the sense of Definition 3.4. Moreover, $y \in \overline{co}(f(\Omega))$.

Remark 3.8. If v is any positive Radon measure on Ω, then some scalar multiple of v is a probability measure. The theorem therefore holds (except for its last sentence) with v in place of μ. It can then be extended to real-valued Radon measures (by the Jordan decomposition theorem) and (if the scalar field of X is $\mathbb{C}$) to complex ones.

Proof. We have to prove that there exists $y \in \overline{\text{co}}(f(\Omega))$ such that

$$\Lambda y = \int_\Omega (\Lambda f) d\mu \quad \text{for every } \Lambda \in X^* . \tag{3.20}$$

Let $L = \{\Lambda_1, \ldots, \Lambda_n\}$ be a finite subset of X^*. Let E_L be the set of all $y \in \overline{\text{co}}(f(\Omega))$ that satisfy (3.20) for every $\Lambda \in L$. Each E_L is closed (by the continuity of Λ) and is therefore compact, since $\overline{\mathrm{co}}(f(\Omega))$ is compact by Theorem 3.39. If no E_L is empty, the collection of all E_L has the finite intersection property. The intersection of all E_L is therefore not empty, and any y in it satisfies (3.20) for every $\Lambda \in X^*$. It is therefore enough to prove $E_L \neq \emptyset$. In other words, there exists $y \in \overline{\mathrm{co}}(f(\Omega))$ with

$$\Lambda_i y = \int_\Omega \Lambda_i f d\mu \quad (1 \leq i \leq n)\,.$$

Without loss of generality, regard X as a real vector space. Regard $L = (\Lambda_1, \ldots, \Lambda_n)$ as a (continuous) mapping from X into $\mathbb{R}^n$. Define

$$m_i = \int_\Omega (\Lambda_i f)\, d\mu \quad (1 \leq i \leq n)\,. \tag{3.21}$$

It suffices to show that $m = (m_1, \ldots, m_n) \in L(\overline{\mathrm{co}}(f(\Omega)))$. In fact,

$$m \in L(\mathrm{co}(f(\Omega))) = \mathrm{co}(L(f(\Omega)))\,.$$

Put $K = L(f(\Omega))$, then K is a compact subset of $\mathbb{R}^n$. If $m \notin \mathrm{co}(K)$, by Theorem 3.42, Hahn-Banach theorem on $\mathbb{R}^n$ and the known form of the linear functionals on $\mathbb{R}^n$, there are real numbers $c_1, \ldots, c_n$ such that

$$\sum_{i=1}^n c_i u_i < \sum_{i=1}^n c_i m_i$$

if $u = (u_1, \ldots, u_n) \in K$. Hence

$$\sum_{i=1}^n c_i \Lambda_i f(\omega) < \sum_{i=1}^n c_i m_i \quad \omega \in \Omega\,.$$

Since μ is a probability measure, integration of the left side gives $\sum c_i m_i < \sum c_i m_i$, that is a contradiction! This completes the proof. $\square$

Remark 3.9. From the proof above, one can see that the theorem still holds provided that X is a locally convex space, f is continuous and the closed convex hull of $f(\Omega)$, $\overline{\mathrm{co}}(f(\Omega))$ is compact. For example, let X be a Banach space equipped with the weak topology τ_w. If $f : \Omega \to (X, \tau_w)$ is continuous, then $\overline{\text{co}}(f(\Omega))$ is compact in (X, τ_w) by Theorem 3.40, and hence the integral $\int_\Omega f d\mu$ exists: for each $\Lambda \in (X, \tau_w)^* = X^*$

$$\Lambda\Big(\int_\Omega f d\mu\Big) = \int_\Omega (\Lambda f) d\mu\,.$$

We will use this fact when dealing with the weak continuous semigroup in Theorem 11.2.

The following proposition is intuitive and easy.

Proposition 3.50. *Suppose Ω is a compact Hausdorff space, X is a Banach space, $f : \Omega \to X$ is continuous, and μ is a positive Borel measure on Ω. Then*

$$\left\|\int_\Omega f d\mu\right\| \leq \int_\Omega \|f\| d\mu$$

Proof. Let $y = \int_\Omega f d\mu$, then by Corollary 1.32,

$$\begin{aligned}\|y\| &= \sup_{\Lambda \in X^*, \|\Lambda\| \leq 1} |\Lambda y| = \sup_{\Lambda \in X^*, \|\Lambda\| \leq 1} \left|\int_\Omega \Lambda f d\mu\right|\,. \\ &\leq \sup_{\Lambda \in X^*, \|\Lambda\| \leq 1} \int_\Omega |\Lambda f| d\mu \leq \sup_{\Lambda \in X^*, \|\Lambda\| \leq 1} \|\Lambda\| \int_\Omega \|f\| d\mu = \int_\Omega \|f\| d\mu\,.\end{aligned}$$

We complete the proof. □

Exercise 3.1. Suppose $(\Omega, \mathcal{F}, \mu)$ is a measure space and X, Y are two Banach spaces.

(a) Let f maps Ω into X and the integral $\int_\Omega f\, \mathrm{d}\mu \in X$ exists. Then for each $A \in \mathcal{B}(X, Y)$, the integral $\int_\Omega Af\, \mathrm{d}\mu \in Y$ exists and

$$\int_\Omega Af\, d\mu = A\left(\int_\Omega f\, d\mu\right)\,.$$

(b) Let F maps Ω into $\mathcal{B}(X)$ and the integral $\int_\Omega F\, \mathrm{d}\mu \in \mathcal{B}(X)$ exists. Then for each $A \in \mathcal{B}(X)$, the integral $\int_\Omega AF\, \mathrm{d}\mu \in \mathcal{B}(X)$, $\int_\Omega FA\, \mathrm{d}\mu \in \mathcal{B}(X)$ exists and

$$\int_\Omega AF\, d\mu = A\left(\int_\Omega F\, d\mu\right)\,;\int_\Omega FA\, d\mu = \left(\int_\Omega F\, d\mu\right) A\,.$$

Exercise 3.2. Suppose μ is a Borel probability measure on a compact Hausdorff space Ω, X is a Fréchet space, and $f : \Omega \to X$ is continuous. A partition of Ω is, by definition, a finite collection of disjoint Borel subsets of Ω whose union is Ω. Prove that to every neighborhood V of 0 in X there corresponds a partition $\{E_i\}$ such that the difference

$$z = \int_\Omega f d\mu - \sum_i \mu(E_i) f(s_i)$$

lies in V for every choice of $s_i \in E_i$. (This exhibits the integral as a strong limit of "Riemann sums".)

Suggestion: Take V convex and balanced. If $\Lambda \in X^*$ and if $|\Lambda x| \leq 1$ for every $x \in V$, then $|\Lambda z| \leq 1$, provided that the sets E_i are chosen so that $f(s) - f(t) \in V$ whenever s and t lie in the same E_i.

With this definition in place all the main results about the one-dimensional Riemann integral in first year analysis carry over to vector valued integrals.

Exercise 3.3 (Properties of the Integral). Let X be a Banach space, fix two real numbers $a < b$, and let $x, y : [a, b] \to X$ be continuous functions. Then the following hold.

(a) The integral is a linear operator $C([a, b], X) \to X$. In particular

$$\int_a^b (x(t) + y(t))dt = \int_a^b x(t)dt + \int_a^b y(t)dt\,.$$

(b) If $a < c < b$, then

$$\int_a^b x(t)dt = \int_a^c x(t)dt + \int_c^b x(t)dt\,.$$

(c) If Y is another Banach space and $A : X \to Y$ is a continuous linear operator, then

$$\int_a^b Ax(t)dt = A\int_a^b x(t)dt\,.$$

(d) Assume $x : [a,b] \to X$ is continuously differentiable, i.e. the limit

$$x'(t) := \lim_{\substack{h \to 0 \\ t+h \in [a,b]}} \frac{x(t+h) - x(t)}{h}.$$

exists for all $t \in [a,b]$ and the derivative $x' : [a,b] \to X$ is continuous. Then

$$\int_a^b x'(t)dt = x(b) - x(a).$$

(e) If $\alpha < \beta$ and $\phi : [\alpha, \beta] \to [a,b]$ is a diffeomorphism, then

$$\int_a^b x(t)dt = \int_\alpha^\beta x(\phi(s))\phi'(s)ds$$

(f) Let $x_0 \in X$ and assume

$$x(t) = x_0 + \int_a^t y(s)ds \quad \text{for } a \le t \le b$$

Then x is continuously differentiable and $x'(t) = y(t)$ for all $t \in [a,b]$

The following further extension of Caratheodory's theorem holds on locally convex spaces:

Theorem 3.51. *Let X be a Fréchet space. Let K be a compact subset of X. Let y be a vector in X. Then $y \in \overline{\mathrm{co}}(K)$ if and only if there is a regular Borel probability measure μ on K such that*

$$y = \int_K x\,\mu(dx).$$

Remark 3.10. The integral is to be understood as in Definition 3.4 with $f(x) = x$. The integral represents every $y \in \overline{\mathrm{co}}(K)$ as a "weighted average" of K, or as the "center of mass" of a certain unit mass distributed over K.

Proof. Regard X again as a real vector space. Let $C(K)$ be the Banach space of all real continuous functions on K, with the supremum norm. The Riesz representation theorem identifies the dual space $C(K)^*$ with the space of all real Borel measures on K that are differences of regular positive ones. With this identification in mind, we define a mapping

$$\phi : C(K)^* \to X \ ; \ \phi(\mu) = \int_K x d\mu(x) \, .$$

Let P be the set of all regular Borel probability measures on K. The theorem asserts that $\phi(P) = \overline{\text{co}}(K)$. For each $x \in K$, the unit mass δ_x concentrated at x belongs to P, since $\phi(\delta_x) = x$, we see that $K \subset \phi(P)$. Since ϕ is linear and P is convex, it follows that $\text{co}(K) \subset \phi(P)$, where $\text{co}(K)$ is the convex hull of K. By Theorem 3.49, $\phi(P) \subset \overline{\text{co}}(K)$. Therefore all that remains to be done is to show that $\phi(P)$ is closed in X. This is a consequence of the following two facts:

(i) P is weak*-compact in $C(Q)^*$. (It's easy to show that P is weak*-closed, then use Theorem 5.15.)

(ii) The mapping ϕ is continuous if $C(Q)^*$ is given its weak*-topology and if X is given its weak topology.

Once we have (i) and (ii), it follows that $\phi(P)$ is weakly compact, hence weakly closed, and since weakly closed sets are strongly closed, we have the desired conclusion. □

Chapter 4

Fundermental Principles of Bounded Operators

4.1 Convergence of Operaters

Let X and Y be normed linear spaces over the same scalar field $\mathbb{F}$. We shall to define several types of convergence on the continuous operators from X to Y.

Definition 4.1. Let $A, A_n (n \geq 1)$ be a sequence of operators in $\mathcal{B}(X, Y)$.

(a) We sya that $\{A_n\}$ *converges uniformly* to A if

$$\|A_n - A\| \to 0\,. \tag{4.1}$$

In this case, A is called the *uniform limit* of the sequence $\{A_n\}$.

(b) We say that $\{A_n\}$ *converges strongly* to A if $A_n x$ converges strongly to Ax in Y for each $x \in X$, i.e.,

$$\|A_n x - Ax\| \to 0\,. \tag{4.2}$$

In this case, A is called the *strong limit* of the sequence $\{A_n\}$, denoted by $A = s\text{-}\lim A_n$ or $A_n \xrightarrow{s} A$.

(c) We say that $\{A_n\}$ *converges weakly* to A if $A_n x$ converges weakly to Ax in Y for each $x \in X$, i.e.,

$$\langle A_n x, y^* \rangle \to \langle Ax, y^* \rangle \quad \text{for each } y^* \in Y^* \,. \tag{4.3}$$

In this case A is called the *weak limit* of the sequence $\{A_n\}$, denoted by $A = w\text{-}\lim A_n$ or $A_n \xrightarrow{w} A$.

It is obvious that uniform convergence implies strong convergence, and strong convergence implies weak convergence. However, the converse does not hold.

Example 4.1. Consider the sequence $\{T_n\}$ on ℓ^2, where for each n, $T_n : \ell^2 \to \ell^2$ is given by

$$T_n(x_1, x_2, \cdots) = (0, 0, \cdots, 0, x_{n+1}, x_{n+2}, \cdots)$$

Then $T_n \to 0$ strongly, but $\|T_n\| = 1$ for all n, so $\{T_n\}$ does not converge to 0 in the uniform topology.

Example 4.2. Consider the sequence $\{S_n\}$ on ℓ^2, where $S : \ell^2 \to \ell^2$ is given by

$$S(x_1, x_2, \cdots) = (0, x_1, x_2, \cdots)$$

and $S_n = S^n$ for $n \in \mathbb{N}$.

Then $S_n \to 0$ weakly, but $\|S_n x\| = \|x\|$ for all x, thus $\{S_n x\}$ does not converge to 0, i.e., $\{S_n\}$ doesn't converge strongly to 0.

Topologies on $\mathcal{B}(X, Y)$. In fact, we can induce some topologies on $\mathcal{B}(X, Y)$ to describe this serval convergence.

(a) the operator norm $\|\cdot\|$ determines a natural norm topology $\mathcal{T}$ (or uniform topology) on $\mathcal{B}(X, Y)$.

(b) For any $x \in X$, define p_x by

$$p_x(A) = \|Ax\|\,, \text{ for all } A \in \mathcal{B}(X,Y)\,. \tag{4.4}$$

Then $\{p_x : x \in X\}$ is a separating family of seminorms on $\mathcal{B}(X,Y)$, inducing a locally convex topology, namely strong operator topology, denoted by $\mathcal{T}_s$. clearly, $\mathcal{T}_s$ is the weakest topology with respect to which $A \mapsto Ax$ is continuous on $\mathcal{B}(X,Y)$ for all $x \in X$, and $A_n \xrightarrow{s} A$ iff $A_n \to A$ in $\mathcal{T}_s$.

(c) For any $x \in X$ and $y^* \in Y^*$, define p_{x,y^*} by

$$p_{x,y^*}(A) = |\langle Ax, y^*\rangle|\,, \text{ for all } A \in \mathcal{B}(X,Y)\,. \tag{4.5}$$

Then $\{p_{x,y^*} : x \in X, y^* \in Y^*\}$ is a separating family of seminorms on $\mathcal{B}(X,Y)$, inducing a locally convex topology, namely weak operator topology, denoted by $\mathcal{T}_w$. Clearly, $\mathcal{T}_w$ is the weakest topology with respect to which $A \mapsto \langle Ax, y^*\rangle$ is continuous on $\mathcal{B}(X,Y)$ for all $x \in X$ and $y^* \in Y^*$, and $A_n \xrightarrow{w} A$ iff $A_n \to A$ in $\mathcal{T}_w$.

Evidently, $\mathcal{T}_w \subset \mathcal{T}_s \subset \mathcal{T}$. Moreover, when $Y = \mathbb{F}$, the scalars field, $\mathcal{B}(X,Y)$ is exactly X^*, then it's easy to find the $\mathcal{T}_s, \mathcal{T}_w$ coincides with $\sigma(X^*, X)$, the weak* topology on X.

Theorem 4.1. *Let X, Y be normed vector spaces over $\mathbb{F}$. If Λ is a linear functional on $\mathcal{B}(X,Y)$, then the following statements hold.*

(a) There exists an integer n, $x_1, \cdots, x_n$ in X and $y_1^*, \cdots, y_n^*$ in Y^* so that

$$\Lambda(A) = \sum_{k=1}^{n} \langle Ax_k, y_k^*\rangle \quad \textit{for all} \quad A \in \mathcal{B}(X,Y)\,.$$

(b) $\Lambda : (\mathcal{B}(X,Y), \mathcal{T}_w) \to \mathbb{F}$ is continuous.

(c) $\Lambda : (\mathcal{B}(X,Y), \mathcal{T}_s) \to \mathbb{F}$ is continuous.

Proof. Trivially (a) $\Rightarrow$ (b) $\Rightarrow$ (c). It suffices to prove that (c) implies (a).

By Theorem 3.15 and the definition of $\mathcal{T}_s$, there exists $x_1, \ldots, x_n$ and positive real numbers C so that

$$|\Lambda(A)| \le C \sum_{k=1}^{n} \|Ax_k\|, \quad \text{for all} \quad A \in \mathcal{B}(X,Y).$$

Thus one can see that $\Lambda(A) = \Lambda(B)$ if and only if $Ax_k = Bx_k$ for all $1 \le k \le n$. Then we can regard Λ as a linear functional on

$$Z \coloneqq \{(Ax_1, \ldots, Ax_n) : A \in \mathcal{B}(X,Y)\},$$

which is a linear subspace of $\oplus_1{}_{k=1}^{n} Y$ (see Section 1.1.3). Specifically, we define

$$\ell(Ax_1, \ldots, Ax_n) = \Lambda(A) \text{ for } (Ax_1, \ldots, Ax_n) \in Z.$$

It's easy to see that ℓ is linear. Indeed ℓ is also continuous since

$$|\ell(Ax_1, \ldots, Ax_n)| = |\Lambda(A)| \le C \sum_{k=1}^{n} \|Ax_k\| = \|(Ax_1, \ldots, Ax_n)\|.$$

Then by the Hahn-Banach extension theorem, we can regard ℓ is a bounded linear functional on $\oplus_1{}_{k=1}^{n} Y$. Then by Exercise 5.1, there are $y_1^*, \ldots, y_n^*$ in Y^* so that

$$\ell(Ax_1, \ldots, Ax_n) = \sum_{k=1}^{n} \langle Ax_k, y_k^* \rangle,$$

and hence the desired result follows. $\square$

4.2 Principle of Uniform Boundedness

Let $\mathcal{A}$ be a subset of $\mathcal{B}(X,Y)$. That is $\mathcal{A}$ is a family of continuous linear operators from X to Y.

(a) We say that $\mathcal{A}$ is *uniformly bounded* or *norm bounded* , if

$$\sup_{A \in \mathcal{A}} \|A\| < \infty.$$

(b) We say that $\mathcal{A}$ is *pointwise bounded* or *bounded for the strong operator topology*, if for each $x \in X$,

$$\sup_{A \in \mathcal{A}} \|Ax\| < \infty\,,$$

i.e., $\mathcal{A}$ is bounded with respect to the strong topology $\mathcal{T}_s$.

(c) We say that $\mathcal{A}$ is *bounded for the weak operator topology*, if for each $x \in X$ and $y^* \in Y^*$,

$$\sup_{A \in \mathcal{A}} |\langle Ax, y^* \rangle| < \infty\,,$$

i.e., $\mathcal{A}$ is bounded with respect to the weak topology $\mathcal{T}_w$.

Obviously, $\mathcal{A}$ is uniformly bounded implies that it is pointwise bounded; and $\mathcal{A}$ is pointwise bounded implies that it is bounded for the weak operator topology. Surprisingly, in the case that X is a Banach space, these three types of boundedness is equivalent. Let's see the famous principle of uniform boundedness first, which asserts that pointwise boundedness is equivalent to uniform boundedness.

Theorem 4.2 (Principle of Uniform Boundedness)**.** *Let X be a Banach space and let Y a normed space. If $\mathcal{A} \subseteq \mathcal{B}(X, Y)$ is pointwise bounded, then $\mathcal{A}$ is uniformly bounded.*

Proof. For each $k \geq 1$, let

$$E_k = \{x \in X : \|Ax\| \leq k \text{ for all } A \in \mathcal{A}\}\ .$$

since A is continuous, E_k is closed. Note that $X = \cup_{k=1}^{\infty} E_k$. By Baire's category theorem, there is an index m such that $\mathrm{int}(E_m) \neq \emptyset$.

That is, there is an $x_0 \in E_m$ and an $\epsilon > 0$ such that $B(x_0, 2\epsilon) \subset E_m$. Then for any $\|x\| \leq 1$, $x_0 + \epsilon x \in B(x_0, 2\epsilon)$, so

$$\|A(x_0 + \epsilon x)\| \leq m\,, \quad \text{for all } A \in \mathcal{A}\,.$$

Thus

$$\|Ax\| = \frac{\|Ax_0\| + m\|}{\epsilon} \le \frac{2m}{\epsilon}\,, \quad \text{for all } A \in \mathcal{A}\,.$$

So $\mathcal{A}$ is uniformly bounded. $\square$

Taking $Y = \mathbb{F}$, we get the following corollary:

Corollary 4.3. *X is a Banach space and $N \subset X^*$, then N is norm bounded iff for every x in X,*

$$\sup_{x^* \in N} |\langle x, x^* \rangle| < \infty\,.$$

In other words, N is weak*-bounded implies N is norm bounded.

To see the relationship between weak boundedness and uniform boundedness in $\mathcal{B}(X, Y)$, we shall to use PUB carefully. It should be emphasized again that for any $x \in X$, we can regard x as a linear functional on X^*:

$$x^* \mapsto \langle x, x^* \rangle \;;\; X^* \to \mathbb{F}\,.$$

In order to ensure rigor, denote by Jx this functional. It is easy to verify that $Jx \in X^{**}$, and $|Jx| = \|x\|$. So using PUB to X^*, we get :

Lemma 4.4. *Let X be a normed vector space. $N \subset X^*$, then N is norm bounded iff for every x^* in X^*,*

$$\sup_{x \in N} |\langle x, x^* \rangle| < \infty\,.$$

In other words, N is weak-bounded implies N is norm bounded.

Using PUB twice, we get the equivalence between boundedness for weak operator topology and for norm topology.

Theorem 4.5. *Let X be a Banach space and let Y be a normed space. If $\mathcal{A} \subset \mathcal{B}(X, Y)$ is weakly bounded, i.e., for every x in X and y^* in Y^**

$$\sup_{A \in \mathcal{A}} |\langle Ax, y^* \rangle| < \infty\,,$$

then $\mathcal{A}$ is uniformly bounded.

Example 4.3. This example shows that the hypothesis that X is complete cannot be removed in PUB. Consider the space $X = \ell_0$ with the supremum norm. We have show that it is not complete, but is a linear subspace of ℓ^∞ whose closure $\bar{\ell_0} = c_0$. Define the linear operators $A_n : \ell_0 \to \ell_0$ and $A : \ell_0 \to \ell_0$ by

$$A_n x := (x_1, 2x_2, \ldots, nx_n, 0, 0, \ldots)\,, \quad Ax := (nx_n)_{n\geq 1}$$

for $n \geq 1$ and $x = (x_n)_{n\geq 1} \in \ell_0$. Then $Ax = \lim_{n\to\infty} A_n x$ for every $x \in X$ and $\|A_n\| = n$ for every $n \geq 1$. Thus the sequence $\{A_n x\}_{n\geq 1}$ is bounded for every $x \in X$, the linear operator A is not bounded, and the sequence A_n converges strongly to A.

Example 4.4 (Bilinear Map)**.** Let X, Y and Z be vector spaces. Let $B : X \times Y \to Z$. Associate to each $x \in X$ and to each $y \in Y$ the mappings

$$B_x : Y \to Z \quad \text{ and } \quad B^y : X \to Z$$

by defining

$$B_x(y) = B(x, y) = B^y(x)\,.$$

B is said to be *bilinear* if every B_x and every B^y are linear. If X, Y, Z are normed vector spaces and if every B_x and every B^y is continuous, then B is said to be *separately continuous*. If B is continuous (relative to the product topology of $X \times Y$) then B is obviously separately continuous. In certain situations, the converse can be proved with the aid of the PUB.

Let X be a Banach space and Y, Z be normed linear space. Then the following propositions are equivalent.

(a) B is *bounded*, i.e. there is a constant $C \geq 0$ such that

$$\|B(x, y)\| \leq C\|x\|\|y\|\,, \quad \text{for all } x \in X \text{ and all } y \in Y\,.$$

(b) B is continuous.

(c) B is obviously separately continuous.

Appications. We will use the PUB to deduce some results about strong convergence.

Proposition 4.6. *Let X be a Banach space and let Y be a normed space. Let $A, A_n \in \mathcal{B}(X, Y)$. If A_n converges to A strongly, then $\{A_n\}$ is uniformly bounded and,*

$$\|A\| \leq \liminf_{n\to\infty} \|A_n\| . \tag{4.6}$$

Proof. It follows from PUB that $\{A_n\}$ is uniformly bounded. Notice that for each $x \in X$,

$$\|Ax\| = \lim_{n\to\infty} \|A_n x\| \leq \liminf_{n\to\infty} \|A_n\| \|x\| .$$

Thus $\|A\| \leq \liminf_n \|A_n\|$. $\square$

The following lemma is quite useful:

Lemma 4.7. *Let X be a Banach space and let Y be a normed space. Let $\{A_n\}$ be a sequence in $\mathcal{B}(X, Y)$. Then $\{A_n\}$ is strongly convergent iff*

(a) *$\{A_n\}$ is uniformly bounded and,*

(b) *$\{A_n x\}$ converges for all $x \in X$.*

Proof. Necessity is obvious. To show sufficiency, we define $A : X \to Y$ by

$$Ax := \lim_{n\to\infty} A_n x .$$

A is linear, clearly. A is continuous by the proof of Proposition 4.6. $\square$

Theorem 4.8 (Banach-Steinhaus)**.** *Let X, Y be two Banach spaces. Let $\{A_n\}$ be a sequence in $\mathcal{B}(X, Y)$. Then $\{A_n\}$ is strongly convergent iff*

(a) *$\{A_n\}$ is uniformly bounded and,*

(b) *there exists a dense subset D of X so that $\{A_n x\}$ converges for each $x \in D$.*

Proof. It suffices to show that $\{A_n x\}$ is a Cauchy sequence for all $x \in X$, and this is why we need Y is complete. For any $x \in X$, and given $\epsilon > 0$, there exists $x' \in D$, depending on ϵ, so that $\|x - x'\| \leq \epsilon$. Then

$$\begin{aligned}\|A_{n+p}x - A_n x\| &\leq \|A_{n+p}x - A_{n+p}x'\| + \|A_{n+p}x' - A_n x'\| + \|A_n x' - A_n x\| \\ &\leq 2\sup_n \|A_n\|\epsilon + \|A_{n+p}x' - A_n x'\|\end{aligned}$$

Since $\{A_n x'\}$ is a Cauchy sequence, it's easy to find that $\{A_n x\}$ is a Cauchy sequence. Then the desired result follows. □

4.3 Open Mappings

Let X and Y be two normed linear spaces. Let $A : X \to Y$ be a linear operator. We call A an *open mapping* if $A(U)$ is open in Y whenever U is open in X.

Theorem 4.9 (Open Mapping Theorem)**.** *Let X, Y be Banach spaces. Let $A \in \mathcal{B}(X, Y)$. If A is surjective, i.e., $R(A) = Y$, then A is an open mapping.*

Proof. *Step 1.* Note that for any $x \in X$, $AB_X(x, r) = Ax + rB_X(0, 1)$, We have only to prove that $AB_X(0, 1)$ is open. To this end, it must be shown that there is constant $r > 0$ such that $B_Y(0, r) \subset AB_X(0, 1)$. However, this is also sufficient, in deed, for any $y = Ax \in AB_X(0, 1)$, take $\epsilon > 0$ such that $\|x\| + \epsilon < 1$, then

$$B_Y(y, \epsilon) = Ax + B_Y(0, \epsilon) \subset AB_X(0, 1)\,,$$

which implies that $AB_X(0, 1)$ is open.

Step 2. We shall show that there is a constant $r > 0$ such that

$$B_Y(0, r) \subset \overline{AB_X(0, 1)}\,.$$

It is easy to see that $X = \bigcup_{n=1}^{\infty} nB_X(0, 1)$. Since T is surjective,

$$Y = AX = A\left(\bigcup_{n=1}^{\infty} nB_X(0, 1)\right) = \bigcup_{n=1}^{\infty} nAB_X(0, 1) = \bigcup_{n=1}^{\infty} n\overline{AB_X(0, 1)}\,.$$

By Baire's category theorem, there is $m \geq 1$ such that $\left(m\overline{AB_X(0,1)}\right)^\circ \neq \emptyset$. This implies that $\left(\overline{AB_X(0,1)}\right)^\circ \neq \emptyset$. Hence, there is a constant $r > 0$ and an element $y_0 \in Y$ such that $B_Y(y_0, 2r) \subset \overline{AB_X(0,1)}$. Since $y_0 \in \overline{AB_X(0,1)}$, it follows, by symmetry, that $-y_0 \in \overline{AB_X(0,1)}$. Therefore

$$B_Y(0, 2r) = B_Y(y_0, 2r) - y_0 \subset \overline{AB_X(0,1)} + \overline{AB_X(0,1)}\,.$$

Since $\overline{TB_x(0,1)}$ is convex, $\overline{AB_X(0,1)} + \overline{AB_X(0,1)} = 2\overline{AB_X(0,1)}$. Hence, $B_Y(0,2r) \subset 2\overline{AB_X(0,1)}$ and, consequently, $B_Y(0,r) \subset \overline{AB_X(0,1)}$.

Step 3. We shall prove that $B_Y(0, r/2) \subset AB_X(0,1)$.

Take any $y \in B_Y(0, r/2)$. From $B_Y(0,r) \subset \overline{AB_X(0,1)}$, we have

$$B_Y(0, r/2^n) \subset \overline{AB_X(0, 1/2^n)}\,,\ n \geq 1.$$

So there is $x_1 \in B_X(0,1/2)$ such that $y - Ax_1 \in B_Y(0, r/2^2)$. By induction there is a suquence $\{x_n\}$ such that $x_n \in B_X(0, 1/2^n)$ and

$$y - A(x_1 + \cdots + x_n) \in B_X(0, 1/2^n)\,.$$

Note that $\sum_{n=1}^{\infty} \|x_n\| < 1$, since X is Banach space, there is $z \in B_X(0,1)$ such that $z = \sum_{n=1}^{\infty} x_n$, and $Az = y$. Thus $B_Y(0, r/2) \subset AB_X(0,1)$. $\square$

Remark 4.1. Conversely, it's evident to see that if $A : X \to Y$ is an open mapping, then A is surjective.

From the open mapping theorem, we get the following theorem directly.

Theorem 4.10 (Inverse Mapping Theorem)**.** *Let X, Y be Banach spaces. Let $A \in \mathcal{B}(X,Y)$. If A is bijective, then $A^{-1} \in \mathcal{B}(Y,X)$. In other words, A is a linear homomorphism.*

In fact, the inverse mapping theorem and the open mapping theorem are equivalent. To see this, suppose $A \in \mathcal{B}(X,Y)$ is a surjection. Observe from Example 1.20 and that

$$\tilde{A} : X/N(A) \to Y\,;\, \tilde{x} \mapsto Ax$$

is bijective continuous linear operator from $X/N(A)$ onto Y. By Theorem 1.17, $X/N(A)$ is Banach space. So the inverse mapping theorem implies that

$$\tilde{A}^{-1} : Y \to X/N(A)\,;\; Ax \mapsto \tilde{x} \quad \text{is continuous}\ .$$

Then we show that there exists $r > 0$ so that $B_Y(0,r) \subset AB_X(0,1)$. Notice that for any $y = Ax \in B_Y(0,r)$

$$\|\tilde{x}\| = \|\tilde{A}^{-1}\tilde{A}\tilde{x}\| \leq \|\tilde{A}^{-1}\|\|y\| \leq \|\tilde{A}^{-1}\|r\,.$$

Let $\|\tilde{A}^{-1}\|r < 1$, then $\|\tilde{x}\| \leq 1$, so there exist $m \in N(A)$ so that $\|x+m\| < 1$ and $Ax = A(x+m) \in AB_X(0,1)$.

As a consequence of the inverse mapping theorem we will see that when X is a Banach space with respect to two different norms then determining whether those two norms are equivalent is simpler.

Theorem 4.11 (Equivalence of Norms on Banach Spaces)**.** *Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be two norms defined on X and let X be a Banach space with respect to both of these norms. Suppose that there exists a constant $C > 0$ such that for all $x \in X$ we have that $\|x\|_2 \leq C\|x\|_1$. Then $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent.*

Proof. Consider the identity mapping $I : (X, \|\cdot\|_1) \to (X, \|\cdot\|_2)$. Trivially, I is a continuous linear bijection from X onto itself. By the inverse mapping theorem, $I^{-1} = I : (X, \|\cdot\|_2) \to (X, \|\cdot\|_1)$ is bounded, which implies the desired result. □

Example 4.5. Let $(X, \|\cdot\|)$ be a Banach space and let $M, N \subset X$ be two closed linear subspaces such that

$$X = M \oplus N$$

i.e., $M \cap N = \{0\}$ and every vector $x \in X$ can be written as $x = m + n$ with $m \in M$ and $n \in N$. Then it follows from Corollary 4.11 that there exists a constant $C \geq 0$ such that

$$\|m\| + \|n\| \leq C\,\|m+n\|$$

for all $m \in M$ and all $n \in N$. In other words, the canonical mapping from $(M \oplus N, \|\cdot\|)$ onto $M \oplus_1 N$, $m + n \mapsto (m, n)$ is a linear homeomorphism.

Example 4.6. This example shows that the hypothesis that X and Y are complete cannot be removed in the open mapping theorem and the inverse mapping theorem. Let $X = \ell_0$, equipped with the supremum norm. Thus X is a normed vector space but is not a Banach space. Define the operator $A : X \to X$ by $Ax := \left(k^{-1}x_k\right)_{k\in\mathbb{N}}$ for $x = (x_k)_{k\in\mathbb{N}} \in X$. Then A is a bijective bounded linear operator but its inverse is unbounded.

Example 4.7. Here is an example where X is complete and Y is not. Let $X = Y = C([0,1])$ be the space of continuous functions $f : [0,1] \to \mathbb{R}$ equipped with the norms

$$\|f\|_X := \sup_{0\leq t\leq 1} |f(t)|, \quad \|f\|_Y := \sqrt{\int_0^1 |f(t)|^2 dt}$$

Then X is a Banach space, Y is a normed vector space, and the identity

$$I : X \to Y$$

is a bijective bounded linear operator with an unbounded inverse.

Exercise 4.1. Let $A \in \mathcal{B}(X,Y)$. Show that $A^{-1} : R(A) \to X$ exists and is continuous, i.e., A is an linear homeomorphism between X and $R(A)$ if and noly if there is a constant $m > 0$ such that

$$\|Ax\| \geq m\|x\| \quad \text{for each } x \in X\,.$$

In this case, if X is a Banach space then $R(A)$ is closed.

4.4 Closed Operators

Let X and Y be two linear spaces. Let A be a linear operator on $D(A) \subset X$ into Y. The *graph* of A, denoted by $G(A)$, is the subset of $X \times Y$ given by

$$G(A) = \{(x, Ax) : x \in X\} .$$

Obviously, $G(A)$ is a linear subspace of $X \times Y$.

Let X and Y be normed linear spaces on the same scalar field. It follows from Proposition 1.3 that we can equip the product space $X \times Y$ so that norm, so that the coordinate projections $(x, y) \mapsto x$ and $(x, y) \to y$ are continuous. Specifically, let $p \in [1, \infty]$, then the norm defined by

$$\|(x, y)\|_p := \begin{cases} [\|x\|^p + \|y\|^p]^{1/p} & \text{if } p < \infty \\ \max\{\|x\|, \|y\|\} & \text{if } p = \infty \end{cases} \quad \text{for all} \quad x \in X\,, y \in Y$$

satisfies the desired properties. It's easy to see that $\|\cdot\|_p$ are equivalent for $p \in [1, \infty]$. Thus we write $X \oplus Y$ for $X \times Y$ endowed $\|\cdot\|_1$ for simplicity; except the case that X and Y are Hilbert space, we will endow $X \times Y$ endowed $\|\cdot\|_2$ to ensure that $X \oplus Y$ becomes an inner product space (see Proposition 2.5). Clearly, if X, Y are complete, then so is $X \oplus Y$.

Definition 4.2. Let X and Y be normed linear spaces, and A a linear operator from $D(A) \subset X$ into Y. Then A is called a *closed linear operator* if it's graph $G(A)$ is a closed subspace of $X \oplus Y$.

The *graph norm* of A on the linear subspace $D(A) \subset X$ is the norm function $D(A) \to [0, \infty) : x \mapsto \|x\|_A$ defined by

$$\|x\|_A := \|(x, Ax)\|_{X \oplus Y} .$$

Note that a linear operator $A : X \supset D(A) \to Y$ is always a continuous linear operator with respect to the graph norm.

Proposition 4.12. *The following statements are equivalent:*

(i) *A is closed.*

(ii) *If (x_n) is a sequence of vectors $x_n \in D(T)$ such that x_n converges to X in X and Ax_n converges to y in Y, then $x \in D(A)$ and $Ax = y$.*

(iii) *$(D(A), \|\cdot\|_A)$ is a Banach space.*

The following proposition shows some easy properties of closed operator. We omit the trivial proof.

Exercise 4.2. Let X and Y be normed linear spaces. Let A be a linear operator on $D(A) \subset X$ into Y.

(a) If A is closed, then the null space $N(A)$ is a closed subspace of X.

(b) If A is injective, then $A^{-1} : Y \supset R(A) \to X$ is also a closed linear operator.

4.4.1 The Closed Graph Theorem

Evidently, if A is a continuous linear operator and $D(A)$ is a closed subspace of X, then A is closed. When Y is a Banach space, every continuous linear operator can be extended as a closed operator.

Lemma 4.13. *Let X be a normed vector space and let Y be a Banach space. Let $A : X \supset D(A) \to Y$ be a continuous linear operator. Then there exists a unique continuous linear operator $\bar{A} : \overline{D(A)} \to Y$ so that*

$$\bar{A}|_{D(A)} = A \quad \textit{and} \quad \|\bar{A}\| = \|A\| \,. \tag{4.7}$$

Proof. For any $x \in \overline{D(A)}$, there exists a sequence $\{x_n\}$ in $D(A)$ so that $x_n \to x$. Since A is a bounded linear operator on $D(A)$, $\{Ax_n\}$ is a Cauchy sequence in Y. By the completeness of Y, $\{Ax_n\}$ is convergent in Y. Define

$$\bar{A}x \coloneqq \lim_{n\to\infty} Ax_n \,.$$

Clearly, $\bar{A}$ is well-defined linear operator on $\overline{D}(A)$ with $\bar{A}|_{D(A)} = A$ and $\|\bar{A}\| = \|A\|$. We noe complete the proof. □

On the contrary, a large kind of closed operators are indeed continuous, this is a consequence of the inverse mapping theorem.

Theorem 4.14 (The Closed Graph Theorem)**.** *Let X and Y be Banach spaces. Let A be a closed linear operator from $D(A) \subset X$ into Y. If $D(A)$ is a closed subspace of X, then A is a continuous.*

Proof. Define the projection

$$P : G(A) \to D(A) \subset X\,;\, (x, Ax) \mapsto x\,.$$

It is easy to check that P is a continuous bijection between $G(A)$ and $D(A)$. Since $G(A)$ is the closed subspace of Banach space $X \oplus Y$, and $D(A)$ is a closed subspace of Banach space X, both $G(A)$ and $D(A)$ are Banach space. By the inverse mapping theorem,

$$P^{-1} : D(A) \to G(A)\;;\; x \mapsto (x, Ax)$$

is continuous. Thus $A : D(A) \to Y$ is the composition of the continuous map $P^{-1} : D(A) \to G(A)$ and the continuous map of $G(A) \to Y$ defined by $(x, Ax) \mapsto Ax$. Therefore, A is continuous too. $\square$

Remark 4.2. Indeed the converse of the closed graph theorem also holds: if the closed linear operator $(A, D(A))$ is also continuous, then $D(A)$ is a closed subspace in X. The proof is trivial. Assume the sequence (x_n) in $D(A)$ converges to some point $x \in X$, then since A is continuous, (Ax_n) is a Cauchy sequence in Y. Let $Ax_n \to y$, using the fact that A is closed, we have $x \in D(A)$ (and $y = Ax$). We are done.

Remark 4.3. In fact, the closed graph theorem is equivalent to the inverse mapping theorem. The proof is trivial. Let $A \in \mathcal{B}(X, Y)$ be a bijection. It suffices to show that A^{-1} is a closed operator. Let $y_n = Ax_n \to y$ and $x_n = A^{-1}y_n \to x$, then it follows from the continuity of A that $Ax = \lim_n Ax_n = \lim_n y_n = y$. So the desired result follows.

Example 4.8 (Hellinger-Toeplitz Theorem)**.** Let H be a Hilbert space over $\mathbb{F}$. Let $A : H \to H$ be a *self-adjoint* linear operator, i.e.,

$$\langle x, Ay\rangle = \langle Ax, y\rangle \quad \text{for all } x, y \in H\,.$$

Then A is continuous. To see this, it suffices to prove that A has a closed graph. Thus assume that $\{x_n\}$ is a sequence in H and $x, y \in H$ are vectors such that

$$\lim_{n\to\infty} x_n = x, \quad \lim_{n\to\infty} Ax_n = y\,.$$

In order that $Ax = y$, notice that for any $z \in H$,

$$\langle y, z\rangle = \lim_{n\to\infty} \langle Ax_n, z\rangle = \lim_{n\to\infty} \langle x_n, Az\rangle = \langle x, Az\rangle = \langle Ax, z\rangle\,.$$

Then the desired result follows.

Example 4.9 (Douglas Factorization)**.** Let X, Y, Z be Banach spaces and let $A : X \to Y$ and $B : Z \to Y$ be bounded linear operators. Assume A is injective. Then the following are equivalent.

(a) $R(B) \subset R(A)$.

(b) There is a continuous linear operator $T : Z \to X$ such that $AT = B$.

We now use the closed graph theorem to prove this. If (b) holds, then $R(B) = R(AT) \subset R(A)$. Conversely, suppose that $R(B) \subset R(A)$ and define

$$T := A^{-1} \circ B : Z \to X$$

Then T is a linear operator and $AT = B$. We prove that T has a closed graph. To see this, let $(z_n)\,d$ be a sequence in Z such that the limits

$$z := \lim_{n\to\infty} z_n, \quad x := \lim_{n\to\infty} Tz_n$$

exist. Then

$$Ax = \lim_{n\to\infty} ATz_n = \lim_{n\to\infty} Bz_n = Bz$$

and hence $x = Tz$. Thus T has a closed graph and $D(T) = Z$, by the closed graph theorem, hence T is continuous.

We will give an easy example of a discontinuous but closed operator.

Example 4.10. Let $X = C[0,1]$ endowed with sup-norm. Let $D = C^1[0,1] \subset C[0,1]$ and let T be the linear operator on D into X defined by

$$(Tx)(t) = \frac{d}{dt}x(t) = x'(t) \qquad \text{for all } t \in [0,1]\,.$$

This T is not continuous, since, for $x_n(t) = t^n$, $\|x_n\| = 1$, but $\|Tx_n\| = n$ for $n \geq 1$. However, T is closed. In fact, let $\{x_n\} \subset D(T)$, $x_n \to x$ and $Tx_n \to y$. Then $x_n'(t)$ converges uniformly to $y(t)$, and $x_n(t)$ converges uniformly to $x(t)$. Hence $x(t)$ must be differentiable with continuous derivative $y(t)$. This proves that $x \in D(T)$ and $Tx = y$.

Besides, the graph norm of T on $D(T) = C^1[0,1]$ agrees with the usual C^1-norm

$$\|f\|_{C^1} = \sup_{0\leq t\leq 1} |f(t)| + \sup_{0\leq t\leq 1} \left|f'(t)\right| \qquad \text{for } f \in C^1[0,1]$$

and then $C^1[0,1]$ is a Banach space with this norm.

Exercise 4.3 (Open Mapping Theorm for Closed Operators). Let X and Y be Banach spaces. Let $A : X \supset D(A) \to Y$ be a closed linear operator. Show that if $R(A)$ is of the second category in Y, then A is an open mapping form $D(A)$ onto Y. In particular, $R(A) = Y$.

4.4.2 Closeable Operator, Closure and Core

For a linear operator that is defined on a proper linear subspace it is an interesting question whether it can be extended to a linear operator with a closed graph. Such linear operators are called *closable*. So, any continuous linear operator A form $D(A) \subset X$ into Y is closable, if Y is complete.

Lemma 4.15 (Characterization of Closeable Operators). *Let X and Y be Banach spaces, let $D(A) \subset X$ be a linear subspace, and let $A : D(A) \to Y$ be a linear operator. Then the following are equivalent.*

(i) A is closable.

(ii) If $\{x_n\}$ is a sequence in $D(A)$ and $y \in Y$ is a vector such that $x_n \to 0$ and $Ax_n \to y$, then $y = 0$.

(iii) $\overline{G(A)}$ is a graph of some operator. i.e., the projection onto the first factor $\pi_X : \overline{G(A)} \to X$, $(x, y) \mapsto x$ is injective.

Proof. (i) implies (ii) is trivial.

We prove that (ii) implies (iii). The closure of any linear subspace of a normed vector space is again a linear subspace. Hence $\overline{G(A)}$ is a linear subspace of $X \times Y$ and the projection $\pi_X : \overline{G(A)} \to X$ is a linear map by definition. By (ii) the kernel of this linear map is the zero subspace and hence it is injective.

We prove that (iii) implies (i). Define

$$D\left(\bar{A}\right) := \pi_X(\overline{G(A)}) \subset X$$

This is a linear subspace and the map $\pi_X : \overline{G(A)} \to D\left(\bar{A}\right)$ is bijective by (iii). Denote its inverse by $\pi_X^{-1} : D\left(\bar{A}\right) \to \overline{G(A)}$ and denote by

$$\pi_Y : \overline{G(A)} \to Y$$

the projection onto the second factor. Then

$$\bar{A} := \pi_Y \circ \pi_X^{-1} : D\left(\bar{A}\right) \to Y \, ; \, (x, y) \mapsto y$$

is a linear operator, since π_X and π_Y are linear. Its graph is the linear subspace

$$G\left(\bar{A}\right) = \overline{G(A)} \subset X \times Y$$

and $\bar{A}|_{D(A)} = A$ holds because $G(A) \subset G\left(\bar{A}\right)$. Thus $\bar{A}$ is a closed extension of A and hence A is closable. $\square$

By Lemma 4.15, if A is closable, then $\overline{G(A)}$ is a graph. Thus there exists a unique closed linear operator $\bar{A}$ with $G(\bar{A}) = \overline{G(A)}$. Clearly, if a closed linear operator T is a extension of A, then $G(A) \subset G(T)$. Since $G(T)$ is closed, $G(\bar{A}) = \overline{G(A)} \subset G(T)$, thus T is a extension of $\bar{A}$. Therefore, $\bar{A}$ is the smallest closed extension of A.

Definition 4.3. For a closable operator A, the operator $\bar{A}$ given above is called the *closure* of A.

Another useful notion is that of a core of an operator.

Definition 4.4. A linear subspace D of $D(A)$ is called a *core* for A if D is dense in $(D(A), \|\cdot\|_A)$, that is, for each $x \in D(A)$, there exists a sequence (x_n) in D such that $x_n \to x$ in X and $Ax_n \to Ax$ in Y.

If A is closed, a linear subspace D of $D(A)$ is a core for A if and only if A is the closure of its restriction $A|_D$. That is, a closed operator can be restored from its restriction to any core. The advantage of a core is that closed operators are often easier to handle on appropriate cores rather than on full domains.

Example 4.11 (Symmetric Operators)**.** Let H be a Hilbert space over $\mathbb{F}$ and let $A : D(A) \to H$ be a linear operator, defined on a *dense* linear subspace $D(A) \subset H$. Suppose A is *symmetric*, i.e.

$$\langle x, Ay \rangle = \langle Ax, y \rangle \quad \text{for all } x, y \in D(A)\,.$$

Then A is closable. To see this, choose a sequence $\{x_n\}$ in $D(A)$ such that $x_n \to 0$ and the sequence $Ax_n \to y \in H$. Then for all $z \in D(A)$,

$$\langle y, z \rangle = \lim_{n\to\infty} \langle Ax_n, z \rangle = \lim_{n\to\infty} \langle x_n, Az \rangle = 0\,.$$

Since $D(A)$ is a dense subspace of H, we deduce that $y = 0$. Thus A is closable.

Example 4.12 (Differential Operators)**.** This example shows that *differential operators are closable.* Let $\Omega \subset \mathbb{R}^n$ be a nonempty open set. Consider the Hilbert space $X := L^2(\Omega; \mathbb{R})$. Then the space

$$D(A) := C_c^\infty(\Omega)$$

of smooth functions $u : \Omega \to \mathbb{R}$ with compact support is a dense linear subspace of $L^2(\Omega)$. Let $m \geq 1$ and, for every multi-index $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0^n$ with $|\alpha| = \alpha_1 + \cdots + \alpha_n \leq m$, let $\phi_\alpha : \Omega \to \mathbb{R}$ be a smooth function. Define the operator $A : C_c^\infty(\Omega) \to L^2(\Omega)$ by

$$Au := \sum_{|\alpha| \leq m} \phi_\alpha \partial^\alpha u \tag{4.8}$$

Here the sum runs over all multi-indices $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0^n$ with $|\alpha| \leq m$ and

$$\partial^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}}.$$

We prove that A is closable. To this end, define the *formal adjoint* of A as the operator $B : C_c^\infty(\Omega) \to L^2(\Omega)$, given by

$$Bv := \sum_{|\alpha| \leq m} (-1)^{|\alpha|} \partial^\alpha (\phi_\alpha v)$$

for $v \in C_c^\infty(\Omega)$. Then using integration by parts, we have

$$\int_\Omega v(Au) = \int_\Omega (Bv)u \tag{4.9}$$

for all $u, v \in C_c^\infty(\Omega)$. Now let $u_k \in C_c^\infty(\Omega)$ be a sequence of smooth functions with compact support and let $v \in L^2(\Omega)$ such that

$$\lim_{k\to\infty} \|u_k\|_{L^2} = 0, \quad \lim_{k\to\infty} \|v - Au_k\|_{L^2} = 0$$

Then, for every test function $\phi \in C_c^\infty(\Omega)$, we have

$$\int_\Omega \phi v = \lim_{k\to\infty} \int_\Omega \phi\,(Au_k) = \lim_{k\to\infty} \int_\Omega (B\phi) u_k = 0$$

since $C_c^\infty(\Omega)$ is dense in $L^2(\Omega)$, this implies that

$$\int_\Omega \phi v = 0 \quad \text{ for all } \phi \in L^2(\Omega)$$

Now take $\phi := \operatorname{sign}(v)|v| \in L^2(\Omega)$ to obtain $\int_\Omega |v|^2 = 0$ and hence v vanishes almost everywhere. Hence the linear operator A is closable, as claimed.

Example 4.13 (Nonclosable Operator)**.** Let D be a linear subspace of a Hilbert space H, and let $e \neq 0$ be a vector of H. Let F be a linear functional on D which is NOT continuous in the Hilbert space norm.

Define the operator T by $D(T) = D$ and $T(x) = F(x)e$ for $x \in D$. We claim that T is not closable.

Indeed, since F is not continuous, there exists a sequence (x_n) from $\mathcal{D}$ such that $x_n \to 0$ in H and $(F(x_n))$ does not converge to zero. By passing to a subsequence if necessary we can assume that there is a constant $c > 0$ such that $|F(x_n)| \geq c$ for all n. Putting $x_n' = F(x_n)^{-1} x_n$, we have $\lim_n x_n' = 0$ and $T(x_n') = F(x_n')e = e \neq 0$. Hence, T is not closable. The preceding proof has also shown that $(0, e) \in \overline{\mathcal{G}(T)}$, so $\overline{\mathcal{G}(T)}$ is not the graph of a linear operator.

Explicit examples of discontinuous linear functionals are easily obtained as follows: If D is the linear span of an orthonormal sequence (e_n) of a Hilbert space, define F on D by $F(e_n) = 1$ for all n. If $\mathcal{H} = L^2(\mathbb{R})$ and $D = C_0^\infty(\mathbb{R})$, define $F(f) = f(0)$ for $f \in D$.

4.5 Projections, Direct Sum Decomposition

Let X be a normed vector space. We say $P \in \mathcal{B}(X)$ is called a *projection* if

$$P^2 = P\,.$$

It's easy to see that if $P \neq 0$, then there must hold

$$\|P\| \geq 1\,,$$

since $\|P\| = \|P^2\| \leq \|P\|^2$.

Proposition 4.16. *Let X be a normed vector space and $P \in \mathcal{B}(X)$ is a projection. Then the following statements hold.*

(a) $I - P$ *is also a projection.*

(b) $N(P) = R(I-P)$, $R(P) = N(I-P)$ *and both* $R(P)$ *and* $N(P)$ *are closed subspaces of* X.

(c) $X = R(P) \oplus N(P)$.

Proof. To show (a), observe that

$$(I-P)^2 = I - 2P + P^2 = I - 2P + P = I - P\,,$$

thus $I - P$ is also an projection.

To show (b), since P is continuous, $N(P)$ is a closed subspace of X. Also,

$$x \in R(P) \Leftrightarrow Px = x \Leftrightarrow (I-P)x = 0 \Leftrightarrow x \in N(I-P).$$

Similarly, $R(I-P) = NP$.

To show (c), note that $x = Px + (I-P)x$. □

From this proposition, we say P is the projection from X to $R(P)$. Moreover, M is a closed subspace of X, we say there exists a projection P from X to M, if there exists projection P such that $R(P) = M$. Clearly, if X is a Hilbert space such a projection exists but is not determined uniquely since the "orthogonality" is not asked. When X is a Banach space, does such projections exist?

Theorem 4.17. *Let X be a Banach space. If M, N are two closed subspace of X, such that $X = M \oplus N$. Then there is a unique projection P satisfying*

$$R(P) = M \text{ and } N(P) = N\,. \tag{4.10}$$

Proof. We show the uniqueness first. Suppose P_1 and P_2 are two projections satisfying (4.10). Then by Proposition 4.16 we have

$$N(P_1) = N(P_2) = R(I - P_2) \ ; \ N(I - P_2) = R(P_2) = R(P_1) \ ;$$
$$N(P_2) = N(P_1) = R(I - P_1) \ ; \ N(I - P_1) = R(P_1) = R(P_2) \ .$$

Thus

$$P_1(I - P_2) = (I - P_2)P_1 = 0 \ ;$$
$$P_2(I - P_1) = (I - P_1)P_2 = 0 \ ,$$

and hence $P_1 = P_1P_2 = P_2P_1 = P_2$.

Next, we give the construction of the projection P. For each $x \in X$, there is a unique composition

$$x = m + n \, .$$

where $m \in M$ and $n \in N$. Define P by

$$Px \coloneqq m \, .$$

P is linear, since the composition is unique. Clearly, $P^2 = P$. It follows from Example 4.5 that P is bounded. So the desired result follows.

Also, we can sue the closed graph theorem. It suffices to show P has a closed graph. Assume $x_n \to x$ and $Px_n \to m$, then

$$x_n - Px_n \to x - m \, .$$

Since M, N are closed, $m \in M$ and $x - m \in N$. Thus we have $Px = m$. So P is bounded. □

Definition 4.5. X is a normed space, a closed subspace M is said to be *complemented in* X, if there exists a closed subspace N such that

$$X = M \oplus N \, .$$

Thus a projection operator is equivalent to a direct sum decomposition of X. At present we need only the following simple facts.

Lemma 4.18. *M is a closed subspace of normed space X. If*

$$\dim M < \infty \quad \textit{or} \quad \operatorname{codim} M \coloneqq \dim(X/M) < \infty\,,$$

there is a projection P form X to M, i.e., M is complemented.

Proof. Case 1. Assume $\dim M = n$, and

$$\{e_1, e_2, \cdots, e_n\}$$

a base for M. For each $x \in M$, there exists unique $\{c_1, \cdots, c_n\}$, depending on x, so that

$$x = c_1(x)e_1 + \cdots + c_n(x)e_n\,.$$

Clearly, $c_j(\cdot)$ is continuous linear functional on M. By Hahn-Banach extension theorem, it can be extended on X as a continuous linear functional. Then define $P : X \to X$ by

$$Px = \sum_{i=1}^{n} c_j(x)e_j\,,$$

and it's easy to check that P is a projection form X to M.

Case 2. Let $\dim X/M = n$, and

$$\{\widetilde{e}_1, \cdots, \widetilde{e}_n\}$$

a base for X/M. Pick any $e_j \in \widetilde{e}_j$, then

$$\{e_1, e_2, \cdots, e_n\}$$

is linear independent. Let

$$N = \operatorname{span}\{e_1, e_2, \cdots, e_n\}\,,$$

then $X = M \oplus N$, and N is closed. □

Schauder Bases. Let X be a separable Banach space and let $(e_j)_{j\geq 1}$ be a *Schauder basis* of X. Recall that this means that, for each element $x \in X$, there exists a *unique* sequence $(\alpha_j(x))$ of scalars such that

$$\lim_{n\to\infty} \left\| x - \sum_{i=1}^{n} \alpha_j(x) e_j \right\| = 0 \tag{4.11}$$

Let $n \geq 1$ and define the map $\Pi_n : X \to X$ by

$$\Pi_n(x) := \sum_{i=1}^{n} \alpha_j(x) e_j$$

for $x \in X$, where $(\alpha_j(x))$ is the unique sequence that satisfies (4.11). Clearly, the operators $\Pi_n : X \to X$ are linear and satisfy

$$\Pi_n \Pi_m = \Pi_m \Pi_n = \Pi_{n\wedge m}$$

for all integers $n, m \geq 1$. In particular, $\Pi_n^2 = \Pi_n$ for all $n \geq 1$. If $\{\Pi_n\}$ were projections, since $\Pi_n \to I$ in the strong operator topology, it follows from the PUB that $\{\Pi_n\}$ is uniformly bounded. Let's check if this is true.

Theorem 4.19. *Define a map $X \to [0, \infty) : x \mapsto \|x\|'$ by the formula*

$$\|x\|' := \sup_{n\geq 1} \|\Pi_n(x)\| \qquad \textit{for all } x \in X\,.$$

Then $(X, \|\cdot\|')$ is a Banach space with $\|x\|' \leq \|x\|$ for all $x \in X$. Hence by Theorem 4.11, there exists a constant $C > 0$ such that

$$\sup_{n\in\mathbb{N}} \|\Pi_n(x)\| \leq C\|x\| \qquad \textit{for all } x \in X\,.$$

In particular, $\{\Pi_n\}$ are projections.

Proof. It's easy to check that $\|\cdot\|'$ is a norm on X and

$$\|x\| = \lim_{n\to\infty} \|\Pi_n(x)\| \leq \sup_{n\geq 1} \|\Pi_n(x)\| = \|x\|'\,. \tag{4.12}$$

It suffices to show that $(X, \|\cdot\|')$ is complete. Let (x_k) be a Cauchy sequence in $(X, \|\cdot\|')$. Then (x_k) is a Cauchy sequence in $(X.\|\cdot\|)$ by (4.12). Suppose

$$\lim_{k\to\infty} \|x_k - x\| = 0\,. \tag{4.13}$$

We have only to show that

$$\lim_{k\to\infty} \|x_k - x\|' = \lim_{k\to\infty} \sup_{n\geq 1} \|\Pi_n(x_k) - \Pi_n(x)\| = 0\,.$$

Firstly, since (x_k) be a Cauchy sequence in $(X, \|\cdot\|')$, for any fixed n, $\{\Pi_n(x_k)\}$ is a Cauchy sequence in $(X, \|\cdot\|)$. Thus there exists a sequence $\{\xi_n\}$ in X with $\|\Pi_n(x_k) - \xi_n\| \to 0$ as $k \to \infty$ uniformly in n; i.e.,

$$\lim_{k\to\infty} \sup_{n\geq 1} \|\Pi_n(x_k) - \xi_n\| = 0\,. \tag{4.14}$$

So it's enough to show that $\xi_n = \Pi_n(x)$ for all n. Clearly $\xi_n \in \operatorname{span}\{e_1, \cdots, e_n\}$. Besides, since Π_m is continuous on $\operatorname{span}\{e_1, \cdots, e_n\}$ for $m \leq n$,

$$\Pi_m \xi_n = \lim_{k\to\infty} \Pi_m \Pi_n(x_k) = \lim_{k\to\infty} \Pi_m(x_k) = \xi_m\,.$$

It's sufficient to show that $\|\xi_n - x\| \to 0$ as $n \to \infty$. To this end, take any $\epsilon > 0$. Notice that for each $n \geq 1$ and each $k \geq 1$,

$$\|\xi_n - x\| \leq \|\xi_n - \Pi_n(x_k)\| + \|\Pi_n(x_k) - x_k\| + \|x_k - x\|\,. \tag{4.15}$$

Combine (4.14) and (4.13), there exists $k_0 = k_0(\epsilon)$ so that

$$\sup_{n\geq 1} \|\xi_n - \Pi_n(x_{k_0})\| \leq \epsilon \text{ and } \|x_{k_0} - x\| \leq \epsilon\,.$$

Recall that $\|\Pi_n(x_{k_0}) - x_{k_0}\| \to 0$ as $n \to \infty$. Taking $k = k_0(\epsilon)$ and letting $n \to \infty$, we get

$$\limsup_{n\to\infty} \|\xi_n - x\| \leq 2\epsilon\,.$$

Since $\epsilon > 0$ is arbitrary, the desired result follows. □

Corollary 4.20. *For each $j \geq 1$, $\alpha_j(\cdot) \in X^*$, where $(\alpha_j(x))$ is the unique sequence that satisfies* (4.11).

4.6 Holomorphic Functions

In the study of Banach algebras, as well as in some other contexts, it is useful to enlarge the concept of holomorphic function from complex-valued ones to vector-valued ones. (Of course, one can also generalize the domains, by going from $\mathbb{C}$ to $\mathbb{C}^n$ and even beyond. But this is another story.) The most important examples in spectral theory are operator valued holomorphic functions. There are at least two very natural definitions of "holomorphic" available in this general setting, a "weak" one and a "strong" one. They turn out to define the same class of functions if the values are assumed to lie in a Fréchet space.

Let Ω be an open set in $\mathbb{C}$ and let X be a complex topological vector space.

(a) A function $f : \Omega \to X$ is said to be *weakly holomorphic* in Ω if Λf is holomorphic in the ordinary sense for every $\Lambda \in X^*$.

(b) A function $f : \Omega \to X$ is said to be *strongly holomorphic* in Ω if

$$f'(z) := \lim_{w \to z} \frac{f(w) - f(z)}{w - z}$$

exists (in the topology of X) for every $z \in \Omega$. Note that the above quotient is the product of the scalar $(w - z)^{-1}$ and the vector $f(w) - f(z)$ in X.

The continuity of the functionals Λ that occur in (a) makes it obvious that every strongly holomorphic function is weakly holomorphic. The converse is true when X is a Fréchet space, but it is far from obvious. (Recall that weakly convergent sequences may very well fail to converge originally.) Nonetheless, in this note, we only deal with the case that X is a Banach space. The *Cauchy theorem* will play an important role in this proof, as will Corollary 4.4.

Theorem 4.21. *Let Ω be an open subset in $\mathbb{C}$. Let X be a complex Banach space. If $f : \Omega \to X$ is weakly holomorphic if and only if f is strong holomorphic.*

Proof. Take any $z_0 \in \Omega$. We shall prove that f is strongly continuous at z_0. Define

$$D(z_0, r) = \{z \in \mathbb{C} : |z - z_0| \le r\}\,.$$

Then $D(z_0, 2r) \subset \Omega$ for some $r > 0$. Let γ be the positively oriented boundary of $D(z_0, 2r)$. Fix $\Lambda \in X^*$. Since Λf is holomorphic, if $0 < |z - z_0| < 2r$,

$$(\Lambda f)(z) - (\Lambda f)(z_0) = \frac{1}{2\pi i}\int_\gamma \frac{(\Lambda f)(\zeta)}{\zeta - z} d\zeta - \frac{1}{2\pi i}\int_\gamma \frac{(\Lambda f)(\zeta)}{\zeta - z_0} d\zeta\,,$$

and hence

$$\frac{(\Lambda f)(z) - (\Lambda f)(z_0)}{z - z_0} = \frac{1}{2\pi i}\int_\gamma \frac{(\Lambda f)(\zeta)}{(\zeta - z)(\zeta - z_0)} d\zeta\,.$$

Intuitively, we guess

$$f'(z_0) = \frac{1}{2\pi i}\int_\gamma \frac{f(\zeta)}{(\zeta - z_0)^2} d\zeta =: y\,. \tag{4.16}$$

The integral is to be understood in the sense of Theorem 3.49. Either one can regard $d\zeta$ as a complex measure on the range of γ (a compact subset of $\mathbb{C}$), or one can parametrize γ and integrate with respect to Lebesgue measure on a compact interval in $\mathbb{R}$. To prove (4.16), note that

$$\begin{aligned}\Lambda\Big(\frac{f(z) - f(z_0)}{z - z_0} - y\Big) &= \frac{(\Lambda f)(z) - (\Lambda f)(z_0)}{z - z_0} - \frac{1}{2\pi i}\int_\gamma \frac{(\Lambda f)(\zeta)}{(\zeta - z_0)^2} d\zeta \\ &= \frac{z - z_0}{2\pi i}\int_\gamma \frac{(\Lambda f)(\zeta)}{(\zeta - z)(\zeta - z_0)^2} d\zeta\,.\end{aligned}$$

Thus

$$\Lambda\Big(\frac{f(z) - f(z_0) - y(z - z_0)}{(z - z_0)^2}\Big) = \frac{1}{2\pi i}\int_\gamma \frac{(\Lambda f)(\zeta)}{(\zeta - z)(\zeta - z_0)^2} d\zeta\,. \tag{4.17}$$

Let $M(\Lambda)$ be the maximum of $|\Lambda f|$ on $D(z_0, 2r)$. If $0 < |z - z_0| \le r$, it follows from (4.17) that

$$\left|\Lambda\Big(\frac{f(z) - f(z_0) - y(z - z_0)}{(z - z_0)^2}\Big)\right| \le \frac{M(\Lambda)}{r^2}\,.$$

The set of all quotients

$$\left\{\frac{f(z) - f(z_0) - y(z - z_0)}{(z - z_0)^2} : 0 < |z - z_0| \le r\right\}$$

is therefore weakly bounded in X. By Corollary 4.4, this set is also norm bounded. In other words, there exists $M > 0$ so that

$$\left\|\frac{f(z) - f(z_0)}{z - z_0} - y\right\| \le M|z - z_0| \ \ \textit{for all} \ \ 0 < |z - z_0| \le r\,.$$

Consequently, $f'(z_0)$ is well-defined and (4.16) holds. □

The next several theorems show that many of the familiar results in complex analysis carry over to the present setting.

Recall that a *chain* $\gamma = (\gamma_1, \cdots, \gamma_n)$ is a "formal sum" of finite many paths $\gamma_1, \cdots, \gamma_n$ in the plane. In other words, γ is a linear functional on $C(\Gamma)$, where $\Gamma = \cup_{j=1}^n \mathrm{Im}(\gamma_j)$, defined by

$$\int_\gamma f(z)dz = \sum_{j=1}^n \int_{\gamma_j} f(z)dz$$

for all $f \in C(\Gamma)$. If each γ_j is a path in some open set Ω, we say that γ is a chain in Ω. If each γ_j is a *closed path*, then γ is called a *cycle*.

Recall that the index of a point $z \in \mathbb{C}$ with respect to a cycle γ that does not pass through z will be denoted by $\mathrm{Ind}_\gamma(z)$. We recall that

$$\mathrm{Ind}_\gamma(z) = \frac{1}{2\pi i}\int_\gamma \frac{d\zeta}{\zeta - z}\,.$$

All paths considered here and later are assumed to be piecewise continuously differentiable, or at least rectifiable.

The Cauchy theorem and the Cauchy formula hold:

Theorem 4.22. *Let Ω be an open subset in $\mathbb{C}$. Let X be a complex Banach space. If $f : \Omega \to X$ is holomorphic. If γ is a cycle in Ω such that* $\operatorname{Ind}_\gamma(w) = 0$ *for every $w \notin \Omega$, then*

$$\int_\gamma f(\zeta)d\zeta = 0\,.$$

If γ_1 and γ_2 are cycles in Ω such that $\operatorname{Ind}_{\gamma_1}(w) = \operatorname{Ind}_{\gamma_2}(w)$ *for every $w \notin \Omega$, then*

$$\int_{\gamma_1} f(\zeta)d\zeta = \int_{\gamma_2} f(\zeta)d\zeta\,.$$

And if $z \in \Omega$ and $\operatorname{Ind}_\gamma(z) = 1$*, then*

$$f(z) = \frac{1}{2\pi i}\int_\gamma \frac{f(\zeta)}{\zeta - z}d\zeta\,.$$

Proof. Since f is holomorphic, f is continuous. Thus all the integrals above exists. These three formulas are correct (by the theory of ordinary holomorphic functions) if f is replaced in them by Λf, where Λ is any member of X^*. The formulas are therefore correct as stated, by Definition 3.4. □

The following extension of Liouville's theorem concerning bounded entire functions does not even depend on Theorem 3.31. It can be used in the study of spectra in Banach algebras.

Proposition 4.23 (Liouville)**.** *Let X be a complex Banach space. Suppose $f : \mathbb{C} \to X$ is weakly holomorphic and $f(\mathbb{C})$ is a weakly bounded subset of X. Then f is constant.*

Proof. For every $\Lambda \in X^*, \Lambda f$ is a bounded (complex-valued) entire function. If $z \in \mathbb{C}$, it follows from the Liouville theorem in complex analysis that

$$\Lambda f(z) = \Lambda f(0)\,.$$

Since X^* separates points on X, we have $f(z) = f(0)$, for every $z \in \mathbb{C}$. □

Chapter 5

Duality in Banach Spaces

5.1 Dual Spaces

Let X be a norm linear space. Recall that we denote by X^* the set of all continuous linear functionals on X. As a corollary of Proposition 1.5, X^* is Banach space. We call X^* the *dual* of X.

Remark 5.1. It should be emphasized that we did NOT assume X is complete. In fact, if $\bar{X}$ is its completion, then X^* and $\bar{X}^*$ are isometrically isomorphic.

Proposition 5.1. *X^* is isometrically isomorphic to a closed subspace of $C_b(B_X)$, where B_X is the closed unit ball in X.*

Proof. To see this, if $f \in X^*$, denote $f|_{B_X}$ as the restriction of f to B_X. Note that $\rho : X^* \to C_b(B_X)\,; f \mapsto f|_{B_X}$ is a linear isometry embedding. $\square$

Theorem 5.2. *X^* is separable, then X is also separable.*

Proof. Let $S = S_{X^*} = \{x^* : \|x^*\| = 1\}$ be the unit sphere in X^*. Then S is separable. Let $\{x_n^*\}$ be a countable dense subset of S.

Hence, for each $n \in \mathbb{N}$ there is an element $x_n \in X$ such that

$$\|x_n\| = 1 \quad \text{and} \quad |\langle x_n, x_n^* \rangle| > \frac{1}{2}\,.$$

We claim that $\overline{\text{span}}\{x_n\} = X$. By Corollary 1.34, it suffices to show that $\ell \in X^*$ vanishing at $\{x_n\}$ implies $\ell = 0$. If not, without loss of generality we assume that $\|\ell\| = 1$. Then there exists x_k^* with $\|x_k^* - \ell\| < 1/2$. Hence $|\ell(x_k) - x_k^*(x_k)| < 1/2$. But $\ell(x_k) = 0$, so $|x_k^*(x_k)| < 1/2$, which is a contradiction! $\square$

The converse of Theorem 5.2 is not true.

Example 5.1. ℓ^1 is separable but $(\ell^1)^* = \ell^\infty$(see Example 5.5) is not. To see this, note that any $x \in (0,1)$ has a *binary representation*, denote by $\{b_n(x)\}$, or in other words

$$0.b_1(x)b_2(x)\cdots$$

where $b_n(x) \in \{0,1\}$. Then $\{b_n(x)\} \in \ell^\infty$ and for $n \neq m$,

$$\|b_n(x) - b_m(x)\|_\infty = 1\,.$$

Thus ℓ^∞ is not separable.

5.1.1 Representation of Dual Spaces

Let X and X' be two metric spaces. Recall that a map $M : X \to X'$ is called an *isometry* if one has

$$d(Mx, My) = d(x,y) \quad \text{for any } x, y \in X\,.$$

Particularly, if X, X' are normed linear space, and the distances are induced by norm, then M is *isometry* if and only if

$$\|Mx\| = \|x\| \quad \text{for any } x \in X\,.$$

Obviously, an isometry is automatically injective, and it can not be surjective.

Example 5.2. Definite $S : \ell^2 \to \ell^2$ by $S(\alpha_1, \alpha_2, \ldots) = (0, \alpha_1, \alpha_2, \ldots)$. Then S is an isometry that is not surjective.

Two linear spaces X and Y over the same field $\mathbb{F}$ are said to be *isomorphic* if there is a bijective linear operator $T \in \mathcal{L}(X, Y)$. If in addition, T is an isometry, then we say that T is an *isometry isomorphism*. Clearly, T is an isometry isomorphism if and only if T is bijective and

$$\|T\| = \|T^{-1}\| = 1\,.$$

In this case, X and Y are said to be *isometrically isomorphic* and we write $X \cong Y$. By abuse of the notation, sometime we write $X = Y$ for short.

Example 5.3. The dual space of c_0 is ℓ_1, i.e., $c_0^* \cong \ell_1$.

To prove this, let $w = (w_n) \in \ell_1$ and define $\Phi : \ell_1 \to c_0^*$ by

$$\langle \Phi w, x\rangle = \sum_{n=1}^{\infty} x_n w_n\,, \quad \text{for any } x = (x_n) \in c_0$$

It's easy to show that Φw is a bounded linear functional on c_0 and

$$\|\Phi w\| = \|w\|_1.$$

To show that Φ is a surjective, consider (e_n), the Schauder basis for c_0, where $e_n = (\delta_{nm})$ has 1 in the n-th position and zeroes elsewhere. Let $f \in c_0^*$ and $x = (x_n) \in c_0$. Then $x = \sum_{n=1}^{\infty} x_n e_n$ and therefore

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

Take any $k \in \mathbb{N}$, let

$$x_n^{(k)} = \begin{cases} |f(e_n)| / f(e_n)\,, & n \leq k \text{ and } f(e_n) \neq 0\,. \\ 0\,, & \text{otherwise}\,. \end{cases}$$

Then $x^{(k)} = (x_n^{(k)}) \in c_0$ and $\|x^{(k)}\| = 1$. So $f(x^{(k)}) = \sum_{n=1}^{k} |f(e_n)| \leq \|f\|$. So we have $\sum_{n=1}^{\infty} |f(e_n)| < \infty$, $(f(e_n)) \in \ell_1$. Therefore, $\Phi(f(e_n)) = f$ and Φ is surjective.

Example 5.4. $L^p(\mu)^* \cong L^q(\mu)$. Specifically, (X, Ω, μ) is a measure space and $p \in (1, \infty)$, $1/p + 1/q = 1$. For any $g \in L^q(X, \Omega, \mu)$, define $\ell_g : L^p(\mu) \to \mathbb{F}$ by

$$\ell_g(f) = \int fg d\mu$$

Then $\ell_g \in L^p(\mu)^*$ and the map $g \mapsto \ell_g$ defines an isometric isomorphism of $L^q(\mu)$ onto $L^p(\mu)^*$.

This theorem and the next have been proved in courses in measure and integration, we omit the proof.

Example 5.5. $L^1(\mu)^* \cong L^\infty(\mu)$, where μ is a σ-finite measure. Specifically, (X, Ω, μ) is a σ-finite measure space. For any $g \in L^\infty(X, \Omega, \mu)$, define $\ell_g : L^1(\mu) \to \mathbb{F}$ by

$$\ell_g(f) = \int fg d\mu \ .$$

Then $\ell_g \in L^1(\mu)^*$ and the map $g \mapsto \ell_g$ defines an isometric isomorphism of $L^\infty(\mu)$ onto $L^1(\mu)^*$.

Example 5.6 (Riesz–Kakutani Representation Theorem)**.** $C_0(X)^* \cong \mathcal{M}(X)$, where X is a locally compact Hausdorff (LCH) space. Specifically, $\mathcal{M}(X)$ denotes the space of all $\mathbb{F}$-valued Radon measures on X with the total variation norm. For any $\mu \in \mathcal{M}(X)$, define $F_\mu : C_0(X) \to \mathbb{F}$ by

$$\ell_\mu(f) = \int f d\mu \ .$$

Then $\ell_\mu \in C_0(X)^*$ and the map $\mu \to \ell_\mu$ is an isometric isomorphism of $\mathcal{M}(X)$ onto $C_0(X)^*$.

There are special cases of these theorems that deserve to be pointed out

- $(l^p)^* \cong l^q$, $(l^1)^* \cong l^\infty$.
- $c_0^* \cong l^1$. In fact, $c_0 = C_0(\mathbb{N})$, where $\mathbb{N}$ is given the discrete topology, and $l^1 = M(\mathbb{N})$.

Exercise 5.1. Let X be a Banach space. Let $n \geq 1$. Then ℓ is a continuous linear functional on $\oplus_{p_{k=1}}^{n} X$ if and only if there are $x_1^*, \ldots, x_n^*$ in X^* so that

$$\ell(x) = \sum_{k=1}^{n} \langle x_k, x_k^* \rangle \text{ for all } x = (x_1, \cdots, x_n) .$$

5.1.2 Dual Spaces of Subspaces and Quotients

Let X be a normed linear space. Let M be a closed subspace in X. If $f \in X^*$, then $f|_M$, the restriction of f to M, belongs to M^* and $\|f|_M\| \leq \|f\|$. According to the Hahn-Banach theorem, each bounded linear functional on M is obtainable as the restriction of a functional from X^*. Indeed, more can be said.

Let now M be a subset of X. The *annihilator* of M, denoted by $M^\perp$, is given by

$$M^\perp := \{x^* \in X^* : \langle x, x^* \rangle = 0 \text{ for all } x \in M\} .$$

Let N be any subset of X^*, then the *annihilator* of N, denoted by ${}^\perp N$, is given by

$${}^\perp N := \{x \in X : \langle x, x^* \rangle = 0 \text{ for all } x^* \in N\} .$$

It is easy to observe that $M^\perp$ and ${}^\perp N$ are closed subspace of X^* and X.

Lemma 5.3. *M is a subspace of X, N is a subspace of X^*, then*

$${}^\perp(M^\perp) = \overline{M} , \; ({}^\perp N)^\perp = \overline{N}^{\sigma(X^*,X)} . \tag{5.1}$$

Proof. Use Hahn-Banach theorem, as in the proof of Theorem 3.37. ☐

Theorem 5.4. *Let M be a closed linear subspace of X. Then $M^* \cong X^*/M^\perp$. Moreover, the map $\rho : X^*/M^\perp \to M^*$ defined by*

$$\rho : \tilde{f} \mapsto f|_M$$

is an isometric isomorphism.

Proof. Clear ρ is a well-defined linear operator. It follows from the Hahn-Banach theorem that ρ is surjective. So we have only to show

$$\|\tilde{f}\| = \|f|_M\| \quad \text{for all } \tilde{f} \in X^*/M^\perp .$$

To this end, notice that $\|f|_M\| = \|g|_M\| \le \|g\|$ for all $g \in \tilde{f}$. Besides, by Hahn-Banach theorem, there exists $g \in \tilde{f}$ so that the equality holds. Then the desired result follows. $\square$

Theorem 5.5. *Let M be a closed linear subspace of X. Let $Q : X \to X/M$ be the natural map. Then the map $\rho : (X/M)^* \to M^\perp$ defined by*

$$\rho : \tilde{f} \mapsto \tilde{f} \circ Q$$

is an isometric isomorphism.

Proof. Clear ρ is a well-defined linear operator. We show that ρ is surjective. For any $\ell \in M^\perp$, define $\tilde{\ell} : X/M \to \mathbb{F}$ by

$$\tilde{\ell}\tilde{x} := \ell x \quad \text{for } \tilde{x} \in X/M .$$

By Example 1.20, $\ell \in (X/M)^*$ and $\ell = \tilde{\ell} \circ Q$. Hence, ρ is surjective. Now, it suffices to show that

$$\|\tilde{f}\| = \|\tilde{f} \circ Q\| .$$

Since $\|Q\| \le 1$, we have $\|\tilde{f} \circ Q\| \le \|f\|$. On the other hand, take any $\lambda \in (0, 1)$. Then we can find $\tilde{x} \in X/M$ with $\|\tilde{x}\| = 1$, satisfying $|\tilde{f}\tilde{x}| \ge \lambda\|\tilde{f}\|$. Take $m \in M$ so that $\|x + m\| \le 1/\lambda$, then

$$\lambda\|\tilde{f}\| \le |\tilde{f}\tilde{x}| = |(\tilde{f} \circ Q)(x + m)| \le \|\tilde{f} \circ Q\|\frac{1}{\lambda} .$$

Letting $\lambda \uparrow 1$, the desired result follows. $\square$

5.2 Bidual Space and Reflexivity

Definition 5.1. The dual space of $(X^*, \|\cdot\|)$ is called the *second dual space* or *bidual space* of X, and denote as X^{**}.

Let $(X, \|\cdot\|)$ be a normed linear space over $\mathbb{F}$. For any fixed $x \in X$, define a functional $Jx : X^* \to \mathbb{F}$ by

$$\langle x^*, Jx\rangle = \langle x, x^*\rangle \quad \text{for all } x^* \in X^* .$$

It is easy to verify that $Jx \in X^{**}$, and $|Jx| = \|x\|$. It now follows that we can define a map

$$J : X \to X^{**},\ x \mapsto Jx. \tag{5.2}$$

Obviously, J_X is linear, and hence J_X is a linear isometry of X *into* its bidual X^{**}, which is called the *canonical* or *natural embedding* of X into its bidual X^{**}. This shows that we can identify X with the subspace JX of X^{**}.

Definition 5.2. $(X, \|\cdot\|)$ is said to be *reflexive* if the canonical embedding is surjective, i.e., $JX = X^{**}$.

Remark 5.2. (a) X is reflexive implies that the canonical embedding J is a isometrical isomorphism to X^{**}, hence X is a Banach space.

(b) Banach space X that is isometriccally isomorphic to X^{**} may be NOT reflexive. See *R.C. James [1951]. A non-reflexive Banach space isometric with its second conjugate space. Proc. Nat. Acad. Sci. USA, 37, 174-177.*

Example 5.7. Every finite-dimensional normed linear space is reflexive.

Example 5.8. For $1 < p < \infty$, $L^p(\mu)$ is reflexive. See Example 5.4.

Example 5.9. c_0, and $C[0,1]$ both are non-reflexive.

(a) $c_0^* = l_1$, so $c_0^{**} = (l_1)^* = l_\infty$. With these identifications, the natural map $c_0 \to c_0^{**}$ is precisely the inclusion map $c_0 \to l_\infty$.

(b) Note that $C[0,1]$ is separable: every continuous function an be approximate by piecewise linear functions with rational nodes and rational ordinates. On the other hand, $C[0,1]^*$ is not separable; the linear functionals ℓ_s defined by

$$\ell_s(f) = f(s), \quad -1 \leq s \leq 1$$

are clearly each bounded by 1, and equally clearly

$$|\ell_s - \ell_t| = 2 \qquad \text{for } s \neq t$$

since the $\{\ell_s\}$ form a nondenumerable collection, $C[0,1]^*$ cannot contain a dense denumerable subset. It follows now that $C[0,1]^{**} \neq C[0,1]$.

Lemma 5.6. *Let M be a closed subspace of X and $J_X : X \to X^{**}$ and $J_M : M \to M^{**}$ be the natural maps. Let $i : M \to X$ is the inclusion map, then there exists a linear isometry embedding $\Phi : M^{**} \to X^{**}$ such that the following diagram commutes.*

$$\begin{tikzcd} X \arrow[r, "J_X"] & X^{**} \\ M \arrow[u, "i"] \arrow[r, "J_M"] & M^{**} \arrow[u, "\Phi"'] \end{tikzcd}$$

Proof. For any $y^{**} \in M^{**}$, define

$$\langle x^*, \Phi y^{**} \rangle = \langle x^*|_M, y^{**} \rangle\,, \quad \text{for any } x^* \in X^*\,.$$

Then it's easy to check that Φ is a linear isometry. Then for any $y \in M$ and $x^* \in X^*$,

$$\langle x^*, \Phi(J_M y) \rangle = \langle x^*|_M, J_M y \rangle = \langle y, x^*|_M \rangle = \langle y, x^* \rangle = \langle x^*, J_X y \rangle\,.$$

So the diagram commutes. $\square$

Theorem 5.7. *A closed linear M of a reflexive space X is reflexive.*

Proof. Using the the lemma above, for any $y^{**} \in M^{**}$, there is a $y \in X$ such that $J_X y = \Phi y^{**}$. We only need to show that $y \in M$.

If not, there is $x^* \in M^\perp$, $\|x^*\| = 1$ and $\langle y, x^* \rangle > 0$. But then $x^*|_M = 0$ so that

$$\langle y, x^* \rangle = \langle x^*, J_X y \rangle = \langle x^*, \Phi y^{**} \rangle = \langle x^*|_M, y^{**} \rangle = 0\,.$$

This is a contradiction. $\square$

Theorem 5.8. *A Banach space X is reflexive iff its dual X^* is reflexive.*

Proof. Assume that X is reflexive. Let $J_X : X \to X^{**}$ and $J_{X^*} : X^* \to X^{***}$ be the canonical embeddings of X and X^* respectively. We must show that J_{X^*} is surjective. To that end, let $\xi \in X^{***}$ and consider the following communicate diagram to define a functional x^* on X by $x^* = \xi \circ J_X$.

$$\begin{tikzcd} X \arrow[r, "J_X"] \arrow[dr, "x^*"'] & X^{**} \arrow[d, "\xi"] \\ & \mathbb{F} \end{tikzcd}$$

It is obvious that x^* is linear since both ξ and J_X are linear. Also, for each $x \in X$,

$$|\langle x, x^* \rangle| = |\langle J_X x, \xi \rangle| \leq \|\xi\| \|x\| \,,$$

so that $x^* \in X^*$. We know show that $J_{X^*} x^* = \xi$. In fact, for any $J_X x \in X^{**}$,

$$\langle J_X x, J_{X^*} x^*, \rangle = \langle x^*, J_X x \rangle = \langle x, x^* \rangle = \langle J_X x, \xi \rangle \,.$$

Suppose now that X^* is reflexive. Then the canonical embedoing J_{X^*} : $X^* \to X^{***}$ is surjective. If $J_X X \neq X^{**}$, take $x^{**} \in X^{**} \backslash J_X X$. Since X is Banach space, $J_X X$ is a closed subspace of X^{**}, it follows from Corollary 1.34 that there is a functional $J_{X^*} x^* \in X^{***}$ such that $\|x^*\| = 1$, and $\langle x^{**}, J_{X^*} x^* \rangle > 0$ and

$$\langle J_X x, J_{X^*} x^* \rangle = 0 \;\; \text{for all } x \in X \,.$$

Hence, for each $x \in X$.

$$0 = \langle J_X x, J_{X^*} x^* \rangle = \langle x^* . J_X x \rangle = \langle x, x^* \rangle$$

Thus $x^* = 0$, which is a contradiction. □

5.3 Weak and Weak-star Topologies

We have made the point that a norm on a vector space X induces a metric. A metric, in turn, induces a topology on X called the norm topology, and X^* also has a natural norm topology. In this section we investigate some of the properties of weak topology, defined in Section 3.7, on normed space X and weak-star topology on X^*.

Proposition 5.9. *Let $(X, \|\cdot\|)$ be a normed space and τ the norm topology.*

(a) $\sigma(X, X^*) \subset \tau$

(b) $\sigma(X, X^*) = \tau$ *if and only if X is finite-dimensional.*

Proof. (a) is obvious.

To show (b), when X is finite-dimensional, we have proved that all the vector topology on X are the same one, see Corollary 3.19, so $\tau_w = \tau$.

Conversely, assume that X is infinite-dimensional. The (open) unit ball

$$B(0,1) = \{x : \|x\| < 1\},$$

is open in τ, but we claim that $B(0,1) \notin \tau_w$. If not, there exists a finite subset Φ of X^* and $\epsilon > 0$ so that

$$V(\Phi, \epsilon) \subset B(0,1).$$

Thus,

$$\bigcap_{x^* \in \Phi} \ker(x^*) \subset B(0,1).$$

Since X is infinite-dimensional, $\cap_{x^* \in \Phi} \ker(x^*) \neq \{0\}$ is a subspace of X, this is a contradiction. ☐

Note that there are three topologies on X^*: the norm topology, the weak topology, denoted by $\sigma(X^*, X^{**})$, and the weak-star topology, denoted by $\sigma(X^*, X)$.

Proposition 5.10. *Let τ^* be the norm topology on X^*. Then*

(a) $\sigma(X^*, X) \subset \sigma(X^*, X^{**}) \subset \tau^*$

(b) $\sigma(X^*, X^{**}) = \tau^*$ *if and only if X is finite-dimensional.*

(c) $\sigma(X^*, X) = \sigma(X^*, X^{**})$ *if and only if X is reflexive.*

Proof. (a), (b) is obvious.

To show (c), “ if ” part is obvious. On the other hand, note that

$$(X^*, \sigma(X^*, X))^* = J(X) \text{ and, } (X^*, \sigma(X^*, X^{**}))^* = X^{**}.$$

Therefore if $\sigma(X^*, X) = \sigma(X^*, X^{**})$, then $J(X) = X^{**}$, X is reflexive. ☐

Theorem 5.11. *Let X be a normed space. Then the following assertions hold.*

(a) For any nonempty convex subset K of X. Then the weak closure of K is equal to its norm closure:

$$\overline{K}^{\sigma(X,X^*)} = \overline{K}\,.$$

(b) If in addition X is reflexive, For any nonempty convex subset K of X^*. Then the weak-star closure of K is equal to its (operator) norm closure:

$$\overline{K}^{\sigma(X^*,X)} = \overline{K}\,.$$

Proof. Part (i) follows trivially from Theorem 3.34. To show part (ii), we use the same method as the proof of Theorem 3.34. It suffices to show that

$$\overline{K}^{\sigma(X^*,X)} \subset \overline{K}\,.$$

Suppose for contradiction that there exists $x_0^* \in X^*$, and

$$x_0^* \in \overline{K}^{\sigma(X^*,X)} \backslash \overline{K}$$

By Hahn-Banach separation theorem and the reflexivity of X, there exist $x \in X$, $\gamma \in \mathbb{R}$ and $\epsilon > 0$ such that, for every $x^* \in K$,

$$\mathrm{Re}\langle x, x_0^* \rangle < \gamma < \gamma + \epsilon < \mathrm{Re}\langle x, x^* \rangle\,.$$

Thus $|\langle x, x_0^* - x^* \rangle| \geq \epsilon$ for all $k \in K$, i.e.,

$$K \cap V(x_0^*; x, \epsilon) = \emptyset\,.$$

Thus $x_0^* \notin \overline{K}^{\sigma(X^*,X)}$, this is a contradiction. $\square$

5.3.1 Convergence

In Example 3.1, we have pointed that

- a net $\{x_i\}$ in X is convergent to some $x \in X$ in $\sigma(X, X^*)$ if and only if
$$\langle x_i, x^* \rangle \to \langle x, x^* \rangle \text{ , for all } x^* \in X^* \text{ ;}$$
- a net $\{x_i^*\}$ in X^* is convergent to some $x^* \in X$ in $\sigma(X^*, X)$ if and only if
$$\langle x, x_i^* \rangle \to \langle x, x^* \rangle \text{ , for all } x \in X \text{ .}$$

However, in this subsection, we care more about convergent sequence. To avoid any confusion we shall sometimes say, $x_n \to x$ in $\sigma(X, X^*)$, or $x_n^* \to x^*$ in $\sigma(X, X^*)$. In order to be totally clear we sometimes emphasize norm convergence by saying, "$x_n \to x$ in norm," meaning that $\|x_n - x\| \to 0$.

Example 5.10. X is finite-dimensional, the weak topology $\sigma(X, X^*)$ and the usual topology are the same. In particular, a sequence (x_n) converges weakly if and only if it converges strongly.

Example 5.11. If a sequence in l^1 converges weakly, it converges in norm (Proposition 5.19). Note that this result demonstrates in a dramatic way that in discussions concerning the weak topology it is essential to consider nets and not just sequences.

Example 5.12. In $L^2[0, 2\pi]$, let $x_n = x_n(t) = \sin nt$. By Riemann-Lebesgue theorem, for any $x \in L^2[0, 2\pi]$,

$$\langle x, x_n \rangle = \int_0^{2\pi} x(t) sin(nt)\, dt \to 0$$

This is $x_n \xrightarrow{w} 0$, but $\|x_n\| = \pi$ for all n, so x_n doesn't converge to 0 in norm.

Generally, H is a Hilbert space and $\{e_n\}$ is a orthonormal system, by Bessel's inequality we know $e_n \xrightarrow{w} 0$ but $\|e_n\| = 1$ for all n.

Proposition 5.12. *Let $\{x_n\}$ be a sequence in the normed linear space X.*

(a) If $x_n \to x$ weakly in $\sigma(X, X^*)$, then $\{x_n\}$ is norm bounded and,

$$\|x\| \leq \liminf_{n\to\infty} \|x_n\| . \tag{5.3}$$

(b) If $x_n \to x$ weakly in $\sigma(X, X^*)$ and if $x_n^* \to x^*$ in norm, then

$$\langle x_n, x_n^* \rangle \to \langle x, x^* \rangle . \tag{5.4}$$

Proof. To show (a), by PUB, we have that $\{x_n\}$ is bounded, and note that

$$|\langle x, x^* \rangle| = \lim_{n\to\infty} |\langle x_n, x^* \rangle| \leq \liminf_{n\to\infty} \|x_n\| \|x^*\|$$

for any $x^* \in X^*$, Thus

$$\|x\| \leq \liminf_{n\to\infty} \|x_n\| .$$

To show (b), note that

$$\begin{aligned} |\langle x_n, x_n^* \rangle - \langle x, x^* \rangle| &\leq |\langle x_n, x_n^* \rangle - \langle x_n, x^* \rangle| + |\langle x_n, x^* \rangle - \langle x, x^* \rangle| \\ &\leq \|x_n^* - x^*\| \sup_n \|x_n\| + |\langle x_n, x^* \rangle - \langle x, x^* \rangle| . \end{aligned}$$

We now complete the proof. □

By using Banach-Steinhaus theorem, we get the following result, which is useful when we discussing sequential compactness.

Theorem 5.13. *Let X be a normed linear space. A sequence $\{x_n\}$ in X is weakly convergent if and only if*

(a) $\{x_n\}$ is norm bounded and,

(b) there exists a dense subset N of X^* so that $\{\langle x_n, x^* \rangle\}$ converges for each $x^* \in N$.

Theorem 5.14. *Let X be a Banach space. A sequence $\{x_n^*\}$ in X^* is convergent relative to the weak-star topology if and only if*

(a) $\{x_n^*\}$ is uniformly (norm) bounded and,

(b) there exists a dense subset M of X so that $\{\langle x, x_n^* \rangle\}$ converges for each $x \in M$.

5.3.2 Compact and Sequentially Compact Sets in Weak-star Topology

Observe that $X^* \subset \mathbb{F}^X = \prod_X \mathbb{F}$ and that the weak-star topology $\sigma(X^*, X)$ on X^* is the relative topology on X^* induced by the product topology on $\prod_X \mathbb{F}$.

Theorem 5.15 (Banach-Alaoglu Theorem)**.** *Let $(X, \|\cdot\|)$ be a normed vector space. Then the closed unit ball in X^* is weak-star compact, i.e.,*

$$B_{X^*} = \{x^* \in X^* : \|x^*\| \leq 1\}$$

is compact with respect to the topology $\sigma(X^*, X)$.

Proof. *Step 1.* For each $x \in X$, let

$$D_x = \{\lambda \in \mathbb{F} : |\lambda| \leq \|x\|\}.$$

Then, for each $x \in X, D_x$ is a closed interval in $\mathbb{R}$ or a closed disk in $\mathbb{C}$. Equipped with the natural topology, D_x is compact for each $x \in X$. Let

$$D = \prod_{x \in X} D_x \,.$$

By Tychonoff's theorem, D is compact with respect to the product topology. The points of D are just $\mathbb{F}$-valued functions (not to be linear) f on X such that $|f(x)| \leq \|x\|$ for each $x \in X$. Obviously, we have

$$B_{X^*} \subset D \,.$$

Step 2. We observe that the topology that D induces on $B(X^*)$ is precisely the weak-star topology on B_{X^*}. It remains to show that B_{X^*} is a closed subset of D. To this end, let $\{x_i^*\}$ be a net in B_{X^*} and $x_i^* \to x^* \in D$ in the product topology. Then $\langle x, x_i^* \rangle \to \langle x, x^* \rangle$ for all $x \in X$. Thus

$$\begin{aligned} &\langle \alpha x + \beta y, x^* \rangle = \lim_i \langle \alpha x + \beta y, x_i^* \rangle \\ &= \lim_i \alpha \langle x, x_i^* \rangle + \beta \langle y, x_i^* \rangle = \alpha \langle x, x^* \rangle + \beta \langle x, x^* \rangle \,. \end{aligned}$$

for all x, y in X and α, β in $\mathbb{F}$, thus x^* is linear. Since

$$|\langle x, x^* \rangle| = \lim_i |\langle x, x_i^* \rangle| \leq \|x\|$$

for all $x \in X$, x^* is continuous, and $\|x^*\| \leq 1$. That is,

$$x^* \in B_{X^*}.$$

Therefore B_{X^*} is closed in D and hence compact. $\square$

Theorem 5.16 (Sequential Compactness)**.** *Let X be a separable Banach space. Then B_{X^*} is weak-star sequentially compact.*

Proof. Assume $\{x_n\}$ is dense in X. Given any sequence $\{x_n^*\}$ in B_{X^*}, by *diagonal process*, we can select a subsequence $\{y_n^*\}$ of $\{x_n^*\}$ so that

$$\langle x_k, y_n \rangle \text{ converges for each } x_k\,.$$

By Theorem 5.14, $\{y_n^*\}$ is converges in weak-star topology. $\square$

Separability and Metrizability. In fact, for a separable Banach space X, the closed unit ball B_{X^*} is metrizable in the weak-star topology. So the weak-star compactness of B_{X^*} implies it's sequential compactness.

We should emphasize that the weak and weak-star topologies on an infinite dimensional Banach space are never metrizable. It is possible, however, to show that under certain conditions these topologies are metrizable when restricted to bounded sets. In applications this is often sufficient.

The next lemma is rather easy, but it will be used so often that it should be explicitly stated and proved.

Lemma 5.17. *Let X, Y be two Hausdorff topological spaces and X is compact. If $f : X \to Y$ is bijective and continuous, then f is a homeomorphism.*

Proof. If F is a closed subset of X, then F is compact. Thus $f(F)$ is compact in Y and hence closed. Since f maps closed sets to closed sets, f^{-1} is continuous. So f is a homeomorphism. $\square$

Theorem 5.18. *If X is a Banach space, then closed unit ball B_{X^*} is weak-star metrizable if and only if X is separable.*

Proof. Assume that X is separable and let $\{x_n\}$ be a countable dense subset of B_{X^*}. For each $n \geq 1$, let $D_n = \{\alpha \in \mathbb{F} : |\alpha| \leq 1\}$. Put $Y = \prod_{n=1}^{\infty} D_n$. Then Y is a compact metric space. So if $(B_{X^*}, \sigma(X^*, X))$ is homeomorphic to a subset of Y, B_{X^*} is weak-star metrizable.

Define $\varphi : B_{X^*} \to X$ by $\varphi(x^*) = \{\langle x_n, x^* \rangle\}$. We show that φ is a homeomorphism. If $\{x_i^*\}$ is a net in B_{X^*} and $x_i^* \to x^*$ in $\sigma(X^*, X)$, then $\langle x_n, x_i^* \rangle \to \langle x_n, x^* \rangle$ for each $n \geq 1$, hence $\varphi(x_i^*) \to \varphi(x^*)$ and φ is continuous. If $\varphi(x^*) = \varphi(y^*)$, $\langle x_n, x^* - y^* \rangle = 0$ for all n, since $\{x_n\}$ is dense, $x^* - y^* = 0$. Thus φ is injective. Since x^* is wk * compact, φ is a homeomorphism onto its image (Lemma 5.17) and B_{X^*} is weak-star metrizable.

Now assume that $(B_{X^*}, \sigma(X^*, X))$ is metrizable. Thus there are open sets $\{U_n\}$ in $(B_{X^*}, \sigma(X^*, X))$ such that $0 \in U_n$ and $\cap_n U_n = \{0\}$. By the definition of the relative weak-star topology on B_{X^*}, for each n there is a finite set F_n contained in X such that

$$\{x^* \in B_{X^*} : |\langle x, x^* \rangle| < 1 \text{ for all } x \text{ in } F_n\} \subset U_n.$$

Let $F = \bigcup_{n=1}^{\infty} F_n$; so F is countable. We claim that $\overline{\operatorname{span}}\{F\} = X$, and hence X is separable. To show this, by Lemma 1.34 it suffices to show that

$$F^{\perp} = \{0\}.$$

If $x^* \in F^{\perp}$, then for each $n \geq 1$ and for each x in F_n, $|\langle x, x^*/\|x^*\| \rangle| = 0 < 1$. Hence $x^*/\|x^*\| \in U_n$ for all $n \geq 1$; thus $x^* = 0$. We now complete the proof. $\square$

Is there a corresponding result for the weak topology? If X^* is separable, then the weak topology on ball X is metrizable. In fact, this follows from Theorem 5.18 if the embedding of X into X^{**} is considered. This result is not very useful since there are few examples of Banach spaces X such that X^* is separable. Of course if X is separable and reflexive, then X^* is separable, but in this case the weak topology on X^* is the same as its weak-star topology when X is identified with X^{**}. Thus Theorem 5.18 is adequate for a discussion of the weak topology on the unit ball of a separable reflexive space. If $X = c_0$, then $X = l^1$ and this is separable but not reflexive. This is one of the few nonreflexive spaces with a separable dual space.

If X is separable, is $(B_X, \sigma(X, X^*))$ metrizable? The answer is no, as the following result of Schur demonstrates. So if $(B(\ell^1), \sigma(\ell^1.\ell^\infty))$ were metrizable, the following proposition would say that the weak and norm topologies on l^1 agree. But this is not the case.

Proposition 5.19. *If a sequence in ℓ^1 converges weakly, it converges in norm.*

Proof. Recall that $\left(l^1\right)^* = l^\infty$. Since l^1 is separable, Theorem 5.18 implies that $B(l^\infty)$ is weak-star metrizable. Indeed, one can check an equivalent metric on $(B(l^\infty), \sigma(\ell^\infty, \ell^1))$ is given by

$$d(\phi, \psi) = \sum_{j=1}^{\infty} \frac{|\phi(j) - \psi(j)|}{2^j} \qquad \text{for } \phi, \psi \in \ell^\infty \,.$$

By Alaoglu's theorem, $B(l^\infty)$ is weak-star compact. Hence $(B(l^\infty), \sigma(\ell^\infty, \ell^1))$ is a complete metric space and the Baire category theorem is applicable.

Let $\{f_n\}$ be a sequence of elements in l^1 such that $f_n \to 0$ weakly. Let $\epsilon > 0$, and for each positive integer m let

$$F_m = \{\phi \in B(l^\infty) : |\langle f_n, \phi\rangle| \le \epsilon \text{ for } n \ge m\} \,.$$

It is easy to see that F_m is weak-star closed in $B(l^\infty)$ and, because $f_n \to 0$ weakly,

$$\bigcup_{m=1}^{\infty} F_m = B(l^\infty) \,.$$

By the theorem of Baire, there exists $m = m_\epsilon$ so that F_m has non-empty weak-star interior. Thus there is a ϕ in F_m and a $\delta = \delta_\epsilon > 0$ such that

$\{\psi \in B(l^\infty) : d(\phi, \psi) < \delta\} \subset F_m$. Given any fixed $n \geq m$, define ψ in l^∞ by

$$\psi(j) = \begin{cases} \phi(j)\,, \text{ for } 1 \leq j \leq J\,; \\ \operatorname{sign}(f_n(j))\,, \text{ for } j > J\,. \end{cases}$$

where J is a large integer that we determine it later. Thus $\psi(j)f_n(j) = |f_n(j)|$ for $j > J$. It is easy to see that $\psi \in B(l^\infty)$. Also,

$$d(\phi, \psi) = \sum_{j>J} 2^{-j}|\phi(j) - \psi(j)| \leq 2 \cdot 2^{-J} < \delta\,.$$

for $J = J_\epsilon$. So $\psi \in F_m$ and hence $|\langle \psi, f_n \rangle| \leq \epsilon$, i.e.,

$$\left| \sum_{j \leq J} \phi(j) f_n(j) + \sum_{j>J} |f_n(j)| \right| \leq \epsilon\,. \tag{5.5}$$

Since $f_n \to 0$ weakly, there is an $m' = m'_\epsilon$ such that for $n \geq m'$,

$$\sum_{j \leq J} |f_n(j)| < \epsilon\,.$$

Combining this with (5.5) gives that for $n \geq \max\{m_\epsilon, m'_\epsilon\}$,

$$\begin{aligned} \|f_n\| &= \sum_{j \leq J} |f_n(j)| + \sum_{j>J} |f_n(j)| \\ &< \epsilon + \left| \sum_{j>J} |f_n(j)| + \sum_{j \leq J} \phi(j) f_n(j) - \sum_{j \leq J} \phi(j) f_n(j) \right| \\ &< \epsilon + \left| \sum_{j>J} |f_n(j)| + \sum_{j \leq J} \phi(j) f_n(j) \right| + \sum_{j \leq J} |\phi(j) f_n(j)| \leq 3\epsilon\,. \end{aligned}$$

So $\|f_n\| \to 0$. The desired result follows. $\square$

5.3.3 Compact and Sequentially Compact Sets in Weak Topology : Reflexivity Revisited

Theorem 5.20 (Goldstine)**.** *Let $(X, \|\cdot\|)$ be a normed linear space. Let J be the natural embedding of X into X^{**}. Let B_X and $B_{X^{**}}$ be the closed* *unit ball in X and X^{**}, respectively. Then JB_X is dense in $B_{X^{**}}$ relative to the weak-star topology $\sigma(X^{**}, X^*)$ on X^{**}, that is,*

$$\overline{JB_X}^{\sigma(X^{**},X^*)} = B_{X^{**}}\,.$$

Proof. Clearly, $JB_X \subset B_{X^{**}}$. By Alaoglu's theorem, $B_{X^{**}}$ is compact with respect to $\sigma(X^{**}, X^*)$. Since the weak-star topology is Hausdorff, so $B_{X^{**}}$ is closed in $\sigma(X^{**}, X^*)$ and

$$\overline{JB_X}^{\sigma(X^{**},X^*)} \subset B_{X^{**}}\,.$$

On the other hand, if there exists $x_0^{**} \notin \overline{JB_X}^{\sigma(X^{**},X^*)}$, since $\overline{JB_X}^{\sigma(X^{**},X^*)}$ is balanced closed convex set in $(X^{**}, \sigma(X^{**}, X^*))$, by Theorem 3.29, there exists $x^* \in X^*$ (use Theorem 3.33 to deduce that $(X^{**}, \sigma(X^{**}, X^*))^* = X^*$) so that

$$|\langle x, x^*\rangle| \leq 1 < |\langle x^*, x_0^{**}\rangle|\ , \text{ for all } x \in B_X\,.$$

Thus $\|x^*\| \leq 1$, and $\|x_0^{**}\| > 1$. So $x_0^{**} \notin B_{X^{**}}$. Therefore,

$$B_{X^{**}} \subset \overline{JB_X}^{\sigma(X^{**},X^*)}\,,$$

and the desired result follows. □

Corollary 5.21. *Let X be a normed linear space. Let J be the canonical embedding of X into X^{**}. Then $J(X)$ is dense in X^{**} relative to the weak-star topology $\sigma(X^{**}, X^*)$ on X^{**}. That is,*

$$\overline{JX}^{\sigma(X^{**},X^*)} = X^{**}\,.$$

Proof. Note that

$$\begin{aligned}\overline{JX}^{\sigma(X^{**},X^*)} &= \overline{\bigcup_{n=1}^{\infty} nJB_X}^{\sigma(X^{**},X^*)} = \bigcup_{n=1}^{\infty} n\overline{JB_X}^{\sigma(X^{**},X^*)}\\ &= \bigcup_{n=1}^{\infty} nB_{X^{**}} = X^{**}\,.\end{aligned}$$

□

Weak compactness and reflexivity. If X is a reflexive Banach space, then the weak and weak-star topologies agree on its dual space X^*, hence the closed unit ball in X^* is weakly compact by the Banach-Alaoglu theorem and so the closed unit ball in X is also weakly compact. The following theorem asserts that this property characterizes reflexivity. It also asserts that weak compactness of the closed unit ball is equivalent to sequential weak compactness.

Theorem 5.22. *Let X be a normed linear space. Let B_X be the closed unit ball in X. Then B_X is weakly compact if and only if X is reflexive.*

Proof. First of all we show that

$$J : (X, \sigma(X, X^*)) \to (J(X), \sigma(X^{**}, X^*)|_{J(X)})$$

is a linear homeomorphism. Obviously J is linear bijection. To see J is homeomorphism, $\{x_i\}$ converges to $x \in X$ with respect to $\sigma(X, X^*)$, $\Leftrightarrow$

$$\langle x_i, x^* \rangle \to \langle x, x^* \rangle \,, \text{ for all } \; x^* \in X^* \,.$$

$\Leftrightarrow$

$$\langle x^*, Jx_i \rangle \to \langle x^*, Jx \rangle \,, \text{ for all } \; x^* \in X^* \,.$$

$\Leftrightarrow$ Jx_i converges to Jx with respect to $\sigma(X^{**}, X^*)|_{J(X)}$. Therefore

- If B_X is weakly compact, since J is continuous, JB_X is cpmpact in X^{**}. By Goldstine theorem, $JB_X = B_{X^{**}}$. Thus $JX = X^{**}$.
- On the other hand, if $J(X) = X^{**}$, note that $B_{X^{**}} = J(B_X)$ is compact with respect to $\sigma(X^{**}, X^*)$. Since J is a linear homeomorphism, B_X is compact with respect to $\sigma(X, X^*)$. □

Corollary 5.23. *X is a reflexive. Let $K \subset X$ be a norm bounded, closed, and convex subset of X. Then K is weakly compact.*

Proof. K is closed and convex, so K is weakly closed subset of a weakly compact set, thus K is compact. □

Weak sequential compactness and reflexity. In connection with the compactness properties of reflexive spaces we also have the following two results about sequential compactness.

Theorem 5.24. *Let X be a reflexive Banach space, B_X is the closed unit ball, then B_X is weakly sequentially compact.*

Proof. Given a sequence $\{x_n\}$ in B_X, let M be the closed subspace spanned by $\{x_n\}$, i.e.,

$$M := \overline{\text{span}}\{x_n\}\,.$$

Then M is separable and by Theorem 5.7 (or Exercise 5.2), M is reflexive. Since $M^{**} = M$ is separable, by Theorem 5.2, M^* is separable. Let $\{y_n^*\}$ dense in M^*. By the *diagonal process*, we can select a subspace $\{z_n\}$ of $\{x_n\}$ so that

$$\lim_{n\to\infty}\langle z_n, y_k\rangle \ \text{ exists for each } \ k \geq 1\,.$$

Then it follows from Theorem 5.14 that $\{z_n\}$ is weakly convergent. $\square$

Remarkably, the converse of Theorem 5.24 is also true. However, these results are among the deepest in the study of weak topologies. and the proof is rather delicate and is omitted. It can be found in Chapter V 13 of *A course in Functional Analysis* by J.Conway.

Theorem 5.25 (Eberlein-Šmulian)**.** *Let be X a Banach space and $A \subset X$, then the following statements are equivalent.*

(i) Each sequence of elements of A has a subsequence that is weakly convergent in X.

(ii) The weak closure of A is weakly compact.

In particular, B_X is weakly sequentially compact if and only if X is reflexive.

Remark 5.3. In order to clarify the connection between Theorem 5.24 and Theorem 5.25, it is useful to recall the following facts:

(a) If X is a metric space, then compactness is equivalent to sequential sompactness. But in infinite-dimensional spaces, the weak topology is never metrizable, so this result is not trivial.

(b) There exist compact topological spaces X and some sequences in X without any convergent subsequence. A typical example is $X = B_{X^*}$, which is compact in the topology $\sigma(X^*, X)$; when $X = \ell^\infty$ it is easy to construct a sequence in X without any convergent subsequence.(see [H.Brezis] exercise 3.18)

(c) If X is a topological space with the property that every sequence admits a convergent subsequence, then X need not be compact.

We give an application of this sequential compactness.

Theorem 5.26 (Best Approximation)**.** *Let X be a reflexive Banach space, K a closed, convex subset of X, For each x in X, there is a point k of K so that*

$$\|x - k\| = dist(x, K) = \inf_{k \in K} \|x - k\| .$$

Remark 5.4. Note that such k may not be unique!

Proof. Without loss of generality, assume $x = 0 \notin K$, then there is $\{k_n\}$ in K so that

$$\|k_n\| \to d = dist(x, K) > 0 .$$

Then there exists some $k \in \overline{K}^{\sigma(X,X^*)} = K$ so that

$$k_n \xrightarrow{w} k .$$

Note that

$$\|k\| \leq \liminf_{n\to\infty} \|k_n\| = d$$

So we get the desired point. □

Exercise 5.2. Use Banach-Alaoglu theorem to show that:

(a) A Banach space X is reflexive iff its dual X^* is reflexive;

(b) A closed linear subspace M of a reflexive space X is reflexive.

5.4 Adjoints

We shall now associate with each $T \in \mathcal{B}(X,Y)$ its adjoint, an operator $T^* \in \mathcal{B}(Y^*, X^*)$, and will see how certain properties of T are reflected in the behavior of T^*.

If X and Y are finite-dimensional, every $T \in \mathcal{B}(X,Y)$ can be represented by a matrix $[T]$; in that case, $[T^*]$ is the transpose of $[T]$, provided that the various vector space bases are properly chosen. No particular attention will be paid to the finite-dimensional case in what follows, but historically linear algebra did provide the background and much of the motivation that went into the construction of what is now known as operator theory.

$$
\begin{tikzcd}
X \arrow[r, "T"] \arrow[dr, dashed, "T^\star y^\star"'] & Y \arrow[d, "y^\star"] \\
& \mathbb{F}
\end{tikzcd}
$$

Figure 5.1: The adjoint of T

Suppose X and Y are normed spaces. To each $T \in \mathcal{B}(X,Y)$ corresponds a unique $T^* \in \mathcal{B}(Y^*, X^*)$ that satisfies

$$\langle Tx, y^* \rangle = \langle x, T^* y^* \rangle \tag{5.6}$$

for all $x \in X$ and all $y^* \in Y^*$. In fact, $T^* y^* = y^* \circ T$, clearly $T^* \in \mathcal{B}(Y^*, X^*)$. T^* is called the *adjoint* (or *dual*) of T. Moreover, one can show that $T^* : (Y^*, \sigma(Y^*, Y)) \to (X^*, \sigma(X^*, X))$ is continuous.

Example 5.13 (Adjoint of Multiplication Operator)**.** Let $(X, \mathcal{F}, \mu)$ and the multiplication operator M_ϕ on L^p, where $1 \leq p < \infty$, be as in Example 1.16. Let $\frac{1}{p} + \frac{1}{q} = 1$, then $M_\phi^* : L^q \to L^q$ is given by

$$M_\phi^* f = \phi f \text{ for all } f \in L^q \,.$$

In other words, the adjoint of multiplication operator with symbol ϕ on L^p is the multiplication operator with symbol ϕ on L^q. To show this, notice that

$$\langle M_\phi f, g\rangle = \int (M_\phi f) g d\mu = \int \phi f g d\mu = \int f(M_\phi^* g) d\mu = \langle f, M_\phi^* g\rangle \,.$$

Example 5.14 (Adjoint of Integral Operator)**.** Let the integral operator K on L^p, where $1 \leq p < \infty$, and kernel k be as in 1.19 or as in Example 1.17 (p=2). Let $\frac{1}{p} + \frac{1}{q} = 1$, then $K^* : L^q \to L^q$ is the integer operator with kernel $k^*(x, y) \coloneqq k(y, x)$. To see this, using Fubini's theorem, we have

$$\begin{aligned}
\langle K f, g\rangle &= \int \left[\int k(x,y) f(y) \mu(dy)\right] g(x)\mu(dx) \\
&= \int \left[\int K(x,y) g(x) \mu(dx)\right] f(y)\mu(dy) = \langle f, K^* g\rangle
\end{aligned}$$

for all $f \in L^p$ and $g \in L^q$.

Example 5.15. Let X be a locally compact Hausdorff space and let $\phi : X \to X$ be a homeomorphism. Let $T : C_0(X) \to C_0(X)$ be the operator defined by $Tf := f \circ \phi$ for $f \in C_0(X)$ (the *pullback* of f under ϕ). Then, under the identification $C_0(X)^* \cong \mathcal{M}(X)$ of the dual space of $C_0(X)$ with the space of $\mathbb{F}$-valued Radon measures with finite total variations on X, the dual operator of T is the operator $T^* : \mathcal{M}(X) \to \mathcal{M}(X)$, which assigns to every measure $\mu : \mathcal{B}(X) \to \mathbb{F}$ its *pushforward* $T^*\mu = \mu \circ \phi^{-1}$ under ϕ. This pushforward is given by $\left(\mu \circ \phi^{-1}\right)(B) := \mu\left(\phi^{-1}(B)\right)$ for every Borel set B in X. To see this, note that

$$\langle Tf, \mu\rangle = \int f \circ \phi d\mu = \int f d\phi_* \mu = \langle f, T^* \mu\rangle$$

for all $f \in C_0(X)$ and $\mu \in \mathcal{M}(X)$.

Proposition 5.27. *X and Y are normed spaces, $T \in \mathcal{B}(X,Y)$. Then $T^* \in \mathcal{B}(Y^*, X^*)$ and $\|T^*\| = \|T\|$.*

Proof. To show $\|T^*\| = \|T\|$, note that

$$\|T\| = \sup_{x \in B_X, y^* \in B_{Y^*}} |\langle Tx, y^* \rangle| = \sup_{x \in B_X, y^* \in B_{Y^*}} |\langle x, T^* y^* \rangle| = \|T^*\| \,,$$

where B_X and B_{Y^*} is the closed unit ball of X and Y^*, respectively. □

Remark 5.5. From this property, we have that $T \mapsto T^{**}$ is an linear isometric embedding form $\mathcal{B}(X,Y)$ to $\mathcal{B}(X^{**}, Y^{**})$. Besides, we can regard T^{**} as a continuous extension of T on X^{**}; indeed,

$$T^{**} \circ J_X = J_Y \circ T \,. \tag{5.7}$$

Obviously, if $A, B \in \mathcal{B}(X,Y)$, and α, β are scalars, then

$$(\alpha A + \beta B)^* = \alpha A^* + \beta B^*$$

If $A \in \mathcal{B}(X,Y)$ and let $B \in \mathcal{B}(Y,Z)$, then

$$(AB)^* = B^* A^* \,.$$

Theorem 5.28 (Duality)**.** *Let X and Y be normed linear spaces. Let $T \in \mathcal{B}(X,Y)$. Then the following assertions hold.*

(a) $R(T)^\perp = N(T^*)$ *and* ${}^\perp R(T^*) = N(T)$.

(b) $\overline{R(T)} = {}^\perp N(T^*)$ *and* $\overline{R(T^*)}^{\sigma(X^*,X)} = N(T)^\perp$.

(c) *T is injective if and only if $R(T^*)$ is weak*-dense in X^*; $R(T)$ is dense in Y if and only if T^* is injective.*

Proof. Note that

$$\begin{aligned} y^* \in R(T)^\perp &\Leftrightarrow \langle Tx, y^* \rangle = 0 \text{ for all } x \\ &\Leftrightarrow \langle x, T^* y^* \rangle = 0 \text{ for all } x \Leftrightarrow T^* y^* = 0 \,; \end{aligned}$$

and

$$x \in {}^{\perp}R(T^*) \Leftrightarrow \langle x, T^*y^* \rangle = 0 \text{ for all } y^*$$
$$\Leftrightarrow \langle Tx, y^* \rangle = 0 \text{ for all } y^* \Leftrightarrow Tx = 0\,,$$

then (a) follows. Trivially, (b), (c) follows from (a). $\square$

Example 5.16. Let X be a normed vector space, let $M \subset X$ be a closed linear subspace, and let $Q : X \to X/M$ be the canonical projection. Then the dual operator $Q^* : (X/M)^* \to X^*$ is the isometric embedding of Theorem 5.5 whose image is $M^\perp$. The dual operator of the inclusion $i : M \to X$ is a surjective operator $i^* : X^* \to M^*$ with kernel $M^\perp$. It descends to the isometric isomorphism $X^*/M^\perp \to M^*$ in Theorem 5.4.

Many of the nontrivial properties of adjoints depend on the completeness of X and Y (the open mapping theorem will play an important role). For this reason, it will be assumed throughout that X and Y are Banach spaces.

Theorem 5.29. *X and Y are Banach spaces, and $T \in \mathcal{B}(X, Y)$. Then T is bijective if and only if T^* is bijective, and in this case*

$$(T^{-1})^* = (T^*)^{-1}\,. \tag{5.8}$$

Particularly, T is an isometry isomorphism if and only if T^* is an isometry isomorphism.

Proof. If T is a bijection, then $T^{-1} \in \mathcal{B}(Y, X)$, so

$$(T^{-1})^* T^* = (TT^{-1})^* = I_Y^* = I_{Y^*}\,,$$

and

$$T^*(T^{-1})^* = (T^{-1}T)^* = I_X^* = I_{X^*}\,.$$

Thus T^* is bijective, and $(T^*)^{-1} = (T^{-1})^*$.

If T^* is bijective, by the preceding proof then T^{**} is, too. By the inverse mapping theorem, T^{**} is a linear homeomorphism, thus T is a injection. On the other hand, since X is Banach space, $J_X(X)$ is Banach space, then $T^{**}J_X(X) = J_Y(TX)$ is a closed subspace in $J_Y(Y)$. Thus TX is a closed subspace of Y. By Theorem 5.28

$$R(T) = {}^{\perp}N(T^*) = Y\,. \qquad \square$$

An example of a Banach space isometry isomorphism is the pullback under a homeomorphism $\phi : X \to X$ of a LCH space, acting on the space of continuous functions on X vanishing at infity, equipped with the supremum norm. Its dual operator is the pushforward under ϕ, acting on the space of $\mathbb{F}$-valued Radon measures on X (see Example 5.15).

5.4.1 Closed Range Theorem

The main theorem of this subsection asserts that a continuous linear operator between two Banach spaces has a closed image if and only if its dual operator has a closed image. A key tool in the proof is that a bounded linear operator $T : X \to Y$ between Banach spaces is surjective if and only if it's an open mapping.

Lemma 5.30. *Let X, Y be Banach spaces. Let $T \in \mathcal{B}(X, Y)$. Then T is surjective if and only if T^* is a linear homeomorphism from Y^* onto $R(T^*)$.*

Proof. Necessity. By Theorem 5.28 we have $N(T^*) = R(T)^{\perp} = \{0\}$. So T^* is a injection. To show $(T^*)^{-1} : R(T^*) \to Y^*$ is continuous, by Exercise 4.1, it suffices to show that there exists some $M > 0$ so that

$$M\|T^*y^*\| \geq \|y^*\|\,, \text{ for all } y^* \in Y^*\,.$$

That is

$$M \sup_{x \in T(B_X)} |\langle x, y^*\rangle| \geq \sup_{y \in B_Y} |\langle y, y^*\rangle|\,, \text{ for all } y^* \in Y^*\,, \tag{5.9}$$

where B_X and B_Y are the closed unit ball in X and Y, respectively. By the open mapping theorem, $T : X \to Y$ is an open mapping, so there exists $r > 0$, such that

$$rB_Y \subset T(B_X)\,.$$

Therefore,

$$\sup_{y \in B_Y} |\langle y, y^* \rangle| = \frac{1}{r} \sup_{y \in rB_Y} |\langle y, y^* \rangle| \leq \frac{1}{r} \sup_{x \in T(B_X)} |\langle x, y^* \rangle|$$

for all $y^* \in Y^*$, and hence (5.9) follows.

Sufficiency. We show that T is an open mapping, which implies that T is a surjective. By the proof of Theorem 4.9, it suffices to show that there exists some $r > 0$ so that

$$B_Y(0, r) \subset \overline{T(B_X(0, 1))}\,.$$

Note that $T(B_X(0, 1))$ is balanced and convex. By Theorem 3.29, for any $y \notin \overline{T(B_X(0, 1))}$, there exists some $y^* \in Y^*$ so that

$$|\langle Tx, y^* \rangle| \leq 1 < \langle y, y^* \rangle\,, \text{ for any } x \in B_X(0, 1)\,.$$

Therefore,

$$\|T^* y^*\| < \langle y, y^* \rangle \leq \|y\| \|y^*\|$$

Since $(T^*)^{-1}$ is continuous, there exises some $M > 0$ so that

$$M\|T^* y^*\| \geq \|y^*\|\,, \text{ for all } y^* \in Y^*\,.$$

Thus

$$\|y\| \geq \frac{1}{M}\,,$$

and we get

$$B_Y(0, 1/M) \subset \overline{T(B_X(0, 1))}\,.$$

We now complete the proof. □

The following consequence is useful in applications.

Corollary 5.31 (Surjection I)**.** *Let X, Y be Banach spaces. Let $T \in \mathcal{B}(X,Y)$. Then the following are equivalent.*

(a) $R(T) = Y$.

(b) *There exists some $M > 0$ so that $M\|T^*y^*\| \geq \|y^*\|$ for all $y^* \in Y^*$.*

(c) $N(T^*) = \{0\}$ *and $R(T^*)$ is norm-closed subspace in X^*.*

Proof. Observe that T^* is a linear homeomorphism from Y^* onto $R(T^*)$ is equivalent to (b) and (c), the desired result follows. □

Observe that $N(T^*) = \{0\}$ implies $\overline{R(T)} = Y$. In fact, $R(T^*)$ is norm-closed implies that $R(T)$ is closed, and hence $R(T) = Y$. This is the so called closed range theorem.

Theorem 5.32 (Closed Range Theorem)**.** *Let X, Y be Banach spaces. Let $T \in \mathcal{B}(X,Y)$. Then each of the following three conditions implies the other two:*

(a) $R(T)$ *is closed in Y ;*

(b) $R(T^*)$ *is weak*-closed in X^* ;*

(c) $R(T^*)$ *is norm-closed in X^* .*

Proof. (a) ⇒ (b). Notice that

$$\overline{R(T^*)}^{\sigma(X^*,X)} = N(T)^{\perp},$$

so it suffices to show that $R(T^*) = N(T)^{\perp}$; i.e., for any $x^* \in N(T)^{\perp}$, there exists $y^* \in Y^*$ so that $x^* = T^*y^*$, that is,

$$\langle x, x^* \rangle = \langle x, T^*y^* \rangle = \langle Tx, y^* \rangle \quad \text{for all } x \in X.$$

Define y^* on $R(T)$ by

$$\langle Tx, y^* \rangle = \langle x, x^* \rangle \quad \text{for all } x \in X.$$

One can show that y^* is well-defined since $x^* \in N(T)^{\perp}$, and y^* is bounded linear functional on $R(T)$. The desired result follows from Hahn-Banach extension theorem.

(b) $\Rightarrow$ (c). This follows trivially from that the weak-star topology is weaker than norm topology.

(c) $\Rightarrow$ (a). Without loss of generality, let $\overline{R(T)} = Y$, or we let $Z = \overline{R(T)}$ and let $S \in \mathcal{B}(X, Z)$ with $Sx = Tx$ for each x, then $R(S) = R(T)$. By Hahn-Banach extension theorem, one can see that $R(S^*) = R(T^*)$.

Since $R(T)^{\perp} = N(T^*)$, ${}^{\perp}N(T^*) = \overline{R(T)} = Y$, hence

$$\overline{N(T^*)}^{\sigma(Y^*,Y)} = N(T^*) = \{0\}.$$

Thus by Theorem 5.31, $R(T) = Y$. □

Exercise 5.3. Let X and Y be real normed vector spaces and let $T : X \to Y$ be a bounded linear operator. Let $x^* \in X^*$. Show that the following are equivalent.

(a) $x^* \in R(T^*)$.

(b) There is a constant $c \geq 0$ such that $|\langle x^*, x\rangle| \leq c\|Tx\|$ for all $x \in X$.

Exercise 5.4 (Surjection II). Let X, Y be Banach spaces. Let $T \in \mathcal{B}(X, Y)$. Show that the following are equivalent.

(a) $R(T^*) = X^*$.

(b) T is a linear homeomorphism form X onto $R(T)$.

(c) There exists some $M > 0$ so that $M\|Tx\| \geq \|x\|$ for all $x \in X$.

(d) $N(T) = \{0\}$ and $R(T)$ is norm-closed.

5.5 Uniformly Convex Space

A norm is called *strictly subadditive* if

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

in strict inequality holds except when x or y is a nonnegative multiple of the other. Furthermore for each of these norms the condition holds uniformly, in the following sense:

For any pair of unit vectors x, y, the norm of $(x+y)/2$ is strictly less than 1 by an amount that depends only on $\|x-y\|$. More explicitly, there is an increasing function $\psi(r)$ defined for positive r in $[0,2]$,

$$\psi(r) > \psi(0) = 0, \text{ for any } r > 0\,, \quad \text{and} \quad \lim_{r \downarrow 0} \psi(r) = 0 \tag{5.10}$$

such that for all x, y such that $\|x\| \leq 1, \|y\| \leq 1$, the inequality

$$\left\| \frac{x+y}{2} \right\| \leq 1 - \psi(\|x-y\|) \tag{5.11}$$

holds.

Definition 5.3. A normed vector space $(X, \|\cdot\|)$ is called *uniformly convex*, if the norm satisfies (5.11) for all vectors x, y in B_X, where $\psi(r)$ is some function satisfying (5.10).

Some other textbooks defined uniformly convex space as follows : For any $\epsilon > 0$, there exists some $\delta > 0$, depending on ϵ so that

$$\left\| \frac{x+y}{2} \right\| \leq 1 - \delta$$

for all x and y in B_X that $\|x-y\| \geq \epsilon$.

The uniform convexity is a geometric property of the unit ball: if we slide a rule of length $\epsilon > 0$ in the unit ball, then its midpoint must stay within a ball of radius $(1-\delta)$ for some $\delta > 0$. In particular, the unit sphere must be "round" and can not include any line segment.

Example 5.17. Let $X = \mathbb{R}^2$. The norm $\|x\|_2 = \sqrt{|x_1|^2 + |x_2|^2}$ is uniformly convex, while the norm $\|x\|_1 = |x_1| + |x_2|$ and the norm $\|x\|_\infty = \max(|x_1|, |x_2|)$ are not uniformly convex. This can be easily seen by staring at the unit balls, as shown in Figure 5.2.

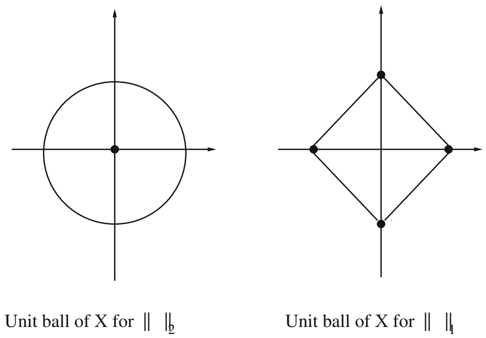

Figure 5.2: Unit ball B_X for $\|\cdot\|_2$ and $\|\cdot\|_1$

Example 5.18. L^p space is uniformly convex for $1 < p < \infty$. In fact, Clarkson's first inequality yields that for $2 \le p < \infty$,

$$\left\|\frac{f+g}{2}\right\|_p^p + \left\|\frac{f-g}{2}\right\|_p^p \le \frac{1}{2}\left(\|f\|_p^p + \|g\|_p^p\right) \quad \forall f, g \in L^p .$$

Besides, Clarkson's second inequality yields that for $1 < p \le 2$:

$$\left\|\frac{f+g}{2}\right\|_p^q + \left\|\frac{f-g}{2}\right\|_p^q \le \left(\frac{1}{2}\|f\|_p^p + \frac{1}{2}\|g\|_p^p\right)^{q/p} \forall f, g \in L^p .$$

Strong convergence and weak convergence. We conclude with a useful property of uniformly convex spaces.

Theorem 5.33. *Let X be a uniformly convex space. Let $x, x_n (n = 1, 2, \cdots)$ in X. Then $\{x_n\}$ converges strongly to x if and only if $\|x_n\| \to \|x\|$ and $\{x_n\}$ converges weakly to x.*

If we assume X is a Hilbert space, then this theorem is very easy to show. In fact, if $x_n \xrightarrow{w} x$ then $\langle x_n, x\rangle \to \|x\|^2$ and hence

$$\begin{aligned}\lim_{n\to\infty} \|x_n - x\|^2 &= \lim_{n\to\infty} \big(\|x_n\|^2 + \|x\|^2 - 2\,\mathrm{Re}\langle x_n, x\rangle\big) \\ &= 2\|x\|^2 - 2\|x\|^2 = 0\end{aligned}$$

as required. In the general case, the proof is following:

Proof. Without loss of generality, we assume $\|x\| > 0$, and $\|x_n\| > 0$ for all n. Note that

$$\epsilon\left(\left\|\frac{x_n}{\|x_n\|} - \frac{x}{\|x\|}\right\|\right) \le 1 - \frac{1}{2}\left\|\frac{x_n}{\|x_n\|} + \frac{x}{\|x\|}\right\| .$$

To show $\|x_n - x\| \to 0$, it suffices to show that

$$\epsilon\left(\left\|\frac{x_n}{\|x_n\|} - \frac{x}{\|x\|}\right\|\right) \to 0 \,, \tag{5.12}$$

since $\epsilon(\cdot)$ is incresing, continuous at 0, $\epsilon(r) > 0$ for any positive r, and $\|x_n\| \to \|x\|$. To show (5.12), it suffices to show that

$$\left\|\frac{x_n + x}{2}\right\| \to \|x\| \,.$$

Since $x_n \xrightarrow{w} x$, we have $(x_n + x)/2 \xrightarrow{w} x$ and hence

$$\|x\| \le \liminf_{n\to\infty} \left\|\frac{x_n + x}{2}\right\|$$

On the other hand,

$$\limsup_{n\to\infty} \left\|\frac{x_n + x}{2}\right\| \le \limsup_{n\to\infty} \frac{\|x_n\| + \|x\|}{2} = \|x\| \,.$$

So we get the desired result. $\square$

Best approximation. As in Theorem 2.6, the best approximation for a complete and convex subset of a uniformly convex space exists and, is unique.

Theorem 5.34 (Existence of the Unique Best Approximation)**.** *Let X be a uniformly convex space. Let K be a complete, convex subset of X. Then each $x \in X$ has a unique best approximation in K, i.e. there is a unique point $k \in K$ satisfying*

$$\|x - k\| = dist(x, K) := \inf_{k \in K} \|x - k\| . \tag{5.13}$$

Proof. Without loss of generality, assume $x = 0 \notin K$, then there is $\{k_n\}$ in K so that

$$\|k_n\| \to d = dist(x, K) > 0 .$$

Assume $\|k_n\| > 0$ for all n, denote

$$x_n = \frac{k_n}{\|k_n\|}$$

Then

$$\epsilon\left(\|x_n - x_m\|\right) \le 1 - \left\|\frac{x_n + x_m}{2}\right\| .$$

Note that

$$\begin{aligned}\frac{x_n + x_m}{2} &= \frac{k_n}{2\|k_n\|} + \frac{k_m}{2\|k_m\|} \\ &= \left(\frac{1}{2\|k_n\|} + \frac{1}{2\|k_m\|}\right)\left[ck_n + (1-c)k_m\right]\end{aligned}$$

where c is a constant in $(0, 1)$. Since $ck_n + (1-c)k_m \in K$.

$$\frac{x_n + x_m}{2} \ge \frac{d}{2}\left(\frac{1}{\|k_n\|} + \frac{1}{\|k_m\|}\right)$$

Thus

$$\lim_{n,m\to\infty} \epsilon\left(\|x_n - x_m\|\right) \to 0 .$$

By the properties of $\epsilon(\cdot)$, we know

$$\lim_{n,m\to\infty} \|x_n - x_m\| \to 0 .$$

Since $\|k_n\| \to d$, then

$$\lim_{n,m\to\infty} \|k_n - k_m\| \to 0 .$$

Thus $\{k_n\}$ is a Cauchy sequence, sence K is complete, there is some $k \in K$ so that $k_n \to k$, thus $\|k\| = d$.

If there exists $k' \in K$ so that $\|k'\| = d$, then

$$\epsilon\left(\left\|\frac{k-k'}{d}\right\|\right) \leq 1 - \left\|\frac{k+k'}{2d}\right\| \leq 0\,.$$

since $k + k'/2 \in K$, so $k = k'$. □

Remark 5.6. By Theorem 5.35, and Theorem 5.26, the existence of k is obvious, the uniform convexity guarantee the uniqueness.

Uniform convexity and reftexity. As everyone knows, every Hilbert space is reflexive. Surprisingly, every uniformly convex Banach space is, too.

Theorem 5.35 (Milman-Pettis)**.** *Every uniformly convex Banach space is reflexive.*

Proof. Suppose that X is a non-reflexive, uniformly convex Banach space. Then for some $\epsilon > 0$ there exists x^{**} in $B_{X^{**}}$ such that the distance between x^{**} and JB_X is 2ϵ. Let $\delta = \psi(\epsilon)$, so if x and y are in X with $\|x\| \leq 1, \|y\| \leq 1$ and $1 - \delta \leq \|\frac{x+y}{2}\|$, then $\|x - y\| \leq \epsilon$. Take x^* in X^*, with $\|x^*\| = 1$ such that $\langle x^*, x^{**}\rangle > 1 - \delta/2$. Let V be the weak-star neighborhood of x^{**} given by

$$V = V(x^{**}, x^*, \delta/2) = \{u^{**} \in X^{**} : |\langle x^*, u^{**} - x^{**}\rangle| < \delta/2\}\ .$$

If Jx and Jy are in JB_X belonging to V, then

$$\left\|\frac{x+y}{2}\right\| \geq \left|\left\langle \frac{x+y}{2}, x^*\right\rangle\right| > 1 - \delta$$

Hence $\|x - y\| \leq \epsilon$. Fixing $Jx \in V \cap JB_X$, we conclude that $V \cap JB_X \subset Jx + \epsilon B_{X^{**}}$. By Goldstein's theorem 5.20, we know that $V \cap JB_X$ is weak-star dense in $V \cap B_{X^{**}}$ which, since $Jx + \epsilon B\,(X^{**})$ is weak-star closed, yields

x^{**} belongs to $Jx + \epsilon B\left(X^{**}\right)$. But this means that the distance between x^{**} and $B(X)$ is less than or equal to ϵ, contradicting our choice of x^{**}. $\square$

Remark 5.7. Uniform convexity is a *geometric property of the norm*; an equivalent norm need not be uniformly convex. On the other hand, reflexivity is a *topological property*: a reflexive space remains reflexive for an equivalent norm. It is a striking feature of this theorem that a geometric property implies a topological property. Uniform convexity is often used as a tool to prove reflexivity; but it is not the ultimate tool, since there are some weird reflexive spaces that admit no uniformly convex equivalent norm!

Chapter 6

Spectral Theory

6.1 Spectrum of Closed Operators

Let T be a linear operator whose domain $D(T)$ and range $R(T)$ both lie in the same complex linear topological space X. We consider the linear operator

$$T_\lambda = \lambda I - T$$

where λ is a complex number and I the identity operator. The distribution of the values of λ for which T_λ has an inverse and the properties of the inverse when it exists, are called the *spectral theory* for the operator T. We shall thus discuss the general theory of the inverse of T_λ.

Definition 6.1. If λ is such that the range $R(T_\lambda)$ is dense in X and T_λ has a continuous inverse $T_\lambda^{-1} = (\lambda I - T)^{-1}$, we say that λ is in the *resolvent set* $\varrho(T)$ of T, and we denote this inverse $(\lambda I - T)^{-1}$ by $R(\lambda; T)$ and call it the resolvent (at λ) of T.

Definition 6.2. All complex numbers λ not in $\varrho(T)$ form a set $\sigma(T)$ called the *spectru*m of T. The spectrum $\sigma(T)$ is decomposed into disjoint sets $\sigma_p(T), \sigma_c(T)$ and $\sigma_r(T)$ with the following properties:

- $\sigma_p(T)$ is the totality of complex numbers λ for which T_λ does not have an inverse ($N(\lambda I - T) \neq \{0\}$); $\sigma_p(T)$ is called the *point spectrum* of T.
- $\sigma_c(T)$ is the totality of complex numbers λ for which T_λ has a discontinuous inverse with domain dense in X ($\overline{R(\lambda I - T)} = X$); $\sigma_c(T)$ is called the *continuous spectrum* of T.
- $\sigma_r(T)$ is the totality of complex numbers λ for which T_λ has an inverse whose domain is not dense in X ($\overline{R(\lambda I - T)} \neq X$); $\sigma_r(T)$ is called the *residual spectrum* of T.

Remark 6.1. Why we need the scalar field be $\mathbb{C}$? The answer can be found in Theorem 6.3 and Theorem 6.8: The spectrum is not empty when X is a complex Banach space and $T \in \mathcal{B}(X)$, and $\lambda \to R(\lambda, T)$ is a operator-valued holomorphic function on the open subset $\varrho(T)$ of $\mathbb{C}$. However, if X is a real Banach space, all the definitions about spectrum can be done similarly.

From these definitions and the linearity of T we have : A necessary and sufficient condition for $\lambda \in \sigma_p(T)$ is that the equation

$$Tx = \lambda x$$

has a nonzero solution $x \in D(T)$. In this case λ is called an *eigenvalue* of T, and x the corresponding *eigenvector*. The null space $N(\lambda I - T)$ is called the *eigenspace* of T corresponding to the eigenvalue λ. It consists of the vector 0 and the totality of eigenvectors corresponding to λ. The dimension of the eigenspace corresponding to λ is called the *multiplicity* of the eigenvalue λ.

In practice, when discussing the spectrum, we always work on a closed (or closable) linear operator T defined on a complex Banach space X.

Lemma 6.1. *Let X be a complex Banach space, and T a closed linear operator with its domain $D(T)$ and range $R(T)$ both in X. Then, the following statements are equivalent.*

(a) $\lambda \in \varrho(T)$*, i.e.,* T_λ^{-1} *exists,* T_λ^{-1} *is densely defined and continuous.*

(b) T_λ *is a bijection of* $D(T)$ *onto* X*.*

(c) T_λ^{-1} *exists and* $T_\lambda^{-1} \in B(X)$*.*

Proof. Recall that the inverse of a closed linear operator is also closed. Then part (c) follows from part (b) by using the closed graph theorem. Part (c) implies part (a) trivially. It remains to show that part (a) implies part (b). By assumption, T_λ^{-1} is a densely defined continuous operator, and is clsoed. From Remark 4.2, $D(T_\lambda^{-1})$ is closed in X. Since $D(T_\lambda^{-1})$ is dense in X, thus $D(T_\lambda^{-1}) = X$ as desired. $\square$

Lemma 6.2. *Let* X *be a complex Banach space, and* $(T, D(T))$ *a closable linear operator in* X*. Then we have*

$$\sigma(\overline{T}) = \sigma(T) \quad \textit{and} \quad \varrho(\overline{T}) = \varrho(T)\,.$$

Proof. It suffices to show that $\varrho(\overline{T}) \subset \varrho(T)$ and $\varrho(T) \subset \varrho(\overline{T})$. To prove the first one, assume $\lambda \in \varrho(\overline{T})$. Then clearly $\lambda I - T$ is injective and $(\lambda I - T)^{-1}$ is continuous. Moreover,

$$X = R(\lambda I - \overline{T}) \subset \overline{R(\lambda I - T)}$$

thus $(\lambda I - T)^{-1}$ is densely defined. So $\lambda \in \varrho(T)$.

On the other hand, assume $\lambda \in \varrho(T)$. Then clearly $R(\lambda I - \overline{T})$ is dense in X. Since $T_\lambda^{-1} = (\lambda I - T)^{-1}$ is continuos, there exists a constant $c_\lambda > 0$ so that

$$\|(\lambda I - T)x\| \geq c_\lambda \|x\| \quad \text{whenever} \quad x \in D(T)\,.$$

Then it's easy to see that

$$\left\|(\lambda I - \bar{T})x\right\| \geq c_\lambda \|x\| \quad \text{whenever} \quad x \in D(\bar{T})\,.$$

Thus $(\lambda I - \bar{T})$ is injective and $(\lambda I - \bar{T})^{-1}$ is continuous. So $\lambda \in \varrho(\bar{T})$ as claimed. $\square$

As a consequence of Lemma 6.1, the partition of the spectrum can be restated as following:

$$\sigma_p(T) = \{\lambda \in \mathbb{C} : N(T_\lambda) \neq \{0\}\}\,;$$
$$\sigma_c(T) = \{\lambda \in \mathbb{C} : N(T_\lambda) = \{0\}\ ,\ \overline{R(T_\lambda)} = X\ ,\ R(T_\lambda) \neq X\}\,;$$
$$\sigma_r(T) = \{\lambda \in \mathbb{C} : N(T_\lambda) = \{0\}\ ,\ \overline{R(T_\lambda)} \neq X\}\,.$$

Henceforth, without specific statement, we always suppose that X is a complex Banach space and T a closed linear operator of $D(T) \subset X$ into X.

The most important cases is that the domain of T is the whole space X. In this case,

$$\varrho(T) = \{\lambda \in \mathbb{C} : (\lambda I - T)^{-1} \in \mathcal{B}(X)\}\,.$$

We note that the resolvent set $\varrho(T)$ consists of all those $\lambda \in \mathbb{C}$ for which the equation

$$\lambda x - Tx = y$$

has a unique solution for each y which furthermore depends continuously on the right hand side y.

Remark 6.2. Let X be a complex Banach space, and T a closed linear operator with its domain $D(T)$ and range $R(T)$ both in X. We call $\lambda \in \mathbb{C}$ an *approximate eigenvalue* for T, if there exists a sequence $x_n \in D(T)$ with $\|x_n\| = 1$ so that $\|Tx_n - \lambda x_n\| \to 0$. We call the set

$$\sigma_{ap}(T) := \{\lambda : \lambda \text{ approximate eigenvalue of } T\}$$

the *approximate point spectrum* of T.

We note that every eigenvalue λ is also an approximate eigenvalue (as we may simply choose $x_n = x$ for an element of $N(\lambda I - T)$ that is normalised to $\|x\| = 1$), so we have

$$\sigma_p(T) \subset \sigma_{ap}(T) \subset \sigma(T)\,.$$

We also remark that λ is an approximate eigenvalue if and only if there exists NO $\delta > 0$ so that

$$\|Tx - \lambda x\| \geq \delta \|x\| \quad \text{for all } x \in X\,, \tag{6.1}$$

as is equivalent to

$$\inf_{x \in X, \|x\|=1} \|Tx - \lambda x\| = 0\,.$$

Moreover, we get directly that $\sigma_c(T) \subset \sigma_{ap}(T)$ from (6.1).

6.1.1 Basic Properties

Our first main result about the spectrum of closed linear operators is

Theorem 6.3. *Let T be a closed linear operator with domain $D(T)$ and range $R(T)$ both in a complex Banach space X. Then the following assertions hold.*

(a) *For each $\lambda_0 \in \varrho(T)$, $B(\lambda_0; \frac{1}{\|R(\lambda_0;T)\|}) \subset \varrho(T)$ and*

$$R(\lambda; T) = \sum_{n=0}^{\infty} (\lambda_0 - \lambda)^n R(\lambda_0; T)^{n+1} \quad \textit{for all} \quad |\lambda - \lambda_0| < \frac{1}{\|R(\lambda_0; T)\|}\,.$$

Particularly, the resolvent set $\varrho(T)$ is an open set of $\mathbb{C}$;

(b) *$R(\lambda; T)$ is a holomorphic function of λ in $\varrho(T)$, and*

$$\frac{d^n}{d\lambda^n} R(\lambda; T) = (-1)^n n! R(\lambda; T)^{n+1} \quad \textit{for all } n \in \mathbb{N}\,.$$

(c) *Let $\lambda_n \in \varrho(T)$ with $\lim\limits_{n\to\infty} \lambda_n = \lambda_0$. Then $\lambda_0 \in \partial\varrho(T) \subset \sigma(T)$ if and only if $\lim\limits_{n\to\infty} \|R(\lambda_n, T)\| = \infty$.*

Proof. By the theorem of the preceding section, $R(\lambda; T)$ for $\lambda \in \varrho(T)$ is an everywhere defined continuous operator. Let $\lambda_0 \in \varrho(T)$ and consider

$$S(\lambda) = \sum_{n=0}^{\infty} (\lambda_0 - \lambda)^n R(\lambda_0, A)^{n+1}\,.$$

The series is convergent in the operator norm whenever

$$|\lambda_0 - \lambda| \, \|R\,(\lambda_0; T)\| < 1\,,$$

and within this circle of the complex plane, the series defines a holomorphic function of λ. Multiplication by $(\lambda I - T) = (\lambda - \lambda_0)\, I + (\lambda_0 I - T)$ on the left or right gives I so that the series $S(\lambda)$ actually represents the resolvent $R(\lambda; T)$. Thus we have proved that a circular neighbourhood of λ_0 belongs to $\varrho(T)$ and $R(\lambda; T)$ is holomorphic in this neighbourhood. Assertion (b) follows immediately from the series representation for the resolvent.

To show (c) we use (a), which implies $\|R(\mu, T)\| \geq 1/\operatorname{dist}(\mu, \sigma(T))$ for all $\mu \in \varrho(T)$. This already proves one implication. For the converse, assume that $\lambda_0 \in \varrho(T)$. Then the continuous resolvent map remains bounded on the compact set $\{\lambda_n : n \geq 0\}$. This contradicts the assumption that $\lim_{n\to\infty} \|R\,(\lambda_n, T)\| = \infty$; hence $\lambda_0 \in \sigma(T)$. $\square$

Theorem 6.4 (Resolvent Equation)**.** *Let T be a closed linear operator with domain and range both in a complex Banach space X. If λ and μ both belong to $\varrho(T)$, then the resolvent equation holds:*

$$\begin{aligned} R(\lambda; T) - R(\mu; T) &= (\mu - \lambda) R(\lambda; T) R(\mu; T) \\ &= (\mu - \lambda) R(\mu; T) R(\lambda; T)\,. \end{aligned}$$

particularly, $R(\lambda; T)$ and $R(\mu; T)$ commute.

Proof. We have

$$\begin{aligned} R(\lambda; T) &= R(\lambda; T)(\mu I - T) R(\mu; T) \\ &= R(\lambda; T)\{(\mu - \lambda) I + (\lambda I - T)\} R(\mu; T) \\ &= (\mu - \lambda) R(\lambda; T) R(\mu; T) + R(\mu; T)\,. \end{aligned}$$

Then the desired result follows. $\square$

Exercise 6.1. Let $\Omega \subset \mathbb{C}$. Suppose $\{R_\lambda\}_{\lambda\in\Omega}$, is a family of bounded linear operators on X satisfying the resolvent equation

$$R_\lambda = R_\mu + (\mu - \lambda)R_\lambda R_\mu \quad \text{for all } \lambda, \mu \in \Omega\,.$$

Then the following statements hold. (i) $\{R_\lambda\}_{\lambda\in\Omega}$ commutes. (ii) The kernel and the range of R_λ is independent of $\lambda \in \Omega$. (iii) If R_λ is injective for some, hence all, $\lambda \in \Omega$, then let $T \coloneqq \lambda I - R_\lambda^{-1}$. The defined of T is independent of λ and T is a closed linear operator.

Proposition 6.5. *Let T be a closed linear operator with domain and range both in a complex Banach space X. Let $\lambda_1, \cdots, \lambda_n$ be distinct eigenvalues of T. If $x_1, \cdots, x_n$ are corresponding non-trivial eigenvectors, then $\{x_1, \cdots, x_n\}$ are linear independent. In particular, the (internal) direct sum*

$$N(\lambda_1 I - T) \oplus \cdots \oplus N(\lambda_n I - T)$$

is well-defined.

Proof. We prove by induction. When $n = 1$, the propostion is trivial. Assume the propostion holds for $n - 1$. If $\alpha_j \in \mathbb{C}$, $j = 1, \cdots, n$ so that

$$\alpha_1 x_1 + \cdots + \alpha_n x_n = 0\,.$$

Let T act on the equality, then we get

$$\alpha_1 \lambda_1 x_1 + \cdots + \alpha_n \lambda_n x_n = 0$$

Thus we get

$$\alpha_1(\lambda_n - \lambda_1)x_1 + \cdots + \alpha_{n-1}(\lambda_n - \lambda_{n-1})x_{n-1} = 0\,.$$

Thus $\alpha_1 = \cdots = \alpha_{n-1} = 0$ since $\{x_1, \cdots, x_{n-1}\}$ are linear independent by induction and $\lambda_1, \cdots, \lambda_n$ are distinct. Then clearly $\alpha_n = 0$ and the desired result follows. $\square$

6.1.2 Examples

In the end of this section we discuss some concrete examples.

Example 6.1. If the space X is of finite dimension, then any bounded linear operator T is represented by a matrix (t_{ij}). It is known that the eigenvalues of T are obtained as the roots of the algebraic equation, the so-called characteristic equation of the matrix (t_{ij}),

$$\det(\lambda\delta_{ij} - t_{ij}) = 0\,.$$

Moreover, $\sigma(T) = \sigma_p(T)$ is the set of eigenvalues and $|\sigma(T)| \leq n$.

Example 6.2 (Multiplication Operators on $C_0(\Omega)$)**.** We start from a locally compact Hausdorff space Ω and define the Banach space (endowed with the sup-norm) $\mathrm{C}_0(\Omega) = \mathrm{C}_0(\Omega;\mathbb{F})$ of all continuous, $\mathbb{F}$-valued functions on Ω that vanish at infinity. With any continuous function $q : \Omega \to \mathbb{F}$ we associate a linear operator M_q on $\mathrm{C}_0(\Omega)$ defined on its "maximal domain" $D(M_q)$ in $\mathrm{C}_0(\Omega)$. Specifically, let $M_q f \coloneqq q \cdot f$, for all f in the domain

$$D(M_q) \coloneqq \{f \in \mathrm{C}_0(\Omega) : q \cdot f \in \mathrm{C}_0(\Omega)\}\ .$$

The main feature of these multiplication operators is that most operator-theoretic properties of M_q can be characterized by analogous properties of the function q. In the following proposition we give some examples for this correspondence:

(i) $(M_q, D(M_q))$ is closed and densely defined.

(ii) $M_q \in \mathcal{B}(\mathrm{C}_0(\Omega))$ if and only if the function q is bounded. In that case, one has $\|M_q\| = \|q\| := \sup\{|q(s)| : s \in \Omega\}$.

(iii) The spectrum of M_q is the closed range of q; i.e., $\sigma(M_q) = \overline{\mathrm{im}(q)}$.

Example 6.3 (Multiplication Operators on L^p)**.** Multiplication operators arise in a natural way in various instances. For example, if one applies the Fourier transform to a linear differential operator on $L^2\left(\mathbb{R}^d\right)$, this operator becomes a multiplication operator on $L^2\left(\mathbb{R}^d\right)$.

Let $(\Omega, \mathcal{F}, \mu)$ be a σ-finite measure space. For fixed $1 \leq p < \infty$, we consider the Banach space $L^p(\mu) = L^p(\Omega, \mathcal{F}, \mu; \mathbb{C})$ of all (equivalence classes of) p-integrable complex functions on Ω. For a measurable function q, we associate a linear operator M_q on $L^p(\mu)$ defined on its "maximal domain" $D\left(M_q\right)$ in $L^p(\mu)$. Specifically, let $M_q f \coloneqq q \cdot f$, for all f in the domain

$$D\left(M_q\right) \coloneqq \left\{f \in L^p(\mu) : q \cdot f \in L^p(\mu)\right\} .$$

We define the *essential range* of q by

$$\text{ess.im}(q) \coloneqq \left\{z \in \mathbb{C} : \text{ for all } r > 0 : \mu(q \in B(z, r)) \neq 0\right\}.$$

In other words $\operatorname{ess.im}(q) = \operatorname{supp}(\mu \circ q^{-1})$. One can easily check that the complement of $\operatorname{ess.im}(q)$, given by

$$\left\{z \in \mathbb{C} : \text{ there exists } r > 0, \mu(q \in B(z, r)) = 0\right\},$$

is open, and since $\mathbb{C}$ is second countable, $\mu(q \in \operatorname{ess.im}(q)^c) = 0$. Hence the essential image $\operatorname{ess.im}(q)$ is always closed. If B is a Borel set in $\mathbb{C}$ disjoint with $\operatorname{ess.im}(q)$, then $\mu(q \in B) = 0$. This fact characterises the essential image: It is the smallest closed subset of $\mathbb{C}$ with this property.

In analogy to the preceding example, we now have the following result.

(i) $\left(M_q, D\left(M_q\right)\right)$ is closed and densely defined.

(ii) $M_q \in \mathcal{B}(\mathrm{C}_0(\Omega))$ if and only if the function q is essentially bounded. In that case, one has $\|M_q\| = \|q\|_\infty := \sup\{\lambda : \lambda \in \text{ess.im}(q)\}$.

(iii) The spectrum of M_q is the essential range of q; i.e., $\sigma\left(M_q\right) = \text{ess.im}(q)$.

Example 6.4. Let $X = L^2(\mathbb{R}, \mathbb{C})$ and let T be defined by

$$(Tx)(t) = tx(t) \text{ for } t \in \mathbb{R}\,;$$

where, $D(T) = \left\{x(t) : x(t) \text{ and } tx(t) \in L^2(\mathbb{R}, \mathbb{C})\right\}$.Then by the example above, $\sigma(T) = \mathbb{R}$. We claim that indeed $\sigma_c(T) = \mathbb{R}$.

To see this, let $\lambda_0 \in \mathbb{R}$. Note that $(\lambda_0 I - T)$ is injective. Thus $(\lambda_0 I - T)^{-1}$ exists. Its' domain $R\,(\lambda_0 I - T)$ comprises those $y(t) \in L^2(\mathbb{R}, \mathbb{C})$ which vanish identically in the neighbourhood of $t = \lambda_0$; the neighbourhood may vary with $y(t)$. Hence $R\,(\lambda_0 I - T)$ is dense in $L^2(\mathbb{R}, \mathbb{C})$. It is easy to see that the operator $(\lambda_0 I - T)^{-1}$ is not bounded on the totality of such $y(t)$'s.

6.2 Spectrum of Bounded Linear Operators

In this section, we always set X a complex Banach space and T a continuous linear operator from X into X.

Theorem 6.6. *Let X be a complex Banach space and $T \in \mathcal{B}(X)$. Then the following limit exists:*

$$\lim_{n\to\infty} \|T^n\|^{\frac{1}{n}} = \inf_{n\geq 1} \|T^n\|^{\frac{1}{n}} =: r_\sigma(T)\,.$$

It is called the spectral radius of T. If $|\lambda| > r_\sigma(T)$ (for example $\lambda > \|T\|$), then the resolvent $R(\lambda; T)$ exists and is given by the series

$$R(\lambda; T) = \sum_{n=1}^{\infty} \frac{1}{\lambda^n} T^{n-1}\,, \tag{6.2}$$

which converges in the norm of operators.

Proof. Set $r = \inf_{n\geq 1} \|T^n\|^{\frac{1}{n}} \geq 0$. It suffices to show that $\limsup_n \|T^n\|^{\frac{1}{n}} \leq r$. For any $\epsilon > 0$, choose k such that $\left\|T^k\right\|^{\frac{1}{k}} \leq r + \epsilon$. For arbitrary n write $n = pk + q$ where $0 \leq q \leq (k-1)$. Then, by $\|AB\| \leq \|A\|\|B\|$, we obtain

$$\|T^n\|^{\frac{1}{n}} \leq \left\|T^k\right\|^{\frac{p}{n}} \cdot \|T\|^{\frac{q}{n}} \leq (r+\epsilon)^{\frac{kp}{n}} \|T\|^{\frac{q}{n}}\,.$$

Since $(kp)/n \to 1$ and $q/n \to 0$ as $n \to \infty$, we have $\limsup_n \|T^n\|^{\frac{1}{n}} \le r + \epsilon$. Since e was arbitrary, $\limsup_n \|T^n\|^{\frac{1}{n}} \le r$.

The series is convergent in the norm of operators when $|\lambda| > r_\sigma(T)$. For, if $|\lambda| \ge r_\sigma(T) + \epsilon$, then

$$\left\| \frac{1}{\lambda^{n+1}} T^n \right\| \le \frac{(r_\sigma(T) + \epsilon/2)^n}{(r_\sigma(T) + \epsilon)^{n+1}} .$$

Thus the series in (6.2) converges. Multiplication by $(\lambda I - T)$ on the left or right of this series gives I so that the series actually represents the resolvent $R(\lambda; T)$. □

Corollary 6.7. *Let X be a complex Banach space and $T \in \mathcal{B}(X)$. Then resolvent set $\varrho(T)$ is not empty.*

One of the most important aspects of the following theorem is that every bounded operator has non-empty spectrum. Here we crucially use that the vector space is over $\mathbb{C}$. The claim is not true if we were to only consider the real spectrum as you already know from Linear Algebra.

Theorem 6.8 (Spectral Radius)**.** *Let X be a complex Banach space and $T \in \mathcal{B}(X)$. Then the spectrum $\sigma(T)$ is non-empty and compact subset of $\mathbb{C}$. In fact,*

$$r_\sigma(T) = \sup_{\lambda \in \sigma(T)} |\lambda| ,$$

and this is why we call $r_\sigma(T)$ the spectral radius of T.

Proof. Step 1. We show that spectrum $\sigma(T)$ is non-empty.

If $\sigma(T) = \emptyset$, then $\varrho(T) = \mathbb{C}$ and $\lambda \mapsto R(\lambda, T)$ is holomorphic on the plane $\mathbb{C}$. Since for $\lambda > \|T\|$, $R(\lambda, T) = \sum_{n=1}^{\infty} \lambda^{-n} T^{n-1}$, we have

$$\|R(\lambda, T)\| \le \frac{1}{\lambda - \|T\|} .$$

Thus $\lambda \mapsto R(\lambda, T)$ is a bounded holomorphic vector-valued function on $\mathbb{C}$. It follows from Theorem 4.23 that $R(\lambda, T)$ is a constant, which is a contradiction!

Step 2. We show that $r_\sigma(T) = \sup_{\lambda\in\sigma(T)} |\lambda|$. Combine this and Theorem 6.3, $\sigma(T)$ is closed and bounded, hence compact.

By Theorem 6.6, we know that $r_\sigma(T) \geq \sup_{\lambda\in\sigma(T)} |\lambda|$. Hence we have only to show that $r_\sigma(T) \leq \sup_{\lambda\in\sigma(T)} |\lambda|$.

By Theorem 6.3, $R(\lambda; T)$ is holomorphic in λ when $|\lambda| > \sup_{\lambda\in\sigma(T)} |\lambda|$. Thus it admits a uniquely determined Laurent expansion in positive and non-positive powers of λ convergent in the operator norm for $|\lambda| > \sup_{\lambda\in\sigma(T)} |\lambda|$. By Theorem 6.6, this Laurent series must coincide with $\sum_{n=1}^\infty \lambda^{-n}T^{n-1}$. Hence $\lim_{n\to\infty} \|\lambda^{-n}T^n\| = 0$ if $|\lambda| > \sup_{\lambda\in\sigma(T)} |\lambda|$, and so for any $\epsilon > 0$,

$$\|T^n\| \leq \left(\epsilon + \sup_{\lambda\in\sigma(T)} |\lambda|\right)^n \qquad \text{for large } n\,.$$

This proves that

$$r_\sigma(T) = \lim_{n\to\infty} \|T^n\|^{\frac{1}{n}} \leq \sup_{\lambda\in\sigma(T)} |\lambda|\,,$$

as desired. $\square$

Remark 6.3. Since $\sigma(T)$ is a compact subset in $\mathbb{C}$, it's not hard to see that indeed

$$r_\sigma(T) = \sup_{\lambda\in\sigma(T)} |\lambda| = \max_{\lambda\in\sigma(T)} |\lambda|\,.$$

Corollary 6.9. *The series $\sum_{n=1}^\infty \lambda^{-n}T^{n-1}$ diverges if $|\lambda| < r_\sigma(T)$.*

Proof. Let r be the smallest non-negative real number such that the series

$$\sum_{n=1}^\infty \lambda^{-n}T^{n-1}$$

converges in the operator norm for $|\lambda| > r$. The existence of such an r is proved as for ordinary power series in λ^{-1}. Then, for $|\lambda| > r$, we have $\lim_{n\to\infty} \|\lambda^{-n}T^n\| = 0$ and so, as in the proof of $r_\sigma(T) \le \sup_{\lambda\in\sigma(T)} |\lambda|$, we must have $\lim_{n\to\infty} \|T^n\|^{1/n} \le r$. This proves that $r_\sigma(T) \le r$. In fact, $r_\sigma(T) = r$. $\square$

We furthermore record the following useful lemma. This lemma has been shown in Exercise 0.2.

Lemma 6.10. *Let X be a vector space, Let $S, T \in \mathcal{L}(X)$. Suppose that $ST = TS$. Then ST is bijective if and only if S and T are bijective.*

Theorem 6.11. *Let X be a complex Banach space, $T \in \mathcal{B}(X)$ and let p be a complex polynomial. Then*

$$\sigma(p(T)) = p(\sigma(T)) := \{p(\lambda) : \lambda \in \sigma(T)\}$$

Here we set $p(T) := \sum_{j=0}^n a_j T^j$ if the polynomial p is given by $p(z) = \sum_{j=0}^n a_j z^j$, with the usual convention that $T^0 = I$.

Proof. We first remark that if p is constant, say $p = c \in \mathbb{C}$, then the spectrum of $p(T) = cI$ is simply $\{c\}$. while the fact that $\sigma(T)$ is non-empty implies that also $p(\sigma(T)) = \{c\}$. So suppose that p has degree $n \ge 1$, let $\mu \in \mathbb{C}$ be any given number. As we are working in $\mathbb{C}$ we can factorise $p(\cdot) - \mu$ and write it as $p(z) - \mu = \alpha\,(z - \beta_1(\mu)) \ldots (z - \beta_n(\mu))$ for some $\alpha \neq 0$ and equally factorise

$$p(T) - \mu I = \alpha\,(T - \beta_1(\mu)I) \ldots (T - \beta_n(\mu)I) \tag{6.3}$$

where we note that all operators on the right hand side commute which will allow us to apply Lemma 6.10. Thus, $\mu \in \varrho(p(T)) \Leftrightarrow \beta_j(\mu) \in \varrho(T)$ for all j. In other words,

$$\mu \in \sigma(p(T)) \Leftrightarrow \exists\; j \text{ so that } \beta_j(\mu) \in \sigma(T)\,.$$

We now note that $\mu \in p(\sigma(T))$ if and only if the equation $p(z) - \mu = 0$ has a root in $\sigma(T)$, in other words,

$$\mu \in p(\sigma(T)) \Leftrightarrow \ j \text{ so that } \beta_j(\mu) \in \sigma(T)\,.$$

Then the desired result follows. □

This theorem can in particular be applied if a given operator can be written as a polynomial of a simpler operator. Indeed we will generalize this result in Theorem 9.14.

As a final result of this section, we prove that there is the following close connection between the spectrum of an operator and the spectrum of its dual operator.

Theorem 6.12. *Let X be a complex Banach space, let $T \in \mathcal{B}(X)$ and let $T^* \in \mathcal{B}(X^*)$ be the corresponding dual operator of T. Then*

(a) $\sigma(T) = \sigma(T^*)$.

(b) The point, residual, and continuous spectra of T and T^* are related by

$$\begin{aligned}
&\sigma_p(T^*) \subset \sigma_p(T) \cup \sigma_r(T), && \sigma_p(T) \subset \sigma_p(T^*) \cup \sigma_r(T^*)\ ;\\
&\sigma_r(T^*) \subset \sigma_p(T) \cup \sigma_c(T), && \sigma_r(T) \subset \sigma_p(T^*)\ ;\\
&\sigma_c(T^*) \subset \sigma_c(T), && \sigma_c(T) \subset \sigma_r(T^*) \cup \sigma_c(T^*)\ .
\end{aligned}$$

(c) If X is reflexive, then $\sigma_c(A^*) = \sigma_c(A)$ and

$$\begin{aligned}
&\sigma_p(A^*) \subset \sigma_p(A) \cup \sigma_r(A), && \sigma_p(A) \subset \sigma_p(A^*) \cup \sigma_r(A^*)\ ;\\
&\sigma_r(A^*) \subset \sigma_p(A), && \sigma_r(A) \subset \sigma_p(A^*)\ .
\end{aligned}$$

(d) $\sigma(T) = \sigma_{ap}(T) \cup \sigma_p(T^*)$.

Proof. To prove part (a), notice that for any $\lambda \in \mathbb{C}$, $(\lambda I_X - T)^* = \lambda I_{X^*} - T^*$, and then (a) follows from Theorem 5.29.

To prove part (b), note tice that by Theorem 5.28, we have

$$\overline{R(\lambda I_X - T)} = {}^{\perp}N(\lambda I_{X^*} - T^*)\,,\ \overline{R(\lambda I_{X^*} - T^*)}^{\sigma(X^*,X)} = N(\lambda I_X - T)^{\perp}\,.$$

Assume first that $\lambda \in \sigma_p(T^*)$. Then $N(\lambda I_{X^*} - T^*) \neq \{0\}$, so $\lambda I_X - T$ does not have a dense image, and hence $\lambda \in \sigma_p(T) \cup \sigma_r(T)$. Next assume $\lambda \in \sigma_r(T^*)$. Then $N(\lambda I_{X^*} - T^*) = \{0\}$, hence $\lambda I_X - T$ has a dense image, and hence $\lambda \in \sigma_p(T) \cup \sigma_c(T)$. Third, assume $\lambda \in \sigma_c(T^*)$. Then $\lambda I_{X^*} - T^*$ is injective and has a dense image and therefore also has a weak-star dense image. Thus $\lambda I_X - T$ is injective and has a dense image, so $\lambda \in \sigma_c(T)$. It follows from these three inclusions that $\sigma_p(T)$ is disjoint from $\sigma_c(T^*)$, that $\sigma_c(T)$ is disjoint from $\sigma_p(T^*)$, and that $\sigma_r(T)$ is disjoint from $\sigma_r(T^*) \cup \sigma_c(T^*)$. This proves part (b).

To prove part (c) observe that in the reflexive case a linear subspace of X^* is weak-star dense if and only if it is dense (Theorem 5.11). Hence $\sigma_c(T) = \sigma_c(T^*)$ whenever X is reflexive. With this understood, the remaining assertions of part (c) follow directly from part (b).

To prove part (d), observe that $\sigma(T) \backslash \sigma_{ap}(T) \subset \sigma_r(T)$. In fact, by (6.1), if $\lambda \in \sigma(T) \backslash \sigma_{ap}(T)$, then there exists $\delta > 0$ with

$$\|(\lambda I - T)x\| \geq \delta \|x\| \quad \text{for all } x \in X .$$

Then it follows from Exercise 4.1 that $\lambda \in \sigma_r(T)$, and the desired result follows from (b). $\square$

We now give some examples.

Example 6.5 (Spectrum of the Shift Operators)**.** Let X be the Hilbert space $\ell^2 = \ell^2(\mathbb{N}, \mathbb{C})$, and define the operators $A, B : \ell^2 \to \ell^2$ by

$$Ax := (x_2, x_3, x_4, \ldots), \quad Bx := (0, x_1, x_2, x_3, \ldots)$$

for $x = (x_i)_{i \in \mathbb{N}} \in \ell^2$. Then

$$\sigma(A) = \sigma(B) = D$$

is the closed unit disc in $\mathbb{C}$ and

$$\begin{array}{lll} \sigma_p(A) = \mathrm{int}(D), & \sigma_r(A) = \emptyset, & \sigma_c(A) = S^1 . \\ \sigma_p(B) = \emptyset, & \sigma_r(B) = \mathrm{int}(D), & \sigma_c(B) = S^1 . \end{array}$$

Example 6.6. Let $X = \ell^2(\mathbb{N}, \mathbb{C})$ and let $(\lambda_i)_{i\in\mathbb{N}}$ be a bounded sequence of complex numbers. Define the bounded linear operator $A : \ell^2 \to \ell^2$ by

$$Ax := (\lambda_i x_i)_{i\in\mathbb{N}} \qquad \text{for } x = (x_i)_{i\in\mathbb{N}} \in \ell^2 .$$

Then

$$\sigma(A) = \overline{\{\lambda_i \mid i \in \mathbb{N}\}}, \quad \sigma_p(A) = \{\lambda_i \mid i \in \mathbb{N}\}, \quad \sigma_r(A) = \emptyset .$$

Thus *every nonempty compact subset of* $\mathbb{C}$ *is the spectrum of a bounded linear operator on an infinite-dimensional Hilbert space.*

Example 6.7. Let X be a complex Banach space and $P \in \mathcal{B}(X)$ is a projection, i.e., $P^2 = P$. Suppose P is not trivial, that is $P \neq 0$ and $P \neq I$. Then

$$\sigma(P) = \sigma_p(P) = \{0, 1\} .$$

Clearly $\{0, 1\} \subset \sigma_p(P)$. It suffices to show that for $\lambda \in \mathbb{C}\backslash\{0, 1\}$, $\lambda \in \varrho(P)$. Firstly, we claim that $N(\lambda I - P) = \{0\}$. If not, there exists $x \neq 0$ so that $Px = \lambda x \neq 0$. Hence $P^2x = Px = \lambda Px$, and we deduce that $\lambda x = 1$, which is a contradiction. Secondly, we show that $R(\lambda I - P) = X$, then the desired result follows from the Banach inverse operator theorem. Take any $y \in X$, if $\lambda x - Px = y$, then $\lambda Px - Px = Py$. So $Px = (\lambda - 1)^{-1}Py$ and

$$x = \frac{1}{\lambda}(y - Px) = \frac{1}{\lambda}\left(y - \frac{1}{\lambda - 1}Py\right) .$$

Example 6.8. Let A be a self-adjoint operator in a Hilbert space H (See Example 4.8). Then the resolvent set $\varrho(A)$ of A comprises all the complex numbers λ with $\mathrm{Im}(\lambda) \neq 0$, and the resolvent $R(\lambda; A)$ is a bounded linear operator with the estimate

$$\|R(\lambda; A)\| \leq \frac{1}{|\,\mathrm{Im}(\lambda)|} . \tag{6.4}$$

We now prove this. If $x \in H$ then $\langle Ax, x\rangle$ is real since $\langle Ax, x\rangle = \langle x, Ax\rangle = \overline{\langle Ax, x\rangle}$. Therefore, we have

$$\|(\lambda I - A)x\|^2 = \|(\mathrm{Re}(\lambda)I - A)x\|^2 + |\,\mathrm{Im}(\lambda)|^2\|x\|^2 \,.$$

As a consequence,

$$\|(\lambda I - A)x\| \geq |\,\mathrm{Im}(\lambda)| \cdot \|x\|, x \in H \,.$$

Hence the inverse $(\lambda I - A)^{-1}$ exists and is continuous if $\mathrm{Im}(\lambda) \neq 0$.

Moreover, the range $R(\lambda I - A)$ is dense in X if $\mathrm{Im}(\lambda) \neq 0$. If otherwise, there would exist a $y \neq 0$ orthogonal to $R(\lambda I - A)$, i.e., $\langle(\lambda I - A)x, y\rangle = 0$ for all $x \in H$ and so $\langle x, (\bar{\lambda} I - A)y\rangle = 0$ for all $x \in H$. We must have $(\bar{\lambda} I - A)y = 0$, that is, $Ay = \bar{\lambda} y$, contrary to the reality of the value $\langle Ay, y\rangle$. Therefore, by Theorem 6.1, we see that, for any complex number λ with $\mathrm{Im}(\lambda) \neq 0$, the resolvent $R(\lambda; A)$ is a bounded linear operator with the estimate (6.4).

6.3 Regular Points and Defect Numbers

In this section we always assume that H is a complex Hilbert space. Let $(T, D(T))$ be a linear operator in H.

Definition 6.3. $\lambda \in \mathbb{C}$ is called a *regular point* for T if there exists a number $c_\lambda > 0$ such that

$$\|(T - \lambda I)x\| \geq c_\lambda \|x\| \quad \text{for all } x \in D(T) \,. \tag{6.5}$$

The *regularity domain* of T, denoted by $\pi(T)$[1], consists of all the regular points of T.

Clearly, $\lambda \in \pi(T)$ if and only if $T - \lambda I$ has a bounded inverse $(T - \lambda I)^{-1}$ defined on $R(T - \lambda I)$. In this case inequality (6.5) holds with $c_\lambda = \left\|(T - \lambda I)^{-1}\right\|^{-1}$. Moreover, one can see that

$$\varrho(T) \subset \pi(T) \subset \varrho(T) \cup \sigma_r(T) \,.$$

[1]There is no unique symbol for the regularity domain of an operator in the literature.

Proposition 6.13. *Show that* $\lambda_0 \in \pi(T), \lambda \in \mathbb{C}$, *and* $|\lambda - \lambda_0| < c_{\lambda_0}$, *where* c_{λ_0} *is a constant satisfying* (6.5) *for* λ_0, *then* $\lambda \in \pi(T)$. *Particularly,* $\pi(T)$ *is an open subset of* $\mathbb{C}$.

Proof. In fact for $x \in D(T)$,

$$\begin{aligned}\|Tx - \lambda x\| &= \|Tx - \lambda_0 x + \lambda_0 x - \lambda x\| \\ &\geq \|Tx - \lambda_0 x\| - |\lambda - \lambda_0|\|x\| \geq (c_{\lambda_0} - |\lambda - \lambda_0|)\|x\| .\end{aligned}$$

Thus $\lambda \in \pi(T)$ as desired. $\square$

Recall that the dimension of a Hilbert space $\mathcal{H}$, denoted by $\dim \mathcal{H}$, is defined by the cardinality of an orthonormal basis of $\mathcal{H}$.

Definition 6.4. For $\lambda \in \pi(T)$, we call the linear subspace $R(T-\lambda I)^{\perp}$ of $\mathcal{H}$ the *deficiency subspace* of T at λ and its dimension $d_\lambda(T) := \dim R(T-\lambda I)^{\perp}$ the *defect number* of T at λ.

Exercise 6.2. Let $(T, D(T))$ be a closed operator in H. Show that (i) $\lambda \in \varrho(T)$ if and only if $\lambda \in \pi(T)$ and $d_\lambda(T) = 0$; (ii) if $\lambda \in \pi(T)$ and $d_\lambda(T) > 0$, then $\lambda \in \sigma_r(T)$.

Deficiency spaces and defect numbers will play a crucial role in the theory of self-adjoint extensions of symmetric operators developed in Chapter 10.

Proposition 6.14. *Let* $(T, D(T))$ *be a closable linear operator in* $\mathcal{H}$. *Then we have* $\pi(\bar{T}) = \pi(T)$, *and for each* $\lambda \in \pi(T)$, $R(\bar{T} - \lambda I) = \overline{R(T - \lambda I)}$ *and* $d_\lambda(\bar{T}) = d_\lambda(T)$.

Proof. For each $x \in D(\overline{T})$, there exists a sequence (x_n) in $D(T)$ so that $x_n \to x$ and $Tx_n \to \overline{T}x$. Thus $\overline{T}x - \lambda x = \lim_{n\to\infty} Tx_n - \lambda x_n$ belongs to $\overline{R(T - \lambda I)}$. Hence we have $R(\bar{T} - \lambda I) \subset \overline{R(T - \lambda I)}$. Besides,

$$\|\overline{T}x - \lambda x\| = \lim_{n\to\infty} \|\overline{T}x_n - \lambda x_n\| \geq \lim_{n\to\infty} c_\lambda \|x_n\| = c_\lambda \|x\| .$$

Thus $\lambda \in \pi(T)$.

Next we show that $\overline{R(T-\lambda I)} \subset R(\bar{T}-\lambda I)$. Take y in $\overline{R(T-\lambda I)}$, then there exists (x_n) in $D(T)$ so that $Tx_n - \lambda x_n$ converges to y. By (6.5) thus (x_n) is a cauchy sequence in H. Let $x_n \to x$ then $Tx_n \to y + \lambda x$. Since T is closable, we have $x \in D(\overline{T})$ and $\overline{T}x = y + \lambda x$. Thus $y \in R(\bar{T}-\lambda I)$.

Observe that

$$\begin{aligned} d_\lambda(\bar{T}) &= \dim R(\bar{T}-\lambda I)^\perp = \dim \overline{R(T-\lambda I)}^\perp \\ &= \dim R(T-\lambda I)^\perp = d_\lambda(T)\,. \end{aligned}$$

We are done. □

The following technical lemma is needed in the proof of the next proposition.

Lemma 6.15. *If $\mathcal{F}$ and $\mathcal{G}$ are closed linear subspaces of a Hilbert space $\mathcal{H}$ such that* $\dim \mathcal{F} < \dim \mathcal{G}$, *then there exists a nonzero vector $y \in \mathcal{G} \cap \mathcal{F}^\perp$*

Proof. In this proof we denote by $|M|$ the cardinality of a set M. First, we suppose that $k = \dim \mathcal{F}$ is finite. We take a $(k+1)$ -dimensional subspace $\mathcal{G}_0$ of $\mathcal{G}$ and define the mapping $\Phi : \mathcal{G}_0 \to \mathcal{F}$ by $\Phi(x) = Px$, where P is the projection of $\mathcal{H}$ onto $\mathcal{F}$. If Φ would be injective, then $k+1 = \dim \mathcal{G}_0 = \dim \Phi(\mathcal{G}_0) \leq \dim \mathcal{F} = k$, which is a contradiction. Hence, there is a nonzero vector $y \in \mathcal{N}(\Phi)$. Clearly, $y \in \mathcal{G} \cap \mathcal{F}^\perp$. Now suppose that $\dim \mathcal{F}$ is infinite. Let $\{f_k : k \in K\}$ and $\{g_l : l \in L\}$ be orthonormal bases of $\mathcal{F}$ and $\mathcal{G}$, respectively. Set $L_k := \{l \in L : \langle f_k, g_l \rangle \neq 0\}$ for $k \in K$ and $L' = \bigcup_{k \in K} L_k$. Since each set L_k is at most countable and $\dim \mathcal{F} = |K|$ is infinite we have $|L'| \leq |K||\mathbb{N}| = |K|$. Since $|K| = \dim \mathcal{F} < \dim \mathcal{G} = |L|$ by assumption, we deduce that $L' \neq L$. Each vector g_l with $l \in L \backslash L'$ is orthogonal to all $f_k, k \in K$ and hence, it belongs to $\mathcal{G} \cap \mathcal{F}^\perp$. □

The next theorem is a classical result of M.A. Krasnosel'skii and M.G. Krein.

Theorem 6.16. *Let $(T, D(T))$ be a closable linear operator in H. Then the defect number $d_\lambda(T)$ is constant on each connected component of the open set $\pi(T)$.*

Proof. By Proposition 6.14, we can assume without loss of generality that T is closed and then $R(T - \lambda I)$ is closed for all $\lambda \in \pi(T)$.

Take $\lambda_0 \in \pi(T)$. We claim that $d_\lambda(T) = d_{\lambda_0}(T)$ for all $\lambda \in \mathbb{C}$ so that $|\lambda - \lambda_0| < c_{\lambda_0}$. (Note that by Exercise 6.13 we have $\lambda \in \pi(T)$.) Then the mapping $\lambda \mapsto d_\lambda(T)$ is locally constant which implies the desired result.

Assume to the contrary that $d_\lambda(T) \neq d_{\lambda_0}(T)$. First suppose that $d_\lambda(T) < d_{\lambda_0}(T)$. By Lemma 6.15, there exists a nonzero vector

$$y \in R(T - \lambda_0 I)^\perp \cap R(T - \lambda I)$$

Then $y = (T - \lambda I)x$ for some nonzero $x \in \mathcal{D}(T)$ and

$$\langle (T - \lambda I)x, (T - \lambda_0 I)\, x\rangle = 0 \tag{6.6}$$

Equation (6.6) is symmetric in λ and λ_0, so it holds also when $d_{\lambda_0}(T) < d_\lambda(T)$.Using (6.6) we derive

$$\begin{aligned} \|(T - \lambda_0 I)\, x\|^2 &= \langle (T - \lambda I)x + (\lambda - \lambda_0)\, x, (T - \lambda_0 I)\, x\rangle \\ &\leq |\lambda - \lambda_0|\, \|x\|\, \|(T - \lambda_0 I)\, x\| \end{aligned}$$

Thus, $\|(T - \lambda_0 I)\, x\| \leq |\lambda - \lambda_0|\, \|x\|$. Since $x \neq 0$ and $|\lambda - \lambda_0| < c_{\lambda_0}$, we obtain

$$|\lambda - \lambda_0|\, \|x\| < c_{\lambda_0}\|x\| \leq \|(T - \lambda_0 I)\, x\| \leq |\lambda - \lambda_0|\, \|x\|$$

by (6.5) which is a contradiction. Thus, we have proved that $d_\lambda(T) = d_{\lambda_0}(T)$. $\square$

Exercise 6.3. The numerical range of a linear operator T in $\mathcal{H}$ is defined by

$$\Theta(T) = \{\langle Tx, x\rangle : x \in \mathcal{D}(T), \|x\| = 1\}$$

A classical result of F. Hausdorff says that $\Theta(T)$ is a convex set. In general, the set $\Theta(T)$ is neither closed nor open for a bounded or closed operator.

Show that if $\lambda \in \mathbb{C}$ is not in the closure of $\Theta(T)$, then $\lambda \in \pi(T)$.

Chapter 7

Compact Operators

7.1 Fundamentals

One of the most important concepts in the study of bounded linear operators is that of a compact operator. In this section, we always suppose that X and Y are two Banach space.

Definition 7.1. Let X and Y be Banach spaces. Let B_X be the closed unit ball in X. A linear map $K : X \to Y$ is said to be *compact* if $K(B_X)$ is relatively compact in Y; i.e., the closure of $K(B_X)$ is compact in Y.

We will denote by $\mathcal{C}(X, Y)$ all the compact linear operator from X into Y, and we write $\mathcal{C}(X)$ for $\mathcal{C}(X, X)$. Since Y is a complete metric space, the relatively compact subsets of Y are precisely the totally bounded ones. So every compact linear operator is bounded, i.e., $\mathcal{C}(X, Y) \subset \mathcal{B}(X, Y)$. Since the sequential compactness and compactness are equivalent in metric space, the notion of a compact operator can be defined in several equivalent ways, as in the following proposition:

Proposition 7.1. *Let X and Y be Banach spaces. Let $K : X \to Y$ be a bounded linear operator. Then the following are equivalent.*

(a) $K \in \mathcal{C}(X, Y)$*, i.e.,* $K(B_X)$ *is a relatively compact subset of* Y*.*

(b) *If* $S \subset X$ *is bounded, then* $K(S)$ *is relatively compact.*

(c) *If* (x_n) *is a bounded sequence in* X, *then the sequence* (Kx_n) *has a convergent subsequence in* Y*.*

This lemma follows trivially from Theorem A.2 and hence we omit the proof.

Many of the operators that arise in the study of integral equations are compact. This accounts for their importance from the standpoint of applications. They are in some respects as similar to linear operators on finite dimensional spaces as one has any right to expect from operators on infinite-dimensional spaces. As we shall see, these similarities show up particularly strongly in their spectral properties.

Example 7.1. Let Ω be a compact subset of $\mathbb{R}^n$. Let $k(x, y)$ be a $\mathbb{F}$-valued continuous function defined on $\Omega \times \Omega$. Denote by $C(\Omega)$ all the continuous $\mathbb{F}$-valued on Ω. Then $C(\Omega)$ endowed with the supremum norm is a Banach space. Then the integral operator $K : C(\Omega) \to C(\Omega)$ defined by

$$(Kf)(x) = \int_\Omega k(x, y) f(y) dy$$

is a compact linear operator. To see this, clearly K mapping $C(\Omega)$ into $C(\Omega)$ is linear. Denote by B the closed unit ball in $C(\Omega)$. In order that $K(B)$ is relatively compact, by the Ascoli-Arzelà theorem, it suffices to show that $K(B)$ is uniformly bounded and equi-continuous.

Since k is continuous, there exists $M > 0$ with $|k(x, y)| \le M$ for all x, y in Ω. Then $\|Kf\| \le (b - a)M\|f\| \le (b - a)M$ for $f \in B$, so that $K(B)$ is uniformly bounded. Observe that for every $f \in B$, $\|f\| \le 1$,

$$|(Kf)(x_1) - (Kf)(x_2)| \le \int_\Omega |K(x_1, y) - K(x_2, y)|\, dy\,.$$

Since k is uniformly continuous in Ω, $K(B)$ is equi-continuous. Therefore, by the Ascoli-Arzelà theorem, the set $K(S)$ is relatively compact in $C(\Omega)$.

Example 7.2. If $K : X \to Y$ is a bounded linear operator between Banach spaces whose image is a closed infinite-dimensional subspace of Y, then K is not compact. Namely, the image of the closed unit ball in X under K contains an open ball in $R(K)$ by the open mapping theorem, and hence does not have a compact closure by Theorem 1.12.

Proposition 7.2. *Let X,Y and Z be Banach spaces. Then the following holds.*

(a) $\mathcal{C}(X, Y)$ *is a closed subspace of* $\mathcal{B}(X, Y)$*.*

(b) *If* $K \in \mathcal{C}(X, Y)$*, then* $R(K)$ *is separable.*

(c) *Let* $A \in \mathcal{B}(X, Y)$ *and* $B \in \mathcal{C}(X, Y)$ *and one of them is compact, then* $BA \in \mathcal{C}(X, Z)$*. In particular,* $\mathcal{C}(X)$ *is two-side ideal in the algebra* $\mathcal{B}(X)$*.*

(d) *If* $K \in \mathcal{C}(X, Y)$ *and* Z *is a subspace of* X*. Then* $K|_Z \in \mathcal{C}(Z, Y)$*.*

Proof. We prove part (a). Obviously, $\mathcal{C}(X, Y)$ is a linear subspace of $\mathcal{B}(X, Y)$. To show that $\mathcal{C}(X, Y)$ is closed, suppose $\{K_n\}$ is a convergent sequence in $\mathcal{C}(X, Y)$, and $K_n \to K$ in operator norm. Let $\{x_k\}$ be a bounded sequence in B_X. By the compact property of each K_n, we can choose, by the *diagonal method*, a subsequence $\{\hat{x}_k\}$ of $\{x_k\}$ such that $\{K_n \hat{x}_k\}$ converges for each fixed n. Then we shall show that $\{K\hat{x}_k\}$ is a Cauchy sequence, and hence is convergent. For fixed $n \geq 1$ and any $k, m \geq 1$, we have

$$\begin{aligned}\|K\hat{x}_m - K\hat{x}_k\| &\leq \|K\hat{x}_m - K_n\hat{x}_k\| + \|K_n\hat{x}_m - K_n\hat{x}_k\| + \|K_n\hat{x}_k - K\hat{x}_k\| \\ &\leq \|K - K_n\| + \|K_n\hat{x}_m - K_n\hat{x}_k\| + \|K_n - K\| \ .\end{aligned}$$

and so

$$\limsup_{m,k\to\infty} \|K\hat{x}_m - K\hat{x}_k\| \leq 2\,\|K - K_n\|$$

Since n is arbitrary, $\{K\hat{x}_k\}$ is a Cauchy sequence in the Y.

We prove part (b). Observe that

$$R(K) = K(X) = K\left(\bigcup_{n\geq 1} nB_X\right) = \bigcup_{n\geq 1} nK\left(B_X\right) .$$

Since $K(B_X)$ is totally bounded, $K(B_X)$ is separable, and hence $nK(B_X)$ is separable. Thus $R(K)$ is separable.

We prove part (c). Let (x_n) be a bounded sequence in X. If A is compact, then there exists a subsequence (x_{n_k}) such that the sequence (Ax_{n_k}) converges, and so does the subsequence (BAx_{n_k}). If B is compact, then, since the sequence (Ax_n) is bounded, there exists a subsequence (Ax_{n_k}) such that the sequence (BAx_{n_k}) converges. This proves (c).

We prove part (d). Since B_Z is a bounded subset in X and $K|_Z(B_Z) = K(B_Z)$, the desired result follows. □

Theorem 7.3 (Schauder)**.** *Let X and Y be Banach space. Then $K \in \mathcal{C}(X,Y)$ if and only if $K^* \in \mathcal{C}(Y^*, X^*)$.*

Proof. Necessity. Suppose that $K \in \mathcal{C}(X,Y)$. We show that $K^* \in \mathcal{C}(Y^*, X^*)$. Let $\{y_n^*\}$ be a sequence in B_{Y^*}. It suffices to show that $\{K^* y_n^*\}$ has a convergent sequence in X^*.

Recall that B_{Y^*} is sequentially compact relative to $\sigma(Y^*, Y)$ if Y is separable (see Theorem 5.16). If Y is not separable, let $Z = \overline{R(K)}$. Then Z is a separable Banach space. Let $\hat{K} : X \to Z$ defined by $\hat{K}x = Kx$ for all x. Then $\hat{K}^* : Z^* \to X^*$, satisfies that $\hat{K}^*(y^*|_Z) = K^* y^*$ for every $y^* \in Y^*$. Hence, without loss of generality we assume Y is separable and so B_{Y^*} is sequentially compact relative to $\sigma(Y^*, Y)$. Let $\{y_{n_k}^*\}$ be a weak-star convergent subsequence of $\{y_n^*\}$ and $y_{n_k}^* \to y^*$ in $\sigma(Y^*, Y)$. We shall show that $K^* y_{n_k}^* \to K^* y^*$ in norm.

$$\|K^* y_{n_k}^* - K^* y^*\| = \sup_{x\in B_X} |\langle Kx, y_{n_k}^* - y^*\rangle| = \sup_{y\in K(B_X)} |\langle y, y_{n_k}^* - y^*\rangle| .$$

Since $K(B_X)$ is totally bounded in Y, for any $\epsilon > 0$, there exists $\{y_1, \cdots, y_m\} \subset K(B_X)$ so that

$$K(B_X) \subset \bigcup_{j=1}^{m} B_Y(y_j, \epsilon)\,.$$

Since $y^*_{n_k} \to y^*$ in $\sigma(Y^*, Y)$,

$$\sup_{1\le j\le m} |\langle y_j, y^*_{n_k} - y^*\rangle| \to 0 \qquad \text{as } k \to \infty\,.$$

Notice that

$$\|K^* y^*_{n_k} - K^* y^*\| = \sup_{y\in K(B_X)} |\langle y, y^*_{n_k} - y^*\rangle| \le \sup_{1\le j\le m} |\langle y_j, y^*_{n_k} - y^*\rangle| + 2\epsilon\,,$$

we get that

$$\limsup_{k\to\infty} \|K^* y^*_{n_k} - K^* y^*\| \le 2\epsilon\,.$$

Since $\epsilon > 0$ is arbitrary, $\|K^* y^*_{n_k} - K^* y^*\| \to 0$. The desired result follows.

Sufficiency. Conversely, suppose that K^* is compact. Then, by what we have just proved, the bidual operator $K^{**} \in \mathcal{C}(X^{**}, Y^{**})$. Since $K^{**}|_X = K$, this implies that K is compact. $\square$

Exercise 7.1. For $y^*_n \in B_{Y^*}$ consider the continuous function

$$\varphi_n : \overline{K(B_X)} \to \mathbb{F}\,;\; y \to \langle y, y^*_n\rangle\,.$$

Use the Arzel‘a–Ascoli theorem to prove Theorem 7.3.

Completely Continuous Operators. Let X and Y be Banach spaces. A bounded linear operator $K \in \mathcal{B}(X, Y)$ is said to be *completely continuous* if the image of every weakly convergent sequence in X under K converges in the norm topology on Y.

Theorem 7.4. *Let X and Y be Banach spaces. Then the following hold.*

(a) Every compact operator $K \in \mathcal{C}(X, Y)$ is completely continuous.

(b) If $K \in \mathcal{C}(X,Y)$ and X is reflexive. Then K is compact.

Proof. We prove part (a). Assume K is compact and let (x_n) be a sequence in X that converges weakly to $x \in X$. By the PUB, (x_n) is (norm) bounded so there exists a subsequence (x_{n_k}) of (x_n) so that (Kx_{n_k}) converges to $y \in Y$. On the other hand, $K : (X, \sigma(X, X^*)) \to (Y, \sigma(Y, Y^*))$ is continuous, so $Kx_{n_k} \xrightarrow{w} Kx$. Thus $y = Kx$ and $Kx_{n_k} \to Kx$. Note that this argument holds for every subsequence of (x_n), so every subsequence of (Kx_n) has a subsequence converging to Kx. Thus the desired result follows.

We prove part (ii). Assume X is reflexive and K is completely continuous. Let (x_n) be a bounded sequence in X. Since X is reflexive, there exists a weakly convergent subsequence (x_{n_k}) by Theorem 5.24. Let $x \in X$ be the limit of that subsequence. Since K is completely continuous, the sequence (Kx_{n_k}) converges strongly to Kx. Thus K is cpmpact. □

Example 7.3. The hypothesis that X is reflexive cannot be removed in part (b) of Theorem 7.4. For example a sequence in ℓ^1 converges weakly if and only if it converges strongly (see Theorem 5.19). Hence the identity operator $\mathrm{id} : \ell^1 \to \ell^1$ is completely continuous. However, it is not a compact operator by Theorem 1.12.

Finite Rank Operator. Let X and Y be Banach spaces. A bounded linear operator $T : X \to Y$ is said to be of *finite rank* if its image $R(T)$ is a finite-dimensional subspace of Y. We will denote by $\mathcal{F}(X,Y)$ all the bound linear operator form X into Y with finite rank. Clearly, $\mathcal{F}(X,Y) \subset \mathcal{C}(X,Y)$, and we write $\mathcal{F}(X)$ for $\mathcal{F}(X,X)$.

Example 7.4. Let X and Y be Banach spaces. For each $x^* \in X^*$ and $y \in Y$, define $y \otimes x^*$ by

$$y \otimes x^* : X \to Y \,;\, x \mapsto \langle x, x^* \rangle y \,.$$

Clearly $y \otimes x^*$ is a finite rank operator. In fact $\dim R(y \otimes x^*) = 1$.

Lemma 7.5. *$T \in \mathcal{F}(X,Y)$ if and only if there exists $n \geq 1$, $y_j \in Y$ and $x_j^* \in X^*$ for $j = 1, \cdots, n$ so that*

$$T = \sum_{j=1}^{n} y_j \otimes x_j^*$$

Proof. Sufficiency is trivial, we only show the necessity. Let $n = \dim R(T)$ and let $\{y_j\}$ be a basis of $R(T)$. Then for any $x \in X$, there exists scalars $\{l_j(x)\}$ so that

$$Tx = \sum_{j=1}^{n} l_j(x) y_j \, .$$

Let x_j^* defined by $\langle x, x_j^* \rangle \mapsto l(x_j)$, then it's easy to check that $x_j^* \in X^*$ and so the desired result follows. $\square$

Clearly, the limit of a sequence of finite rank operators in the norm topology is a compact operator. It is a natural question to ask whether, conversely, every compact operator can be approximated in the norm topology by a sequence of finite rank operators; i.e.,

$$\overline{\mathcal{F}(X,Y)} = \mathcal{C}(X,Y) \, . \tag{7.1}$$

The answer to this question was an open problem in functional analysis for many years. It was eventually shown that the answer depends on the Banach space in question. For example, if $Y = H$ is a Hilbert space, then (7.1) holds.

To see this, note that for any given $K \in \mathcal{C}(X,H)$, $\overline{R(K)}$ is a separable Hilbert space. Let $\{e_n\}$ be a Hilbert basis of $\overline{R(K)}$. Let P_n be the orthogonal projection from H onto $\text{span}\{e_1, \cdots, e_n\}$. Define

$$K_n \coloneqq P_n \circ K \quad \text{for all } n \geq 1 \, .$$

Clearly $K_n \to K$ in the strong operator topology. W show that

$$\|K_n \to K\| \to 0 \, .$$

Given any $\epsilon > 0$, since $K(B_X)$ is totally bounded, there exists $x_1, \cdots, x_m$ in B_X so that for every $x \in B_X$, there exists x_j so that $\|x - x_j\| \leq \epsilon$, and

$$\begin{aligned}\|K_n x - Kx\| &\leq (\|K_n x - K_n x_j\| + \|K_n x_j - Kx_j\| + \|Kx_j - Kx\|) \\ &\leq 2\epsilon + \sup_{1 \leq j \leq m} \|K_n x_j - Kx_j\| .\end{aligned}$$

Then the desired result follows.

Let Y be a Banach space. If for every Banach space X, (7.1) holds, we say that Y has the *approximation property.* Every Hilbert space has this property. The follows lemma due to Grothendieck implies that the whether (7.1) hols or not depends only on Y.

Lemma 7.6. *Let Y be a Banach space. Then Y has the approximation property if and only if for every compact subset $C \subset Y$ and every $\epsilon > 0$ there is $T \in \mathcal{F}(Y)$ such that $\|y - Ty\| \leq \epsilon$ for all $y \in C$.*

Remark 7.1. The above condition can be restated as: the identity operator $I : X \to X$ can be approximated, uniformly on every compact subset C of X, by linear operators of finite rank.

Proof. We prove the sufficiency. Given any Banach space X and any $K \in \mathcal{C}(X, Y)$. Since $\overline{K(B_X)}$ is compact, by assumption, for $\epsilon > 0$, there exists $T \in \mathcal{F}(Y)$ so that $\|Kx - TKx\| \leq \epsilon$ for all $x \in B_X$, and hence $\|K - TK\| \leq \epsilon$. Since $TK \in \mathcal{F}(X, Y)$ and ϵ is arbitrary, $K \in \overline{\mathcal{F}(X, Y)}$. The proof of necessity is not easy, see here. □

In fact, using Theorem 4.19 and the same argument we can show that:

Theorem 7.7. *Every Separable Banach space with a Schauder basis has the approximation property.*

Example 7.5. Fix a number $1 \leq p \leq \infty$ and a bounded sequence of scalars $\lambda = (\lambda_j)$. For $j \in \mathbb{N}$ let $e_j := (\delta_{ij}) \in \ell^p$. Clearly (e_j) is a Schauder basis in

ℓ^p. Define the bounded linear operator $K_\lambda : \ell^p \to \ell^p$ by

$$K_\lambda x := (\lambda_j x_j) \qquad \text{for } x = (x_j) \in \ell^p$$

Then

$$K_\lambda \in \mathcal{C}(\ell^p) \Leftrightarrow \lim_{j\to\infty} \lambda_j = 0\,.$$

The condition $\lambda_j \to 0$ is necessary for compactness because, if there exist a constant $\delta > 0$ and a sequence $1 \leq n_1 < n_2 < n_3 < \cdots$ such that $|\lambda_{n_k}| \geq \delta$ for all $k \in \mathbb{N}$, then the sequence $Ke_{n_k} = \lambda_{n_k} e_{n_k}, k \in \mathbb{N}$, in ℓ^p has no convergent subsequence. The condition $\lambda_j \to 0$ implies compactness because then K can be approximated by a sequence of finite rank operators in the norm topology.

Example 7.6. If $(X, \mathcal{F}, \mu)$ is a measure space and $k \in L^2(X \times X, \mathcal{F} \times \mathcal{F}, \mu \times \mu)$ then

$$(Kf)(x) = \int k(x,y) f(y) \mu(dy) \tag{7.2}$$

is a compact operator and $\|K\| \leq \|k\|_2$.

The following lemma is useful for proving this proposition. The lemma is intuitive so the proof is omitted.

Lemma. *If $\{e_i\}_{i\in I}$ is a Hilbert basis for $L^2(X, \mathcal{F}, \mu)$ and*

$$\phi_{ij}(x,y) = e_j(x)\overline{e_i(y)} \qquad \textit{for } i,j \in I \textit{ and } x,y \in X\,,$$

then $\{\phi_{ij}\}_{i,j\in I}$ is an orthonormal set in $L^2(X \times X, \mathcal{F} \times \mathcal{F}, \mu \times \mu)$. If k and K are as in (7.2)*, then $\langle k, \phi_{ij}\rangle = \langle Ke_j, e_i\rangle$.*

It follows from Example 1.17 that K is a bounded linear operator on L^2 with $\|K\| \leq \|k\|_2$. Now let $\{e_i\}$ be a basis for $L^2(\mu)$ and define ϕ_{ij} as in the preceding lemma. Thus by Bessel inequality (see Theorem 2.14),

$$\|k\|^2 \geq \sum_{i,j} |\langle k, \phi_{ij}\rangle|^2 = \sum_{i,j} |\langle Ke_j, e_i\rangle|^2\ .$$

Since $k \in L^2(\mu\times\mu)$, there are at most a countable number of i and j such that $\langle k, \phi_{ij}\rangle \neq 0$; denote these by $\{\psi_{km} : 1 \le k, m < \infty\}$. Note that $\langle Ke_j, e_i\rangle = 0$, unless $\phi_{ij} \in \{\psi_{km}\}$. Wiout loss of generality, let $\psi_{km}(x,y) = e_k(x)\overline{e_m(y)}$. Let P_n be the orthogonal projection onto span $\{e_k : 1 \le k \le n\}$, and put

$$K_n = KP_n + P_nK - P_nKP_n\,;$$

so K_n is a finite rank operator. We will show that $\|K - K_n\| \to 0$ as $n \to \infty$, thus showing that K is compact.

Let $f \in L^2(\mu)$ with $\|f\|^2 \le 1$; so $f = \sum_j \alpha_j e_j$. Hence

$$\begin{aligned}
\|Kf - K_nf\|^2 &= \sum_i |\langle Kf - K_nf, e_i\rangle|^2 \\
&= \sum_i \left| \sum_j \alpha_j \langle (K - K_n)\, e_j, e_i\rangle \right|^2 \\
&= \sum_k \left| \sum_m \alpha_m \langle (K - K_n)\, e_m, e_k\rangle \right|^2 \\
&\le \sum_k \left[\sum_m |\alpha_m|^2\right] \left[\sum_m |\langle (K - K_n)\, e_m, e_k\rangle|^2\right] \\
&= \|f\|^2 \sum_{k,m} |\langle (K - K_n)\, e_m, e_k\rangle|^2 \;,
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{k,m} |\langle (K - K_n)\, e_m, e_k\rangle|^2 \\
&= \sum_{k,m} |\,\langle Ke_m, e_k\rangle - \langle KP_ne_m, P_ne_k\rangle - \langle KP_ne_m, P_ne_k\rangle + \langle KP_ne_m, P_ne_k\rangle|^2 \\
&= \sum_{k,m>n} |\langle Ke_m, e_k\rangle|^2 = \sum_{k,m>n} |\langle k, \psi_{km}\rangle|^2 \;.
\end{aligned}$$

Since $\sum_{k,m} |\langle k, \psi_{km}\rangle|^2 < \infty, n$ can be chosen sufficiently large such that for any $\epsilon > 0$ this last sum will be smaller than ϵ^2. Thus $\|K - K_n\| \to 0$.

7.2 Riesz-Fredholm Theory

Let X be a Banach space over $\mathbb{F}$ and $T \in \mathcal{C}(X)$, i.e., T is a compact linear operator from X into X. Let λ be a scalar (we always set $\lambda \neq 0$). In this section, we shall discuss the following equation

$$(\lambda I - T)x = y\,.$$

We shall denote $T_\lambda \coloneqq \lambda I - T$.

Fredholm studied the following integral equations. Let $K(x,y) \in C([0,1]\times[0,1])$ be an integral kernel, and consider the inhomogeneous equation and homogeneous equation

$$x(t) - \int_0^1 K(t,s)x(s)ds = y(t)\,, \tag{7.3}$$

with it's dual equation

$$f(t) - \int_0^1 K(s,t)f(s)ds = g(t)\,, \tag{7.4}$$

where $x, y, f, g \in L^2[0,1] = L^2([0,1], \mathbb{R})$. He got the following result:

Result 1: About the equation (7.3)((7.4)), exactly one of the following holds:

(i) For every $y\,(g) \in L^2[0,1]$, the equation (7.3)((7.4)) has a unique solution.

(ii) For $y\,(g) = 0$ the equation (7.3)((7.4)) has a non-trivial solution.

Result 2: If (7.3) falls into the first case, then so is (7.4); If (7.3) falls into the second case, then so is (7.4). Besides, in the second case, the homogeneous equations of (7.3) and (7.4) have the same finite numbers of the linear independent solutions.

Result 3: The equation (7.3) has a solution if and only if

$$\int_0^1 f(t)y(t)dt = 0$$

for all f satisfying $f(t) - \int_0^1 K(s,t)f(s)ds = 0$. The equation (7.4) has a solution if and only if

$$\int_0^1 g(t)x(t)dt = 0$$

for all x satisfying $x(t) - \int_0^1 K(t,s)x(s)ds = 0$.

As we have pointed in Example 7.6, if we let T be the integral operator on $L^2[0,1]$ with kernel K, then T is a compact operator on $L^2[0,1]$, and the previous equation is exactly

$$(I-T)x = y$$

and

$$(I-T)^*f = g\,.$$

We will extend the above results to the general situation.

Lemma 7.8 (F.Riesz). *Let X be a Banach space over $\mathbb{F}$ and $T \in \mathcal{C}(X)$. Then for any $\lambda \in \mathbb{F}\backslash\{0\}$, $R(T_\lambda) = R(\lambda I - T)$ is closed.*

Naturally, set $\lambda x_n - Tx_n \to y$, and we will check that if we can find x so that $\lambda x - Tx = y$. If we suppose that $\{x_n\}$ is (norm) bounded in X, then without loss of generality we assume that $\{Tx_n\}$ converges, and since $\lambda x_n = y - Tx_n$, $\lambda \neq 0$, so $\{x_n\}$ converges. Clearly, let $x_n \to x$, then $T_\lambda x_n \to T_\lambda x = y$ and the desired result holds.

Proof. Let $\widetilde{T_\lambda} : X/N(T_\lambda) \to R(T_\lambda)$ be given by

$$\widetilde{T_\lambda}\widetilde{x} \coloneqq T_\lambda x \quad \text{for every } \widetilde{x} \in X/N(T_\lambda)\,.$$

By Example 1.20, $\widetilde{T_\lambda}$ is a well-defined continuous linear operator. Moreover, $\widetilde{T_\lambda}$ is bijective. So we only need to show that $\widetilde{T_\lambda}^{-1}$ is continuous, then we can suppose that the sequence $\{x_n\}$ is bounded in norm. However, indeed, $\widetilde{T_\lambda}$ is a linear homomorphism between $X/N(T_\lambda)$ and $R(T_\lambda)$. Since $X/N(T_\lambda)$ is a Banach space, then $R(T_\lambda)$ muse be a closed subspace in Y.

If $\widetilde{T_\lambda}^{-1}$ is not continuous, by Exercise 4.1, there exists a sequence $\{\widetilde{x}_n\}$ in $X/N(T_\lambda)$ so that $\|\widetilde{x}_n\| = 1$ and $\|\widetilde{T}_\lambda x_n\| = \|\lambda x_n - Tx_n\| \to 0$. Without loss of generality, suppose $\{x_n\}$ is (norm) bounded and $Tx_n \to z$. Then $\lambda x_n \to z$, so $Tx_n \to \frac{1}{\lambda}Tz$, and hence $T_\lambda z = (\lambda z - Tz) = 0$. So $\widetilde{z} = \widetilde{0}$. But $x_n \to z$ implies $\widetilde{x}_n \to \widetilde{z} = \widetilde{0}$, which contradicts to $\|\widetilde{x}_n\| = 1$. We now complete the proof. $\square$

Theorem 7.9 (Fredholm Alternative)**.** *Let X be a Banach space over $\mathbb{F}$ and $T \in \mathcal{C}(X)$. Let $\lambda \in \mathbb{F}\backslash\{0\}$. Then $N(T_\lambda) = \{0\}$ if and only if $R(T_\lambda) = X$.*

Remark 7.2. The Fredholm alternative claim that either there is a non-zero solution $x \in X$ to the equation $Tx = \lambda x$ or the operator $T_\lambda = \lambda I - T$ has a continuous inverse $R(\lambda, T) = (\lambda I - T)^{-1}$ on X.

Proof. The strategy here is to try to contradict the compactness of T by exhibiting a bounded set whose image under T is not totally bounded.

Firstly, we assume that $N(T_\lambda) = \{0\}$. For each $n \in \mathbb{N}$, let

$$X_n = R(T_\lambda^n) = R((\lambda I - T)^n)\,,$$

and let $X_0 = X$. Obviously $X_{n+1} \subset X_n$ for $n \geq 0$. Suppose for contradiction that $X_1 = R(T_\lambda) \subsetneqq X$, then we have $X_{n+1} \subsetneqq X_n$ for every $n \geq 0$, since $N(T_\lambda) = \{0\}$ impleis $N(T_\lambda^n) = \{0\}$. Moreover, by Lemma 7.8, X_{n+1} is a closed of X_n, since $(\lambda I - T)^n = \lambda^n I + \sum_{j=1}^n \binom{n}{j}(-T)^j = \lambda^n I -$ (a compact operator).

$$X_0 \supsetneqq X_1 \supsetneqq X_2 \supsetneqq \cdots \supsetneqq X_n \supsetneqq X_{n+1} \supsetneqq \cdots$$

By Riesz's lemma (Lemma 1.11), there exists $x_n \in X_n$ with $\|x_n\| = 1$ and $d(x_n, X_{n+1}) > 1/2$.

Now let $n \geq 0$ and $p \geq 1$. By construction, x_{n+p}, $T_\lambda x_{n+p}$, $T_\lambda x_n$ all lie in X_{n+1}, and thus

$$\begin{aligned} Tx_n - Tx_{n+p} &= (\lambda I - T_\lambda)(x_n - x_{n+p}) \\ &= \lambda x_n + (-\lambda x_{n+p} - T_\lambda x_n + T_\lambda x_{n+p}) \in \lambda x_n + X_{n+1}\,. \end{aligned}$$

Since x_n lies at a distance at least $1/2$ from X_{n+1}, we conclude the separation proeprty

$$\|Tx_{n+p} - Tx_n\| \geq \frac{|\lambda|}{2}$$

But this implies that the sequence $\{Tx_n\}$ is not totally bounded, contradicting the compactness of T. Thus $N(T_\lambda) = \{0\}$ implies $R(T_\lambda) = X$.

Secondly, we assume that $R(T_\lambda) = X$. For each $n \in \mathbb{N}$, let

$$Y_n = N(T_\lambda^n) = N((\lambda I - T)^n)\,,$$

and let $Y_0 = \{0\}$. Obviously $Y_n \subset Y_{n+1}$ for $n \geq 0$. Suppose for contradiction that $\{0\} \subsetneqq N(T_\lambda) = Y_1$, then we have $Y_n \subsetneqq Y_{n+1}$ for every $n \geq 0$, since $R(T_\lambda) = X$ impleis $R(T_\lambda^n) = X$. Moreover, Y_{n+1} is a closed of Y_n.

$$Y_0 \subsetneqq Y_1 \subsetneqq Y_2 \subsetneqq \cdots \subsetneqq Y_n \subsetneqq Y_{n+1} \subsetneqq \cdots$$

By Riesz's lemma (Lemma 1.11), there exists $y_n \in Y_n$ with $\|y_n\| = 1$ and $d(y_n, Y_{n-1}) > 1/2$.

Now let $n \geq 0$ and $p \geq 1$. By construction, y_n, $T_\lambda y_{n+p}$, $T_\lambda y_n$ all lie in Y_{n+p-1}, and thus

$$\begin{aligned}
Ty_{n+p} - Ty_n &= (\lambda I - T_\lambda)(y_{n+p} - y_n) \\
&= \lambda y_{n+p} + (-\lambda y_n - T_\lambda y_{n+p} + T_\lambda y_n) \in \lambda y_{n+p} + Y_{n+p-1}\,.
\end{aligned}$$

Since y_{n+p} lies at a distance at least $1/2$ from Y_{n+p-1}, we conclude the separation proeprty

$$\|Ty_{n+p} - Ty_n\| \geq \frac{|\lambda|}{2}$$

But this implies that the sequence $\{Ty_n\}$ is not totally bounded, contradicting the compactness of T. Thus $R(T_\lambda) = X$ implies $N(T_\lambda) = \{0\}$. $\square$

Remark 7.3. A hypothesis such as compactness is necessary; the shift operator U on the complex Hilbert space $\ell^2(\mathbb{N}, \mathbb{C})$, for instance, has no eigenfunctions, but $\lambda I - U$ is not invertible for any unit complex number λ (see

Example 6.5). The claim is also false when $\lambda = 0$; consider for instance the multiplication operator $(Tx)(n) := x(n)/n$ on $\ell^2(\mathbb{N}, \mathbb{C})$, which is compact and has no eigenvalue at zero, but is not invertible (see Example 6.6).

Theorem 7.10 (Riesz-Fredholm)**.** *Let X be a Banach space over $\mathbb{F}$ and $T \in \mathcal{C}(X)$. Denote by I the identity operator on X. Then for every $\lambda \in \mathbb{F}\backslash\{0\}$,*

(a) $N(\lambda I - T) = \{0\}$ *if and only if* $R(\lambda I - T) = X$.

(b) $R(\lambda I - T) = {}^{\perp}N(\lambda I^* - T^*)$, *and* $R(\lambda I^* - T^*) = N(\lambda I - T)^{\perp}$.

(c) $\sigma(T) = \sigma(T^*)$ *and* $\dim N(\lambda I - T) = \dim N(\lambda I^* - T^*) < \infty$. *Moreover,* $\dim N(\lambda I - T) = \operatorname{codim} R(\lambda I - T)$.

Proof. As before, denote $T_\lambda = \lambda I - T$. Part (a) is the Riesz-Fredholm alternative. To show part (b), observing that $R(T_\lambda)$ is closed by Lemma 7.8, using Theorem 5.28, we get $R(T_\lambda) = {}^{\perp}N(T_\lambda^*)$. Besides, by Theorem 5.32, $R(T_\lambda)$ is closed implies that $R(T_\lambda^*)$ is weak-star closed in X^*. Then using Theorem 5.28 again, we get $R(T_\lambda^*) = N(T_\lambda^*)^{\perp}$.

We now prove part (c). By Theorem 6.12, $\sigma(T) = \sigma(T^*)$. Since $T|_{N(T_\lambda)} = \lambda I|_{N(T_\lambda)}$, the identity operator from $N(T_\lambda)$ onto $N(T_\lambda)$ is compact. Thus the closed unit ball in $N(T_\lambda)$ is compact. By Theorem 1.12, there must holds that $\dim N(T_\lambda) < \infty$. Since T^* is also compact ($T^* \in \mathcal{C}(X^*)$) by Theorem 7.3, we get $\dim N(T_\lambda^*) < \infty$. We only need to show that $\dim N(T_\lambda) = \dim N(T_\lambda^*)$. By Theorem 5.28, $N(T_\lambda^*) = R(T_\lambda)^{\perp}$; and by Theorem 5.5, $R(T_\lambda)^{\perp}$ is isometrically isomorphic to $(X/R(T_\lambda))^*$. So it suffices to show that $\dim N(T_\lambda) = \dim(X/R(T_\lambda))^*$. Since $\dim N(T_\lambda) < \infty$, the desired result holds if and only if

$$\dim N(T_\lambda) = \dim X/R(T_\lambda) \equiv \operatorname{codim} R(T_\lambda) = \dim N(T_\lambda^*)\,.$$

To this end, we firstly show the following lemma:

Lemma 7.11. $\operatorname{codim} R(T_\lambda) \equiv \dim(X/R(T_\lambda)) \leq \dim N(T_\lambda) < \infty$.

Proof of Lemma 7.11. Suppose for contradiction that

$$\mathrm{codim} R(T_\lambda) > \dim N(T_\lambda) = n\,.$$

Then we can find $\{\widetilde{x}_1, \cdots, \widetilde{x}_{n+1}\}$ in $X/R(T_\lambda)$ which is linear independent. Trivially, $\{x_1, \cdots, x_{n+1}\}$ is linear independent. Let $N = \mathrm{span}\{x_1, \cdots, x_n\}$. Clearly $N \cap R(T_\lambda) = \{0\}$ and $x_{n+1} \notin R(T_\lambda) \oplus N$. Let $V : N(T_\lambda) \to N$ be a linear isomorphism. Let $P : X \to N(P)$ be a projection on X, that is, $P \in \mathcal{B}(X)$, $P^2 = P$ and $R(P) = N(T_\lambda)$. (The existence of such a projection is contained in Theorem 2.7 and Theorem 4.18.) Now define

$$S \coloneqq T_\lambda + VP = \lambda I - (T - VP)\,.$$

Firstly, $N(S) = \{0\}$ since $N \cap R(T_\lambda) = \{0\}$. Secondly, observe that $T - VP \in \mathcal{C}(X)$, then we get $R(S) = X$ by the Fredholm alternative. However, $x_{n+1} \notin R(S)$ since $R(S) \subset R(T_\lambda) \oplus N$, which is a contradiction! $\square$

Using Lemma 7.11, since $\dim X/R(T_\lambda) < \infty$, we have

$$\begin{aligned}
\dim N(T_\lambda^*) &= \dim R(T_\lambda)^\perp = \dim(X/R(T_\lambda))^* \\
&= \dim X/R(T_\lambda) \le \dim N(T_\lambda)\,.
\end{aligned}$$

For the same reason, $\dim N(T_\lambda^{**}) \le \dim N(T_\lambda^*)$. But $\dim N(T_\lambda) \le \dim N(T_\lambda^{**})$ since we can regard T_λ^{**} as an extension of T on X^{**} (see Remark 5.5), then the desired result follows. $\square$

7.3 Riesz-Schauder Theory

In this section, we will deal with the following two subjects: the spectrum of compact operators; the construction of compact operators. Corresponding to matrices, each question has a clear answer: 1, matrices have eigenvalues whose number $\le$ the dimension of the space. 2, using a series of invariant subspaces, a matrix can be transformed into a Jordan canonical form.

7.3.1 Spectrum of Compact Operators

The spectral theory of compact operators is considerably simpler than that of general bounded linear operators. In particular, every nonzero spectral value is an eigenvalue, the generalized eigenspaces are all finite-dimensional, and zero is the only possible accumulation point of the spectrum (i.e. each nonzero spectral value is an isolated point of the spectrum).

Theorem 7.12 (Riesz-Schauder). *Let X be a Banach space over $\mathbb{F}$ and $T \in \mathcal{C}(X)$, i.e., T is a compact operator on X. Then the following holds*

(a) *Every non-zero spectral point of T is an eigenvalue, in other words,*

$$\sigma(T)\backslash\{0\} = \sigma_p(T)\,.$$

(b) *If the dimension of X is not finite, then $\sigma(T)$ must contain 0.*

(c) *$\sigma(T)$ is at most countably infinite. If $\sigma(T)$ is countably infinite, let $\{\lambda_n\}$ be all the non-zero eigenvalues, then $\lim_n \lambda_n = 0$.*

Proof. Part (a) is the Fredholm alternative. Part (b) is the easy: If not, then $T : X \to X$ is a linear homeomorphism. Denote by B_X the closed unit ball of X. Then $T^{-1}(B_X)$ is a bounded subset of X. As a consequence of the compactness of T, $T(T^{-1}(B_X)) = B_X$ is compact , deducing that X has finite dimension. This is a contradiction.

We now show part (c). It sufficies to show that for each $\delta > 0$, there are only finite many eigenvalues outside $B(0, \delta)$. Suppose for contradiction that there is a sequence $\{\lambda_n\}$ su that $|\lambda_n| > \delta$ and $\lambda_n \neq \lambda_m$ for every $n \neq m$. Let $\{x_n\}$ be the corresponding non-trivial eigenvectors of $\{\lambda_n\}$. For each $n \in \mathbb{N}$, let

$$X_n = \operatorname{span}\{x_1, \cdots, x_n\}\,,$$

and $X_0 = \{0\}$. Then clearly $X_n \subsetneqq X_{n+1}$ for $n \geq 0$. By Riesz-Fredholm theory, X_n has finite dimension. So it follows from Lemma 1.11 that there exists $y_n \in X_n$ with

$$d(y_n, X_{n-1}) \geq \frac{1}{2} \quad \text{and} \quad \|y_n\| = 1\,.$$

Assume that $y_n = \sum_{j=1}^n \alpha_j^n x_j$, then $Ty_n = \sum_{j=1}^n \alpha_j^n \lambda_j x_j$, and hence

$$Ty_n - \lambda_n y_n \in X_{n-1}\,.$$

For $n > m \geq 0$, since $Ty_m \in X_m \subset X_{n-1}$, we have

$$Ty_n - Ty_m = \lambda_n y_n + [(Ty_n - \lambda_n y_n) - Ty_m] \in \lambda_n y_n + X_{n-1}\,.$$

Then

$$\|Ty_n - Ty_m\| \geq d(\lambda_n y_n, X_{n-1}) = |\lambda_n| d(y_n, X_{n-1}) \geq \frac{\delta}{2}\,.$$

But this implies that the sequence $\{Tx_n\}$ is not totally bounded, contradicting the compactness of T. We now complete the proof. $\square$

7.3.2 Construction of Compact Operators

The classical result for square matrices is the Jordan canonical form, which states the following:

Theorem. *Let A be an $n \times n$ complex matrix, i.e. A a linear operator acting on $\mathbb{C}^n$. If $\lambda_1, \cdots, \lambda_k$ are the distinct eigenvalues of A, then $\mathbb{C}^n$ can be decomposed into the invariant subspaces of A*

$$\mathbb{C}^n = N(\lambda_1 I - A)^{r_1} \oplus \cdots \oplus N(\lambda_k I - A)^{r_k}$$

where r_j is the algebraic multiplicity of eigenvalue λ_j satisfying $N\left(\lambda_i - A\right)^m = N\left(\lambda_i - A\right)^{m+1}$.

Theorem. *Let V be a finite dimensional vector space. Let L be a nilpotent transformation on V and q is the minimal positive integer satisfying $L^q = 0$.*

Then there exists positive integer r, integers $1 \le q_1 \le \cdots \le q_r \le q$, and vectors $x_1, \cdots, x_r$ such that

$$\begin{array}{cccc} \{x_1, & Lx_1, & \cdots, & L^{q_1-1}x_1 \\ x_2, & Lx_2, & \ldots, & L^{q_2-1}x_2 \\ \vdots & \vdots & & \vdots \\ x_r, & Lx_r, & \cdots, & L^{q_r-1}x_r\} \end{array}$$

is a basis of V with $L^{q_1}x_1 = \cdots = L^{q_r}x_r = 0$. Under this basis, the matrix of L is $\operatorname{diag}\{J_1, \cdots, J_r\}$*, where*

$$J_k = \begin{pmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}_{q_k \times q_k} .$$

We will try to extend the result to compact linear operators on Banach space X. Firstly, we point out the following fact. Let $A \in \mathcal{B}(X)$ be a bounded linear transformation on X. As we all know, (set $A^0 = I$)

$$\begin{gathered} \{0\} = N(A^0) \subset N(A) \subset N(A^2) \subset \cdots \subset N(A^n) \subset \cdots, \\ X = R(A^0) \supset R(A) \supset R(A^2) \supset \cdots \supset R(A^n) \supset \cdots. \end{gathered}$$

If $N(A^r) = N(A^{r+1})$ for some integer $r \ge 1$, then $N(A^r) = N(A^{r+k})$ for all $k \in \mathbb{N}$. Similarly, if $R(A^r) = R(A^{r+1})$ for some integer $m \ge 1$, then $R(A^r) = R(A^{r+k})$ for all $k \in \mathbb{N}$. However, for general $A \in B(X)$, we can not guarantee the existence of such a r, unless $A = T_\lambda = \lambda I - T$ for some $T \in \mathcal{C}(X)$ and $\lambda \ne 0$, as the following lemma states.

Lemma 7.13. *Let X be a Banach space over $\mathbb{F}$. Let $T \in \mathcal{C}(X)$ be a compact operator. Let $\lambda \in \mathbb{F}\backslash\{0\}$. Then*

$$\begin{aligned} r &:= \inf\{m \ge 0 : N(\lambda I - T)^m = N(\lambda I - T)^{m+1}\} \\ &= \inf\{m \ge 0 : R(\lambda I - T)^m = R(\lambda I - T)^{m+1}\} < \infty. \end{aligned}$$

Proof. Let $p \coloneqq \inf\{m \geq 0 : N(\lambda I - T)^m = N(\lambda I - T)^{m+1}\}$ and $q =$ *coloneqq* $\inf\{m \geq 0 : R(\lambda I - T)^m = R(\lambda I - T)^{m+1}\}$. It follows from the proof of Theorem 7.9 that p, q are finite. We shall show that $p \leq q$ and $q \leq p$, respectively.

To show that $p \leq q$, notice that $R(\lambda I - T)^q = R(\lambda I - T)^{q+1}$. By Theorem 7.10 (c), since $(\lambda I - T)^q = \lambda^q I + \sum_{j=1}^{n} \binom{q}{j}(-T)^j = \lambda^q I -$ (a compact operator), we have

$$\begin{aligned}\dim N(\lambda I - T)^q &= \operatorname{codim} R(\lambda I - T)^q \\ &= \operatorname{codim} R(\lambda I - T)^{q+1} = \dim N(\lambda I - T)^{q+1}\,.\end{aligned}$$

Because $\dim N(\lambda I - T)^q < \infty$, we have $N(\lambda I - T)^q = N(\lambda I - T)^{q+1}$. Thus it follows from the definition that $p \leq q$. Similarly, notice that $N(\lambda I - T)^p = N(\lambda I - T)^{p+1}$. By Theorem 7.10 (c) we have

$$\begin{aligned}\operatorname{codim} R(\lambda I - T)^p &= \dim N(\lambda I - T)^p \\ &= \operatorname{codim} N(\lambda I - T)^{p+1} = \operatorname{codim} R(\lambda I - T)^{p+1}\,.\end{aligned}$$

Because $\operatorname{codim} R(\lambda I - T)^q < \infty$, we have $R(\lambda I - T)^q = R(\lambda I - T)^{q+1}$. Thus it follows from the definition that $q \leq p$. Hence $p = q < \infty$. □

The null space $N(\lambda I - T)^r$ is called the *generalized eigenspace* of T associated to the non-zero eigenvalue λ. Recall that $\dim N(\lambda I - T)$ is called the *geometric multiplicity* of the eigenvalue λ and $\dim N(\lambda I - T)^r$ is called the *algebraic multiplicity* of the eigenvalue λ.

Theorem 7.14. *Let X be a Banach space and let $T \in \mathcal{C}(X)$ be a compact operator. Let λ be an non-zero eigenvalue of T. Then*

$$X = N(\lambda I - A)^r \oplus R(\lambda I - A)^r$$

where r is defined in Lemma 7.13.

Proof. Firstly, we show that $N(\lambda I - A)^r \cap R(\lambda I - A)^r = \{0\}$. In fact, if $y \in N(\lambda I - A)^r \cap R(\lambda I - A)^r$, then $(\lambda I - A)^r y = 0$ and there exists $x \in X$ so that $y = (\lambda I - A)^r x$. So $(\lambda I - A)^{2r} x = 0$. Since $N(\lambda I - A)^{2r} = N(\lambda I - A)^r$, we have $y = N(\lambda I - A)^r x = 0$.

Secondly, we show that $N(\lambda I - A)^r \oplus R(\lambda I - A)^r = X$. To this end, given any $x \in X$, we need to find some $z \in X$ with $x - (\lambda I - A)^r z \in N(\lambda I - A)^r$. That is

$$(\lambda I - A)^r x = (\lambda I - A)^{2r} z \,.$$

Since $R(\lambda I - A)^r = R(\lambda I - A)^{2r}$, such a z exists. We now complete the proof. $\square$

Remark 7.4. Since $(\lambda I - T)$ and T communicates, $N(\lambda I - T)^r, N(\lambda I - T)^r$ are invariant subspaces of T. Let $\widehat{T}_\lambda$ be the restriction of $\lambda I - T$ on $R(\lambda I - T)^r$ with it's range in $R(\lambda I - T)^r$. Then $\widehat{T}_\lambda$ is a linear homeomorphism.

Since $N(\lambda I - T)^r$, $R(\lambda I - T)^r$ both are closed subspace of X, let P_λ be the corresponding projection from X onto $N(\lambda I - T)^r$. Then we have a decomposition as following:

$$T = TP_\lambda + T(I - P_\lambda)\,.$$

Denote $S = TP_\lambda$ and $R = T(I - P_\lambda)$.

Corollary 7.15. *Let X be a Banach space and let $T \in \mathcal{C}(X)$ be a compact operator. Let λ be a non-zero eigenvalue of T. Then there exists a finite rank operator S on X and a compact operator R on X, satisfying the following assertions.*

(a) $T = S + R$ *and* $SR = RS = 0$.

(b) λ *is an eigenvalue of* S*. For* $\mu \neq 0$ *and* $\mu \neq \lambda$, $\mu \in \varrho(S)$.

(c) $\lambda \in \varrho(R)$*. For* $\mu \neq 0$ *and* $\mu \neq \lambda$, μ *is an eigenvalue of* R *if and only if* μ *is an eigenvalue of* T.

Proof. To show part (a), it suffices to show that T communicates with P_λ. This is trivial since $R(P_\lambda) = N(\lambda I - T)^r$ is an invariant subspace of T.

To show part (b), take any $\mu \neq 0$. If μ is an egivalue of S, then there exists $x \neq 0$ so that $(\mu I - S)x = 0$. Let $y = P_\lambda x$ and $z = (I - P_\lambda x)$, then we get

$$(\mu I - S)(y + z) = (\mu I - T)y + \mu z = 0\,.$$

Thus $z = 0$ and $Ty = \mu y$. Then $(\lambda I - T)^r y = (\lambda - \mu)^r y = 0$ forces that $\mu = \lambda$. Thus λ is the only non-zero eigenvalue of S.

We now show part (c). Take any $\mu \neq 0$. If μ is an egivalue of R, there exists $x \neq 0$ so that $(\mu I - R)x = 0$. Let $y = P_\lambda x$ and $z = (I - P_\lambda x)$, then we get

$$(\mu I - R)(y + z) = \mu y + (\mu I - T)z = 0\,.$$

Thus $y = 0, z \neq 0$ and $Tz = \mu z$. So we must have $\mu \in \sigma_p(T)$ and $\mu \neq \lambda$ (since if $\mu = \lambda$ then $z = 0$). Thus $\lambda \in \varrho(R)$, and for $\mu \neq 0$ and $\mu \neq \lambda$, if μ is an eigenvalue of R then μ is an eigenvalue of T.

On the other hand, if $\mu \neq 0$, $\mu \neq \lambda$ is an eigenvalue of T. There exists $x \neq 0$ so that $(\mu I - T)x = 0$. Let $y = P_\lambda x$ and $z = (I - P_\lambda x)$, then we get

$$(\mu I - T)(y + z) = (\mu - T)y + (\mu I - T)z = 0\,.$$

Then $Ty = \mu y$ and $Tz = Rz = \mu z$. Since $(\lambda I - T)^r y = (\lambda - \mu)^r y = 0$, we must have $y = 0$, so $z \neq 0$. Thus μ is an eigenvalue of R. $\square$

Chapter 8

Bounded Operators on Hilbert Space

A large area of current research interest is centered around the theory of operators on Hilbert space. There is a marked contrast here between Hilbert spaces and the Banach spaces. Essentially all of the information about the geometry of Hilbert space is contained in the preceding chapter. The geometry of Banach space lies in darkness and has attracted the attention of many talented research mathematicians. However, the theory of linear operators (linear transformations) on a Banach space has very few general results, whereas Hilbert space operators have an elegant and well-developed general theory. Indeed, the reason for this dichotomy is related to the opposite status of the geometric considerations. Questions concerning operators on Hilbert space don't necessitate or imply any geometric difficulties.

8.1 The Adjoint of an Operator

In this chapter, if not specifically declared, we always suppose that H and K be Hilbert spaces over the field $\mathbb{F}$.

8.1.1 Sesquilinear Form

A function $u : H \times K \to \mathbb{F}$ is called a *sesquilinear form* if for any h, g in H, k, f in K, and α, β in $\mathbb{F}$,

(a) $u(\alpha h + \beta g, k) = \alpha u(h, k) + \beta u(g, k)$;

(b) $u(h, \alpha k + \beta f) = \bar{\alpha} u(h, k) + \bar{\beta} u(h, f)$.

The prefix "sesqui" is used because the function is linear in one variable but (for $\mathbb{F} = \mathbb{C}$) only conjugate linear in the other. "Sesqui" means "one-and-a-half."

A sesquilinear form is *bounded* if there is a constant $M > 0$ such that $|u(h, k)| \leq M\|h\|\|k\|$ for all h in H and k in K. When u is bounded,

$$\|u\| := \sup_{\|h\| \leq 1, \|k\| \leq 1} |u(h, k)| \tag{8.1}$$

is called the *norm* of u. As the linear operator, a sesquilinear form u is bounded if and only if u is continuous on $H \times K$. In fact, if "only if" part is trivial; if u is continuous but not bounded, for any n, there exists $h_n \in H$ and $k_n \in K$ such that

$$\|h_n\| = \|k_n\| = 1 \, , |u(h_n, k_n)| \geq n^2 \, .$$

Then h_n/n and k_n/n tend to 0, but $|u(h_n/n, k_n/n)| \geq 1$, which a contradiction.

Example 8.1. Sesquilinear forms are used to study operators.

(a) If $A \in \mathcal{B}(H, K)$, then $u(h, k) := \langle Ah, k \rangle$ is a bounded sesquilinear form, and $\|u\| = \|A\|$.

(b) Similarly, if $B \in \mathcal{B}(K, H)$, then $u(h, k) := \langle h, Bk \rangle$ is a bounded sesquilinear form, and $\|u\| = \|B\|$.

A natural question is that are there any more bounded sesquilinear form? Are these two forms related?

Theorem 8.1. *Let H, K be Hilbert spaces. Let $u : H \times K \to \mathbb{F}$ be a continuous sesquilinear form. Then there are unique operators $A \in \mathcal{B}(H, K)$ and $B \in \mathcal{B}(K, H)$ such that*

$$u(h, k) = \langle Ah, k \rangle = \langle h, Bk \rangle \,, \tag{8.2}$$

for all $h \in H$ and $k \in K$ and $\|u\| = \|A\| = \|B\|$.

Proof. Only the existence and uniqueness of A will be shown since the same argument still holds for B. For each h in H, define a functional L_h on K by

$$L_h(k) = \overline{u(h, k)} \,, \text{ for all } k \in K \,.$$

Then it's easy to check that L_h is linear and,

$$|L_h(k)| \le \|u\|\|h\|\|k\|.$$

By the Riesz representation theorem, there is a unique vector f_h in H such that $\langle k, f_h \rangle = L_h(k) = \overline{u(h, k)}$ and $\|f_h\| \le \|u\|\|h\|$.

Define $A : H \to K$; $h \mapsto f_h$. Then A is linear by the uniqueness part of the Riesz representation theorem. Besides, for any $h \in H$ and $k \in K$,

$$\langle Ah, k \rangle = \overline{\langle k, Ah \rangle} = \overline{\langle k, f_h \rangle} = u(h, k) \,,$$

which implies that A is bounded. From Example 8.1 we have $\|A\| = \|u\|$.

If $A_1 \in \mathcal{B}(H, K)$ and $u(h, k) = \langle A_1 h, k \rangle$, then $\langle Ah - A_1 h, k \rangle = 0$ for all k, thus $Ah - A_1 h = 0$ for all h, and hence $A_1 = A$. So the uniqueness follows. ☐

A variant of F.Riesz's representation theorem, formulated by P. Lax and A.N.Milgram, is a useful tool for the discussion of the existence of solutions of linear partial differential equations of elliptic type.

Theorem 8.2 (The Lax-Milgram Theorem)**.** *Let H be Hilbert spaces. Let $u : H \times H \to \mathbb{F}$ be a continuous sesquilinear form. If u is also coercive, that is, there is some $c > 0$ so that*

$$u(x, x) \ge c\|x\|^2 \quad \textit{for each} \quad x \in H \,.$$

Then there exists a uniquely determined $S \in \mathcal{B}(H)$ such that $\langle x, y\rangle = u(x, Sy)$ whenever $x, y \in H$, and $\|S\| \leq c^{-1}$.

Proof. By Theorem 8.1, there exists $B \in \mathcal{B}(H)$ so that $u(x, y) = \langle x, By\rangle$ whenever $x, y \in H$. If we can show that B is bijective, then letting $S = B^{-1}$, clearly we have $\langle x, y\rangle = u(x, Sy)$ whenever $x, y \in H$. Moreover, since for every $y \in H$,

$$\|Sy\|\|y\| \geq |\langle Sy, y\rangle| = |u(Sy, Sy)| \geq c\|Sy\|^2 \,,$$

we get that $\|S\| \leq c^{-1}$. We now have to show that B is bijective. Obviously, B is injective since if $By = 0$ then $0 = |\langle y, By\rangle| = |u(y, y)| \geq c\|y\|^2$, so $y = 0$. Next, we show that $R(B)^{\perp} = \{0\}$, from which we deduce that $R(B)$ is dense in H. Note that if $x \in R(B)^{\perp}$, then $0 = |\langle x, Bx\rangle| = |u(x, x)| \geq c\|x\|^2$, and hence $x = 0$. Finally, we show that $R(B)$ is closed, which follows trivially from that $S = B^{-1} : R(B) \to H$ is a continuous linear operator and hence $R(B)$ and H are linear homeomorphic. $\square$

Example 8.2. Here is a sketch of the typical application of Lax-Milgram to elliptic PDEs (we don' t need to know all notions in advance, just grab the main message). The task is always to identify the "good" Hilbert space of functions among which we look for solutions to a given PDE, and to check the validity of the assumptions on u. The sesquilinear form emerges naturally when testing the PDE against some test functions (in fact, one looks for *weak solutions*). Coercivity encodes some kind of Sobolev embedding.

Let Ω be a bounded region in $\mathbb{R}^d$. Let $f \in L^2(\Omega)$ be given. Let $a \geq 0$. Consider the boundary-value problem

$$\begin{cases} -\Delta u + au = f & \text{in } \Omega\,; \\ u = 0 & \text{on } \partial\Omega\,. \end{cases} \tag{8.3}$$

The claim is: there exists a unique"weak" solutions u to (8.3) in the space

$H_0^1(\Omega)$, the completion of $C_0^\infty(\Omega)$ with respect to the norm

$$\|v\|_{H_0^1} := \sqrt{\|v\|_2^2 + \sum_{j=1}^{n} \|\partial_j v\|_2^2}\,.$$

Note that the condition $u \in H_0^1(\Omega)$ encodes the vanishing of $u \in H_0^1(\Omega)$ at the boundary of Ω. By weak solution to (8.3) one means a function u that satisfies

$$\langle \nabla v, \nabla u\rangle + a\langle v, u\rangle = \langle v, f\rangle \quad \forall v \in H_0^1(\Omega) \tag{8.4}$$

where $\langle\cdot,\cdot\rangle$ denotes the scalar product in L^2 as usual. (∇u for $u \in H_0^1(\Omega)$ is a well defined function in L^2 via a limiting procedure- recall the definition of $H_0^1(\Omega)$). In fact, if u and v were smooth and vanished on $\partial\Omega$ then, owing to Green's identity,

$$\begin{aligned}\int_\Omega -(\Delta u)(x)v(x)dx &= \int_\Omega f(x)v(x)dx - \int_{\partial\Omega} \frac{\partial u}{\partial n} v d\sigma \\ &= \int_\Omega f(x)v(x)dx\,.\end{aligned}$$

Thus, in fact, (8.4) is the PDE (8.3) ”tested” against v. Thus, (8.4) would be certainly satisfied by a “classical” solution u to (8.3) which might not exist in this case, though, because f is a priori not smooth enough. (8.4) suggests that the appropriate bilinear form in this case is

$$B(v,u) := \langle \nabla v, \nabla u\rangle + a\langle v, u\rangle\,.$$

Such a B is bounded on $H_0^1(\Omega)$ because $\forall u, v \in H_0^1(\Omega)$

$$|B(v,u)| \le \|\nabla v\|_2 \|\nabla u\|_2 + a\|v\|_2\|u\|_2 \le (1+a)\|v\|_{H_0^1}\|u\|_{H_0^1}$$

and is coercive on $H_0^1(\Omega)$ because $\forall v \in H_0^1(\Omega)$

$$|B(v,v)| = \|\nabla v\|_2^2 + a\|v\|_2^2 \ge \|\nabla v\|_2^2 \ge c\|v\|_2$$

where the last step is the “deep” one and follows from the Poincaré's inequality for functions $g \in H_0^1(\Omega)$

$$\|g\|_{L^2(\Omega)} \le c_\Omega \|\nabla g\|_{L^2(\Omega)} \quad (\text{ when } |\Omega| < \infty)$$

and $c := \min\left\{\frac{1}{2}, \frac{1}{2c_\Omega^2}\right\} > 0$. Therefore Lax-Milgram says that there exists uniquely a $u \in H_0^1(\Omega)$ such that $B(v,u) = \langle v, f\rangle$ which means precisely that there is a unique weak solution u to (8.3).

8.1.2 Adjoints

For any $A \in \mathcal{B}(H,K)$, Theorem 8.1 asserts that there exists an unique operator B in $\mathcal{B}(K,H)$ satisfying

$$u(h,k) = \langle Ah, k\rangle = \langle h, Bk\rangle\,,$$

for all $h \in H$ and $k \in K$. We say B is the *adjoint* of A, and we prefer to denote the adjoint of A by A^*. Recall that Theorem 8.1 implies that $\|A\| = \|A^*\|$. Moreover, we have

$$\|A^*A\|^2 = \|A\|^2\,.$$

This is the so-called C^*-identity and will be used later. Indeed, clearly $\|A^*A\|^2 \le \|A\|^2$. On the contrary,

$$\begin{aligned}\|A^*A\|^2 &= \sup_{x,y\in B_H} |\langle A^*Ax, y\rangle| \ge \sup_{x\in B_H} |\langle A^*Ax, x\rangle| \\ &= \sup_{x\in B_H} \|Ax\|^2 = \|A\|^2\,.\end{aligned}$$

Thus the desired result holds.

Remark 8.1. we have defined the adjoint (dual) of A before, by A^*_{Banach} : $K^* \to H^*$, $k^* \mapsto k^* \circ A$. In fact, set $\phi_H : H \to H^*$, $h \mapsto \langle \cdot, h\rangle$ and the same for ϕ_K, then the following diagram is commutative.

$$\begin{tikzcd} H \arrow[d, "\phi_H"'] & K \arrow[l, "A^*_{Hilbert}"'] \arrow[d, "\phi_K"] \\ H^* & K^* \arrow[l, "A^*_{Banach}"] \end{tikzcd}$$

In other words, $A^*_{Hilbert} = \phi_H^{-1} A^*_{Banach}\, \phi_K$.

Example 8.3. If a linear transform $A : \mathbb{C}^n \to \mathbb{C}^m$ is presented by a matrix

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

then its adjoint A^* is represented by the conjugate transpose of the matrix, that is

$$\begin{bmatrix} \bar{a}_{11} & \bar{a}_{21} & \cdots & \bar{a}_{m1} \\ \bar{a}_{12} & \bar{a}_{22} & \cdots & \bar{a}_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{a}_{1n} & \bar{a}_{2n} & \cdots & \bar{a}_{mn} \end{bmatrix}.$$

Example 8.4. Let (X, Ω, μ) be a σ-finite measure space and let M_ϕ be the multiplication operator with symbol ϕ on Hilbert space $L^2(\mu)$ (Example 1.16). Then

$$M_\phi^* = M_{\overline{\phi}}$$

is the multiplication operator with symbol $\overline{\phi}$, since for all $f, g \in L^2(\mu)$,

$$\langle M_\phi f, g\rangle = \int f\phi\bar{g}\, d\mu = \int f\, \overline{\bar{\phi}g}\, d\mu = \langle f, M_{\bar{\phi}} g\rangle\,.$$

Example 8.5. Let (X, Ω, μ) be a σ-finite measure space and K is the integral operator with kernel k on $L^2(\mu)$ (Example 1.17). Then K^* is the integral operator with kernel k^*, given by

$$k^*(x, y) \equiv \overline{k(y, x)}\,,\ \text{for } (x, y) \in X^2\,.$$

since for all $f, g \in L^2(\mu)$,

$$\begin{aligned} \langle Kf, g\rangle &= \int\int k(x, y)f(y)\mu(dy)\overline{g(x)}\mu(dx) \\ &= \int\int k(x, y)\overline{g(x)}\mu(dx)f(y)\mu(dy) \\ &= \int\overline{\int \overline{k(x, y)}g(x)\mu(dx)}f(y)\mu(dy) = \langle f, K^*g\rangle\,. \end{aligned}$$

Example 8.6. If $S : \ell^2 \to \ell^2$ is defined by $S(x_1, x_2, \ldots) = (0, x_1, x_2, \ldots)$, which is called the *unilateral shift*, then S is an isometry embedding and

$$S^*(x_1, x_2, \ldots) = (x_2, x_3, \ldots)$$

S^* is the *backward shift operator*. To see this, note that for all $x, y \in \ell^2$,

$$\langle Sx, y\rangle = \sum_{n=1}^{\infty} x_n \bar{y}_{n+1} = \langle x, S^* y\rangle .$$

Example 8.7. Let P be a projection on Hilbert space H, then $P^* = P$. Indeed, for any $x, y \in H$, we have

$$\langle Px, y\rangle = \langle Px, Py\rangle = \langle x, Py\rangle .$$

Example 8.8. Let U be an isometric isomorphism between Hilbert spaces H and K , then $U^* = U^{-1}$. In fact, for $h \in H$ and $k \in K$, we have

$$\langle Uh, k\rangle = \langle Uh, UU^{-1}k\rangle = \langle h, U^{-1}k\rangle .$$

From now on we will examine and prove results for the adjoint of operators in $\mathcal{B}(H)$. Often, as in the next proposition, there are analogous results for the adjoint of operators in $\mathcal{B}(H, K)$. This simplification is justified, however, by the cleaner statements that result. Also, there is no trouble formulating the more general statement when it is needed. The proof of the next proposition is easy and hence is omitted.

Proposition 8.3. *Let $A, B \in \mathcal{B}(H)$ and $\alpha, \beta \in \mathbb{F}$, then the following assertions hold.*

(a) $A^{**} := (A^*)^* = A$.

(b) $(\alpha A + \beta B)^* = \bar{\alpha} A^* + \bar{\beta} B^*$.

(c) $(AB)^* = B^* A^*$.

The operation of taking the adjoint of an operator is, as the reader may have seen from the examples above, analogous to taking the conjugate of a complex number. It is good to keep the analogy in mind, but do not become too religious about it.Next, we conclude with a very important, though easily proved, result.

Theorem 8.4. *Let $A \in \mathcal{B}(H)$, then*

$$\begin{aligned} R(A)^{\perp} = N(A^*) \quad &\textit{and} \quad R(A^*)^{\perp} = N(A)\,; \\ \overline{R(A)} = N(A^*)^{\perp} \quad &\textit{and} \quad \overline{R(A^*)} = N(A)^{\perp}\,. \end{aligned}$$

In particular, A is injective if and only if $R(A^*)$ is dense in H; $R(A)$ is dense in H if and only if A^* is injective.

Proof. First of all, note that

$$\begin{aligned} y \in R(A)^{\perp} &\Leftrightarrow \langle Ax, y\rangle = 0 \text{ for all } x \\ &\Leftrightarrow \langle x, A^*y^*\rangle = 0 \text{ for all } x \Leftrightarrow A^*y^* = 0\,, \end{aligned}$$

so $R(A)^{\perp} = N(A^*)$ and hence $R(A^*)^{\perp} = N(A^{**}) = N(A)$. Since for any subspace S of Hilbert space H we have $S^{\perp\perp} = \overline{S}$ (Corollary 2.10), we get $\overline{R(A)} = R(A)^{\perp\perp} = N(A^*)^{\perp}$ and $\overline{R(A^*)} = R(A^*)^{\perp\perp} = N(A)^{\perp}$. □

Corollary 8.5. *Let $A \in \mathcal{B}(H)$, then $H = N(A) \oplus \overline{R(A^*)}$ with $N(A) \perp \overline{R(A^*)}$.*

Example 8.9. Let A be a continuous linear transform on ℓ^2 defined by

$$(x_k) \mapsto \left(\frac{x_k}{k}\right)$$

This operator is self-adjoint, (i.e., $A = A^*$), injective, and has a dense image, but is not surjective. For example $(k^1) \in \ell^2 \backslash R(A)$. Therefore,

$$R(A) \subsetneqq \ell^2 = {}^{\perp}N\left(A^*\right)\,.$$

Theorem 8.6. *Let $A \in \mathcal{B}(H)$. Then the following assertions holds.*

(a) *A is bijective if and only if A^* is, and in this case $(A^*)^{-1} = \left(A^{-1}\right)^*$.*

(b) *A is an isometric isomorphism if and only if A^* is.*

(c) *$R(A)$ is a closed subspace of H if and only if $R(A^*)$ is.*

(d) *A is compact if and only if A^* is.*

Proof. Part (a) follows directly from Proposition 8.3. Part (b) follows directly from part (a). As we have emphasized before, for $A \in \mathcal{B}(H)$

$$A^*_{Hilbert} = \phi_H^{-1} A^*_{Banach} \phi_H \,,$$

where $\phi_H : h \mapsto \langle \cdot, h\rangle ; H \mapsto H^*$ is a conjugate-linear isometric surjection. Thus part (c) and part (d) follows directly from Theorem 5.32 and Theorem 7.3, respectively. □

Theorem 8.7. *Let H be a complex Hilbert space. Let $A \in \mathcal{B}(H)$. Then*

$$\sigma\left(A^*\right) = \{\bar{\lambda} \in \mathbb{C} : \lambda \in \sigma(A)\} \,.$$

Moreover

$$\begin{aligned} &\sigma_p\left(A^*\right) \subset \{\bar{\lambda} \mid \lambda \in \sigma_p(A) \cup \sigma_r(A)\} \,; \\ &\sigma_c\left(A^*\right) = \{\bar{\lambda} \mid \lambda \in \sigma_c(A)\} \,; \\ &\sigma_r\left(A^*\right) \subset \{\bar{\lambda} \mid \lambda \in \sigma_p(A)\} \,. \end{aligned}$$

Proof. Observe that

$$\lambda I - A^* = (\bar{\lambda} I - A)^* \,,$$

and the part (a) of Theorem 8.6, we get that $\lambda \in \varrho(A)$ if and only if $\bar{\lambda} \in \varrho(A)$. Thus $\lambda \in \sigma(A)$ if and only if $\bar{\lambda} \in \sigma(A)$. The rest relationships can be shown by the same argument as the proof Theorem 6.12. □

8.2 Three Kinds of Operators on Hilbert Space

Definition 8.1. Let H be a Hilbert space. Let A be a continuous linear transform on H, i.e., $A \in \mathcal{B}(H)$.

- A is called *normal* if $AA^* = A^*A$.
- A is called *self-adjoint* if $A^* = A$.
- A is called *unitary* if $AA^* = A^*A = I$.

In the analogy between the adjoint and the complex conjugate, self-adjoint operators become the analogues of real numbers and, unitaries are the analogues of complex numbers of modulus 1. Normal operators, as we shall see, are the true analogues of complex numbers. Notice that every self-adjoint operator and every unitary operator both are normal.

Example 8.10. Let (X, Ω, μ) be a σ-finite measure space and let M_ϕ be the multiplication operator with symbol ϕ on Hilbert space $L^2(\mu)$ (Example 1.16). Recall that in Example 8.4 we have pointed that $M_\phi^* = M_{\bar{\phi}}$.

(a) M_ϕ is normal, since multiplication operators commute.

(b) M_ϕ is self-adjoint if and only if ϕ is real-valued μ-a.e..

(c) M_ϕ is unitary if and only if $|\phi| = 1$ μ-a.e..

Example 8.11. Let (X, Ω, μ) be a σ-finite measure space and K is the integral operator with kernel k on $L^2(\mu)$ (Example 1.17). Recall that in Example 8.5 we have pointed that K^* is the integral operator with kernel k^* given by

$$k^*(x, y) = \overline{k(y, x)}\,, \text{ for all } x \in X\,.$$

Thus K is self-adjoint if and only if

$$k(x, y) = \overline{k(x, y)}\,, \ \mu \times \mu\text{-a.e.}.$$

Example 8.12. Obviously, the unilateral shift S on ℓ^2 defined in (8.6) is not normal. In fact, we have

$$S^*S = I$$

and

$$(SS^*)(x_1, x_2, \cdots) = (0, x_2, \cdots)\,, \text{ for } (x_n) \in \ell^2\,.$$

Thus S is not normal. It is an isometric embedding but is not unitary.

8.2.1 Self-adjoint Operators

A bivariate function $u : H \times H \to \mathbb{F}$ is called *hermitian* if,

$$u(h_1, h_2) = \overline{u(h_2, h_1)} \text{ for any } h_1, h_2 \in H\,.$$

From Theorem 8.1 we know there is a one-to-one correspondence between continuous sesquilinear hermitian form and self-adjoint operators. In fact, if u is a continuous sesquilinear hermitian form, then there exists unique $A \in \mathcal{B}(H)$, which is so that

$$u(h_1, h_2) = \langle Ah_1, h_2 \rangle\,.$$

Since u is hermitian, we have

$$\langle Ah_1, h_2 \rangle = \overline{\langle Ah_2, h_1 \rangle} = h_1, Ah_2 \rangle\,,$$

for all h_1, h_2 in H. Thus A is self-adjoint. We get the following theorem.

Theorem 8.8. *Let $u : H \times H \to \mathbb{F}$ be a continuous sesquilinear hermitian form. Then there are unique self-adjoint operators $A \in \mathcal{B}(H)$ so that*

$$u(h_1, h_2) = \langle Ah_1, h_2 \rangle \tag{8.5}$$

for all $h_1, h_2 \in H$, and $\|u\| = \|A\|$.

Lemma 8.9 (Polar Indenity)**.** *Let H be a complex Hilbert space. Let u be a sesquilinear form on H. Then the following "polar identity" holds for any $h_1, h_2 \in H$.*

$$\begin{aligned} u(h_1, h_2) = & \frac{u(h_1 + h_2, h_1 + h_2) - u(h_1 - h_2, h_1 - h_2)}{4} \\ & + i \times \frac{u(h_1 + ih_2, h_1 + ih_2) - u(h_1 - ih_2, h_1 - ih_2)}{4} \,. \end{aligned} \tag{8.6}$$

Remark 8.2. If u is a sesquilinear hermitian form on $\mathbb{F}$-Hilbert space H, the "polar identity" holds without doubt.

Proof. Let $v(h) \coloneqq u(h, h)$ for all $h \in H$. Then

$$\begin{aligned} v(h_1 + h_2) = v(h_1) + v(h_2) + u(h_1, h_2) + u(h_2, h_1)\,; \\ v(h_1 - h_2) = v(h_1) + v(h_2) - u(h_1, h_2) - u(h_2, h_1)\,. \end{aligned}$$

Hence

$$\begin{aligned} v(h_1 + h_2) - v(h_1 - h_2) = 2u(h_1, h_2) + 2u(h_2, h_1)\,; \\ i[v(h_1 + ih_2) - v(h_1 - ih_2)] = 2u(h_1, h_2) - 2u(h_2, h_1)\,. \end{aligned}$$

So

$$v(h_1 + h_2) - v(h_1 - h_2) + i[v(h_1 + ih_2) - v(h_1 - ih_2)] = 4u(h_1, h_2)\,. \quad \square$$

Proposition 8.10. *Let H be a complex Hilbert space. u is a sesquilinear form H. Then u is hermitian if and only if*

$$u(h, h) \in \mathbb{R}\,, \quad \textit{for all } h \in H\,. \tag{8.7}$$

Particularly, $A \in \mathcal{B}(H)$, then A is self-adjoint if and only if $\langle Ah, h \rangle \in \mathbb{R}$ for any h in H.

Proof. If u is hermitian, then for each $h \in H$,

$$u(h,h) = \overline{u(h,h)}\,,$$

thus $u(h,h) \in \mathbb{R}$. Using the polar identity with $v(h) \coloneqq u(h,h)$,

$$\begin{aligned} u(h_1,h_2) &= \frac{v(h_1+h_2) - v(h_1-h_2)}{4} + i\,\frac{v(h_1+ih_2) - v(h_1-ih_2)}{4}\,; \\ u(h_2,h_1) &= \frac{v(h_2+h_1) - v(h_2-h_1)}{4} + i\,\frac{v(h_2+ih_1) - v(h_2-ih_1)}{4}\,. \end{aligned}$$

Observing that $v(h) \in \mathbb{R}$ by assumption, $v(-h) = v(h)$ and $v(ih) = v(h)$ for all h, we have

$$\begin{aligned} v(h_1+h_2) - v(h_1-h_2) &= v(h_2+h_1) - v(h_2-h_1)\,; \\ v(h_1+ih_2) - v(h_1-ih_2) &= v(h_2-ih_1) - v(h_2+ih_1)\,. \end{aligned}$$

That is $\operatorname{Re} u(h_1,h_2) = \operatorname{Re} u(h_2,h_1)$ and $\operatorname{Im} u(h_1,h_2) = -\operatorname{Im} u(h_2,h_1)$. So u is hermitian, as desired. □

Example 8.13 (Counterexample)**.** The preceding proposition is false if it is only assumed that H is an real Hilbert space. For example, if

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

on $\mathbb{R}^2$, then $\langle Ah, h\rangle = 0$ for all h in $\mathbb{R}^2$. However, A^* is the transpose of A and so $A^* \neq A$. Indeed, for any operator A on an $\mathbb{R}$-Hilbert space, $\langle Ah_1, h_2\rangle \in \mathbb{R}$.

Theorem 8.11. *Let H be a Hilbert space. Let $A \in \mathcal{B}(H)$ be self-adjoint. Then*

$$\|A\| = \sup_{\|h\|=1} |\langle Ah, h\rangle|\,. \tag{8.8}$$

Proof. Let $u(h_1, h_2) = \langle Ah_1, h_2 \rangle$ and $v(h_1) = u(h_1, h_1)$ for $h_1, h_2 \in H$. We only need to prove that

$$M := \sup_{\|h\|=1} |v(h)| \geq \|A\| = \|u\| \,.$$

For any $h_1, h_2 \in H$ such that $\|h_1\|, \|h_2\| \leq 1$. Using "polar identity" we know

$$\operatorname{Re} u(h_1, h_2) = \frac{v(h_1 + h_2) - v(h_1 - h_2)}{4} \,,$$

So

$$|\operatorname{Re} u(h_1, h_2)| \leq \frac{M}{4}(\|h_1 + h_2\|^2 + \|h_1 - h_2\|^2) = \frac{M}{2}(\|h_1\|^2 + \|h_2\|^2) \leq M \,.$$

Let $u(h_1, h_2) = |u(h_1, h_2)|e^{i\theta}$. So $u(e^{-i\theta}h_1, h_2) = |u(h_2, h_2)|$ is real. Thus,

$$|u(h_1, h_2)| = u(e^{-i\theta}h_1, h_2) = \operatorname{Re} u(e^{-i\theta}h_1, h_2) \leq M \,.$$

Therefore, $\|A\| = \|u\| \leq M$. □

Corollary 8.12. *Let $A \in \mathcal{B}(H)$ be self-adjoint. Then $A = 0$ if and only if*

$$\langle Ah, h \rangle = 0 \text{ for all } h \in H \,.$$

The preceding corollary is not true unless $A = A^*$, as the example given in Example 8.13 shows. However, if a complex Hilbert space is present, this hypothesis can be deleted, as a consequence of the polar identity of sesquilinear form on complex Hilbert space.

Exercise 8.1. Let H be a complex Hilbert space and $A \in \mathcal{B}(H)$ such that $\langle Ah, h \rangle = 0$ for all h in H, then $A = 0$.

Corollary 8.13. *Let $A \in \mathcal{B}(H)$. Then A^*A is a self-adjoint operator and*

$$\|A^*A\| = \|A\|^2 \,.$$

8.2.2 Normal Operators

Theorem 8.14. *$A \in \mathcal{B}(H)$, then A is normal if and only if*

$$\|Ah\| = \|A^*h\| \ \text{for all} \ h \in H. \tag{8.9}$$

Proof. If A is normal, then for any $h \in H$, we have

$$\begin{aligned}\|Ah\|^2 &= \langle Ah, Ah\rangle = \langle A^*Ah, h\rangle \\ &= \langle AA^*h, h\rangle = \langle A^*h, A^*h\rangle = \|A^*h\|^2 .\end{aligned}$$

Thus (8.9) holds. On the other hand, suppose (8.9) holds, then by the argument above we have

$$\langle (A^*A - AA^*)h, h\rangle = 0 \text{ for all } h \in H .$$

Since A^*A and AA^* both are self-adjoint, by Corollary 8.12, $AA^* - A^*A = 0$, and hence A is normal. $\square$

Exercise 8.2. Let $A \in \mathcal{B}(H)$. Then A is normal if and only if for any $B \in \mathcal{B}(H)$, $\|AB\| = \|A^*B\|$ and $N(AB) = N(A^*B)$.

Corollary 8.15. *Let $A \in \mathcal{B}(H)$ be normal, then the following statements hold.*

(a) $N(A) = N(A^*)$ *and* $\overline{R(A)} = \overline{R(A^*)}$.

(b) $H = N(A) \oplus \overline{R(A)}$ *with* $N(A) \perp \overline{R(A)}$.

Proof. By (8.9) we get $N(A) = N(A^*)$. Since by Theorem 8.4,

$$\overline{R(A^*)} = N(A)^{\perp} = N(A^*)^{\perp} = \overline{R(A)}$$

and hence

$$H = N(A) \oplus \overline{R(A)} \ \text{ with } \ N(A) \perp \overline{R(A)} . \qquad \square$$

Let H be a complex Hilbert space, $A \in \mathcal{B}(H)$. Define

$$B = \frac{A + A^*}{2}\,, \text{ and } C = \frac{A - A^*}{2i}\,. \tag{8.10}$$

Then it's easy to see that B, C are self-adjoint with

$$A = B + iC \quad \text{and} \quad A^* = B - iC\,.$$

The operators B, C are called the *real* and *imaginary parts* of A, respectively.

Proposition 8.16. *Let H be a complex Hilbert space. Let $A \in \mathcal{B}(H)$. Then A is normal if and only if the real and imaginary parts of A commute.*

Proof. If B, C are real and imaginary parts of A, then a calculation yields

$$A^*A = B^2 - iCB + iBC + C^2$$
$$AA^* = B^2 + iCB - iBC + C^2$$

Hence $A^*A = AA^*$ if and only if $CB = BC$, as desired. $\square$

8.2.3 Unitaries

Lemma 8.17. *Let $A \in \mathcal{B}(H)$. Then A is an isometry if and only if $A^*A = I$.*

Proof. Note that for $h \in H$,

$$\|Ah\|^2 = \langle Ah, Ah\rangle = \langle A^*Ah, h\rangle$$

Thus $\|Ah\| = \|h\|$ if and only if $\langle (A^*A - I)h, h\rangle = 0$ for all $h \in H$. Since A^*A is self-adjoint, by Corollary 8.12, $\|Ah\| = \|h\|$ if and only if $A^*A = I$. $\square$

Theorem 8.18. *Let $A \in \mathcal{B}(H)$. Then the following statements are equivalent.*

(a) *A is unitary; i.e., $A^*A = AA^* = I$.*

(b) *A is a normal isometry.*

(c) A *and* A^* *are isometric; i.e.,* $\|Ah\| = \|A^*h\| = \|h\|$ *for all* h.

(d) A *is an isometric isomorphism.*

Proof. By the preceding lemma, it's easy to see that (a), (b), (c) are equivalent and (a) implies (d). To show that (d) implies (a), since A is an isometric, $A^*A = I$. Since A is an isomorphism, by the uniqueness of the inverse of A, $A^* = A^{-1}$ and hence (a) holds. □

Theorem 8.19. *Let* U *be a unitary operator on the complex Hilbert space* H*. Then* $\sigma(U)$ *is a compact subset of the unit circle* S^1*.*

Proof. It suffices to show that for every $\lambda \in \mathbb{C}$ so that $|\lambda| \neq 1$, $\lambda \in \varrho(U)$. Observe that

$$\|\lambda h - Uh\|^2 \geq \big||\lambda| - 1\big|^2 \|h\|^2 \quad \text{for all} \quad h \in H\,,$$

We have $N(\lambda I - U) = \{0\}$. Moreover, $R(\lambda I - U)$ is dense in H. Indeed if $y \perp R(\lambda I - U)$, then $\langle(\lambda I - U)x, y\rangle = 0$ for all $x \in H$ and so $\langle x, (\bar{\lambda} I - U^*)y\rangle = 0$ for all $x \in H$. We must have $(\bar{\lambda} I - U^*)y = 0$, that is, $U^*y = \bar{\lambda}y$, contrary to the fact that $\|U^*y\| = \|y\|$. □

8.3 Projections and Invariant Subspaces

Let M be any closed subspace of the Hilbert space H. We have given the definition of (*orthogonal*) *projection* P_M in the preceding chapter (see (2.7)). When we say a operator P is projection, we means that there exists a closed linear subspace M of H such that $P = P_M$. Obviously, We have $R(P) = M$ and $N(P) = M^\perp$.

If P is a projection on H, then clearly $P^2 = P$. In Banach space, we use this property to define projection since there doesn't exist the orthogonality(see Section 4.5). However, in Hilbert space, this property characterizes "projection" which is not orthogonal but skewed. So we distinguish it from orthogonal projection and rename it.

Definition 8.2. An *idempotent* on H is a bounded linear operator E on H such that $E^2 = E$.

It is not difficult to construct an idempotent that is not a projection on some Hilbert space.

Example 8.14. Let H be the two-dimensional Euclid space $\mathbb{R}^2$, let

$$M = \left\{(x, 0) \in \mathbb{R}^2 : x \in \mathbb{R}\right\}, \ N = \{(x, x\tan\theta) : x \in \mathbb{R}\},$$

where $0 < \theta < \frac{1}{2}\pi$. There exists an idempotent E_θ with $R(E_\theta) = M$ and $N(E_\theta) = N$. Clearly E_θ must not be a orthogonal projection since M and N are not orthogonal.

We restate Proposition 4.16 and Theorem 4.17 as follows.

Proposition 8.20. *Let H be a Hilbert space and $E \in \mathcal{B}(H)$.*

(a) *E is an idempotent if and only if $I - E$ is an idempotent.*

(b) *$R(E) = N(I - E)$, $N(E) = R(I - E)$ and both ran E and ker E are closed subspaces of H.*

(c) *$H = R(E) \oplus N(E)$.*

Proposition 8.21. *If M, N are two closed subspace of Hilbert space H such that $H = M \oplus N$. Then there is a unique idempotent $E \in \mathcal{B}(H)$ satisfying*

$$R(E) = M \text{ and } N(E) = N\,.$$

Now we turn our attention to orthogonal projections, which are peculiar to Hilbert space. A natural question is what conditions are given can make an idempotent become a projection?

Theorem 8.22. *Let H be a Hilbert space. Let $E \in \mathcal{B}(H)$ be an idempotent and $E \neq 0$. Then the following statements are equivalent.*

(a) *E is a projection.*

(b) *$R(E) \perp N(E)$.*

(c) *E is self-adjoint.*

(d) *E is normal.*

(e) *$\|E\| = 1$.*

(f) *$\langle Eh, h\rangle \geq 0$ for all h in H.*

Proof. (a) $\Leftrightarrow$ (b). Clearly (a) implies (b). If $R(E) \perp N(E)$ and $X = R(E) \oplus N(E)$, and $R(E)$, $N(E)$ are closed, there exists unique projection P so that $R(P) = R(E)$ and $N(P) = R(E)^{\perp} = N(E)$. By the uniqueness part of Proposition 8.21, $E = P$. So E is a projection.

(a) $\Leftrightarrow$ (c). In Example 8.15 we have pointed that every projection is self-adjoint. On the other hand, if E is self-adjoint, then for all $Ex \in R(E)$ and $y \in N(E)$,

$$\langle Ex, y\rangle = \langle x, Ey\rangle = 0\,.$$

Thus $R(E) \perp N(E)$. So (c) implies (a).

(a) $\Leftrightarrow$ (d). Clearly (a) implies (d). On the other hand, observe that $R(E)$ is closed since E is idempotent. If E is normal then by Corollary 8.15, we get

$$N(E) \perp \overline{R(E)} = R(E)\,.$$

So (d) implies (a).

(a) $\Leftrightarrow$ (e). Clearly (a) implies (e). On the other hand, for any $x \in R(E)$ and $y \in N(E)$, we have $\|E(x + y)\| = \|x\| \leq \|x + y\|$; thus

$$2\,\mathrm{Re}\langle x, y\rangle \geq \|y\|^2\,.$$

Replace y by ty with $t > 0$, we have $2\operatorname{Re}\langle x, y\rangle \geq t\|y\|^2$. Letting $t \downarrow 0$, we get

$$\operatorname{Re}\langle x, y\rangle \geq 0 \quad \text{for all } x \in R(E)\,, y \in N(E)\,.$$

Since $-y$ is also in $N(E)$, so $\operatorname{Re}\langle x, -y\rangle = -\operatorname{Re}\langle x, y\rangle \geq 0$. $\operatorname{Re}\langle x, y\rangle \leq 0$. Thus we have

$$\operatorname{Re}\langle x, y\rangle = 0 \quad \text{for all } x \in R(E)\,, y \in N(E)\,.$$

If H is real Hilbert space, we get $R(E) \perp N(E)$. If H is complex Hilbert space, then observe that $iy \in N(E)$ and $\operatorname{Im}\langle x, y\rangle = \operatorname{Re}\langle x, iy\rangle = 0$. There still holds $R(E) \perp N(E)$. Thus (e) implies (a).

(a) $\Leftrightarrow$ (f). Clearly (a) implies (f). On the other hand, for any $x \in R(E)$ and $y \in N(E)$, we have $\langle E(x+y), x+y\rangle = \langle x, x+y\rangle \geq 0$, then

$$\langle x, y\rangle \geq -\|x\|^2 \text{ for all } x \in R(E) \text{ and } y \in N(E)\,.$$

Replace x by tx with $t > 0$, we get $\langle x, y\rangle \geq -t\|x\|^2$. Letting $t \downarrow 0$ we get

$$\langle x, y\rangle \geq 0 \text{ for all } x \in R(E) \text{ and } y \in N(E)\,.$$

Since $-x$ is also in $R(E)$, so $\langle -x, y\rangle = -\operatorname{Re}\langle x, y\rangle \geq 0$. $\langle x, y\rangle \leq 0$. Thus

$$\langle x, y\rangle = 0 \text{ for all } x \in R(E) \text{ and } y \in N(E)\,.$$

So (f) implies (a). $\square$

Exercise 8.3. Let H be a complex Hilbert space and $P \in \mathcal{B}(H)$. Then P is a projection if and only if

$$\|Ph\|^2 = \langle Ph, h\rangle \text{ for all } h \in H\,.$$

8.3.1 The Operations of Projections

Let H be a Hilbert space. Let $\{M_\alpha\}_{\alpha\in\Lambda}$ be a collection of closed linear subspaces of H. We denote by $\vee_\alpha M_\alpha$ the closed linear subspace generated by $\{M_\alpha\}_{\alpha\in\Lambda}$; in other words,

$$\bigvee_{\alpha\in\Lambda} M_\alpha := \overline{\operatorname{span}}\{\bigcup_{\alpha\in\Lambda} M_\alpha\}\,.$$

Moreover, if in addition that $\{M_\alpha\}$ are orthogonal to each other, i.e., $M_\alpha \perp M_\beta$ for $\alpha \neq \beta$, then

$$\bigvee_{\alpha\in\Lambda} M_\alpha = \left\{\sum_{\alpha\in\Lambda} x_\alpha : x_\alpha \in M_\alpha,\ \sum_{\alpha\in\Lambda} \|x_\alpha\|^2 < \infty\right\}\,. \tag{8.11}$$

The series $\sum_\alpha x_\alpha$ converges provided that $\{x_\alpha\}$ are pairwise orthogonal and $\sum_\alpha \|x_\alpha\|^2 < \infty$, since for any $\epsilon > 0$, there exists a finite subset S of Λ so that for any finite subset J disjoint with S,

$$\sum_{\alpha\in J} \|x_\alpha\|^2 = \|\sum_{\alpha\in J} x_\alpha\|^2 < \epsilon^2\,,$$

and hence $\|\sum_{\alpha\in J} x_\alpha\| < \epsilon$. The convergence of $\sum_\alpha x_\alpha$ follows form the completeness of H.

Indeed, in order to show (8.11), it suffices to show that the right-hand-side is closed, which follows directly from Proposition 2.5.

Lemma 8.23. *Let $\{M_\alpha\}_{\alpha\in\Lambda}$ be a collection of closed linear subspaces of Hilbert space H. Then*

$$\bigcap_{\alpha\in\Lambda} M_\alpha^\perp = \left(\bigvee_{\alpha\in\Lambda} M_\alpha\right)^\perp \quad \textit{and} \quad \left(\bigcap_{\alpha\in\Lambda} M_\alpha\right)^\perp = \bigvee_{\alpha\in\Lambda} M_\alpha^\perp\,.$$

Proof. Since $M_\alpha \subset \vee_\alpha M_\alpha$, we have $(\vee_\alpha M_\alpha)^\perp \subset M_\alpha^\perp$ for each i. Thus

$$\left(\bigvee_{\alpha\in\Lambda} M_\alpha\right)^\perp \subset \bigcap_{\alpha\in\Lambda} M_\alpha^\perp\,.$$

If $x \in \cap_\alpha M_\alpha^\perp$, then $x \perp (\cup_\alpha M_\alpha)$; i.e., $x \in (\cup_\alpha M_\alpha)^\perp$. By (2.5) we get

$$\bigcap_{\alpha\in\Lambda} M_\alpha^\perp \subset \left(\bigvee_{\alpha\in\Lambda} M_\alpha\right)^\perp\,.$$

Therefore, the first equality holds. To show the second one, replace M_α by $M_\alpha^\perp$ in the first one, we get

$$\bigcap_{\alpha\in\Lambda} M_\alpha = \left(\bigvee_{\alpha\in\Lambda} M_\alpha^\perp\right)^\perp ,$$

and hence

$$\left(\bigcap_{\alpha\in\Lambda} M_\alpha\right)^\perp = \bigvee_{\alpha\in\Lambda} M_\alpha^\perp . \qquad \square$$

Proposition 8.24. *Let P and Q be projections on Hilbert space H. Then $P+Q$ is a projection if and only if*

$$R(P) \perp R(Q) .$$

In this case, then $R(P+Q) = R(P) \oplus R(Q)$ and $N(P+Q) = N(P) \cap N(Q)$.

Proof. Note that $P+Q$ are self-adjoint. Thus $P+Q$ is projection if and only if $P+Q$ is idempotent. Since P and Q are idempotent, $P+Q$ is idempotent if and only if $PQ+QP=0$. If $R(P) \perp R(Q)$, then $PQ = QP = 0$, hence $PQ+QP=0$ is a projection. If $PQ+QP=0$, then

$$QPQ + QP = 0 \ ; \ PQ + QPQ = 0 .$$

So $PQ = QP$, and hence $PQ = QP = 0$. Thus $R(P) \perp R(Q)$. Obviously, in this case $R(P+Q) = R(P) \oplus R(Q)$ and $N(P+Q) = N(P) \cap N(Q)$. $\square$

Proposition 8.25. *Let $\{P_\alpha\}_{\alpha\in\Lambda}$ be a collection of projections on Hilbert space H such that $P_\alpha P_\beta = 0$ if $\alpha \neq \beta$. Then there exists a projection P such that*

$$P = s\text{-}\sum_{\alpha\in\Lambda} P_\alpha ;$$

i.e., $Ph = \sum_{\alpha\in\Lambda} P_\alpha h$ for any $h \in H$. Moreover, if we denote by M_α the range of P_α, then

$$R(P) = \bigoplus_{\alpha\in\Lambda} M_\alpha = \left\{\sum_{\alpha\in\Lambda} x_\alpha : x_\alpha \in M_\alpha, \ \sum_{\alpha\in\Lambda} \|x_\alpha\|^2 < \infty\right\} .$$

Remark 8.3. However, $\sum_{\alpha\in\Lambda} P_\alpha$ does not converge to P in the norm of $\mathcal{B}(H)$. In fact, it never does unless the index set Λ is finite.

Proof. First of all we show that for each h, $\sum_{\alpha\in\Lambda} P_\alpha h$ converges. To see this, note that for any finite subset S of Λ, $\sum_{\alpha\in S} P_\alpha$ is a projection on H and hence

$$\sum_{\alpha\in S} \|P_\alpha h\|^2 = \left\|\sum_{\alpha\in S} P_\alpha h\right\|^2 \leq \|h\|^2 .$$

Taking the limit, we get

$$\sum_{\alpha\in\Lambda} \|P_\alpha h\|^2 \leq \|h\|^2 < \infty .$$

Thus $\sum_{\alpha\in\Lambda} P_\alpha h$ converges. Define $P : H \to H$ by

$$Ph := \sum_{\alpha\in\Lambda} P_\alpha h \quad \text{for all} \quad h \in H .$$

Obviously P is linear. Note that

$$\|Ph\|^2 = \sum_{\alpha\in\Lambda} \|P_\alpha h\|^2 \leq \|h\|^2 ,$$

P is bounded with $\|P\| \leq 1$. Now we have only to show that P is idempotent, then P is a projection by Theorem 8.22. Since $Ph - P_\alpha h \perp M_\alpha$, we have $P_\alpha P h = P_\alpha h$; and hence

$$P^2 h = \sum_{\alpha\in\Lambda} P_\alpha P h = \sum_{\alpha\in\Lambda} P_\alpha h = Ph .$$

By the preceding argument and (8.11), we have

$$R(P) \subset \oplus_\alpha M_\alpha = \left\{ \sum_{\alpha\in\Lambda} x_\alpha : x_\alpha \in M_\alpha,\ \sum_{\alpha\in\Lambda} \|x_\alpha\|^2 < \infty \right\} .$$

On the other hand, it's easy to see that for each i, $M_\alpha \subset R(P)$. Since $R(P)$ is a closed linear subspace, we have

$$\bigoplus_{\alpha\in\Lambda} M_\alpha \subset R(P) ,$$

as desired. $\square$

Proposition 8.26. *Let P, Q be projections on Hilbert space H. Then PQ is a projection if and only if*

$$PQ = QP\,.$$

In this case, $R(PQ) = R(P) \cap R(Q)$ and $N(PQ) = N(P) + N(Q)$.

Proof. If $PQ = QP$, then

$$(PQ)^2 = PQPQ = P^2Q^2 = PQ\,.$$
$$(PQ)^* = Q^*P^* = QP = PQ\,.$$

Thus PQ is self-adjoint and idempotent. Obviously the converse is true. In this case, clearly $R(PQ) \subset R(P) \cap R(Q)$ and $N(P) + N(Q) \subset N(PQ)$.

If $y \in R(P) \cap R(Q)$, then there exists x so that $y = Qx \in R(P)$. So $y = Py = PQx$, and hence $y \in R(PQ)$. So $R(PQ) = R(P) \cap R(Q)$.

If $x \in N(PQ)$ but $x \notin N(P) + N(Q)$, then there exists

$$y \in \{N(P) + N(Q)\}^{\perp} = N(P)^{\perp} \cap N(Q)^{\perp} = R(P) \cap R(Q) = R(PQ)$$

so that $\langle x, y\rangle \neq 0$. Let $y = PQz$, then $\langle x, PQz\rangle = \langle PQx, z\rangle \neq 0$. This is absurd since $x \in N(PQ)$. So $N(P) + N(Q) = N(PQ)$. $\square$

Let M and N are two closed linear subspaces of Hilbert space H. Define

$$M \ominus N = M \cap N^{\perp}\,,$$

which is called the *orthogonal difference* of M and N. Clearly $M \ominus N$ is a closed subspace of H.

Proposition 8.27. *If P and Q are projections, then the following statements are equivalent.*

(a) $P - Q$ *is a projection.*

(b) $R(Q) \subset R(P)$.

(c) $PQ = Q$.

(d) $QP = Q$.

In this case, $R(P-Q) = R(P) \ominus R(Q)$ *and* $N(P-Q) = R(Q) \oplus N(P)$.

Proof. Obviously $P-Q$ is self-adjoint, and $P-Q$ is idempotent if and only if

$$PQ + QP = 2Q\,. \tag{8.12}$$

It not hard to see that properties (b), (c), (d) are equivalent and imply (8.12). On the other hand,if (8.12) holds, then

$$PQ + QPQ = 2Q\ ;\ QPQ + QP = 2Q\,.$$

So $PQ = QP$ and hence $PQ = QP = Q$.

For $y \in R(P) \ominus R(Q) = R(P) \cap N(Q)$, there exists x so that $y = Px \in N(Q)$. So $QPx = Qx = 0$ and hence $y = (P-Q)x \in R(P-Q)$. Thus $R(P) \ominus R(Q) \subset R(P-Q)$. For $y \in R(P-Q)$. there exists x so that $y = Px - Qx$, so $y = Px - PQx = P(x - Qx) \in R(P)$. Since $Qy = QPx - Qx = 0$, $y \in N(Q) = R(Q)^{\perp}$. Thus $R(P-Q) \subset R(P) \ominus R(Q)$.

Clearly $R(Q) \subset N(P-Q)$ and $N(P) \subset N(P-Q)$, so $R(Q) \oplus N(P) \subset N(P-Q)$. If $x \in N(P-Q)$, then $Px = Qx$. Hence $x = Qx + (x - Px)$ with $Qx \in R(Q)$ and $(x - Px) \in N(P)$, So $N(P-Q) \subset R(Q) \oplus N(P)$. □

Exercise 8.4. $\{P_n\}$ are projections and is monotonous, i.e. $P_n \leq P_{n+1}$ for all n or $P_n \geq P_{n+1}$ for all n, then there exists a projection P such that $P_n \xrightarrow{s} P$.

8.3.2 Invariant and Reducing Subspaces

Let H be a Hilbert space. Let A be a bounded linear operator on H. Let M be a closed linear subspace of H.

- Say that M is an *invariant subspace* for A if

$$Ah \in M \quad \text{whenever } h \in M\,.$$

In other words, $A(M) \subset M$.

- Say that M is a *reducing subspace* for A if

$$Ah \in M \quad \text{for all } h \in M \quad \text{and} \quad Ah \in M^{\perp} \quad \text{for all } h \in M^{\perp}.$$

In other words, $A(M) \subset M$ and $A(M^{\perp}) \subset M^{\perp}$.

Note that $H = M \oplus M^{\perp}$. If $A \in \mathcal{B}(H)$, then A can be written as a 2×2 matrix with operator entries,

$$A = \begin{bmatrix} W & X \\ Y & Z \end{bmatrix}, \tag{8.13}$$

where

$$W = P_M A|_M \in \mathcal{B}(M) \ ; \ X = P_M A|_{M^{\perp}} \in \mathcal{B}\left(M^{\perp}, M\right)$$
$$Y = (I - P_M)A|_M \in \mathcal{B}\left(M, M^{\perp}\right) \ ; \ Z = (I - P_M)A|_{M^{\perp}} \in \mathcal{B}\left(M^{\perp}\right).$$

Then M is invariant for A if and only if $Y = 0$, $W = A|_M$. That is

$$A = \begin{bmatrix} W & X \\ 0 & Z \end{bmatrix}.$$

M reduces A if and only if $Y = Z = 0$ and $W = A|_M$, $Z = A|_{M^{\perp}}$. That is

$$A = \begin{bmatrix} W & 0 \\ 0 & Z \end{bmatrix}.$$

This is the reason for the terminology.

Proposition 8.28. *Let M be a closed linear subspace of Hilbert space H and $A \in \mathcal{B}(H)$. Then M is invariant for A if and only if*

$$P_M A P_M = A P_M.$$

Proof. If M is invariant for A, then $R(AP_M) \subset M$, and hence $P_M A P_M = AP_M$. Conversely, for each $h \in M$ we have $P_M Ah = Ah$, so $Ah \in M$. Thus M is invariant for A. $\square$

Proposition 8.29. *Let M be a closed linear subspace of Hilbert space H and $A \in \mathcal{B}(H)$. Then the following statements are equivalent.*

(a) *M is a reducing subspace for A.*

(b) *$P_M A = AP_M$.*

(c) *M is invariant for both A and A^*.*

Proof. (a) $\Rightarrow$ (b). Since M is a reducing subspace for A, we have $P_M A P_M = AP_M$ and $P_M A(I - P_M) = 0$. Thus

$$P_M A = P_M[AP_M + A(I - P_M)] = AP_M \, .$$

(b) $\Rightarrow$ (c). Since $P_M A = AP_M$, we have $P_M A P_M = AP_M^2 = AP_M$ and hence M is invariant for A. Since $P_M A = AP_M$, we have $A^* P_M = P_M A^*$. So M is invariant for A^*.

(c) $\Rightarrow$ (a). It suffices to show that for $h \in M^\perp$, $Ah \in M^\perp$. Take any $g \in M$, then

$$\langle Ah, g \rangle = \langle h, A^* g \rangle = 0$$

since $A^* g \in M$. Notice that g was an arbitrary vector in M, so $Ah \in M^\perp$. □

Remark 8.4. In fact, though the preceding proof we have shown that M is invariant for A^* if and only if $M^\perp$ is invariant for A. In fact, if A has the form (8.13), then

$$A^* = \begin{bmatrix} W^* & Y^* \\ X^* & Z^* \end{bmatrix} .$$

8.4 The Spectrum of a Normal Operator

In this section, we always set H a complex Hilbert space. We begin by characterizing the diagonalizable operator.

A *partition of the identity* on H is a family $\{P_\alpha\}$ of pairwise orthogonal projections on H such that $I = \text{s-}\sum_\alpha P_\alpha$. By Proposition 8.25, this might be indicated by $\vee_\alpha R(P_\alpha) = H$.

An operator A on H is *diagonalizable* if there is a partition of the identity $\{P_\alpha\}$ on H, and a family of scalars $\{\lambda_\alpha\}$ such that $\sup_\alpha |\lambda_\alpha| < \infty$ and

$$Ah = \lambda_\alpha h \quad \text{whenever } h \in R(P_\alpha)\,.$$

Clearly, in this case, $Ah = A\sum_\alpha P_\alpha h = \sum_\alpha AP_\alpha h = \sum_\alpha \lambda_\alpha P_\alpha h$ for all $h \in H$; in other words,

$$A = \text{s-}\sum_\alpha \lambda_\alpha P_\alpha\,. \tag{8.14}$$

Note that it was not assumed that the scalars $\{\lambda_\alpha\}$ in are distinct. There is no loss in generality in assuming this, however. In fact, if $\lambda_\alpha = \lambda_\beta$, then we can replace P_α and P_β with $P_\alpha + P_\beta$.

Exercise 8.5. An operator A on H is diagonalizable if and only if there is an orthonormal basis for H consisting of eigenvectors for A.

It's not hard to see that every diagonalizable operator is normal. Indeed, if A has the form (8.14), then

$$A^* = \text{s-}\sum_\alpha \bar{\lambda}_\alpha P_\alpha\,,$$

and

$$A^*A = AA^* = \text{s-}\sum_\alpha |\lambda_\alpha|^2 P_\alpha\,.$$

Surprisingly, the converse is true if in addition A is compact. The proof requires a few preliminary results.

Theorem 8.30. *Let H be a complex Hilbert space and let $A \in \mathcal{B}(H)$ be a normal operator. Then the following hold.*

(a) $\|A^n\| = \|A\|^n$ *for all* $n \in \mathbb{N}$

(b) $r_\sigma(A) = \max_{\lambda\in\sigma(A)} |\lambda| = \|A\|$.

(c) $\sigma_p(A^*) = \{\bar{\lambda} \mid \lambda \in \sigma_p(A)\}$ *and* $\sigma_r(A^*) = \sigma_r(A) = \emptyset$.

(d) $\lambda \in \varrho(A)$ *if and only if there exits a constant* $c > 0$ *so that*

$$\|\lambda h - Ah\| \geq c\|h\| \quad \textit{for all } h \in H\,. \tag{8.15}$$

(e) $\lambda \in \sigma(A)$ *if and only if there exits a sequence* $\{h_n\}$ *in* H *so that* $\|h_n\| = 1$ *and*

$$\|\lambda h_n - Ah_n\| \to 0 \quad \textit{as} \quad n \to \infty\,.$$

Proof. To show part (a), we prove $\|A^2\| = \|A\|^2$ first. By Theorem 8.14 we have

$$\|A^2 h\| = \|A^* A h\| \text{ for all } h \in H\,.$$

Thus $\|A^2\| = \|A^*A\| = \|A\|^2$ by Corollary 8.13. Hence it follows by induction that such that $n < 2^m$, and deduce that

$$\|A\|^{2^m - n}\|A\|^n = \left\|A^{2^m}\right\| \leq \|A^n\|\,\|A\|^{2^m - n}$$

Hence $\|A\|^n \leq \|A^n\| \leq \|A\|^n$ and so $\|A^n\| = \|A\|^n$. This proves part (a).

Part (b) follows directly from part (a) and Theorem 6.6.

To prove part (c), fix an element $\lambda \in \mathbb{C}$. Then $(\lambda I - A)^* = \bar{\lambda} I - A^*$ and hence $\lambda I - A$ is normal. It follows from Corollary 8.15 that $\sigma_p(A^*) = \{\bar{\lambda} \mid \lambda \in \sigma_p(A)\}$ and $\sigma_r(A^*) = \sigma_r(A) = \emptyset$.

Finally we show part (d), since part (e) follows directly form part (d). Clearly if $\lambda \in \varrho(A)$, then (8.15) holds. To prove the converse, note that (8.15) implies that $N(\lambda I - A) = \{0\}$ and $R(\lambda I - A)$ is closed. Since $\lambda I - A$ is also normal, by Corollary 8.15, $R(\lambda I - A) = H$. Hence $\lambda \in \varrho(A)$ by the inverse mapping theorem. $\square$

Remark 8.5. By part (b) of Theorem 8.30 and the Riesz-Schauder theory, if A is a compact normal operator on the complex Hilbert space H, then either $\pm\|A\|$ is an eigenvalue of A.

Theorem 8.31. *Let A be a normal operator on the complex Hilbert space H. Let $\lambda \in \sigma_p(A)$. Then the following hold.*

(a) The eigenspace $N(\lambda I - A)$ reduces A.

(b) Each generalized eigenvector is an eigenvector, i.e.,

$$N(\lambda I - A)^n = N(\lambda I - A) \quad \textit{for all} \quad n \in \mathbb{N}\,.$$

Proof. To prove part (a), observe that $N(\lambda I - A)$ is invariant for A since A and $\lambda I - A$ commutes. Since A is normal, A^* commutes with $\lambda I - A$, and hence $N(\lambda I - A)$ is invariant for A^*. Thus $N(\lambda I - A)$ reduces A.

To prove part (b), it suffices to show that $N(\lambda I - A)^2 = N(\lambda I - A)$. By Theorem 8.14 (or Exercise 8.2), since $\lambda I - A$ is also a normal operator, for $x \in N(\lambda I - A)^2$, we have

$$\left(\bar{\lambda} I - A^*\right)(\lambda I - A)x = 0\,.$$

Hence

$$\|\lambda x - Ax\|^2 = \langle \lambda x - Ax, \lambda x - Ax\rangle = \left\langle x, \left(\bar{\lambda} I - A^*\right)(\lambda I - A)x\right\rangle = 0$$

and hence $x \in N(\lambda I - A)$, as desired. □

Theorem 8.32. *Let A be a normal operator on the complex Hilbert space H. Then, the eigenspaces of A are pairwise orthogonal: for $\lambda, \mu \in \sigma_p(A)$ such that $\lambda \neq \mu$, we have*

$$N(\lambda I - A) \perp N(\mu I - A)\,.$$

Proof. Choose $x, y \in H$ such that $Ax = \lambda x$ and $Ay = \mu y$,. We shall show that $\langle x, y\rangle = 0$. By Theorem 8.14, since $\mu I - A$ is also a normal operator and $(\mu I - A)^* = \bar{\mu} I - A^*$, we have $A^* y = \bar{\mu} y$. Then

$$\lambda\langle x, y\rangle = \langle Ax, y\rangle = \langle x, A^* y\rangle = \mu\langle x, y\rangle\,.$$

Since $\lambda \neq \mu$, we have $\langle x, y\rangle = 0$. This proves the desired result. □

Theorem 8.33 (Spectral Theorem for Compact Normal Operators)**.** *Let A be a compact normal operator on the complex Hilbert space H. Then A is diagonalizable. More precisely, if $\{\lambda_n\}_{n\in\Lambda} \subset \mathbb{C}$ are the distinct nonzero eigenvalues of A so that Λ either equal to $\mathbb{N}$ or equal to $\{1, 2, \cdots, l\}$ for some $l \in \mathbb{N}$, and P_n is the projection of H onto $N(\lambda_n I - A)$, then $P_n P_m = P_m P_n = 0$ if $n \neq m$ and*

$$A = \sum_{n\in\Lambda} \lambda_n P_n\,, \tag{8.16}$$

where this series converges to A in the operator norm on $\mathcal{B}(H)$ if $\Lambda = \mathbb{N}$.

Proof. Step 1. Let $\lambda_0 = 0$ and P_0 the projection of H onto $N(A)$. It suffices to show that

$$\bigoplus_{n\in\Lambda\cup\{0\}} N(\lambda_n I - A) = H\,. \tag{8.17}$$

Then $\{P_n\}_{n\in\Lambda\cup\{0\}}$ is a partition of identity on H by Theorem 8.32 and Proposition 8.25, and hence

$$A = \text{s-}\sum_{n\in\Lambda\cup\{0\}} \lambda_n P_n = \text{s-}\sum_{n\in\Lambda} \lambda_n P_n\,.$$

On the other hand, if Λ is finite then clearly $A = \sum_{n\in\Lambda} \lambda_n P_n$; if $\Lambda = \mathbb{N}$, then by the Riesz-Schauder theory $|\lambda_n| \to 0$ as $n \to \infty$, and hence $\sum_{n\in\mathbb{N}} \lambda_n P_n$ converges to A in operator norm since

$$\left\|\sum_{n=k}^{k+p} \lambda_n P_n\right\| \le \sup_{n\ge k} |\lambda_n|\,, \text{ for all } k, p \ge 1.$$

Step 2. Now we have only to show (8.17). To this end, it suffices to show that

$$H_0 \coloneqq \left(\bigoplus_{n\in\Lambda} N(\lambda_n I - A)\right)^{\perp} = N(A)\,. \tag{8.18}$$

Clearly $N(A) \subset H_0$ by Theorem 8.32. To prove the converse, observe that H_0 is a closed A-invariant subspace of H and

$$A_0 \coloneqq A|_{H_0} : H_0 \to H_0$$

is a compact normal operator. Suppose, by contradiction, that $A_0 \neq 0$. Then it follows from part (b) of Theorem 8.30 and the Riesz-Schauder theory that A_0 has a nonzero eigenvalue. This contradicts the definition of H_0 and proves (8.18). □

Remark 8.6. If A is a compact normal operator, then there is a sequence $\{\lambda_n\}_{n\in\Lambda}$ of nonzero complex numbers with

$$\#\{n \in \Lambda : \lambda_n = \lambda\} = \dim N(\lambda I - A) \qquad \text{for all } \lambda \in \mathbb{C}\backslash\{0\};$$

and an orthonormal basis $\{e_n\}_{n\in\Lambda}$ for $N(A)^{\perp} = N(A^*)^{\perp} = \overline{R(A)}$ such that

$$Ae_n = \lambda_n e_n \quad \text{for all } n \in \Lambda,$$

and

$$Ah = \sum_{n\in\Lambda} \lambda_n \langle h, e_n\rangle e_n \quad \text{for all } h \in H.$$

Corollary 8.34. *If A is a compact operator on a complex Hilbert space, then A is diagonalizable if and only if A is normal.*

Remark 8.7. If A is a normal operator which is not necessarily compact, there is a spectral theorem for A which has a somewhat different form. This theorem states that A can be represented as an integral with respect to a measure whose values are not numbers but projections on a Hilbert space. Theorem 8.33 will be a consequence of this more general theorem and correspond to the case in which this projection-valued measure is "atomic".

8.5 The Spectrum of a Self-Adjoint Operator

Let H be a Hilbert space and $A \in \mathcal{B}(H)$ is self-adjoint. Recall that we have shown that $\sigma(A) \subset \mathbb{R}$ in Example 6.8. In this section, we give more results about the spectrum $\sigma(A)$.

Theorem 8.35. *Let H be a complex Hilbert space and let $A \in \mathcal{B}(H)$ be a self-adjoint operator. Then $\sigma(A)$ is a compact subset of $\mathbb{R}$. Moreover,*

$$\sup \sigma(A) = \max \sigma(A) = \sup_{\|h\|=1} \langle Ah, h\rangle\,;$$

$$\inf \sigma(A) = \min \sigma(A) = \inf_{\|h\|=1} \langle Ah, h\rangle\,.$$

Proof. Take any $\lambda \in \mathbb{C}\backslash\mathbb{R}$, then for all $h \in H$, we have

$$\langle(\lambda I - A)h, h\rangle = \lambda\|h\|^2 - \langle Ax, h\rangle = \mathrm{Re}(\lambda)\|h\|^2 - \langle Ah, h\rangle + i\,\mathrm{Im}(\lambda)\|h\|^2\,.$$

Hence by Cauchy-Schwarz inequality,

$$\|(\lambda I - A)h\| \cdot \|h\| \geq |\langle(\lambda I - A)h, h\rangle| \geq |\,\mathrm{Im}(\lambda)| \cdot \|h\|^2\,,$$

which implies that

$$\|\lambda h - Ah\| \geq |\,\mathrm{Im}\,\lambda|\|h\|\,, \text{ for all } h \in H\,.$$

By part (d) of Theorem 8.30, $\lambda \in \varrho(A)$.

Let $m \coloneqq \inf_{\|h\|=1}\langle Ah, h\rangle$ and $M = \inf_{\|h\|=1}\langle Ah, h\rangle$. Using the preceding argument, we can show that $\sigma(A) \subset [m, M]$. For example, if $\lambda > M$, then for all $h \in H$,

$$\|(\lambda I - A)h\| \cdot \|h\| \geq \langle(\lambda I - A)h, h\rangle = \lambda\|h\|^2 - \langle Ah, h\rangle \geq \lambda\|h\|^2 - M\|h\|^2\,.$$

Hence

$$\|\lambda h - Ah\| \geq (\lambda - M)\|h\|\,, \text{ for all } h \in H\,.$$

By part (d) of Theorem 8.30, $\lambda \in \varrho(A)$.

It suffices to show that $m, M \in \sigma(A)$. To this end, observe that if $m = 0$, i.e., $\langle Ah, h\rangle \geq 0$ for all h, then by Theorem 9.52 in this case $M = \|A\|$. By $\sigma(A) \subset [0, M]$ and part (b) of Theorem 8.30, $M = \|A\| \in \sigma(A)$. If $m < 0$, then $T = A - mI$ is a self-adjoint operator with $\langle Th, h\rangle \geq 0$. Moreover,

$$\sup_{\|h\|=1} \langle Th, h\rangle = \sup_{\|h\|=1} \langle Ah, h\rangle - m = M - m\,.$$

By the preceding argument, we get $M-m \in \sigma(T) = \sigma(A)-m$, so $M \in \sigma(A)$. So in any case we have $M \in \sigma(A)$. On the other hand, since $-A$ is self-adjoint and

$$\sup_{\|h\|=1} \langle -Ah, h\rangle = -\sup_{\|h\|=1} \langle Ah, h\rangle = -m\,,$$

by the preceding argument, we get $-m \in \sigma(-A) = -\sigma(A)$, so $m \in \sigma(A)$. $\square$

The following spectral theorem for self-adjoint operators follows directly form Theorem 8.33, and note that all the eigenvalues of self-adjoint operator are real numbers.

Theorem 8.36 (Spectral Theorem for Compact Self-Adjoint Operators)**.** *Let A be a compact self-adjoint operator on the complex Hilbert space H. Suppose $\{\lambda_n\}_{n\in\Lambda} \subset \mathbb{R}$ are the distinct nonzero eigenvalues of A so that Λ either equal to $\mathbb{N}$ or equal to $\{1, 2, \cdots, l\}$ for some $l \in \mathbb{N}$, and P_n is the projection of H onto $N(\lambda_n I - A)$, then $P_nP_m = P_mP_n = 0$ if $n \neq m$ and*

$$A = \sum_{n\in\Lambda} \lambda_n P_n\,,$$

where this series converges to A in the operator norm on $\mathcal{B}(H)$ if $\Lambda = \mathbb{N}$.

Remark 8.8. Without loss of generality, we can number the distinct nonzero eigenvalues in decreasing order of absolute value; that is

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_{n-1}| \geq |\lambda_n| \geq \cdots$$

In this case, the eigenvalues have the following extreme properties:

$$|\lambda_1| = \sup_{\|h\|=1} |\langle Ah, h\rangle|\,,$$

and

$$|\lambda_n| = \sup\left\{|\langle Ah, h\rangle| : h \perp R(P_m), 1 \leq m < n; \|h\| = 1\right\}\,.$$

Exercise 8.6. We can arrange the eigenvalues by positive and negative values, the eigenvalue will be written the same times in succession as it's multiplicity, denoting them by

$$\lambda_1^+ \geq \lambda_2^+ \geq \cdots \geq 0\,;$$
$$\lambda_1^- \leq \lambda_2^- \leq \cdots < 0\,.$$

Then

$$\lambda_n^+ = \inf_M \sup_{\substack{h \in M^\perp \\ \|h\|=1}} \langle Ah, h\rangle\,;$$

$$\lambda_n^- = \sup_M \inf_{\substack{h \in M^\perp \\ \|h\|=1}} \langle Ah, h\rangle\,.$$

where M runs over all closed subspaces of H with dimension $n-1$.

Singular Value Decomposition Let X, Y be complex Hilbert spaces and $T \in \mathcal{B}(X,Y)$. A real number $\lambda \geq 0$ is called a *singular value* of T if

$$\lambda^2 \in \sigma\left(T^*T\right)\,.$$

Thus the singular values of T are the square roots of the (nonnegative) spectral values of the self-adjoint operator $T^*T \in \mathcal{B}(H)$. By Theorem 9.52,

$$\|T^*T\| = \sup_{\|x\|=1} \langle T^*Tx, x\rangle = \sup_{\|x\|=1} \|Tx\|^2 = \|T\|^2\,.$$

As a consequence of part (b) of Theorem 8.30, the maximum of the singular values is the norm of T, since

$$\max\{\lambda \geq 0 : \lambda^2 \in \sigma\left(T^*T\right)\} = r_\sigma(T^*T)^{1/2} = \|T^*T\|^{1/2} = \|T\|\,.$$

Theorem 8.37 (Singular Value Decomposition for Compact Operators)**.** *Let X, Y be complex Hilbert spaces. Let $K \in \mathcal{B}(X,Y)$ and $K \neq 0$. If K is compact, then there exists a set $\Lambda \subset \mathbb{N}$, either equal to $\mathbb{N}$ or equal to $\{1,\ldots,l\}$ for some $l \in \mathbb{N}$, orthonormal sequences $(x_n)_{n\in\Lambda}$ in X and $(y_n)_{n\in\Lambda}$ in Y, and a sequence $(\lambda_n)_{n\in\Lambda}$ of all positive singular values of K such that*

$$Kx = \sum_{n\in\Lambda} \lambda_n \langle x, x_n\rangle\, y_n \quad \text{for all } x \in X\,. \tag{8.19}$$

Proof. Step 1. We shall find the sequences $(\lambda_n)_{n\in\Lambda}$, $(x_n)_{n\in\Lambda}$ and $(y_n)_{n\in\Lambda}$.

Consider the operator $K^*K \in \mathcal{B}(X)$. This operator is self-adjoint and is compact. Hence $\sigma(K^*K)\setminus\{0\}$ is a discrete subset of the positive real axis $(0,\infty)$ by the Riesz-Schauder theory. Write $\sigma(K^*K)\setminus\{0\} = (\lambda_n)_{n\in\Lambda}$ where $\Lambda = \mathbb{N}$ when the spectrum is infinite and $\Lambda = \{1,\dots,l\}$ otherwise, the λ_n are chosen positive, and

$$\#\{n\in\Lambda : \lambda_n = \lambda\} = \dim N\left(\lambda^2 I - K^*K\right) \qquad \text{for all } \lambda > 0\,.$$

By Remark 8.6, there exists an orthonormal basis $(x_n)_{n\in\Lambda}$ for $N(K^*K)^\perp \subset X$ such that

$$K^*Kx_n = \lambda_n^2 x_n \quad \text{for all} \quad n\in\Lambda\,,$$

and

$$K^*Kx = \sum_{n\in\Lambda} \lambda_n^2 \langle x, x_n\rangle\, x_n \quad \text{for all} \quad x\in X\,.$$

Observe that

$$\langle Kx_n, Kx_m\rangle = \langle K^*Kx_n, x_m\rangle = \lambda_n^2\langle x_n, x_m\rangle = \lambda_n^2\delta_{nm}\,,$$

thus $(Kx_n)_{n\in\Lambda}$ is an orthogonal sequence in Y. We normalize it by defining

$$y_n := \frac{1}{\lambda_n} Kx_n \quad \text{for all} \quad n\in\Lambda\,.$$

Then $(y_n)_{n\in\Lambda}$ is an orthonormal sequence in Y with $K^*y_n = \lambda_n x_n$ for all n.

Step 2. We now show (8.6). It suffices to show that

$$M := \overline{\operatorname{span}}\{y_n : n\in\Lambda\} = \overline{R(K)} \tag{8.20}$$

Then by Theorem 2.15, for every $x\in X$,

$$Kx = \sum_{n\in\Lambda}\langle Kx, y_n\rangle y_n = \sum_{n\in\Lambda}\langle x, K^*y_n\rangle y_n = \sum_{n\in\Lambda}\lambda_n\langle x, x_n\rangle y_n\,.$$

In order to prove (8.20), since $M\subset\overline{R(K)}$ it suffices to show that

$$\overline{R(K)}\ominus M = \overline{R(K)}\cap M^\perp = \{0\}\,.$$

If $z \in \overline{R(K)} \ominus M$, then $\langle z, y_n \rangle = 0$ for all n and hence $\langle K^* z, x_n \rangle = 0$ for all n. In other words,

$$K^* z \in \overline{\operatorname{span}}\{x_n : n \in \Lambda\}^{\perp} = N(K^*K)^{\perp\perp} = N(K^*K) = N(K)\,.$$

Thus $KK^*z = 0$; i.e.,

$$z \in N(KK^*) = N(K^*)\,.$$

However, since $\overline{R(K)}^{\perp} = N(K^*)$, so

$$z \in \overline{R(K)} \cap \overline{R(K)}^{\perp} = \{0\}\,,$$

as desired. We now complete the proof. $\square$

Remark 8.9. In fact, to show (8.19), we can check directly that

$$\begin{aligned}
&\|Kx - \sum_{n\in\Lambda} \lambda_n \langle x, x_n\rangle\, y_n\|^2 \\
&= \|Kx\|^2 + \sum_{n\in\Lambda} \lambda_n^2 \,|\langle x, x_n\rangle|^2 - 2\sum_{n\in\Lambda} \lambda_n \operatorname{Re}\Big(\overline{\langle x, x_n\rangle}\,\langle Kx, y_n\rangle\Big) \\
&= \|Kx\|^2 - \sum_{n\in\Lambda} \lambda_n^2 \,|\langle x, x_n\rangle|^2 = 0\,,
\end{aligned}$$

since

$$\|Kx\|^2 = \langle K^*Kx, x\rangle = \left\langle \sum_{n\in\Lambda} \lambda_n^2 \langle x, x_n\rangle\, x_n, x \right\rangle = \sum_{n\in\Lambda} \lambda_n^2 \,|\langle x, x_n\rangle|^2\ .$$

Chapter 9

Banach Algebra and C* Algebra

9.1 Definition and Examples

Recall that we say $(\mathscr{A}, +, \cdot)$ is an (*associative*) *algebra* over the filed $\mathbb{F}$, if the following statements holds

(i) $(\mathscr{A}, +)$ is a vector space over $\mathbb{F}$.

(ii) $(\mathscr{A}, +, \cdot)$ is a ring.

(iii) The scalar multiplication and vector multiplication satisfy that

$$\lambda(x \cdot y) = (\lambda x) \cdot y = x \cdot (\lambda y)\,,$$

for all $x, y \in \mathscr{A}$ and $\lambda \in \mathbb{F}$.

If in addition, $(\mathscr{A}, +, \cdot)$ is a commutative ring, we say $(\mathscr{A}, +, \cdot)$ is a *commutative algebra*. An element $e \in \mathscr{A}$ is a *unit element* (or *identity*) if

$$e \cdot x = x \cdot e = x \quad \text{for all } x \in \mathscr{A}\,.$$

In this case, we say $\mathscr{A}$ is *unital*. It's easy to see that the unit element must be unique in $\mathscr{A}$. In this note, we always assume that a unital algebra is nonzero, i.e., the unital element is not equal to zero.

Let $\mathscr{A}$ be an algebra with a unit e. Then $x \in \mathscr{A}$ is called *invertible* if there exists some $y \in \mathscr{A}$, called the *inverse* of x, so that

$$x \cdot y = y \cdot x = e\,.$$

It's easy to see that if x is invertible, then the inverse of x is unique. If in addition, every nonzero element in $\mathscr{A}$ is invertible, then $\mathscr{A}$ is called a *division algebra*.

Let $\mathscr{A}, \mathscr{B}$ be two unital algebra over $\mathbb{F}$. We say ψ is a (algebra) *homomorphism* of $\mathscr{A}$ into $\mathscr{B}$, if ψ protects the algebraic operations. In other words for all $x, y \in \mathscr{A}$ and $\alpha, \beta \in \mathbb{F}$,

$$\psi(\alpha x + \beta y) = \alpha\psi(x) + \beta\psi(y)\ ;\ \psi(xy) = \psi(x)\psi(y)\,.$$

If in addition ψ is bijective, we call it an *isomorphism*. In this case, we can see that ψ maps the identity in $\mathscr{A}$ into the identity in $\mathscr{B}$ if ψ is nonzero.

Let $\mathscr{A}$ be a algebra over $\mathbb{F}$. $\mathscr{B} \subset \mathscr{A}$ is called a *subalgebra*, if $\mathscr{B}$ equipped with the operations inherited form $\mathscr{A}$ is an algebra. In other words, a subalgebra of an algebra is a subset of elements that is closed under addition, multiplication, and scalar multiplication.

Definition 9.1. Let $\mathscr{A}$ be an algebra over $\mathbb{F}$, $\|\cdot\|$ a norm on $\mathscr{A}$. We say that $(\mathscr{A}, \|\cdot\|)$ is a *normed algebra* over $\mathbb{F}$, if the norm satisfies the multiplicative inequality:

$$\|x \cdot y\| \leq \|x\|\|y\| \tag{9.1}$$

for all $x, y \in \mathscr{A}$. If in addition, $\mathscr{A}$ is complete with respect to the norm, we call $\mathscr{A}$ a *Banach algebra*.

Remark 9.1. If $\mathscr{A}$ is a normed algebra with a unit e, then clearly $\|e\| \geq 1$. We can always assume that

$$\|e\| = 1$$

(without change the topology on $\mathscr{A}$). We will explain this in Example 9.4.

Remark 9.2. The inequality (9.1) makes multiplication a continuous operation in $\mathscr{A}$. This means that if $x_n \to x$ and $y_n \to y$ then $x_n y_n \to xy$.

On the contrary, if $\mathscr{A}$ is an algebra endowed with a norm $\|\cdot\|$ so that the multiplication is continuous, then there exists a constant $C > 0$ so that

$$\|xy\| \leq C\|x\|\|y\|$$

for all $x, y \in \mathscr{A}$. In fact if not, for each $n \in \mathbb{N}$, there exists x_n, y_n so that $\|x_n\| = 1$, $\|y_n\| = 1$ and $\|x_n y_n\| \geq n^2$. Then $\frac{x_n}{n} \to 0$, $\frac{y_n}{n} \to 0$ but $\|\frac{x_n y_n}{n^2}\| \geq 1$, which contradicts to that the multiplication is continuous.

Remark 9.3. Every normed algebra $\mathscr{A}$ can be regarded as an subalgebra of some unital algebra. Indeed, let $\mathscr{A}_1 = \mathscr{A} \times \mathbb{F}$. Define algebraic operations on $\mathscr{A}_1$ by

(i) $(x, \alpha) + (y, \beta) = (x + y, \alpha + \beta)$;

(ii) $\beta(x, \alpha) = (\beta x, \beta\alpha)$;

(iii) $(x, \alpha)(y, \beta) = (xy + \alpha y + \beta x, \alpha\beta)$.

Define

$$\|(x, \alpha)\| = \|x\| + |\alpha| \,.$$

Then $\mathscr{A}_1$ with this norm and the algebraic operations defined in (i), (ii), and (iii) is a normed algebra with identity $(0, 1)$ and $\|(0, 1)\| = 1$. Moreover, $\mathscr{A}$ is a Banach algebra, then so is $\mathscr{A}_1$; also $\mathscr{A}_1$ if commutative if $\mathscr{A}$ is.

The mapping $x \mapsto (x, 0)$ is an *isometric isomorphism* of $\mathscr{A}$ onto a subspace of $\mathscr{A}_1$ (in fact, onto a closed two-sided ideal of $\mathscr{A}_1$) whose codimension is 1. If x is identified with $(x, 0)$ then $\mathscr{A}_1$ is simply $\mathscr{A}$ plus the one-dimensional vector space generated by e. For concrete examples, see Example 9.2 and 9.5.

We now give some examples for Banach algebra.

Example 9.1. Let X be a compact Hausdorff space. Then the Banach space $C(X, \mathbb{F})$ of all complex continuous functions on X, with the supremum norm and multiplication defined in the usual way: $(fg)(x) = f(x)g(x)$, is a unital commutative Banach algebra. The constant function 1 is the unit element.

If X is a finite set with the discrete topology, consisting of, say, d points, then $C(X, \mathbb{F})$ is simply $\mathbb{F}^d$, with coordinatewise multiplication. In particular, when $d = 1$, we obtain the simplest Banach algebra, namely $\mathbb{F}$, with the absolute value as norm.

Example 9.2. If X is a LCH space, then the Banach space $\mathscr{A} = C_0(X; \mathbb{F})$ of all complex continuous functions on X vanishing at infinity, with the supremum norm, is a Banach algebra when the multiplication is defined pointwise as in the preceding example. $\mathscr{A}$ is commutative, but if X is not compact, $\mathscr{A}$ does not have an identity.

One can adjoin a uint element by the abstract procedure outlined in Remark 9.3 or one can do it more concretely by enlarging $C_0(X; \mathbb{F})$ to $C(X_\infty; \mathbb{F})$, where X_∞ is the one-point compactification of X. Indeed, for each $f \in C_0(X; \mathbb{F})$, we can extend f on X_∞ by defined $f(\infty) = 0$ and then the extended function f belongs to $C(X_\infty; \mathbb{F})$. Note that for each $f \in C(X_\infty; \mathbb{F})$, $g_f = f - f(\infty) \in C_0(X; \mathbb{F})$, and then

$$f = g + f(\infty) \cdot 1$$

where 1 is the constant 1 function.

Example 9.3. Let (X, Ω, μ) be a σ-finite measure space and $\mathscr{A} = L^\infty(\mu; \mathbb{F})$. Then $\mathscr{A}$ is an commutative Banach algebra with identity if the operations are defined pointwise.

Example 9.4. Let X be a Banach space over $\mathbb{F}$. Then $\mathcal{B}(X)$, the algebra of all bounded linear operators on X, is a unital Banach algebra, with respect to the usual operator norm. The identity operator I is its unit element. If X has finite dimension, then $\mathcal{B}(X)$ is (isomorphic to) the algebra of all complex $n \times n$-matrices. If $\dim X \geq 2$, then $\mathcal{B}(X)$ is not commutative.

Every closed subalgebra of $\mathcal{B}(X)$ that contains I is also a unital Banach algebra. In fact, that every unital Banach algebra is isomorphic and homeomorphism to one of these:

Let $\mathscr{A}$ be a Banach algebra. Assign to each $x \in \mathscr{A}$ the left-multiplication operator M_x defined by

$$M_x(a) = xa \quad \text{for} \quad a \in \mathscr{A} \, .$$

Clearly $M_x \in \mathcal{B}(\mathscr{A})$, the Banach space of all continuous linear operators on $\mathscr{A}$, with $\|M_x\| \leq \|x\|$.

Let $\tilde{\mathscr{A}}$ be the set of all M_x. It is clear that $x \to M_x$ is linear. The associative law implies that $M_{xy} = M_x M_y$. On the other hand, note that

$$\|x\| = \|xe\| = \|M_x e\| \leq \|M_x\| \, \|e\| \, ,$$

we get

$$\|M_x\| \geq \frac{1}{\|e\|} \|x\| \, .$$

Thus the mapping

$$x \mapsto M_x \ ; \ \mathscr{A} \to \tilde{\mathscr{A}}$$

is both a isomorphism and a homeomorphism of $\mathscr{A}$ onto $\tilde{\mathscr{A}} \subset \mathcal{B}(\mathscr{A})$. Then it follows that $\tilde{\mathscr{A}}$ is a closed unital subalgebra of $\mathcal{B}(\mathscr{A})$.

Moreover, the unit element in $\tilde{a}$ is $M_e = I$, and hence $\|M_e\| = \|I\| = 1$. As a consequence, we can always assume that the unit element in a Banach algebra has norm 1 (without change the topology on it).

Example 9.5. $L^1\left(\mathbb{R}^d;\mathbb{F}\right)$, with convolution as multiplication, is a Banach algebra has no unit element. One can adjoin one by the abstract procedure outlined in Remark 9.3 or one can do it more concretely by enlarging $L^1\left(\mathbb{R}^d;\mathbb{F}\right)$ to the algebra of all complex Borel measures μ on $\mathbb{R}^d$ of the form

$$d\mu = f dm + \lambda d\delta\,,$$

where $f \in L^1\left(\mathbb{R}^d;\mathbb{F}\right)$, m is the Lebesgue measure $\mathbb{R}^d$, δ is the Dirac measure on $\mathbb{R}^d$, and $\lambda \in \mathbb{F}$ is a scalar.

Example 9.6. Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ be the one-dimensional tours. We denote by $A(\mathbb{T})$ the space of complex-valued continuous functions on $\mathbb{T}$ having an absolutely convergent Fourier series. That is,

$$A(\mathbb{T}) := \left\{ f \in C(\mathbb{T};\mathbb{C}) : f(t) = \sum_{k\in\mathbb{Z}} \hat{f}(k)e^{2\pi ikt}\,, \sum_{k\in\mathbb{Z}} |\hat{f}(k)| < \infty \right\}.$$

under the usual addition of multiplication. The mapping $f \mapsto \hat{f} = \{\hat{f}(k)\}_{k\in\mathbb{Z}}$ of $A(\mathbb{T})$ into ℓ^1 is clearly linear and one-to-one. If $\sum |a_k| < \infty$ the series $\sum a_k e^{ikt}$ converges uniformly on $\mathbb{T}$ and, denoting its sum by g, we have $a_k = \hat{g}(k)$. It follows that the mapping above is an algebra isomorphism of $A(\mathbb{T})$ onto ℓ^1. We introduce a norm to $A(\mathbb{T})$ by

$$\|f\|_{A(\mathbb{T})} = \|\hat{f}\|_{\ell^1} = \sum_{-\infty}^{\infty} |\hat{f}(k)|\,.$$

We emphasize that the norm on $A(\mathbb{T})$ is not the supremum norm! With this norm $A(\mathbb{T})$ is a Banach space isometric to ℓ^1. We now claim it is a Banach algebra.

Assume that $f, g \in A(\mathbb{T})$, we are going to show that $fg \in A(\mathbb{T})$ and

$$\|fg\|_{A(\mathbb{T})} \le \|f\|_{A(\mathbb{T})}\|g\|_{A(\mathbb{T})}\,.$$

We have $f(t) = \sum \hat{f}(k)e^{ikt}, \quad g(t) = \sum \hat{g}(k)e^{ikt}$ and since both series converge absolutely:

$$f(t)g(t) = \sum_k \sum_m \hat{f}(k)\hat{g}(m)e^{i(k+m)t}$$

Collecting the terms for which $k+m=l$ we obtain

$$f(t)g(t)=\sum_l\sum_k \hat{f}(k)\hat{g}(l-k)e^{ilt}$$

so that $\widehat{fg}(l)=\sum_k \hat{f}(k)\hat{g}(l-k)$; hence

$$\sum_l |\widehat{fg}(l)| \le \sum_l\sum_k |\hat{f}(k)||\hat{g}(l-k)| = \sum_k |\hat{f}(k)| \sum_l |\hat{g}(l)| \, .$$

Example 9.7. Let $\mathbb{D}$ be the unit open disc in the complex plane. Define the disc algebra $A(\mathbb{D})$ by

$$A(\mathbb{D}) \coloneqq \left\{ f \in C(\overline{\mathbb{D}}) : f \in H(\mathbb{D}) \right\}$$

equipped with the supremum norm, under the usual addition and multiplication. It's easy to see that $A(\mathbb{D})$ is a subalgebra of $C(\overline{\mathbb{D}};\mathbb{C})$, and one can easily verify that $A(\mathbb{D})$ is closed in $C(\overline{\mathbb{D}};\mathbb{C})$, thus $A(\mathbb{D})$ is a unital commutative Banach algebra.

9.2 Basic Properties of Spectra

Let $\mathscr{A}$ be a Banach algebra over $\mathbb{F}$ with identity e. Let $G(\mathscr{A})$ be the set of all invertible elements of $\mathscr{A}$. It's easy to verify that $(G(\mathscr{A}),\cdot)$ is a group, where $\cdot$ is the multiplication inherited form $\mathscr{A}$.

The following lemma is easy but very useful, and the proof is omitted.

Lemma 9.1. *Let $\mathscr{A}$ be a Banach algebra over $\mathbb{F}$ with identity e. Then for each $\|x\|<1$, $e-x$ is invertible, and*

$$(e-x)^{-1}=\sum_{n=0}^{\infty} x^n \, .$$

Proposition 9.2. *Let $\mathscr{A}$ be a Banach algebra over $\mathbb{F}$ with identity e. Then* *(a)* *For fixed $x \in G(\mathscr{A})$, if $\|h\| < \frac{1}{\|x^{-1}\|}$ then $x - h \subset G(\mathscr{A})$.*

(b) *$G(\mathscr{A})$ is an open subset of $\mathscr{A}$, and the mapping $x \mapsto x^{-1}$ is a homeomorphism of $G(\mathscr{A})$ onto $G(\mathscr{A})$.*

Proof. For each $x \in G(\mathscr{A})$ and $h \in \mathscr{A}$, we have

$$x - h = x(e - x^{-1}h)\,.$$

Then use the last lemma, we can see that $B(x, \frac{1}{\|x^{-1}\|}) \subset G(\mathscr{A})$. Thus $G(\mathscr{A})$ is open. Moreover,

$$\begin{aligned}
&\|(x-h)^{-1} - x^{-1}\| = \|(e - x^{-1}h)^{-1}x^{-1} - x^{-1}\| \\
&\leq \left\|\sum_{n=0}^{\infty}(x^{-1}h)^n - e\right\| \, \|x^{-1}\| \\
&\leq \sum_{n=1}^{\infty} \|x^{-1}h\|^n \, \|x^{-1}\| = \frac{\|x^{-1}h\|}{1 - \|x^{-1}h\|} \, \|x^{-1}\|\,.
\end{aligned}$$

Thus for each $x \in G(\mathscr{A})$,

$$\lim_{h \to 0} \|(x-h)^{-1} - x^{-1}\| = 0$$

Then clearly $x \mapsto x^{-1}$ is continuous. Since $x \to x^{-1}$ maps $G(\mathscr{A})$ onto $G(\mathscr{A})$ and since it is its own inverse, it is a homeomorphism. $\square$

Let $\mathscr{A}$ be a Banach algebra over $\mathbb{F}$ with identity e. The *spectrum* of $a \in \mathscr{A}$, denoted by $\sigma(a)$, is defined by

$$\sigma(a) = \{\lambda \in \mathbb{F} : \lambda e - a \text{ is not invertible}\}\,.$$

The *resolvent set* of a, denoted by $\varrho(a)$, is defined by

$$\varrho(a) = \{\lambda \in \mathbb{F} : \lambda e - a \text{ is invertible}\} = \mathbb{F}\backslash\sigma(a)\,.$$

The following proposition is similar to Theorem 6.6 and the proofs are basically the same.

Proposition 9.3. *Let $\mathscr{A}$ be a Banach algebra over $\mathbb{F}$ and $x \in \mathscr{A}$. Then the following limit exists:*

$$\lim_{n\to\infty} \|x^n\|^{\frac{1}{n}} = \inf_{n\geq 1} \|x^n\|^{\frac{1}{n}} =: r_\sigma(x)\,,$$

called the spectral radius of x. For each $|\lambda| > r_\sigma(x)$ (for example $\lambda > \|x\|$), $\lambda e - x$ is invertible and its converse is given by the series

$$(\lambda e - x)^{-1} = \sum_{n=1}^{\infty} \frac{1}{\lambda^n} x^{n-1}\,.$$

Example 9.8. Let X be a compact Hausdorff space. For each $f \in C(X;\mathbb{F})$, we have

$$\sigma(f) = f(X) = \{f(x) : x \in X\}\,.$$

In fact, if $\lambda = f(x_0)$, then $\lambda - f$ has a zero and cannot be invertible. So $f(X) \subset \sigma(f)$. On the other hand, if $\lambda \notin f(X), \lambda - f$ is a non-vanishing continuous function on X. Hence $\frac{1}{\lambda - f} \in C(X;\mathbb{F})$ and so $\lambda - f$ is invertible. Thus $\lambda \in \varrho(f)$.

Example 9.9. If X is a Banach space over $\mathbb{F}$, and $T \in \mathcal{B}(X)$, then

$$\begin{aligned}\rho(T) &= \{\lambda \in \mathbb{F} : \lambda I - T \text{ is bijective}\}\,;\\ \sigma(T) &= \{\lambda \in \mathbb{F} : \text{ either } N(\lambda I - T) \neq \{0\} \text{ or } R(T-\lambda) \neq X\}\,.\end{aligned}$$

In fact, if $\lambda \in \rho(T)$, clearly, $\lambda I - T$ is bijective. On the other hand, if $\lambda I - T$ is bijective, $(\lambda I - T)^{-1} \in \mathcal{B}(X)$ by the inverse mapping theorem.

Example 9.10. If $\mathscr{A} = M_2(\mathbb{R})$ and $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, then

$$\sigma(A) = \emptyset$$

In fact, $\lambda I - A$ is not invertible if and only if $0 = \det(\lambda I - A) = \lambda^2 + 1$, which is impossible in $\mathbb{R}$.

There are several reasons for restricting our attention to Banach algebras over the complex field $\mathbb{C}$.

- The first reason is that, the phenomenon of the last example does not occur if $\mathscr{A}$ is a Banach algebra over $\mathbb{C}$.
- Another reason is that certain elementary facts about holomorphic functions play an important role in the foundations of the subject. This will be observed in the next section when we discussing the spectrum, and becomes even more obvious in the Holomorphic functional calculus.
- The last reason - one whose implications are not quite so obvious - is that $\mathbb{C}$ has a natural nontrivial *involution* (see Definition 9.6), namely, *conjugation*, and that many of the deeper properties of certain types of Banach algebras depend on the presence of an involution. For the same reason, the theory of complex Hilbert spaces is richer than that of real ones.

By the same argument of Theorem 6.8, we can get the following result.

Theorem 9.4. *Let $\mathscr{A}$ be a complex unital Banach algebra. Let $x \in \mathscr{A}$. Then the spectrum $\sigma(x)$ is a non-empty compact subset of $\mathbb{C}$. In fact,*

$$\sup_{\lambda \in \sigma(x)} |\lambda| = r_\sigma(x) \,,$$

and this is why we call $r_\sigma(x)$ the spectral radius of x.

Thus form now on, without special announcement, we always suppose that all Banach algebras are over the complex field $\mathbb{C}$.

Remark 9.4. Whether an element of $\mathscr{A}$ is or is not invertible in $\mathscr{A}$ is a purely *algebraic property*. The spectrum and the spectral radius of an $x \in \mathscr{A}$ are thus defined in terms of the *algebraic structure* of $\mathscr{A}$, regardless of any metric (or topological) considerations. On the other hand, $\lim \|x^n\|^{\frac{1}{n}}$

depends obviously on the *norm* of $\mathscr{A}$. This is one of the remarkable features of the spectral radius formula: It asserts the equality of certain quantities which arise in entirely different ways.

Remark 9.5. Our algebra $\mathscr{A}$ may have a subalgebra $\mathscr{B}$ containing the unit element, and it may then very well happen that some $x \in \mathscr{B}$ is not invertible in $\mathscr{B}$ but is invertible in $\mathscr{A}$. The spectrum of x depends on the algebra, and the inclusion $\sigma_{\mathscr{A}}(x) \subset \sigma_{\mathscr{B}}(x)$ holds (the notation is self-explanatory). The two spectra can be different (see Example 9.12). The spectral radius is, however, unaffected by the passage from $\mathscr{A}$ to $\mathscr{B}$, since the spectral radius formula expresses it in terms of metric properties of powers of x, and these are independent of anything that happens outside $\mathscr{A}$. Later we will describe the relation between $\sigma_{\mathscr{A}}(x)$ and $\sigma_{\mathscr{B}}(x)$ in greater detail.

Example 9.11. Let $\mathscr{A} = M_2(\mathbb{C})$ and let $A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$. Then $A^2 = 0$ and $\sigma(A) = \{0\}$; so $r_\sigma(A) = 0$. So it is possible to have $r(A) = 0$ with $A \neq 0$.

The non-emptiness of spectrum leads to an easy characterization of those Banach algebras that are division algebras.

Theorem 9.5 (Gelfand-Mazur)**.** *If $\mathscr{A}$ is a complex unital Banach algebra in which every nonzero element is invertible, then $\mathscr{A}$ is (isometrically isomorphic to) the complex field $\mathbb{C}$.*

Proof. If $x \in \mathscr{A}$ and $\lambda_1 \neq \lambda_2$, then at most one of the elements $\lambda_1 e - x$ and $\lambda_2 e - x$ is 0; hence at least one of them is invertible. Since $\sigma(x)$ is not empty, it follows that $\sigma(x)$ consists of exactly one point, say $\lambda(x)$, for each $x \in \mathscr{A}$. since $\lambda(x)e - x$ is not invertible, it is 0. Hence $x = \lambda(x)e$.

The mapping $x \to \lambda(x)$ is therefore an isomorphism of $\mathscr{A}$ onto $\mathbb{C}$, which is also an isometry, since $|\lambda(x)| = \|\lambda(x)e\| = \|x\|$ for every $x \in \mathscr{A}$. $\square$

It is natural to ask whether the spectra of two elements x and y of $\mathscr{A}$

are close together, in some suitably defined sense, if x and y are close to each other. The next proposition gives a very simple answer.

Proposition 9.6 (Perturbation of Spectrum). *Let $\mathscr{A}$ be a complex unital Banach algebra, $x \in \mathscr{A}$, Ω is an open set in $\mathbb{C}$ containing $\sigma(x)$. Then there exists $\delta > 0$ so that for ever $y \in \mathscr{A}$ with $\|y\| < \delta$, we have $\sigma(x+y) \subset \Omega$.*

Proof. Since $\left\|(\lambda e - x)^{-1}\right\|$ is a continuous function of λ in the complement of $\sigma(x)$, and since this norm tends to 0 as $\lambda \to \infty$, there is a number $M < \infty$ such that

$$\left\|(\lambda e - x)^{-1}\right\| < M \text{ for all } \lambda \in \Omega^c .$$

If $y \in \mathscr{A}, \|y\| < 1/M$, and $\lambda \in \Omega^c$, it follows that

$$\lambda e - (x+y) = (\lambda e - x)\left[e - (\lambda e - x)^{-1} y\right]$$

is invertible in $\mathscr{A}$, since $\left\|(\lambda e - x)^{-1} y\right\| < 1$; hence $\lambda \notin \sigma(x+y)$. This gives the desired conclusion, with $\delta = 1/M$. $\square$

Exercise 9.1. Let $\mathscr{A}$ be a complex Banach algebra with identity e. Let x, y in $\mathscr{A}$. Then the following statements hold.

(a) $e - xy$ is invertible if and only if $e - yx$ is.

(b) $\sigma(xy)\backslash\{0\} = \sigma(yx)\backslash\{0\}$.

(c) $r_\sigma(xy) = r_\sigma(yx)$.

(d) If $xy = yx$, then $r_\sigma(x+y) \le r_\sigma(x) + r_\sigma(y)$ and $r_\sigma(xy) \le r_\sigma(x) r_\sigma(y)$.

9.2.1 Dependence of the Spectrum on the Algebra

We begin with an example.

Example 9.12. Let $\partial\mathbb{D} = \{z \in \mathbb{C} : |z| = 1\}$ and $\mathscr{A} = C(\partial\mathbb{D}; \mathbb{C})$. In Example 9.8, we have computed the spectrum of (the identity mapping) z as an element of $\mathscr{A}$:

$$\sigma_{\mathscr{A}}(z) = \partial\mathbb{D} .$$

Let $\mathscr{B}$ = the uniform closure of the polynomials in $\mathscr{A}$. (Here "polynomial" means a polynomial in z.) Now $z \in \mathscr{B}$ and so it has a spectrum as an element of this algebra; denoted by $\sigma_{\mathscr{B}}(z)$. As we have pointed, $\sigma_{\mathscr{A}}(z) \subset \sigma_{\mathscr{B}}(z)$. But there is no reason to believe that $\sigma_{\mathscr{B}}(z) = \sigma_{\mathscr{A}}(z)$. In fact, they are not equal:

$$\sigma_{\mathscr{B}}(z) = \overline{\mathbb{D}}\,.$$

To see this first note that $\|z\| = 1$, so that $\sigma_{\mathscr{B}}(z) \subset \overline{\mathbb{D}}$.

If $|\lambda| \leq 1$ and $\lambda \notin \sigma_{\mathscr{B}}(z)$, there is an f in $\mathscr{B}$ such that $(\lambda - z)f(z) = 1$ for $z \in \partial\mathbb{D}$. Note that this implies that $|\lambda| < 1$. Because $f \in \mathscr{B}$, there is a sequence of polynomials $\{p_n\}$ such that $p_n \to f$ uniformly on $\partial\mathbb{D}$. Thus for every $\epsilon > 0$ there is a $N = N_\epsilon$ such that for $m, n \geq N$,

$$\sup\{|p_n(z) - p_n(z)| : z \in \partial\mathbb{D}\} \leq \epsilon\,.$$

By the maximum principle,

$$\sup\{|p_n(z) - p_n(z)| : z \in \mathbb{D}\} \leq \epsilon\,.$$

Thus $g(z) = \lim p_n(z)$ is analytic on $\mathbb{D}$ and continuous on $\overline{\mathbb{D}}$; also, $g|_{\partial\mathbb{D}} = f$. By the same argument, since $p_n(z)(\lambda - z) \to 1$ uniformly on $\partial\mathbb{D}$, $p_n(z)(\lambda - z) \to 1$ uniformly on $\mathbb{D}$. Thus $g(z)(\lambda - z) = 1$ on $\mathbb{D}$. But $g(\lambda)(\lambda - \lambda) = 0$, a contradiction. Thus, $\overline{\mathbb{D}} \subset \sigma_{\mathscr{B}}(z)$.

Thus the spectrum not only depends on the element of the algebra, but also on the algebra. Precisely how this dependence occurs is given below. We begin with the following lemma.

Lemma 9.7. *Suppose V and W are open sets in some topological space X, $V \subset W$, and W contains no boundary point of V. Then V is a union of maximal connected components of W.*

Proof. Let Ω be a maximal connected component of W that intersects V. Since W contains no boundary point of V, Ω is the union of the two disjoint open sets $\Omega \cap V$ and $\Omega \backslash \overline{V}^c$. Since Ω is connected, $\Omega \backslash \overline{V}^c$ is empty. Thus $\Omega \subset V$. $\square$

Lemma 9.8. *Let $\mathscr{A}$ be a complex unital Banach algebra. Let x be a boundary point of $G(\mathscr{A})$ with $x_n \in G(\mathscr{A})$ for $n \geq 1$ so that $x_n \to x$ as $n \to \infty$. Then $\left\|x_n^{-1}\right\| \to \infty$ as $n \to \infty$.*

Proof. If the conclusion is false, there exists $M < \infty$ such that $\left\|x_n^{-1}\right\| < M$ for infinitely many n. For one of these, $\|x_n - x\| < 1/M$. For this n

$$\left\|e - x_n^{-1}x\right\| = \left\|x_n^{-1}\left(x_n - x\right)\right\| < 1$$

so that $x_n^{-1}x \in G(\mathscr{A})$. Since $x = x_n\left(x_n^{-1}x\right)$ and $G(\mathscr{A})$ is a group, it follows that $x \in G(\mathscr{A})$. This contradicts the hypothesis, since $G(\mathscr{A})$ is open. □

Let $\mathscr{A}$ be a unital Banach algebra and $\mathscr{B}$ a closed subalgebra in $\mathscr{A}$ contains the unit element of $\mathscr{A}$. Then $G(\mathscr{B})$ and $\mathscr{B} \cap G(\mathscr{A})$ are two open sets in $\mathscr{B}$. By the previous lemmas, it's easy to see that $G(\mathscr{B})$ contains no boundary point of $\mathscr{B} \cap G(\mathscr{A})$ and hence $G(\mathscr{B})$ is a union of maximal connected components of $\mathscr{B} \cap G(\mathscr{A})$. Moreover, we have :

Theorem 9.9. *Let $\mathscr{A}$ be a complex unital Banach algebra and $\mathscr{B}$ a closed subalgebra in $\mathscr{A}$ contains the unit element of $\mathscr{A}$. Let x be an element in $\mathscr{B}$. Then the following statements hold.*

(a) $\partial\varrho_{\mathscr{B}}(x) = \partial\sigma_{\mathscr{B}}(x) \subset \sigma_{\mathscr{A}}(x)$.

(b) $\varrho_{\mathscr{B}}(x)$ *is a union of maximal connected components of* $\varrho_{\mathscr{A}}(x)$.

(c) $\sigma_{\mathscr{B}}(x)$ *is the union of* $\sigma_{\mathscr{A}}(x)$ *and a (possibly empty) collection of bounded maximal connected components of* $\varrho_{\mathscr{A}}(x)$.

Proof. Note that by Lemma 9.8, $\varrho_{\mathscr{B}}(x)$ contains no boundary point of $\varrho_{\mathscr{A}}(x)$. Thus $\partial\varrho_{\mathscr{B}}(x) = \partial\sigma_{\mathscr{B}}(x) \subset \mathbb{C}\backslash\sigma_{\mathscr{A}}(x) = \varrho_{\mathscr{A}}(x)$. Then by Lemma 9.7, part (b) and part (c) follows. □

Remark 9.6. Intuitively, if $\sigma_{\mathscr{B}}(x)$ is larger than $\sigma_{\mathscr{A}}(x)$, then $\sigma_{\mathscr{B}}(x)$ is obtained from $\sigma_{\mathscr{A}}(x)$ by "*filling in some holes*" in $\sigma_{\mathscr{A}}(x)$. In particular, if $\varrho_{\mathscr{A}}(x)$ is connected (for example $\sigma_{\mathscr{A}}(x)$ contains only real numbers) or if $\sigma_{\mathscr{B}}(x)$ has no interior point, then we have $\sigma_{\mathscr{A}}(x) = \sigma_{\mathscr{B}}(x)$.

As another application of Lemma 9.8, we now give an exercise whose conclusion is the same as that of the Gelfand-Mazur theorem, although its consequences are not nearly so important.

Exercise 9.2. If $\mathscr{A}$ is a complex unital Banach algebra and if there exists $M < \infty$ such that

$$\|x\|\|y\| \leq M\|xy\| \quad (x \in \mathscr{A}, y \in \mathscr{A})$$

then $\mathscr{A}$ is isometrically isomorphic to $\mathbb{C}$. (Hint: Show that $\partial G(\mathscr{A}) = \{0\}$, as a consequence, for each $x \in \mathscr{A}$, $\sigma(x)$ consists of a single point.)

9.2.2 Complex Homomorphisms

Among the important mappings from one Banach algebra into another are the *homomorphisms*.

Definition 9.2. Suppose $\mathscr{A}$, $\mathscr{B}$ be two complex algebra and $\psi : \mathscr{A} \to \mathscr{B}$ is a linear mapping. If

$$\psi(xy) = \psi(x)\psi(y)$$

for all $x, y \in \mathscr{A}$, then ψ is called a *homomorphism*. If in addition ψ is a bijection, then we say it is an *isomorphism*. We say a homomorphism ψ is *nonzero* or *nontrivial*, if ψ is not identically zero.

Of particular interest is the case in which the range is the simplest of all Banach algebras, namely, $\mathbb{C}$ itself. Many of the significant features of the commutative theory depend crucially on a sufficient supply of homomorphisms onto $\mathbb{C}$.

A homomorphism ϕ taking values in $\mathbb{C}$ is called a *complex homomorphism.* For every complex algebra $\mathscr{A}$, trivially, $\phi \equiv 0$ is of course a complex homomorphism on $\mathscr{A}$.

Proposition 9.10. *If ϕ is a nonzero complex homomorphism on a complex algebra $\mathscr{A}$ with unit e, then $\phi(e) = 1$, and $\phi(x) \neq 0$ for every invertible $x \in \mathscr{A}$.*

Proof. For some $y \in \mathscr{A}, \phi(y) \neq 0$. Since

$$\phi(y) = \phi(ye) = \phi(y)\phi(e)$$

it follows that $\phi(e) = 1$. If x is invertible, then

$$\phi(x)\phi\left(x^{-1}\right) = \phi\left(xx^{-1}\right) = \phi(e) = 1$$

so that $\phi(x) \neq 0$. □

Theorem 9.11. *Let $\mathscr{A}$ be a complex unital Banach algebra. Then every complex homomorphism on $\mathscr{A}$ is a contraction; that is,*

$$|\phi(x)| \leq \|x\| \quad \textit{for all } x \in \mathscr{A}\,.$$

Proof. Without loss of generality, let $\phi(e) = 1$ where e is the unit element in $\mathscr{A}$. It suffices to show that $|\phi(x)| \leq 1$ for $\|x\| = 1$. In other words, for any $\lambda \in \mathbb{C}$ with $|\lambda| > 1$,

$$\lambda - \phi(x) = \phi(\lambda e - x) \neq 0\,.$$

Observe that since $\|x\| = 1$, $\lambda e - x = \lambda(e - \frac{x}{\lambda})$ is invertible by Lemma 9.1, and then by Proposition 9.10 we have $\phi(\lambda e - x) = \neq 0$ as desired. □

We now interrupt the main line of development and insert a theorem which shows, for Banach algebras, that Theorem 9.20 actually characterizes the complex homomorphisms among the linear functionals. This striking result has apparently found no interesting applications as yet.

Theorem (Gleason, Kahane, Zelazko). *If ϕ is a linear functional on the complex unital Banach algebra $\mathscr{A}$, such that $\phi(e) = 1$ and $\phi(x) \neq 0$ for every invertible $x \in \mathscr{A}$, then ϕ is a complex homomorphism:*

$$\phi(xy) = \phi(x)\phi(y) \quad \textit{for all } x, y \in \mathscr{A}\,.$$

The proof is not easy so we omit it, whereas it can be found in 10.9 Theorem, *Functional Analysis* by W.Rudin.

9.3 Holomorphic Functional Calculus

9.3.1 Introduction

Let $\mathscr{A}$ be a complex Banach algebra with an unit element e. If $x \in \mathscr{A}$ and if $f(\lambda) = \alpha_0 + \alpha_1\lambda + \cdots + \alpha_n\lambda^n$ is a polynomial with complex coefficients α_i, there can be no doubt about the meaning of the symbol $f(x)$; it obviously denotes the element of $\mathscr{A}$ defined by

$$f(x) = \alpha_0 e + \alpha_1 x + \cdots + \alpha_n x^n\,.$$

The defines a mapping from the algebra of polynomials into the algebra $\mathscr{A}$, that is, clearly, a homomorphism. This homomorphism can be extended to a larger class of functions than polynomials; for instance, we can define

$$\exp\{x\} := \sum_{n=0}^{\infty} \frac{1}{n!}\, x^n\,.$$

More generally, if $f(\lambda) = \sum_{n=0}^{\infty} \alpha_n\lambda^n$ is any *entire function* in $\mathbb{C}$, it is natural to define $f(x) \in \mathscr{A}$ by

$$f(x) = \sum_{n=0}^{\infty} \alpha_n x^n\,,$$

this series always (absolutely) converges. Another example is given by the meromorphic functions

$$f(\lambda) = \frac{1}{\alpha - \lambda}\,,$$

In this case, the natural definition of $f(x)$ is

$$f(x) = (\alpha e - x)^{-1}$$

which makes sense for all x whose spectrum does not contain α.

One is thus led to the conjecture that $f(x)$ should be definable, within $\mathscr{A}$, whenever f is holomorphic in an open set that contains $\sigma(x)$. This turns out to be correct and can be accomplished by a version of the Cauchy formula that converts complex functions defined in open subsets of $\mathbb{C}$ to $\mathscr{A}$-valued ones defined in certain open subsets of $\mathscr{A}$. (Just as in classical analysis, the Cauchy formula is a much more adaptable tool than the power series representation.)

In certain algebras one can go further. For instance, if $\mathscr{A} = \mathcal{B}(H)$ and $\mathscr{A}$ is a bounded normal operator on a complex Hilbert space H, the symbol $f(\mathscr{A})$ can be interpreted as a bounded normal operator on H when f is any continuous complex function on $\sigma(\mathscr{A})$, and even when f is any complex bounded Borel function on $\sigma(\mathscr{A})$. Later we shall see how this leads to an efficient proof of a very general form of the spectral theorem.

9.3.2 The Riesz-Dunford Integral

Let K be a compact subset of an open $\Omega \subset \mathbb{C}$. It is well known that exists a cycle (see Section 4.6) $\gamma = (\gamma_1, \cdots, \gamma_n)$ in $\Omega \backslash K$ so that

$$\operatorname{Ind}_\gamma(z) = \frac{1}{2\pi i} \int_\gamma \frac{d\zeta}{\zeta - z} = \begin{cases} 1\,, & \text{if } z \in K\,; \\ 0\,, & \text{if } z \notin \Omega\,. \end{cases}$$

In other words, γ winds once around every point in K but winds zero time around any point of the complement of Ω. We shall describe this situation briefly by saying that *the cycle γ surrounds K in Ω*. As we know in this case, the Cauchy formula

$$f(z) = \frac{1}{2\pi i} \int_\gamma (\zeta - z)^{-1} f(\zeta) d\zeta$$

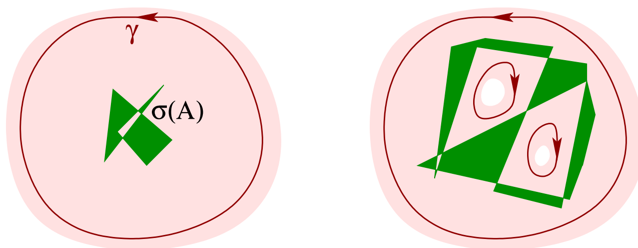

Figure 9.1: A cycle surrounding the spectrum

then holds for every holomorphic function f in Ω and for every $z \in K$.

Note that neither K nor Ω nor the union of the intervals γ_i has been assumed to be connected.

Definition 9.3. Let $\mathscr{A}$ be a complex unital Banach algebra and $x \in \mathscr{A}$. Let Ω be an open set in $\mathbb{C}$ containing $\sigma(x)$. For every holomorphic function f on Ω, i.e., $f \in H(\Omega)$, we define

$$\tilde{f}(x) := \frac{1}{2\pi i} \int_{\gamma} f(\zeta)(\zeta e - x)^{-1} d\zeta \, .$$

where γ is a cycle surrounding $\sigma(x)$ in Ω.

Remark 9.7. The existence of the integral is provided by Theorem 3.49. The vector $f(x)$ is independent of the choice of the cycle γ in $\Omega \backslash \sigma(x)$ surrounding $\sigma(x)$ in Ω. In fact, let γ and $\tilde{\gamma}$ be two cycles in $\Omega \backslash \sigma(x)$ that surrounding $\sigma(x)$ in Ω. Then since $\zeta \mapsto f(\zeta)(\zeta e - x)^{-1}$; $\Omega \backslash \sigma(x) \to \mathscr{A}$ is holomorphic, by Theorem 4.22, we get

$$\int_{\gamma} f(\zeta)(\zeta e - x)^{-1} d\zeta = \int_{\tilde{\gamma}} f(\zeta)(\zeta e - x)^{-1} d\zeta \, .$$

Lemma 9.12. *Let $\mathscr{A}$ be a complex unital Banach algebra and $x \in \mathscr{A}$. Then for any cycle γ surrounding $\sigma(x)$ in $\mathbb{C}$ and for any $n \geq 0$, we have*

$$\frac{1}{2\pi i} \int_{\gamma} \zeta^n (\zeta e - x)^{-1} d\zeta = x^n \, .$$

Proof. By Remark 9.7, without loss of generality, we can suppose γ is the positive oriented circle with radius strictly lager than $\|x\|$. Then, for each $n \geq 0$,

$$\begin{aligned}
&\frac{1}{2\pi i}\int_\gamma \zeta^n(\zeta e - x)^{-1} d\zeta = \frac{1}{2\pi i}\int_\gamma \zeta^n \sum_{k=0}^{\infty} \frac{1}{\zeta^{k+1}}\, x^k d\zeta \\
&= \sum_{k=0}^{\infty} \frac{1}{2\pi i}\int_\gamma \zeta^n \frac{1}{\zeta^{k+1}}\, x^k d\zeta = \sum_{k=0}^{\infty} \frac{1}{2\pi i}\int_\gamma \frac{1}{\zeta^{k+1-n}}\, d\zeta\ x^k = x^n\,,
\end{aligned}$$

as desired. $\square$

Definition 9.4. Suppose $\mathscr{A}$ is a Banach algebra, Ω is an open set in $\mathbb{C}$, and $H(\Omega)$ is the algebra of all complex holomorphic functions in Ω. By Proposition 9.6,

$$\mathscr{A}_\Omega := \{x \in \mathscr{A} : \sigma(x) \subset \Omega\}$$

is an open subset of $\mathscr{A}$.

We define $\tilde{H}\left(\mathscr{A}_\Omega\right)$ to be the set of all $\mathscr{A}$-valued functions $\tilde{f}$, with domain $\mathscr{A}_\Omega$, that arise from an $f \in H(\Omega)$ by the formula

$$\tilde{f}(x) = \frac{1}{2\pi i}\int_\gamma f(\zeta)(\zeta e - x)^{-1} d\zeta\,, \tag{9.2}$$

where γ is any cycle that surrounds $\sigma(x)$ in Ω.

This definition calls for some comments.

Remark 9.8. If $x = \lambda e$ and $\lambda \in \Omega$, (9.2) becomes

$$\tilde{f}(\lambda e) = f(\lambda) e\,. \tag{9.3}$$

Note that $\sigma(\lambda e) = \lambda$, so $\lambda e \in \mathscr{A}_\Omega$ if and only if $\lambda \in \Omega$. If we identity $\lambda \in \mathbb{C}$ with $\lambda e \in \mathscr{A}$, every $f \in H(\Omega)$ may be regarded as mapping a certain subset of $\mathscr{A}_\Omega$ (namely, the intersection of $\mathscr{A}_\Omega$ with the one-dimensional subspace of $\mathscr{A}$ generated by e) into $\mathscr{A}$, and then (9.3) shows that $\tilde{f}$ may be regarded as an extension of f. In most treatments of this topic, $f(x)$ is written in place of our $\tilde{f}(x)$. The notation $\tilde{f}$ is used here because it avoids certain ambiguities that might cause misunderstandings.

Remark 9.9. If S is any set and $\mathscr{A}$ is any algebra, the collection of all $\mathscr{A}$-valued functions on S is an algebra, if scalar multiplication, addition, and multiplication are defined pointwise. For instance, if u and v map S into $\mathscr{A}$, then

$$(uv)(s) = u(s)v(s) \quad (s \in S)\,.$$

This will be applied to $\mathscr{A}$-valued functions defined in $\mathscr{A}_\Omega$.

Theorem 9.13. *Suppose $\mathscr{A}$, $H(\Omega)$, and $\tilde{H}(\mathscr{A}_\Omega)$ are as in Definition 9.4. Then clearly $H(\Omega)$, $\tilde{H}(\mathscr{A}_\Omega)$ both are complex algebras.*

- *(a)* *The mapping $f \to \tilde{f}$ is an isomorphism of $H(\Omega)$ onto $\tilde{H}(\mathscr{A}_\Omega)$.*
- *(b)* *The mapping $f \to \tilde{f}$ is continuous in the following sense: if $f_n \in H(\Omega)$ and $f_n \to f$ uniformly on compact subsets of Ω, then*

$$\tilde{f}(x) = \lim_{n\to\infty} \tilde{f}_n(x) \quad \textit{for} \quad x \in \mathscr{A}_\Omega\,.$$

Remark 9.10. Since $H(\Omega)$ is obviously a commutative algebra, Theorem 9.13 implies that $\tilde{H}(\mathscr{A}_\Omega)$ is also commutative. This may be surprising, because $\tilde{f}(x)$ and $\tilde{f}(y)$ need not commute. However, $\tilde{f}(x)$ and $\tilde{g}(x)$ do commute in $\mathscr{A}$ for every $x \in \mathscr{A}_\Omega$.

Proof. Take $x \in \mathscr{A}_\Omega$. The assertion that $\widetilde{f+g}(x) = \tilde{f}(x) + \tilde{g}(x)$ follows directly from the definition.

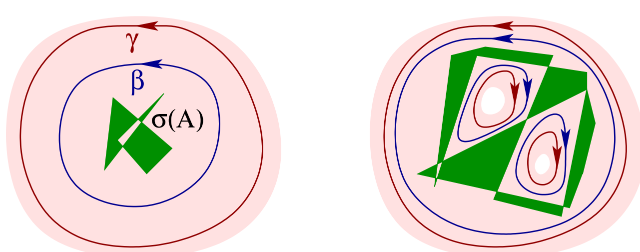

Figure 9.2: Two cycles encircling the spectrum

To prove $\widetilde{fg}(x) = \tilde{f}(x)\tilde{g}(x)$, choose two cycles β and γ in $\Omega\backslash\sigma(x)$ that both surround $\sigma(x)$ in Ω have disjoint images so that

$$\mathrm{im}(\beta) \cap \mathrm{im}(\gamma) = \emptyset$$

and such that the image of β is encircled by γ, i.e.

$$\begin{aligned} \mathrm{Ind}_\gamma(w) = 1 &\quad \text{for all } w \in \mathrm{im}(\beta) \\ \mathrm{Ind}_\gamma(z) = 0 &\quad \text{for all } z \in \mathrm{im}(\gamma) \end{aligned}$$

Then, by the resolvent identity,

$$\begin{aligned} f(x)g(x) =& \frac{1}{2\pi \mathrm{i}} \int_\beta f(w)(we-x)^{-1} dw \ \frac{1}{2\pi \mathrm{i}} \int_\gamma g(z)(ze-x)^{-1} dz \\ =& \frac{1}{2\pi \mathrm{i}} \frac{1}{2\pi \mathrm{i}} \int_\beta \int_\gamma f(w)g(z) \frac{(we-x)^{-1} - (ze-x)^{-1}}{z-w} dz dw \\ =& \frac{1}{2\pi \mathrm{i}} \int_\beta f(w) \left(\frac{1}{2\pi \mathrm{i}} \int_\gamma \frac{g(z)dz}{z-w} \right) (we-x)^{-1} dw \\ &+ \frac{1}{2\pi \mathrm{i}} \int_\gamma g(z) \left(\frac{1}{2\pi \mathrm{i}} \int_\beta \frac{f(w)dw}{w-z} \right) (ze-x)^{-1} dz \\ =& \frac{1}{2\pi \mathrm{i}} \int_\beta f(w)g(w)(we-x)^{-1} dw = (fg)(x) \,. \end{aligned}$$

Therefore, $f \mapsto \tilde{f}$ is a homomorphism.

If $\tilde{f} = 0$, then $f(\lambda)e = \tilde{f}(\lambda e) = 0$ for all $\lambda \in \Omega$ so that $f = 0$. Thus $f \to \tilde{f}$ is one-to-one. By the definition on $\tilde{H}(\mathscr{A}_\Omega)$, $f \to \tilde{f}$ is a surjection. Thus $f \to \tilde{f}$ is an algebra isomorphism of $H(\Omega)$ onto $\tilde{H}(\Omega)$.

The asserted continuity follows directly from the integral (9.2). Since for fixed x, $\left\|(\zeta e - x)^{-1}\right\|$ is bounded on $\zeta \in \gamma$. Use the same cycle γ in $\Omega\backslash\sigma(x)$ for all representation of $\tilde{f}_n$, and apply Proposition 3.50, we get the desired result. $\square$

Theorem 9.14 (Spectral Mapping Theorem)**.** *Suppose $x \in \mathscr{A}_\Omega$ and $f \in H(\Omega)$. Then $\tilde{f}(x)$ is invertible in $\mathscr{A}$ if and only if $f(\lambda) \neq 0$ for every $\lambda \in \sigma(x)$. Moreover,*

$$\sigma(\tilde{f}(x)) = f(\sigma(x)) \,.$$

Proof. If f has no zero on $\sigma(x)$, then $g = 1/f$ is holomorphic in an open set U such that $\sigma(x) \subset U \subset \Omega$. Since $fg = 1$ in U, Theorem 9.13 (with U in place of Ω) shows that $\tilde{f}(x)\tilde{g}(x) = e$, and thus $\tilde{f}(x)$ is invertible.

Conversely, if $f(\lambda) = 0$ for some $\lambda \in \sigma(x)$, then there exists $h \in H(\Omega)$ such that

$$f(z) = (\lambda - z)h(z) \text{ for } z \in \Omega\,;$$

which, by Theorem 9.13, implies that

$$\tilde{f}(x) = (\lambda e - x)\tilde{h}(x) = \tilde{h}(x)(x - \lambda e) \tag{9.4}$$

Since $\lambda e - x$ is not invertible in A, neither is $\tilde{f}(x)$ by (9.4).

Now fix $\lambda \in \mathbb{C}$. By definition, $\lambda \in \sigma(\tilde{f}(x))$ if and only if $\lambda e - \tilde{f}(x)$ is not invertible in $\mathscr{A}$. By the previous argument, applied to $f - \lambda$ is place of f, this happens if and only if $f - \lambda$ has a zero in $\sigma(x)$, that is, if and only if $\lambda \in f(\sigma(x))$. □

The spectral mapping theorem makes it possible to include composition of functions among the operations of the symbolic calculus.

Theorem 9.15. *Suppose $x \in \mathscr{A}_\Omega$, $f \in H(\Omega)$, U is an open set containing $f(\Omega)$, $g \in H(U)$, and $(g \circ f)(\lambda) = g(f(\lambda))$ for $\lambda \in \Omega$. Then*

$$\widetilde{(g \circ f)}(x) = \tilde{g}(\tilde{f}(x))\,.$$

Proof. Note first that the operator $g(f(x))$ is well defined, because $\sigma(f(x)) = f(\sigma(x)) \subset f(\Omega) \subset U$ by the spectrum mapping theorem. Fix a contour Γ_1 that surrounds $f(\sigma(x))$ in U. Then, by definition,

$$\tilde{g}(\tilde{f}(x)) = \frac{1}{2\pi i}\int_{\Gamma_1} g(\zeta)[\zeta e - \tilde{f}(x)]^{-1} d\zeta$$

For fixed $\zeta \in \mathrm{im}(\Gamma_1)$, we want to represent $(\zeta e - f(x))^{-1}$ by Cauchy integral and hence we need the mapping $\lambda \to [\zeta e - \tilde{f}(x)]^{-1}$ is holomorphic on some open neighborhood of $\sigma(x)$. It suffices to request that $f(\lambda) \cap \mathrm{im}(\Gamma_1) =$

$\emptyset$. Note that Γ_1 surrounds $f(\sigma(x))$ in U, we can choose an open set W, with $f(\sigma(x)) \subset W \subset U$, so small that

$$\mathrm{Ind}_{\Gamma_1}(z) = 1 \text{ for } z \in W\,.$$

Thus $W \cap \mathrm{im}(\Gamma_1) = \emptyset$ and $\sigma(x) \subset f^{-1}(W) \subset \Omega$. Now, since $\lambda \to (\zeta - f(\lambda))^{-1}$ is holomorphic on $f^{-1}(W)$, fix a contour Γ_0 that surrounds $\sigma(x)$ in $f^{-1}(W)$, we have

$$[\zeta e - \tilde{f}(x)]^{-1} = \frac{1}{2\pi i}\int_{\Gamma_0} [\zeta - f(z)]^{-1}(ze - x)^{-1} dz \quad (\zeta \in \Gamma_1)$$

Thus

$$\begin{aligned}
\tilde{g}(\tilde{f}(x)) &= \frac{1}{2\pi i}\int_{\Gamma_1} g(\zeta)[\zeta e - \tilde{f}(x)]^{-1} d\zeta \\
&= \frac{1}{2\pi i}\int_{\Gamma_0} \frac{1}{2\pi i}\int_{\Gamma_1} g(\zeta)[\zeta - f(z)]^{-1} d\zeta (ze - x)^{-1} dz \\
&= \frac{1}{2\pi i}\int_{\Gamma_0} g(f(z))(ze - x)^{-1} dz = \widetilde{(g \circ f)}(x)\,,
\end{aligned}$$

as desired. □

9.3.3 Spectral Projection

Now we consider the complex Banach algebra $\mathscr{A} = \mathcal{B}(X)$, the Banach algebra of all bounded linear operators on the complex Banach space X.

Suppose that the spectrum of $A \in \mathcal{B}(X)$ can be decomposed as the union of N pairwise disjoint closed components:

$$\sigma(A) = \Sigma_1 \cup \cdots \cup \Sigma_N, \quad \Sigma_j \cap \Sigma_k = \emptyset \ \text{ if } k \neq j\,.$$

Let $U_j \subset \mathbb{C}$ be disjoint open sets such that $\Sigma_j \subset U_j$ for $j = 0, 1, \cdots, N$. Let $U = \cup_{j=1}^{N} U_j$ and define the function $f_j : U \to \mathbb{C}$ by

$$f|_{U_j} \coloneqq 1 \ \text{ and } \ f|_{U_k} \coloneqq 0 \ \text{ for } k \neq j\,.$$

Trivially, f_j is a holomorphic function on $U \supset \sigma(A)$. Define

$$P_j \coloneqq f_j(A)\,.$$

In other words, if a cycle γ surrounds $\sigma(A)$ in U, there must exists cycles γ_j surrounding Σ_j in U_j so that $\gamma = \gamma_1 + \cdots + \gamma_N$. Then it's easy to see that

$$P_j = \int_{\gamma_j} (\zeta I - A)^{-1} d\zeta \,.$$

Theorem 9.16. *Let $\{P_j\}_{j=1}^N$ be defined above. Then*

(a) The $\{P_j\}_{j=1}^N$ are disjoint projections that is,

$$P_j^2 = P_j \quad \textit{and} \quad P_j P_k = 0 \quad \textit{for} \quad j \neq k \,;$$

and $\sum_{j=1}^N P_j = I$.

(b) For each j, P_j and A commute. Thus $X_j \coloneqq R(P_j)$ are closed A-invariant subspaces of X such that $X = X_1 \oplus \cdots \oplus X_N$. The spectrum of the operator $A_j \coloneqq A|_{X_j} : X_j \to X_j$ is given by $\sigma(A_j) = \Sigma_j$.

Proof. Since $f_j f_k = \delta_{jk} f_j$ it follows from that $P_j P_k = \delta_{jk} P_j$. Moreover P_j commutes with A by definition.

Let $\alpha \in \mathbb{C}$. Consider the operator

$$\tilde{A}_j \coloneqq \begin{bmatrix} \alpha I_{X_1} & & & & \\ & \ddots & & & \\ & & A_j & & \\ & & & \ddots & \\ & & & & \alpha I_{X_N} \end{bmatrix}$$

Define $g : U \to \mathbb{C}$ by $g(z) = z$ for $z \in U$ and let $c \in \mathbb{C}$. Then, we can see that

$$\tilde{A} = (\alpha(1 - f) + gf)(A) \;, \quad \sigma(\tilde{A}) = \{\alpha\} \cup \Sigma_j \,.$$

If $\lambda \in \mathbb{C} \backslash \Sigma_j$, it follows that the operator $\lambda I - \tilde{A}$ is bijective for $\alpha \neq \lambda$ and so $\lambda I_{X_j} - A_j$ is bijective. Conversely, suppose $\lambda \in \Sigma_j$. Then $\lambda I - \tilde{A}$ is not bijective and, for $\alpha \neq \lambda$, this implies that $\lambda I_{X_j} - A_j$ is not bijective. Thus $\sigma(A_j) = \Sigma_j$. $\square$

Example 9.13 (Spectral Projection)**.** Let X be a complex Banach space, let $A \in \mathcal{C}(X)$ be a compact linear operator on X, let $\lambda \in \sigma(A)$ be a nonzero eigenvalue of A, and choose $r \in \mathbb{N}$ such that

$$E_\lambda \coloneqq N(\lambda I - A)^r = N(\lambda I - A)^{r+1} .$$

The proof of Lemma 7.13 and Theorem 7.14 shows that such an integer $r \geq 1$ exists, that E_λ is a finite-dimensional linear subspace of X, that the operator $(\lambda I - A)^r$ has a closed image, and that

$$X = N(\lambda I - A)^r \oplus R(\lambda I - A)^r .$$

Hence the formula

$$P_\lambda (x_0 + x_1) = x_0 \text{ for } x_0 \in N(\lambda I - A)^r \;\text{ and }\; x_1 \in R(\lambda I - A)^r \tag{9.5}$$

defines a bounded linear operator $P_\lambda : X \to X$ which is an A-invariant projection onto E_λ, i.e.

$$P_\lambda^2 = P_\lambda, \quad P_\lambda A = AP_\lambda, \quad R\,(P_\lambda) = E_\lambda .$$

The operator P_λ is uniquely determined by (9.5) and is called the *spectral projection* associated to the eigenvalue λ.

It can also be written in the form

$$P_\lambda = \frac{1}{2\pi i}\int_\gamma (\zeta I - A)^{-1} d\zeta ,$$

where $\gamma(t)$ is the positive oriented circle centering at λ with radius r so that $D(\lambda, r) \cap \sigma(A) = \{\lambda\}$.

9.4 The Gelfand Theory

This chapter deals primarily with the Gelfand theory of commutative Banach algebras, although some of the results of this theory will be applied to noncommutative situations.

The main result of this section is that, for each commutative Banach algebra $\mathscr{A}$, there is a compact Hausdorff space Δ, and a continuous homomorphism $\Gamma : \mathscr{A} \to C(\Delta)$. Indeed, Δ consists of all the nonzero complex homomorphisms on $\mathscr{A}$.

We begin with some preliminary works.

9.4.1 Ideals and Homomorphisms

Definition 9.5. A subset J of a complex algebra $\mathscr{A}$ is said to be an *ideal* if

(a) J is a subspace of $\mathscr{A}$ (in the vector space sense), and

(b) xy and yx in J whenever $x \in \mathscr{A}$ and $y \in J$.

If $J \neq \mathscr{A}$, we say J is a *proper ideal.* If J is a proper ideal which is not contained in any larger proper ideal, we say J is a *maximal ideal.*

Given a ideal of $\mathscr{A}$, we can define the *quotient algebra* $\mathscr{A}/J$ as follows. Clearly $\mathscr{A}/J$ has a natural linear structure. We define the multiplication on $\mathscr{A}/J$ by

$$\tilde{x} \cdot \tilde{y} \coloneqq \widetilde{xy} \qquad \text{for all} \;\; \tilde{x}, \tilde{y} \in \mathscr{A}/J\,.$$

The it's easy to see that $\mathscr{A}/J$ is an algebra.

The following proposition is trivial, so we omit the proof.

Proposition 9.17. *Let $\mathscr{A}$ be a complex algebra with unit e. Let J be a proper ideal of $\mathscr{A}$. Then*

(a) $e \notin J$, and

(b) if a is invertible, then $a \notin J$.

By Zorn's lemma, we have the following assertion.

Proposition 9.18. *Every proper ideal is contained in a maximal ideal.*

Proof. Let J be a proper ideal of $\mathscr{A}$. Let $\mathscr{P}_J$ be the collection of all proper ideals of $\mathscr{A}$ that contain J. Partially order $\mathscr{P}$ by set inclusion, let $\mathscr{I}$ be a totally ordered subset of $\mathscr{P}$, and let M be the union of all members of $\mathscr{I}$. Being the union of a totally ordered collection of ideals, M is an ideal. Obviously $J \subset M$, and $M \neq \mathscr{A}$. Thus $\mathscr{I}$ has an upper bounded. By zorn's lemma, $\mathscr{P}_J$ has a maximal element, as required. $\square$

Proposition 9.19. *Let $\mathscr{A}$ be a complex Banach algebra with unit e. Then the following statements hold.*

(a) Let J be a ideal of $\mathscr{A}$, then so is $\overline{J}$.

(b) If J is a maximal ideal of $\mathscr{A}$, then J is closed.

Proof. We begin with part (a). Clearly $\overline{J}$ is a linear subspace of $\mathscr{A}$, it suffices to show the absorption law. Take $x \in \overline{J}$ and $y \in \mathscr{A}$, then there exists (x_n) in J so that $x_n \to x$. Then $x_n y \to xy$ and $yx_n \to yx$. Since $x_n y$ and yx_n belong to J, thus xy and yx belong to $\overline{J}$. Thus $\overline{J}$ is an ideal.

Now show that part (b). If J is a maximal ideal, to show $J = \overline{J}$, it suffices to show that $\overline{J} \neq \mathscr{A}$. Observe that an ideal I of $\mathscr{A}$ equals $\mathscr{A}$ iff $I \cap G(\mathscr{A}) \neq \emptyset$. So we have only to show that $\overline{J} \cap G(\mathscr{A}) = \emptyset$. However, since J is maximal,

$$J \cap G(\mathscr{A}) = \emptyset\,.$$

Since $G(\mathscr{A})$ is open, we get $\overline{J} \cap G(\mathscr{A}) = \emptyset$, as desired. $\square$

The following result highly depends on the commutativity of $\mathscr{A}$ and the fact that $\mathscr{A}$ has a unit element.

Theorem 9.20. *Let $\mathscr{A}$ be a complex commutative algebra with unit element e. Then the following statement hold.*

(a) An element $x \in \mathscr{A}$ is invertible if and only if x is not contained in any maximal ideal of $\mathscr{A}$.

(b) *A proper ideal J of $\mathscr{A}$ is maximal if and only if $\mathscr{A}/J$ is a division algebra.*

Proof. (a). Clearly if x is invertible, then any proper ideal does not contain x. If x is not invertible, then consider

$$x\mathscr{A}\,,$$

the ideal generated by $\mathscr{A}$. Since x is not invertible, so $x\mathscr{A} \neq \mathscr{A}$. Thus $x\mathscr{A}$ is a proper ideal containing x.

(b). Note that a proper ideal J is maximal iff for each $x \notin J$, the ideal generated by $J \cup \{x\}$ equals $\mathscr{A}$. Since $\mathscr{A}$ is unital and commutative, the ideal generated by $J \cup \{x\}$ is given by

$$x\mathscr{A} + J\,.$$

On the other hand, $x\mathscr{A} + J = \mathscr{A}$ if and only if $\tilde{x}$ is invertible in $\mathscr{A}/J$, and hence the desired result follows. $\square$

Homomorphisms and quotient algebras If $\mathscr{A}$ and $\mathscr{B}$ are Banach algebras and ψ is a nonzero continuous homomorphism of $\mathscr{A}$ into $\mathscr{B}$, then

$$\ker(\psi)$$

the null space or kernel of ψ, is obviously an *closed proper ideal* in $\mathscr{A}$.

Conversely, dose every closed proper ideal J in $\mathscr{A}$ is the kernel of some homomorphism of $\mathscr{A}$? First of all, $\mathscr{A}/J$ is a Banach space, with respect to the quotient norm, and also an algebra. Indeed it's a Banach algebra, since

$$\begin{aligned}\|\tilde{x}\tilde{y}\| = \|\widetilde{xy}\| &= \inf\{\|xy + j\| : j \in J\} \\ &\leq \inf\{\|(x + j_1)(y + j_2)\| : j_1, j_2 \in J\} \\ &\leq \inf\{\|(x + j_1)\|\|(y + j_2)\| : j_1, j_2 \in J\} = \|\tilde{x}\|\|\tilde{y}\|\,.\end{aligned}$$

Then clearly the quotient map $Q : \mathscr{A} \to \mathscr{A}/J$ is a continuous homomorphism so that $\ker(Q) = J$.

Complex homomorphisms and maximal ideals Let $\mathscr{A}$ be a complex Banach algebra with unit e. If ϕ is a nonzero complex homomorphism of $\mathscr{A}$ onto $\mathbb{C}$, then $\ker(\phi)$ must be a maximal ideal of $\mathscr{A}$ since

$$\mathscr{A}/\ker(\phi) \ \text{ ad } \ \mathrm{Im}(\phi) = \mathbb{C} \ \text{ are isomorphism } .$$

We denote by Δ all the nonzero complex homomorphism on $\mathscr{A}$ an by $\mathfrak{M}$ all the maximal ideals of $\mathscr{A}$. Then the mapping

$$\Delta \to \mathfrak{M} \ ; \ \phi \to \ker(\phi) \tag{9.6}$$

is a injection. Indeed, if $\phi_1, \phi_2 \in \Delta$ so that $\ker(\phi_1) = \ker(\phi_2)$, then by Lemma 0.1, there exists constant $c \in \mathbb{C}$ so that $\phi_1 = c\phi_2$. Since $\phi_1(e) = \phi_2(e) = 1$, thus $c = 1$ and $\phi_1 = \phi_2$.

It's natural to ask in which case the mapping (9.6) is a bijection? We will give an affirmative answer in the case that $\mathscr{A}$ is a complex commutative unital Banach algebra. This ono-on-one correspondence (9.6) is one of the key facts of the whole theory.

Theorem 9.21. *Let $\mathscr{A}$ be a complex commutative Banach algebra with unit element, and let Δ be the set of all complex homomorphisms of $\mathscr{A}$.*

(a) The mapping $\phi \mapsto \ker(\phi)$; $\Delta \to \mathfrak{M}$ is a bijection.

(b) $x \in \mathscr{A}$ is invertible if and only if $\phi(x) \neq 0$ for all $\phi \in \Delta$.

(c) The spectrum of x is given by $\sigma(x) = \{\phi(x) : \phi \in \Delta\}$.

Proof. (a). It suffices to show that the mapping $\phi \mapsto \ker(\phi)$; $\Delta \to \mathfrak{M}$ is a surjection. Let $J \in \mathfrak{M}$ be a maximal ideal of $\mathscr{A}$. Then by Theorem 9.19, J is closed and $\mathscr{A}/J$ is therefore a Banach algebra. It follows from Theorem 9.20 that $\mathscr{A}/J$ is division algebra. By the Gelfand-Mazur theorem, there is an isometrical (algebra) isomorphism h of $\mathscr{A}/J$ onto $\mathbb{C}$. Put $\phi_J = h \circ \pi$. Then $\phi \in \Delta$, and J is the kernel of ϕ_J, as desired.

(b). If x is invertible in $\mathscr{A}$, then by Proposition 9.10, $\phi(x) \neq 0$ for all $\phi \in \Delta$. If x is not invertible in $\mathscr{A}$, by Theorem 9.20, there is a maximal ideal J of $\mathscr{A}$ containing x. Then ϕ_J defined in the proof (a) satisfies $\phi_J(x) = 0$ and hence part (b) follows.

(c). Apply part (b) to $\lambda e - x$ in place of x. □

Now we are ready for the Gelfand representation.

9.4.2 The Gelfand Representation

In this subsection, unless otherwise specified, we suppose that $\mathscr{A}$ is a complex commutative unital Banach algebra. Let Δ be the set of all nonzero complex homomorphisms of $\mathscr{A}$. Since every complex homomorphism is a contraction linear operator (see Theorem 9.11), $\Delta \subset \mathscr{A}^*$.

The formula

$$\hat{x}(\phi) = \phi(x) \quad (\phi \in \Delta)$$

assigns to each $x \in \mathscr{A}$ a function $\hat{x} : \Delta \to \mathbb{C}$. We call $\hat{x}$ the *Gelfand representation* of x and also denote it by Γx. Indeed, $\hat{x}$ is the restriction of $J_{\mathscr{A}} x$ on $\Delta \subset \mathscr{A}^*$, where $J_{\mathscr{A}}$ is the natural embedding of $\mathscr{A}$ into $\mathscr{A}^{**}$.

The *Gelfand topology* of Δ is the weakest topology that makes every $\hat{x} \in \hat{\mathscr{A}}$ continuous. In other words, regarding Δ as a subset of $\mathscr{A}^*$ the Gelfand topology is the subspace topology inherited form $(\mathscr{A}^*, \tau_{w^*})$.

Let $\hat{\mathscr{A}}$ be the set of all $\hat{x}$, for $x \in \mathscr{A}$. Since Δ is equipped with the Gelfand topology, we can see that $\hat{\mathscr{A}} \subset C(\Delta)$, the algebra of all complex continuous functions on Δ. The mapping

$$\Gamma : \mathscr{A} \to \hat{\mathscr{A}} \subset C(\Delta)$$

is called the *Gelfand representation* of $\mathscr{A}$.

Lemma 9.22. *Endowed with the Gelfand topology, Δ is a compact Hausdorff space.*

Proof. As we have pointed, the Gelfand topology on Δ is the subspace topology inherited form $(\mathscr{A}^*, \tau_{w^*})$. Note that by Theorem 9.11, $\Delta \subset B_{\mathscr{A}^*}$, the closed unit ball in $\mathscr{A}^*$. By Theorem 5.15, it suffices to show that Δ is a weak-star closed subset of $B_{\mathscr{A}^*}$.

So it suffices to show that if (ϕ_α) is a net in Δ converging to $\phi \in \mathscr{A}^*$ relative to the weak star topology, then $\phi \in \Delta$. To this end, since $\phi \in \mathscr{A}^*$, we have only to show that ϕ preserve the multiplication. This is trivial. Indeed, for each x, y in $\mathscr{A}$, since $\phi_\alpha \to \phi$ in $\mathscr{A}^*$, we have $\phi_\alpha(z) \to \phi(z)$ for all z, and hence

$$\begin{aligned}\phi(xy) &= \lim_\alpha \phi_\alpha(xy) = \lim_\alpha \phi_\alpha(x)\phi_\alpha(y)\\ &= \lim_\alpha \phi_\alpha(x) \lim_\alpha \phi_\alpha(y) = \phi(x)\phi(y)\,,\end{aligned}$$

as desired. □

Remark 9.11. Since there is a one-to-one correspondence between Δ and $\mathfrak{M}$, all the maximal ideals of $\mathscr{A}$, we can also define the Gelfand topology on $\mathfrak{M}$. Equipped with its Gelfand topology, $\mathfrak{M}$ is a compact Hausdorff space, and is usually called the *maximal ideal space* of $\mathscr{A}$.

Theorem 9.23. *Let Δ be the all the nonzero complex homomorphisms on $\mathscr{A}$ equipped with the Gelfand topology.*

(a) The Gelfand representation is a continuous homomorphism of $\mathscr{A}$ onto a subalgebra $\hat{\mathscr{A}}$ of $C(\Delta)$.

(b) For each $x \in \mathscr{A}$, the range of $\hat{x} = \Gamma x$ is the spectrum $\sigma(x)$. Hence

$$\|\Gamma x\|_{C(\Delta)} = r_\sigma(x) \le \|x\|\,.$$

Proof. It's trivial to check that the mapping $x \to \hat{x}$ is a homomorphism and $\hat{\mathscr{A}}$ is a subalgebra of $C(\Delta)$. Observe that by Theorem 9.21, the range of $\hat{x}$

$$\{\hat{x}(\phi) : \phi \in \Delta\} = \{\phi(x) : \phi \in \Delta\} = \sigma(x)\,;$$

and hence

$$\|\hat{x}\|_{C(\Delta)} = \sup_{\phi\in\Delta} |\phi(x)| = \sup_{\lambda\in\sigma(x)} |\lambda| = r_\sigma(x) \le \|x\| \, .$$

Thus Γ is continuous. $\square$

Next, we are going to discuss the following questions.

- When does the Gelfand representation is an isomorphism of $\mathscr{A}$ onto $\hat{\mathscr{A}}$?
- When does the Gelfand representation is an isometry of $\mathscr{A}$ into $\hat{\mathscr{A}}$?
- When does the Gelfand representation is an isometric isomorphism of $\mathscr{A}$ onto $\hat{\mathscr{A}}$ and $\mathscr{A} = C(\Delta)$?

The third question is rather difficult, we will encounter it when talking about the C^* algebra. Let's see the first two questions.

When does the Gelfand representation is an isomorphism of $\mathscr{A}$ onto $\hat{\mathscr{A}}$? It's easy to see that

$$\begin{aligned} \hat{x} = 0 &\Leftrightarrow \hat{x}(\phi) = 0 \text{ for all } \phi \in \Delta \\ &\Leftrightarrow \phi(x) = 0 \text{ for all } \phi \in \Delta \\ &\Leftrightarrow x \in \bigcap_{\phi\in\Delta} \ker(\phi) \Leftrightarrow x \in \bigcap_{J\in\mathfrak{M}} J \, . \end{aligned}$$

Thus

$$\ker(\Gamma) = \bigcap_{\phi\in\Delta} \ker(\phi) = \bigcap_{J\in\mathfrak{M}} J \, .$$

In the study of algebra, for any algebra $\mathscr{A}$,

$$\operatorname{rad}(\mathscr{A}) \coloneqq \bigcap_{J\in\mathfrak{M}} J \, ,$$

the intersection of all maximal ideals of $\mathscr{A}$, is called the *radical* of $\mathscr{A}$. If $\operatorname{rad}(\mathscr{A}) = \{0\}$, then $\mathscr{A}$ is called *semisimple*. Clearly, $\Gamma : \mathscr{A} \to \hat{\mathscr{A}}$ is an isomorphism if and only if $\mathscr{A}$ is semisimple.

Exercise 9.3. Let $\mathscr{A}$ be a complex commutative unital Banach algebra. Then the following statements are equivalent.

(a) $\mathscr{A}$ is semisimple.

(b) Δ separates the points on $\mathscr{A}$, i.e., $\cap_{\phi\in\Delta}\ker(\phi) = \{0\}$.

(c) For each $x \in \mathscr{A}$, $r_\sigma(x) = \{0\}$ if and only if $x = 0$.

(d) For each $x \in \mathscr{A}$, $\sigma(x) = \{0\}$ if and only if $x = 0$.

Semisimple algebras have an important property which was earlier proved for the complex field $\mathbb{C}$ in Theorem 9.11:

Proposition 9.24. *Let $\mathscr{A}$, $\mathscr{B}$ be two complex unital Banach algebra. Suppose $\mathscr{B}$ is commutative and semisimple. Then each homomorphism $\psi : \mathscr{A} \to \mathscr{B}$ is continuous.*

Proof. By the closed graph theorem, it suffices to show that ψ is a closed operator. Suppose $x_n \to x$ in $\mathscr{A}$ and $\psi(x_n) \to y$ in $\mathscr{B}$. In order to show $y = \psi(x)$, since $\mathscr{B}$ is commutative and semisimple, it suffices to show that $\phi(y) = \phi(\psi(x))$ for each homomorphism $\phi : \mathscr{B} \to \mathbb{C}$.

By Theorem 9.20, since ϕ is continuous ,$\phi(\psi(x_n)) \to \phi(y)$. On the other hand, note that $\phi \circ \psi$ is also a complex homomorphism (of $\mathscr{A}$), $\phi \circ \psi$ is continuous, hence $\phi(\psi(x_n)) \to \phi(\psi(x))$. Thus $\phi(y) = \phi(\psi(x))$. $\square$

Remark 9.12. Every isomorphism between two semisimple commutative unital Banach algebras is a homeomorphism. In particular, this is true of every automorphism of a semisimple commutative unital Banach algebra. *The topology of such an algebra is therefore completely determined by its algebraic structure.*

When does the Gelfand representation is an isometry of $\mathscr{A}$ into $\hat{\mathscr{A}}$? If $\|\Gamma x\| \equiv r_\sigma(x) = \|x\|$ for all $x \in \mathscr{A}$, then we have $\|x^2\|^{1/2} \geq r_\sigma(x) = \|x\|$, thus $\|x^2\| \geq \|x\|^2$. Recall that $\left\|x^2\right\| \leq \|x\|^2$ is always true, then we get $\|x^2\| = \|x\|^2$.

On the contrary, if $\|x^2\| = \|x\|^2$ for all $x \in \mathscr{A}$, then by induction we have $\|x^{2^n}\| = \|x\|^{2^n}$, thus

$$\|\Gamma x\| \equiv r_\sigma(x) = \lim_{n\to\infty} \|x^{2^n}\|^{\frac{1}{2^n}} = \|x\|\,.$$

Therefore we get

Proposition 9.25. *The Gelfand transform is an isometry iff*

$$\left\|x^2\right\| = \|x\|^2$$

for all $x \in \mathscr{A}$.

In the end of this subsection, we point out that whether the Gelfand representation Γ is a homeomorphism of $\mathscr{A}$ onto $\hat{\mathscr{A}}$ can be decided by comparing $\left\|x^2\right\|$ with $\|x\|^2$, for all $x \in \mathscr{A}$.

Lemma 9.26. *Let $\mathscr{A}$ be a complex commutative unital Banach algebra. Then*

$$\inf_{\|x\|=1} \left\|x^2\right\| \leq \inf_{\|x\|=1} \|\Gamma x\| \leq \left(\inf_{\|x\|=1} \left\|x^2\right\|\right)^{1/2}.$$

Proof. Set $r = \inf_{\|x\|=1} \left\|x^2\right\|$. Then we have $\|x^2\| \geq r\|x\|^2$ for all $x \in \mathscr{A}$. By induction

$$\|x^{2^n}\| \geq r^{2^n-1}\|x\|^{2^n}$$

for all $n \geq 1$. Thus for each x,

$$\|\Gamma x\| = \lim_{n\to\infty} \|x^{2^n}\|^{\frac{1}{2^n}} \geq \lim_{n\to\infty} r^{1-\frac{1}{2^n}}\|x\| = r\|x\|\,.$$

To show another inequality, set $s = \inf_{\|x\|=1} \|\hat{x}\|$, then $\|\Gamma x\| \geq s\|x\|$ for all $x \in \mathscr{A}$. Since $\|\Gamma x\| = \inf_{n\geq 1} \|x^n\|^{\frac{1}{n}}$, we have $\|x^2\| \geq s^2\|x\|^2$. and hence $s^2 \leq \inf_{\|x\|=1} \left\|x^2\right\|$. $\square$

Theorem 9.27. *Let $\mathscr{A}$ be a commutative unital Banach algebra. $\mathscr{A}$ is semisimple and $\hat{\mathscr{A}}$ is closed in $C(\Delta)$ iff* $\inf_{\|x\|=1} \|x\|^2 > 0$.

Proof. If $\inf_{\|x\|=1} \|x\|^2 > 0$, then by the previous lemma $\inf_{\|x\|=1} \|\Gamma x\| > 0$ thus $\Gamma : \mathscr{A} \to \hat{\mathscr{A}}$ is a homeomorphism. Thus $\mathscr{A}$ is semisimple and $\hat{\mathscr{A}}$ is closed. Conversely, if $\mathscr{A}$ is semisimple, then Γ is injective, if $\hat{\mathscr{A}}$ is closed, then $\hat{\mathscr{A}}$ is a Banach space. By the conversing mapping theorem, Γ^{-1} is continuous and hence $\inf_{\|x\|=1} \|x\|^2 > 0$. □

9.4.3 Noncommutative Algebras

Noncommutative algebras always contain commutative ones. Their presence can sometimes be exploited to extend certain theorems from the commutative situation to the noncommutative one. On a trivial level, we have already done this: In the elementary discussion of spectra, our attention was usually fixed on one element $x \in \mathscr{A}$; the (closed) subalgebra $\mathscr{A}_x$ of $\mathscr{A}$ that x generates is commutative, and much of the discussion took place within $\mathscr{A}_x$. One possible difficulty was that x might have different spectra with respect to $\mathscr{A}$ and $\mathscr{A}_x$. There is a simple construction (Theorem 9.29) that circumvents this. Another device (Theorem 9.38) can be used when $\mathscr{A}$ has an involution.

If S is a subset of a complex unital Banach algebra $\mathscr{A}$, the *centralizer*, also called the *commutator* of S is the set

$$S' = C(S) \coloneqq \{x \in \mathscr{A} : xs = sx \text{ for every } s \in S\}.$$

Clearly S' is not empty since it always contains the unit element. We say that S *commutes* if any two elements of S commute with each other. Clearly, S commutes if and only if

$$S \subset S'.$$

We shall use the following simple properties of centralizers.

Proposition 9.28. *Let S be a subset of a complex unital Banach algebra $\mathscr{A}$. Then the following statements holds:*

(a) S' *is a closed subalgebra of* $\mathscr{A}$*;*

(b) $S \subset S''$*, and if S commutes, then S'' commutes.*

Proof. Indeed, if x and y commute with every $s \in S$, so do $\lambda x, x+y$, and xy. Since multiplication is continuous in $\mathscr{A}$, S' is closed. This proves (a).

Since every $s \in S$ commutes with every $x \in S'$, we have $S \subset S''$.

If S commutes, then $S \subset S'$. Since S is a subset of S', $S'' \subset S'$. By the same reason, $S'' \subset S'''$, thus S'' commutes. (b) holds. □

Theorem 9.29. *Suppose $\mathscr{A}$ is a complex unital Banach algebra, $S \subset \mathscr{A}$, S commutes, and $\mathscr{B} = S''$. Then $\mathscr{B}$ is a commutative unital Banach algebra, $S \subset \mathscr{B}$, and $\sigma_{\mathscr{B}}(x) = \sigma_{\mathscr{A}}(x)$ for ever $x \in \mathscr{B}$.*

Proof. Since $e \in \mathscr{B}$, the last proposition shows that $\mathscr{B}$ is a commutative unital Banach algebra that contains S. To show that $\sigma_{\mathscr{B}}(x) = \sigma_{\mathscr{A}}(x)$ for ever $x \in \mathscr{B}$, it suffices to show that for $\mathscr{B} \cap G(\mathscr{A}) \subset G(\mathscr{B})$.

Suppose $x \in \mathscr{B}$ and x is invertible in $\mathscr{A}$. We have to show that $x^{-1} \in \mathscr{B}$. Since $x \in \mathscr{B}$, $xy = yx$ for every $y \in S'$; hence $y = x^{-1}yx$, $yx^{-1} = x^{-1}y$. This says that $x^{-1} \in S'' = \mathscr{B}$. □

We give an application of Theorem 9.29.

Proposition 9.30. *Suppose $\mathscr{A}$ is a complex unital Banach algebra, $x \in \mathscr{A}, y \in \mathscr{A}$, and $xy = yx$. Then*

$$\sigma(x+y) \subset \sigma(x) + \sigma(y) \quad \textit{and} \quad \sigma(xy) \subset \sigma(x)\sigma(y)\,.$$

Proof. To see this, put $S = \{x, y\}$; put $\mathscr{B} = S''$. Then $x+y \in \mathscr{B}, xy \in \mathscr{B}$, and Theorem 9.29 shows that we have to prove that

$$\sigma_{\mathscr{B}}(x+y) \subset \sigma_{\mathscr{B}}(x) + \sigma_{\mathscr{B}}(y) \quad \text{and} \quad \sigma_{\mathscr{B}}(xy) \subset \sigma_{\mathscr{B}}(x)\sigma_{\mathscr{B}}(y)\,.$$

Since $\mathscr{B}$ is commutative, $\sigma_{\mathscr{B}}(z)$ is the range of the Gelfand transform $\hat{z}$, for every $z \in \mathscr{B}$. (The Gelfand transforms are now functions on the maximal ideal space of $\mathscr{B}$.) Since

$$(x+y)\hat{} = \hat{x} + \hat{y} \quad \text{and} \quad (xy)\hat{} = \hat{x}\hat{y}\,.$$

we have the desired conclusion. □

9.5 Examples

In some cases, the maximal ideal space of a given commutative Banach algebra can easily be described explicitly. In others, extreme pathologies occur. We shall now give some examples to illustrate this. However, we begin with the following useful lemma.

Lemma 9.31. *Let X be a compact Hausdorff space and Y be a Hausdorff space. If $\Phi : X \to Y$ is a continuous bijection, then Φ is a homeomorphism.*

Proof. It remains to show that Φ^{-1} is continuous. Observe that for each F closed in X, F is compact and hence $(\Phi^{-1})^{-1}(F) = \Phi(F)$ is compact in Y hence closed. Then the desired result follows. □

Example 9.14 (Continuous Functions on Compact Space)**.** Let X be a compact Hausdorff space, put $C(X) = C(X; \mathbb{C})$, with the supremum norm. Then $C(X)$ becomes a commutative unital Banach algebra. Then the maximal ideal space Δ of $C(X)$ is X itself. Specifically, associate each $x \in X$ the evaluation mapping $\phi_x : \mathscr{A} \to \mathscr{A}; f \mapsto f(x)$, then clearly $\phi_x \in \Delta$. The mapping

$$\Phi : x \mapsto \phi_x$$

gives a homeomorphism of X onto Δ. As a consequence, we can regard X as the maximal ideal space of $C(X)$; and then the Gelfand representation is the identity mapping on $C(X)$.

Firstly, since $C(X)$ separates points on X (Urysohn's lemma), $x \neq y$ implies $\phi_x \neq \psi_y$ if $x \neq y$, thus Φ is injective.

Secondly, we show that Φ is surjective. If this is false, there exists a $\psi \in \Delta$ so that for each $x \in X$, there exists $f_x \in C(X)$ so that

$$\phi_x(f_x) = f_x(x) \neq \psi(f_x)\,.$$

By the continuity of f_x, we can choose an open neighborhood U_x of x so that $|f_x(y) - \psi(f_x)| > 0$ for all $y \in U_x$. Then $\{U_x : x \in X\}$ is an open cover of the compact space X. Suppose $\{U_{x_1}, \dots, U_{x_k}\}$ is a finite subcover, then

$$f(y) = \sum_{j=1}^{k} |f_{x_j}(y) - \psi(f_{x_j})|^2 > 0 \;\text{ for all }\; y \in X\,.$$

On the one hand, $f \in C(X)$ and hence $1/f \in C(X)$, thus f is invertible in $C(X)$. On the other hand, since

$$f = \sum_{j=1}^{k} |f_{x_j} - \psi(f_{x_j})|^2 = \sum_{j=1}^{k} \left[f_{x_j} - \psi(f_{x_j})\right]\left[\overline{f}_{x_j} - \overline{\psi(f_{x_j})}\right]$$

we have $\psi(f) = 0$, which contradicts to the fact that f is invertible. Thus Φ must be surjective.

Finally we show that Φ is a homeomorphism. To show that continuity of Φ, let (x_α) be a net in X converging to x. Then $f(x_\alpha) \to f(x)$ for each $f \in C(X)$; in other words, $\phi_{x_\alpha}(f) \to \phi_x(f)$ for each $f \in C(X)$. Thus (ϕ_{x_α}) converges to ϕ_x in Δ and hence Φ is continuous. By Lemma 9.31, Φ is a homeomorphism as claimed.

Example 9.15 (Absolutely Convergent Fourier Series)**.** Let $A(\mathbb{T})$ be the Banach algebra of all absolutely convergent trigonometric series, as in Example 9.6. Then the maximal ideal space Δ of $A(\mathbb{T})$ is $\mathbb{T}$. Specifically, for each $t \in \mathbb{T}$ let $\phi_t(f) = f(t)$ for all $f \in A(\mathbb{T})$, then clearly $\phi \in \Delta$. The mapping

$$\Phi : t \mapsto \phi_t$$

gives a homeomorphism of $\mathbb{T}$ onto Δ. As a consequence, we can regard $\mathbb{T}$ itself as its the maximal ideal space; and the Gelfand representation is the embedding mapping of $A(\mathbb{T})$ into $C(\mathbb{T})$.

As an consequence, we have

Theorem (Wiener)**.** *Suppose $f \in A(\mathbb{T})$ so that $f(t) \neq 0$ for all $t \in \mathbb{T}$. Then $1/f \in A(\mathbb{T})$, i.e., $1/f$ has absolutely convergent Fourier series.*

Example 9.16. Let $A(\mathbb{D})$ be the disc algebra of all holomorphic functions on the open disk $\mathbb{D}$ that can be extended on $\overline{\mathbb{D}}$ continuously, as in Example 9.7. Then the maximal ideal space Δ of $A(\mathbb{D})$ is $\overline{\mathbb{D}}$. Specifically, for each $z \in \overline{\mathbb{D}}$, let $\phi_z(f) = f(z)$ for all $f \in A(\mathbb{D})$, then clearly $\phi \in \Delta$. The mapping

$$\Phi : z \mapsto \phi_z$$

gives a homeomorphism of $\overline{\mathbb{D}}$ onto Δ. As a consequence, we can regard $\overline{\mathbb{D}}$ itself as its the maximal ideal space; and the Gelfand representation is the embedding mapping of $A(\mathbb{D})$ into $C(\overline{\mathbb{D}})$.

As an consequence, we have

Theorem. *Suppose $\{f_1, \cdots, f_n\} \subset A(\mathbb{D})$ is non-vanishing on $\overline{\mathbb{D}}$, that is, for each $x \in \overline{\mathbb{D}}$, al least one of $\{f_1(x), \cdots, f_n(x)\}$ is nonzero. Then there exists $\{g_1, \cdots, g_n\} \subset A(\mathbb{D})$ so that*

$$f_1 g_1 + \cdots + f_n g_n \equiv 1\,.$$

Exercise 9.4. Let

$$\mathscr{A} = \left\{ f : \mathbb{Z} \to \mathbb{C} : \|f\| = \sum_{k=-\infty}^{\infty} |f(k)| 2^{|k|} < \infty \right\}$$

under the usual addition of scalar multiplication and the following multiplication

$$f * g(k) = \sum_{l=-\infty}^{\infty} f(k-l) g(l)$$

Show that

(a) $\mathscr{A}$ is a commutative Banach algebra;

(b) Let $K = \left\{z \in \mathbb{C} : \frac{1}{2} \leq |z| \leq 2\right\}$ then K is one-to-one correspondent with Δ and the Gelfand representation of $\mathscr{A}$ is the Laurent series that are absolutely convergent on K.

9.6 Involution and C*-Algebra

Definition 9.6. A mapping $x \mapsto x^*$ of a complex (not necessarily commutative) algebra $\mathscr{A}$ into $\mathscr{A}$ is called an *involution* if it is a *conjugate-linear anti-automorphism of period 2*. In other words, it has the following properties: for all $x, y \in \mathscr{A}$, and $\lambda \in \mathbb{C}$:

$$(x+y)^* = x^* + y^* \ , \ (\lambda x)^* = \bar{\lambda} x^* \, ;$$
$$(xy)^* = y^* x^* \ ; \ x^{**} = x \, .$$

We say $x \in \mathscr{A}$ is *hermitian*, or *self-adjoint* if$x^* = x$.

There are two classical example of the Banach algebra with an involution.

Example 9.17. Let X be a compact Hausdorff space and $\mathscr{A} = C(X, \mathbb{C})$, then

$$f \to \bar{f}$$

is an involution. Clearly f is hermitian iff f is real-valued.

Example 9.18. Let H be a Hilbert space and $\mathscr{A} = \mathcal{B}(H)$, then

$$A \mapsto A^*$$

where A^* is the adjoint operator of $\mathscr{A}$, is an involution on $\mathcal{B}(H)$. Clearly A is hermitian if and only if A is a self-adjoint operator. In practice, we will be most concerned with this example.

The following lemma is easy so we omit the proof.

Lemma 9.32. *If $\mathscr{A}$ is a Banach algebra with an involution, and if $x \in \mathscr{A}$, then*

(a) $x + x^*, i\,(x - x^*)$, *and* xx^* *are hermitian;*

(b) x *has a unique representation* $x = u + iv$, *with* $u, v \in \mathscr{A}$ *and both* u *and* v *hermitian;*

(c) the unit e *is hermitian;*

(d) x *is invertible in* $\mathscr{A}$ *if and only if* x^* *is, in which case* $(x^*)^{-1} = \left(x^{-1}\right)^*$; *and*

(e) $\lambda \in \sigma(x)$ *if and only if* $\bar{\lambda} \in \sigma\,(x^*)$.

Theorem 9.33. *Let $\mathscr{A}$ be a semisimple commutative unital Banach algebra. Then every involution on $\mathscr{A}$ is continuous.*

Proof. Although the involution is conjugate-linear, the closed graph theorem holds in this case. So it suffices to show that if $x_n \to x$ and $x_n^* \to y$ in $\mathscr{A}$, then $y = x^*$. Since $\mathscr{A}$ is commutative and semisimple, it suffices to show that

$$\phi(x^*) = \phi(y) \quad \text{for all} \quad \phi \in \Delta\,.$$

Clearly $\phi(y) = \lim_n \phi(x_n^*)$. Note that if we define φ by $\varphi(x) = \overline{\phi(x^*)}$, then $\varphi \in \Delta$. Thus $\varphi(x_n) = \overline{\phi(x_n^*)} \to \varphi(x) = \overline{\phi(x^*)}$, and hence $\phi(x^*) = \lim_n \phi(x_n^*)$. Then we get $\phi(x^*) = \phi(y)$ as desired. $\square$

Definition 9.7. A complex unital Banach algebra $\mathscr{A}$ with an involution $x \mapsto x^*$ that satisfies

$$\|xx^*\| = \|x\|^2 \quad \text{for all } x \in \mathscr{A} \tag{9.7}$$

is called a B^*-*algebra.* The identity (9.7) is called the B^*-condition.

If $x \in \mathscr{A}$ is hermitian, then we have $\|x^2\| = \|x\|^2$.

Note that $\|x\|^2 = \|xx^*\| \leq \|x\| \, \|x^*\|$ implies $\|x\| \leq \|x^*\|$, hence also

$$\|x^*\| \leq \|x^{**}\| = \|x\|$$

Thus

$$\|x^*\| = \|x\| \tag{9.8}$$

in every B^* -algebra. It also follows that

$$\|xx^*\| = \|x\| \, \|x^*\| \qquad \text{for all } x \in \mathscr{A} . \tag{9.9}$$

Identity (9.9) is often called the *C^*-condition*. Complex unital Banach algebras with an involution that satisfy the C^*-condition (9.9) is called a *C^*-algebra*. It's easy to see that (9.8) and (9.9) implies (9.7).

Remark 9.13. The term B^*-algebra was introduced by C. E. Rickart in 1946 to describe Banach algebras with an involution. Clearly a B*-algebra is also a C*-algebra. Conversely, the C^*-condition implies the B^*-condition. This is nontrivial, and can be proved without using the condition (9.8). For these reasons, the term B^*-algebra is rarely used in current terminology, and has been replaced by the term 'C^*-algebra'. The term C*-algebra was introduced by I. E. Segal in 1947 to describe norm-closed subalgebras of $\mathcal{B}(H)$, namely, the space of bounded operators on some Hilbert space H. 'C' stood for 'closed'. In his paper Segal defines a C^*-algebra as a "uniformly closed, self-adjoint algebra of bounded operators on a Hilbert space".

Definition 9.8. Let $\mathscr{A}$ and $\mathscr{B}$ be two complex algebra with involutions. Then $\psi : \mathscr{A} \to \mathscr{B}$ is called a *$*$-isomorphism*, if ψ is an algebra isomorphism preserving involution, i.e., $\psi(x^*) = \psi(x)^*$ for all $x \in \mathscr{A}$.

9.6.1 Commutative C*-algebra

It's easy to see that the algebra $C(X)$, consisting of complex-valued continuous functions on compact Hausdorff space X, is a commutative C^*

algebra. Surprisingly, the converse is true. This is the key to the proof of the spectral theorem that will be given later.

We need the following lemma, which does not depend on the commutativity conditions.

Lemma 9.34. *Let $\mathscr{A}$ be a C^*-algebra and $x \in \mathscr{A}$ is hermitian. Then*

$$\phi(x) \in \mathbb{R} \quad \textit{for all} \quad \phi \in \Delta\,.$$

Proof. We shall show that $\mathrm{Im}(\phi(x)) = 0$. Consider $x + ite$, where $t \in \mathbb{R}$. Since

$$|\phi(x+ite)| = |\phi(x) + it| \leq \|x + ite\|\,,$$

and $(x+ite)^* = x - ite$, by C^*-condition we have

$$\begin{aligned}
|\phi(x)+it|^2 &= |\,\mathrm{Re}(\phi(x))|^2 + |\,\mathrm{Im}(\phi(x)) + t|^2 \\
&= |\,\mathrm{Re}(\phi(x))|^2 + |\,\mathrm{Im}(\phi(x))|^2 + t^2 + 2\,\mathrm{Im}(\phi(x))t \\
&\leq \|x+ite\|^2 = \|(x+ite)(x-ite)\| \leq \|x^2\| + t^2\,.
\end{aligned}$$

Then

$$|\,\mathrm{Re}(\phi(x))|^2 + |\,\mathrm{Im}(\phi(x))|^2 + 2\,\mathrm{Im}(\phi(x))t \leq \|x^2\|\,.$$

Since $t \in \mathbb{R}$ is arbitrary, it must be $\mathrm{Im}(\phi(x)) = 0$ as required. $\square$

Theorem 9.35 (Gelfand-Naimark)**.** *Let $\mathscr{A}$ be a commutative C^*-algebra. Let Δ be all the nonzero complex homomorphisms on X. Then the Gelfand representation Γ is an isometric $*$-isomorphism of $\mathscr{A}$ onto $C(\Delta)$.*

Proof. It suffices to show that (i) Γ preserves the involution; (ii) Γ is an isometry, and; (iii) $\Gamma : \mathscr{A} \to C(\Delta)$ is surjective.

(i). Since the involution on $C(\Delta)$ is the conjugation, to show that $\Gamma(x^*) = \overline{\Gamma(x)}$, we have to show

$$\phi(x^*) = \overline{\phi(x)}\,.$$

Let $x = u + iv$ with u and v hermitian, then since $\phi(u)$ and $\phi(v)$ are real number

$$\begin{aligned}\phi(x^*) &= \phi(u - iv) = \phi(u) - i\phi(v) \\ &= \overline{\phi(u) + i\phi(v)} = \overline{\phi(x)}\,,\end{aligned}$$

as desired.

(ii). In order that Γ is an isometry, we shall show that for each $x \in \mathscr{A}$,

$$\|x^2\| = \|x\|^2\,.$$

Since x and x^* commutes, and x^*x is hermitian, we have

$$\begin{aligned}\|x^2\|^2 &= \|(x^2)^*x^2\| = \|(x^*)^2x^2\| \\ &= \|(x^*x)^2\| = \|x^*x\|^2 = \|x\|^4\,,\end{aligned}$$

and then the desired result follows.

(iii). Since we have shown Γ is an isometry, then $\hat{\mathscr{A}}$ is a closed subalgebra in $C(\Delta)$. So it suffices to show that $\hat{\mathscr{A}}$ is dense in $C(\Delta)$. We will prove this by Theorem 3.48. Clearly the constant function $1 = \Gamma(e) \in \hat{\mathscr{A}}$, thus $\hat{\mathscr{A}}$ vanishes nowhere. Since $\overline{\Gamma(x)} = \Gamma(x^*)$, $\hat{\mathscr{A}}$ is closed under complex conjugation. For $\phi_1 \neq \phi_2$, there exists $x \in \mathscr{A}$ so that $\phi(x_1) \neq \phi(x_2)$, thus $\Gamma(x_1)(\phi) \neq \Gamma(x_2)(\phi)$ and hence $\hat{\mathscr{A}}$ separates the points. Thus by the Stone-Weierstrass theorem (see Theorem 3.48), $\hat{\mathscr{A}}$ is dense in $C(\Delta)$. $\square$

9.6.2 Noncommutative C*-algebra

Clearly, $\mathcal{B}(H)$ is a C^*-algebra. Besides, every closed subalgebra of $\mathcal{B}(H)$ which is closed under involution is also a C^*-algebra. Surprisingly, the converse is true. This is a deep result and we will not show it.

Theorem (Gelfand-Naimark)**.** *Let $\mathscr{A}$ be a C^*-algebra. Then there exists a closed unital subalgebra $\mathscr{B}$ of $\mathcal{B}(H)$, which is also closed under $*$ operation, i.e., $\mathscr{B}^* = \mathscr{B}$, so that $\mathscr{A}$ isometrically $*$-isomorphic to $\mathscr{B}$.*

Let $\mathscr{A}$ be a C^*-algebra. Now we concern the following question as before. To use the result for the commutative case, we want to find some commutative sub-C^*-algebra $\mathscr{B}$ of $\mathscr{A}$ so that for $x \in \mathscr{B}$ there holds $\sigma_{\mathscr{B}}(x) = \sigma_{\mathscr{A}}(x)$.

Definition 9.9. Let $\mathscr{A}$ be a C^*-algebra. $x \in \mathscr{A}$ is said to be *normal* if $xx^* = x^*x$. A set $S \subset \mathscr{A}$ is said to be *normal* if S commutes and if $x^* \in S$ whenever $x \in S$.

Clearly, a closed unital subalgebra $\mathscr{B}$ of $\mathscr{A}$ is a commutative C^*-algebra if and only if it is normal.

Since the union of a family of totally ordered normal subsets is still normal, by Zorn's lemma, every normal subset S of $\mathscr{A}$ is contained in a maximal one. The following theorem asserts that the maximal normal subset of $\mathscr{A}$ satisfies our requests.

Theorem 9.36. *Let $\mathscr{A}$ be a C^*-algebra, and $\mathscr{B}$ is a maximal normal subset of $\mathscr{A}$. Then $\mathscr{B}$ is a commutative sub-C^*-algebra of $\mathscr{A}$, and*

$$\sigma_{\mathscr{B}}(x) = \sigma_{\mathscr{A}}(x) \quad \text{for every } x \in \mathscr{B}\,.$$

Proof. We begin with a simple criterion for membership in $\mathscr{B}$:

If $x \in \mathscr{A}$ is normal and if x commutes with $\mathscr{B}$, i.e., $xy = yx$ for all $y \in \mathscr{B}$, then $x \in \mathscr{B}$.

Indeed, if x satisfies these conditions, we also have $xy^* = y^*x$ for all $y \in \mathscr{B}$, since $\mathscr{B}$ is normal, and therefore $x^*y = yx^*$. It follows that $\mathscr{B} \cup \{x, x^*\}$ is normal. Hence $x \in \mathscr{B}$, since $\mathscr{B}$ is maximal.

This criterion makes it clear that sums and products of members of $\mathscr{B}$ are in $\mathscr{B}$, the unit element is in $\mathscr{B}$. Thus $\mathscr{B}$ is a commutative unital algebra.

Suppose $x_n \in \mathscr{B}$ and $x_n \to x$. since $x_n y = y x_n$ for all $y \in \mathscr{B}$, and multiplication is continuous, we have $xy = yx$ and therefore also

$$x^*y = (y^*x)^* = (xy^*)^* = yx^*$$

In particular, $x^*x_n = x_nx^*$ for all n, which leads to $x^*x = xx^*$. Hence $x \in \mathscr{B}$, by the above criterion. This proves that $\mathscr{B}$ is closed.

To prove $\sigma_{\mathscr{B}}(x) = \sigma_{\mathscr{A}}(x)$ for every $x \in \mathscr{B}$, it suffices to show that $G(\mathscr{B}) = \mathscr{B} \cap G(\mathscr{A})$. Assume $x \in \mathscr{B}, x^{-1} \in \mathscr{A}$. since x is normal, so is x^{-1}, and since x commutes with every $y \in \mathscr{B}$, so does x^{-1}. Hence $x^{-1} \in \mathscr{B}$. $\square$

Our next application of Theorem 9.36 will extend some consequences of Theorem 9.35 to arbitrary (not necessarily commutative) C^* algebras.

In a C^*-algebra, the statement "$x \geq 0$" means that $x = x^*$ and that $\sigma(x) \subset [0, \infty)$.

Proposition 9.37. *Let $\mathscr{A}$ be a C^* algebra. Then the following holds.*

(a) If $x \in \mathscr{A}$ is Hermitian, then $\sigma_{\mathscr{A}}(x) \subset \mathbb{R}$.

(b) If $x \in \mathscr{A}$ is normal, then $r_\sigma(x) = \|x\|$. In particular, $r_\sigma(y^*y) = \|y\|^2$ for all $y \in \mathscr{A}$.

(c) If $u, v \in \mathscr{A}$, $u, v \geq 0$, then $u + v \geq 0$.

(d) If $y \in \mathscr{A}$, then $y^*y \geq 0$. In particularly, $e + y^*y$ is invertible in $\mathscr{A}$.

Proof of (a), (b). Every normal $x \in \mathscr{A}$ lies in a maximal normal set $\mathscr{B} \subset \mathscr{A}$. By the preceding theorem, $\mathscr{B}$ is a commutative C^*-algebra which is isometrically $*$-isomorphic to its Gelfand transform $\hat{\mathscr{B}} = C(\Delta)$ and which has the property that

$$\sigma(z) = \hat{z}(\Delta) \text{ for } z \in \mathscr{B}. \tag{9.10}$$

Here $\sigma(z)$ is the spectrum of z relative to $\mathscr{A}$, Δ is the maximal ideal space of $\mathscr{B}$, and $\hat{z}(\Delta)$ is the range of the Gelfand transform of z, regarded as an element of $\mathscr{B}$.

If $x = x^*$, Lemma 9.34 shows that $\hat{x}$ is a real-valued function on Δ. Hence (9.38) implies (a).

For any normal x, (9.38) implies $r_\sigma(x) = \|\hat{x}\|_{C(\Delta)}$. Also, $\|\hat{x}\|_{C(\Delta)} = \|x\|$, since $\mathscr{B}$ and $\hat{\mathscr{B}}$ are isometric. Thus $r_\sigma(x) = \|x\|$. If $y \in \mathscr{A}$, then yy^* is hermitian, and hence $r_\sigma(yy^*) = \|yy^*\| = \|y\|^2$. This proves (b). $\square$

We can not use the previous argument to show (c), (d) since we can not say that u and v commute or y and y^* commute.

Proof of (c). Since $\sigma(u) \subset [0, \|u\|]$, so that

$$\sigma(\|u\|e - u) \subset [0, \|u\|]$$

and this implies therefore that $\big\|\|u\|e - u\big\| \leq \|u\|$. For the same reason, $\big\|\|v\|e - v\big\| \leq \|v\|$. Hence

$$\big\|\|u\|e + \|v\|e - (u+v)\big\| \leq \|u\| + \|v\|\,.$$

since $\|u\|e + \|v\|e - (u+v)$ is hermitian, (a) implies that its spectrum is a subset of $\mathbb{R}$. Thus

$$\sigma(\|u\|e + \|v\|e - (u+v)) \subset [-\|u\| - \|v\|, \|u\| + \|v\|]\,.$$

Therefore,

$$\sigma(u+v) \subset [0, 2\|u\| + 2\|v\|]\,,$$

which deduce that $u + v \geq 0$. $\square$

Proof of (d). We turn to the proof of (d). Put $x = y^*y$. Then x is hermitian, and if $\mathscr{B}$ is chosen as in the proof of (a),(b), then $\hat{x}$ is a real-valued function on Δ. By (9.38), we have to show that $\hat{x} \geq 0$ on Δ.

We begin with a classic mistake: For each $\phi \in \Delta$,

$$\hat{x}(\phi) = \phi(x) = \phi(y^*y) = \phi(y^*)\phi(y) = \overline{\phi(y)}\phi(y) = |\phi(y)|^2 \geq 0\,.$$

This is wrong since ϕ is a complex homomorphism on $\mathscr{B}$, and we can not say $y, y^* \in \mathscr{B}$. Thus $\phi(y)$ and $\phi(y^*)$ may not make sense and $\phi(y^*y) = \phi(y^*)\phi(y)$, $\phi(y^*) = \overline{\phi(y)}$ may not hold.

Next, we give the proof. Since $\hat{\mathscr{B}} = C(\Delta)$, there exists $z \in \mathscr{B}$ such that

$$\hat{z} = |\hat{x}| - \hat{x} \quad \text{on } \Delta\,. \tag{9.11}$$

Note that $\hat{z}$ is real, then $z = z^*$. Put

$$yz = w = u + iv\,,$$

where u and v are hermitian elements of $\mathscr{A}$. Then

$$w^*w = z^*y^*yz = zxz = z^2x \tag{9.12}$$

and therefore

$$ww^* = 2u^2 + 2v^2 - w^*w = 2u^2 + 2v^2 - z^2x\,. \tag{9.13}$$

Since $u = u^*$, $\sigma(u)$ is real, by (a), hence $u^2 \geq 0$, by the spectral mapping theorem. Likewise $v^2 \geq 0$. By (9.11), $\hat{z}^2\hat{x} \leq 0$ on Δ. Since $z^2x \in \mathscr{B}$, it follows from that $-z^2x \geq 0$. Now (9.13) and (c) imply that

$$ww^* \geq 0\,.$$

But $\sigma(w^*w) \subset \sigma(ww^*) \cup \{0\}$ (Exercise 9.1). Hence $w^*w \geq 0$. By (9.12), this means that $\hat{z}^2\hat{x} \geq 0$ on Δ. By (9.11), this last inequality holds only when $\hat{x} = |\hat{x}|$. Thus $\hat{x} \geq 0$ as desired. □

Remark 9.14. In fact, (d) is a key step in the proof of the Gelfand-Naimark theorem. On the other hand, if we admit the Gelfand-Naimark theorem, (d) is a trivial consequence.

Equality of spectra can now be proved in yet another situation, in which commutativity plays no role.

Theorem 9.38. *Let $\mathscr{A}$ be a C^*-algebra. Let $\mathscr{B}$ be a sub-C^*-algebra of $\mathscr{A}$. Then*

$$\sigma_{\mathscr{B}}(x) = \sigma_{\mathscr{A}}(x) \quad \textit{for every } x \in \mathscr{B}\,.$$

Proof. We shall show that $G(\mathscr{B}) = \mathscr{B} \cap G(\mathscr{A})$. Suppose $x \in \mathscr{B}$ and x has an inverse in $\mathscr{A}$. We have to show that $x^{-1} \in \mathscr{B}$.

Since x is invertible in $\mathscr{A}$, so is x^*, hence also x^*x, and therefore $0 \notin \sigma_{\mathscr{A}}(x^*x)$. Observe that x^*x is hermitian, by Proposition 9.37, $\sigma_{\mathscr{A}}(x^*x) \subset \mathbb{R}$, so that $\varrho_{\mathscr{A}}(x^*x)$ is connected in $\mathbb{C}$. Then Theorem 9.9 shows now that $\sigma_{\mathscr{B}}(x^*x) = \sigma_{\mathscr{A}}(x^*x)$. Hence $(x^*x)^{-1} \in \mathscr{B}$. Finally since $x^* \in \mathscr{B}$,

$$x^{-1} = (x^*x)^{-1} x^* \in \mathscr{B}$$

as desired. □

9.6.3 Continuous Functional Calculus for Normal Operators

The construction of the continuous functional calculus for normal operators is based on several lemmas. Assume throughout that H is a complex Hilbert space and that $N \in \mathcal{B}(H)$ is a normal operator. Let $\mathscr{A}_N \subset \mathcal{B}(H)$ be the smallest sub-C^*-algebra that contains N; i.e.,

$$\mathscr{A}_N = \operatorname{cl}\{p(N, N^*) : p \text{ is a polynomial in two variables}\}. \tag{9.14}$$

Clearly, $\mathscr{A}_N$ is a commutative sub-C^*-algebra of $\mathcal{B}(H)$, the spectrum of N relative to $\mathcal{B}(H)$ and $\mathscr{A}_N$ coincide by Theorem 9.38. Therefore we always denote it by $\sigma(N)$.

Theorem 9.39. *Let the commutative C^*-algebra $\mathscr{A}_N$ be given above. Let Δ be all the nonzero complex homomorphisms of $\mathscr{A}_N$. Let $\hat{N}$ be the Gelfand representation of N. Then the following hold.*

(a) *$\hat{N}$ is a homeomorphism of Δ onto $\sigma(N)$.*

(b) *For each $f \in C(\sigma(N))$, let $\tilde{f}(N) \in \mathscr{A}_N \subset \mathcal{B}(H)$ satisfy $\phi(\tilde{f}(N)) = f(\phi(N))$ for all $\phi \in \Delta$. Then the mapping*

$$f \mapsto \tilde{f}(N)\ ;\ C(\sigma(N)) \to \mathscr{A}_N$$

is an isometric $*$-isomorphism.

Since $\mathscr{A}_N$ is a commutative C^*-algebra, the Gelfand representation $\Gamma : \mathscr{A}_N \to \Delta$ is a isomorphic $*$-isomorphism. Thus $\tilde{f}(N)$ is well defined. Indeed,

$$\tilde{f}(N) = \Gamma^{-1}(f(\Gamma(N)))$$

Proof. As we know, $\hat{N}$ is a continuous function on Δ whose range is $\sigma(N)$, i.e., $\hat{N}$ is a continuous surjection of Δ onto $\sigma(N)$.

To see that $\hat{N}$ is injective, suppose $\phi_1 \in \Delta, \phi_2 \in \Delta$, and $\hat{N}\left(\phi_1\right) = \hat{N}\left(\phi_2\right)$, that is, $\phi_1(N) = \phi_2(N)$. Then that $\phi_1\left(N^*\right) = \overline{\phi_1(N)} = \overline{\phi_2(N)} = \phi_2\left(N^*\right)$. If p is any polynomial in two variables, it follows that

$$\phi_1\left(p\left(N, N^*\right)\right) = \phi_2\left(p\left(N, N^*\right)\right) ,$$

since ϕ_1 and ϕ_2 are homomorphisms. Since elements of the form $p\left(N, N^*\right)$ are dense in $\mathscr{A}_N$, the continuity of ϕ_1 and ϕ_2 implies therefore $\phi_1 = \phi_2$. We have proved that $\hat{N}$ is one-to-one.

Since Δ and $\sigma(N)$ both are compact, it follows that $\hat{N}$ is a homeomorphism of Δ onto $\sigma(N)$, and (a) follows.

The mapping $f \to f \circ \hat{N}$ is therefore an isometric $*$-isomorphism of $C(\sigma(N))$ onto $C(\Delta)$. Since the inverse of the Gelfand representation Γ^{-1} is also an isometric $*$-isomorphism of $C(\Delta)$ onto $\mathscr{A}_N$, then the mapping $f \to \Gamma^{-1} \circ f \circ \hat{N}$ is therefore an isometric $*$-isomorphism of $C(\sigma(N))$ onto $\mathscr{A}_N$, as desired. □

With these preparations in place we are ready to establish the continuous functional calculus for normal operators on Hilbert spaces.

Theorem 9.40 (Continuous Functional Calculus)**.** *Let H be a complex Hilbert space, let $N \in \mathcal{B}(H)$ be a bounded normal operator. Then the isometric $*$-isomorphism*

$$f \mapsto \tilde{f}(N) \ ; \ C(\sigma(N)) \to \mathscr{A}_N \subset \mathcal{B}(H) \tag{9.15}$$

satisfies the following axioms.

(a) *(Normalization) If* $f(\lambda)=\lambda$ *for all* $\lambda \in \sigma(N)$ *then* $\tilde{f}(N)=N$.

(b) *(Commutative) If* $A \in \mathcal{B}(H)$ *satisfies* $NA = AN$ *and* $N^*A = AN^*$ *then* $\tilde{f}(N)A = A\tilde{f}(N)$ *for all* $f \in C(\sigma(N))$.

(c) *(Eigenvector) If* $\lambda \in \sigma_p(N)$ *and* $x \in H$ *satisfy* $Nx = \lambda x$ *then*

$$\tilde{f}(N)x = f(\lambda)x \quad \textit{for all } f \in C(\sigma(N)) \,.$$

(d) *(Spectrum) For every* $f \in C(\sigma(N))$ *the operator* $\tilde{f}(N)$ *is normal and*

$$\sigma(\tilde{f}(N)) = f(\sigma(N)) \,.$$

(e) *(Positive) If* $f \in C(\sigma(N), \mathbb{R})$ *and* $f \geq 0$ *then* $\tilde{f}(N) = \tilde{f}(N)^* \geq 0$.

(f) *(Composition) If* $f \in C(\sigma(N))$ *and* $g \in C(f(\sigma(N)))$ *then*

$$\widetilde{(g \circ f)}(N) = \tilde{g}(\tilde{f}(N)) \,.$$

Proof. We prove part (a). If $f(\lambda) = \lambda$ for all $\lambda \in \sigma(N)$, then $f \circ \hat{N} = \hat{N}$. Thus $\tilde{f}(N) = N$.

We prove part (b). In fact, by the definition of $\mathscr{A}_N$, if $AN = NA$ then A commutes with $p(N, N^*)$ for each polynomial p in two variables. By the continuity of multiplication, A commutes with $\mathscr{A}_N$ and hence (b) follows.

We prove part (c). Let $\lambda \in \sigma_p(N)$ and $x \in H$ such that $Nx = \lambda x$. Then $N^*x = \bar{\lambda}x$ by Theorem 8.14. Hence $p(N, N^*)\,x = p(\lambda, \overline{\lambda})x$ for every polynomial p in z and $\bar{z}$. By the Stone-Weierstrass theorem, $\{p(z, \bar{z}) :$ p is a polynomial in two variables$\}$ is dense in $C(\sigma(N))$, using $f \mapsto \tilde{f}(N)$ is isometric we have $\tilde{f}(N)x = f(\lambda)x$, as required.

We prove part (d). By the Gelfand representation, letting Δ be all the nonzero complex homomorphism on $\mathscr{A}_N$, then

$$\begin{aligned}\sigma(\tilde{f}(N)) &= \left\{\widehat{\tilde{f}(N)}(\phi) : \phi \in \Delta\right\} = \left\{f(\hat{N}(\phi)) : \phi \in \Delta\right\} \\ &= \{f(\phi(N)) : \phi \in \Delta\} = \{f(\lambda) : \lambda \in \sigma(N)\} = f(\sigma(N)) \,.\end{aligned}$$

(e) follows directly from (d) and the fact that $f \mapsto f(N)$ preserve the involution.

We prove part (f). By the definition

$$\Big(\widetilde{(g\circ f)}(N)\Big)^\wedge = g\circ f\circ \hat{N}\,;$$

and

$$\Big(\tilde{g}(\tilde{f}(N))\Big)^\wedge = g\circ(\widehat{\tilde{f}(N)}) = g\circ f\circ \hat{N}\,;$$

thus (f) holds. $\square$

Remark 9.15. Let H be an infinite-dimensional complex Hilbert space. It is useful to examine the special case of Theorem 9.40 where the normal operator $N \in \mathcal{B}(H)$ is compact, which we now assume. By Theorem 8.33, there are complex numbers $\{\lambda_n\}$ projections $\{P_n\}$ with $P_nP_m = 0$ if $n \neq m$ so that

$$A = \text{s-}\sum_n \lambda_n P_n\,.$$

Then it's easy to see that for a polynomial p in two variables

$$p(A, A^*) = \text{s-}\sum_n p(\lambda_n, \bar{\lambda}_n) P_n\,.$$

Then by the Stone-Weierstrass theorem, $\{p(z,\bar{z}) : p \text{ is a polynomial in two variables}\}$ is dense in $C(\sigma(N))$, using $f \mapsto \tilde{f}(N)$ is isometric we have

$$f(A) = \text{s-}\sum_n f(\lambda_n) P_n\,.$$

We end this section with a commutativity theorem, as an application of the continuous functional calculus.

Theorem 9.41. *Let $N \in \mathscr{B}(H)$ be normal. Then for each $T \in \mathscr{B}(H)$ commutes with N, T must commutes with N^*.*

Proof. Since $NT = TN$ holds, then by induction we have $N^k T = TN^k$ for integer $k \geq 0$. Hence

$$\exp(N)T = T\exp(N)\,.$$

Thanks to the continuous functional calculus $\exp(N)$ has inverse $\exp(-N)$. Thus

$$T = \exp(-N)T\exp(N)\,.$$

By the continuous functional calculus, we have

$$\exp\left(N^*\right)T\exp\left(-N^*\right) = \exp(N^* - N)T\exp(N - N^*)$$

Since $|e^{z-\bar{z}}| = 1$ for all $z \in \sigma(N)$, by the continuous functional calculus, we have

$$\|\exp(N^* - N)\| = \|\exp(N^* - N)\| = 1\,.$$

thus

$$\|\exp\left(N^*\right)T\exp\left(-N^*\right)\| \leq \|T\|\,.$$

We now define

$$f(\lambda) = \exp\left(\lambda N^*\right)T\exp\left(-\lambda N^*\right) \text{ for } \lambda \in \mathbb{C}\,.$$

Then f is an analytic $\mathcal{B}(H)$-valued function, since $\lambda \mapsto \exp\left(\lambda N^*\right)$ and $\lambda \mapsto \exp\left(-\lambda N^*\right)$ are. The hypotheses of the theorem hold with $\bar{\lambda}N$ in place of N. Therefore we have

$$\|f(\lambda)\| \leq \|T\| \text{ for all } \lambda \in \mathbb{C}\,.$$

Then f is a bounded entire function. By Liouville's theorem, f is a constant function, so $f(\lambda) = f(0) = T$, for every $\lambda \in \mathbb{C}$. Hence

$$\exp\left(\lambda N^*\right)T = T\exp\left(\lambda N^*\right) \text{ for all } \lambda \in \mathbb{C}\,.$$

Compute the derivative of the both side, we obtain $N^*T = TN^*$. $\square$

Remark 9.16. Following this result, commutative axiom (part (b)) of Theorem 9.40 can be reduced as following: If $A \in \mathcal{B}(H)$ satisfies $NA = AN$, then $\tilde{f}(N)A = A\tilde{f}(N)$ for all $f \in C(\sigma(N))$.

In fact, a more general result is true:

Exercise 9.5 (Fuglede-Putnam-Rosenblum). Assume that $M, N, T \in \mathscr{B}(H)$, M and N are normal, and $MT = TN$, then $M^*T = TN^*$.

9.7 Spectral Decomposition of Normal Operators on Hilbert Space

[Spectral Decomposition of Normal Operators]

Let H be a complex Hilbert space. The main results of this section are the following. Firstly, if $\mathscr{A}$ is a commutative sub-C^*-algebra of $\mathcal{B}(H)$, we can give a specific formula of the inverse of the Gelfand transform.

The second is a important consequence of the first one. If $N \in \mathcal{B}(H)$ is normal, then there exists complex Borel measures $\{E_{x,y} : x, y \in H\}$ so that

$$\langle Nx, y\rangle = \int_{\sigma(N)} \lambda\, E_{x,y}(d\lambda)\,.$$

Moreover, $\{E_{x,y} : x, y \in H\}$ is very special. To make it clear, we begin some preliminaries.

9.7.1 Projection Valued Measures

Denote by $\mathcal{P}(H)$ all the orthogonal projections on H. Let $(\Omega, \mathcal{F})$ be a measurable space. Indeed we always assume that Ω is a Hausdorff space and $\mathcal{F}$ is the Borel algebra on Ω.

Definition 9.10. A *projection valued measure* P on $(\Omega, \mathcal{F})$ is a map $\mathcal{F} \to \mathcal{P}(H)$ which assigns to every measurable set A a projection $P(A)$ on H and satisfies the following axioms.

(a) (Normalization) $P(\emptyset) = 0,\ P(\Omega) = I$.

(b) (Intersection) If $A_1,\ A_2$ are in $\mathcal{F}$ then

$$P(A_1 \cap A_2) = P(A_1)P(A_2)\,.$$

(c) (σ-Additive) If $\{A_n\}$ is a sequence of pairwise disjoint sets in $\mathcal{F}$ then

$$P\big(\bigcup_{n=1}^{\infty} A_n\big) = \text{s-}\sum_{n=1}^{\infty} P(A_n)\,.$$

Remark 9.17. One should note that (b) implies that $P(A_1)$ and $P(A_2)$ commute for every $A_1,\ A_2$ in $\mathcal{F}$. Moreover, in fact (a) and (c) implies (b).

First we note that $P\left(A_1\right) P\left(A_2\right) = 0$ if $A_1, A_2 \in \mathcal{F}$ are disjoint. Indeed since $P\left(A_1\right) + P\left(A_2\right) = P(A_1 \cup A_2)$ is a projection. By Proposition 8.24 we have $P\left(A_1\right) P\left(A_2\right) = 0$. Now for each A_1, A_2 in $\mathcal{F}$, we have

$$\begin{aligned} P(A_1)P(A_2) &= \left[P(A_1 \cap A_2) + P(A_1 \backslash A_2)\right]\left[P(A_1 \cap A_2) + P(A_2 \backslash A_1)\right] \\ &= P(A_1 \cap A_2)^2 = P(A_1 \cap A_2)\,. \end{aligned}$$

Remark 9.18. In fact, if the mapping $P : \mathcal{F} \to \mathcal{P}(H)$ satisfies (a) and (b), then (c) can be replaced by

(c$'$) For all $x, y \in H,\ A \mapsto \langle P(A)x, y\rangle$ is a *complex measure* on $(\Omega, \mathcal{F})$.

To see this, trivially, (a) and (c) implies (c$'$). On the contrary, by part (b), if $\{A_n\} \subset \mathcal{F}$ is pairwise disjoint, then $\{P(A_n)x\}$ is pairwise orthogonal. Thus the series

$$\sum_{n=1}^{\infty} P(A_n)x$$

converges absolutely. By part (c$'$),

$$\langle \sum_{n=1}^{\infty} P(A_n)x, y\rangle = \sum_{n=1}^{\infty} \langle P(A_n)x, y\rangle = \langle P(\bigcup_{n=1}^{\infty} A_n)x, y\rangle$$

for all $y \in H$. Thus

$$P(\bigcup_{n=1}^{\infty} A_n)x = \sum_{n=1}^{\infty} P(A_n)x\,.$$

Since $x \in H$ is arbitrary, (c) follows.

Henceforth, for a projection valued measure P, we denote by $P_{x,y}$ the complex measure $A \mapsto \langle P(A)x, y\rangle$. Then clearly the following holds.

(a) For each $x \in H$, $P_{x,x}$ is a finite positive measure on $(\Omega, \mathcal{F})$; for $A \in \mathcal{F}$, we have $P_{x,x}(A) = \|P(A)x\|^2$. So if $\|x\| = 1$, then $P_{x,x}$ is a probability measure on $(\Omega, \mathcal{F})$.

(b) For each $x, y \in H$, and $A \in \mathcal{F}$, we have

$$|P_{x,y}(A)| \leq \|y\||P_{x,x}(A)|^{1/2}\,.$$

Exercise 9.6. Let P be a projection valued measure on $(\Omega, \mathcal{F})$. Let $A \in \mathcal{F}$. Then the following statements are equivalent.

(a) $P(A) = 0$.

(b) $P_{x,y}(A) = 0$ for all x, y in H.

(c) $|P_{x,y}|(A) = 0$ for all x, y in H, where $|P_{x,y}|$ is the total variance of $P_{x,y}$.

(d) $P_{x,x}(A) = 0$ for all x in H.

The Essential Bounded Functions Next, we will define the integrals of measurable functions on the projection-valued measure space $(\Omega, \mathcal{F}, P)$. To this end, we need some preliminaries.

Let f be a measurable, complex-valued function on $(\Omega, \mathcal{F}, P)$. Then the *essential range* of f is defined to be the set:

$$\text{ess.}\,\text{im}(f) = \{z \in \mathbb{C} : \text{ for all } r > 0 : P(f \in B(z, r)) \neq 0\}\,.$$

In other words $\text{ess.}\,\text{im}(f) = \text{supp}(P \circ f^{-1})$. Thus, one can easily check that:

(a) The complement of $\operatorname{ess.im}(f)$, given by

$$\{z \in \mathbb{C} : \text{ there exists } r > 0, P(f \in B(z,r)) = 0\} ,$$

is open, and since $\mathbb{C}$ is second countable, $P(f \in \operatorname{ess.im}(f)^c) = 0$.

(b) The essential image $\operatorname{ess.im}(f)$ is always closed.

(c) If B is a Borel set in $\mathbb{C}$ disjoint with $\operatorname{ess.im}(f)$, then $P(f \in B) = 0$. This fact characterises the essential image: It is the smallest closed subset of $\mathbb{C}$ with this property.

(d) The essential image cannot be used to distinguish functions that are almost everywhere equal: If $f = g$ holds P-almost everywhere, then $\operatorname{ess.im}(f) = \operatorname{ess.im}(g)$.

(e) If $\operatorname{ess.im}(f) = \{0\}$, then $f = 0$ P-almost everywhere.

(f) The essential image of f the biggest set contained in the closures of $\operatorname{im}(g)$ for all g that are a.e. equal to f:

$$\operatorname{ess.im}(f) = \bigcap_{f=g \text{ a.e.}} \overline{g(\Omega)} \subset \overline{f(\Omega)} .$$

(g) If f is continuous and $\operatorname{supp}(P) = \Omega$, then $\operatorname{ess.im}(f) = \overline{f(\Omega)}$.

We say that f is *essentially bounded* if its essential range is bounded, hence compact. In that case, the largest value of $|\lambda|$, as λ runs through the essential range of f, is called the *essential supremum* $\|f\|_\infty$ of f:

$$\|f\|_\infty \coloneqq \sup_{\lambda \in \operatorname{ess.im}(f)} |\lambda| .$$

Denote by $L^\infty(P) = L^\infty(\Omega, \mathcal{F}, P; \mathbb{C})$ all the essentially bounded complex valued measurable functions on $(\Omega, \mathcal{F}, P)$. We agree that in $L^\infty(P)$ two functions f, g are the same if $\|f - g\|_\infty = 0$, i.e., $f = g$ P-a.e.. Then $L^\infty(P)$ is a commutative C^*-algebra with the essential supremum norm and with the involution given by complex conjugation.

If easy to see that $f \in L^\infty(P)$ is invertible if and only if $0 \notin \operatorname{ess.im}(f)$. As a consequence, the spectrum of f is exactly the essential range of f:

$$\sigma(f) = \operatorname{ess.im}(f)\,.$$

Remark 9.19. Denote by $L^\infty(\Omega, \mathcal{F})$ all the bounded complex valued measurable functions on $(\Omega, \mathcal{F})$. Then $L^\infty(\Omega, \mathcal{F})$ is also a commutative C^*-algebra with the uniform norm (supremum norm)

$$\|f\|_u \coloneqq \sup_{x\in\Omega} |f(x)| \quad \text{for} \quad f \in L^\infty(\Omega, \mathcal{F})\,.$$

and with the involution given by complex conjugation. However, the essential supremum $\|\cdot\|_\infty$ is only a seminorm on $L^\infty(\Omega, \mathcal{F})$. Fixing a projection valued measure P, one sees easily that

$$N = \{f \in L^\infty(\Omega, \mathcal{F}) : \|f\|_\infty = 0\}$$

is an ideal of $L^\infty(\Omega, \mathcal{F})$ which is closed. Then one can check that $L^\infty(\Omega, \mathcal{F})/N$, equipped with the quotient norm, is exactly $L^\infty(P)$. As is usually done in measure theory, the distinction between f and its equivalence class $f + N$ will be ignored.

Integrals For each $f \in L^\infty(P)$, we shall define $\int_\Omega f\,\mathrm{d}P$, the *integral* of f with respect to P, be the bounded linear operator on H satisfies that

$$\langle \int_\Omega f\,\mathrm{d}P\,x, y\rangle = \int_\Omega f\,\mathrm{d}P_{x,y} \quad \text{for all } x, y \in H\,. \tag{9.16}$$

First of all, $\int_\Omega f\,\mathrm{d}P$ is well-defined, in the since that if $f = g$ P-a.e., then $\int_\Omega f\,\mathrm{d}P = \int_\Omega g\,\mathrm{d}P$, since $\int_\Omega f\,\mathrm{d}P_{x,y} = \int_\Omega g\,\mathrm{d}P_{x,y}$ for all x, y in H. Secondly, the integral must be unique if it exists. The existence is provided by the next theorem.

Theorem 9.42. *Let P be a projection valued measure on $(\Omega, \mathcal{F})$. Then for all $f \in L^\infty(P)$, the integral $\int_\Omega f dP \in \mathcal{B}(H)$, satisfying* (9.16)*, does exist. Moreover,*

$$\mathscr{A}_P \coloneqq \overline{\operatorname{span}}\,\{P(A) : A \in \mathcal{F}\} \subset \mathcal{B}(H)\,.$$

is a sub-C^*-algebra of $\mathcal{B}(H)$, and the mapping

$$L^\infty(P) \to \mathscr{A}_P \;;\; f \mapsto \int_\Omega f dP \tag{9.17}$$

is an isometric $*$-isomorphism.

Proof. Step 1. Let $S(\Omega, \mathcal{F})$ be the collection of all the the simple measurable functions on $(\Omega, \mathcal{F})$, as a subspace of $L^\infty(P)$. We shall show for each $\phi \in S(\Omega, \mathcal{F})$, $\int_\Omega \phi \, \mathrm{d}P$ exists, and the mapping

$$\phi \mapsto \int_\Omega \phi \, \mathrm{d}P$$

is a isometric $*$-isomorphism of $S(\Omega, \mathcal{F})$ onto span $\{P(A) : A \in \mathcal{F}\}$.

Suppose $\phi = \sum_{k=1}^n \alpha_k 1_{A_k}$ so that $A_k \cap A_m = \emptyset$ for $k \neq m$ and A_k's are measurable. Then, following the intuition, we define

$$\int_\Omega \phi \, \mathrm{d}P \coloneqq \sum_{k=1}^n \alpha_k P(A_k) \in \operatorname{span}\{P(A) : A \in \mathcal{F}\} \tag{9.18}$$

Thus, clearly for x, y in H,

$$\begin{aligned} \langle \int_\Omega \phi \, \mathrm{d}P \, x, y \rangle &= \langle \sum_{k=1}^n \alpha_k P(A_k) \, x, y \rangle = \sum_{k=1}^n \alpha_k \langle P(A_k) \, x, y \rangle \\ &= \sum_{k=1}^n \alpha_k P_{x,y}(A_k) = \int_\Omega \phi \, \mathrm{d}P_{x,y} \, . \end{aligned}$$

Thus our definition satisfies (9.16). This also show that the definition (9.18) does not depend on the particular representation of ϕ.

- Since $\{A_k\}$ are pairwise disjoint, $\{P(A_k)\}$ are pairwise orthogonal, then

$$\begin{aligned} \| \int_\Omega \phi \, \mathrm{d}P \| &= \| \sum_k \alpha_k P(A_k) \| \\ &= \max\{|\alpha_k| : P(A_k) \neq 0\} = \|\phi\|_\infty \, . \end{aligned} \tag{9.19}$$

- Trivially, for simple function ϕ, ψ an complex numbers α, β we have
$$\int_\Omega \alpha\phi + \beta\psi \,\mathrm{d}P = \alpha \int_\Omega \phi \,\mathrm{d}P + \beta \int_\Omega \psi \,\mathrm{d}P\,.$$
- If ψ is also a simple function and $\psi = \sum_{m=1}^{n'} \beta_m 1_{B_m}$ so that $B_k \cap B_m = \emptyset$ for $k \neq m$ and B_m's are measurable, then $\phi\psi = \sum_{k,m} \alpha_k \beta_m 1_{A_k \cap B_m}$. By (9.18), we have
$$\begin{aligned}\int_\Omega \phi\psi \,\mathrm{d}P &= \sum_{k,m} \alpha_k \beta_m P(A_k \cap B_m) = \sum_{k,m} \alpha_k \beta_m P(A_k) P(B_m) \\ &= \sum_k \alpha_k P(A_k) \ \sum_{k,m} \alpha_k \beta_m P(A_k) P(B_m) = \int_\Omega \phi \,\mathrm{d}P \int_\Omega \psi \,\mathrm{d}P\,.\end{aligned}$$
- Since $\bar{\phi} = \sum_{k=1}^n \overline{\alpha}_k 1_{A_k}$, we have
$$\begin{aligned}\int_\Omega \bar{\phi} \,\mathrm{d}P &= \sum_{k=1}^n \overline{\alpha}_k P(A_k) = \sum_{k=1}^n (\alpha_k P(A_k))^* \\ &= \sum_{k=1}^n (\alpha_k P(A_k))^* = \left(\sum_{k=1}^n \alpha_k P(A_k)\right)^* = (\int_\Omega \phi \,\mathrm{d}P)^*\,.\end{aligned}$$

Step 2. Now suppose $f \in L^\infty(P)$. Then there is a sequence of simple measurable functions ϕ_n that converges to f uniformly, i.e., in the norm of $L^\infty(P)$. By (9.19) the corresponding operators $\int_\Omega \phi_n \,\mathrm{d}P$ form a Cauchy sequence in the Banach space $\mathcal{B}(H)$. Set
$$\int_\Omega f \,\mathrm{d}P \coloneqq \lim_{n\to\infty} \int_\Omega \phi_n \,\mathrm{d}P \in \mathscr{A}_P\,.$$
Note that this definition does not depend on the particular choice of $\{\phi_n\}$. To check (9.16), given x, y in H, we have
$$\begin{aligned}\langle \int_\Omega f \,\mathrm{d}P\, x, y\rangle &= \lim_{n\to\infty} \langle \int_\Omega \phi_n \,\mathrm{d}P\, x, y\rangle \\ &= \lim_{n\to\infty} \int_\Omega \phi_n \,\mathrm{d}P_{x,y} = \int_\Omega f \,\mathrm{d}P_{x,y}\,.\end{aligned}$$
We are going to show that the mapping (9.17) is a isometric $*$-isomorphism of $L^\infty(P)$ onto $\mathscr{A}_P$.

Obviously (9.19) leads to

$$\| \int_\Omega f \,\mathrm{d}P \| = \lim_{n\to\infty} \| \int_\Omega \phi_n \,\mathrm{d}P \| = \lim_{n\to\infty} \|\phi_n\|_\infty = \|f\|_\infty \,.$$

Thus $f \mapsto \int_\Omega f \,\mathrm{d}P$ is a isometric. It's easy to see that the mapping (9.17) is a $*$-homomorphism. Then the rangle of the mapping (9.17) is a closed subspace of $\mathscr{A}_P$ containing $\operatorname{span}\{P(A) : A \in \mathcal{F}\}$. Thus the mapping (9.17) is surjective. $\square$

Corollary 9.43. *If $T \in \mathcal{B}(H)$ commutes with $\{P(A) : A \in \mathcal{F}\}$, then T commutes with $\{\int_\Omega f \,\mathrm{d}P : f \in L^\infty(P)\}$.*

Corollary 9.44. *Let P be a projection valued measure on $(\Omega, \mathcal{F})$. Let $f \in L^\infty(P)$. Then the following statements holds.*

(a) *For each $x \in H$,*

$$\Big\|\big(\int_\Omega f \,\mathrm{d}P\big)\, x\Big\|^2 = \int_\Omega |f|^2 \,\mathrm{d}P_{x,x}\,;$$

and hence

$$\|f\|_\infty = \Big\|\big(\int_\Omega f \,\mathrm{d}P\big)\, x\Big\| = \sup_{\|x\|\le 1} \Big[\int_\Omega |f|^2 \,\mathrm{d}P_{x,x}\Big]^{1/2} \,.$$

(b) *If $\{f_n\}$ is a bounded sequence in $L^\infty(P)$, and $f_n(\omega) \to f(\omega)$ P-a.e., we have $f \in L^\infty(P)$, and*

$$\int f \,\mathrm{d}P = s\text{-}\lim_{n\to\infty} \int f_n \,\mathrm{d}P \,.$$

Proof. To show (a), note that

$$\begin{aligned}\Big\|\big(\int_\Omega f \,\mathrm{d}P\big)\, x\Big\|^2 &= \langle \int_\Omega f \,\mathrm{d}P\, x, \int_\Omega f \,\mathrm{d}P, x\rangle = \langle \int_\Omega \bar{f} \,\mathrm{d}P \int_\Omega f \,\mathrm{d}P\, x, x\rangle \\ &= \langle \int_\Omega |f|^2 \,\mathrm{d}P\, x, x\rangle = \int_\Omega |f|^2 \,\mathrm{d}P_{x,x}\,.\end{aligned}$$

(b) is a direct consequence of (a) and the dominated convergence theorem. $\square$

Spectrum of the Integrals In this paragraph we will discuss the spectrum of the integrals and its classification.

Theorem 9.45. *Let P be a projection valued measure on $(\Omega, \mathcal{F})$. Let $f \in L^\infty(P)$. Then*

$$\sigma\left(\int_\Omega f \, \mathrm{d}P\right) = \operatorname{ess.im}(f) \subset \overline{f(\Omega)}\,.$$

In particular, if f is continuous and $\operatorname{supp}(P) = \Omega$*, then*

$$\sigma\left(\int_\Omega f \, \mathrm{d}P\right) = \overline{f(\Omega)}\,.$$

Proof. Note that since $\mathscr{A}_P$ is a commutative sub-C^*-algebra of $\mathcal{B}(H)$, by Theorem 9.38,

$$\sigma\left(\int_\Omega f \, \mathrm{d}P\right) = \sigma_{\mathscr{A}_P}\left(\int_\Omega f \, \mathrm{d}P\right)\,.$$

Since the mapping (9.17) is a $*$-isomorphism, we have

$$\sigma\left(\int_\Omega f \, \mathrm{d}P\right) = \sigma_{L^\infty(P)}(f) = \operatorname{ess.im}(f) \subset \overline{f(\Omega)}\,.$$

If f is continuous and $\operatorname{supp}(P) = \Omega$, then $\operatorname{ess.im}(f) = \overline{f(\Omega)}$ and the desired result holds. □

Lemma 9.46. *Let P be a projection valued measure on $(\Omega, \mathcal{F})$. Let $f \in L^\infty(P)$. Then*

$$\operatorname{Ker}(\int_\Omega f \, \mathrm{d}P) = \operatorname{Ran}(P(f=0))\,.$$

Proof. Firstly, for $x \in \operatorname{Ran}(P(f=0))$ we have $P(f=0)x = x$, then

$$(\int_\Omega f \, \mathrm{d}P)x = (\int_\Omega f \, \mathrm{d}P)P(f=0)x = \left(\int_\Omega f 1_{\{f=0\}} \, \mathrm{d}P\right) x = 0.$$

On the other hand, if $x \in \operatorname{Ker}(\int_\Omega f \, \mathrm{d}P)$, then

$$\left\|(\int_\Omega f \, \mathrm{d}P)x\right\|^2 = \int_\Omega |f|^2 \, \mathrm{d}P_{x,x} = 0\,.$$

Thus $P_{x,x}(|f| > 0) = 0$, i.e, $P(|f| > 0)x = 0$. Therefore we have

$$x = P(|f| > 0)x + P(f = 0)x = P(|f| > 0)x \in \mathrm{Ran}(P(f = 0)) .$$

We are done. $\square$

Theorem 9.47. *Let P be a projection valued measure on $(\Omega, \mathcal{F})$. Let $f \in L^\infty(P)$. Then*

$$\sigma\left(\int_\Omega f \,\mathrm{d}P\right) = \mathrm{ess.\,im}(f) \ ; \ \sigma_r\left(\int_\Omega f \,\mathrm{d}P\right) = \emptyset ;$$
$$\sigma_p\left(\int_\Omega f \,\mathrm{d}P\right) = \{\lambda \in \mathrm{ess.\,im}(f) : P(f = \lambda) \neq 0\} \ ;$$
$$\sigma_c\left(\int_\Omega f \,\mathrm{d}P\right) = \{\lambda \in \mathrm{ess.\,im}(f) : P(f = \lambda) = 0\} \ .$$

Proof. It has been shown in Theorem 9.45 that $\sigma\left(\int_\Omega f \,\mathrm{d}P\right) = \mathrm{ess.\,im}(f)$.

Since $\int_\Omega f \,\mathrm{d}P$ is normal, we have $\sigma_r\left(\int_\Omega f \,\mathrm{d}P\right) = \emptyset$ by Theorem 8.30.

It remains to show that $\lambda \in \sigma_p(\int_\Omega f \,\mathrm{d}P)$ if and only if $\lambda \in \mathrm{ess.\,im}(f)$ and $P(f = \lambda) \neq 0$. Observe that by Lemma 9.46,

$$\mathrm{Ker}(\lambda I - \int_\Omega f \,\mathrm{d}P) = \mathrm{Ker}(\int_\Omega \lambda - f \,\mathrm{d}P) = P(f = \lambda)$$

then the desired result follows. $\square$

9.7.2 The Spectral Theorem

The principal assertion of the spectral theorem is that every bounded normal operator N on a Hilbert space induces (in a canonical way) a projection valued measure E on the Borel subsets of its spectrum $\sigma(N)$ and that N can be reconstructed from E by the integral of the type discussed above. A large part of the theory of normal operators depends on this fact.

It should perhaps be stated explicitly that the spectrum $\sigma(T)$ of an operator $T \in \mathcal{B}(H)$ will always refer to the full algebra $\mathcal{B}(H)$. In other words, $\lambda \in \sigma(T)$ if and only if $\lambda I - T$ has no inverse in $\mathcal{B}(H)$. Sometimes we shall also be concerned with the sub-C^*-algebra $\mathscr{A}$ of $\mathcal{B}(H)$. Suppose that $N \in \mathscr{A}$, then by Theorem 9.38, $\sigma_{\mathscr{A}}(N) = \sigma(N)$. Thus N has the same spectrum relative to all sub-C^*-algebras in $\mathcal{B}(\mathrm{H})$ that contain N.

Recall that $\mathscr{A}$ is a commutative sub-C^*-algebra of $\mathcal{B}(H)$ iff $\mathscr{A}$ is a closed normal subalgebra of $\mathcal{B}(H)$ containing the identity operator I.

Theorem 9.48. *Let $\mathscr{A}$ be a commutative sub-C^*-algebra of $\mathcal{B}(H)$. Let Δ be all the nonzero complex homomorphisms of $\mathscr{A}$. Then the inverse of the Gelfand representation $\Gamma^{-1} : \hat{\mathscr{A}} = C(\Delta) \to \mathscr{A}$ has a specific formular. In fact, there exists a unique projection valued measure P on $(\Delta, \mathcal{B}(\Delta))$, so that*

$$N = \int_\Delta \hat{N}\,\mathrm{d}P\,, \quad \text{for all } N \in \mathscr{A}\,, \tag{9.20}$$

where $\hat{N}$ is the Gelfand transform of N; and $\operatorname{supp}(P) = \Delta$.

Recall that by the Gelfand-Naimark theorem, the Gelfand representation Γ is an isometric $*$-isomorphism of $\mathscr{A}$ onto $\hat{\mathscr{A}} = C(\Delta)$. Thus $\Gamma^{-1} : \hat{\mathscr{A}} = C(\Delta) \to \mathscr{A}$ is well-defined.

Proof. Recall that (9.20) is an abbreviation for

$$\langle Nx, y\rangle = \int_\Delta \hat{N}\,\mathrm{d}P_{x,y} \ \text{ for all } \ x, y \in H\,.$$

Step 1. Our first task is to find the complex measures $\{P_{x,y} : x, y \in H\}$. Fix now $x, y \in H$, define

$$\ell(\hat{N}) = \langle Nx, y\rangle \ \text{ for all } \ \hat{N} \in \hat{\mathscr{A}} = C(\Delta)\,.$$

Then

$$|\ell(\hat{N})| \le \|N\|\|x\|\|y\| = \|\hat{N}\|\|x\|\|y\|$$

since the $\Gamma : \mathscr{A} \to C(\Delta)$ is isometric. Thus ℓ is a bounded continuous linear functional on $C(\Delta)$. By the Riesz-Markov-Kakutani representation theorem, there exists a unique complex Radon measure, namely $\mu_{x,y}$, on $(\Delta, \mathcal{B}(\Delta))$, so that

$$|\ell(\hat{N})| = \langle Nx, y\rangle = \int_\Delta \hat{N}\,\mathrm{d}\mu_{x,y} \tag{9.21}$$

for all $\hat{N} \in C(\Delta)$; and the total variance

$$\|\mu_{x,y}\|_{TV} = \|\ell\| \leq \|x\|\|y\|\,.$$

It suffices now to show that there exists a projection valued measure P on $(\Delta, \mathcal{B}(\Delta))$ so that

$$\mu_{x,y}(A) = \langle P(A)x, y\rangle \tag{9.22}$$

for all A in $\mathcal{B}(\Delta)$ and x, y in H and $\operatorname{supp}(P) = \Delta$.

Step 2. We shall show that for each fixed $f \in L^\infty(\Delta, \mathcal{B}(\Delta))$, there exists a self-adjoint operator $P(f)$ so that $\int_\Delta f\,\mathrm{d}\mu_{x,y} = \langle P(f)x, y\rangle$. Then $P(A) \coloneqq P(1_A)$ satisfies (9.22). By Theorem 8.8, we have only to show the mapping

$$(x, y) \mapsto \int_\Delta f\,\mathrm{d}\mu_{x,y}$$

is a bounded sesquilinear hermitian form on H. For given α, β in $\mathbb{C}$ and x_1, x_2, y in H, since

$$\begin{aligned}\int_\Delta \hat{N}\,\mathrm{d}\mu_{\alpha x_1+\beta x_2,y} &= \langle N(\alpha x_1 + \beta x_2), y\rangle \\ &= \alpha\langle Nx_1, y\rangle + \beta\langle Nx_2, y\rangle \\ &= \int_\Delta \hat{N}\,\mathrm{d}(\alpha\mu_{x_1,y} + \beta\mu_{x_2,y})\end{aligned}$$

for all $\hat{N} \in C(\Delta)$, we get

$$\mu_{\alpha x_1+\beta x_2,y} = \alpha\mu_{x_1,y} + \beta\mu_{x_2,y}\,.$$

Similarly, for given x, y in H, since

$$\begin{aligned}\int_\Delta \hat{N}\,\mathrm{d}\mu_{x,y} &= \langle x, N^*y\rangle = \overline{\langle N^*y, x\rangle} \\ &= \int_\Delta \widehat{N^*}\,\mathrm{d}\mu_{y,x} = \int_\Delta \overline{\hat{N}}\,\mathrm{d}\mu_{y,x} = \overline{\int_\Delta \hat{N}\,\mathrm{d}\mu_{y,x}}\end{aligned}$$

for all $\hat{N} \in C(\Delta)$, this implies that $\mu_{x,y} = \overline{\mu_{y,x}}$. Moreover,

$$\left\| \int_\Delta f \,\mathrm{d}\mu_{x,y} \right\| \leq \|f\|_u \|\mu_{x,y}\|_{TV} \leq \|f\|_u \|x\| \|y\| \,;$$

so we get the existence of $\|P(f)\|$, with

$$\|P(f)\| \leq \|f\|_u \,.$$

Besides, note that by (9.21)

$$P(\hat{N}) = N \text{ for all } \hat{N} \in C(\Delta)\,.$$

Step 3. Trivially $P(\emptyset) = 0$. We also have $P(\Delta) = I$, since for all x, y in H there holds

$$\langle P(\Delta)x, y\rangle = \int_\Delta 1 \,\mathrm{d}\mu_{x,y} = \int_\Delta \hat{I} \,\mathrm{d}\mu_{x,y} = \langle x, y\rangle \,.$$

In order that P is a projection valued measure, we have only to show that

$$P(A \cap B) = P(A)P(B) \quad \text{for all} \quad A, B \in \mathcal{B}(\Delta)\,. \tag{9.23}$$

Indeed, this implies $P(A) = P(A)^2$, then combine this with that $P(A)$ is self-adjoint, we get that $P(A)$ is a (orthogonal) projection. Set

$$P : \mathcal{B}(\Delta) \to \mathcal{P}(H) \;;\; A \mapsto P(A)\,,$$

then the mapping P satisfies part (a), (b) of Definition 9.10. Note that for all x, y in H, $\langle P(A)x, y\rangle = \mu_{x,y}(A)$ is a complex measure on $\mathcal{B}(\Delta)$. Thus (c$'$) is true for P. So P is a projection valued measure with $P_{x,y} = \mu_{x,y}$.

Thus for $f \in L^\infty(\Delta, \mathcal{B}(\Delta))$,

$$\langle P(f)x, y\rangle = \int_\Delta f \,\mathrm{d}\mu_{x,y} = \int_\Delta f \,\mathrm{d}P_{x,y} = \langle \int_\Delta f \,\mathrm{d}P \, x, y\rangle \,;$$

and hence

$$P(f) = \int_\Delta f \,\mathrm{d}P\,.$$

Let $f = \hat{N} \in C(\Delta)$, since $P(\hat{N}) = N$, we get

$$N = \int_\Delta \hat{N} \, \mathrm{d}P \, .$$

Step 4. We enhance (9.23) to

$$P(fg) = P(f)P(g) \quad \text{for all } f, g \in L^\infty(\Delta, \mathcal{B}(\Delta)) \, .$$

To this end, observe that

$$P(\hat{N}\hat{M}) = P(\widehat{NM}) = NM = P(\hat{N})P(\hat{M})$$

for all $\hat{N}$, $\hat{M}$ in $C(\Delta)$. For fixed x, y in H, thus

$$\int \hat{N}\hat{M} \, \mathrm{d}\mu_{x,y} = \int \hat{N} \, \mathrm{d}\mu_{Mx,y} \, .$$

By Lusin's theorem, and the dominated convergence theorem, we get

$$\begin{aligned} \int f\hat{M} \, \mathrm{d}\mu_{x,y} &= \int f \, \mathrm{d}\mu_{Mx,y} = \langle P(f)Mx, y \rangle \\ &= \langle Mx, P(f)^* y \rangle = \int \hat{M} \, \mathrm{d}\mu_{x,P(f)^* y} \, , \end{aligned}$$

for all $f \in L^\infty(\Delta, \mathcal{B}(\Delta))$. Again, by Lusin's theorem and the dominated convergence theorem,

$$\int fg \, \mathrm{d}\mu_{x,y} = \int g \, \mathrm{d}\mu_{x,P(f)^* y} = \langle P(g)x, P(f)^* y \rangle = \langle P(f)P(g)x, y \rangle \, ,$$

that is

$$\langle P(fg)x, y \rangle = \langle P(f)P(g)x, y \rangle \, .$$

Since x, y is arbitrary, we get $P(fg) = P(f)P(g)$ as desired.

Step 5. Finally, we show that $\operatorname{supp}(P) = \Omega$. To this end, we have to show that for every open sets U in Δ, $P(U) \neq 0$. Suppose for contradiction that there exists an open set U in Δ with $P(U) = 0$. Then, by Urysohn's lemma, for fixed $\phi \in U$, the exists $\hat{N} \in C(\Delta)$ with $\operatorname{supp}(\hat{N}) \subset U$ and $\hat{N}(\phi) = \phi(N) = 1$. However, then

$$N = \int_\Delta \hat{N} \, \mathrm{d}P = \int_\Delta \hat{N} 1_U \, \mathrm{d}P = 0$$

and hence $\phi(N) = 0$, which is a contradiction! $\square$

We now specialize this theorem to a single operator.

Theorem 9.49 (The Spectrum Theorem). *If $N \in \mathcal{B}(H)$ and N is normal, then there exists a unique projection-valued measure E on $(\sigma(N), \mathcal{B}(\sigma(N)))$ so that* $\operatorname{supp}(E) = \sigma(N)$ *and*

$$N = \int_{\sigma(N)} \lambda \, E(\mathrm{d}\lambda)\,.$$

We shall refer to this E as the spectral decomposition of N.

Proof. Let $\mathscr{A}_N \subset \mathcal{B}(H)$ be the smallest sub-C^*-algebra that contains N; i.e.,

$$\mathscr{A}_N = \operatorname{cl}\{p(N, N^*) : p \text{ is a polynomial in two variables}\}\,.$$

Clearly, $\mathscr{A}_N$ is a commutative sub-C^*-algebra of $\mathcal{B}(H)$. Let Δ be all the nonzero complex homomorphisms of $\mathscr{A}_N$. Then by Theorem 9.48, there is a unique projection valued measure P on $(\Delta, \mathcal{B}(\Delta))$ so that $\operatorname{supp}(P) = \Delta$, and

$$N = \int_\Delta \hat{N} \, \mathrm{d}P\,.$$

By (9.24), $\hat{N}$ is a homeomorphism of Δ onto $\sigma(N)$. Let $E = P \circ \hat{N}^{-1}$ be the image measure on $(\sigma(N), \mathcal{B}(\sigma(N)))$, then we get $\operatorname{supp}(E) = \sigma(N)$, and

$$N = \int_\Delta \lambda E(\mathrm{d}\lambda)\,.$$

In fact E is unique is guaranteed by the fact that P is unique. To give a direct proof, note that for any polynomial in two variables p,

$$p(N, N^*) = \int_\Delta p(\lambda, \bar{\lambda}) E(\mathrm{d}\lambda)\,. \tag{9.24}$$

By the Stone-Weierstrass theorem, these polynomials form a dense subalgebra in $C(\sigma(T))$. Thus for each given x, y, $E_{x,y}$ are therefore uniquely determined by the integrals (9.24), hence by N. $\square$

Remark 9.20. For normal operator $N \in \mathcal{B}(H)$, we sometimes assume it's spectrum measure E is defined on the complex plane $\mathbb{C}$ equipped with the Borel algebra by setting $E(A) = E(A \cap \sigma(N))$ for all Borel sets A in $\mathbb{C}$. If N is self-adjoint, as we know then $\sigma(N) \subset \mathbb{R}$, so we shall think that E is defined on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and if N is unitary then $\sigma(N) \subset S^1$, we can think that E is defined on $(S^1, \mathcal{B}(S^1)$.

Bounded Measurable Functional Calculus If E is the spectral decomposition of a normal operator $N \in \mathcal{B}(H)$, and if f is a bounded Borel function on $\sigma(N)$, it is customary to define

$$\tilde{f}(N) := \int_{\sigma(N)} f \, \mathrm{d}E \,.$$

Using this notation, part of the content of Theorems 9.42 to 9.49 can be summarized as follows:

Exercise 9.7 (Bounded Measurable Functional Calculus). Let E be the spectral decomposition of a normal operator N in $\mathcal{B}(H)$. Recall that

$$\mathscr{A}_E := \overline{\operatorname{span}} \left\{ E(A) : A \in \mathcal{B}(\sigma(N)) \right\} \,.$$

Then, the mapping

$$L^\infty(E) \to \mathscr{A}_E \ ; \ f \mapsto \tilde{f}(N)$$

is an isometric $*$-isomorphism satisfying the following statements.

(a) (Normalization) If $f(\lambda) = \lambda$ for all $\lambda \in \sigma(N)$ then $\tilde{f}(N) = N$.

(b) (Commutative) If $T \in \mathcal{B}(H)$ satisfies $NT = TN$, then $\tilde{f}(N)T = T\tilde{f}(N)$ for all $f \in L^\infty(E)$.

(c) (Convergence) If $f_n \in L^\infty(E)$ be a bounded sequence in $L^\infty(E)$, and $f_n \to f$ P-a.e., then $f \in L^\infty(P)$ and

$$\tilde{f}(N) = \text{s-}\lim_{n\to\infty} \tilde{f}_n(N) \,.$$

(d) (Image) $\mathscr{A}_E$ is the closure of $\mathscr{A}_N$ relative to the strong operator topology, where $\mathscr{A}_N$ is defined in (9.14).

(e) (Eigenvalues) If $\lambda \in \sigma_p(N)$ and $x \in H$ satisfy $Nx = \lambda x$ then

$$\tilde{f}(N)x = f(\lambda)x \quad \text{for all } f \in L^\infty(E)\,.$$

(f) (Spectrum) For every $f \in L^\infty(E)$ the operator $\tilde{f}(N)$ is normal and

$$\sigma(\tilde{f}(N)) = \operatorname{ess.im}(f) \subset \overline{f(\sigma(N))}\,.$$

In particular, if f is continuous, then $\sigma(\tilde{f}(N)) = f(\sigma(N))$.

(g) (Positive) If $f \in L^\infty(E)$ and $f \geq 0$ then $\tilde{f}(N) = \tilde{f}(N)^* \geq 0$.

(h) (Composition) If $f \in C(\sigma(N))$ and $g \in L^\infty(f(\sigma(N)))$ then

$$\widetilde{(g \circ f)}(N) = \tilde{g}(\tilde{f}(N))\,.$$

9.7.3 Discrete Spectrum and Essential Spectrum

If $N \in \mathcal{B}(H)$ is normal, its eigenvalues bear a simple relation to its spectral decomposition E. This will be derived from the fact that for each $f \in L^\infty(E)$,

$$\operatorname{Ker}(f(N)) = \operatorname{Ran}\left(E\left(f = 0\right)\right)\,.$$

Clearly this is a restatement of Lemma 9.46.

Theorem 9.50. *Suppose $N \in \mathcal{B}(H)$ is a normal operator with spectral decomposition E.*

(a) $E(\{\lambda\})$ is the projection of H onto $\operatorname{Ker}(\lambda I - N)$, and hence λ is an eigenvalue of N if and only if $E(\{\lambda\}) \neq 0$.

(b) Every isolated point of $\sigma(N)$ is an eigenvalue of N.

(c) *If* $\sigma(N) = \{\lambda_1, \lambda_2, \lambda_3, \ldots\}$ *is at most countable, then* $E(\lambda_k)$ *is the projection of* H *onto* $\mathrm{Ker}(\lambda_k I - N)$ *and*

$$\tilde{f}(N) = \sum_k f(\lambda_k) E(\{\lambda_k\}) \quad \textit{for all} \quad f \in L^\infty(E)\,.$$

Statements (a) and (b) explain the term *point spectrum* of N for the set of all eigenvalues of N.

Definition 9.11. For a normal operator $N \in \mathcal{B}(H)$, the *discrete spectrum* $\sigma_d(N)$ of N is the set of all eigenvalues of N of finite multiplicities which are isolated points of the spectrum $\sigma(N)$. The complement set $\sigma_{\text{ess}}(N) := \sigma(N) \backslash \sigma_d(N)$ is called the *essential spectrum* of N.

By the Theorem 9.50 an isolated point of the spectrum $\sigma(N)$ is always an eigenvalue of N. Therefore, a number belongs to $\sigma_{\text{ess}}(N)$ of N if and only if it is an accumulation point of $\sigma(N)$ or an eigenvalue of infinite multiplicity.

Theorem 9.51. *Let* $N \in \mathcal{B}(H)$ *be a normal operator with spectrum decomposition* E. *Then* $\lambda_0 \in \sigma_d(N)$ *if and only if there is an open neighborhood* U *of* λ_0*,*

$$\dim(\mathrm{Ran}E(U)) < \infty\,;$$

and $\lambda_0 \in \sigma_{ess}(N)$ *if and only if for each open neighborhood* U *of* λ_0*,*

$$\dim(\mathrm{Ran}E(U)) = \infty\,.$$

Proof. It suffices to show that $\lambda_0 \in \sigma_d(N)$ if and only if there is an open neighborhood U of λ_0,

$$\dim(\mathrm{Ran}E(U)) < \infty\,;$$

The necessity is trivial, since if $\lambda_0 \in \sigma_d(N)$, then take an open neighborhood U of λ_0 so that $U \cap \sigma(N) = \{\lambda_0\}$. Thanks to $\mathrm{supp}(E) = \sigma(N)$, we have $E(U) = E(\{\lambda_0\})$. By part (a) of Theorem 9.50, then the desired result follows.

To prove the sufficiency, assume that $\lambda_0 \in \sigma_d(N)$ and there exists an open neighborhood U of λ_0 so that $\dim(\mathrm{Ran}E(U)) < \infty$. If λ_0 is not an isolated point of $\sigma(N)$, then there is a sequence $\{\lambda_n\}$ of distinct spectral points of N so that $\lambda_n \to \lambda_0$. We get take a sequence $\{U_n\}$ of pairwise disjoint open sets so that $\lambda_n \in U_n \subset U$. Since $\mathrm{supp}(E) = \sigma(N)$, $E(U_n) \neq 0$. Thus

$$\dim R(E(U)) \geq \sum_n \dim R(E(U_n)) = \infty$$

which is absurd. Thus λ_0 is an isolated point of $\sigma(N)$ and hence $E(U) = E(\{\lambda_0\})$. By Theorem 9.50, then $\dim\mathrm{Ker}(\lambda_0 - I) = \dim\mathrm{Ker}(E(\{\lambda_0\})) < \infty$ follows as desired. □

9.7.4 Applications of Functional Calculus

The following proof, containing our first application of this symbolic calculus, generalizes Theorem 9.52.

Theorem 9.52. *Let $N \in \mathcal{B}(H)$ be normal. Then*

$$\|N\| = \sup_{\|x\|=1} |\langle Nx, x\rangle| \,.$$

Proof. Choose $\epsilon > 0$. It is clearly enough to show that

$$|\langle Nx_0, x_0\rangle| > \|N\| - \epsilon$$

for some $x_0 \in H$ with $\|x_0\| = 1$. Indeed, since $\|N\| = r_\sigma(N)$, there is $\lambda_0 \in \sigma(N)$ so that $|\lambda_0| = \|N\|$. Then since $\lambda_0 \in \mathrm{supp}(E)$, $E(B(\lambda_0, \epsilon)) \neq 0$. Take $x_0 \in \mathrm{Ran}(E(B(\lambda_0, \epsilon)))$ and $\|x_0\| = 1$, then

$$\begin{aligned}
\|Nx_0 - \lambda_0 x_0\|^2 &= \int_{\sigma(N)} |\lambda - \lambda_0|^2 E_{x_0,x_0}(\mathrm{d}\lambda) \\
&= \int_{\sigma(N)} |\lambda - \lambda_0|^2 1_{B(\lambda_0,\epsilon)} E_{x_0,x_0}(\mathrm{d}\lambda) \leq \epsilon^2 \,;
\end{aligned}$$

and hence

$$|\langle Nx_0, x_0\rangle - \lambda_0| \leq \epsilon \,.$$

Thus $|\langle Nx_0, x_0\rangle| > |\lambda_0| - \epsilon = \|N\| - \epsilon$ as desired. □

Remark 9.21. To see that normality is needed here, let T be the linear operator on $\mathbb{C}^2$ (with basis $\{e_1, e_2\}$) given by $Te_1 = 0, Te_2 = e_1$. It has $\|T\| = 1$, but $|(Tx, x)| \leq \frac{1}{2}$ if $\|x\| \leq 1$.

We give a application of the functional calculus for normal operators.

Theorem 9.53. *Let N be a normal operator on H.*

(a) *N is self-adjoint if and only if $\sigma(N) \subset \mathbb{R}$.*

(b) *N is non-negative (that is N is self-adjoint and $\langle Nx, x\rangle \geq 0$ for all $x \in H$) if and only if $\sigma(N) \subset [0, \infty)$, i.e., $N \geq 0$.*

(c) *N is unitary if and only if $\sigma(N) \subset S^1$.*

Proof. We show now part (a). If N is self-adjoint, then by Example 6.8 we get $\sigma(N) \subset \mathbb{R}$. If $N \in \mathcal{B}(H)$ is normal and $\sigma(N) \subset \mathbb{R}$, using the continuous functional calculus, let $f(z) = \bar{z}$ for $z \in \sigma(N)$, then $N^* = \tilde{f}(N)$. Since $\sigma(N) \subset \mathbb{R}$, so that $f(z) = z$, we get $N^* = N$.

We show now part (b). If N is nonnegative, then by the same method in Example 6.8, or by Theorem 8.35, we get $(-\infty, 0) \subset \varrho(N)$. Conversely, if $\sigma(N) \subset [0, \infty)$, by part (a) N is self-adjoint. By Theorem 8.35 or by the spectrum decomposition theorem, $\langle Nx, x\rangle \geq 0$ for all $x \in H$.

If N is unitary, then by Theorem 8.19, $\sigma(N) \subset S^1$. On the contrary if $\sigma(N) \subset S^1$, then

$$NN^* = N^*N = \int_{\sigma(N)} \lambda E(\mathrm{d}\lambda) \int_{\sigma(N)} \overline{\lambda} E(\mathrm{d}\lambda) = \int_{\sigma(N)} |\lambda|^2 E(\mathrm{d}\lambda) = I\,.$$

Thus N is unitary. $\square$

Square Roots

Theorem 9.54. *Every nonnegative $T \in \mathcal{B}(H)$ has a unique nonnegative square root $\sqrt{T} \in \mathcal{B}(H)$. If T is invertible in $\mathcal{B}(H)$, so is $\sqrt{T}$.*

Proof. Let $f(z) = \sqrt{z}$ for $z \in \sigma(T)$. By the continuous functional calculus, we denote $\tilde{f}(T)$ by $\sqrt{T}$. Then $\sqrt{T} \in \mathscr{A}_T \subset \mathcal{B}(H)$ is self-adjoint with $\sigma(\sqrt{T}) \subset [0, \infty)$ and $\sqrt{T}\sqrt{T} = T$. (Recall that for a normal operator N, $\mathscr{A}_N$ is the smallest sub-C^*-algebra of $\mathcal{B}(H)$ containing N given by (9.14).) Clearly if $T = S^2$ is invertible, then $\operatorname{Ker}(S) = \{0\}$ and $\operatorname{Ran}(S) = H$ and hence S is invertible.

If $S \in \mathcal{B}(H)$ is also a nonnegative square root of T. Then $T = S^2$ and hence $\mathscr{A}_T \subset \mathscr{A}_S$. Thus there is some $g \in C(\sigma(S))$ is nonnegative so that

$$\sqrt{T} = \int_{\sigma(S)} g(\lambda) E^S(\mathrm{d}\lambda)$$

Then

$$T = \int_{\sigma(S)} g(\lambda)^2 E^S(\mathrm{d}\lambda) = \int_{\sigma(S)} \lambda^2 E^S(\mathrm{d}\lambda)$$

Since the functional calculus is a isometric $*$-isomorphism, we get $g(\lambda)^2 = \lambda$ a.e.-E^S. Thus $g(\lambda) = \lambda$ a.e.-E^S and $\sqrt{T} = S$. □

Proposition 9.55. *If $T \in \mathcal{B}(H)$, then the nonnegative square root of T^*T is the only nonnegative operator $S \in \mathcal{B}(H)$ that satisfies $\|Sx\| = \|Tx\|$ for ever $x \in H$.*

Proof. Note first that T^*T is self-adjoint and

$$\langle T^*Tx, x\rangle = \langle Tx, Tx\rangle = \|Tx\|^2 \geq 0$$

so that $T^*T \geq 0$. (In the more abstract setting of part (d) of Proposition 9.37 this was much harder to prove!) Next, if $S \in \mathcal{B}(H)$ and $S = S^*$, then

$$\langle S^2x, x\rangle = \|Sx\|^2 \quad (x \in H)$$

It follows that $\|Sx\| = \|Tx\|$ for every $x \in H$ if and only if

$$\langle (S^2 - T^*T)x, x\rangle = 0 \text{ for all } x \in H\,.$$

By Corollary 8.12 the desired result follows. □

Polar Decomposition The fact that every complex number λ can be factored in the form $\lambda = \alpha|\lambda|$, where $|\alpha| = 1$, suggests the problem of trying to factor $T \in \mathcal{B}(H)$ in the form $T = US$, with U unitary and $S \geq 0$. When this is possible, we call US a *polar decomposition* of T.

Note that U, being unitary, is an isometry. Proposition 9.55 shows therefore that S is uniquely determined by T.

Theorem 9.56. *Let $T \in \mathcal{B}(H)$. Then the following holds.*

(a) If $T \in \mathcal{B}(H)$ is invertible, then T has a unique polar decomposition $T = US$

(b) If $T \in \mathcal{B}(H)$ is normal, then T has a polar decomposition $T = US$ in which U and S commute with each other and with T.

Proof. If T is invertible, so are T^* and T^*T, and the nonnegative square root S of T^*T is also invertible. Put $U = TS^{-1}$. Then U is invertible, and

$$U^*U = S^{-1}T^*TS^{-1} = S^{-1}S^2S^{-1} = I$$

so that U is unitary. Since S is invertible, it is obvious that TS^{-1} is the only possible choice for U.

If $T \in \mathcal{B}(H)$ is normal, Put $s(\lambda) = |\lambda|, u(\lambda) = \lambda/|\lambda|$ if $\lambda \neq 0, u(0) = 1$. Then s and u are bounded Borel functions on $\sigma(T)$. Put $S = s(T), U = u(T)$. Since $s \geq 0$, we have $S \geq 0$. Since $u\bar{u} = 1, UU^* = U^*U = I$. Since $\lambda = u(\lambda)s(\lambda)$, the relation $T = US$ follows from the functional calculus. □

Remark 9.22. In (a), no two of T, U, S need to commute. For example,

$$\begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

Remark 9.23. It is NOT true that every $T \in \mathcal{B}(H)$ has a polar decomposition. However, if S is the positive square root of T^*T, then $\|Sx\| = \|Tx\|$ for every $x \in H$; hence $Tx = Ty$ if $Sx = Sy$ by linearity. The formula

$$VSx = Tx$$

defines a linear isometry V of $\operatorname{Ran}(S)$ onto $\operatorname{Ran}(T)$, which has a continuous extension to a linear isometry of the closure of $\operatorname{Ran}(S)$ onto the closure of $\operatorname{Ran}(T)$. If there is a linear isometry of $\operatorname{Ran}(S)^\perp$ onto $\operatorname{Ran}(T)^\perp$, then V can be extended to a unitary operator on H, and then T has a polar decomposition. This always happens when $\dim H < \infty$, since $\operatorname{Ran}(S)$ and $\operatorname{Ran}(T)$ have then the same codimension.

If V is extended to a member of $\mathcal{B}(H)$ by defining $Vy = 0$ for all $y \in \operatorname{Ran}(S)^\perp$, then V is called a partial isometry.

Every $T \in \mathcal{B}(H)$ thus has a factorization $T = VS$ in which S is positive and V is a partial isometry.

In combination with Exercise 9.5 the polar decomposition leads to an interesting result concerning similarity of normal operators. The following theorem thus asserts that similar normal operators are actually unitarily equivalent.

Theorem 9.57. *Suppose $M, N, T \in \mathcal{B}(H), M$ and N are normal, T is invertible, and*

$$M = TNT^{-1}$$

If $T = US$ is the polar decomposition of T, then

$$M = UNU^{-1}$$

Proof. By hypotheses, we have $M = (US)N(US)^{-1} = USNS^{-1}U^{-1}$. It suffices to show that N and $S = \sqrt{T^*T}$ commutes. To this end, we show that N and T^*T commutes, then by Remark 9.5 the desired result holds.

Since $MT = TN$ we have $T^*TN = T^*MT$. On the other hand, by Exercise 9.5, we have $M^*T = TN^*$, Consequently,

$$T^*M = (M^*T)^* = (TN^*)^* = NT^*$$

and hence $NT^*T = T^*M$ as required. $\square$

Chapter 10

Unbounded Self-adjoint Operators in Hilbert Space

10.1 Adjoint Operators

Let H be a Hilbert space. By an operator in H we now mean a linear mapping T whose domain $D(T)$ and range $R(T)$ lie in H. We wish to associate a (Hilbert space) *adjoint* T^* to T. Its domain is defined by

$$D\left(T^*\right) := \{y \in H : \text{ the map } x \mapsto \langle Tx, y\rangle \text{ is continuous on } D(T)\}\ .$$

If $y \in D\left(T^*\right)$, then the Hahn-Banach theorem extends the functional $x \mapsto \langle Tx, y\rangle$ to a continuous linear functional on H, and therefore there exists an element $T^*y \in H$ that satisfies

$$\langle Tx, y\rangle = \langle x, T^*y\rangle \text{ for all } x \in D(T)\,. \tag{10.1}$$

On the contrary, if there is $z \in H$ satisfying

$$\langle Tx, y\rangle = \langle x, z\rangle \text{ for all } x \in D(T)\,.$$

then $y \in D(T^*)$ and $T^*y = z$.

We should emphasize that T^*y will be uniquely determined by (10.1) if and only if $D(T)$ is dense in H, that is, if and only if T is *densely defined*. The only operators T that will be given an adjoint T^* are therefore the densely defined ones. Thus in this chapter we shall assume all the operators are densely defined. Routine verifications show then that T^* is also an operator in H, that is, that $D(T^*)$ is a subspace of H and that T^* is linear.

Definition 10.1. The linear operator $(T^*, D(T^*))$ is called the *adjoint operator* of the densely defined $(T, D(T))$ on H.

Note that if $T \in \mathscr{B}(H)$, then the definition of T^* given here coincides with that given in Section 8.1.2. In particular, $D(T^*) = H$ and $T^* \in \mathscr{B}(H)$.

Proposition 10.1. *Let $(T, D(T))$ be a densely defined operator in H. Then $(T^*, D(T^*))$ is an closed operator in H.*

Proof. Suppose (y_n) is a sequence in $D(T^*)$ so that $y_n \to y$ in H and $T^*y_n \to z$ in H. Then for all $x \in D(T)$, we have

$$\langle Tx, y\rangle = \lim_{n\to\infty} \langle Tx, y_n\rangle = \lim_{n\to\infty} \langle x, T^*y_n\rangle = \langle x, z\rangle .$$

Thus $y \in D(T^*)$ and $z = T^*y$ as desired. $\square$

Exercise 10.1. Let S and T be two densely defined operators in H. Show that the following statements hold.

(a) $(\lambda T)^* = \lambda T^*$ for $\lambda \in \mathbb{C}$

(b) If $T + S$ is densely defined, then $T^* + S^* \subset (T + S)^*$. If, in addition, $S \in \mathscr{B}(H)$, then $T^* + S^* = (T + S)^*$.

(c) If ST is densely defined then $T^*S^* \subset (ST)^*$. If, in addition, $S \in \mathscr{B}(H)$, then $T^*S^* = (ST)^*$.

(d) If $T \subset S$, then $S^* \subset T^*$.

Proposition 10.2. *Let $(T, D(T))$ be a densely defined closable operator in H. Then $\overline{T}^* = T^*$.*

Proof. Since $T \subset \overline{T}$ we have $\overline{T}^* \subset T^*$. It remains to show that $D(T^*) \subset D(\overline{T}^*)$. Take $y \in D(\overline{T}^*)$, then for each $x \in D(\overline{T})$ there is a sequence (x_n) in $D(T)$ sa that $x_n \to x$ and $Tx_n \to \overline{T}x$. Thus

$$\langle \overline{T}x, y \rangle = \lim_{n\to\infty} \langle Tx_n, y \rangle = \lim_{n\to\infty} \langle x_n, T^* y \rangle = \langle x, T^* y \rangle .$$

So $y \in D(\overline{T}^*)$ and $\overline{T}^* y = T^* y$ as desired. □

10.1.1 The Graph of T^* and T

A number of important and useful properties of closable and adjoint operators are derived by the so-called *graph method.*

Let H be a Hilbert space. Recall that $H \times H$ can be made into a Hilbert space by defining the inner product of two elements (x_1, y_1) and (x_2, y_2) of $H \times H$ to be

$$\langle (x_1, y_1), (x_2, y_2) \rangle = \langle x_1, y_1 \rangle + \langle x_2, y_2 \rangle .$$

In particular, the corresponding norm in $H \times H$ is given by

$$\|(x, y)\|^2 = \|x\|^2 + \|y\|^2 , \text{ for } (x, y) \in H \times H .$$

Define the operator V on $H \times H$ by

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

Then V is a unitary operator on $H \times H$, which satisfies

$$V^2 = -I$$

Thus $V^2 M = M$ if M is any subspace of $H \times H$. This operator yields a remarkable description of T^* in terms of T:

Theorem 10.3. *If T is a densely defined operator in H, then*

$$G(T^*) = [VG(T)]^\perp ,$$

the orthogonal complement of $VG(T)$ in $H \times H$.

In particular, $G(T^*)$ is closed in $H \times H$ as hence T^* is a closed operator.

Proof. Note that

$$\begin{aligned}
(y,z) \in G(T^*) &\Leftrightarrow y \in D(T^*), z = T^* y \\
&\Leftrightarrow \langle Tx, y\rangle = \langle x, z\rangle \quad \forall x \in D(T) \\
&\Leftrightarrow -\langle Tx, y\rangle + \langle x, z\rangle = 0 \quad \forall x \in D(T) \\
&\Leftrightarrow \langle (-Tx, x), (y,z)\rangle = 0. \quad \forall x \in D(T) \\
&\Leftrightarrow (y,z) \perp V(G(T)),
\end{aligned}$$

then we are done. $\square$

Corollary 10.4. *If T is a densely defined operator in H, then*

$$H \times H = V\overline{G(T)} \ominus G(T^*) = \overline{G(T)} \ominus VG(T^*) .$$

If in addition T is closable, using Corollary 10.4 for T and $\overline{T}$ we get $H \times H = VG(\overline{T}) \ominus G(T^*)$ and $H \times H = VG(\overline{T}) \ominus G(\overline{T}^*)$, thus $G(T^*) = G(\overline{T}^*)$, i.e., $T^* = \overline{T}^*$.

Exercise 10.2. Let T be a densely defined operator in H. Show that if $a \in H$ and $b \in H$, the system of equations

$$\begin{aligned}
-Tx + y &= a \\
x + T^* y &= b
\end{aligned}$$

has a unique solution with $x \in D(T)$ and $y \in D(T^*)$.

10.1.2 Does $T^{**} = T$?

As we know for $T \in \mathcal{B}(H)$ we have $T^{**} = T$. Does this equality holds for unbounded operators? First of all, $T^{**} := (T^*)^*$ is well-defined iff $D(T^*)$ is dense. Secondly, if $T^{**} = T$ then T must be closed. The following theorem show that if T is closed (closable) then $D(T^*)$ is dense and $T^{**} = T$ ($T^{**} = \overline{T}$) by using the graph method.

Theorem 10.5. *Let T be a densely defined linear operator in H. Then*

(a) T^* is densely defined iff T is closable, in this case $T^{**} = \overline{T}$.

(b) If T is closed, then $T = T^{**}$.

Proof. It suffices to prove part (a). Assume that T is closable, then we show $D(T^*)^{\perp} = \{0\}$. Let $u \in D\left(T^*\right)^{\perp}$. Then one can see that

$$(u, 0) \in \mathcal{G}\left(T^*\right)^{\perp} = VG(\overline{T}).$$

where the second equality holds by Theorem 10.3. Thus $V(u, 0) = (0, -u) \in G(\overline{T})$, i.e., $u = -\overline{T}(0) = 0$ as desired.

On the contrary, assume that T^* is densely defined. Then using Theorem 10.4 for T^* we get

$$H \times H = VG\left(T^*\right) \oplus G\left(T^{**}\right)$$

On the other hand, Theorem 10.4 asserts that

$$H \times H = \overline{G(T)} \oplus VG\left(T^*\right).$$

Thus we have

$$\overline{G(T)} = G\left(T^{**}\right).$$

Since T^{**} is closed, thus T is closable and $\overline{T} = T^{**}$. □

Proposition 10.6. *If T is a densely defined operator in $\mathcal{H}$, then*

$$R(T)^{\perp} = N\left(T^*\right); \overline{R(T)} = N\left(T^*\right)^{\perp}$$

If in addition T is dosed, then

$$R\left(T^*\right)^{\perp} = N(T); \quad \overline{R\left(T^*\right)} = N(T)^{\perp}.$$

Proof. Note that

$$\begin{aligned} y \in R(T)^{\perp} &\Leftrightarrow \langle x, y \rangle = 0, \ \forall x \in D(T) \\ &\Leftrightarrow y \in D\left(T^*\right), \ T^* y = 0 \Leftrightarrow y \in N\left(T^*\right). \end{aligned}$$

then we get $R(T)^{\perp} = N(T^*)$. As a consequence, $\overline{R(T)} = R(T)^{\perp\perp} = N(T^*)^{\perp}$. If in addition T is closed, the desired result follows from the previous result and the fact that $T^{**} = T$. $\square$

Exercise 10.3. Let T be a densely defined linear operator in H. Show that

(a) If $R(T)$ is dense in H, then T^* is injective.

(b) If $N(T) = \{0\}$ and $R(T)$ is dense in H, then T^* is injective and

$$\left(T^{-1}\right)^* = (T^*)^{-1} .$$

(c) If T is closable and $N(T) = \{0\}$. Then the inverse T^{-1} of T is closable if and only if $N(\bar{T}) = \{0\}$. If this holds, then

$$(\bar{T})^{-1} = \overline{(T^{-1})}$$

10.1.3 Examples: Differential Operators

In the first paragraph we discuss the notions from the preceding sections for the differentiation operator $-\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}x}$ and its square $-\frac{\mathrm{d}^2}{\mathrm{d}x^2}$ on various intervals. We develop these examples and their continuations in later sections in great detail and as elementarily as possible by using absolutely continuous functions. In the second paragraph distributions are used to define maximal and minimal operators for linear partial differential expressions.

Differentiation Operators on Bounded Interval Suppose that $a, b \in \mathbb{R}, a < b$. Let T be the linear operator on the complex Hilbert space $L^2((a,b);\mathbb{C})$ defined by $Tf = -\mathrm{i}f'$ for f in

$$D(T) = H_0^1(a,b) \coloneqq \left\{f \in H^1(a,b) : f(a) = f(b) = 0\right\} .$$

Clearly, its square T^2 acts as $T^2 f = -f''$ for f in

$$D\left(T^2\right) = H_0^2(a,b) \coloneqq \left\{f \in H^2(a,b) : f(a) = f(b) = f'(a) = f'(b) = 0\right\} .$$

Obviously, $D(T)$ and $D\left(T^2\right)$ are dense in $L^2(a,b)$. Our aim in this example is to describe the adjoints of both operators T and T^2 and a core for T.

Lemma 10.7. *$R(T)^{\perp} = \operatorname{span}\{1\}$ and $R\left(T^2\right)^{\perp} = \operatorname{span}\{1,x\}$, here 1 and x denote the functions $f_1(x) = 1$ and $f_2(x) = x$ for $x \in [a,b]$, respectively.*

Proof. One can check that $\operatorname{span}\{1\} \subset R(T)^{\perp}$ and $\operatorname{span}\{1,x\} \subset R\left(T^2\right)^{\perp}$. We now prove that $\operatorname{span}\{1\}^{\perp} \subset R(T)$ and $\operatorname{span}\{1,x\}^{\perp} \subset R\left(T^2\right)$, then the desired result follows.

Suppose that $h_1 \in \operatorname{span}\{1\}^{\perp}$ and $h_2 \in \operatorname{span}\{1,x\}^{\perp}$. We define functions on $[a,b]$ by

$$k_1(x) = \int_a^x h_1(t)\,\mathrm{d}t, \quad k_2(x) = \int_a^x \left(\int_a^t h_2(s)\,\mathrm{d}s\right)\mathrm{d}t\,. \tag{10.2}$$

Since $h_1, h_2 \in L^2(a,b) \subset L^1(a,b)$, we conclude that $k_1 \in H^1(a,b), k_1' = h_1$, and $k_2 \in H^2(a,b), k_2'' = h_2$. Obviously, $k_1(a) = k_2(a) = k_2'(a) = 0$. Moreover, $k_1(b) = \langle h_1, 1\rangle = 0$ and $k_2'(b) = \langle h_2, 1\rangle = 0$. Hence, $k_2'(x)x|_a^b = 0$, and using the integration by parts formula we derive

$$k_2(b) = \int_a^b k_2'(t)\,\mathrm{d}t = \left\langle k_2', 1\right\rangle = -\left\langle k_2'', x\right\rangle = -\left\langle h_2, x\right\rangle = 0$$

Thus, we have shown that $k_1 \in D(T)$ and $k_2 \in D\left(T^2\right)$, so $T\left(\mathrm{i}k_1\right) = h_1 \in R(T)$ and $T^2\left(-k_2\right) = h_2 \in R\left(T^2\right)$. □

Proposition 10.8. *$D\left(T^*\right) = H^1(a,b)$ and $T^*g = -\mathrm{i}g'$ for $g \in D\left(T^*\right)$.*

$D\left((T^*)^2\right) = D((T^*)^2) = H^2(a,b)$ and $\left(T^2\right)^* g = -g''$ for $g \in D\left((T^*)^2\right)$.

Proof. First let $g \in H^1(a,b)$. Let $f \in D(T)$. Since $f(a) = f(b) = 0$, using the integration by parts formula we obtain

$$\langle Tf, g\rangle = -\mathrm{i}\left\langle f', g\right\rangle = \mathrm{i}\left\langle f, g'\right\rangle = \left\langle f, -\mathrm{i}g'\right\rangle\,.$$

By the definition of T^* it follows that $g \in D\left(T^*\right)$ and $T^*g = -\mathrm{i}g'$. Now let $g \in H^2(a,b)$. Then, by the definition of $H^2(a,b)$, g and g' are in $H^1(a,b)$.

Applying the result of the preceding paragraph twice, first to g and then to $T^*g = -\mathrm{i}g'$, we conclude that $g \in D\left((T^*)^2\right)$ and $(T^*)^2 g = -g''$.

Conversely, suppose that $g_1 \in D(T^*)$ and $g_2 \in D\left((T^*)^2\right)$. We set $h_1 \coloneqq T^*g_1$ and $h_2 \coloneqq \left(T^2\right)^* g_2$ and define functions k_1 and k_2 by (10.2). As noted above, we then have $k_1 \in H^1(a,b), k_2 \in H^2(a,b), k_1' = h_1$, and $k_2'' = h_2$.

Let $f_1 \in D(T)$ and $f_2 \in D\left(T^2\right)$. Since the boundary values $f_1(a)$, $f_1(b)$, $f_2(a)$ $f_2(b)$, $f_2'(a)$, and $f_2'(b)$ vanish, it follows from teh integration by parts formula that

$$-\left\langle f_1', k_1\right\rangle = \left\langle f_1, k_1'\right\rangle = \langle f_1, h_1\rangle = \langle f_1, T^*g_1\rangle = \langle Tf_1, g_1\rangle = \left\langle -\mathrm{i}f_1', g_1\right\rangle$$
$$\left\langle f_2'', k_2\right\rangle = \left\langle f_2, k_2''\right\rangle = \langle f_2, h_2\rangle = \left\langle f_2, \left(T^2\right)^* g_2\right\rangle = \left\langle T^2 f_2, g_2\right\rangle = \left\langle -f_2'', g_2\right\rangle$$

Hence, $\langle -\mathrm{i}f_1', g_1 - \mathrm{i}k_1\rangle = 0$ and $\langle f_2'', g_2 + k_2\rangle = 0$, so that $g_1 - \mathrm{i}k_1 \in \mathcal{R}(T)^\perp = \operatorname{span}\{1\}$ and $g_2 + k_2 \in \mathcal{R}\left(T^2\right)^\perp \subset \operatorname{span}\{1, x\}$ by the preceding lemma. Since the functions k_1 and 1 are in $H^1(a,b)$ and $k_2, 1$, and x are in $H^2(a,b)$, we conclude that $g_1 \in H^1(a,b)$ and $g_2 \in H^2(a,b)$.

By the preceding, we have proved the assertions about T^* and the relations $H^2(a,b) \subset D\left((T^*)^2\right), (T^*)^2 g = -g''$ for $g \in H^2(a,b)$, and $D\left((T^*)^2\right) \subset H^2(a,b)$, this implies the assertions concerning $\left(T^2\right)^*$ and $(T^*)^2$. □

10.2 Symmetric Operators, Self-adjoint Operators

The notion of a closed operator is too general to develop a deeper theory. This section is devoted to some important classes of closed or closable operators: symmetric operators and self-adjoint operators. The interplay between symmetric and self-adjoint operators, or more precisely, the problem of when a symmetric operator is self-adjoint is one of the main themes in this chapter.

From now on, we shall assume that H is a complex Hilbert space. Let $(T, D(T))$ be a densely defined linear operator on H.

10.2.1 Symmetric Operators

Definition 10.2. The densely defined operator $(T, D(T))$ is called *symmetric* if

$$\langle Tx, y\rangle = \langle x, Ty\rangle \quad \text{for all } x, y \in D(T)\,.$$

Remark 10.1. Indeed the definition of symmetric operator does not need the density of the domain. For convenience, we only deal with densely defined symmetric operator, unless specifically stated.

Clearly we can conclude that T is symmetric if and only if

$$T \subset T^*\,.$$

Since T^* is closed, it follows that the symmetric operator T is closable with $\overline{T} \subset T^*$. Besides, as $(\bar{T})^* = T^*$, its closure $\bar{T}$ is again symmetric, and we have

$$T \subset \bar{T} = T^{**} \subset T^* = (\bar{T})^*\,.$$

If the domain $D(T)$ is the whole Hilbert space H, then $T = T^*$ and hence T is a closed operator with domain H. By the closed graph theorem $T \in \mathcal{B}(H)$. This is the Hellinger-Toeplitz theorem proved in Example 4.8.

Exercise 10.4. Let $(T, D(T))$ be a densely defined linear operator in H. Show that T is symmetric if and only if $\langle Tx, x\rangle$ is real for all $x \in D(T)$.

Definition 10.3. Let $(T, D(T))$ be a densely defined symmetric operator. T is said to be *lower semibounded* if there exists a real number m, called a *lower bound* for T, satisfying that

$$\langle Tx, x\rangle \geq m\|x\|^2 \ \text{ for all } \ x \in D(T)\,.$$

Similarly, we can define the *upper semibounded* symmetric operator. If T is lower semibounded or upper semibounded, T is called *semibounded*.

We say that T is nonnegative and write $T \geq 0$ if $\langle Tx, x\rangle \geq 0$ for all $x \in D(T)$. If $\langle Tx, x\rangle > 0$ for all nonzero $x \in D(T)$, then T is called positive.

Clearly, each lower semibounded operator T has a greatest lower bound given by

$$m_T := \inf\{\langle Tx, x\rangle; x \in D(T), \|x\| = 1\}\ .$$

The next proposition deals with the regularity domain of a symmetric operator.

Lemma 10.9. *Let $(T, D(T))$ be a densely defined symmetric operator on H, then the following statements hold.*

(a) *For $x \in D(T)$ and $\lambda \in \mathbb{C}\backslash\mathbb{R}$ we have*

$$\|(T - \lambda I)x\|^2 = \|(T - \operatorname{Re}\lambda I)x\|^2 + |\operatorname{Im}\lambda|^2\|x\|^2\,.$$

In particular, $\mathbb{C}\backslash\mathbb{R} \subset \pi(T)$.

(b) *T is closed if and only if $R(T - \lambda I)$ is closed for $\lambda \in \mathbb{C}\backslash\mathbb{R}$.*

(c) *If T is lower semibounded with lower bound m, then $(-\infty, m) \subset \pi(T)$.*

Proof. We now prove part (a). For $x \in D(T)$,

$$\begin{aligned}
\|(T - \lambda I)x\|^2 &= \|(Tx - \operatorname{Re}\lambda x) - \mathrm{i}\operatorname{Im}\lambda x\|^2 \\
&= \|(T - \operatorname{Re}\lambda I)x\|^2 + |\operatorname{Im}\lambda|^2\|x\|^2 - 2\operatorname{Re}\langle Tx - \operatorname{Re}\lambda x, \mathrm{i}\operatorname{Im}\lambda x\rangle \\
&= \|(T - \operatorname{Re}\lambda I)x\|^2 + |\operatorname{Im}\lambda|^2\|x\|^2 - 2\operatorname{Im}\lambda\operatorname{Im}\langle Tx - \operatorname{Re}\lambda x, x\rangle \\
&= \|(T - \operatorname{Re}\lambda I)x\|^2 + |\operatorname{Im}\lambda|^2\|x\|^2\,,
\end{aligned}$$

since T is symmetric $\langle Tx, x\rangle$ is real. Then we have

$$\|(T - \lambda I)x\| \geq |\operatorname{Im}\lambda|\|x\| \qquad \text{for } x \in D(T), \lambda \in \mathbb{C}$$

Therefore, $\lambda \in \pi(T)$ if $\operatorname{Im}\lambda \neq 0$.

To show part (b), observe that, by part (a), $(Tx_n - \lambda x_n)$ is a Cauchy sequence in H if and only if (x_n), (Tx_n) are Cauchy sequence in H. Then we are done.

We prove the assertion for lower semibounded T. For $\lambda \in (-\infty, m)$,

$$(m-\lambda)\|x\|^2 \leq \langle (T-\lambda I)x, x\rangle \leq \|(T-\lambda I)x\|\|x\|$$

and hence $(m-\lambda)\|x\| \leq \|(T-\lambda I)x\|$ for $x \in D(T)$. Since $m - \lambda > 0$, we get $\lambda \in \pi(T)$ as desired. □

Recall from Theorem 6.16 that for a closable operator T, the defect number $d_\lambda(T)$ is constant on connected subsets of the regularity domain $\pi(T)$. Therefore, if T is closable and symmetric, by Lemma 10.9 the number $d_\lambda(T)$ is constant on the upper half-plane and on the lower half-plane. If, in addition, T is semibounded, then $\pi(T)$ is connected, and hence $d_\lambda(T)$ is constant on the whole set $\pi(T)$.

Definition 10.4. The *deficiency indices* (or the *defect numbers*) of a densely defined symmetric operator T are the cardinal numbers

$$\begin{aligned} d_+(T) &:= \dim R(T+iI)^\perp = \dim N\left(T^* - \mathrm{i}I\right) ; \\ d_-(T) &:= \dim R(T-iI)^\perp = \dim N\left(T^* + \mathrm{i}I\right) . \end{aligned}$$

By Theorem 6.16 there holds

$$\begin{aligned} d_+(T) &= \dim R(T-\bar\lambda I)^\perp, \quad \operatorname{Im}\lambda > 0\,; \\ d_-(T) &= \dim R(T-\bar\lambda I)^\perp, \quad \operatorname{Im}\lambda < 0\,. \end{aligned}$$

Remark 10.2. Let T be a densely defined symmetric operator. If T is semibounded or more generally if $\pi(T)$ contains a real number, then we have $d_+(T) = d_-(T)$. Since in both cases, $\pi(T)$ is connected and the assertion follows from Theorem 6.16.

Definition 10.5. Let $(T, D(T))$ be a symmetric operator on H. We say T is *maximal symmetric* if T has no proper symmetric extension; i.e., if S is a symmetric operator on H such that $T \subset S$, then $T = S$.

Proposition 10.10. *Let $(T, D(T))$ be a densely defined closed symmetric operator in H. Then if $d_+(T) = 0$ or $d_-(T) = 0$, then T is maximal symmetric.*

Proof. If $d_+(T) = 0$ or $d_-(T) = 0$, there exists $\lambda \in \mathbb{C}\backslash\mathbb{R}$ so that $\dim R(T - \lambda I)^\perp = 0$. Since T is closed, $R(T - \lambda I)$ is closed and hence $R(T - \lambda I) = H$. Note that if S is an symmetric extension of T, then $\lambda \in \pi(S)$ and $R(T - \lambda I) = H$. Thus we have $D(T) = D(S)$ as desired. $\square$

Proposition 10.11 (Spectrum of Symmetric Operator)**.** *Let $(T, D(T))$ be a densely defined symmetric operator in H. Then the spectrum of T can only be one of the following four cases:*

(i) $\sigma(T) = \{\lambda \in \mathbb{C} : \operatorname{Im}\lambda \geq 0\}$ *if* $d_+(T) = 0$ *and* $d_-(T) > 0$.

(ii) $\sigma(T) = \{\lambda \in \mathbb{C} : \operatorname{Im}\lambda \leq 0\}$ *if* $d_+(T) > 0$ *and* $d_-(T) = 0$.

(iii) $\sigma(T) = \mathbb{C}$ *if* $d_+(T) > 0$ *and* $d_-(T) > 0$.

(iv) $\sigma(T) \subset \mathbb{R}$ *if* $d_+(T) = d_-(T) = 0$.

Proof. Take $\lambda \in \mathbb{C}$ with $\operatorname{Im}\lambda > 0$ we have $\dim R(T - \bar{\lambda} I)^\perp = d_+(T)$. Combine this with the fact that $\bar{\lambda} \in \pi(T)$ we conclude that $\bar{\lambda} \in \varrho(T)$ if $d_+(T) = 0$ and $\bar{\lambda} \in \sigma_r(T)$ if $d_+(T) > 0$.

Similarly, for $\lambda \in \mathbb{C}$ with $\operatorname{Im}\lambda < 0$, we have $\bar{\lambda} \in \varrho(T)$ if $d_-(T) = 0$ and $\bar{\lambda} \in \sigma_r(T)$ if $d_-(T) > 0$. Note that the spectrum $\sigma(T)$ is always a closed subset of $\mathbb{C}$, the desired result follows. $\square$

Exercise 10.5. Let $(T, D(T))$ be a densely defined symmetric operator in H. Show that

(a) Any eigenvalue of T is real.

(b) Eigenvectors belonging to different eigenvalues of T are mutually orthogonal.

(c) Suppose $T \geq 0$. If $\langle Tx, x\rangle = 0$ for some $x \in D(T)$, then $Tx = 0$. (Hint: use the Cauchy-Schwarz inequality.)

10.2.2 Self-adjoint Operators

Self-adjointness is the most important notion on unbounded operators in this chapter. The main results about self-adjoint operators are the spectral theorem proved later and the corresponding functional calculus based on it.

Definition 10.6. A densely defined symmetric operator $(T, D(T))$ in H is called *self-adjoint* if $T = T^*$; and *essentially self-adjoint* if $\bar{T}$ is self-adjoint or equivalently if $\bar{T} = T^*$.

Let us state some simple consequences that will be often used without mention.

- A self-adjoint operator T is symmetric and closed, since T^* is always closed.
- Let T be a densely defined symmetric operator. Since then T is self-adjoint if and only if $D(T) = D(T^*)$ Likewise, T is essentially self-adjoint if and only if $D(\bar{T}) = D(T^*)$.
- Any self-adjoint operator T on H is *maximal symmetric*. Indeed, $T \subset S$ implies that $S^* \subset T^* = T$. Combined with $S \subset S^*$ this yields $S \subset T$, so that $T = S$.

The next proposition characterizes the numbers of the resolvent set of a self-adjoint operator. Condition (ii) therein is often useful to detect the spectrum of the operator.

Proposition 10.12. *Let T be a self-adjoint operator on a Hilbert space H. For any complex number λ, the following conditions are equivalent:*

(i) $\lambda \in \rho(T)$;

(ii) $\lambda \in \pi(T)$, *that is, there exists a constant* $c_\lambda > 0$ *such that* $\|(T-\lambda I)x\| \geq c_\lambda \|x\|$ *for all* $x \in D(T)$*;*

(iii) $R(T - \lambda I) = H$.

Moreover, if $\lambda \in \mathbb{C}\backslash\mathbb{R}$*, then* $\lambda \in \rho(T)$ *and* $\|R(\lambda, T)\| \leq |\operatorname{Im}\lambda|^{-1}$.

When does a symmetric operator is self-adjoint or essentially self-adjoint? Such self-adjointness criteria follow easily from the next result.

Lemma 10.13 (von Neumann's Formula)**.** *Let* $(T, D(T))$ *be a densely defined symmetric operator. Then we have*

$$D\left(T^*\right) = D(\bar{T}) \oplus \mathcal{N}\left(T^* - \lambda I\right) \oplus \mathcal{N}\left(T^* - \bar{\lambda} I\right) . \tag{10.3}$$

Proof. Without loss of generality we assume that T is closed and $\operatorname{Im}\lambda > 0$. Set $D_+ = N\left(T^* - \lambda I\right)$ and $D_- = N\left(T^* - \bar{\lambda} I\right) = R\left(T - \lambda I\right)^{\perp}$.

Firstly we prove that $D(T) \oplus \mathcal{N}\left(T^* - \lambda I\right) \oplus \mathcal{N}\left(T^* - \bar{\lambda} I\right)$ is well-defined. Assume $x_0 \in D(T)$ and $x_\pm \in D_\pm$ so that

$$x_0 + x_+ + x_- = 0\,.$$

Then we have

$$Tx_0 + \lambda x_+ + \bar{\lambda} x_- = 0\,;$$

and hence $2\operatorname{Im}\lambda x_- = (T - \lambda I)x_0$. We get

$$x_- \in R\left(T - \lambda I\right)^{\perp} \cap R(T - \lambda) = \{0\}\,.$$

Similarly $x_+ = 0$ and so $x_0 = 0$ as desired.

It remains to show that $D(T^*) \subset D(T) \oplus \mathcal{N}\left(T^* - \lambda I\right) \oplus \mathcal{N}\left(T^* - \bar{\lambda} I\right)$. That is for $x \in D(T^*)$ we need to find $x_0 \in D(T)$ and $x_\pm \in D_\pm$ so that $x = x_0 + x_+ + x_-$. Then we must have

$$T^*x = Tx_0 + \lambda x_+ + \bar{\lambda} x_-\,;$$

and hence

$$(T^* - \lambda I)x = (T - \lambda I)x_0 - 2(\operatorname{Im} \lambda)x_- . \tag{10.4}$$

However, since $D_- = R\,(T - \lambda I)^{\perp}$ we have $H = R\,(T - \lambda I) \oplus D_-$. Thus such $x_0 \in D(T)$ and $x_- \in D_-$ satisfying (10.4) exist. Now let $x_+ = x - x_0 - x_-$, we check that if $x_+ \in D_+$. By (10.4)

$$(T^* - \lambda I)x_+ = (T^* - \lambda I)(x - x_0 - x_-) = 0$$

as desired. □

Recall that a densely defined symmetric operator T is essentially self-adjoint if and only if $D(\bar{T}) = D\,(T^*)$. That is, using von Neumann's formula, we obtain the following:

Theorem 10.14 (Deficiency Indices)**.** *Let $(T, D(T))$ be a densely defined symmetric operator. Then the following the statements hold.*

(a) T is essentially self-adjoint if and only if $d_+(T) = d_-(T) = 0$.

(b) T is self-adjoint if and only if T is closed and $d_+(T) = d_-(T) = 0$.

Combine Theorem 10.14 and Proposition 10.11 we have:

Corollary 10.15. *A closed symmetric linear operator T on $\mathcal{H}$ is self-adjoint if and only if $\sigma(T) \subset \mathbb{R}$.*

Remark 10.3. Since the deficiency indices are constant on connected subsets of $\pi(T)$, Let T be a closed symmetric operator. Suppose that $\pi(T)$ contains a real number. Then T is self-adjoint if and only if $d_\lambda(T) = 0$ (equivalently, $\lambda \in \rho(T)$, or equivalently, $\mathcal{R}(T - \lambda I) = \mathcal{H}$) for one, hence all, $\lambda \in \pi(T)$.

Now we give some self-adjointness criteria that do not assume that the symmetric operator is densely defined. They are essentially based on the following proposition. In particular, the preceding result implies again Theorem 10.14.

Proposition 10.16. *Let $(T, D(T))$ be a symmetric operator in H (here we do assume need $D(T)$ is dense in H). If there is $\lambda \in \mathbb{C}$ so that $R(T-\lambda I) = R(T-\bar{\lambda} I) = H$ then T is self-adjoint and $\lambda, \bar{\lambda} \in \varrho(T)$.*

Proof. We first show that $D(T)$ is dense in H. Let $y \in D(T)^{\perp}$ and we aim to prove that $y = 0$. Since $R(T-\lambda I) = H$, there exists a vector $u \in D(T)$ such that $y = (T-\lambda I)u$. Therefore, we have

$$0 = \langle y, x\rangle = \langle (T-\lambda I)u, x\rangle = \langle u, (T-\bar{\lambda} I)x\rangle$$

for all $x \in D(T)$, so $u \in R(T-\bar{\lambda} I)^{\perp}$. Since $R(T-\bar{\lambda} I)$ is dense, $u = 0$ and hence $y = 0$. Thus, $D(T)$ is dense and hence T^* is well defined.

It remains to show that $D(T^*) \subset D(T)$. Let $w \in D(T^*)$, we shall find some $v \in D(T)$ so that $v = w$. Since $R(T-\bar{\lambda} I)^{\perp} = N(T^* - \lambda I) = \{0\}$, it suffices to find some $v \in D(T)$ satisfying

$$(T^* - \lambda I)v = (T-\lambda I)v = (T^* - \lambda I)w\,.$$

Applying once more the assumption $R(T-\lambda I) = H$, such $v \in D(T)$ exists as required. □

The case $\lambda = 0$ in Proposition 10.16 yields the following:

Exercise 10.6. If $(T, D(T))$ is a densely defined symmetric operator on H such that $R(T) = H$, then T is self-adjoint, and its inverse T^{-1} is a bounded self-adjoint operator on H.

10.3 The Cayley Transform

The mapping

$$t \to \frac{t-i}{t+i}\,, \quad t \in \mathbb{R}$$

sets up a one-to-one correspondence between the real line $\mathbb{R}$ and the unit circle S^1 minus the point 1. Indeed, one can check that the inverse mapping

is
$$e^{i\theta} \mapsto i\frac{1+e^{i\theta}}{1-e^{i\theta}}\,, \quad e^{i\theta} \in S^1\backslash\{1\}\,.$$

Let now $T \in \mathcal{B}(H)$ be self-adjoint. As we know, $\sigma(T) \subset \mathbb{R}$ and hence $\pm i \in \varrho(T)$. Let
$$U = (T - iI)(T + iI)^{-1} \in \mathcal{B}(H)\,.$$

We claim that U is unitary and $1 \in \varrho(U)$. To this end, we use the tool of functional calculus. Denote by E^T the spectrum decomposition of T, then we have
$$U = \int_{\sigma(T)} \frac{t-i}{t+i}\, E^T(\mathrm{d}t)\,.$$

By the spectrum mapping theorem $\sigma(U) \subset S^1\backslash\{1\}$, thus U is unitary with $1 \in \varrho(U)$ and it follows from part (f) of Theorem 9.40 that

$$T = i(I + U)(I - U)^{-1}\,.$$

On the contrary, let $U \in \mathcal{B}(H)$ be a unitary with $1 \in \varrho(U)$. Then define

$$T = i(I + U)(I - U)^{-1} \in \mathcal{B}(H)\,.$$

Similarly, by the functional calculus, let E^U be the spectrum decomposition of U then
$$T = \int_{\sigma(U)} i\frac{1+e^{i\theta}}{1-e^{i\theta}}\, E^U(\mathrm{d}\theta)\,.$$

Thus $\sigma(T) \subset \mathbb{R}$ and hence $T \in \mathcal{B}(H)$ is self-adjoint. Moreover, we have

$$U = (T - iI)(T + iI)^{-1}\,.$$

In summation, the Cayley transform introduce a bijection

$$\{T \in \mathcal{B}(H) : T = T^*\} \to \{U \in \mathcal{B}(H) : UU^* = U^*U = I, 1 \in \varrho(U)\}\,.$$

This relation will now be extended to a one-to-one correspondence between symmetric operators, on the one hand, and isometries, on the other.

Let $(T, D(T))$ be a densely defined symmetric operator in H. Recall that we have

$$\|Tx + ix\|^2 = \|x\|^2 + \|Tx\|^2 = \|Tx - ix\|^2$$

for all $x \in D(T)$. Hence there is an isometry U, with $D(U) = R(T + iI)$, $R(U) = R(T - iI)$ defined by

$$U(Tx + ix) = Tx - ix \quad \text{for} \quad x \in D(T)\,.$$

Since $(T + iI)^{-1}$ maps $D(U)$ onto $D(T), U$ can also be written in the form

$$U = (T - iI)(T + iI)^{-1}\,.$$

This operator U is called the *Cayley transform* of T. It will lead to an easy proof of the spectral theorem for self-adjoint (not necessarily bounded) operators.

Lemma 10.17. *Suppose U is an operator in H which is an isometry, i.e., $\|Ux\| = \|x\|$ for every $x \in D(U)$. Then the following statements hold:*

(a) $\langle Ux, Uy\rangle = \langle x, y\rangle$ *for all* $x, y \in D(U)$.

(b) *If $R(I - U)$ is dense in H, then $I - U$ is one-to-one.*

(c) *If any one of the three spaces $D(U), R(U)$, and $G(U)$ is closed, so are the other two.*

Proof. Part (a) follows from the polarization identity trivially.

To prove (b), suppose $x \in D(U)$ and $(I - U)x = 0$, i.e., $x = Ux$. Then

$$\langle x, (I - U)y\rangle = \langle x, y\rangle - \langle x, Uy\rangle = \langle Ux, Uy\rangle - \langle x, Uy\rangle = 0$$

for all $y \in D(U)$. Thus $x \in R(I - U)^{\perp}$ so that $x = 0$.

The proof of (c) is a consequence of the relations

$$\|Ux - Uy\| = \|x - y\| = \frac{1}{\sqrt{2}}\|(x, Ux) - (y, Uy)\|$$

which hold for all $x, y \in D(U)$. $\square$

Theorem 10.18. *Suppose U is the Cayley transform of a densely defined symmetric operator $(T, D(T))$ in H. Then the following statements are true.*

(a) $R(I-U) = D(T)$, *$I-U$ is one-to-one, and T can be reconstructed from U by the formula*

$$T = i(I+U)(I-U)^{-1}.$$

The Cayley transforms of distinct symmetric operators are therefore distinct.

(b) *U is closed if and only if T is closed.*

(c) *U is unitary if and only if T is self-adjoint.*

(d) *U is unitary and $1 \in \varrho(U)$ if and only if $T \in \mathcal{B}(H)$ is self-adjoint.*

Conversely, if V is an operator in H which is an isometry, and if $I-V$ is one-to-one, then V is the Cayley transform of a symmetric operator in H.

Proof. Step 1. We prove part (a). Take $x \in D(T)$, Let $y = (T+iI)x \in R(T+iI) = D(U)$, then

$$\begin{aligned}(I-U)y &= [I-(T-iI)(T+iI)^{-1}]y \\ &= (T+iI)(T+iI)^{-1}y - (T-iI)(T+iI)^{-1}y \\ &= 2i(T+iI)^{-1}y = 2ix.\end{aligned}$$

Thus $R(I-U) = D(T)$ is dense. By the preceding lemma, $I-U$ is one-to-one and hence

$$\begin{aligned}(I+U)y &= [I+(T-iI)(T+iI)^{-1}]y \\ &= (T+iI)(T+iI)^{-1}y + (T-iI)(T+iI)^{-1}y \\ &= 2T(T+iI)^{-1}y = 2Tx.\end{aligned}$$

Then

$$Tx = \frac{1}{2}(I+U)y = \frac{1}{2}(I+U)(I-U)^{-1}2ix = i(I+U)(I-U)^{-1}x.$$

To prove part (b), note that T is closed iff $R(T+iI) = D(U)$ is closed. By the preceding lemma, $D(U)$ is closed iff U is closed, and thus the desired result hold.

We now show part (c). Note that U is unitary iff $D(U) = R(U) = H$, i.e., $R(T+iI) = R(T-iI) = H$. By Proposition 10.16, T is self-adjoint iff $R(T+iI) = R(T-iI) = H$. Thus (c) follows.

To prove part (d), combine the fact that $D(T) = R(I-U)$ and (c), the desired result follows.

Step 2. Let V be as in the statement of the converse. Then $(I-V)^{-1} : R(I-V) \to D(V)$ exists. Define $D(S) = R(I-V)$ and

$$S = i(I+V)(T-V)^{-1}\,.$$

Then S is densely defined.

We claim that S is symmetric. Indeed, for $(I-V)x$, $(I-V)y$ in $D(S) = R(I-V)$,

$$\begin{aligned}
\langle S(I-V)x, (I-V)y\rangle &= \langle i(I+V)x, (I-V)y\rangle = i\langle Vx, y\rangle - i\langle x, Vy\rangle\,;\\
\langle (I-V)x, S(I-V)y\rangle &= \langle i(I-V)x, (I+V)y\rangle = i\langle Vx, y\rangle - i\langle x, Vy\rangle\,.
\end{aligned}$$

We show that $R(S+iI) = D(V)$. In fact, for $y = (I-V)x \in D(S) = R(I-V)$, we have

$$\begin{aligned}
(S+iI)y &= Sy + iy = i(I+V)x + iy\\
&= ix + Vx + i(I-V)x = 2ix\,.
\end{aligned}$$

Finally we show that V is the Cayley transform of S, that is

$$V = (S-iI)(S+iI)^{-1}\,.$$

For $y = (S+iI)x \in D(V) = R(S+iI)$, $x = (I-V)z \in D(S) = R(I-V)$, we have

$$\begin{aligned}
(S-iI)(S+iI)^{-1}y &= (S-iI)x = Sx - ix\\
&= i(I+V)z - i(I-V)z = 2iz\,.
\end{aligned}$$

We are done. $\square$

In summation, the Cayley transform establish a one-on-one correspondence form the left hand side of the follwoing table onto the right:

$(T, D(T))$	$(U, D(U))$
densely defined symmetric operator	isometry, $R(I-U)$ dense in H
self-adjoint operator	unitary, $R(I-U)$ dense in H
bonded and self-adjoint operator	unitary, $1 \in \varrho(U)$

10.4 Spectral Decompositions of Unbounded Self-adjoint Operators

Given a self-adjoint operator $(T, D(T))$ in H, by the Cayley transform, we know that $U = (T+iI)(T-iI)^{-1}$ is a unitary operator. Then by the spectral theorem in the last chapter, we have

$$U = \int_{\sigma(U)} e^{i\theta} E^U(\mathrm{d}\theta)$$

where E^U is the spectrum measure of U. Observing that $T = i(I+U)(I-U)^{-1}$, naturally, we would like to ask if

$$T = \int_{\sigma(U)} i\frac{1+e^{i\theta}}{1-e^{i\theta}} E^U(\mathrm{d}\theta) \ ?$$

Thus we need to defined the integral of unbounded functions w.r.t. a projection-valued measure.

10.4.1 Integral for unbounded functions

Let Ω be a Hausdorff space and let $\mathcal{F}$ be the Borel algebra on Ω. Assume that P is a projection-valued measure on $\mathcal{F}$. In the last chapter, for $f \in L^\infty(P)$, we have defined the integral $\int_\Omega f \,\mathrm{d}P$ as a bounded linear operator in $\mathcal{B}(H)$ determined by

$$\langle \int_\Omega f \,\mathrm{d}P x, y\rangle = \int_\Omega f \,\mathrm{d}P_{x,y}$$

for all $x, y \in H$.

Lemma 10.19. *Let $f : \Omega \to \mathbb{C}$ be measurable. Put*

$$D_f = \left\{ x \in H : \int_\Omega |f|^2 \, \mathrm{d}P_{x,x} < \infty \right\} .$$

Then D_f is a dense subspace of H. Besides, for $x \in D_f$ and $y \in H$ we have

$$\left| \int_\Omega f \, \mathrm{d}P_{x,y} \right| \le \int_\Omega |f| \, \mathrm{d}\, |P_{x,y}| \le \|y\| \left\{ \int_\Omega |f|^2 \, \mathrm{d}P_{x,x} \right\}^{1/2} . \tag{10.5}$$

Noe that if $f = g$ P-a.e., then $f = g$ $P_{x,x}$-a.e. for all $x \in H$ and hence $D_f = D_g$. To prove Lemma 10.19, let's see a lemma first.

Lemma 10.20. *Suppose $f \in L^\infty(P)$. Let $y \in H$ and $z = (\int_\Omega f \, \mathrm{d}P)y$. Then for each $x \in H$,*

$$\mathrm{d}P_{z,x} = f \, \mathrm{d}P_{y,x} \ ; \ \mathrm{d}P_{x,z} = \bar{f} \, \mathrm{d}P_{x,y} .$$

Proof. Note that for all $B \in \mathcal{F}$ we have

$$\begin{aligned} P_{z,x}(B) &= \langle P(B)z, x \rangle = \langle P(B)(\int_\Omega f \, \mathrm{d}P)y, x \rangle \\ &= \langle (\int_\Omega f 1_B \, \mathrm{d}P)y, x \rangle = \int_\Omega f 1_B \, \mathrm{d}P_{y,x} . \end{aligned}$$

Thus $\mathrm{d}P_{z,x} = f \, \mathrm{d}P_{y,x}$. By the same argument $\mathrm{d}P_{x,z} = \bar{f} \, \mathrm{d}P_{x,y}$. □

We now give the proof of Lemma 10.19.

Proof. We first show that D_f is a linear subspace of H. If $x, y \in D_f$, then, evidently $\alpha x \in D_f$ for all $\alpha \in \mathbb{C}$ since $P_{\alpha x, \alpha x} = |\alpha|^2 P_{x,x}$. It remains to show that $z = x + y \in D_f$. Note that for $B \in \mathcal{F}$,

$$\begin{aligned} &P_{z,z}(B) = \|P(B)x + P(B)y\|^2 \\ &\le 2[\|P(B)x\|^2 + \|P(B)y\|^2] = 2[P_{x,x}(B) + P_{y,y}(B)] . \end{aligned}$$

Thus $P_{z,z} \le 2(P_{x,x} + P_{y,y})$ and hence $z \in D_f$.

We show that D_f is dense in H. Fix $x \in H$. For each $n \ge 1$, define $x_n = P(|f| \le n)x$. On the one hand, since $P(|f| \le n) \xrightarrow{s} I$, we have $x_n \to x$.

On the other hand, by Lemma 10.20, $\mathrm{d}P_{x_n,x_n} = 1_{\{|f|\le n\}}\,\mathrm{d}P_{x,x}$. Then we get $x_n \in D_f$ since

$$\int_\Omega |f|^2\,\mathrm{d}P_{x_n,x_n} = \int_\Omega |f|^2 1_{\{|f|\le n\}}\,\mathrm{d}P_{x,x} < \infty\,.$$

Thus D_f is dense in H.

Finally we show that (10.5) holds for $x \in D_f$ and $y \in H$. Note that for all $g \in L^\infty(P)$

$$\begin{aligned}
&\int_\Omega |g|\,\mathrm{d}\,|P_{x,y}| = \int_\Omega |g| \frac{\mathrm{d}\,|P_{x,y}|}{\mathrm{d}P_{x,y}}\,\mathrm{d}P_{x,y} \\
&= \langle (\int_\Omega |g| \frac{\mathrm{d}\,|P_{x,y}|}{\mathrm{d}P_{x,y}}\,\mathrm{d}P)x, y\rangle \le \|y\| \left\{ \int_\Omega |g|^2\,\mathrm{d}P_{x,x} \right\}^{1/2}.
\end{aligned}$$

Thus for (unbounded) f, by the monotone convergence theorem,

$$\begin{aligned}
&\int_\Omega |f|\,\mathrm{d}|P_{x,y}| = \lim_{n\to\infty} \int_\Omega |f|\,1_{\{|f|\le n\}}\,\mathrm{d}|P_{x,y}| \\
&\le \lim_{n\to\infty} \|y\| \left\{ \int_\Omega |f|^2 1_{\{|f|\le n\}}\,\mathrm{d}P_{x,x} \right\}^{1/2} = \|y\| \left\{ \int_\Omega |f|^2\,\mathrm{d}P_{x,x} \right\}^{1/2}
\end{aligned}$$

as desired. $\square$

Now we are going to define the integral $\int_\Omega f\,\mathrm{d}P$, as an operator in H, for a unbounded complex-valued measurable function f on Ω. Of course we want the equality

$$\langle (\int_\Omega f\,\mathrm{d}P)x, y\rangle = \int_\Omega f\,\mathrm{d}P_{x,y}$$

holds for as many x, y as possible. Naturally, we ask $x \in D_f$ and $y \in H$ to provide that $\int_\Omega f\,\mathrm{d}P_{x,y}$ makes sense. In this case, since

$$\left| \int_\Omega f\,\mathrm{d}P_{x,y} \right| \le \|y\| \left\{ \int_\Omega |f|^2\,\mathrm{d}P_{x,x} \right\}^{1/2},$$

the mapping $y \mapsto \int_\Omega f\,\mathrm{d}P_{x,y}$ is a continuous linear functional on H. By Riesz's theorem, there is a unique point, denoted by $(\int_\Omega f\,\mathrm{d}P)x$, satisfying (10.4.1). Therefore the operator

$$\int_\Omega f\,\mathrm{d}P \quad \text{with domain} \quad D_f$$

is well-defined. It's easy to see that this is a linear operator of D_f into H.

Theorem 10.21. *For each $f : \Omega \to \mathbb{C}$ measurable, the operator $(\int_\Omega f \, \mathrm{d}P, D_f)$ is a densely defined closed operator in H, characterized by*

$$\langle (\int_\Omega f \, \mathrm{d}P)x, y \rangle = \int_\Omega f \, \mathrm{d}P_{x,y} \text{ for } x \in D_f \text{ and } y \in H\,.$$

Moreover, we have

(a) $(\int_\Omega f \, \mathrm{d}P)x = \lim_{n\to\infty} (\int_\Omega f 1_{\{|f|\le n\}} \, \mathrm{d}P)x$ *for* $x \in D_f$*; and*

(b) $\|(\int_\Omega f \, \mathrm{d}P)x\|^2 = \int_\Omega |f|^2 \, \mathrm{d}P_{x,x}$ *for* $x \in D_f$.

Proof. We show that $(\int_\Omega f 1_{\{|f|\le n\}} \, \mathrm{d}P)x$ converges to $(\int_\Omega f \, \mathrm{d}P)x$ for $x \in D_f$. Firstly note that $\{(\int_\Omega f 1_{\{|f|\le n\}} \, \mathrm{d}P)x\}_{n\ge 1}$ is a Cauchy sequence since

$$\left\| \int_\Omega f 1_{\{n<|f|\le n+k\}} \, \mathrm{d}Px \right\|^2 = \int_\Omega |f|^2 1_{\{n<|f|\le n+k\}} \, \mathrm{d}P_{x,x} \to 0$$

uniformly in $k \ge 1$ as $n \to \infty$. Moreover, by the dominated convergence theorem, we have

$$\begin{aligned} \langle \lim_{n\to\infty} (\int_\Omega f 1_{\{|f|\le n\}} \, \mathrm{d}P)x, y \rangle &= \lim_{n\to\infty} \langle (\int_\Omega f 1_{\{|f|\le n\}} \, \mathrm{d}P)x, y \rangle \\ &= \lim_{n\to\infty} \int_\Omega f 1_{\{|f|\le n\}} \, \mathrm{d}P_{x,y} = \int_\Omega f \, \mathrm{d}P_{x,y}\,. \end{aligned}$$

Thus part (a) holds. Part (b) is a easy consequence of part (a), indeed,

$$\begin{aligned} \|(\int_\Omega f \, \mathrm{d}P)x\|^2 &= \lim_{n\to\infty} \|(\int_\Omega f 1_{\{|f|\le n\}} \, \mathrm{d}P)x\|^2 \\ &= \lim_{n\to\infty} \int_\Omega |f|^2 1_{\{|f|\le n\}} \, \mathrm{d}P_{x,x} = \int_\Omega |f|^2 \, \mathrm{d}P_{x,x}\,. \end{aligned}$$

Finally we show that $(\int_\Omega f \, \mathrm{d}P, D_f)$ is closed. Let $\{x_n\}$ be a sequence in D_f with $x_n \to x$ and $(\int_\Omega f \, \mathrm{d}P)x_n \to y$. Take arbitrarily $M > 0$, then

$$\begin{aligned} &\int_\Omega |f|^2 1_{\{|f|\le M\}} \, \mathrm{d}P_{x,x} = \lim_{n\to\infty} \int_\Omega |f|^2 1_{\{|f|\le M\}} \, \mathrm{d}P_{x_n,x_n} \\ &\le \lim_{n\to\infty} \int_\Omega |f|^2 \, \mathrm{d}P_{x_n,x_n} = \lim_{n\to\infty} \left\| (\int_\Omega f \, \mathrm{d}P)x_n \right\|^2 = \|y\|^2\,. \end{aligned}$$

Letting $M \to \infty$ we get $x \in D_f$. Note that for each $z \in D_f$,

$$\begin{aligned}\langle y, z\rangle &= \lim_{n\to\infty} \langle (\int_\Omega f \,\mathrm{d}P)x_n, z\rangle = \lim_{n\to\infty} \langle x_n, (\int_\Omega \bar{f} \,\mathrm{d}P)z\rangle \\ &= \langle x, (\int_\Omega \bar{f} \,\mathrm{d}P)z\rangle = \langle (\int_\Omega f \,\mathrm{d}P)x, z\rangle ,\end{aligned}$$

then we get $y = (\int_\Omega f \,\mathrm{d}P)x$ since D_f is dense in H as desired. $\square$

Lemma 10.22. *Suppose $f : \Omega \to \mathbb{C}$ is measurable. Let $y \in D_f$ and $z = (\int_\Omega f \,\mathrm{d}P)y$. Then for each $x \in H$,*

$$\mathrm{d}P_{z,x} = f \,\mathrm{d}P_{y,x} \ ; \ \mathrm{d}P_{x,z} = \bar{f} \,\mathrm{d}P_{x,y} \,.$$

Proof. Note that for all $B \in \mathcal{F}$ we have

$$\begin{aligned}P_{z,x}(B) &= \langle P(B)z, x\rangle = \langle P(B)(\int_\Omega f \,\mathrm{d}P)y, x\rangle \\ &= \lim_{n\to\infty} \langle P(B)(\int_\Omega f 1_{\{|f|\le n\}} \,\mathrm{d}P)y, x\rangle = \lim_{n\to\infty} \langle (\int_B f 1_{\{|f|\le n\}} \,\mathrm{d}P)y, x\rangle \\ &= \lim_{n\to\infty} \int_B f 1_{\{|f|\le n\}} \,\mathrm{d}P_{y,x} = \int_B f \,\mathrm{d}P_{y,x} \,,\end{aligned}$$

where the last equality follows from the dominated convergence theorem. Thus $\mathrm{d}P_{z,x} = f \,\mathrm{d}P_{y,x}$. By the same argument $\mathrm{d}P_{x,z} = \bar{f} \,\mathrm{d}P_{x,y}$. $\square$

Theorem 10.23. *Let f, g be complex-valued measurable function on Ω. Then the following statements hold.*

(a) *For nonzero complex number α, β, we have*

$$\alpha \int_\Omega f \,\mathrm{d}P + \beta \int_\Omega g \,\mathrm{d}P \subset \int_\Omega \alpha f + \beta g \,\mathrm{d}P \,.$$

(b) $D\left(\int_\Omega f \,\mathrm{d}P \int_\Omega g \,\mathrm{d}P\right) = D_g \cap D_{fg}$, *and*

$$\int_\Omega f \,\mathrm{d}p \int_\Omega g \,\mathrm{d}P \subset \int_\Omega fg \,\mathrm{d}P \,.$$

Hence the equality holds ifd $D_{fg} \subset D_g$, for example $g \in L^\infty(P)$.

(c) $D_f = H$ *if and only if* $f \in L^\infty(P)$.

(d) *There holds*

$$\left(\int_\Omega f \, \mathrm{d}P\right)^* = \int_\Omega \bar{f} \, \mathrm{d}P.$$

Particularly, if f *is real-valued, then* $(\int_\Omega f \, \mathrm{d}P, D_f)$ *is self-adjoint.*

Proof. We prove part (a). Clearly $D_f \cap D_g \subset D_{f+g}$. For $x \in D_f \cap D_g$ and $y \in H$ we have

$$\begin{aligned}\langle \int_\Omega (\alpha f + \beta g) \, \mathrm{d}Px, y\rangle &= \int_\Omega (\alpha f + \beta g) \, \mathrm{d}P_{x,y} \\ = \alpha \int_\Omega f \, \mathrm{d}P_{x,y} + \beta \int_\Omega f \, \mathrm{d}P_{x,y} &= \langle \alpha \int_\Omega f \, \mathrm{d}Px + \beta \int_\Omega g \, \mathrm{d}Px, y\rangle \,.\end{aligned}$$

Thus part (a) holds.

We prove part (b). Take $x \in D_g$ and let $y = (\int_\Omega g \, \mathrm{d}P)x$. Then by Lemma 10.22, $\mathrm{d}P_{y,y} = |g|^2 \, \mathrm{d}P_{x,x}$ and hence

$$\int_\Omega |f|^2 \, \mathrm{d}P_{y,y} = \int_\Omega |fg|^2 \, \mathrm{d}P_{x,x} \,.$$

So $y \in D_f$ if and only if $x \in D_{fg}$. That is $D\left(\int_\Omega f \, \mathrm{d}P \int_\Omega g \, \mathrm{d}P\right) = D_g \cap D_{fg}$. Now assume in addition that $x \in D_{fg}$, then for $z \in H$, by Lemma 10.22 again

$$\langle (\int_\Omega f \, \mathrm{d}P \int_\Omega g \, \mathrm{d}P)x, z\rangle = \int_\Omega f \, \mathrm{d}P_{y,z} = \int_\Omega fg \, \mathrm{d}P_{x,z} = \langle (\int_\Omega fg \, \mathrm{d}P)x, z\rangle$$

Thus $(\int_\Omega f \, \mathrm{d}P \int_\Omega g \, \mathrm{d}P)x = (\int_\Omega fg \, \mathrm{d}P)x$ as desired.

We prove part (c). It suffice to show that if $D_f = H$ then $f \in L^\infty(P)$. By the closed graph theorem, if $D_f = H$ then $\int_\Omega f \, \mathrm{d}P \in \mathcal{B}(H)$. Note that by part (b) for $n \geq 1$,

$$\int_\Omega f 1_{\{|f| \leq n\}} \, \mathrm{d}P = (\int_\Omega f \, \mathrm{d}P) \; P(|f| \leq n) \,.$$

Thus

$$\|f 1_{\{|f| \leq n\}}\|_\infty = \left\| \int_\Omega f 1_{\{|f| \leq n\}} \, \mathrm{d}P \right\| \leq \left\| \int_\Omega f \, \mathrm{d}P \right\| \,.$$

Letting $n \to \infty$, we are done.

We prove (d). On the one hand for $y \in D_{\bar{f}} = D_f$, there holds, for each $x \in D_f$

$$\begin{aligned}\langle (\int_\Omega f \,\mathrm{d}P)x, y\rangle &= \lim_{n\to\infty} \langle (\int_\Omega f 1_{\{|f|\le n\}} \,\mathrm{d}P)x, y\rangle \\ &= \lim_{n\to\infty} \langle x, (\int_\Omega \bar{f} 1_{\{|f|\le n\}} \,\mathrm{d}P)y\rangle = \langle x, (\int_\Omega \bar{f} \,\mathrm{d}P)y\rangle\,,\end{aligned}$$

and hence $\int_\Omega \bar{f} \,\mathrm{d}P \subset \left(\int_\Omega f \,\mathrm{d}P\right)^*$. On the other hand, for $y \in D\left(\int_\Omega f \,\mathrm{d}P\right)^*$, we have

$$\begin{aligned}&\int_\Omega |f^2| 1_{\{|f|\le n\}} \,\mathrm{d}P_{y,y} = \left\| (\int_\Omega \bar{f} 1_{\{|f|\le n\}} \,\mathrm{d}P)y \right\|^2 \\ &= \left\| (\int_\Omega f 1_{\{|f|\le n\}} \,\mathrm{d}P)^* y \right\|^2 = \left\| (\int_\Omega f \,\mathrm{d}P\, P(|f| \le n))^* y \right\|^2 .\end{aligned}$$

By part (c) of Exercise 10.1, we have

$$P(|f| \le n)(\int_\Omega f \,\mathrm{d}P)^* \subset (\int_\Omega f \,\mathrm{d}P\, P(|f| \le n))^* .$$

Combine this with the hypothesis that $y \in D\left(\int_\Omega f \,\mathrm{d}P\right)^*$ we get

$$\int_\Omega |f^2| 1_{\{|f|\le n\}} \,\mathrm{d}P_{y,y} = \left\| P(|f| \le n)(\int_\Omega f \,\mathrm{d}P)^* y \right\|^2 .$$

Letting $n \to \infty$ we get $y \in D_f$ as desired. We are done. $\square$

Lemma 10.24. *Let f be a complex-valued measurable function on Ω. Then*

$$\mathrm{Ker}(\int_\Omega f \,\mathrm{d}P) = \mathrm{Ran}(P(f = 0))\,.$$

Proof. Firstly, for $x \in \mathrm{Ran}(P(f = 0))$ we have $P(f = 0)x = x$, then

$$(\int_\Omega f \,\mathrm{d}P)x = (\int_\Omega f \,\mathrm{d}P)P(f = 0)x = \left(\int_\Omega f 1_{\{f=0\}} \,\mathrm{d}P\right) x = 0.$$

On the other hand, if $x \in \mathrm{Ker}(\int_\Omega f \,\mathrm{d}P)$, then

$$\left\| (\int_\Omega f \,\mathrm{d}P)x \right\|^2 = \int_\Omega |f|^2 \,\mathrm{d}P_{x,x} = 0\,.$$

Thus $P_{x,x}(|f|>0)=0$, i.e, $P(|f|>0)x=0$. Therefore we have

$$x=P(|f|>0)x+P(f=0)x=P(|f|>0)x\in \mathrm{Ran}(P(f=0))\,.$$

We are done. $\square$

Theorem 10.25. *Let f be a complex-valued measurable function on Ω. Then*

$$\sigma\left(\int_\Omega f\,\mathrm{d}P\right)=\mathrm{ess.\,im}(f)\ ;\ \sigma_r\left(\int_\Omega f\,\mathrm{d}P\right)=\emptyset\,;$$
$$\sigma_p\left(\int_\Omega f\,\mathrm{d}P\right)=\{\lambda\in\mathrm{ess.\,im}(f):P(f=\lambda)\neq 0\}\ ;$$
$$\sigma_c\left(\int_\Omega f\,\mathrm{d}P\right)=\{\lambda\in\mathrm{ess.\,im}(f):P(f=\lambda)=0\}\ .$$

Proof. Step 1. We show that $\sigma\left(\int_\Omega f\,\mathrm{d}P\right)\subset\mathrm{ess.\,im}(f)$. Indeed for $\lambda\notin\mathrm{ess.\,im}(f)$, there exists $r>0$ so that $P(f\in B(\lambda,r))=0$. Then we define

$$g=\frac{1}{\lambda-f}1_{\{|\lambda-f|\geq r\}}\,.$$

Clear $g\in L^\infty(P)$, and $(\lambda-f)g=1$ P-a.e. on Ω. By part (b) of Theorem 10.23,

$$(\lambda I-\int_\Omega f\,\mathrm{d}P)\int_\Omega g\,\mathrm{d}P=I\,,$$

and

$$\int_\Omega g\,\mathrm{d}P\ (\lambda I-\int_\Omega f\,\mathrm{d}P)=I_{D_f}\,.$$

Thus $\lambda\in\varrho(\int_\Omega g\,\mathrm{d}P)$.

Step 2. If $\lambda\in\mathrm{ess.\,im}(f)$ and $P(f=\lambda)\neq 0$, then λ is an eigenvalue of T by Lemma 10.24 since

$$\mathrm{Ker}(\lambda I-\int_\Omega f\,\mathrm{d}P)=\mathrm{Ran}(P(f=\lambda))\,.$$

Step 3. We show that if $\lambda\in\mathrm{ess.\,im}(f)$ and $P(f=\lambda)=0$, then $\lambda\in\sigma_c(\int_\Omega f\,\mathrm{d}P)$. Firstly $\mathrm{Ker}(\lambda I-\int_\Omega f\,\mathrm{d}P)=\mathrm{Ran}(P(f=\lambda))=\{0\}$, and

$\operatorname{Ker}(\bar{\lambda} I - \int_\Omega \bar{f}\,\mathrm{d}P) = \operatorname{Ran}(P(f=\lambda)) = \{0\}$. Then we have

$$\operatorname{Ran}(\lambda I - \int_\Omega f\,\mathrm{d}P)^\perp = \operatorname{Ker}(\bar{\lambda} I - \int_\Omega f\,\mathrm{d}P) = \{0\}$$

and hence $\operatorname{Ran}(\lambda I - \int_\Omega f\,\mathrm{d}P)$ is dense in H.

It remains to show that $(\lambda I - \int_\Omega f\,\mathrm{d}P)^{-1}$ is not bounded. For each $n \geq 1$, since $P(|f-\lambda| \leq \frac{1}{n}) \neq 0$, there is $x_n \in \operatorname{Ran}(P(|f-\lambda| \leq \frac{1}{n}))$ with $\|x_n\| = 1$. Then

$$(\lambda I - \int_\Omega f\,\mathrm{d}P)x_n = (\int_\Omega (\lambda - f)1_{\{|f-\lambda|\leq \frac{1}{n}\}}\,\mathrm{d}P)x_n$$

and hence

$$\left\|(\lambda I - \int_\Omega f\,\mathrm{d}P)x_n\right\|^2 = \int_\Omega |\lambda - f|^2 1_{\{|f-\lambda|\leq\frac{1}{n}\}}\,\mathrm{d}P_{x_n,x_n} \leq \frac{1}{n^2}\,.$$

We are done. $\square$

The following theorem is sometimes called the change of measure principle.

Lemma 10.26. *Let $\phi : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ be measurable. Then $E := P \circ \phi^{-1}$ defined by*

$$E(A) := P(\phi^{-1}(A)) \quad \textit{for all} \quad x \in \mathcal{F}\,,$$

is a projection-valued measure on $(\Omega', \mathcal{F}')$. Moreover, for measurable function $f : (\Omega', \mathcal{F}') \to \mathbb{C}$, we have

$$\int_{\Omega'} f\,\mathrm{d}E = \int_\Omega f \circ \phi\,\mathrm{d}P\,.$$

Proof. We only show the two operators are the same one. First of all, note that for each $x, y \in H$,

$$E_{x,y}(A) = P_{x,y}(\phi^{-1}(A)) \quad \text{for all} \quad A \in \mathcal{F}\,.$$

Then

$$\int_{\Omega'} |f|^2\,\mathrm{d}E_{x,x} = \int_\Omega |f \circ \phi|^2\,\mathrm{d}P_{x,x}\,.$$

Thus the two operator has the same domain. On the other hand, for $x \in D_f$ and $y \in H$

$$\langle (\int_{\Omega'} f \,\mathrm{d}E)x, y\rangle = \int_{\Omega'} f \,\mathrm{d}E_{x,y} = \int_{\Omega} f \circ \phi \,\mathrm{d}P_{x,y}\,,$$

and the desired result follows. $\square$

10.4.2 The Spectrum Theorem

Now we are ready to give the central result of this section.

Theorem 10.27. *To every self-adjoint operator* $(T, D(T))$ *in* H *corresponds a unique projection-valued measure* E *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *such that*

$$T = \int_{\mathbb{R}} t\, E(\mathrm{d}t)\,.$$

Moreover, $\mathrm{supp}(E) = \sigma(T)$*.*

Proof. Let U be the Cayley transform of T. Then U is a unitary operator on H. Denote by P the spectrum decomposition of U. Then $(S^1, \mathcal{B}(S^1), P)$ is a spectrum measure space,

$$U = \int_{S^1} e^{i\theta} P(\mathrm{d}\theta)\,;$$

and $\mathrm{supp}(P) = \sigma(U)$. Since $\mathrm{Ker}(I - U) = \{0\}$, we have $P(\{1\}) = 0$ by Theorem 9.47. Thus we set

$$f(e^{i\theta}) = i\frac{1 + e^{i\theta}}{1 - e^{i\theta}}\; 1_{S^1\setminus\{1\}}(e^{i\theta})\,.$$

Then $f : S^1 \to \mathbb{R}$ is measurable. Let $E \coloneqq P \circ f^{-1}$ be the projection-valued measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and

$$\tilde{T} \coloneqq \int_{S^1} f(e^{i\theta}) P(\mathrm{d}\theta) = \int_{\mathbb{R}} t\; E(\mathrm{d}t)\,.$$

Then by Theorem 10.23, $\tilde{T}$ is self-adjoint. Moreover, since $f(e^{i\theta})(1 - e^{i\theta}) = i(1 + e^{i\theta})$ P-a.e. on S^1, by Theorem 10.23 again we have

$$\tilde{T}(I - U) = i(I + U)\,.$$

Since U is the Cayley transform of T,we have $R(I-U) = D(T)$ and $T(I-U) = i(I+U)$, thus we get

$$T \subset \tilde{T}\,.$$

Since self-adjoint operator is maximal symmetric, we conclude that $T = \tilde{T} = \int_{S^1} f(e^{i\theta})P(\mathrm{d}\theta) = \int_{\mathbb{R}} t\ E(\mathrm{d}t)$. Besides, by Theorem 10.25

$$\operatorname{supp}(E) = \operatorname{supp}(P \circ f^{-1}) = \operatorname{ess.im}(f) = \sigma(T)\,.$$

The uniqueness of E follows immediately from the uniqueness of P. □

Remark 10.4. A densely defined closed linear operator $(N, D(N))$ in H is said to be normal if $NN^* = N^*N$. Indeed there are corresponding spectrum theorem for (unbounded) normal operators in Hilbert space, see Theorem 13.33 of *Functional Analysis* by Rudin.

As an easy consequence of the spectrum theorem is the following analogy to part (b) of Theorem 9.53:

Exercise 10.7. Assume that T is a self-adjoint on H. Then T is nonnegative, that is $\langle Tx, x\rangle \geq 0$ for all $x \in D(T)$, if and only if $\sigma(T) \subset [0, \infty)$.

Exercise 10.8. Assume that T is a self-adjoint on H with spectrum decomposition E. Use Theorem 10.23 to show that for each $k \geq 1$,

$$T^k = \int_{\mathbb{R}} t^k\, E(\mathrm{d}t)\,.$$

Recall that in Section 9.3.3 we have discussed the spectrum projection. Now we can give a more beautiful result for self-adjoint operators.

Proposition 10.28. *Let $(T, D(T))$ be a self-adjoint in H with spectrum decomposition E. Suppose that $\sigma(T)$ can be decomposed as the union of N pairwise disjoint closed components:*

$$\sigma(T) = \Sigma_1 \cup \cdots \cup \Sigma_N, \quad \Sigma_j \cap \Sigma_k = \emptyset \quad \textit{if } k \neq j\,.$$

Let $U_j \subset \mathbb{C}$ be disjoint open sets such that $\Sigma_j \subset U_j$ for $j = 0, 1, \cdots, N$. Let γ_j surrounds Σ_j in U_j, then

$$E(\Sigma_j) = \frac{1}{2\pi i}\int_{\gamma_j} (\zeta I - T)^{-1}\,\mathrm{d}\zeta$$

Proof. Note that for all $x, y \in H$, by Fubini's theorem

$$\begin{aligned}
&\langle(\frac{1}{2\pi i}\int_{\gamma_j}(\zeta I - T)^{-1}\,\mathrm{d}\zeta)x, y\rangle = \frac{1}{2\pi i}\int_{\gamma_j}\langle(\zeta I - T)^{-1}x, y\rangle\,\mathrm{d}\zeta \\
&= \frac{1}{2\pi i}\int_{\gamma_j}\int_{\mathbb{R}}\frac{1}{\zeta - t}E_{x,y}(\mathrm{d}t)\,\mathrm{d}\zeta = \int_{\mathbb{R}}\frac{1}{2\pi i}\int_{\gamma_j}\frac{1}{\zeta - t}\,\mathrm{d}\zeta E_{x,y}(\mathrm{d}t) \\
&= \int_{\mathbb{R}} 1_{\Sigma_j}(t)E_{x,y}(\mathrm{d}t) = E_{x,y}(\Sigma_j) = \langle E(\Sigma_j)x, y\rangle,
\end{aligned}$$

then we are done. □

The following proposition is an immediate consequence of Theorem 10.25 and Theorem 10.27.

Proposition 10.29. *Let $(T, D(T))$ be a self-adjoint in H with spectrum decomposition E. Then*

(a) $\sigma_r(T) = \emptyset$.

(b) $\lambda \in \sigma_p(T)$ *if and only if* $E(\{\lambda\}) \neq 0$.

(c) $\lambda \in \sigma_c(T)$ *if and only if* $E(\{\lambda\}) = 0$ *and* $\lambda \in \operatorname{supp}(E) = \sigma(T)$.

Moreover,

(d) $E(\{\lambda\})$ *is the projection of H onto* $\operatorname{Ker}(\lambda I - T)$, *and hence λ is an eigenvalue of T if and only if* $E(\{\lambda\}) \neq 0$.

(e) Every isolated point of $\sigma(T)$ is an eigenvalue of T.

Definition 10.7. For a self-adjoint operator $(T, D(T))$ in H, the *discrete spectrum* $\sigma_d(T)$ of T is the set of all eigenvalues of T of finite multiplicities which are isolated points of the spectrum $\sigma(T)$. The complement set $\sigma_{\text{ess}}(T) \coloneqq \sigma(T)\backslash\sigma_d(T)$ is called the *essential spectrum* of T.

Recall that an isolated point of the spectrum $\sigma(T)$ is always an eigenvalue of T. Therefore, a number belongs to $\sigma_{\text{ess}}(T)$ of T if and only if it is an accumulation point of $\sigma(T)$ or an eigenvalue of infinite multiplicity.

The following criterion of discrete (essential) spectrum , proved in Theorem 9.51 for bounded normal operators, still holds for unbounded self-adjoint operators.

Exercise 10.9. Let $(T, D(T))$ be a self-adjoint operator in H with spectrum decomposition E. Then $\lambda_0 \in \sigma_d(T)$ if and only if there is an open neighborhood U of λ_0,

$$\dim(\mathrm{Ran}E(U)) < \infty \,.$$

The next proposition contains *Weyl's criterion* for the essential spectrum. It is based on the notion of a singular sequence.

Definition 10.8. Let $\lambda \in \mathbb{R}$. A singular sequence for $(T, D(T))$ at λ is a sequence (x_n) of in $D(T)$ such that

$$\liminf_{n\to\infty} \|x_n\| > 0, \quad \text{w-}\lim_{n\to\infty} x_n = 0, \quad \lim_{n\to\infty} (\lambda I - T)x_n = 0$$

Theorem 10.30 (Weyl's Criterion)**.** *Let $(T, D(T))$ be a self-adjoint in H. For any $\lambda \in \mathbb{R}$, the following statements are equivalent:*

(a) $\lambda \in \sigma_{ess}(T)$.

(b) There exists an orthonormal singular sequence for T at λ.

(c) There exists a singular sequence for T at λ.

Proof. (i) $\Rightarrow$ (ii): If λ is an eigenvalue of infinite multiplicity, then any orthonormal sequence from $N(T - \lambda I)$ is a singular sequence for T at λ.

Now let λ be an accumulation point of $\sigma(T)$. Then there is a sequence (t_n) in $\sigma(T)$ such that $t_n \neq t_k$ if $n \neq k$ and $t_n \to \lambda$. We choose a positive null sequence (ϵ_n) such that the intervals $J_n = (t_n - \epsilon_n, t_n + \epsilon_n)$ are pairwise disjoint. Since $t_n \in \sigma(T)$, $E(J_n) \neq 0$ where $(\mathbb{R}, \mathcal{B}(\mathbb{R}), E)$ is the spectrum decomposition for T. Hence, we can find unit vectors $x_n \in \operatorname{Ran}(E(J_n))$. Since $J_n \cap J_k = \emptyset$ if $n \neq k$, the sequence (x_n) is orthonormal, and hence $x_n \xrightarrow{w} 0$ by Bessel's inequality. Then

$$\begin{aligned}\|(\lambda I - T)x_n\|^2 = \|(\lambda I - T)E(J_n)x_n\|^2 &= \int_{J_n} (\lambda - t)^2 E_{x_n,x_n}(\mathrm{d}t) \\ &\leq \int_{J_n} (|\lambda - t_n| + \epsilon_n)^2 E_{x_n,x_n}(\mathrm{d}t) \leq (|t_n - \lambda| + \epsilon_n)^2 \to 0\end{aligned}$$

as $n \to \infty$. Thus, (x_n) is an orthonormal singular sequence for T at λ.

(ii) $\Rightarrow$ (iii) is trivial.

(iii) $\Rightarrow$ (i): If there exists a singular sequence (x_n) for T at λ, then clearly $\lambda \in \sigma(T)$. Assume to the contrary that $\lambda \in \sigma_d(T)$, then there is $\epsilon > 0$ so that

$$\dim(\operatorname{Ran}(E(J)) < \infty$$

where $J = (\lambda - \epsilon, \lambda + \epsilon)$. Thus $E(J)$, having finite rank, is a compact operator. Since $x_n \xrightarrow{w} 0$ we have $E(J)x_n \to 0$. Note that

$$\|x_n\|^2 = \|E(J)x_n\|^2 + \|E(\mathbb{R}\backslash J)x_n\|^2$$

We compute

$$\begin{aligned}&\|E(\mathbb{R}\backslash J)x_n\|^2 = \int_{\mathbb{R}\backslash J} 1 \; E_{x_n,x_n}(\mathrm{d}t) \\ &\leq \int_{\mathbb{R}} \frac{|t-\lambda|^2}{\epsilon^2} E_{x_n,x_n}(\mathrm{d}t) = \frac{1}{\epsilon^2}\|(\lambda I - T)x_n\|^2 \to 0 \quad \text{as} \quad n \to \infty\,.\end{aligned}$$

Thus $x_n \to 0$, which contracts to the hypothesis $\liminf_n \|x_n\| > 0$. $\square$

10.5 Self-adjoint Extension of Symmetric Operators

(Needs to be modified)

10.6 Perturbations of Self-adjointness and Spectra

(Needs to be modified)

Chapter 11

Operator Semigroups

Introduction

Generally speaking, a *dynamical system* is a family $(T(t))_{t\geq 0}$ of mappings on a set X satisfying

$$\begin{cases} T(t+s) = T(t)T(s) \text{ for all } t, s \geq 0 \\ T(0) = I \end{cases}$$

where I is the identity mapping on X. Here X is viewed as the set of all states of a system, $t \in \mathbb{R}_+ := [0, \infty)$ as time and $T(t)$ as the map describing the change of a state $x \in X$ at time 0 into the state $T(t)x$ at time t. In the linear context, the state space X is a vector space, each $T(t)$ is a linear operator on X, and $(T(t))_{t\geq 0}$ is called a (one-parameter) *operator semigroup*.

The standard situation in which such operator semigroups naturally appear are so-called *Abstract Cauchy Problems* (ACP).

$$\begin{cases} u'(t) = Au(t) \quad \text{for } t \geq 0 \\ u(0) = x \end{cases},$$

where $(A, D(A))$ is a linear operator on a Banach space X. Here, the problem consists in finding a differentiable function $u : \mathbb{R}_+ \to X$ such that (ACP) holds. If for each initial value $x \in X$ a unique solution $u(\cdot, x)$ exists, then

$$T(t)x := u(t, x), \quad t \geq 0, x \in X$$

defines an operator semigroup.

For the "working mathematician," (ACP) is the problem, and $(T(t))_{t\geq 0}$ the solution to be found. The opposite point of view also makes sense: given an operator semigroup (i.e., a dynamical system) $(T(t))_{t\geq 0}$, under what conditions can it be "described" by a differential equation (ACP), and how can the operator A be found?

In some simple and concrete situations the relation between $(T(t))_{t\geq 0}$ and A is given by the formulas

$$T(t) = \mathrm{e}^{tA} \quad \text{and} \quad A = \frac{d}{dt}T(t)\bigg|_{t=0}.$$

In general, a comparably simple relation seems to be out of reach. However, miraculously as it may seem, a simple continuity assumption on the semigroup produces, in the usual Banach space setting, a rich and beautiful theory with a broad and almost universal field of applications. It is the aim of this chapter to develop this theory.

11.1 Strongly Continuous Semigroups

In this chapter we always assume that X is a Banach space over the scalar field $\mathbb{F}$. The following is our basic definition.

Definition 11.1. A family $(T(t))_{t\geq 0}$ of bounded linear operators on X is called a *strongly continuous (one-parameter) semigroup* if it satisfies the functional equation

$$\begin{cases} T(t+s) = T(t)T(s) & \text{for all } t, s \geq 0 \\ T(0) = I \end{cases} \tag{FE}$$

and is strongly continuous in the following sense: For every $x \in X$ the orbit maps

$$\xi_x : t \mapsto \xi_x(t) := T(t)x \tag{SC}$$

are continuous from $\mathbb{R}_+$ into X.

Remark 11.1. The property (SC) can also be expressed by saying that the map $t \mapsto T(t)$ is continuous from $\mathbb{R}_+$ into the space $(\mathcal{B}(X), \mathcal{T}_s)$ of all bounded operators on X endowed with the strong operator topology $\mathcal{T}_s$.

If these properties hold for $\mathbb{R}$ instead of $\mathbb{R}_+$, we call $(T(t))_{t\in\mathbb{R}}$ a *strongly continuous (one-parameter) group* on X.

11.1.1 Basic Properties

Our first goal is to facilitate the verification of the strong continuity (SC) required in Definition 11.1. This is possible thanks to the uniform boundedness principle.

As an easy consequence of this lemma, in combination with the functional equation (FE), we obtain that the continuity of the orbit maps

$$\xi_x : t \mapsto T(t)x$$

at each $t \geq 0$ and for each $x \in X$ is already implied by much weaker properties.

Proposition 11.1. *For a semigroup $(T(t))_{t\geq 0}$ on a Banach space X, the following assertions are equivalent.*

- *(a)* *$(T(t))_{t\geq 0}$ is strongly continuous.*
- *(b)* *$(T(t))_{t\geq 0}$ is strongly continuous at* 0*, in other words,* $\lim_{t\downarrow 0} T(t)x = x$ *for all $x \in X$.*
- *(c)* *There exist $t_0 > 0$, $M \geq 1$, and a dense subset $D \subset X$ such that (ci) $\|T(t)\| \leq M$ for all $t \in [0, t_0]$; (cii)* $\lim_{t\downarrow 0} T(t)x = x$ *for all $x \in D$.*

Proof. Clearly (a) implies (b).

We claim that (b) and (c) are equivalent. Indeed if (c) holds, then by the Banach-Steinhaus theorem, we conclude that $T(t)$ converges strongly to I as $t \downarrow 0$. Thus (b) follows. Now if (b) holds, we claim that for each $t_0 > 0$, $\{T(t)\}_{0 \le t \le t_0}$ is uniformly bounded. To see this, fix $x \in X$. Since $T(t) \to x$ as $t \downarrow 0$, there is $\delta_x > 0$ so that

$$\|T(t)x\| \le \|x\| + 1 \text{ for all } 0 \le t \le \delta_x \,.$$

For each $t \in [0, t_0]$, there are $n \in \mathbb{N}$ and $0 \le r < \delta_x$ so that $t = n\delta_x + r$. Then

$$\|T(t)x\| \le \|T(\delta_x)\|^n \|T(r)x\| \le \|T(\delta_x)\|^n (\|x\| + 1)\,.$$

Therefore we conclude that $\sup_{t \in [0,t_0]} \|T(t)x\| < \infty$. By the PUB, $\{T(t)\}_{0 \le t \le 1}$ is uniformly bounded.

It remains to show that (b) implies (a). Assume (b) holds, fix $x \in X$ and $t_0 > 0$. We will show that $t \to T(t)x$ is continuous at t_0. Then (a) follows. The right continuity is easy: we compute

$$\lim_{h \downarrow 0} \|T(t_0 + h)\, x - T(t_0)\, x\| \le \|T(t_0)\| \cdot \lim_{h \downarrow 0} \|T(h)x - x\| = 0\,.$$

For the right continuity, we have the estimate

$$\|T(t_0 - h)\, x - T(t_0)\, x\| \le \|T(t_0 - h)\| \cdot \|x - T(h)x\|, \text{ for } h \ge 0\,.$$

This implies the left continuity since $\{T(t)\}_{0 \le t \le t_0}$ remains uniformly bounded for $t \in [0, t_0]$ by the preceding argument. $\square$

Remark 11.2. Because in many cases the uniform boundedness of the operators $T(t)$ for $t \in [0, t_0]$ is obvious, one obtains strong continuity by checking (right) continuity of the orbit maps ξ_x at $t = 0$ for a dense set of "nice" elements $x \in X$ only. We demonstrate the advantage of this procedure in the examples discussed below.

The following surprising result, asserts that if we use the weak operator topology instead of the strong operator topology in Remark 11.1 will not change our class of semigroups.

Theorem 11.2. *A semigroup $(T(t))_{t\geq 0}$ on a Banach space X is strongly continuous if and only if it is weakly continuous, i.e., if the mappings*

$$t \mapsto \langle T(t)x, x^* \rangle \ \ ; \ \ [0,\infty) \to \mathbb{F}$$

are continuous for each $x \in X, x^* \in X^*$.

Proof. We have only to show that weak continuity implies strong continuity. As a first step, for any $t_0 > 0$, $(T(t) : t \in [0, t_0])$ is bounded for the weak operator topology. By PUB, it is uniformly bounded. Using Proposition 11.1 (c), it suffices to show that

$$E := \left\{ x \in X : \lim_{t \downarrow 0} \|T(t)x - x\| = 0 \right\}$$

is a (strongly) dense subspace of X.

Fix $x \in X$ and $r > 0$. By Remark 11.1, since $t \mapsto T(t)x$ is a continuous mapping $[0, r]$ into (X, τ_w),the integral

$$x_r = \frac{1}{r} \int_0^r T(s)x \, \mathrm{d}s \in X$$

is well-defined in the sense that for each $x^* \in X^*$,

$$\langle x_r, x^* \rangle = \frac{1}{r} \int_0^r \langle T(s)x, x^* \rangle \, \mathrm{d}s \, .$$

Then by Exercise 3.1, for each $t \geq 0$ we have

$$T(t)x_r = \frac{1}{r} \int_0^r T(s+t)x \, \mathrm{d}s = \frac{1}{r} \int_t^{r+t} T(s)x \, \mathrm{d}s \, .$$

We compute

$$
\begin{aligned}
\|T(t)x_r - x_r\| &= \sup_{\|x^*\|\le 1} |\langle T(t)x_r, x^*\rangle - \langle x_r, x^*\rangle| \\
&= \sup_{\|x^*\|\le 1} \left| \frac{1}{r}\int_t^{r+t} \langle T(s)x, x^*\rangle \,\mathrm{d}s - \frac{1}{r}\int_0^r \langle T(s)x, x^*\rangle \,\mathrm{d}s \right| \\
&\le \sup_{\|x^*\|\le 1} \left(\left| \frac{1}{r}\int_r^{r+t} \langle T(s)x, x^*\rangle \,\mathrm{d}s \right| + \left| \frac{1}{r}\int_0^t \langle T(s)x, x^*\rangle \,\mathrm{d}s \right| \right) \\
&\le \frac{2t}{r}\|x\| \sup_{0\le s\le r+t} \|T(s)\| \to 0 \quad \text{as } t \downarrow 0\,.
\end{aligned}
$$

In other words $\lim_{t\downarrow 0} T(t)x_r = x_r$. Now let $D = \{x_r : x \in X, r > 0\}$. Then $D \subset E$.

It remains to show that E is dense in X. Suppose for contradiction that E is not dense in X. Note that E is a linear subspace of X, then by Corollary 1.34, there is a nonzero $x^* \in X^*$ so that $x^* \in E^\perp$. Since $D \subset E$, then for each fixed $x \in X$, we have

$$
\langle x_r, x^*\rangle = \frac{1}{r}\int_0^r \langle T(s)x, x^*\rangle \,\mathrm{d}s = 0 \qquad \text{for all} \;\; r > 0\,.
$$

Letting $r \downarrow 0$, since $s \mapsto \langle T(s)x, x^*\rangle$ is continuous, we conclude that $\langle x, x^*\rangle = 0$. As x is arbitrary, there must be the case that $x^* = 0$, which is a contradiction. Thus E is (strongly) dense in X as required.

Indeed there is an much easier proof for the (strong) density of E. Observe that D is weakly dense in X, we conclude that E is weakly dense in X. Since E is convex, by Theorem 3.34, E is strongly dense in X. □

Exercise 11.1. Show that the semigroup $\{T(t)\}_{t\ge 0}$ is strongly continuous if and only if $t \mapsto T(t)$ from $\mathbb{R}_+$ into $(\mathcal{B}(X), \mathcal{T}_w)$ is weakly continuous at 0.

Exercise 11.2. Let X be a Banach space and let $(T(t))_{t\ge 0}$ be a strongly continuous semigroup. Then the following holds.

(a) The operator $T(t)$ is injective for some $t > 0$ if and only if it is injective for all $t > 0$.

(b) The operator $T(t)$ is surjective for some $t > 0$ if and only if it is surjective for all $t > 0$.

(c) The operator $T(t)$ has a dense image for some $t > 0$ if and only if it has a dense image for all $t > 0$.

Exercise 11.3. Let $(T(t))_{t\geq 0}$ be a semigroup on X. Suppose that $T(t)$ is bijective for some, and hence all, $t > 0$. Let $T(-t) = T(t)^{-1}$ for $t \geq 0$. Then $(T(t))_{t\in\mathbb{R}}$ is a group on X.

11.1.2 Growth Bound

We repeat that for a strongly continuous semigroup $(T(t))_{t\geq 0}$ the finite orbits

$$\{T(t)x : t \in [0, t_0]\}$$

are continuous images of a compact interval, hence compact and therefore bounded for each $x \in X$. So by the uniform boundedness principle each strongly continuous semigroup is uniformly bounded on each compact interval, a fact that implies exponential boundedness on $\mathbb{R}_+$.

Proposition 11.3. *For every strongly continuous semigroup* $(T(t))_{t\geq 0}$,

$$\lim_{t\downarrow 0} \frac{\log \|T(t)\|}{t} = \inf_{t>0} \frac{\log \|T(t)\|}{t} := \omega_0 \in [-\infty, \infty)\,. \tag{11.1}$$

Moreover, we have

$$\omega_0 = \inf \left\{ w \in \mathbb{R} : \begin{array}{c} \textit{there exists } M_w \geq 1 \textit{ such that} \\ \|T(t)\| \leq M_w \mathrm{e}^{wt} \textit{ for all } t \geq 0 \end{array} \right\} \tag{11.2}$$

is called the growth bound (or type) of $\{T(t)\}_{t\geq 0}$.

Proof. Observe that for each $t, s \geq 0$, we have

$$\log \|T(s+t)\| \leq \log \|T(s)\| + \log \|T(t)\|\,.$$

Now fix $t_0 > 0$, for each $t > 0$, there is unique $n \in \mathbb{N}$ and $0 \leq r < t_0$ so that $t = nt_0 + r$. Then

$$\frac{\log \|T(t)\|}{t} \leq \frac{n}{t} \log \|T(t_0)\| + \frac{\log \|T(r)\|}{t}\,.$$

Letting $t \to \infty$ and noting that $t/n \to t_0$, we get

$$\limsup_{t \uparrow \infty} \frac{\log \|T(t)\|}{t} \leq \frac{\log \|T(t_0)\|}{t_0}\,.$$

Since t_0 is arbitrary, we get

$$\limsup_{t \uparrow \infty} \frac{\log \|T(t)\|}{t} \leq \inf_{t_0 > 0} \frac{\log \|T(t_0)\|}{t_0} \leq \liminf_{t \uparrow \infty} \frac{\log \|T(t)\|}{t}\,,$$

and then (11.1) follows. (11.2) is an easy consequence of (11.1). $\square$

It becomes clear in the discussion below, but is presently left as a challenge to the reader that

- $\omega_0 = -\infty$ may occur,
- the infimum in (11.2) may not be attained; i.e, it might happen that no constant M exists such that $\|T(t)\| \leq M\mathrm{e}^{\omega_0 t}$ for all $t \geq 0$, and
- Constants $M > 1$ may be necessary; i.e., no matter how large $w \geq \omega_0$ is chosen, $\|T(t)\|$ will not be dominated by e^{wt} for all $t \geq 0$.

Definition 11.2. A semigroup $(T(t))_{t\geq 0}$ is called *bounded* if there is $M > 0$ so that

$$\|T(t)\| \leq M \qquad \text{for all} \;\; t \geq 0\,;$$

and *contractive* if $M = 1$ is possible. Finally, $(T(t))_{t\geq 0}$ is called *isometric* if $\|T(t)x\| = \|x\|$ for all $t \geq 0$ and $x \in X$.

11.2 Uniformly Continuous Semigroups

In order to create a feeling for the concepts introduced so far, we discuss first the case in which the semigroup $(T(t))_{t\geq 0}$ can be represented as an operator-valued exponential function $\left(\mathrm{e}^{tA}\right)_{t\geq 0}$. Due to this representation, we later consider this case as rather trivial.

For $A \in \mathcal{B}(X)$ we define

$$\mathrm{e}^{tA} := \sum_{n=0}^{\infty} \frac{t^n A^n}{n!} \tag{11.3}$$

for each $t \geq 0$ or $t \in \mathbb{R}$. It follows from the completeness of X that e^{tA} is a well-defined bounded operator on X.

Proposition 11.4. $\left(\mathrm{e}^{tA}\right)_{t\geq 0}$ *is a semigroup on X such that $t \mapsto \mathrm{e}^{tA}\,; \mathbb{R}_+ \to (\mathcal{B}(X), \|\cdot\|)$ is continuous.*

Proof. Because the series $\sum_{k=0}^{\infty} t^k \|A\|^k / k!$ converges, one can show, as for the Cauchy product of scalar series, that

$$\begin{aligned}\sum_{k=0}^{\infty} \frac{t^k A^k}{k!} \cdot \sum_{k=0}^{\infty} \frac{s^k A^k}{k!} &= \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{t^{n-k} A^{n-k}}{(n-k)!} \cdot \frac{s^k A^k}{k!} \\ &= \sum_{n=0}^{\infty} \frac{(t+s)^n A^n}{n!}\,.\end{aligned}$$

This proves that $(\mathrm{e}^{tA})_{t\geq 0}$ is a semigroup. In order to show that $t \mapsto \mathrm{e}^{tA}$ is continuous, we first observe that

$$\mathrm{e}^{(t+h)A} - \mathrm{e}^{tA} = \mathrm{e}^{tA}\left(\mathrm{e}^{hA} - I\right)$$

for all $t, h \in \mathbb{R}$. Therefore, it suffices to show that $\lim_{h\to 0} \mathrm{e}^{hA} = I$. This follows from the estimate

$$\begin{aligned}\left\|\mathrm{e}^{hA} - I\right\| &= \left\|\sum_{k=1}^{\infty} \frac{h^k A^k}{k!}\right\| \\ &\leq \sum_{k=1}^{\infty} \frac{|h|^k \cdot \|A\|^k}{k!} = \mathrm{e}^{|h|\cdot\|A\|} - 1\,.\end{aligned}$$

$\square$

Remark 11.3. In fact, there is no need to restrict the (time) parameter t to $\mathbb{R}_+$. The definition, the continuity, and the functional equation hold for any real and even complex t. Then the map

$$T(\cdot) : t \mapsto \mathrm{e}^{tA}$$

extends to a continuous homomorphism from the additive group $(\mathbb{R}, +)$ into the multiplicative group $\mathscr{G}$ of all invertible elements in the Banach algebra $\mathcal{B}(X)$. We call $\left(\mathrm{e}^{tA}\right)_{t\in\mathbb{R}}$ the (one-parameter) group generated by A.

Semigroups having the continuity property stated in the preceding Proposition are called uniformly continuous. Specifically:

Definition 11.3. Semigroup $\{T(t)\}_{t\geq 0}$ on X is called uniformly continuous if $t \mapsto T(t)$ is a continuous mapping from $\mathbb{R}_+$ into $(\mathcal{B}(X), \|\cdot\|)$.

Exercise 11.4. Show that the semigroup $\{T(t)\}_{t\geq 0}$ is uniformly continuous if and only if $\|T(t) - I\| \to 0$ as $t \downarrow 0$.

Proposition 11.5. *Let $A \in \mathcal{B}(X)$. Then the map $t \mapsto T(t) \coloneqq \mathrm{e}^{tA}$; $\mathbb{R}_+ \to (\mathcal{B}(X), \|\cdot\|)$ is differentiable and satisfies the differential equation*

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}T(t) = AT(t) \textit{ for } t \geq 0\,, \\ T(0) = I\,. \end{cases} \tag{11.4}$$

Conversely, every differentiable function $T(\cdot) : \mathbb{R}_+ \to (\mathcal{B}(X), \|\cdot\|)$ satisfying (11.4) *is already of the form $T(t) = \mathrm{e}^{tA}$ for $A = \frac{\mathrm{d}}{\mathrm{d}t}T(0) \in \mathcal{B}(X)$.*

Proof. We only show that $T(\cdot)$ satisfies (11.4). Because the functional equation implies that, for all $t, h \in \mathbb{R}$,

$$\frac{T(t+h) - T(t)}{h} = \frac{T(h) - I}{h} \cdot T(t)\,.$$

(11.4) is proved if $\lim_{h\to 0} \frac{T(h)-I}{h} = A$. This, however, follows because

$$\begin{aligned}\left\|\frac{T(h)-I}{h} - A\right\| &\leq \sum_{k=2}^{\infty} \frac{|h|^{k-1}\cdot\|A\|^k}{k!} \\ &= \frac{\mathrm{e}^{|h|\cdot\|A\|}-1}{|h|} - \|A\| \to 0 \quad \text{ as } h\to 0\end{aligned}$$

Conversely, if $\{T(t)\}_{t\geq 0}$ solves the differential equation, then we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left(T(t)e^{-tA}\right) &= \left(\frac{\mathrm{d}}{\mathrm{d}t}T(t)\right)e^{-tA} + T(t)\left(\frac{\mathrm{d}}{\mathrm{d}t}e^{-tA}\right) \\ &= T(t)Ae^{-tA} + T(t)(-A)e^{-tA} = 0\,.\end{aligned}$$

Combine this with the initial value condition, we then get that

$$T(t)e^{-tA} = I \quad \text{ for all } t\geq 0\,.$$

Then the desired result follows. □

Theorem 11.6. *Let $(T(t))_{t\geq 0}$ be a uniformly continuous semigroup on X. Then $t\mapsto T(t)$ is a differentiable mapping from $\mathbb{R}_+$ into $(\mathcal{B}(X), \|\cdot\|)$, with the form*

$$T(t) = \mathrm{e}^{tA}, \quad t\geq 0\,,$$

where $A = \frac{\mathrm{d}}{\mathrm{d}t}T(0)\in\mathcal{B}(X)$.

Proof. By the uniform continuity of $s\mapsto T(s)$, for each $h>0$, the integral

$$V(t) := \int_0^t T(s)\,\mathrm{d}s \in \mathcal{B}(X)$$

is well-defined by Theorem 3.49. Besides, we have

$$\frac{1}{t}V(t)\to I \;\text{ in } (\mathcal{B}(X),\|\cdot\|) \quad \text{ as } t\downarrow 0\,.$$

Thus for sufficiently small $h>0$, $V(h)$ is invertible in $\mathcal{B}(X)$.

By Exercise 3.1, we have

$$\begin{aligned}
(T(t)-I)V(h) &= \int_0^h T(t)T(s)\,\mathrm{d}s - \int_0^h T(s)\,\mathrm{d}s \\
&= \int_t^{h+t} T(s)\,\mathrm{d}s - \int_0^h T(s)\,\mathrm{d}s = \int_h^{h+t} T(s)\,\mathrm{d}s - \int_0^t T(s)\,\mathrm{d}s \\
&= \int_0^t (T(h)-I)T(s)\,\mathrm{d}s\,.
\end{aligned}$$

Since $V(h)$ is invertible, by Exercise 3.1, we have

$$T(t)-I = \int_0^t T(s)(T(h)-I)V(h)^{-1}\,\mathrm{d}s\,.$$

Thus we can see that $t \mapsto T(t)$ is a differentiable mapping from $\mathbb{R}_+$ into $\mathcal{B}(X)$. Let $A = (T(h)-I)V(h)^{-1} \in \mathcal{B}(X)$. By Exercise 3.1 again, one can see that A and $(T(t))_{t\geq 0}$ commutes. Then we conclude that

$$\frac{\mathrm{d}}{\mathrm{d}t}T(t) = AT(t) = AT(t)\,.$$

By Proposition 11.5, the desired result follows. $\square$

Remark 11.4. Because definition for e^{tA} works also for $t \in \mathbb{R}$, it follows that each uniformly continuous semigroup can be extended to a uniformly continuous group $\left(\mathrm{e}^{tA}\right)_{t\in\mathbb{R}}$.

From the differentiability of $t \mapsto T(t)$ it follows that for each $x \in X$ the orbit map $\mathbb{R}_+ \ni t \mapsto T(t)x \in X$ is differentiable as well. Therefore, the map $\xi_x(t) := T(t)x$ is the unique solution of the abstract Cauchy problem

$$\begin{cases} u'(t) = Au(t)\ , \text{ for } t \geq 0\,; \\ u(0) = x\,. \end{cases}$$

If we set $X = \mathbb{C}^n$ and A is a $n \times n$ complex matrix, this is what we have learned in the course of ODE.

11.3 Generators and Resolvents

We recall that for a one-parameter semigroup $(T(t))_{t\geq 0}$ on a Banach space X uniform continuity implies differentiability of the map $t \mapsto T(t); \mathbb{R}_+ \to (\mathcal{B}(X), \|\cdot\|)$. The derivative of $T(\cdot)$ at $t = 0$ then yields a bounded operator A for which $T(t) = \mathrm{e}^{tA}$ for all $t \geq 0$.

We now hope that strong continuity of a semigroup $(T(t))_{t\geq 0}$ still implies some differentiability of the orbit maps

$$\xi_x : t \mapsto T(t)x \ ; \ \mathbb{R}_+ \to X \, .$$

In order to pursue this idea we first show, in analogy to Proposition 11.1 and Exercise 11.1 that differentiability of ξ_x is already implied by the differentiability at $t = 0$.

Lemma 11.7. *Let $(T(t))_{t\geq 0}$ be a strongly continuous semigroup on X and fix $x \in X$. Then the orbit map ξ_x differentiable on $\mathbb{R}_+$ if and only if it is differentiable at $t = 0$.*

Proof. If ξ_x is differentiable at 0, then for fixed $t_0 > 0$,

$$\lim_{h\to 0} \frac{T(t_0 + h)x - T(t_0)x}{h} = T(t_0) \lim_{h\to 0} \frac{T(h)x - x}{h} = T(t_0)\xi_x'(0).$$

Thus ξ_x is right differentiable at t_0. On the other hand, by that the uniform boundedness of $\{T(t)\}_{0\leq t\leq t_0}$, we compute

$$\begin{aligned}
&\lim_{h\to 0} \frac{T(t_0)x - T(t_0 - h)x}{h} = T(t_0 - h) \lim_{h\to 0} \frac{T(h)x - x}{h} \\
&= T(t_0)\xi_x'(0) + [T(t_0 - h) - T(t_0)] \lim_{h\to 0} \frac{T(h)x - x}{h} \\
&= T(t_0)\xi_x'(0) + [T(t_0 - h) - T(t_0)] \lim_{h\to 0} [\frac{T(h)x - x}{h} - \xi_x'(0)] \\
&= T(t_0)\xi_x'(0) \, .
\end{aligned}$$

Then the desired result holds. □

On the subspace of X consisting of all those x for which the orbit maps ξ_x are differentiable, the right derivative at $t = 0$ then yields an operator A from which we obtain, in a sense to be specified later, the operators $T(t)$ as the "exponentials e^{tA}". This is already expressed in the choice of the term "generator" in the following definition.

In the following, we set $A_h := \frac{T(h)-I}{h}$ for $h > 0$. Then clearly $A_h \in \mathcal{B}(X)$.

Definition 11.4. The *generator* $(A, D(A))$ for strongly continuous semigroup $(T(t))_{t\geq 0}$ is the operator given by

$$Ax = \lim_{h\downarrow 0} A_h x = \lim_{h\downarrow 0} \frac{T(h)x - x}{h} = \xi_x'(0)\,;$$
$$D(A) = \{x : \xi_x \text{ is differentiable}\} = \{x : \lim_{h\downarrow 0} A_h x \text{ exists }\}\,.$$

It's easy to see that $(A, D(A))$ is a linear operator. To ensure that the operator $(A, D(A))$ has reasonable properties, we proceed as in Theorem 11.2 and Theorem 11.6, we need to look at "average" elements of the orbit map:

$$x_t := \frac{1}{t}\int_0^t \xi_x(s)\,\mathrm{d}s = \frac{1}{t}\int_0^t T(s)x\,\mathrm{d}s \qquad \text{for } x \in X, t > 0\,.$$

Thus first of all we discuss the integral $\int_0^t T(s)x\,\mathrm{d}s$.

Lemma 11.8. *For the generator $(A, D(A))$ of a strongly continuous semigroup $(T(t))_{t\geq 0}$, the following properties hold.*

(i) For every $t \geq 0$ and $x \in X$, one has $\int_0^t T(s)x\,\mathrm{d}s \in D(A)$ and

$$T(t)x - x = A\int_0^t T(s)x\,\mathrm{d}s\,. \tag{11.5}$$

(ii) If $x \in D(A)$, then $T(t)x \in D(A)$ and

$$\frac{\mathrm{d}}{\mathrm{d}t}T(t)x = T(t)Ax = AT(t)x \qquad \textit{for all } t \geq 0.$$

Particularly,

$$T(t)x - x = \int_0^t T(s)Ax\,\mathrm{d}s \qquad \textit{for all } t \geq 0. \tag{11.6}$$

(iii) Let x, y in X. If

$$T(t)x - x = \int_0^t T(s)y \, \mathrm{d}s \qquad \text{for all } t \geq 0,$$

then we have $x \in D(A)$ and $y = Ax$.

Proof. To show part (i), note that

$$\begin{aligned} A_h \int_0^t T(s)x \, \mathrm{d}s &= \frac{1}{h}\left(\int_0^t T(s+h)x \, \mathrm{d}s - \int_0^t T(s)x \, \mathrm{d}s\right) \\ &= \frac{1}{h}\int_h^{t+h} T(s)x \, \mathrm{d}s - \frac{1}{h}\int_0^t T(s)x \, \mathrm{d}s \\ &= \frac{1}{h}\int_t^{t+h} T(s)x \, \mathrm{d}s - \frac{1}{h}\int_0^h T(s)x \, \mathrm{d}s \, . \end{aligned}$$

Thus letting $h \downarrow 0$, by Exercise 3.1 and the strong continuity, we get

$$\lim_{h \downarrow 0} A_h \int_0^t T(s)x \, \mathrm{d}s = T(t)x - x \, .$$

Hence part (i) follows.

To show part (ii), note that $A_h T(t) = T(t) A_h$ for all $h, t > 0$. Thus for $x \in D(A)$, we have

$$\lim_{h \downarrow 0} A_h T(t)x = T(t) \lim_{h \downarrow 0} A_h x = T(t)Ax \, .$$

By the definition of $(A, D(A))$, $T(t)x \in D(A)$ and $AT(t)x = T(t)Ax$. Since $x \in D(A)$, $t \mapsto T(t)x$ is differentiable on $\mathbb{R}_+$ and

$$\frac{\mathrm{d}}{\mathrm{d}t} T(t)x = \lim_{h \downarrow 0} \frac{T(t+h)x - T(t)x}{h} = \lim_{h \downarrow 0} T(t) A_h x = T(t)Ax \, .$$

Then by Newton-Leibniz formula formula,

$$T(t)x - x = \int_0^t T(s)Ax \, \mathrm{d}s \, .$$

Indeed we can use (11.5) to show (11.6). Since A and $(T(t))_{t \geq 0}$ commute on $D(A)$, it's easy to see that for $x \in D(A)$, $t \mapsto T(t)x$ is a continuous mapping form $\mathbb{R}_+$ into $(D(A), \|\cdot\|_A)$. Thus by Exercise 3.1, since A is a bounded linear operator on $(D(A), \|\cdot\|_A)$, we conclude that

$$\int_0^t T(t)Ax\,\mathrm{d}s = A\int_0^t T(t)x\,\mathrm{d}s\,.$$

as desired.

To show part (iii), since $s \mapsto T(s)y$ is continuous, we have

$$\lim_{h\downarrow 0} A_h x = \lim_{h\downarrow 0} \frac{1}{h}\int_0^h T(s)y\,\mathrm{d}s = y\,.$$

Thus by definition $x \in D(A)$ and $Ax = y$. □

Part (iii) is an easier way to show some element x in X belongs to $D(A)$. For example, we shall use it to prove that for a semigroup, weak differentiability is equivalent to the strong differentiability:

Proposition 11.9. *Let $(T(t))_{t\geq 0}$ be a strongly continuous semigroup on X with generator $(A, D(A))$. Let x, y in X. If*

$$w\text{-}\lim_{h\downarrow 0} \frac{T(h)x - x}{h} = y\,.$$

Then $x \in D(A)$ and $Ax = y$.

Proof. By the same argument in the proof of Lemma 11.7, we conclude that the orbits map $\xi_x : t \mapsto T(t)x$ is weakly differentiable, i.e., for each $x^* \in X^*$, $t \mapsto \langle T(t)x, x^*\rangle$ is a differentiable mapping from $\mathbb{R}_+$ into the scalar field of X, with

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle T(t)x, x^*\rangle = \langle T(t)y, x^*\rangle \qquad \text{for all } t \geq 0\,.$$

By Lemma 11.8, it suffices to show that

$$T(t)x - x = \int_0^t T(s)y\,\mathrm{d}s$$

for all $t \geq 0$. Observe that for all $x^* \in X^*$

$$\begin{aligned}\langle T(t)x - x, x^*\rangle = \langle T(t)x, x^*\rangle - \langle x, x^*\rangle &= \int_0^t \frac{\mathrm{d}}{\mathrm{d}s}\langle T(s)x, x^*\rangle\,\mathrm{d}s \\ &= \int_0^t \langle T(s)y, x^*\rangle\,\mathrm{d}s = \langle \int_0^t T(s)y\,\mathrm{d}s, x^*\rangle\,.\end{aligned}$$

then the desired result follows. $\square$

Lemma 11.10. *Let $J \subset \mathbb{R}$ be an interval and $P, Q : J \to \mathcal{B}(X)$ be two strongly continuous operator-valued functions. If $P(\cdot)x : J \to X$ and $Q(\cdot)x : J \to X$ are differentiable for all $x \in D$ for some subspace D of X and D invariant under Q. Then $(PQ)(\cdot)x : J \to X; t \mapsto P(t)Q(t)x$ is differentiable for every $x \in D$ and for $t_0 \in J$,*

$$\frac{\mathrm{d}}{\mathrm{d}t}(P(\cdot)Q(\cdot)x)(t_0) = \frac{\mathrm{d}}{\mathrm{d}t}\left(P(\cdot)Q(t_0)x\right)(t_0) + P(t_0)\left(\frac{\mathrm{d}}{\mathrm{d}t}Q(\cdot)x\right)(t_0)$$

Proof of Lemma 11.10. Fix $x \in D$ and $t_0 \in J$. For $h \in \mathbb{R}$ with $t_0 + h \in J$, note that

$$\begin{aligned}&\frac{P(t_0+h)Q(t_0+h)x - P(t_0)Q(t_0)x}{h}\\ &= P(t_0+h)\frac{Q(t_0+h)x - Q(t_0)x}{h} + \frac{P(t_0+h) - P(t_0)}{h}Q(t_0)x\end{aligned}$$

On the one hand, since D is invariant under Q, we have $Q(t_0)x \in D$ and hence

$$\frac{P(t_0+h) - P(t_0)}{h}Q(t_0)x \to \frac{\mathrm{d}}{\mathrm{d}t}\left(P(\cdot)Q(t_0)x\right)(t_0)\ .$$

as $h \to 0$ with $t_0 + h \in J$. On the other hand, we compute

$$\begin{aligned}&P(t_0+h)\frac{Q(t_0+h)x - Q(t_0)x}{h}\\ &= P(t_0)\frac{Q(t_0+h)x - Q(t_0)x}{h} + [P(t_0+h) - P(t_0)]\left(\frac{\mathrm{d}}{\mathrm{d}t}Q(\cdot)x\right)(t_0)\\ &+ [P(t_0+h) - P(t_0)]\left[\frac{Q(t_0+h)x - Q(t_0)x}{h} - \left(\frac{\mathrm{d}}{\mathrm{d}t}Q(\cdot)x\right)(t_0)\right]\ ;\end{aligned}$$

As $h \to 0$ with $t_0 + h \in J$, since $x \in D$ we have

$$P(t_0)\frac{Q(t_0+h)x - Q(t_0)x}{h} \to P(t_0)\left(\frac{\mathrm{d}}{\mathrm{d}t}Q(\cdot)x\right)(t_0)\ ;$$

since $P(\cdot) : J \mapsto \mathcal{B}(X)$ are strongly continuous, we have

$$[P(t_0+h) - P(t_0)]\left(\frac{\mathrm{d}}{\mathrm{d}t}Q(\cdot)x\right)(t_0) \to 0\,;$$

since the strong continuity of $P(\cdot)$ implies the uniform boundeness of $P(\cdot)$ on compact intervals and $x \in D$ we have

$$[P(t_0+h) - P(t_0)] \left[\frac{Q(t_0+h)\,x - Q(t_0)\,x}{h} - \left(\frac{\mathrm{d}}{\mathrm{d}t} Q(\cdot)x \right)(t_0) \right] \to 0\,.$$

Then the desired result follows. □

With the help of Lemma 11.8 and Lemma 11.10, we now show that the generator, although unbounded in general, has nice properties.

Theorem 11.11. *The generator* $(A, D(A))$ *of a strongly continuous semigroup* $(T(t))_{t\geq 0}$ *is a densely defined closed linear operator that determines the semigroup uniquely.*

Proof. Firstly, we show that $D(A)$ is dense in X. For each fixed $x \in X$, consider

$$x_t := \frac{1}{t} \int_0^t T(s)x \,\mathrm{d}s \qquad \text{for } t > 0\,.$$

Then by Lemma 11.8, $x_t \in D(A)$. By the strong continuity of $(T(t))_{t\geq 0}$, $x_t \to= x$ in X as $t \downarrow 0$. Then the density of $D(A)$ follows.

To show that A is closed, assume that (x_n) is a sequence in $D(A)$ and $x_n \to x$, $Ax_n \to y$. By Lemma 11.8, for each n we have

$$T(t)x_n - x_n = \int_0^t T(s)Ax_n \,\mathrm{d}s\,.$$

Then letting $n \to \infty$ we get,

$$T(t)x - x = \int_0^t T(s)y \,\mathrm{d}s\,.$$

By Exercise 11.8 we conclude that $x \in D(A)$ and $y = Ax$ as required.

Finally we show that the generator determines the semigroup uniquely. Suppose that $(S(t))_{t\geq 0}$ is also a strongly continuous semigroup with generator $(A, D(A))$. We show that $T(t) = S(t)$ for all $t > 0$. To this end, since $T(t), S(t) \in \mathcal{B}(X)$ and $D(A)$ is dense in X, it suffices to show that

$$T(t)x = S(t)x \qquad \text{for all } x \in D(A)\,.$$

Observe that, formally we have

$$T(t)x - S(t)x = T(t-s)S(s)x\,\Big|_{s=0}^{t} = \int_0^t \frac{\mathrm{d}}{\mathrm{d}s} T(t-s)S(s)x\,\mathrm{d}s\,, \tag{11.7}$$

Thus we shall consider the mapping $s \mapsto T(t-s)S(s)$ form $[0,t]$ into $\mathcal{B}(X)$. By Lemma 11.10, since $D(A)$ is invariant under $S(s)$, for all $x \in D(A)$, the orbits map $s \mapsto T(t-s)S(s)x$ is differentiable with

$$\frac{\mathrm{d}}{\mathrm{d}s} T(t-s)S(s)x = -T(t-s)AS(s)x + T(t-s)AS(s)x = 0 \quad \text{for all } 0 \le s \le t\,.$$

Thus by (11.7) we get $T(t)x = S(t)x$ as required. □

Combining these properties of the generator with the closed graph theorem gives a new characterization of uniformly continuous semigroups, thus complementing Theorem 11.6.

Corollary 11.12. *For a strongly continuous semigroup $(T(t))_{t\ge0}$ with generator $(A, D(A))$, the following assertions are equivalent.*

(a) The generator A is bounded; i.e., there exists $M > 0$ such that $\|Ax\| \le M\|x\|$ for all $x \in D(A)$.

(b) The domain $D(A)$ is all of X.

(c) The semigroup $(T(t))_{t\ge0}$ is uniformly continuous.

In each case, the semigroup is given by

$$T(t) = \mathrm{e}^{tA} := \sum_{n=0}^{\infty} \frac{t^n A^n}{n!}, \quad t \ge 0\,.$$

11.3.1 The Core for the Generator

Property (b) of the Corollary 11.12 indicates that the domain of the generator contains important information about the semigroup and therefore has to be taken into account carefully. However, in many examples it is often routine to compute the expression Ax for some or even many elements in the domain $D(A)$, although it is difficult to identify $D(A)$ precisely. In these situations, the concept of core helps to distinguish between "small" and "large" subspaces of $D(A)$.

Recall that a subspace D of the domain $D(A)$ of is called a *core* for A if D is dense in $D(A)$ for the graph norm $\|x\|_A := \|x\| + \|Ax\|$. In this case, $(A|_D, D)$ is closable and its closure is exactly $(A, D(A))$. We now state a useful criterion for subspaces to be a core for the generator.

Lemma 11.13. *Let $(A, D(A))$ be the generator of a strongly continuous semigroup $(T(t))_{t\geq 0}$ on X. A subspace D of $D(A)$ that is dense in X and invariant under the semigroup $(T(t))_{t\geq 0}$ is always a core for A.*

Proof. Since A is closed, the $\|\cdot\|_A$-closure of D, namely $\overline{D}^{\|\cdot\|_A}$, is contained in $D(A)$. It remains to show that $D(A) \subset \overline{D}^{\|\cdot\|_A}$.

Fix $x \in D(A)$. As we have pointed in the proof of Lemma 11.8, since $AT(t)x = T(t)Ax$ for all $t \geq 0$, the orbit map $t \mapsto T(t)x \in (D(A), \|\cdot\|_A)$ is continuous. Thus $\int_0^t T(s)x \,\mathrm{d}s \in D(A)$ for each $t \geq 0$ and

$$\left\| x - \frac{1}{t}\int_0^t T(s)x \,\mathrm{d}s \right\|_A \to 0 \quad \text{ as } t \downarrow 0\,.$$

It follows that we have only to show that $\int_0^t T(s)x \,\mathrm{d}s \in \overline{D}^{\|\cdot\|_A}$ for all $t \geq 0$.

Since D is dense in X, we can find a sequence (x_n) in D such that $\|x_n - x\| \to 0$. By the hypothesis that D is invariant under $(T(t))_{t\geq 0}$, we have $T(t)x_n \in D$ for all $t \geq 0$. Then we deduce that the map $s \mapsto T(s)x_n \in (D, \|\cdot\|_A)$ is continuous and hence for each $t \geq 0$,

$$\int_0^t T(s)x_n \,\mathrm{d}s \in \overline{D}^{\|\cdot\|_A}\,.$$

Fix $t > 0$. We claim that

$$\int_0^t T(s)x_n \,\mathrm{d}s \to \int_0^t T(s)x \,\mathrm{d}s \ \text{ in } \ (D(A), \|\cdot\|_A)\,,$$

then the desired result follows.

Let $M = \sup\{\|T(s)\| : 0 \le s \le t\}$. By Lemma 11.8, we compute

$$\begin{aligned}
&\left\| \int_0^t T(s)x_n \,\mathrm{d}s - \int_0^t T(s)x \,\mathrm{d}s \right\|_A \\
&\le tM\|x_n - x\| + \left\| A\int_0^t T(s)x_n \,\mathrm{d}s - A\int_0^t T(s)x \,\mathrm{d}s \right\| \\
&\le tM\|x_n - x\| + \|T(t)x_n - x_n - T(t)x + x\| \\
&\le (tM + M + 1)\|x_n - x\| \to 0 \quad \text{as} \quad n \to \infty\,.
\end{aligned}$$

We are done. □

Important examples of cores are given by the domains $D(A^n)$ of the powers A^n of a generator A.

Theorem 11.14. *For the generator $(A, D(A))$ of a strongly continuous semigroup $(T(t))_{t\ge 0}$ the space*

$$D(A^\infty) := \bigcap_{n\in\mathbb{N}} D(A^n)\,,$$

hence each $D(A^n) := \left\{x \in D\left(A^{n-1}\right) : A^{n-1}x \in D(A)\right\}$, is a core for A.

Proof. Because the space $D(A^\infty)$ is a $(T(t))_{t\ge 0}$ -invariant subspace of $D(A)$, by the preceding lemma, it remains to show that it is dense in X. Let

$$D := \left\{ x_\varphi := \int_0^\infty \varphi(s)T(s)x \,\mathrm{d}s : x \in X, \varphi \in C_c^\infty(\mathbb{R}) \text{ with } \operatorname{supp}(\varphi) \subset (0,\infty) \right\}.$$

Note that since ϕ has compact support, x_φ is well-defined by Theorem 3.49. We claim that (i) $D \subset D(A^\infty)$; (ii) D is dense in X. Then the desired result follows.

To demonstrate (i), note that for $h > 0$, by Exercise 3.1 we have

$$\begin{aligned}
\frac{T(h) - I}{h}x_\varphi &= \frac{1}{h}\int_0^\infty \varphi(s)(T(s+h) - T(s))x \,\mathrm{d}s \\
&= \frac{1}{h}\int_h^\infty \varphi(s-h)T(s)x \,\mathrm{d}s - \frac{1}{h}\int_0^\infty \varphi(s)T(s)x \,\mathrm{d}s \\
&= \int_0^\infty \frac{1}{h}(\varphi(s-h) - \varphi(s))T(s)x \,\mathrm{d}s\,.
\end{aligned}$$

Letting $h \downarrow 0$, by the dominated convergence theorem, we get $x_\varphi \in D(A)$ and

$$Ax_\varphi = -\int_0^\infty \varphi'(s)T(s)x\,\mathrm{d}s = x_{\varphi'}$$

Thus we can see that $x_\varphi \in D(A^\infty)$ and claim (i) follows.

To show (ii), fix $x \in X$. For each n, choose $\varphi_n \in C_c^\infty(\mathbb{R})$ with $\operatorname{supp}(\varphi_n) \subset (0, \frac{1}{n})$, $\varphi_n \geq 0$ and $\int_{\mathbb{R}} \varphi_n = 1$. Then by the dominated convergence theorem,

$$\begin{aligned}\|x - x_{\varphi_n}\| &= \left\| \int_0^\infty \varphi_n(s)[T(s)x - x]\,\mathrm{d}s \right\| \\ &\leq \int_0^\infty \varphi_n(s)\,\|T(s)x - x\|\,\mathrm{d}s \to 0 \quad \text{as } n \to \infty\,.\end{aligned}$$

Thus D is dense in X. We are done. □

11.3.2 Resolvent of the Generator

Our starting points are the following two identities, which are easily derived from their predecessors in Theorem 11.8. We stress that these identities will be used very frequently throughout these notes.

Lemma 11.15. *Let $(A, D(A))$ be the generator of a strongly continuous semigroup $(T(t))_{t\geq 0}$. Then, for every $\lambda \in \mathbb{C}$ and $t > 0$,*

$$\begin{aligned}\mathrm{e}^{-\lambda t}T(t)x - x &= (A - \lambda I)\int_0^t \mathrm{e}^{-\lambda s}T(s)x\,\mathrm{d}s \quad \text{if } x \in X\,; \\ &= \int_0^t \mathrm{e}^{-\lambda s}T(s)(A - \lambda I)x\,\mathrm{d}s \quad \text{if } x \in D(A)\,.\end{aligned}$$

Proof. Observe that the rescaled semigroup $(\mathrm{e}^{-\lambda t}T(t))_{t\geq 0}$ has generator $(A - \lambda I, D(A))$. Then the desired result follows from Lemma 11.8. □

Next, we give an important formula relating the semigroup to the resolvent of its generator.

Theorem 11.16. *Let $(T(t))_{t\geq 0}$ be a strongly continuous semigroup on X and take constants $\omega \in \mathbb{R}, M \geq 1$ such that $\|T(t)\| \leq M\mathrm{e}^{\omega t}$ for $t \geq 0$. For the* *generator $(A, D(A))$ of $(T(t))_{t\geq 0}$ and scalar λ with $\operatorname{Re}\lambda > \omega$, the following properties hold.*

(i) For each $x \in X$, the integral $\int_0^\infty \mathrm{e}^{-\lambda t} T(t)x \,\mathrm{d}t$ is well-defined.

(ii) $\lambda \in \varrho(A)$, and

$$R(\lambda; A)x = \int_0^\infty \mathrm{e}^{-\lambda t} T(t)x \,\mathrm{d}t\,.$$

for each $x \in X$. Moreover,

$$\|R(\lambda, A)\| \leq \frac{M}{\operatorname{Re}\lambda - \omega}\,.$$

(iii) Let $n \in \mathbb{N}$. Then

$$R(\lambda, A)^n x = \frac{1}{(n-1)!}\int_0^\infty t^{n-1}\mathrm{e}^{-\lambda t} T(t)x \,\mathrm{d}t$$

for all $x \in X$. In particular, the estimates

$$\|R(\lambda, A)^n\| \leq \frac{M}{(\operatorname{Re}\lambda - w)^n}\,.$$

The formula for $R(\lambda, A)$ in (ii) is called the *integral representation* of the resolvent. Indeed, by Theorem 4.1, we can deduce that

$$R(\lambda, A) = \int_0^\infty \mathrm{e}^{-\lambda t} T(t) \,\mathrm{d}t \in (\mathcal{B}(X), \mathcal{T}_t)\,.$$

Proof. To prove (i), note that for $n \in \mathbb{N}$, $\int_0^n \mathrm{e}^{-\lambda t} T(t)x \,\mathrm{d}t$ is well-defined, since $t \mapsto \mathrm{e}^{-\lambda t} T(t)x$ is a continuous mapping from $[0, n]$ into X. Observe that

$$\left\|\int_n^{n+p} e^{-\lambda t} T(t)x \,\mathrm{d}t\right\| \leq M\|x\| \int_n^{n+p} e^{-(\operatorname{Re}\lambda - \omega)t} \,\mathrm{d}t$$

for all $n, p \geq 0$. Thus $\left\{\int_0^n \mathrm{e}^{-\lambda t} T(t)x \,\mathrm{d}t\right\}_{n\geq 1}$ is a Cauchy sequence in X. Let

$$\int_0^\infty \mathrm{e}^{-\lambda t} T(t)x \,\mathrm{d}t := \lim_{n\to\infty} \int_0^n \mathrm{e}^{-\lambda t} T(t)x \,\mathrm{d}t\,. \tag{11.8}$$

Then for each $x^* \in X^*$,

$$\begin{aligned}\langle \int_0^\infty \mathrm{e}^{-\lambda t} T(t)x\,\mathrm{d}t, x^*\rangle &= \lim_{n\to\infty} \langle \int_0^n \mathrm{e}^{-\lambda t} T(t)x\,\mathrm{d}t, x^*\rangle \\ = \lim_{n\to\infty} \int_0^n \mathrm{e}^{-\lambda t} \langle T(t)x, x^*\rangle\,\mathrm{d}t &= \int_0^\infty \mathrm{e}^{-\lambda t}\langle T(t)x, x^*\rangle\,\mathrm{d}t\end{aligned}$$

as desired. Thus assertion (i) follows.

To prove (ii), we need to that $\lambda I - A : D(A) \to X$ is bijective, which implies $\lambda \in \varrho(A)$ since A is closed. We have only to check that the inverse of $\lambda I - A$ is given by $x \mapsto \int_0^\infty \mathrm{e}^{-\lambda t} T(t)x\,\mathrm{d}t$. That is

- for each $x \in X$, $\int_0^\infty \mathrm{e}^{-\lambda t} T(t)x\,\mathrm{d}t \in D(A)$ and
$$(\lambda I - A)\int_0^\infty \mathrm{e}^{-\lambda t} T(t)x\,\mathrm{d}t = x\,; \tag{11.9}$$
- for each $x \in D(A)$,
$$\int_0^\infty \mathrm{e}^{-\lambda t} T(t)(\lambda I - A)x\,\mathrm{d}t = x\,. \tag{11.10}$$

By (11.8) and Lemma 11.15, we have

$$\begin{aligned}(\lambda I - A)\int_0^\infty \mathrm{e}^{-\lambda t} T(t)x\,\mathrm{d}t &= \lim_{n\to\infty}(\lambda I - A)\int_0^n \mathrm{e}^{-\lambda t} T(t)x\,\mathrm{d}t \\ &= x - \lim_{n\to\infty} \mathrm{e}^{-\lambda n} T(n)x = x\,; \\ \int_0^\infty \mathrm{e}^{-\lambda t} T(t)(\lambda I - A)x\,\mathrm{d}t &= \lim_{n\to\infty} \int_0^n \mathrm{e}^{-\lambda t} T(t)(\lambda I - A)x\,\mathrm{d}t \\ &= x - \lim_{n\to\infty} \mathrm{e}^{-\lambda n} T(n)x = x\,,\end{aligned}$$

since $\|\mathrm{e}^{-\lambda n} T(n)x\| \le M\mathrm{e}^{-(\operatorname{Re}\lambda - \omega)n}\|x\|$, as required. For the norm of the resolvent, we compute

$$\|R(\lambda; A)\| \le M\int_0^\infty \mathrm{e}^{-(\operatorname{Re}\lambda-\omega)t}\,\mathrm{d}t = \frac{M}{\operatorname{Re}\lambda - \omega}\,.$$

To show (iii), by part (b) of Theorem 6.3, and the dominated convergence theorem, we have

$$\begin{aligned} R(\lambda, A)^n x &= \frac{(-1)^{n-1}}{(n-1)!} \cdot \frac{\mathrm{d}^{n-1}}{\mathrm{d}\lambda^{n-1}} R(\lambda, A)x \\ &= \frac{(-1)^{n-1}}{(n-1)!} \cdot \frac{\mathrm{d}^{n-1}}{\mathrm{d}\lambda^{n-1}} \int_0^\infty \mathrm{e}^{-\lambda t} T(t)x \,\mathrm{d}t \\ &= \frac{1}{(n-1)!} \int_0^\infty t^{n-1} \mathrm{e}^{-\lambda t} T(t)x \,\mathrm{d}t \,. \end{aligned}$$

Finally, the estimate follows from

$$\|R(\lambda, A)^n\| \leq \frac{M}{(n-1)!} \int_0^\infty t^{n-1} \mathrm{e}^{-(\mathrm{Re}\,\lambda - \omega)t} \,\mathrm{d}t = \frac{M}{|\,\mathrm{Re}\,\lambda - \omega|^n} \,.$$

We are done. □

Remark 11.5. Property (ii) in Theorem 11.16 says that the spectrum of a semigroup generator is always contained in a left half-plane. The number determining the smallest such half-plane is an important characteristic of any linear operator and is called the *spectral bound*, defined by

$$\mathrm{s}(A) := \sup\{\mathrm{Re}\,\lambda : \lambda \in \sigma(A)\}$$

Then for a strongly continuous semigroup $(T(t))_{t\geq 0}$ with generator A, one has $-\infty \leq \mathrm{s}(A) \leq \omega_0 < +\infty$, where ω_0 is the growth bound.

The following exercise, though trivial, is very useful when we conclude that an operator is the generator of some semigroup.

Exercise 11.5. Let $(A, D(A))$, $(B, D(B))$ be two closed linear operator and $A \subset B$. If $\varrho(A) \cap \varrho(B) \neq \emptyset$, then $A = B$. (Hint: Use Exercise 0.1.)

Proposition 11.17. *Let $(T(t))_{t\geq 0}$ be a strongly continuous semigroup on X with generator $(A, D(A))$. Then for each $t > 0$ one has*

$$T(t) = s\text{-}\lim_{n\to\infty} \left(I - \frac{t}{n}A\right)^{-n} = s\text{-}\lim_{n\to\infty} \left[\frac{n}{t} R(\frac{n}{t}, A)\right]^n$$

Proof. First of all, we show that the operators $\left[\frac{n}{t}R(\frac{n}{t},A)\right]^n$ are uniformly bounded. Take $\omega \in \mathbb{R}$ and $M > 0$ so that $\|T(t)\| \le Me^{\omega t}$ for all $t \ge 0$. Then by Theorem 11.16, for large n,

$$\left\|\left[\frac{n}{t}R(\frac{n}{t},A)\right]^n\right\| \le M\left(\frac{n/t}{(n/t-\omega)}\right)^n = M\left(1-\frac{\omega t}{n}\right)^{-n}$$

Since $\left(1-\frac{\omega t}{n}\right)^{-n}$ converges as $n \to \infty$, we conclude that $\left[\frac{n}{t}R(\frac{n}{t},A)\right]^n$ are uniformly bounded. Then by the Banach-Steinhaus theorem, it suffices to show that

$$T(t)x = \lim_{n\to\infty}\left(I-\frac{t}{n}A\right)^{-n}x \qquad \text{for } x \in D\,,$$

where D is a dense subspace of X.

Note that formally we have

$$\begin{aligned}\left(I-\frac{t}{n}A\right)^{-n}x - T(t)x &= T(t-s)\left(I-\frac{s}{n}A\right)^{-n}x\Big|_{s=0}^{t} \\ &= \int_0^t \frac{\mathrm{d}}{\mathrm{d}s}\left[T(t-s)\left(I-\frac{s}{n}A\right)^{-n}x\right]\mathrm{d}s\,.\end{aligned}$$

By Lemma 11.10 and part (ii) of Theorem 6.3, since $D(A)$ is invariant under $R(n/t;A)$, for fixed $x \in D(A)$, then the mapping $s \to T(t-s)\left(I-\frac{s}{n}A\right)^{-n}x$ from $[0,t]$ into X is differentiable with

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}s}\left[T(t-s)\left(I-\frac{s}{n}A\right)^{-n}x\right] \\ &= -T(t-s)A\left(I-\frac{s}{n}A\right)^{-n}x + T(t-s)\left(I-\frac{s}{n}A\right)^{-(n+1)}Ax \\ &= T(t-s)\left(I-\frac{s}{n}A\right)^{-(n+1)}\left[\left(I-\frac{s}{n}A\right)-I\right]Ax \\ &= -\frac{s}{n}T(t-s)\left(I-\frac{s}{n}A\right)^{-(n+1)}A^2x\,,\end{aligned}$$

if we assume in addition that $x \in D(A^2)$. Therefore,

$$\begin{aligned}&\left\|\left(I-\frac{t}{n}A\right)^{-n}x - T(t)x\right\| \\ &\le \int_0^t \frac{s}{n}\left\|T(t-s)\left(I-\frac{s}{n}A\right)^{-(n+1)}A^2x\right\|\mathrm{d}s \\ &\le \frac{1}{n}\cdot t\cdot t\cdot Me^{|\omega|t}\cdot K\cdot\|A^2x\|\,,\end{aligned}$$

where K is a real number given by

$$\begin{aligned}\sup_{s\in[0,t];n\geq 1}\left\|\left(I-\frac{s}{n}A\right)^{-(n+1)}\right\| &\leq M\sup_{s\in[0,t];n\geq 1}\left(1-\frac{\omega s}{n}\right)^{-(n+1)}\\ &\leq M\sup_{n\geq 1}\left(1-\frac{|\omega|t}{n}\right)^{-(n+1)} := K<\infty\,.\end{aligned}$$

Letting $n\to\infty$ we get hat

$$T(t)x=\lim_{n\to\infty}\left(I-\frac{t}{n}A\right)^{-n}x \qquad \text{for } x\in D(A^2)\,,$$

Since $D(A^2)$ is dense in X by Theorem 11.14, the desired result follows. $\square$

Remark 11.6. Indeed one can show that for $x\in X$, $\left(I-\frac{t}{n}A\right)^{-n}x$ converges to $T(t)x$ as $h\downarrow 0$ uniformly for t in compact intervals.

To conclude this section, we collect in a diagram the information obtained so far on the relations between a semigroup, its generator, and its resolvent.

$$\begin{tikzcd}
& \big(T(t)\big)_{t\geq 0} \arrow[dl, "Ax=\lim\limits_{t\downarrow 0}\frac{T(t)x-x}{t}"'] \arrow[dr, shift left, "R(\lambda,A)=\int_0^\infty e^{-\lambda t}T(t)\,dt"] & \\
\big(A,D(A)\big) \arrow[rr, shift left, "R(\lambda,A)=(\lambda-A)^{-1}"] & & \big(R(\lambda,A)\big)_{\lambda\in\rho(A)} \arrow[ll, shift left, "A=\lambda-R(\lambda,A)^{-1}"] \arrow[ul, shift left, "T(t)=\lim\limits_{n\to\infty}[{}^n/_t R({}^n/_t,A)]^n"]
\end{tikzcd}$$

Figure 11.1: Semigroup, generator, and resolvent

11.4 Standard Constructions

In this section, we explain how one can construct in various ways new strongly continuous semigroups from a given one. In each case, we try to identify the corresponding generator, its spectrum and resolvent, so that our abstract definitions gain a more concrete meaning.

Example 11.1 (Rescaled Semigroups)**.** Let $(T(t))_{t\geq 0}$ to be a strongly continuous semigroup on X. For any scalar λ and any $\alpha > 0$, we define the *rescaled semigroup* $(S(t))_{t\geq 0}$ by

$$S(t) := \mathrm{e}^{\lambda t} T(\alpha t) \quad \text{ for } \; t \geq 0\,.$$

It's easy to check that $(S(t))_{t\geq 0}$ a strongly continuous semigroup on X. Denote by $(B, D(B))$ the generator of $(S(t))_{t\geq 0}$. Then we asserts that

$$B = \alpha A + \lambda I \quad \text{ with domain } \quad D(A) = D(B)\,.$$

Moreover, $\sigma(B) = \alpha\sigma(A) + \lambda$ and

$$R(\mu; B) = \frac{1}{\alpha} R(\frac{\mu - \lambda}{\alpha}; A) \quad \text{ for } \mu \in \varrho(B)\,.$$

This shows that we can switch quite easily between the original and the rescaled objects.

To prove the assertion, by the symmetry it suffices to show that $D(A) \subset D(B)$ and $Bx = \alpha Ax + \lambda x$ for $x \in D(A)$. We compute

$$\begin{aligned}
&\lim_{h\downarrow 0} \frac{e^{\lambda h} T(\alpha h)x - x}{h} = \lim_{h\downarrow 0} e^{\lambda h} \frac{[T(\alpha h)x - x]}{h} + \lim_{h\downarrow 0} \frac{e^{\lambda h} - 1}{h} x \\
&= \alpha Ax + \lambda x\,, \quad \text{for all } x \in D(A)
\end{aligned}$$

as required.

Taking $\lambda = -\omega_0$ (or $\lambda < -\omega_0$) and $\alpha = 1$ the rescaled semigroup will have growth bound equal to (or less than) zero. This is an assumption we make without loss of generality in many situations.

Example 11.2 (Product Semigroups.)**.** Let $(T(t))_{t\geq 0}$, $(S(t))_{t\geq 0}$ be two strongly continuous semigroup on X with generators $(A, D(A))$ and $(B, D(B))$ respectively. If $(T(t))_{t\geq 0}$ and $(S(t))_{t\geq 0}$ commutes; i.e., $S(t)T(t) = T(t)S(t)$ for all $t \geq 0$, then

(i) The operators $U(t) := S(t)T(t)$ form a strongly continuous semigroup $(U(t))_{t\geq 0}$, called the *product semigroup* of $(T(t))_{t\geq 0}$ and $(S(t))_{t\geq 0}$.

(ii) Denote by $(C, D(C))$ the generator of $(U(t))_{t\geq 0}$. Then $D(A)\cap D(B)$ is a core for C and

$$Cx = Ax + Bx \quad \text{for all} \quad x \in D(A)\cap D(B)\,.$$

In other words, the generator of $(U(t))_{t\geq 0}$ is the closure of $(D(A)\cap D(B), A+B)$.

To prove (i), it suffices to show that $T(s)$ and $S(r)$ commute for all $s, r \geq 0$. To this end, we first take $r = p_1/q$ and $s = p_2/q \in \mathbb{Q}_+$. Then

$$\begin{aligned} S(r)T(s) &= S(1/q)^{p_1} \cdot T(1/q)^{p_2} \\ &= T(1/q)^{p_2} \cdot S(1/q)^{p_1} = T(s)S(r)\,. \end{aligned}$$

Now fix $x \in X$. Since $(s,r) \mapsto S(r)T(s)x$ and $(s,r) \mapsto T(s)S(r)x$ both are continuous mapping from $\mathbb{R}_+^2$ in to X, and they coincide on $\mathbb{Q}_+^2$, we conclude that

$$S(r)T(s)x = T(s)S(r)x \qquad \text{for all} \quad s, r \geq 0$$

as desired.

To prove (ii), firstly we show if $x \in D(A)\cap D(B)$ then $x \in D(C)$ and $Cx = Ax + Bx$. We compute

$$\begin{aligned} \lim_{h\downarrow 0} \frac{T(h)S(h)x - x}{h} &= \lim_{h\downarrow 0} T(h)\frac{S(h)x - x}{h} + \lim_{h\downarrow 0} \frac{T(h)x - x}{h} \\ &= Bx + Ax \end{aligned}$$

as desired. It remains to show that $D(A)\cap D(B)$ is a core for C.

By Lemma 11.13, we have only to prove that $D(A)\cap D(B)$ is invariant under $(U(t))_{t\geq 0}$ and is dense in X. The former is trivial since $T(t)$ and $S(t)$ commutes. To prove the latter, take λ large enough, then we have the representations: $R(\lambda; A) = \int_0^\infty \mathrm{e}^{-\lambda s} T(t)\, \mathrm{d}t$ and $R(\lambda; B) = \int_0^\infty \mathrm{e}^{-\lambda s} S(t)\, \mathrm{d}t$. From these we deduce that

$$R(\lambda; A)R(\lambda; B) = R(\lambda; B)R(\lambda; A)\,.$$

Therefore, $R(\lambda; B)R(\lambda; A)X$ is contained in $D(A) \cap D(B)$. Because both $R(\lambda; A)$ and $R(\lambda; B)$ are continuous and have dense range, we conclude that $D(A) \cap D(B)$ is dense in X as desired.

Example 11.3 (Adjoint Semigroups)**.** Let $(T(t))_{t\geq 0}$ to be a strongly continuous semigroup on X. Then it's easy to see that $(T(t)^*)_{t\geq 0}$ consisting of all adjoint operators $T(t)^*$ on the dual space X^* is a semigroup.In general, it is NOT strongly continuous. An example is provided by the (left) translation group on $L^1(\mathbb{R})$. Its adjoint is the (right) translation group on $L^\infty(\mathbb{R})$, which is not strongly continuous.

However, it is easy to see that $(T(t)^*)_{t\geq 0}$ is always weak*-continuous in the sense that the maps

$$t \mapsto \langle x, T(t)^* x^* \rangle = \langle T(t)x, x^* \rangle$$

are continuous for all $x \in X$ and $x^* \in X^*$. Because on the dual of a reflexive Banach space weak and weak* topology coincide, if we assume X is reflexive, then the adjoint semigroup $(T(t)^*)_{t\geq 0}$ is weakly, and hence by Theorem 11.2 strongly, continuous.

We claim that if X is a Hilbert space and the generator of $(T(t))_{t\geq 0}$ is $(A, D(A))$, then the generator of the adjoint semigroup $(T(t)^*)_{t\geq 0}$ is $(A^*, D(A^*))$.

Denote by $(B, D(B))$ the generator of $(T(t)^*)_{t\geq 0}$. On the one hand, for each $y \in D(B)$,

$$\langle Ax, y \rangle = \lim_{h\downarrow 0} \langle \frac{T(h) - I}{h} x, y \rangle = \lim_{h\downarrow 0} \langle x, \frac{T(h)^* - I}{h} y \rangle = \langle x, By \rangle$$

for all $x \in D(A)$. Thus we have $B \subset A^*$. On the other hand, for $y \in D(A^*)$, we asserts that

$$T(t)^* y - y = \int_0^t T(s)^* A^* y \, \mathrm{d}s \qquad \text{for all } t \geq 0 \,. \tag{11.11}$$

Then by Lemma 11.8, we conclude that $y \in D(B)$ and $By = A^* y$ as desired. To show (11.11), note that for all $x \in D(A)$,

$$\begin{aligned} \langle x, T(t)^* y - y \rangle &= \langle T(t)x - x, y \rangle = \langle \int_0^t AT(s)x \, \mathrm{d}s, y \rangle \\ &= \int_0^t \langle AT(s)x, y \rangle \, \mathrm{d}s = \int_0^t \langle x, T(s)^* A^* y \rangle \, \mathrm{d}s \\ &= \langle x, \int_0^t T(s)^* A^* y \, \mathrm{d}s \rangle \,. \end{aligned}$$

Since $D(A)$ is dense, (11.11) holds. We are done.

Example 11.4 (Similar Semigroups)**.** Let X, Y be two Banach spaces with a linear homeomorphism V from Y onto X. Let $(T(t))_{t \geq 0}$ to be a strongly continuous semigroup on X with generator $(A, D(A))$. Then we obtain a new strongly continuous semigroup $(S(t))_{t \geq 0}$ on Y by defining

$$S(t) := V^{-1} T(t) V \qquad \text{for } t \geq 0 \,.$$

Its generator is

$$B = V^{-1} A V \qquad \text{with domain} \qquad D(B) = \{ y \in Y : Vy \in D(A) \}$$

Equality of the spectra

$$\sigma(A) = \sigma(B)$$

is clear, and the resolvent of B is $R(\lambda; B) = V^{-1} R(\lambda; A) V$ for $\lambda \in \rho(A)$.

Without explicit reference to the linear homeomorphism V, we call the two semigroups $(T(t))_{t \geq 0}$ and$(S(t))_{t \geq 0}$ *similar*. Two such semigroups have the same topological properties; e.g., they have the same growth bound.

11.5 The Stone Theorem

This section is devoted to describe the characteristic strongly continuous unitary group, and give an application of it. First of all, we describe which operators are generators of strongly continuous groups. In order to make this more precise we first adapt the definition in Section to this situation.

11.5.1 Generators of Groups

Let $(T(t))_{t\in\mathbb{R}}$ be a strongly continuous group on the Banach space X. Recall that the orbit maps is defined by

$$\xi_x : t \mapsto T(t)x \ ; \ \mathbb{R}_+ \to X$$

for each given $x \in X$. Similar to Lemma 11.7 we have:

Exercise 11.6. Then the orbit map ξ_x differentiable on $\mathbb{R}$ if and only if it is differentiable at $t = 0$.

Definition 11.5. The *generator* $(A, D(A))$ for the strongly continuous semigroup $(T(t))_{t\in\mathbb{R}}$ is the operator given by

$$Ax = \lim_{h\to 0} A_h x = \lim_{h\to 0} \frac{T(h)x - x}{h} = \xi_x'(0)\,;$$
$$D(A) = \{x : \xi_x \text{ is differentiable}\} = \{x : \lim_{h\to 0} A_h x \text{ exists }\}\,.$$

Given a strongly continuous group $(T(t))_{t\in\mathbb{R}}$ with generator $(A, D(A))$ we can define

$$T_+(t) := T(t) \ \text{ and } \ T_-(t) := T(-t) \ \text{ for } \ t \geq 0\,.$$

Then, from the previous definition, it's clear that $(T_+(t))_{t\geq 0}$ and $(T_-(t))_{t\geq 0}$ are strongly continuous semigroups with generators $(A, D(A))$, $(-A, D(A))$ respectively. Therefore, if A is the generator of a group, then both A and $-A$ generate strongly continuous semigroups. The next result shows that the converse of this statement is also true.

Proposition 11.18. *Let $w \in \mathbb{R}$ and $M \geq 1$ be constants. For a linear operator $(A, D(A))$ on a Banach space X the following properties are equivalent.*

(a) *$(A, D(A))$ generates a strongly continuous group $(T(t))_{t\in\mathbb{R}}$ satisfying the growth estimate*

$$\|T(t)\| \leq M\mathrm{e}^{w|t|} \quad \textit{for } t \in \mathbb{R}\,. \tag{11.12}$$

(b) *$(A, D(A))$ and $(-A, D(A))$ are the generators of strongly continuous semigroups $(T_+(t))_{t\geq 0}$ and $(T_-(t))_{t\geq 0}$, respectively, which satisfy*

$$\|T_+(t)\|\,, \|T_-(t)\| \leq M\mathrm{e}^{wt} \quad \textit{for all } t \geq 0\,. \tag{11.13}$$

Proof. It remains to show that (b) $\Rightarrow$ (a). Let

$$T(t) = \begin{cases} T_+(t) & \text{if } t \geq 0\,; \\ T_-(-t) & \text{if } t \leq 0\,. \end{cases}$$

Then we asserts that $(T(t))_{t\in\mathbb{R}}$ forms a operator group. By Exercise 11.3, it suffices to show that

$$T_+(t)T_-(t) = T_-(t)T_+(t) = I \quad \text{for } t \geq 0\,.$$

To this end, firstly we show that $T_+(t)$ and $T_-(t)$ commutes. Observe that the Yosida approximants $A_{+,n}$ and $A_{-,n}$ of A and $-A$, respectively, commute; and we have

$$T_+(t)x = \lim_{n\to\infty} \exp\{tA_{+,n}\}x \quad \text{and} \quad T_-(t)x = \lim_{n\to\infty} \exp\{tA_{-,n}\}x$$

for all $x \in X$, we see that $T_+(t)$ and $T_-(t)$ commute. We have only to show that $T_+(t)T_-(t) = I$ for all $t \geq 0$.

Note that, formally, there holds

$$T_+(t)T_-(t)x - x = \int_0^t \frac{\mathrm{d}}{\mathrm{d}s} T_+(s)T_-(s)x\,\mathrm{d}s\,.$$

By Lemma 11.10, since $D(A)$ is invariant under $(T_-(t))_{t\geq 0}$, for $x \in D(A)$, the map $s \mapsto T_+(s)T_-(s)x$ is differentiable form $\mathbb{R}_+$ into X with

$$\frac{\mathrm{d}}{\mathrm{d}s}T_+(s)T_-(s)x = -T_+(s)AT_-(s)x + T_+(s)AT_-(s)x = 0\,.$$

Thus $T_+(t)T_-(t)x = x$ for all $x \in D(A)$. Since $T_+(t)T_-(t) \in B(X)$ and $D(A)$ is dense in X, we get $T_+(t)T_-(t) = I$.

Finally, the estimation (11.12) follows from (11.13); the strong continuity of $(T(t))_{t\in\mathbb{R}}$ follows form the strong continuity of $(T_+(t))_{t\geq 0}$ and $(T_-(t))_{t\geq 0}$; by the definition $(T(t))_{t\in\mathbb{R}}$ has generator $(A, D(A))$. We are done. □

11.5.2 Stone's Theorem

In the following, we always assume that H is a complex Hilbert space.

Definition 11.6. An operator group $\{U(t)\}_{t\in\mathbb{R}}$ on H is called a *unitary group* if for each t, $U(t)$ is a unitary operator on H.

Remark 11.7. If $\{U(t)\}_{t\geq 0}$ is a strongly continuous unitary semigroup, since $U(t)^{-1} = U(t)^* \in \mathcal{B}(X)$ for each t, define $U(-t) = U(t)^* = U(t)^{-1}$ for $t \geq 0$. Then $\{U(t)\}_{t\in\mathbb{R}}$ forms a strongly continuous unitary group by Exercise 11.3 and the fact that

$$\|U(-t)x - x\| = \|U(t)^*x - x\| = \|U(t)x - x\| \ \text{ for all } \ t \geq 0, x \in H\,.$$

Remark 11.8. The weak continuity of a unitary group $\{U(t)\}_{t\in\mathbb{R}}$ implies the continuity. There is an proof much easier than Theorem 11.2. Indeed if $U(h)x \xrightarrow{w} x$ as $h \to 0$, since $\|U(h)x\| = \|x\|$ for all h, we conclude that $U(h)x \to x$ as $h \to 0$ by Theorem 5.33.

Example 11.5. Let $(A, D(A))$ be a self-adjoint operator on the complex Hilbert H. Let E^A be the spectrum decomposition for A. For each $t \in \mathbb{R}$, set

$$U(t) \coloneqq e^{itA} \equiv \int_{\mathbb{R}} e^{it\lambda} E^A(\mathrm{d}\lambda)\,.$$

By the functional calculus, it's easy to see that $\{U(t)\}_{t\in\mathbb{R}}$ is a unitary group. We show that it is weakly continuous and hence, strongly continuous. Indeed, for each x, y in H,

$$t \mapsto \langle U(t)x, y\rangle = \int_{\mathbb{R}} e^{it\lambda} E^A_{x,y}(\mathrm{d}\lambda)$$

is a continuous function by the dominated convergence theorem as required.

We claim that the generator of $\{U(t)\}_{t\in\mathbb{R}}$ is $(iA, D(A))$.

Denote by $(B, D(B))$ the generator of $\{U(t)\}_{t\in\mathbb{R}}$. Then fix $x \in D(A)$, there holds

$$\left\| \frac{U(h)x - x}{h} x - iAx \right\|^2 = \int_{\mathbb{R}} \left| \frac{e^{ih\lambda} - 1}{h} - i\lambda \right|^2 E^A_{x,x}(\mathrm{d}\lambda) \to 0$$

as $h \to 0$, $h \neq 0$. Thus $iA \subset B$. On the other hand note that

$$\varrho(iA) \cap \varrho(B) \neq \emptyset$$

we conclude that $B = iA$ by Exercise 11.5.

Theorem 11.19 (Stone)**.** *Let $(B, D(B))$ be a densely defined operator on H. Then $(B, D(B))$ generates a strongly continuous unitary group $\{U(t)\}_{t\in\mathbb{R}}$ if and only if B is skew-adjoint, i.e., $B^* = -B$. In this case, $A = -iB$ is self-adjoint and*

$$U(t) = e^{itA} \quad \text{for all } t \in \mathbb{R}\,.$$

Proof. If B is skew-adjoint, let $A = -iB$ then $A^* = iB^* = -iB = A$, so A is self-adjoint. By the example above, the desired result follows.

If $(B, D(B))$ generates a strongly continuous unitary group $\{U(t)\}_{t\in\mathbb{R}}$, then by Proposition 11.18, $(-B, D(B))$ generator the semigroup $\{U_(t) = U(t)^*\}_{t\geq 0}$. However, by Example 11.3, the generator of the adjoint semigroup $\{U(t)^*\}_{t\geq 0}$ is $(B^*\ D(B^*)$. Thus we conclude that B is skew-adjoint

$$B^* = -B\,.$$

By the example above, since $\{U(t)\}_{t\geq 0}$ and $\{e^{itA}\}_{t\geq 0}$ has the same generator, by Theorem 11.11 there must be

$$U(t) = e^{itA}$$

for all $t \in \mathbb{R}$. We are done. $\square$

Finally we give an application of Stone's theorem. We shall answer the following question.

> A characteristic function is defined to be the Fourier transform of a probability measure. Can it be characterized by some other properties?

Several answers are known, but the following characterization, due to Bochner and Herglotz, is most useful. It plays a basic role in harmonic analysis and in the theory of second-order stationary processes.

Definition 11.7. A complex-valued function f defined on $\mathbb{R}$ is called *positive definite* iff for any finite set of real numbers t_j and complex numbers z_j (with conjugate complex $\bar{z}_j$), $1 \leq j \leq n$, we have

$$\sum_{j=1}^{n}\sum_{k=1}^{n} f(t_j - t_k)\, z_j \bar{z}_k \geq 0\,. \tag{11.14}$$

Theorem 11.20 (Bochner)**.** $f : \mathbb{R} \to \mathbb{C}$ *is a characteristic function if and only if it is positive definite and uniformly continuous on* $\mathbb{R}$ *with* $f(0) = 1$.

Remark 11.9. Indeed if f is positive definite then there holds $f(-t) = \overline{f(t)}$, $|f(t)| \leq f(0)$ for all $t \in \mathbb{R}$ and; if f is continuous at 0 then f is uniformly continuous on $\mathbb{R}$. See Theorem 6.5.1 in *A Course in Probability Theory* by KaiLai Chung.

Proof. If f is the characteristic function of the probability measure μ, then we need only verify that it is positive definite. This is immediate, since

$$\begin{aligned}\sum_{1\le j,k\le n} f\left(t_j-t_k\right) z_j\bar{z}_k &= \int \sum_{1\le j,k\le n} e^{i(t_j-t_k)x} z_j\bar{z}_k\mu(\mathrm{d}x)\\ &= \int \big(\sum_{j=1}^{n} e^{it_jx}z_j\big)^2\mu(\mathrm{d}x)\ge 0\,.\end{aligned}$$

On the contrary, assume that f is positive definite and uniformly continuous on $\mathbb{R}$. Let

$$L=\{x(t):\mathbb{R}\to\mathbb{C}:x(t)=0 \text{ for all except finitely many } t\in\mathbb{R}\}$$

and define

$$\langle x,y\rangle_f := \sum_{s,t\in\mathbb{R}} f\left(t-s\right)x(t)\overline{y}(s) \quad \text{for } \ x,y\in L\,.$$

Here for convenience we assume f is strictly positive definite, that is the equality in (11.14) holds if and only if $z_j=0$ for all $1\le j\le n$. Then $(L,\langle\cdot,\cdot\rangle_f)$ is a complex inner product space. (If f is only positive definite, then $\langle\cdot,\cdot\rangle_f$ is only a semi-inner product, we need to consider the quotient space L modulo $\langle\cdot,\cdot\rangle_f$.) Denote by $(H,\langle\cdot,\cdot\rangle_f)$ the completion of $(L,\langle\cdot,\cdot\rangle_f)$. Then L is a dense subspace in H.

For $\tau\in\mathbb{R}$, let $e_\tau := 1_{\{\tau\}}(t)\in L$. Then

$$f(\tau)=\sum_{s,t\in\mathbb{R}} f\left(t-s\right)1_{\{\tau\}}(t)1_{\{0\}}(s)=\langle e_\tau,e_0\rangle_f\,.$$

Define $U_\tau\in B(H)$ by

$$(U_\tau x)(t) := x(t-\tau) \ \text{ for } t\in\mathbb{R}, x\in L\,.$$

Since L is dense in H, U_τ is well-defined. Observe that $U_\tau e_0=e_\tau$. Then

$$f(\tau)=\langle U_\tau e_0,e_0\rangle_f\,.$$

In fact $\{U_\tau\}_{\tau\in\mathbb{R}}$ is a strongly continuous unitary group on H. It's clear that $\{U_\tau\}_{\tau\in\mathbb{R}}$ is an operator group. Moreover, U_τ protects the norm

$$\|U_\tau x\|_f = \|x\|_f \quad \text{for } x \in L\,;$$

and by the uniform continuity of f,

$$\lim_{h\to 0} \|U_h x - x\|_f^2 = \lim_{h\to 0} \sum_{s,t\in\mathbb{R}} \left[f\left(t-s\right) - f\left(t-s+h\right)\right] x(t)\overline{x}(s) = 0\,.$$

for $x \in L$. Therefore by Stone's theorem, there is self-adjoint operator $(A, D(A))$ on H with spectrum decomposition E^A so that

$$f(\tau) = \langle U_\tau e_0, e_0\rangle_f = \int_{\mathbb{R}} \mathrm{e}^{it\tau} E^A_{e_0,e_0}(\mathrm{d}t)\,.$$

Since $\|e_0\|_f = 1$, $E^A_{e_0,e_0}$ is a probability measure on $\mathbb{R}$ as desired. $\square$

11.6 Examples (Needs to be modified)

In order to convince us that new and interesting phenomena appear for semigroups on infinite-dimensional Banach spaces, we first discuss several classes of semigroups on concrete spaces. These semigroups are not uniformly continuous anymore and hence not of the form $\left(\mathrm{e}^{tA}\right)_{t\geq 0}$ for some bounded operator A. On the other hand, they are not "pathological" in the sense of being completely unrelated to any analytic structure. In addition, these semigroups accompany us through the further development of the theory and provide a source of illuminating examples and counterexamples.

11.6.1 Multiplication Semigroups on $\mathrm{C}_0(\Omega)$

Multiplication operators can be considered as an infinite-dimensional generalization of *diagonal matrices*. They are extremely simple to construct, and most of their properties are evident. Nevertheless, their value should not be underestimated. They appear, for example, naturally in the context of Fourier analysis or when one applies the spectral theorem for self-adjoint operators on Hilbert spaces. We therefore strongly recommend that any first attempt to illustrate a result or disprove a conjecture on semigroups should be made using multiplication semigroups.

11.7 Hille-Yosida Theorems

We now turn to the fundamental problem of semigroup theory, which is to find arrows in Figure 11.1 leading from the generator to the semigroup. This means that we discuss the following problem:

> Characterize those linear operators that are generators of some strongly continuous semigroup, and describe how the semigroup is generated.

11.7.1 Exponential Formulas

To tackle the above problem, it is helpful to recall the results from Section 11.2 and to think of the semigroup generated by an operator A as an "exponential function"

$$t \mapsto \mathrm{e}^{tA} \,.$$

This is also implied by the following proposition.

Proposition 11.21. *Let $(T(t))_{t\geq 0}$ be a strongly continuous semigroup with generator $(A, D(A))$. Then*

$$T(t) = s\text{-}\lim_{h\downarrow 0} \mathrm{e}^{tA_h} \,.$$

Proof. By the Banach-Steinhaus theorem, it suffices to show that e^{tA_h} are uniformly bounded for $h \in (0,1]$; and $\mathrm{e}^{tA_h}x \to T(t)x$ as $h \downarrow 0$ for each $x \in D(A)$. To prove the first assertion, note that

$$\mathrm{e}^{tA_h} = e^{-\frac{t}{h}I} e^{\frac{t}{h}T(h)} = e^{-\frac{t}{h}} e^{\frac{t}{h}T(h)} = e^{-\frac{t}{h}} \sum_{n=0}^{\infty} \frac{1}{n!} \frac{t^n}{h^n} T(nh) \,.$$

By Proposition 11.3, there are $\omega, M > 0$ so that $\|T(t)\| \leq M\mathrm{e}^{\omega t}$ for all $t \geq 0$. Hence, set $K := \sup_{0<h\leq 1} \frac{1}{h}\left(e^{\omega h} - 1\right)$, and we compute for all $0 < h \leq 1$,

$$\|\mathrm{e}^{tA_h}\| \leq e^{-\frac{t}{h}} \sum_{n=0}^{\infty} \frac{1}{n!} \frac{t^n}{h^n} M\mathrm{e}^{\omega n h} = M \exp\left\{\frac{t}{h}\left(e^{\omega h} - 1\right)\right\} \leq M \exp\left\{tK\right\} .$$

Thus the uniform boundedness follows.

Observe that, formally we have

$$\mathrm{e}^{tA_h}x - T(t)x = \mathrm{e}^{(t-s)A_h}T(s)x\,\Big|_{s=0}^{t} = \int_0^t \frac{\mathrm{d}}{\mathrm{d}s}\mathrm{e}^{(t-s)A_h}T(s)x\,\mathrm{d}s\,.$$

By Lemma 11.10, since $D(A)$ is invariant under $T(s)$, then $s \to \mathrm{e}^{(t-s)A_h}T(s)x$ is differentiable mapping from $[0,t]$ into X with

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\mathrm{e}^{(t-s)A_h}T(s)x &= -\mathrm{e}^{(t-s)A_h}A_hT(s)x + \mathrm{e}^{(t-s)A_h}AT(s)x \\ &= \mathrm{e}^{(t-s)A_h}T(s)[Ax - A_hx]\,.\end{aligned}$$

Therefore we get for all $0 < h \leq 1$,

$$\begin{aligned}\|\mathrm{e}^{tA_h}x - T(t)x\| &\leq \int_0^t \|\mathrm{e}^{(t-s)A_h}\|\|T(s)\|\|Ax - A_hx\|\,\mathrm{d}s \\ &\leq tM\exp\{tK\}M\exp\{\omega t\}\|Ax - A_hx\|\,.\end{aligned}$$

Letting $h \downarrow 0$, the desired result follows. □

We pursue this idea by recalling the various ways by which we can define "exponential functions." Each of these formulas and each method is then checked for a possible generalization to infinite-dimensional Banach spaces and, in particular, to unbounded operators. Here are some more or less promising formulas for "e^{tA}".

(i) We might use the power series and define

$$\mathrm{e}^{tA} := \sum_{n=0}^{\infty} \frac{t^n}{n!}A^n$$

However, for unbounded A, it is unrealistic to expect convergence of this series.

(ii) We might use the Cauchy integral formula and define

$$e^{tA} := \frac{1}{2\pi i} \int_\gamma e^{\lambda t} R(\lambda; A)\, d\lambda$$

As already noted, the generator A, hence also its spectrum $\sigma(A)$, may be unbounded. Therefore, the path γ surrounding $\sigma(A)$ will be unbounded, and so we need extra conditions to make the integral converge. For a class of semigroups, this approach does work. See Section II.4 in *A Short Course on Operator Semigroups* by K.Engel and R.Nagel.

(iii) At least in the one-dimensional case, the formulas

$$e^{tA} = \lim_{n\to\infty} \left(1 + \frac{t}{n}A\right)^n = \lim_{n\to\infty} \left(1 - \frac{t}{n}A\right)^{-n}$$

are well known. Whereas the first formula again involves powers of the unbounded operator A and therefore will rarely converge, we can rewrite the second. This yields a formula involving only powers of bounded operators. It was Hille's idea (in 1948) to use this formula and to prove that under appropriate conditions, the limit exists and defines a strongly continuous semigroup.

(iv) Because it is well understood how to define the exponential function for bounded operators, one can try to approximate A by a sequence $(A_n)_{n\geq 1}$ of bounded operators and hope that

$$e^{tA} := \lim_{n\to\infty} e^{tA_n}$$

exists and is a strongly continuous semigroup. This was Yosida's idea (also in 1948) and is now examined in detail in order to obtain strongly continuous semigroups.

We start with an important convergence property for the resolvent under the assumption that $\|\lambda R(\lambda; A)\|$ remains bounded as $\lambda \to \infty$. It suggests immediately which bounded operators A_n should be chosen to approximate the unbounded operator A.

Lemma 11.22. *Let $(A, D(A))$ be a closed, densely defined operator. Suppose there exist $w \in \mathbb{R}$ and $M > 0$ such that $[w, \infty) \subset \rho(A)$ and $\|\lambda R(\lambda; A)\| \leq M$ for all $\lambda \geq w$. Then the following convergence statements hold for $\lambda \to \infty$.*

(a) $\lambda R(\lambda; A)x \to x$ *for all* $x \in X$

(b) $\lambda AR(\lambda; A)x = \lambda R(\lambda; A)Ax \to Ax$ *for all* $x \in D(A)$

Proof. The second statement is an immediate consequence of the first one. Note that indeed the first one assets that $\lambda R(\lambda; A) \xrightarrow{s} I$ as $\lambda \to \infty$. Since $\|\lambda R(\lambda; A)\| \leq M$ for all $\lambda \geq w$, by the Banach-Steinhaus theorem, it suffices to show that $\lambda R(\lambda; A)x \to x$ for all $x \in D(A)$. Observe that

$$\|\lambda R(\lambda; A)x - x\| = \|R(\lambda; A)Ax\| \leq \frac{M}{\lambda}\|Ax\|,$$

the desired result follows. □

11.7.2 Generation Theorem : Contraction Case

Because for contraction semigroups the technical details of the subsequent proof become much easier (and because the general case can then be deduced from this one), we first give the characterization theorem for generators in this special case.

Theorem 11.23 (Contraction Case, Hille, Yosida)**.** *For a linear operator $(A, D(A))$ on a Banach space X, the following properties are all equivalent.*

(a) $(A, D(A))$ *generates a strongly continuous contraction semigroup.*

(b) $(A, D(A))$ *is closed, densely defined, and for every $\lambda > 0$ one has $\lambda \in \rho(A)$ and*

$$\|R(\lambda, A)\| \leq \frac{1}{\lambda}.$$

Proof. In view of Theorem 11.16 it suffices to show (b) $\Rightarrow$ (a). To that purpose, we define the so-called *Yosida approximants*

$$A_n := nAR(n, A) = n^2 R(n, A) - nI \in \mathcal{B}(X)$$

which are mutually commuting bounded operators for each $n \in \mathbb{N}$. Consider then the uniformly continuous semigroups given by

$$T_n(t) := \mathrm{e}^{tA_n}, \quad t \geq 0$$

Because A_n converges to A pointwise on $D(A)$ by Lemma 11.22, we anticipate that the following properties hold. By establishing these statements we complete the proof.

(i) $T(t)x := \lim_n T_n(t)x$ exists for each $x \in X$.

(ii) $(T(t))_{t \geq 0}$ is a strongly continuous contraction semigroup on X.

(iii) This semigroup has generator $(A, D(A))$.

Step 1. We shall show that for each fixed t, $\{T_n(t)\}_n$ converges strongly. Observe that each $(T_n(t))_{t \geq 0}$ is a contraction semigroup, because

$$\|T_n(t)\| \leq \mathrm{e}^{-nt} \mathrm{e}^{\|n^2 R(n,A)\| t} \leq \mathrm{e}^{-nt} \mathrm{e}^{nt} = 1\,.$$

So by the Banach-Steinhaus theorem, it suffices to prove that for each $x \in D(A)$,

$$\{T_n(t)x\}_{n=1}^{\infty} \text{ is a Cauchy sequence.}$$

Observe that, formally we have

$$T_n(t)x - T_m(t)x = T_m(t-s)T_n(s)x \,\Big|_{s=0}^{t} = \int_0^t \frac{\mathrm{d}}{\mathrm{d}s} T_m(t-s)T_n(s)x \,\mathrm{d}s\,.$$

By Lemma 11.10, the map $s \to T_m(t-s)T_n(s)x$ from $[0,t]$ into X is differentiable with

$$\frac{\mathrm{d}}{\mathrm{d}s} T_m(t-s)T_n(s)x = T_m(t-s)T_n(s)\left(A_n x - A_m x\right)\,.$$

Therefore we get

$$\begin{aligned}\|T_n(t)x - T_m(t)x\| &\le \int_0^t \|T_m(t-s)T_n(s)\left(A_n x - A_m x\right)\| \,\mathrm{d}s \\ &\le t\,\|A_n x - A_m x\| \,.\end{aligned}$$

By Lemma 11.22, $\{A_n x\}$ is a Cauchy sequence for each $x \in D(A)$. Therefore, $\{T_n(t)x\}$ converges for each $x \in D(A)$ as required.

Moreover, one can see that for fixed $x \in X$, $\{T_n(t)x\}$ converges uniformly for t in each compact interval $[0, t_0]$ as $n \to \infty$.

Step 2. The pointwise convergence of $(T_n(t)x)_n$ implies that the limit family $(T(t))_{t\ge 0}$ satisfies the functional equation, hence is a semigroup, and consists of contractions. Moreover, for each $x \in X$, the corresponding orbit map

$$\xi : t \mapsto T(t)x, \quad 0 \le t \le t_0$$

is the uniform limit of continuous functions

$$\xi_n : t \mapsto T_n(t)x, \quad 0 \le t \le t_0$$

and so is continuous itself. The strong continuity of $(T(t))_{t\ge 0}$ follows.

Step 3. Denote by $(B, D(B))$ the generator of $(T(t))_{t\ge 0}$ and fix $x \in D(A)$. We first show that $A \subset B$. Fix $x \in D(A)$, then by Lemma 11.8,

$$T_n(t)x - x = \int_0^t T_n(s)A_n x \,\mathrm{d}s \qquad \text{for } t \ge 0\,.$$

Letting $n \to \infty$, by the fact that $T_n(t)x \to T(t)x$ and $T_n(s)A_n x \to T(s)Ax$ and the dominated convergence theorem, we conclude that

$$T(t)x - x = \int_0^t T(s)Ax \,\mathrm{d}s \qquad \text{for } t \ge 0\,.$$

By Lemma 11.8 again, $x \in D(B)$ and $Bx = Ax$. On the other hand, since

$$\varrho(A) \cap \varrho(B) \supset (0, \infty) \ne \emptyset$$

there must be $A = B$ by Exercise 11.5. We are done. $\square$

If a strongly continuous semigroup $(T(t))_{t\geq 0}$ with generator A satisfies, for some $\omega \in \mathbb{R}$, an estimate

$$\|T(t)\| \leq \mathrm{e}^{\omega t} \quad \text{for } t \geq 0\,,$$

then we say $(T(t))_{t\geq 0}$ is *quasi-contractive*, and we can apply the above characterization to the rescaled contraction semigroup given by

$$S(t) := \mathrm{e}^{-\omega t} T(t) \quad \text{for } t \geq 0\,.$$

Because the generator of $(S(t))_{t>0}$ is $B = A - \omega I$.

Corollary 11.24. *Let $\omega \in \mathbb{R}$. For a linear operator $(A, D(A))$ on a Banach space X the following conditions are equivalent.*

(a) *$(A, D(A))$ generates a strongly continuous semigroup $(T(t))_{t\geq 0}$ satisfying*

$$\|T(t)\| \leq \mathrm{e}^{\omega t} \quad \textit{for } t \geq 0\,.$$

(a) *$(A, D(A))$ is closed, densely defined, and for each $\lambda > \omega$ one has*

$$\|R(\lambda, A)\| \leq \frac{1}{\lambda - \omega}\,.$$

In the end of this section we characterize the resolvents of the generator of a contraction semigroup. This is also an easy application of the Hille-Yosida theorem and Exercise 6.1.

Exercise 11.7. Let $\{R_\lambda\}_{\lambda>0}$ be a family of bounded linear operators on X. Then $\{R_\lambda\}_{\lambda>0}$ is the resolvents of $(A, D(A))$ which generates a strongly continuous contraction semigroup, iff the following statements hold.

(a) The resolvent equation holds: $R_\lambda - R_\mu + (\lambda - \mu)R_\lambda R_\mu = 0$ for each $\lambda, \mu > 0$.

(b) $\{R_\lambda\}_{\lambda>0}$ is strongly continuous in the sense that $\lambda R_\lambda \xrightarrow{s} I$ as $\lambda \to \infty$.

(c) $\{R_\lambda\}_{\lambda>0}$ is contractive, that is, $\|\lambda R_\lambda\| \leq 1$ for all $\lambda > 0$.

11.7.3 Generation Theorem : Contraction Case

The characterization of generators of arbitrary strongly continuous semigroups can be deduced from the above method for contraction semigroups. However, norm estimates for all powers of the resolvent are needed.

Proposition 11.25 (General Case, Feller, Miyadera, Phillips)**.** *Let $(A, D(A))$ be a linear operator on a Banach space X and let $w \in \mathbb{R}$, $M \geq 1$ be constants. Then the following properties are equivalent.*

(a) $(A, D(A))$ *generates a strongly continuous semigroup* $(T(t))_{t\geq 0}$ *satisfying*

$$\|T(t)\| \leq M\mathrm{e}^{\omega t} \quad \textit{for } t \geq 0 \,.$$

(b) $(A, D(A))$ *is closed, densely defined, and for every* $\lambda > \omega$ *one has* $\lambda \in \rho(A)$ *and*

$$\|R(\lambda, A)^n\| \leq \frac{M}{(\lambda - \omega)^n} \quad \textit{for all } n \in \mathbb{N} \,.$$

Proof. It suffices to show (b) $\Rightarrow$ (a). Without loss of generality we assume $\omega = 0$. Define the Yosida approximants

$$A_n := nAR(n, A) = n^2 R(n, A) - nI \;\text{ and }\; T_n(t) := \mathrm{e}^{tA_n} \,.$$

We shall establish these statements.

(i) $T(t)x := \lim_n T_n(t)x$ exists for each $x \in X$.

(ii) $(T(t))_{t\geq 0}$ is a strongly continuous contraction semigroup on X.

(iii) This semigroup has generator $(A, D(A))$.

We only show (i) since (ii) and (iii) can be proved by the same argument as before. Fix t, wo show the uniformly boundedness of $(T_n(t))_n$, note that

$$T_n(t) = \mathrm{e}^{-nt}\mathrm{e}^{n^2 R(n,A)t} = \mathrm{e}^{-nt} \sum_{m=0}^{\infty} \frac{(n^2 t)^m}{m!} R(n, A)^m$$

and hence for all n,

$$\|T_n(t)\| \le e^{-nt} \sum_{m=0}^{\infty} \frac{(n^2 t)^m}{m!} \frac{M}{n^m} = e^{-nt} M e^{nt} = M\,.$$

So by the Banach-Steinhaus theorem, it suffices to prove that for each $x \in D(A)$,

$$\{T_n(t)x\}_{n=1}^{\infty} \quad \text{is a Cauchy sequence.}$$

As before we have

$$\begin{aligned}
\|T_n(t)x - T_m(t)x\| &\le \int_0^t \left\| \frac{\mathrm{d}}{\mathrm{d}s} T_m(t-s) T_n(s) x \right\| \mathrm{d}s \\
&\le \int_0^t \| T_m(t-s) T_n(s) \left(A_n x - A_m x \right) \| \, \mathrm{d}s \\
&\le M^2 t \, \| A_n x - A_m x \| \ .
\end{aligned}$$

By Lemma 11.22, $\{A_n x\}$ is a Cauchy sequence for each $x \in D(A)$. Therefore, $\{T_n(t)x\}$ converges for each $x \in D(A)$ as required. □

Remark 11.10. As a general rule, we point out that for an operator $(A, D(A))$ to be a generator one needs

- Conditions on the location of $\sigma(A)$ in some left half-plane and
- Growth estimates of the form

  $$\|R(\lambda, A)^n\| \le \frac{M}{(\operatorname{Re}\lambda - \omega)^n}$$

  for all powers of the resolvent $R(\lambda, A)$ in some right half-plane or on some semiaxis (ω, ∞). We show emphasize that the estimate with only $n = 1$ does not suffice (if $M \neq 1$).

This last condition is rather complicated and can be checked for nontrivial examples only in the (quasi) contraction case, i.e., only if $n = 1$ is sufficient in the case $M = 1$.

11.8 The Lumer–Phillips Theorem

Due to their importance, we now return to the study of contraction semigroups and look for a characterization of their generator that does not require explicit knowledge of the resolvent. The following is a key notion towards this goal.

Definition 11.8. A linear operator $(A, D(A))$ on a Banach space X is called *dissipative* if

$$\|(\lambda I - A)x\| \geq \lambda \|x\| \tag{11.15}$$

for all $\lambda > 0$ and $x \in D(A)$.

To familiarize ourselves with these operators we state some of their basic properties.

Proposition 11.26. *For a dissipative operator* $(A, D(A))$ *the following properties hold.*

(i) $\lambda I - A$ *is injective for all* $\lambda > 0$ *and*

$$\left\|(\lambda I - A)^{-1} y\right\| \leq \frac{1}{\lambda}\|y\| \quad \textit{for all} \quad y \in \operatorname{Ran}(\lambda I - A)\,.$$

(ii) $\lambda I - A$ *is surjective for some* $\lambda > 0$ *if and only if it is surjective for each* $\lambda > 0$. *In that case, one has* $(0, \infty) \subset \rho(A)$.

(iii) A *is closed if and only if the range* $\operatorname{Ran}(\lambda I - A)$ *is closed for some (hence all)* $\lambda > 0$.

(iv) If A *is densely defined, then* A *is closable. Its closure* $\bar{A}$ *is again dissipative and satisfies* $\operatorname{Ran}(\lambda I - \bar{A}) = \overline{\operatorname{Ran}(\lambda I - A)}$ *for all* $\lambda > 0$.

Proof. (i) is just a reformulation of estimate (11.15).

Tp prove (ii), assume that $\lambda_0 I - A$ is surjective for some $\lambda_0 > 0$. Then $\lambda_0 \in \varrho(A)$ and $\|R(\lambda_0; A)\| \leq \frac{1}{\lambda}$ by (i). Note that for $\lambda > 0$,

$$\begin{aligned}\lambda I - A &= \lambda_0 I - A - (\lambda_0 - \lambda) I \\ &= [I - (\lambda_0 - \lambda) R(\lambda_0; A)] (\lambda_0 I - A) \ .\end{aligned}$$

Thus for all $|\lambda - \lambda_0| < \lambda_0$ we have $|\lambda - \lambda_0| \| R(\lambda_0; A) \| < 1$, which implies $\lambda \in \varrho(A)$. From this we can see that indeed $(0, \infty) \subset \varrho(A)$.

To prove property (iii) observe that

$$\lambda \|x\| \leq \|(\lambda I - A)x\| \leq \|Ax\| + \lambda \|x\|$$

for all $x \in D(A)$. Thus $R(\lambda I - A)$ is closed if and only if A is closed.

To prove property (iv), take a sequence (x_n) in $D(A)$ satisfying $x_n \to 0$ and $Ax_n \to y$. By Lemma 4.15, we have to show that $y = 0$. For each $\epsilon > 0$, let $x \in D(A)$ so that $\|x - y\| < \epsilon$. Then for fixe $\lambda > 0$,

$$\|(\lambda I - A)x_n - (\lambda I - A)\frac{x}{\lambda}\| \geq \lambda \|x_n - \frac{x}{\lambda}\|$$

Passing to the limit as $n \to \infty$ yields

$$\left\| -y + x - \frac{1}{\lambda} Ax \right\| \geq \|x\|$$

For $\lambda \to \infty$ we obtain that $\|x\| \leq \|x - y\|$ Thus

$$\|y\| \leq \|y - x\| + \|x\| \leq 2\|y - x\| \leq 2\epsilon \, .$$

Since ϵ is arbitrary, we conclude that $y = 0$.

In order to verify that $\bar{A}$ is dissipative, take $x \in D(\bar{A})$. By definition of the closure of a linear operator, there exists a sequence (x_n) in $D(A)$ satisfying $x_n \to x$ and $Ax_n \to \bar{A}x$. Because A is dissipative and the norm is continuous, this implies that $\|(\lambda I - \bar{A})x\| \geq \lambda \|x\|$ for all $\lambda > 0$. Hence $\bar{A}$ is dissipative. Finally, observe that $\mathrm{Ran}(\lambda I - A)$ is dense in $\mathrm{Ran}(\lambda I - \bar{A})$. Because by assertion (iii) $\mathrm{Ran}(\lambda I - \bar{A})$ is closed in X, we obtain the final assertion in (iv). $\square$

From the resolvent estimate in Theorem 11.23, it is evident that the generator of a contraction semigroup satisfies the estimate (11.15), and hence is dissipative. On the other hand, many operators can be shown directly to be dissipative and densely defined. We therefore reformulate Generation Theorem 11.23 in such a way as to single out the property that ensures that a densely defined, dissipative operator is a generator.

Theorem 11.27 (Lumer, Phillips)**.** *For a densely defined, dissipative operator* $(A, D(A))$ *on a Banach space* X *the following statements are equivalent.*

(a) *The closure* $\bar{A}$ *of* A *generates a contraction semigroup.*

(b) $\operatorname{Ran}(\lambda I - A)$ *is dense in* X *for some (hence all)* $\lambda > 0$.

Proof. (a) $\Rightarrow$ (b). Theorem 11.23 implies that $\operatorname{Ran}(\lambda I - \bar{A}) = X$ for all $\lambda > 0$. By Proposition 11.26, $\operatorname{Ran}(\lambda - \bar{A}) = \overline{\operatorname{Ran}(\lambda - A)}$, then we obtain (b).

(b) $\Rightarrow$ (a). By the same argument, the density of the range $\operatorname{Ran}(\lambda - A)$ implies that $(\lambda - \bar{A})$ is surjective. Property (ii) of Proposition 11.26 shows that $(0, \infty) \subset \rho(\bar{A})$, and dissipativity of A implies the estimate

$$\|R(\lambda, \bar{A})\| \leq \frac{1}{\lambda} \quad \text{for } \lambda > 0\,.$$

This was required in Theorem 11.23 to assure that $\bar{A}$ generated a contraction semigroup. $\square$

Remark 11.11*.* The above theorem gains its significance when viewed in the context of the abstract Cauchy problem associated with an operator A. Assume that the operator A is known to be closed, densely defined, and dissipative. Then in order to solve the (time-dependent) initial value problem

$$x'(t) = Ax(t), x(0) = x$$

for all $x \in D(A)$, it is sufficient to solve the (stationary) resolvent equation

$$x - Ax = y$$

for all y in some dense subset in the Banach space X.

There is a simpler method that works particularly well in concrete function spaces such as $\mathrm{C}_0(\Omega)$ or $\mathrm{L}^p(\mu)$.

To introduce this method we start with a Banach space X and its dual space X^*. By the Hahn-Banach theorem, for every $x \in X$ there exists $x^* \in X^*$ such that

$$\langle x, x^* \rangle = \|x\|^2 = \|x^*\|^2 .$$

Hence, for every $x \in X$ the following set, called its *duality set*,

$$\mathcal{J}(x) := \left\{ x^* \in X^* : \langle x, x^* \rangle = \|x\|^2 = \|x^*\|^2 \right\}$$

is nonempty. Particular if X is a Hilbert space, then the duality set is always a singleton $\mathcal{J}(x) = \{y \in X : \langle x, y \rangle = \|x\|^2 = \|y\|^2\} = \{x\}$. Such sets allow a new characterization of dissipativity.

Theorem 11.28. *An operator $(A, D(A))$ is dissipative if and only if for every $x \in D(A)$ there exists $j(x) \in \mathcal{J}(x)$ such that*

$$\mathrm{Re}\langle Ax, j(x) \rangle \leq 0 . \tag{11.16}$$

If A is the generator of a strongly continuous contraction semigroup, then (11.16) *holds for all $x \in D(A)$ and arbitrary $x^* \in \mathcal{J}(x)$.*

Proof. Assume (11.16) is satisfied for $x \in D(A), \|x\| = 1$, and some $j(x) \in \mathcal{J}(x)$. Then $\langle x, j(x) \rangle = \|j(x)\|^2 = 1$ and

$$\begin{aligned} \|\lambda x - Ax\| &\geq |\langle \lambda x - Ax, j(x) \rangle| \\ &\geq \mathrm{Re}\langle \lambda x - Ax, j(x) \rangle \geq \lambda \end{aligned}$$

for all $\lambda > 0$. This proves one implication.

To show the converse, we take $x \in D(A), \|x\| = 1$, and assume that $\|\lambda x - Ax\| \geq \lambda$ for all $\lambda > 0$. Choose $y_\lambda^* \in \mathcal{J}(\lambda x - Ax)$ and consider the normalized elements $z_\lambda^* := y_\lambda^* / \|y_\lambda^*\|$.

Then the inequalities

$$\lambda \le \|\lambda x - Ax\| = \langle \lambda x - Ax, z_\lambda^* \rangle = \lambda \operatorname{Re} \langle x, z_\lambda^* \rangle - \operatorname{Re} \langle Ax, z_\lambda^* \rangle$$

are valid for each $\lambda > 0$. Since $\|x\| \le 1$ and $\|z_\lambda^*\| = 1$, we have $\operatorname{Re} \langle x, z_\lambda^* \rangle \le 1$. Then this yields

$$\operatorname{Re} \langle Ax, z_\lambda^* \rangle \le 0 \,.$$

Moreover,

$$\operatorname{Re} \langle x, z_\lambda^* \rangle \ge 1 + \frac{1}{\lambda} \operatorname{Re} \langle Ax, z_\lambda^* \rangle \ge 1 - \frac{1}{\lambda} \|Ax\| \,.$$

Regarding $\{z_\lambda\}_{\lambda>0}$ as a net in weak*-compact set B_{X^*}, the closed unit ball in X^*, we can choose $z \in B_{X^*}$ a weak-star accumulation point of z_λ^* as $\lambda \to \infty$. Then

$$\|z^*\| \le 1, \quad \operatorname{Re} \langle Ax, z^* \rangle \le 0, \qquad \text{and} \qquad \operatorname{Re} \langle x, z^* \rangle \ge 1$$

Combining these facts, it follows that z^* belongs to $\mathcal{J}(x)$ and satisfies (11.16).

Finally, assume that A generates a contraction semigroup $(T(t))_{t\ge 0}$ on X. Then, for every $x \in D(A)$ and arbitrary $x^* \in \mathcal{J}(x)$, we have

$$\begin{aligned} \operatorname{Re} \langle Ax, x^* \rangle &= \lim_{h \downarrow 0} \left(\frac{\operatorname{Re} \langle T(h)x, x^* \rangle}{h} - \frac{\operatorname{Re} \langle x, x^* \rangle}{h} \right) \\ &\le \overline{\lim_{h \downarrow 0}} \left(\frac{\|T(h)x\| \cdot \|x^*\|}{h} - \frac{\|x\|^2}{h} \right) \le 0. \end{aligned}$$

This completes the proof. □

11.9 Evolution Equations

We turn our attention to what could have been, in a certain perspective, our starting point: We want to solve the abstract Cauchy problem (ACP) associated with $(A, D(A))$ and the initial value x:

$$\begin{cases} u'(t) = Au(t)\,, \text{ for } t \ge 0\,; \\ u(0) = x\,, \end{cases} \tag{ACP}$$

where the independent variable t represents time, $u(\cdot)$ is a function on $\mathbb{R}_+$ with values in a Banach space X, $(A, D(A))$ a linear operator on X, and $x \in X$ the initial value.

Definition 11.9. A function $u : \mathbb{R}_+ \to X$ is called a *(classical) solution* of (ACP) if u is continuously differentiable, $u(t) \in D(A)$ for all $t \geq 0$, and (ACP) holds.

If u is a classical solution of (ACP), combine the fact that $Au(t) = u'(t)$ and $u : \mathbb{R}_+ \to X$ is continuous differentiable, there holds

$$u \in C^1(\mathbb{R}_+; X) \cap C(\mathbb{R}_+; (D(A), \|\cdot\|_A)) .$$

Theorem 11.29. *Let* $(A, D(A))$ *be the generator of the strongly continuous semigroup* $(T(t))_{t\geq 0}$*. Then, for every* $x \in D(A)$*, the function*

$$u : t \mapsto u(t) := T(t)x$$

is the unique classical solution of (ACP)*.*

Proof. It suffices to show the uniqueness of the solution. Let $v(t)$ be another classical solution of (ACP). Fix $t > 0$, then

$$v(t) - u(t) = v(t) - T(t)x = T(t-s)v(s) \,\Big|_{s=0}^{t} .$$

It's easy to show that the mapping $s \mapsto T(t-s)v(s)$ form $\mathbb{R}_+$ into X is differentiable with

$$\frac{\mathrm{d}}{\mathrm{d}s} T(t-s)v(s) = -T(t-s)Av(s) + T(t-s)Av(s) = 0 .$$

By the Newton-Leibniz formula we get $v(t) = u(t)$. We are done. $\square$

The important point is that (classical) solutions exist if (and, by the definition of $D(A)$, only if) the initial value x belongs to $D(A)$. However, one might *substitute the differential equation by an integral equation*, thereby obtaining a more general concept of "solution".

Definition 11.10. A continuous function $u : \mathbb{R}_+ \to X$ is called a *mild solution* of (ACP) if $\int_0^t u(s)\,\mathrm{d}s \in D(A)$ for all $t \geq 0$ and

$$u(t) = A \int_0^t u(s)\,\mathrm{d}s + x\,.$$

Every classical solution u of (ACP) is a mild solution. Indeed since $u \in C(\mathbb{R}_+; (D(A), \|\cdot\|_A))$ and $A : (D(A), \|\cdot\|_A) \to X$ is continuous, using the Newton-Leibniz formula we have

$$u(t) = \int_0^t u'(s)\,\mathrm{d}s + x = \int_0^t Au(s)\,\mathrm{d}s + x = A\int_0^t u(s)\,\mathrm{d}s + x$$

for all t. Conversely, if u is a mild solution of (ACP) and if $u \in C(\mathbb{R}_+; (D(A), \|\cdot\|_A))$ then u must be a classical solution since

$$u(t) = A\int_0^t u(s)\,\mathrm{d}s + x = \int_0^t Au(s)\,\mathrm{d}s + x\,.$$

Then we conclude that $u \in C^1(\mathbb{R}_+; X)$ with $u'(t) = Au(t)$ and $u(0) = x$.

It follows from our previous results (Lemma 11.8) that for A being the generator of a strongly continuous semigroup, mild solutions exist for every initial value $x \in X$ and are again given by the semigroup.

Proposition 11.30. *Let* $(A, D(A))$ *be the generator of the strongly continuous semigroup* $(T(t))_{t\geq 0}$*. Then, for every* $x \in X$*, the orbit map*

$$u : t \mapsto u(t) := T(t)x$$

is the unique mild solution of the associated abstract Cauchy problem (ACP).

Proof. We only have to show the uniqueness of the zero solution for the initial value 0. To this end, assume u to be a mild solution of (ACP) for $x = 0$. Then

$$u(t) = A\int_0^t u(s)\,\mathrm{d}s \quad \text{for} \quad t \geq 0\,.$$

Let $v(t) = \int_0^t u(s)\,\mathrm{d}s$ for $t \geq 0$. Since $u \in C(\mathbb{R}_+; X)$ we have $v \in C^1(\mathbb{R}_+; X)$ with

$$v'(t) = u(t) = A\int_0^t u(s)\,\mathrm{d}s = Av(t) \quad \text{for} \quad t \geq 0\,.$$

Thus v is a classical solution of (ACP) with initial value $x = 0$. By Theorem 11.29, $v(t) = 0$ for all $t \geq 0$ and hence $u(t) = 0$ for all $t \geq 0$ as desired. $\square$

The above two propositions are just reformulations of results on strongly continuous semigroups. They might suggest that the converse holds. The following example shows that this is not true.

Example 11.6. Let $(B, D(B))$ be a closed and unbounded operator on X. On the product space $X \times X$, consider the operator $(A, D(A))$ written in matrix form as

$$A := \begin{pmatrix} 0 & B \\ 0 & 0 \end{pmatrix} \quad \text{with domain} \quad D(A) = X \times D(B)\,.$$

Then $t \mapsto u(t) := \binom{x+tBy}{y}$ is the unique solution of (ACP) associated with A for every $\binom{x}{y} \in D(A)$. However, the operator A does not generate a strongly continuous semigroup, because for every $\lambda \in \mathbb{C}$, one has

$$\mathrm{Ran}(\lambda I - A) = \left\{ \begin{pmatrix} \lambda x - By \\ \lambda y \end{pmatrix} : x \in X, y \in D(B) \right\} \subset X \times D(B) \neq X \times X\,,$$

and hence $\sigma(A) = \mathbb{C}$ and $\varrho(A) = \emptyset$.

11.9.1 Well-Posedness

We now show which properties of the solutions $u(\cdot, x)$ or of the operator $(A, D(A))$ have to be added in order to characterize semigroup generators.

Definition 11.11. The abstract Cauchy problem (ACP) is called *well-posed* if the following statements hold.

- (Existence and Uniqueness; EU) There exists a unique classical solution $u(\cdot, x)$ of (ACP) for every initial value $x \in D(A)$.
- (Continuous Dependence; CD) For every $T > 0$, there is an $M > 0$ such that

$$\|u(t, x)\| \leq M\|x\| \text{ for all } t \in [0, T] \text{ and } x \in D(A)\,.$$

Exercise 11.8. Assume the property for existence and uniqueness holds. Show that the following statement is equivalent to "Continuous Dependence".

- (Continuous Dependence') For every sequence $(x_n)_{n\geq 1}$ in $D(A)$ so that $x_n \to 0$, one has $u(t, x_n) \to 0$ uniformly in $t \in [0, T]$ for each $T > 0$.

(Hint: Note that by the uniqueness we have $u(t, \frac{x}{n}) = \frac{1}{n}u(t, x)$.)

Theorem 11.31. *Let* $(A, D(A))$ *be a densely defined closed operator on* X*. Then for the associated abstract Cauchy problem* (ACP) *the following properties are equivalent.*

(a) $(A, D(A))$ *generates a strongly continuous semigroup.*

(b) (EU) holds and $\rho(A) \neq \emptyset$.

(c) The abstract Cauchy problem is well-posed.

Proof. From the basic properties of semigroup generators, it follows that (a) implies (b) and (c).

Step 1. For (b) $\Rightarrow$ (c), we first show that for all $x \in X$ there exists a unique mild solution of (ACP) with initial value x.

By the hypothesis there is $\lambda \in \varrho(A)$. We set $y = R(\lambda; A)x \in D(A)$. Then there is a unique classical solution $u(\cdot, y)$ of (ACP) with initial value y. Define

$$v(t, x) := (\lambda I - A)u(t, y) \quad \text{for} \quad t \geq 0\,.$$

We asserts that $v(\cdot, x)$ defines a mild solution for the initial value x.

In fact, since $u(\cdot, y) \in C^1(\mathbb{R}_+; X) \cap C(\mathbb{R}_+; (D(A), \|\cdot\|_A))$, we conclude that $v(\cdot, x) : \mathbb{R}_+ \to X$ is continuous. We compute for each $t \geq 0$

$$\begin{aligned}\int_0^t v(s, x)\,\mathrm{d}s &= \int_0^t (\lambda I - A)u(s, y)\,\mathrm{d}s = \lambda \int_0^t u(s, y)\,\mathrm{d}s - \int_0^t Au(s, y)\,\mathrm{d}s \\ &= \lambda \int_0^t u(s, y)\,\mathrm{d}s - (u(t, y) - y)\,.\end{aligned}$$

Thus

$$\begin{aligned}A\int_0^t v(s,x)\,\mathrm{d}s &= \lambda A\int_0^t u(s,y)\,\mathrm{d}s - A(u(t,y)-y)\\ &= \lambda\int_0^t Au(s,y)\,\mathrm{d}s - A(u(t,y)-y)\\ &= (\lambda I - A)(u(t,y)-y) = v(t,x)-x\,.\end{aligned}$$

In order to prove uniqueness let $u(\cdot,0)$ be a mild solution to the initial value 0. Then $v(t,0) := \int_0^t u(s,0)\,\mathrm{d}s$ is the classical solution for the initial value 0, (see the proof of Proposition 11.30), hence $v(\cdot,0)=0$ and consequently $u(\cdot,0)=0$ as well.

Step 2. For (b) $\Rightarrow$ (c), We now show the continuous dependence upon the initial data.

We consider for fixed $T>0$ the linear map

$$\Phi : X \to \mathrm{C}\left([0,T]\,,X\right), \quad x \mapsto u(\cdot,x)$$

where $u(\cdot,x)$ is the mild solution of (ACP) with initial data $x\in X$. By the existence and uniqueness of mild solution Φ is a well-defined linear operator. It remains to show that Φ is bounded.

By the closed graph theorem, it's sufficient to prove that Φ is closed. In fact, if $x_n\to x$ and $u\left(\cdot,x_n\right)\to v(\cdot)\in \mathrm{C}\left([0,T]\,,X\right)$. Then $v(0)=x$. We have only to show that $v(t)=u(t,x)$ for all $0\le t\le T$.

Observe that $t\in[0,T]$

$$D(A)\ni\int_0^t u\left(s,x_n\right)\mathrm{d}s \to \int_0^t v(s)\,\mathrm{d}s\,,$$

and

$$A\int_0^t u\left(s,x_n\right)\mathrm{d}s = u\left(t,x_n\right)-x_n \to v(t)-x\,.$$

Hence, by the closedness of A we conclude that $\int_0^t v(s)\,\mathrm{d}s\in D(A)$ and

$$A\int_0^t v(s)\,\mathrm{d}s = v(t)-x.$$

for all $0 \leq t \leq T$. Consequently $v(\cdot)$ is the unique mild solution of (ACP) with initial value x if we define $v(t) = u(t - T, v(T))$ for $t > T$. Therefore $v(t) = u(t, x)$ for all $0 \leq t \leq T$ as required.

Step 3. Finally we show that (c) $\Rightarrow$ (a).

Firstly for each t, since $x \mapsto u(t, x)$, $D(A) \to X$ is bounded and $D(A)$ is dense in X, the operator $T(t) \in \mathcal{B}(X)$ is well-defined by

$$T(t)x := u(t, x) \quad \text{for all } x \in D(A)\,.$$

Moreover, $\sup_{0 \leq t \leq T} \|T(t)\| < \infty$, since by the hypothesis

$$\sup_{0 \leq t \leq T} \|T(t)\| = \sup_{0 \leq t \leq T} \sup_{x \in D(A), \|x\| \leq 1} \|u(t, x)\| < \infty\,.$$

The uniqueness of the solutions implies

$$T(t + s)x = T(t)T(s)x$$

for each $x \in D(A)$ and all $t, s \geq 0$. Thus $(T(t))_{t \geq 0}$ is a semigroup on X. Since $T(t)x = u(t, x) \to x$ as $t \downarrow 0$ for all $x \in D(A)$, by Proposition 11.1 $(T(t))_{t \geq 0}$ is strongly continuous.

Now it suffices to show that the generator of $(T(t))_{t \geq 0}$ is exactly $(A, D(A))$. We denote it by $(B, D(B))$ temporarily. Since for $x \in D(A)$,

$$\frac{\mathrm{d}}{\mathrm{d}t} u(t, x) = Au(t, x)\,.$$

We get $x \in D(B)$ and $Bx = Ax$, that is $A \subset B$. On the other hand note that $D(A)$ is invariant under $(T(t))_{t \geq 0}$, by Lemma 11.13, $D(A)$ is a core for B. Since A is closed, we get $A = B$ as desired. $\square$

11.9.2 Regularity of the Solution

We now assume that $(A, D(A))$ is the generator of a strongly continuous semigroup $\{T(t)\}_{t \geq 0}$ on the Banach space X. We shall prove here that the classical solution u of the homogeneous equation

$$\begin{cases} u'(t) = Au(t)\,, \text{ for } t \geq 0\,; \\ u(0) = x\,, \end{cases} \tag{ACP}$$

is more regular than just $C^1(\mathbb{R}_+; X) \cap C(\mathbb{R}_+; (D(A), \|\cdot\|_A))$, provided one makes additional assumptions on the initial data x.

For each $k \geq 1$ and we define the *k-norm* on $D\left(A^k\right)$ by

$$\|x\|_k = \left[\|x\|^2 + \|Ax\|^2 + \ldots + \|A^k x\|^2\right]^{1/2} \text{ for } x \in D(A^k)\,.$$

We set $D(A^0) = X$ and $A^0 = I$. Then $\big(D(A^k), \|\cdot\|_k\big)$ is a Banach space. Indeed if (x_n) is a Cauchy sequence in $\big(D(A^k), \|\cdot\|_k\big)$, then there are x, y_j, $1 \leq j \leq k$ in X so that $x_n \to x$ and $A^j x_n \to y_j$. Then since A is closed, by induction, we conclude that $x \in D(A^k) = \big(D(A^k), \|\cdot\|_k\big)$ and $y_j = A^j x$ for $1 \leq j \leq k$. Thus $x_n \to x$ in $\big(D(A^k), \|\cdot\|_k\big)$.

$(T(t)|_{D(A^k)})_{t\geq 0}$ is a strongly continuous semigroup on $\big(D(A^k), \|\cdot\|_k\big)$. By Lemma 11.8, $T(t)$ maps $D(A^k)$ into $D(A^k)$. $T(t)$ is bounded since

$$\begin{aligned} \|T(t)x\|_k^2 &= \sum_{j=0}^{k} \|A^j T(t)x\|^2 = \sum_{j=0}^{k} \|T(t)A^j x\|^2 \\ &\leq \|T(t)\|^2 \sum_{j=0}^{k} \|A^j x\|^2 = \|T(t)\|^2 \|x\|_k^2\,. \end{aligned}$$

for all $x \in D(A^k)$. The strong continuity follows from the fact that

$$\|T(h)x - x\|_k^2 = \sum_{j=0}^{k} \|T(h)A^j x - A^j x\|^2 \to 0 \ \text{ as } \ h \downarrow 0$$

for each $x \in D(A^k)$.

The generator of $(T(t)|_{D(A^k)})_{t\geq 0}$ is exactly $A|_{D(A^{k+1})}$. Indeed, for $x, y \in D(A^k)$, there holds

$$\begin{aligned} \lim_{h\downarrow 0} \left\|\frac{T(h)x - x}{h} - y\right\|_k = 0 &\iff \lim_{h\downarrow 0} \sum_{j=0}^{k} \left\|\frac{T(h)A^j x - A^j x}{h} - A^j y\right\| = 0 \\ &\iff x \in D(A^{k+1}) \text{ and } y = Ax\,. \end{aligned}$$

Moreover, it's clear that the graph norm of the generator $A|_{D(A^{k+1})}$, namely $\|\cdot\|_{k,A|_{D(A^{k+1})}}$ on $D(A^{k+1})$ is equivalent to the norm $\|\cdot\|_{k+1}$.

Proposition 11.32. *Assume $x \in D(A^k)$ for some integer $k \geq 1$. Then the classical solution of* (ACP) *$u(t) = T(t)x$ satisfies*

$$u \in C^j \Big(\mathbb{R}_+; (D(A^{k-j}), \|\cdot\|_{k-j})\Big) \quad \text{for} \;\; 0 \leq j \leq k\,.$$

In particular, if $x \in D(A^\infty)$, then

$$u \in C^\infty \Big(\mathbb{R}_+; (D(A^l), \|\cdot\|_l)\Big) \quad \text{for all} \;\; l \geq 0\,. \tag{11.17}$$

Proof. The assumption $x \in D(A^k)$ means that the initial value x belongs to the domain of the generator of the semigroup $(T(t)|_{D(A^{k-1})})_{t\geq 0}$. Since $u(t) = T(t)x = T(t)|_{D(A^{k-1})}x$ for all $t \geq 0$, by Theorem 11.29,

$$\begin{aligned} u &\in C^1(\mathbb{R}_+; (D(A^{k-1}), \|\cdot\|_{k-1})) \cap C(\mathbb{R}_+; (D(A^k), \|\cdot\|_{k-1,A|_{D(A^k)}})) \\ &= C^1(\mathbb{R}_+; (D(A^{k-1}), \|\cdot\|_{k-1})) \cap C(\mathbb{R}_+; (D(A^k), \|\cdot\|_k))\,. \end{aligned}$$

Also, we have

$$u'(t) = T(t)|_{D(A^{k-1})}A|_{D(A^k)}x = T(t)Ax$$

where u' is the derivative of $u : \mathbb{R}_+ \to C^1(\mathbb{R}_+; (D(A^{k-1}), \|\cdot\|_{k-1}))$.

By induction, suppose that for some $1 \leq j < k$ we have

$$u \in C^j \Big(\mathbb{R}_+; (D(A^{k-j}), \|\cdot\|_{k-j})\Big)\,, u^{(j)}(t) = T(t)A^jx\,.$$

Then $A^jx \in D(A^{k-j})$, i.e., A^jx belongs to the domain of the generator of the semigroup $(T(t)|_{D(A^{k-j-1})})_{t\geq 0}$. Since $u^{(j)}(t) = T(t)A^jx = T(t)|_{D(A^{k-j-1})}A^jx$, by Theorem 11.29 again,

$$\begin{aligned} u^{(j)} &\in C^1(\mathbb{R}_+; (D(A^{k-j-1}), \|\cdot\|_{k-j-1})) \cap C(\mathbb{R}_+; (D(A^k), \|\cdot\|_{k-j-1,A|_{D(A^{k-j})}})) \\ &= C^1(\mathbb{R}_+; (D(A^{k-j-1}), \|\cdot\|_{k-j-1})) \cap C(\mathbb{R}_+; (D(A^{k-j}), \|\cdot\|_{k-j}))\,. \end{aligned}$$

Therefore

$$u \in C^{(j+1)}(\mathbb{R}_+; (D(A^{k-j-1}), \|\cdot\|_{k-j-1}))$$

and $u^{(j+1)}(t) = T(t)|_{D(A^{k-j-1})}A|_{D(A^{k-j})}A^jx = T(t)A^{j+1}x$. Now the desired result follows. $\square$

The Self-Adjoint Case Let H be a complex Hilbert space and $(A, D(A))$ be a self-adjoint operator on H so that $A \leq 0$, i.e., $\langle Ax, x\rangle \leq 0$ for all $x \in D(A)$. It's also easy to show that A is dissipative. For $\lambda > 0$, note that

$$\operatorname{Ran}(\lambda I - A)^{\perp} = \operatorname{Ker}(\lambda I - A^*) = \operatorname{Ker}(\lambda I - A) = \{0\}\,.$$

Thus $\operatorname{Ran}(\lambda I - A)$ is dense in H. By the Lumer–Phillips theorem, $(A, D(A))$ generates a strongly continuous contraction semigroup $(T(t))_{t\geq 0}$ on H. Moreover, using the same argument in Example 11.5, we have

$$T(t) = \mathrm{e}^{tA} := \int_{\sigma(A)} \mathrm{e}^{t\lambda}\ E^A(\mathrm{d}\lambda) \qquad \text{for } t \geq 0\,, \tag{11.18}$$

where E^A is the spectrum decomposition of A.

Theorem 11.33. *For every $x \in H$, the abstract Cauchy equation*

$$\begin{cases} u'(t) = Au(t) & \textit{for} \quad t > 0\,; \\ u(0) = x\,. \end{cases} \tag{ACP$'$}$$

has a unique solution u in

$$C(\mathbb{R}_+; H) \cap C^1((0, +\infty); H) \cap C((0, +\infty); (D(A), \|\cdot\|_A))\,, \tag{11.19}$$

given by $u(t) = T(t)x = \mathrm{e}^{tA}x$ for all $t \geq 0$. Moreover,

(i) $u \in C^\infty\big((0, +\infty); (D(A^l), \|\cdot\|_l)\big)$*; and*

(ii) for each $k \geq 0$ there is a constant C_k so that

$$\left\|u^{(k)}(t)\right\| = \left\|A^k u(t)\right\| \leq \frac{C_k}{t^k}\|x\| \quad \textit{for all} \quad t > 0\,.$$

Proof. Step 1. We show the uniqueness for the solution of (ACP′) in (11.19). It suffices to show that the solution u in (11.19) of (ACP′) with initial value 0 is $u(t) = 0$ for all $t \geq 0$. To this end we compute

$$\frac{\mathrm{d}}{\mathrm{d}t}\|u(t)\|^2 = \frac{\mathrm{d}}{\mathrm{d}t}\langle u(t), u(t)\rangle = 2\langle Au(t), u(t)\rangle \leq 0 \quad \text{for all} \quad t > 0\,.$$

Thus $t \mapsto \|u(t)\|^2$ is decreasing on $(0,\infty)$. Combine this with the fact that $t \mapsto \|u(t)\|^2$ is continuous on $[0,\infty)$, we get $u(t) = 0$ as desired.

Step 2. We now that the theorem holds provided that the initial value $x \in D(A^\infty)$. By Theorem 11.29, $u(t) = T(t)x = \mathrm{e}^{tA}x$ is the solution of (ACP) and hence must be the solution of (ACP′). Part (i) follows form by Proposition 11.32. It remains to show part (ii).

The case for $k = 0$ is trivial. Let $k \geq 1$, by part (ii) of Lemma 11.8, $u^{(k)}(t) = T(t)A^k x = \mathrm{e}^{tA}A^k x$. Thus by (11.18),Theorem 10.23, Exercise 10.8 and Theorem 10.21, we have

$$\begin{aligned}
\|u^{(k)}(t)\|^2 &= \int_{\sigma(A)} |\lambda^k \mathrm{e}^{t\lambda}|^2 \, E^A_{x,x}(\mathrm{d}\lambda) \\
&= \int_{\sigma(A)} \big(|\lambda|^k \mathrm{e}^{-t|\lambda|}\big)^2 \, E^A_{x,x}(\mathrm{d}\lambda) \\
&\leq \int_{\sigma(A)} \big(\frac{C_k}{t^k}\big)^2 \, E^A_{x,x}(\mathrm{d}\lambda) = \big(\frac{C_k}{t^k}\|x\|\big)^2,
\end{aligned}$$

where $C_k = \max\{\mathrm{e}^{-s}s^k : s > 0\} < \infty$. We are done.

Step 3. We now handle the general case and assume that $x \in H$. Let $u(t) = T(t)x$. By Theorem 11.14, $D(A^\infty)$ is dense in H, thus we can choose a sequence (x_n) in $D(A^\infty)$ so that $\|x_n - x\| \to 0$. Let $u_n(t) := T(t)x_n$. Since $\{T(t)\}_{t\geq 0}$ is a contraction semigroup

$$\|u_n(t) - u(t)\| = \|T(t)x - T(t)x_n\| \leq \|x - x_n\|$$

for all $t \geq 0$. Thus (u_n) converges uniformly to u on $\mathbb{R}_+$.

By Step 1, fix $k \geq 0$ and $t > 0$, we have

$$\begin{aligned}
&\left\|u_n^{(k)}(t) - u_m^{(k)}(t)\right\|_l = \sum_{j=0}^{l} \left\|A^j u_n^{(k)}(t) - A u_m^{(k)}(t)\right\| \\
&= \sum_{j=0}^{l} \left\|u_n^{(k+j)}(t) - u_m^{(k+j)}(t)\right\| \leq \sum_{j=0}^{l} \frac{C_{k+j}}{t^{k+j}} \|x_n - x_m\|
\end{aligned}$$

for all $m, n \geq 1$ and $l \geq 1$. It's clear that $\{u_n^{(k)}(t)\}$ is a Cauchy sequence in $(D(A^l), \|\cdot\|_l)$.

Suppose that $u_n^{(k)}(t) \to v_k(t)$ in $(D(A^l), \|\cdot\|_l)$. (The limit $v_k(t)$ does not depend on l since $(D(A^l), \|\cdot\|_l)$ can be embedded into $(D(A^{l-1}), \|\cdot\|_{l-1})$ continuously.) Thn it must be $v_0(t) = u(t)$, and hence $u_n(t) \to u(t)$ in $(D(A^l), \|\cdot\|_l)$ for each l. As a consequence,

$$A^k u(t) = \lim_{n\to\infty} A^k u_n(t) = \lim_{n\to\infty} u_n^{(k)}(t) = v_k(t)\,.$$

Fix $l \geq 1$. Observe that $\|u_n^{(k)}(t) - v_k(t)\|_l \to 0$ uniformly for $t \in [\delta, \infty)$ for every $\delta > 0$. We can conclude that the map $u : (0, \infty) \to (D(A^l), \|\cdot\|_l); t \mapsto u(t)$ is C^∞-differentiable with derivatives $u^{(k)}(t) = v_k(t) = A^k u(t)$; that is

$$u \in C^\infty \Big((0, +\infty); (D(A^l), \|\cdot\|_l) \Big)\,.$$

Trivially u is the solution of (ACP$'$) and part (ii) holds. □

11.9.3 Nonhomogeneous Evolution Equations

We are going to handle the nonhomogeneous evolution equations

$$\begin{cases} u'(t) = Au(t) + f(t)\,, \text{ for } t \geq 0\,; \\ u(0) = x\,, \end{cases} \tag{11.20}$$

where $(A, D(A))$ generates a strongly continuous semigroup $(T(t))_{t\geq 0}$ on the Banach space X, $f : \mathbb{R}_+ \to X$ and $x \in X$. A function $u : \mathbb{R}_+ \to X$ is called a *(classical) solution* of (11.20) if u is continuously differentiable, $u(t) \in D(A)$ for all $t \geq 0$, and (11.20) holds. Clearly any classical solution belongs to

$$C^1(\mathbb{R}_+; X) \cap C(\mathbb{R}_+; (D(A), \|\cdot\|_A))\,.$$

By the Duhamel's principle or the variation of constant method, we guess the solution has the form

$$u(t) = T(t)x + \int_0^t T(t-s) f(s)\,\mathrm{d}s\,.$$

To verify this, we need some assumption on the function f.

Theorem 11.34. *The nonhomogeneous evolution equations* (11.20) *has a unique classical solution* $u : \mathbb{R}_+ \to X$ *given by*

$$u(t) = T(t)x + \int_0^t T(t-s)f(s)\,\mathrm{d}s \quad \textit{for all} \quad t \geq 0\,,$$

provided that

$$f \in C^1(\mathbb{R}_+; X) \quad \textit{or} \quad f \in C(\mathbb{R}_+; (D(A), \|\cdot\|_A))\,.$$

Proof. The uniqueness of the solution follows form Theorem 11.29. It remains to show that u is a classical solution. To this end, let

$$v(t) := \int_0^t T(t-s)f(s)\,\mathrm{d}s = \int_0^t T(s)f(t-s)\,\mathrm{d}s \quad \text{for} \quad t \geq 0\,.$$

It suffices to show that v is a classical solution of (11.20) with the initial value $x = 0$.

Step 1. We show that $v \in C(\mathbb{R}_+; (D(A), \|\cdot\|_A))$.

If $f \in C^1(\mathbb{R}_+; X)$, then we compute, for $h > 0$,

$$\begin{aligned}
\frac{T(h)-I}{h}v(t) &= \frac{1}{h}\left[\int_0^t T(t+h-s)f(s)\,\mathrm{d}s - \int_0^t T(t-s)f(s)\,\mathrm{d}s\right] \\
&= \frac{1}{h}\left[\int_h^{t+h} T(t)f(t-s+h)\,\mathrm{d}s - \int_0^t T(s)f(t-s)\,\mathrm{d}s\right] \\
&= \frac{1}{h}\int_t^{t+h} T(s)f(t-s+h)\,\mathrm{d}s - \frac{1}{h}\int_0^h T(s)f(t-s)\,\mathrm{d}s \\
&\quad + \int_h^t T(s)\frac{f(t-s+h)-f(t-s)}{h}\,\mathrm{d}s\,.
\end{aligned}$$

Letting $h \downarrow 0$, by the dominated convergence theorem, we get $v(t) \in D(A)$ and

$$Av(t) = T(t)f(0) - f(t) + \int_0^t T(s)f'(t-s)\,\mathrm{d}s\,. \tag{11.21}$$

From this we can see that $v \in C(\mathbb{R}_+; (D(A), \|\cdot\|_A))$.

If $f \in C(\mathbb{R}_+; (D(A), \|\cdot\|_A))$, then the mapping $s \mapsto T(t-s)f(s); [0,t] \to (D(A), \|\cdot\|_A)$ is continuous. By the definition of the integral $v(t) \in D(A)$

for all $t \geq 0$. Moreover,

$$Av(t) = \int_0^t T(t-s)Af(s)\,\mathrm{d}s\,. \tag{11.22}$$

From this we can see that $v \in C(\mathbb{R}_+; (D(A), \|\cdot\|_A))$.

Step 2. We show that $v \in C^1(\mathbb{R}_+; X)$ and $v'(t) = f(t) + Av(t)$ for $t \geq 0$.

We compute, for $t \geq 0$ and $h > 0$,

$$\begin{aligned}
&\frac{v(t+h)-v(t)}{h} \\
&= \frac{1}{h}\left[\int_0^{t+h} T(t+h-s)f(s)\,\mathrm{d}s - \int_0^t T(t-s)f(s)\,\mathrm{d}s\right] \\
&= \frac{1}{h}\left[\int_t^{t+h} T(t+h-s)f(s)\,\mathrm{d}s + \int_0^t [T(t+h)-T(t)]f(s)\,\mathrm{d}s\right] \\
&= \frac{1}{h}\int_0^h T(s)f(t+h-s)\,\mathrm{d}s + \frac{T(h)-I}{h}v(t)\,.
\end{aligned}$$

Letting $h \downarrow 0$, by the dominated convergence theorem and the hypothesis on f, we get

$$\lim_{h\downarrow 0} \frac{v(t+h)-v(t)}{h} = f(t) + Av(t)\,.$$

That is v is right-differentiable on $\mathbb{R}_+$. We then show the left-differentiability.

If $f \in C^1(\mathbb{R}_+; X)$, we compute, for $t > 0$ and $h > 0$,

$$\begin{aligned}
&\frac{v(t)-v(t-h)}{h} \\
&= \frac{1}{h}\left[\int_0^t T(s)f(t-s)\,\mathrm{d}s - \int_0^{t-h} T(s)f(t-s-h)\,\mathrm{d}s\right] \\
&= \frac{1}{h}\int_{t-h}^t T(s)f(t-s)\,\mathrm{d}s + \int_0^{t-h} T(s)\frac{f(t-s)-f(t-s-h)}{h}\,\mathrm{d}s\,.
\end{aligned}$$

Letting $h \downarrow 0$, by the dominated convergence theorem and (11.21), we get

$$\lim_{h\downarrow 0} \frac{v(t)-v(t-h)}{h} = T(t)f(0) + \int_0^t T(s)f'(t-s)\,\mathrm{d}s = f(t) + Av(t)\,.$$

If $f \in C(\mathbb{R}_+; (D(A), \|\cdot\|_A))$, we compute, for $t > 0$ and $h > 0$,

$$\begin{aligned}
&\frac{v(t) - v(t-h)}{h} \\
&= \frac{1}{h}\left[\int_0^t T(t-s)f(s)\,\mathrm{d}s - \int_0^{t-h} T(t-s-h)f(s)\,\mathrm{d}s\right] \\
&= \frac{1}{h}\int_{t-h}^t T(t-s)f(s)\,\mathrm{d}s + \int_0^{t-h} \frac{T(t-s) - T(t-s-h)}{h} f(s)\,\mathrm{d}s
\end{aligned}$$

Letting $h \downarrow 0$, by the dominated convergence theorem and (11.22), we get

$$\lim_{h\downarrow 0} \frac{v(t) - v(t-h)}{h} = f(t) + \int_0^t T(t-s)Af(s)\,\mathrm{d}s = f(t) + Av(t)\,.$$

We are done. □

In the ned, we consider the semi-linear equation

$$\begin{cases} u'(t) = Au(t) + f(t, u(t))\,, \text{ for } 0 \le t \le T\,; \\ u(0) = x\,, \end{cases} \tag{11.23}$$

where $(A, D(A))$ generates a strongly continuous semigroup $(T(t))_{t\ge 0}$ on the Banach space X, $f : [0, T] \times X \to X$ is continuous and $x \in X$.

A function $u : [0, T] \to X$ is called a *(classical) solution* of (11.23) if u is continuously differentiable, $u(t) \in D(A)$ for all $t \in [0, T]$ and (11.20) holds. Clearly any classical solution belongs to

$$C^1([0, T]; X) \cap C([0, T]; (D(A), \|\cdot\|_A))\,.$$

A function $u : [0, T] \to X$ is called a *mild solution* of (11.23) if u satisfies the integral equation

$$u(t) = T(t)x + \int_0^t T(t-s)f(s, u(s))\,\mathrm{d}s\,.$$

Lemma 11.35. *If $f(t, x) \in C([0, T] \times X; X)$ satisfies the Lipschitz condition for x, that is, there is some constant $L > 0$ so that*

$$\|f(t, x_1) - f(t, x_2)\| \le L\,\|x_1 - x_2\| \ \textit{ for all } \ t \in [0, T], x_1, x_2 \in X\,.$$

Then for any initial value x, there exists a unique mild solution for (11.23) *.*

Proof. We shall use the Banach fixed point theorem. Define $F : C([0,T];X) \to C([0,T];X)$ by

$$(Fu)(t) = T(t)x + \int_0^t T(t-s)f(s,u(s))\,\mathrm{d}s \quad \text{for} \quad 0 \le t \le T\,.$$

Then for u, v in $C([0,T];X)$ and $t \in [0,T]$ we have

$$\begin{aligned} \|(Fu)(t) - (Fv)(t)\| &\le \int_0^t \|T(t-s)\|\|f(s,u(s)) - f(s,v(s))\|\,\mathrm{d}s \\ &\le MtL\|u-v\|_{C([0,T];X)} \end{aligned}$$

where $M \coloneqq \sup_{t\in[0,T]} \|T(t)\|$. Then by induction,

$$\|(F^n u)(t) - (F^n v)(t)\| \le \frac{(MLt)^n}{n!}\|u-v\|_{C([0,T];X)}\,.$$

Thus we get

$$\|F^n u - F^n v\| \le \frac{(MLT)^n}{n!}\|u-v\|_{C([0,T];X)}\,.$$

Since the series $\sum_n \frac{(MLT)^n}{n!} < \infty$, by the Banach fixed point theorem, there exists a unique fixed point $u \in C^1([0,T])$ for F. We are done. □

Remark 11.12. Indeed, the mill solution depends continuously on the integral value. To see this, let $u(t,x)$ be the mild solution with initial value $x \in X$. Then

$$\begin{aligned} &\|u(t,x) - u(t,y)\| \\ &\le \|T(t)x - T(t)y\| + \int_0^t \|T(t-s)\|\|f(s,u(s,x)) - f(s,u(s,y))\|\,\mathrm{d}s \\ &\le M\|x-y\| + ML\int_0^t \|u(s,x) - u(s,y)\|\,\mathrm{d}s \end{aligned}$$

By Grönwall's inequality

$$\|u(t,x) - u(t,y)\| \le Me^{MLt}\|x-y\|\,.$$

Thus $\|u(\cdot,x) - u(\cdot,y)\| \le Me^{MLT}\|x-y\|$ as desired.

Exercise 11.9. Suppose the assumption in the preceding lemma holds. Let $g \in C([0,T],X)$. Then show that the integral equation

$$y(t) = g(t) + \int_0^t T(t-s)f(s,y(s))\,\mathrm{d}s$$

has a unique solution $y \in C([0,T],X)$.

Theorem 11.36. *Suppose that $f \in C^1([0,T]\times X;X)$. Then for each initial data $x \in D(A)$, the mill solution u a classical solution.*

Proof. Let $u(t)$ be the unique mild solution with initial data $x \in D(A)$. Then

$$u(t) = T(t)x + \int_0^t T(s)f(t-s,u(t-s))\,\mathrm{d}s\,.$$

Step 1. We shall compute the derivative of u formally. Then we conclude that u' should satisfy the integral equation

$$\begin{aligned} u'(t) =& T(t)Ax + T(t)f(0,x) + \int_0^t T(t-s)\partial_1 f(s,u(s))\,\mathrm{d}s \\ &+ \int_0^t T(t-s)\partial_2 f(s,u(s))u'(s)\,\mathrm{d}s\,. \end{aligned}$$

Note that $\partial_2 f(s,u(s))$ is a bounded linear operator in $\mathcal{B}(X)$. To make the equation look simple, since u is given, we set

$$g(t) \coloneqq T(t)Ax + T(t)f(0,x) + \int_0^t T(t-s)\partial_1 f(s,u(s))\,\mathrm{d}s\,.$$

Then we guess that u' is the (unique) solution of the integral equation

$$y(t) = g(t) + \int_0^t T(t-s)\partial_2 f(s,u(s))y(s)\,\mathrm{d}s\,. \tag{11.24}$$

Step 2. We show that (11.24) has a unique solution $y(t) \in C([0,T];X)$.

Since $f \in C^1([0,T]\times X;X)$ we get $g \in C([0,T];X)$. On the other hand, by the hypothesis $s \mapsto \partial_2 f(s,u(s)); [0,T] \to \mathcal{B}(X)$ is continuous, thus $(st,y) \mapsto \partial_2 f(s,u(s))y$ is continuous and satisfies the Lipschitz condition for y.Thus, by Exercise 11.9, there is unique $y(t) \in C([0,T];X)$ so that

$$y(t) = g(t) + \int_0^t T(t-s)\partial_2 f(s, u(s)) y'(s)\,\mathrm{d}s\,.$$

Step 3. We show that $u \in C^1([0,T;X])$ with $u'(t) = y(t)$ for $t \in [0,T]$.

Fix $t \in [0,T]$, for $h > 0$, we have

$$\begin{aligned}\frac{u(t+h)-u(t)}{h} &= \frac{T(t+h)x - T(t)x}{h} \\ &+ \frac{1}{h}\int_0^{t+h} T(t+h-s) f(s,u(s))\,\mathrm{d}s - \frac{1}{h}\int_0^t T(t-s) f(s,u(s))\,\mathrm{d}s\,.\end{aligned}$$

The second line can be rewritten as

$$\begin{aligned}&\frac{1}{h}\int_0^h T(t+h-s) f(s,u(s))\,\mathrm{d}s \\ &\quad + \frac{1}{h}\left\{\int_h^{t+h} T(t+h-s) f(s,u(s))\,\mathrm{d}s - \int_0^t T(t-s) f(s,u(s))\,\mathrm{d}s\right\} \\ &= \frac{1}{h}\int_0^h T(t+h-s) f(s,u(s))\,\mathrm{d}s \\ &\quad + \frac{1}{h}\left\{\int_0^t T(t-s)[f(s+h,u(s+h)) - f(s,u(s))]\,\mathrm{d}s\right\}\end{aligned}$$

Since $f \in C^1([0,T] \times X; X)$,

$$\begin{aligned}&f(s+h, u(s+h)) - f(s,u(s)) \\ &= \partial_1 f(s,u(s))h + \partial_2 f(s,u(s))[u(s+h)-u(s)] + r(h; u(s+h)-u(s))\,.\end{aligned}$$

By the finite-increment theorem (Section 10.4, Theorem 1 in *Mathematical Analysis II* by Zorich),

$$\|r(h; u(s+h)-u(s))\| \to 0 \quad \text{as} \quad h \downarrow 0 \quad \text{uniformly for} \quad s \in [0,T]\,.$$

We define $\Delta_h(t) := \frac{u(t+h)-u(t)}{h} - y(t)$ for $h > 0$. Then

$$\Delta_h(t) = \frac{T(t+h)x - T(t)x}{h} - T(t)Ax \tag{11.25}$$

$$+ \frac{1}{h}\int_0^h T(t+h-s)f(s,u(s))\,\mathrm{d}s - T(t)f(0,x) \tag{11.26}$$

$$+ \int_0^t T(t-s)\partial_2 f(s,u(s))\Delta_h(t)\,\mathrm{d}s \tag{11.27}$$

$$+ \frac{1}{h}\int_0^t T(t-s)r(h;u(s+h)-u(s))\,\mathrm{d}s\,. \tag{11.28}$$

Note that, by the dominated convergence theorem, (11.25), (11.26) and (11.27) converges to zero as $h \downarrow 0$, we get

$$\|\Delta_h(t)\| \leq \epsilon(h) + M'\int_0^t \|\Delta_h(s)\|\,\mathrm{d}s\,.$$

where

$$\lim_{h\downarrow 0}\epsilon(h) = 0 \quad \text{and} \quad M' := M\sup_{t\in[0,T]}\|\partial_2 f(s,u(s))\|\,.$$

Thus again by Grönwall's inequality, we get

$$\|\Delta_h(t)\| \leq \epsilon(h)e^{M't} \text{ for all } t \in [0,T]\,.$$

Letting $h \downarrow 0$, we conclude that

$$\lim_{h\downarrow 0}\|\Delta_h(t)\| = \lim_{h\downarrow 0}\left\|\frac{u(t+h)-u(t)}{h} - y(t)\right\| = 0\,.$$

By the same argument we can show that

$$\lim_{h\downarrow 0}\left\|\frac{u(t)-u(t-h)}{h} - y(t)\right\| = 0\,.$$

Then the desired result follows.

Step 4. Finally we show that u is a classical solution.Note that for $h > 0$ and $t \in [0,T]$,

$$\frac{T(h)-I}{h}u(t) = \frac{u(t+h)-u(t)}{h} - \frac{1}{h}\int_t^{t+h} T(t+h-s)f(s,u(s))\,\mathrm{d}s\,.$$

Since we have shown $u \in C^1([0,T];X)$, then $f(s,u(s)) \in C^1([0,T];X)$. Let $h \downarrow 0$, we get $u(t) \in D(A)$ and

$$Au(t) = u'(t) - f(t,u(t))\,.$$

We are done. $\square$

Appendix A

Basic Topology

A.1 Zorn's Lemma

Sometimes one wants to prove the existence of a mathematical object (which can be viewed as a maximal element in some partially ordered oset). One could try proving the existence of such an object by assuming there is no maximal element and using transfinite induction and the assumptions of the situation to get a contradiction. Zorn's lemma tidies up the conditions a situation needs to satisfy in order for such an argument to work. Therefore Zorn's lemma enables mathematicians to not have to repeat the transfinite induction argument by hand each time, but just check the conditions of Zorn's lemma.

Definition A.1. A binary relation $\preceq$ on a nonempty set P is a *partial order* if it satisfies the following properties: For all $x, y, z \in P$

(a) $\preceq$ is *reflexive*: $x \preceq x$;

(b) $\preceq$ is *antisymmetric*: if $x \preceq y$ and $y \preceq x$, then $x = y$;

(c) $\preceq$ is *transitive*: if $x \preceq y$ and $y \preceq z$, then $x \preceq z$.

In this case, say that $(P, \preceq)$ is a *partially ordered set*.

We give some examples for partial ordered set. Let $P = \mathbb{R}$ and take $\preceq$ to be $\leq$, the usual less than or equal to relation on $\mathbb{R}$; Let $P = \mathcal{P}(X)$ the power set of a set X and take $\preceq$ to be $\subseteq$, the usual set inclusion relation. Let $P = \mathcal{C}[0,1]$, the space of continuous real-valued functions on the interval [0,1] and take $\preceq$ to be the relation $\leq$ given by $f \leq g$ if and only if $f(x) \leq g(x)$ for each $x \in [0,1]$.

Definition A.2. Let C be a subset of a partially ordered set $(P, \preceq)$. An element $u \in P$ is an *upper bound* of C if $x \preceq u$ for every $x \in C$. An element $m \in C$ is said to be *maximal* if for any element $y \in C$, the relation $m \preceq y$ implies that $m = y$.

Definition A.3. Let $(P, \preceq)$ be a partially ordered set and $x, y \in P$. We say that x and y are *comparable* if either $x \preceq y$ or $y \preceq x$. Otherwise, x and y are *incomparable*. A partial order $\preceq$ is called a *total order* if any two elements of P are comparable. In this case we say that $(P, \preceq)$ is a or totally ordered set. A totally ordered set is also called a *chain*.

Theorem A.1 (Zorn's lemma)**.** *Let $(P, \preceq)$ be a partially ordered set. If each totally ordered subset of P has an upper bound, then P has a maximal element.*

Proof. A sketch of the proof of Zorn's lemma follows, assuming the axiom of choice. Suppose for contradiction that the lemma is false. Then for every element in P there is another element bigger than it. For every totally ordered subset T we may then define a bigger element $b(T)$, because T has an upper bound, and that upper bound has a bigger element. To actually define the function b, we need to employ the axiom of choice.

Using the function b, we are going to define elements $a_0 \prec a_1 \prec a_2 \prec a_3 \prec \dots$ in P (where $u \prec v$ means that $u \preceq v$ and $u \neq v$). This sequence is really long: the indices are not just the natural numbers, but all ordinals. In fact, the sequence is too long for the set P, there are too many ordinals (a proper class), more than there are elements in any set, and the set P will be exhausted before long and then we will run into the desired contradiction.

The a_i are defined by transfinite recursion: we pick a_0 in P arbitrary, and for any other ordinal w we set $a_w = b\left(\{a_v : v < w\}\right)$. Because the a_v are totally ordered, this is a well-founded definition. $\square$

Remark A.1. This proof shows that actually a slightly stronger version of Zorn's lemma is true: If P is a poset in which every well-ordered subset has an upper bound, and if x is any element of P, then P has a maximal element greater than or equal to x. That is, there is a maximal element which is comparable to x.

A.2 Metric space

A.2.1 Compact Subset of a Metric Space

Definition A.4. Let A be a subset of a metric space (X, d). we say A is *bounded* if A is contained in a ball of finite radius, i.e. there exists some $x \in X$ and $r > 0$ such that $A \subset B(x, r)$.

Definition A.5. A is a subset of a metric space (X, d) and $\varepsilon > 0$. A subset $F_\varepsilon \subset X$ is called an *ε-net* for A if each $x \in A$ there is an element $y \in F_\varepsilon$ such that $d(x, y) < \varepsilon$.

Definition A.6. A subset A of a metric space (X, d) is *totally bounded* if for any $\varepsilon > 0$ there is a *finite* ε-net $F_\varepsilon \subset X$ for A. That is, there is a finite set $F_\varepsilon \subset X$ such that

$$A \subset \bigcup_{x \in F_\varepsilon} B(x, \varepsilon).$$

Remark A.2. A subset A of a metric space (X, d) is totally bounded if and only if for any $\varepsilon > 0$ there is a finite ε-net $F_\varepsilon \subset A$ for A.

Obviously, every toally bounded set of a metric space is bounded. The following examples shows that boundedness does not, in general, imply total boundedness.

Example A.1. Let $X = \ell_2$ and $B = B(0,1) = \{x \in X : \|x\| \leq 1\}$. B is bounded but not totally bounded.

Theorem A.2. *A subset K of a metric space (X,d) is totally bounded if and only if every sequence in K has a Cauchy subsequence.*

Proof. Assume that K is totally bounded and let (x_n) be an infinite sequence in K. There is a finite set of points $\{y_{11}, y_{12}, \ldots, y_{1r}\}$ in K such that

$$K \subset \bigcup_{j=1}^{r} B\left(y_{1j}, \frac{1}{2}\right).$$

At least one of the balls $B\left(y_{1j}, \frac{1}{2}\right), j = 1, 2, \ldots, r$, contains an infinite subsequence (x_{n1}) of (x_n). Again, there is a finite set $\{y_{21}, y_{22}, \ldots, y_{2s}\}$ in K such that

$$K \subset \bigcup_{j=1}^{s} B\left(y_{2j}, \frac{1}{2^2}\right).$$

At least one of the balls $B\left(y_{2j}, \frac{1}{2^2}\right), j = 1, 2, \ldots, s$, contains an infinite subsequence (x_{n2}) of (x_{n1}) Continuing in this way, at the m-th step, we obtain a subsequence (x_{nm}) of $\left(x_{n(m-1)}\right)$ which is contained in a ball of the form $B\left(y_{mj}, \frac{1}{2^m}\right)$.

Claim: The diagonal subsequence (x_{nn}) of (x_n) is Cauchy.
Indeed, if $m > n$, then both x_{nn} and x_{mm} are in the ball of radius 2^{-n}. Hence, by the triangle inequality,

$$\|x_{nn} - x_{mm}\| < 2^{1-n} \to 0 \text{ as } n \to \infty.$$

Conversely, assume that every sequence in K has a Cauchy subsequence and that K is not totally bounded. Then, for some $\epsilon > 0$, no finite ϵ -net exists for K. Hence, if $x_1 \in K$, then there is an $x_2 \in K$ such that

$\|x_1 - x_2\| \geq \epsilon$. Otherwise, $\|x_1 - y\| < \epsilon$ for all $y \in K$ and consequently $\{x_1\}$ is a finite ϵ -net for K, a contradiction.) Similarly, there is an $x_3 \in K$ such that

$$\|x_1 - x_3\| \geq \epsilon \text{ and } \|x_2 - x_3\| \geq \epsilon$$

Continuing in this way, we obtain a sequence (x_n) in K such that $\|x_n - x_m\| \geq \epsilon$ for all $m \neq n$. Therefore (x_n) cannot have a Cauchy subsequence, a contradiction. $\square$

Definition A.7. A metric space (X, d) is called *sequentially compact* if every sequence in X has a convergent subsequence, a subset K of X is sequentially compact when (K, d) is sequentially compact.

Theorem A.3. *A subset of a metric space is sequentially compact if and only if it is totally bounded and complete.*

Proof. Let K be a sequentially compact subset of a normed linear space (X, d), and (x_n) be a sequence in K. By sequential compactress of K, (x_n) has a subsequence (x_{n_k}) which converges in K. since every convergent sequence is Cauchy, by Theorem A.2, K is totally bounded.

Conversely, assume that K is a totally bounded and complete subset of a normed linear space (X, d). Let (x_n) be a sequence in K. By Theorem A.2, (x_n) has a Cauchy subsequence (x_{n_k}) . since K is complete, (x_{n_k}) converges in K. Hence K is sequentially compact. $\square$

Definition A.8. A metric space is called *compact* if each of its open covers has a finite subcover, a subset K of X is compact when (K, d) is compact.

Theorem A.4. (X, d) *is a metric space and* $K \subset X$. *Then* K *is compact if and only if* K *is sequentially compact.*

Proof. $\square$

A.2.2 Baire's Category Theorem

Definition A.9. A subset S of a metric space (X, d) is called *nowhere dense* in X if the closure of S contains no interior points.

Theorem A.5. *Let (X, d) be a complete metric space.*

(a) If (G_n) is a sequence of nonempty, open and dense subsets of X then $G = \cap_{n\in\mathbb{N}} G_n$ is dense in X.

(b) If (F_n) is a sequence of closed, nowhere dense subsets of X, then $F = \cup_{n\in\mathbb{N}} F_n$ cotains no interior points.

Proof. We ony need to show 1. Let $x \in X$ and $\epsilon > 0$. Since G_1 is dense in X, there is a point x_1 in the open set $G_1 \cap B(x, \epsilon)$. Let r_1 be a number such that $0 < r_1 < \frac{\epsilon}{2}$ and $B(x_1, r_1) \subset G_1 \cap B(x, \epsilon)$. By induction, we obtain a sequence (x_n) in X and a sequence (r_n) of radii such that for each n, $0 < r_n < \frac{\epsilon}{2^n}$, and

$$\overline{B(x_{n+1}, r_{n+1})} \subset G_{n+1} \cap B(x_n, r_n) \text{ and } \overline{B(x_1, r_1)} \subset G_1 \cap B(x, \epsilon)$$

Hence, (x_n) is a Cauchy sequence in X. since X is complete, there is a $y \in X$ such that $x_n \to y$ as $n \to \infty$. since x_k lies in the closed set $\overline{B(x_n, r_n)}$ if $k > n$, it follows that y lies in each $\overline{B(x_n, r_n)}$. Hence y lies in each G_n. That is, $G = \bigcap_{n\in\mathbb{N}} G_n \neq \emptyset$. It is also clear that $y \in B(x, \epsilon)$. $\square$

A subset S of a metric space (X, d) is said to be

(a) of *first category* or *meagre* in X if S can be written as a countable union of sets which are nowhere dense in X.

(b) of *second category* or *nonmeagre* in X if it is not of first category in X.

Theorem A.6 (Baire's Catagory Theorem)**.** *A complete metric space (X, d) is of second category in itself.*

A.3 Sum of Uncountable Series

Definition A.10. X 是一赋范线性空间, $\{\alpha_i\}_{i\in I}$ 是 X 中一族向量, 记 I 的有限子集全体为 $\mathcal{I}$. 容易知道 $(\mathcal{I},\subset)$ 是一定向集, 故 $\{\sum_{i\in S}\alpha_i : S\in\mathcal{I}\}$ 是 X 中的网, 若网 (依范数诱导的拓扑) 收敛 β , 就称级数 $\sum_{i\in I}\alpha_i$ 收敛 (有时也称为无条件收敛,unconditionally convergent) 于 β, 记为 $\beta=\sum_{i\in I}\alpha_i$.

Remark A.3. 显然, $\sum_{i\in I}\alpha_i$ 收敛于 β 即是: 对任意 $\epsilon>0$, 存在 I 的有限子集 S, 使得 I 任何包含 S 的有限子集 T , 有

$$\|\sum_{i\in T}\alpha_i-\beta\|\le\epsilon$$

Proposition A.7. *X 是一赋范线性空间,$\alpha_i,\beta_i\in X(i\in I)$, 且 $\sum_{i\in I}\alpha_i$, $\sum_{i\in I}\beta_i$ 皆收敛. 则*

(a) $\sum_{i\in I}(\alpha_i+\beta_i)$ 收敛, 且 $\sum_{i\in I}(\alpha_i+\beta_i)=\sum_{i\in I}\alpha_i+\sum_{i\in I}\beta_i$.

(b) 任意 $\lambda\in\mathbb{F}$, $\sum_{i\in I}\lambda\alpha_i$ 收敛, 且 $\sum_{i\in I}\lambda\alpha_i=\lambda\sum_{i\in I}\alpha_i$.

(c) A 是 X 到一赋范线性空间 Y 的有界线性算子, 则 $\sum_{i\in I}A\alpha_i$ 是 Y 收敛的级数, 且 $\sum_{i\in I}A\alpha_i=A(\sum_{i\in I}\alpha_i)$.

Proof. 利用定义易证. □

Proposition A.8. *X 是赋范线性空间,$\alpha_i\in X, i\in I$, 且 $\sum_{i\in I}\alpha_i$ 收敛. 则对任意 $\epsilon>0$, 存在 I 的有限子集 S, 使得 I 任何与 S 不交的有限子集 J, 有*

$$\|\sum_{i\in J}\alpha_i\|\le\epsilon$$

当 X 是 Banach 空间时, 逆命题也成立.

Proof. 逆命题证明时, 构造柯西列利用空间完备性. □

Corollary A.9. *$\sum_{i\in I}\alpha_i$ 收敛, 则只有至多可数项 α_i 非零.*

Proof. 利用上述命题, 对任何 n, 有 $S_n \in \mathcal{I}$ 对任何与 S_n 不交的 J, 有

$$\|\sum_{i\in J}\alpha_i\| \leq \frac{1}{n}$$

令 $S = \bigcup_{n=1}^{\infty} S_n$. 可见只有 $i \notin S$ 时必然有 $\alpha_i = 0$. □

Proposition A.10. X 是一赋范线性空间,$\alpha_i \in X(i \in I)$, 且 I 是可数集. 则 $\sum_{i\in I}\alpha_i$ 收敛于 β 当且仅当对任意 $\mathbb{N}$ 到 I 的双射 f, 级数 $\sum_{k=1}^{\infty}\alpha_{f(k)}$ 皆收敛于 β .

Proof. 充分性用反证法构造矛盾, 必要性易证 □

Proposition A.11. $x_i \in \mathbb{R}$, 且 $x_i \geq 0\, i \in I$. 则 $\sum_{i\in I} x_i$ 收敛于 $y \in \mathbb{R}$ 当且仅当

$$y = \sup\{\sum_{i\in S} x_i : S \text{ 是} I \text{ 有限子集}\}.$$