Functional Analysis: Spectral Theory

V.S. Sunder
Institute of Mathematical Sciences
Madras 600113
INDIA

July 31, 2000

ha ha

ha ha

ha ha

ha ha

Preface

This book grew out of a course of lectures on functional analysis that the author gave during the winter semester of 1996 at the Institute of Mathematical Sciences, Madras. The one difference between the course of lectures and these notes stems from the fact that while the audience of the course consisted of somewhat mature students in the Ph.D. programme, this book is intended for Master's level students in India; in particular, an attempt has been made to fill in some more details, as well as to include a somewhat voluminous Appendix, whose purpose is to (at least temporarily) fill in the possible gaps in the background of the prospective Master's level reader.

The goal of this book is to begin with the basics of normed linear spaces, quickly specialise to Hilbert spaces and to get to the spectral theorem for (bounded as well as unbounded) operators on separable Hilbert space.

The first couple of chapters are devoted to basic propositions concerning normed vector spaces (including the usual Banach space results - such as the Hahn-Banach theorems, the Open Mapping theorem, Uniform boundedness principle, etc.) and Hilbert spaces (orthonormal bases, orthogonal complements, projections, etc.).

The third chapter is probably what may not usually be seen in a first course on functional analysis; this is no doubt influenced by the author's conviction that the only real way to understand the spectral theorem is as a statement concerning representations of commutative C^*-algebras. Thus, this chapter begins with the standard Gelfand theory of commutative Banach algebras, and proceeds to the Gelfand-Naimark theorem on commutative C^*-algebras; this is then followed by a discussion of representations of (not necessarily commutative) C^*-algebras (including the GNS construction which establishes the correspondence between cyclic representations and states on the C^*-algebra, as well as the so-called 'non-commutative Gelfand Naimark theorem' which asserts that C^*-algebras admit faithful representations on Hilbert space); the final section of this chapter is devoted to the Hahn-Hellinger classification of separable representations of a commutative C^*-algebra (or equivalently, the classification of separable spectral measures on the Borel σ-algebra of a compact Hausdorff space).

The fourth chapter is devoted to more standard 'operator theory' in Hilbert space. To start with, the traditonal form of the spectral theorem for a normal operator on a separable Hilbert space is obtained as a special case of the theory discussed in Chapter 3; this is followed by a discussion of the polar decomposition of operators; we then discuss compact operators and the spectral decomposition of normal compact operators, as well as the singular value decomposition of general compact operators. The final section of this chapter is devoted to the classical facts concerning Fredholm operators and their 'index theory'.

The fifth and final chapter is a brief introduction to the theory of unbounded operators on Hilbert space; in particular, we establish the spectral and polar decomposition theorems.

A fairly serious attempt has been made at making the treatment almost self-contained. There are seven sections in the Appendix which are devoted to: (a) discussing and establishing some basic results in each of the following areas: linear algebra, transfinite considerations (including Zorn's lemma and a 'naive' treatment of infinite cardinals), general topology, compact and locally compact Hausdorff spaces, and measure theory; and (b) a proof of the Stone-Weierstrass theorem, and finally, a statement of the Riesz Representation theorem (on measures and continuous functions). Most statements in the appendix are furnished with proofs, the exceptions to this being the sections on measure theory and the Riesz representation theorem.

The intended objective of the numerous sections in the Appendix is this: if a reader finds that (s)he does not know some 'elementary' fact from, say linear algebra, which seems to be used in the main body of the book, (s)he can go to the pertinent section in the Appendix, to attempt a temporary stop-gap filler. The reader who finds the need to pick up a lot of 'background material' from some section of the Appendix, should, at the earliest opportunity, try to fill in the area in which a lacuna in one's training has been indicated. In particular, it should be mentioned that the treatment in Chapter 3 relies heavily on various notions from measure theory, and the reader should master these prerequisites before hoping to master the material in this chapter. A few 'standard' references have been listed in the brief bibliography; these should enable the reader to fill in the gaps in the appendix.

Since the only way to learn mathematics is to do it, the book is punctuated with a lot of exercises; often, the proofs of some propositions in the text are relegated to the exercises; further, the results stated in the exercises are considered to be on the same footing as 'properly proved' propositions, in the sense that we have freely used the statements of exercises in the subsequent treatment in the text. Most exercises, whose solutions are not immediate, are furnished with fairly elaborate hints; thus, a student who is willing to sit down with pen and paper and 'get her hands dirty' should be able to go through the entire book without too much difficulty.

Finally, all the material in this book is very 'standard' and no claims for originality are made; on the contrary, we have been heavily influenced by the treatment of [Sim] and [Yos] in various proofs; thus, for instance, the proofs given here for the Open Mapping Theorem, Urysohn's lemma and Alexander's subbase theorem, are more or less the same as the ones found in [Sim], while the proofs of the Weierstrass as well as the Stone-Weierstrass theorems are almost identical to the ones in [Yos]; furthermore, the treatment in §A.2 has been influenced a little by [Hal], but primarily, the proofs of this section on cardinal numbers are a specialisation of the proofs in [MvN] for corresponding statements in a more general context.

Acknowledgements :

The author would like to thank the Institute of Mathematical Sciences for the excellent facilities it provided him during his preparation of these notes; he also wishes to thank the students of his class - in particular, M. Rajesh - for having made the process of giving the course an enjoyable experience; he would also like to thank R. Bhatia and M. Krishna for their constant encouragement in this venture.

Finally, the author would like to thank Paul Halmos for having taught him most of this material in the first instance, for having been a great teacher, and later, colleague and friend. This book is fondly dedicated to him and to the author's parents (for reasons too numerous to go into here).

Contents

Chapter 1

Normed spaces

1.1 Vector spaces

We begin with the fundamental notion of a vector space.

DEFINITION 1.1.1 *A* **vector space** *is a (non-empty) set V that comes equipped with the following structure:*

(a) There exists a mapping $V \times V \to V$, denoted $(x, y) \mapsto x + y$, referred to as **vector addition***, which satisfies the following conditions, for arbitrary $x, y, z \in V$:*

(i) (commutative law) $x + y = y + x$;

(ii) (associative law) $(x + y) + z = x + (y + z)$;

(iii) (additive identity) there exists an element in V, always denoted by 0 (or 0_V, if it is important to draw attention to V), such that $x + 0 = x$;

(iv) (additive inverses) for each x in V, there exists an element of V, denoted by $-x$, such that $x + (-x) = 0$.

(b) There exists a map $\mathbb{C} \times V \to V$, denoted by $(\alpha, x) \mapsto \alpha x$, referred to as **scalar multiplication***, which satisfies the following conditions, for arbitrary $x, y \in V$, $\alpha, \beta \in \mathbb{C}$:*

(i) (associative law) $\alpha(\beta x) = (\alpha\beta)x$;

(ii) (distributive law) $\alpha(x + y) = \alpha x + \alpha y$, and $(\alpha + \beta)x = \alpha x + \beta x$; and

(iii) (unital law) $1x = x$.

Remark 1.1.2 Strictly speaking, the preceding axioms define what is usually called a 'vector space over the field of complex numbers' (or simply, a complex vector space). The reader should note that the definition makes perfect sense if every occurrence of $\mathbb{C}$ is replaced by $\mathbb{R}$; the resulting object would be a vector space over the field of real numbers (or a 'real vector space'). More generally, we could replace $\mathbb{C}$ by an abstract field $\mathbb{K}$, and the result would be a vector space over $\mathbb{K}$, and $\mathbb{K}$ is called the 'underlying field' of the vector space. However, in these notes, we shall always confine ourselves to **complex vector spaces**, although we shall briefly discuss vector spaces over general fields in §A.1.

Furthermore, the reader should verify various natural consequences of the axioms, such as: (a) the 0 of a vector space is unique; (b) additive inverses are unique - meaning that if $x, y \in V$ and if $x + y = 0$, then necessarily $y = -x$; more generally, we have the cancellation law, meaning: if $x, y, z \in V$ and if $x + y = x + z$, then $y = z$; (c) thanks to associativity and commutativity of vector addition, the expression $\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$ has an unambiguous (and natural) meaning, whenever $x_1, \cdots, x_n \in V$. In short, all the 'normal' rules of arithmetc hold in a general vector space. □

Here are some standard (and important) examples of vector spaces. The reader should have little or no difficulty in checking that these are indeed examples of vector spaces (by verifying that the operations of vector addition and scalar multiplication, as defined below, do indeed satisfy the axioms listed in the definition of a vector space).

Example 1.1.3 (1) $\mathbb{C}^n$ is the (unitary space of all $n-tuples$ of complex numbers; its typical element has the form $(\xi_1, \cdots, \xi_n)$, where $\xi_i \in \mathbb{C} \ \forall \ 1 \leq i \leq n$. Addition and scalar multiplication are defined co-ordinatewise: thus, if $x = (\xi_i)$, $y = (\eta_i)$, then $x + y = (\xi_i + \eta_i)$ and $\alpha x = (\alpha \xi_i)$.

(2) If X is any set, let $\mathbb{C}^X$ denote the set of all complex-valued functions on the set X, and define operations co-ordinatewise, as before, thus: if $f, g \in \mathbb{C}^X$, then $(f + g)(x) = f(x) + g(x)$ and

$(\alpha f)(x) = \alpha f(x)$. (The previous example may be regarded as the special case of this example with $X = \{1, 2, \cdots, n\}$.)

(3) We single out a special case of the previous example. Fix positive integers m, n; recall that an $m \times n$ **complex matrix** is a rectangular array of numbers of the form

$$A = \left[\begin{array}{cccc} a_1^1 & a_2^1 & \cdots & a_n^1 \\ a_1^2 & a_2^2 & \cdots & a_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_1^m & a_2^m & \cdots & a_n^m \end{array} \right] .$$

The horizontal (resp., vertical) lines of the matrix are called the **rows** (resp., **columns**) of the matrix; hence the matrix A displayed above has m rows and n columns, and has entry a_j^i in the unique position shared by the i-th row and the j-th column of the matrix.

We shall, in the sequel, simply write $A = ((a_j^i))$ to indicate that we have a matrix A which is related to a_j^i as stated in the last paragraph.

The collection of all $m \times n$ complex matrices will always be denoted by $M_{m\times n}(\mathbb{C})$ - with the convention that we shall write $M_n(\mathbb{C})$ for $M_{n\times n}(\mathbb{C})$.

It should be clear that $M_{m\times n}(\mathbb{C})$ is in natural bijection with $\mathbb{C}^X$, where $X = \{(i, j) : 1 \leq i \leq m, 1 \leq j \leq n\}$.

□

The following proposition yields one way of constructing lots of new examples of vector spaces from old.

PROPOSITION 1.1.4 *The following conditions on a non-empty subset $W \subset V$ are equivalent:*

(i) $x, y \in W,\ \alpha \in \mathbb{C} \ \Rightarrow \ x + y, \alpha x \ \in \ W$;

(ii) $x, y \in W,\ \alpha \in \mathbb{C} \ \Rightarrow \ \alpha x + y \ \in \ W$;

(iii) W is itself a vector space if vector addition and scalar multiplication are restricted to vectors coming from W.

A non-empty subset W which satisfies the preceding equivalent conditions is called a **subspace** *of the vector space V.*

The (elementary) proof of the proposition is left as an exercise to the reader. We list, below, some more examples of vector spaces which arise as subspaces of the examples occurring in Example 1.1.3.

EXAMPLE 1.1.5 (1) $\{x = (\xi_1, \cdots, \xi_n) \in \mathbb{C}^n : \sum_{i=1}^n \xi_n = 0\}$ is a subspace of $\mathbb{C}^n$ and is consequently a vector space in its own right.

(2) The space $C[0,1]$ of continuous complex-valued functions on the unit interval [0,1] is a subspace of $\mathbb{C}^{[0,1]}$, while the set $C^k(0,1)$ consisting of complex-valued functions on the open interval (0,1) which have k continuous derivatives is a subspace of $\mathbb{C}^{(0,1)}$. Also, $\{f \in C[0,1] : \int_0^1 f(x)dx = 0\}$ is a subspace of $C[0,1]$.

(3) The space $\ell^\infty = \{(\alpha_1, \alpha_2, \cdots, \alpha_n, \cdots) \in \mathbb{C}^{\mathbb{N}} : sup_n|\alpha_n| < \infty\}$ (of bounded complex sequences), where $\mathbb{N}$ denotes the set of natural numbers. may be regarded as a subspace of $\mathbb{C}^{\mathbb{N}}$, ; similarly the space $\ell^1 = \{(\alpha_1, \alpha_2, \cdots, \alpha_n, \cdots) : \alpha_n \in \mathbb{C}, \sum_n |\alpha_n| < \infty\}$ (of absolutely summable complex sequences) may be regarded as a subspace of ℓ^∞ (and consequently of $\mathbb{C}^{\mathbb{N}}$.)

(4) The examples in (3) above are the two extremes of a continuum of spaces defined by $\ell^p = \{(\alpha_1, \alpha_2, \cdots, \alpha_n, \cdots) \in \mathbb{C}^{\mathbb{N}} : \sum_n |\alpha_n|^p < \infty\}$, for $1 \leq p < \infty$. It is a (not totally trivial) fact that each ℓ^p is a vector subspace of $\mathbb{C}^{\mathbb{N}}$.

(5) For $p \in [1, \infty]$, $n \in \mathbb{N}$, the set $\ell^p_n = \{\alpha \in \ell^p : \alpha_k = 0 \ \forall \ k > n\}$ is a subspace of ℓ^p which is in natural bijection with $\mathbb{C}^n$. $\square$

1.2 Normed spaces

Recall that a set X is called a **metric space** if there exists a function $d : X \times X \to [0, \infty)$ which satisfies the following conditions, for all $x, y, z \in X$:

$$d(x,y) = 0 \Leftrightarrow x = y \tag{1.2.1}$$

$$d(x,y) = d(y,x) \tag{1.2.2}$$

$$d(x,y) \quad \leq \quad d(x,z) \ + \ d(z,y) \ . \tag{1.2.3}$$

The function d is called a **metric** on X and the pair (X, d) is what is, strictly speaking, called a metric space.

The quantity $d(x, y)$ is to be thought of as the distance between the points x and y. Thus, the third condition on the metric is the familiar *triangle inequality.*

The most familiar example, of course, is $\mathbb{R}^3$, with the metric defined by

$$d((x_1, x_2, x_3), (y_1, y_2, y_3)) \ = \ \sqrt{\sum_{i=1}^{3} (x_i - y_i)^2}.$$

As with vector spaces, any one metric space gives rise to many metric spaces thus: if Y is a subset of X, then $(Y, d|_{Y\times Y})$ is also a metric space, and is referred to as a (metric) subspace of X. Thus, for instance, the unit sphere $S^2 \ = \ \{x \in \mathbb{R}^3 : d(x, 0) = 1\}$ is a metric space; and there are clearly many more interesting metric subspaces of $\mathbb{R}^3$. It is more customary to write $||x|| \ = \ d(x, 0)$, refer to this quantity as the **norm** of x and to think of S^2 as the set of vectors in $\mathbb{R}^3$ of unit norm.

The set $\mathbb{R}^3$ is an example of a set which is simultaneously a metric space as well as a vector space (over the field of real numbers), where the metric *arises from a norm.* We will be interested in more general such objects, which we now pause to define.

DEFINITION 1.2.1 *A norm on a vector space V is a function $V \ni x \ \mapsto \ ||x|| \in [0, \infty)$ which satisfies the following conditions, for all $x, y \in V, \ \alpha \in \mathbb{C}$:*

$$||x|| \ = \ 0 \ \Leftrightarrow \ x = 0 \tag{1.2.4}$$

$$||\alpha x|| \ = \ |\alpha| \ ||x|| \tag{1.2.5}$$

$$||x + y|| \ \leq \ ||x|| + ||y|| \tag{1.2.6}$$

and a **normed vector space** *is a pair $(V, || \cdot ||)$ consisting of a vector space and a norm on it.*

EXAMPLE 1.2.2 (1) It is a fact that ℓ^p (and hence ℓ^p_n, for any $n \in \mathbb{N}$) is a normed vector space with respect to the norm defined by

$$||\alpha||_p = \left(\sum_k |\alpha_k|^p \right)^{\frac{1}{p}} . \tag{1.2.7}$$

This is easy to verify for the cases $p = 1$ and $p = \infty$; we will prove in the sequel that $|| \cdot ||_2$ is a norm; we will not need this fact for other values of p and will hence not prove it. The interested reader can find such a proof in [Sim] for instance; it will, however, be quite instructive to try and prove it directly. (The latter exercise/effort will be rewarded by a glimpse into notions of duality between appropriate pairs of convex maps.)

(2) Given a non-empty set X, let $B(X)$ denote the space of bounded complex-valued functions on X. It is not hard to see that $B(X)$ becomes a normed vector space with respect to the norm (denoted by $|| \cdot ||_\infty$) defined by

$$||f||_\infty = \sup\{|f(x)| \ : \ x \in X\} . \tag{1.2.8}$$

(The reader, who is unfamiliar with some of the notions discussed below, might care to refer to §A.4 and §A.6 in the Appendix, where (s)he will find definitions of compact and locally compact spaces, respectively, and will also have the pleasure of getting acquainted with 'functions vanishing at infinity'.)

It should be noted that if X is a compact Hausdorff space, then the set $C(X)$ of continuous complex-valued functions on X is a vector subspace of $B(X)$ (since continuous functions on X are necessarily bounded), and consequently a normed vector space in its own right).

More generally, if X is a locally compact Hausdorff space, let $C_0(X)$ denote the space of all complex-valued continuous functions on X which 'vanish at infinity'; thus $f \in C_0(X)$ precisely when $f : X \to \mathbb{C}$ is a continuous function such that for any positive ϵ, it is possible to find a compact subset $K \subset X$ such that $|f(x)| < \epsilon$ whenever $x \notin K$. The reader should verify that $C_0(X) \subset B(X)$ and hence, that $C_0(X)$ is a normed vector space with respect to $|| \cdot ||_\infty$.

(3) Let $C_b^k(0,1) = \{f \in C^k(0,1) : ||f^{(j)}||_\infty < \infty$ for all $0 \leq j \leq k\}$, where we write $f^{(j)}$ to denote the j-th derivative of the function f - with the convention that $f^{(0)} = f$. This becomes a normed space if we define

$$||f|| = \sum_{j=0}^{k} ||f^{(j)}||_\infty . \qquad (1.2.9)$$

More generally, if Ω is an open set in $\mathbb{R}^n$, let $C_b^k(\Omega)$ denote the space of complex-valued functions on Ω which admit continuous partial derivatives of order at most k which are uniformly bounded functions. (This means that for any multi-index $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_n)$ where the α_i are any non-negative integers satisfying $|\alpha| = \sum_{j=1}^n \alpha_j \leq k$, the partial derivative $\partial^\alpha f = \frac{\partial^{|\alpha|}}{\partial_{x_1}^{\alpha_1} \partial_{x_2}^{\alpha_2} \cdots \partial_{x_n}^{\alpha_n}} f$ exists, is continuous - with the convention that if $\alpha_j = 0 \; \forall \; j$, then $\partial^\alpha f = f$ - and is bounded.) Then, the vector space $C_b^k(\Omega)$ is a normed space with respect to the norm defined by

$$||f|| = \sum_{\{\alpha : 0 \leq |\alpha| \leq k\}} ||\partial^\alpha f||_\infty .$$

$\square$

Note that - verify ! - a normed space is a vector space which is a metric space, the metric being defined by

$$d(x,y) = ||x - y|| . \qquad (1.2.10)$$

It is natural to ask if the 'two structures'- i.e., the vector (or linear) and the metric (or topological) structures - are compatible; this amounts to asking if the vector operations are continuous.

Recall that a function $f : X \to Y$ between metric spaces is said to be **continuous** at $x_0 \in X$ if, whenever $\{x_n\}$ is a sequence in X which converges to x_0 - i.e., if $\lim_n d_X(x_n, x_0) = 0$ - then also the sequence $\{f(x_n)\}$ converges to $f(x_0)$; further, the function f is said to be continuous if it is continuous at each point of X.

EXERCISE 1.2.3 *(1) If X, Y are metric spaces, there are several ways of making $X \times Y$ into a metric space; if d_X, d_Y denote* *the metrics on X, Y respectively, verify that both the following specifications define metrics on $X \times Y$:*

$$\begin{aligned} d_1(\ (x_1, y_1), (x_2, y_2)\ ) &= d_X(x_1, x_2) + d_Y(y_1, y_2) \\ d_\infty(\ (x_1, y_1), (x_2, y_2)\ ) &= max\{d_X(x_1, x_2), d_Y(y_1, y_2)\}\ . \end{aligned}$$

Also show that a sequence $\{(x_n, y_n)\}$ in $X \times Y$ converges to a point (x_0, y_0) with respect to either of the metrics $d_i, i \in \{1, \infty\}$ if and only if we have 'co-ordinate-wise convergence' - i.e., if and only if $\{x_n\}$ converges to x_0 with respect to d_X, and $\{y_n\}$ converges to y_0 with respect to d_Y.

(2) Suppose X is a normed space; show that each of the following maps is continuous (where we think of the product spaces in (a) and (c) below as being metric spaces via equation 1.2.10 and Exercise (1) above):
(a) $X \times X \ \ni \ (x, y) \ \mapsto \ (x + y) \in \ X$;
(b) $X \ni x \ \mapsto \ -x \in X$;
(c) $\mathbb{C} \times X \ \ni \ (\alpha, x) \ \mapsto \ \alpha x \ \in \ X$.

(3) Show that a composite of continuous functions between metric spaces - which makes sense in a consistent fashion only if the domains and targets of the maps to be composed satisfy natural inclusion relations - is continuous, as is the restriction to a subspace of the domain of a continuous function.

1.3 Linear operators

The focus of interest in the study of normed spaces is on the appropriate classes of mappings between them. In modern parlance, we need to identify the proper class of *morphisms* in the category of normed spaces. (It is assumed that the reader is familiar with the rudiments of linear algebra; the necessary material is quickly discussed in the appendix - see §A.1 - for the reader who does not have the required familiarity.)

If we are only interested in vector spaces, we would concern ourselves with **linear transformations** between them - i.e., if V, W are vector spaces, we would be interested in the class $L(V, W)$ of mappings $T : V \to W$ which are linear in the sense that

$$T(\alpha x + \beta y) \;=\; \alpha Tx + \beta Ty \;,\; \forall\, \alpha, \beta \;\in \mathbb{C},\; x, y \in V \quad (*)$$

When $V = W$, we write $L(V) \;=\; L(V, V)$.

We relegate some standard features (with the possible exception of (2)) of such linear transformations in an exercise.

EXERCISE 1.3.1 *(1) If* $T \in L(V, W)$, *show that* T *preserves collinearity, in the sense that if* x, y, z *are three points (in* V*) which lie on a straight line, then so are* Tx, Ty, Tz *(in* W*).*

(2) If $V = \mathrm{R}^3$ *(or even* R^n*), and if* $T : V \to V$ *is a mapping which (i) preserves collinearity (as in (1) above), and (ii)* T *maps* V *onto* W*, then show that* T *must satisfy the algebraic condition displayed in (*). (This may be thought of as a reason for calling the above algebraic condition 'linearity'.)*

(3) Show that the following conditions on a linear transformation $T \in L(V, W)$ *are equivalent:*

(i) T *is 1-1 ; i.e.,* $x, y \in V,\; Tx = Ty \;\Rightarrow\; x = y$;

(ii) ker T $(= \;\{x \in V : Tx = 0\}) \;= \{0\}$;

(iii) T *'preserves linear independence', meaning that if* $X = \{x_i : i \in I\} \;\subset V$ *is a linearly independent set, so is* $T(X) = \{Tx_i : i \in I\} \;\subset W$*. (A set* X *as above is said to be linearly independent if, whenever* $x_1, \cdots, x_n \in X$*, the only choice of scalars* $\alpha_1, \cdots, \alpha_n \in \mathbb{C}$ *for which the equation* $\alpha_1 x_1 + \cdots + \alpha_n x_n = 0$ *is satisfied is the trivial one:* $\alpha_1 = \cdots = \alpha_n = 0$*. In particular, any subset of a linearly independent set is also linearly independent. A set which is not linearly independent is called 'linearly dependent'. It must be noted that* $\{0\}$ *is a linearly dependent set, and consequently, that any set containing 0 is necessarily linearly dependent.)*

(4) Show that the following conditions on a linear transformation $T \in L(V, W)$ *are equivalent:*

(i) T *is* **invertible** *- i.e., there exists* $S \in L(W, V)$ *such that* $ST \;=\; id_V$ *and* $TS = \; id_W$*, where we write juxtaposition (*ST*) for composition (*$S \circ T$*);*

(ii) T *is 1-1 and onto;*

(iii) if $X = \{x_i : i \in I\}$ is a basis for V - equivalently, a maximal linearly independent set in V, or equivalently, a linearly independent spanning set for V - then $T(X)$ is a basis for W. (A linearly independent set is maximal if it is not a proper subset of a larger linearly independent set.)

(5) When the equivalent conditions of (4) are satisfied, T is called an **isomorphism** *and the spaces V and W are said to be* **isomorphic***; if T is an isomorphism, the transformation S of (i) above is unique, and is denoted by T^{-1}.*

(6) Show that $GL(V) = \{T \in L(V) : T \text{ is invertible}\}$ is a group under multiplication.

(7) (i) If $B_V = \{x_1, x_2, \cdots, x_n\}$ is a basis for V, and $B_W = \{y_1, y_2, \cdots, y_m\}$ is a basis for W, show that there is an isomorphism between the spaces $L(V,W)$ and $M_{m\times n}(\mathbb{C})$ given by $L(V,W) \ni T \mapsto [T]_{B_V}^{B_W}$, where the matrix $[T]_{B_V}^{B_W} = ((t_j^i))$ is defined by

$$Tx_j = \sum_{i=1}^{m} t_j^i y_i \ . \tag{1.3.11}$$

(ii) Show that $[ST]_{B_{V_1}}^{B_{V_3}} = [S]_{B_{V_2}}^{B_{V_3}}[T]_{B_{V_1}}^{B_{V_2}}$ where this makes sense. (Recall that if $A = ((a_k^i))$ is an $m \times n$ matrix, and if $B = ((b_j^k))$ is an $n \times p$ matrix, then the product AB is defined to be the $m \times p$ matrix $C = ((c_j^i))$ given by

$$c_j^i = \sum_{k=1}^{n} a_k^i b_j^k \ , \ \forall\ 1 \le i \le m, 1 \le j \le p \ .)$$

(iii) If we put $W = V$, $y_i = x_i$, show that the isomorphism given by (i) above maps $GL(V)$ onto the group $Gl(n,\mathbb{C})$ of invertible $n \times n$ matrices.

(iv) The so-called **standard basis** *$B_n = \{e_i : 1 \le i \le n\}$ for $V = \mathbb{C}^n$ is defined by $e_1 = (1, 0, \cdots, 0), \cdots, e_n = (0, 0, \cdots, 1)$; show that this is indeed a basis for $\mathbb{C}^n$.*

(v) (In the following example, assume that we are working over the real field $\mathbb{R}$.) Compute the matrix $[T]_{B_2}^{B_2}$ of each of the following (obviously) linear maps on $\mathbb{R}^2$:

(a) T is the rotation about the origin in the counter-clockwise direction by an angle of $\frac{\pi}{4}$;

(b) T is the projection onto the x-axis along the y-axis;
(c) T is mirror-reflection in the line $y = x$.

When dealing with normed spaces - which are simultaneously vector spaces and metric spaces - the natural class of mappings to consider is the class of linear transformations which are continuous.

PROPOSITION 1.3.2 *Suppose V, W are normed vector spaces, and suppose $T \in L(V, W)$; then the following conditions on T are equivalent:*
(i) T is continuous at 0 $(= 0_V)$;
(ii) T is continuous;
(iii) The quantity defined by

$$||T|| \;=\; \sup\{||Tx|| : x \in V, ||x|| \leq 1\} \tag{1.3.12}$$

is finite; in other words, T maps bounded sets of V into bounded sets of W. (A subset B of a metric space X is said to be bounded if there exist $x_0 \in X$ and $R > 0$ such that $d(x, x_0) \leq R \; \forall \; x \in B$.)

Proof: Before proceeding to the proof, notice that

$$T \in L(V, W) \;\Rightarrow\; T0_V = 0_W \; . \tag{1.3.13}$$

In the sequel, we shall simply write 0 for the zero vector of any vector space.

$(i) \Rightarrow (ii)$:

$$\begin{aligned} x_n \to x \text{ in } V \;&\Rightarrow\; (x_n - x) \;\to\; 0 \\ &\Rightarrow\; T(x_n - x) \;\to\; T0 = 0 \\ &\Rightarrow\; (Tx_n - Tx) \;\to\; 0 \\ &\Rightarrow\; Tx_n \;\to\; Tx \text{ in } W \; . \end{aligned}$$

$(ii) \Rightarrow (iii)$: To say that $||T|| = \infty$ means that we can find a sequence $\{x_n\}$ in X such that $||x_n|| \leq 1$ and $||Tx_n|| \geq n$. It follows from equation 1.3.13 that if x_n are as above, then the sequence $\{\frac{1}{n}x_n\}$ would converge to 0 while the sequence $\{T(\frac{1}{n}x_n)\}$ would not converge to 0, contradicting the assumption (ii).

$(iii) \Rightarrow (i)$: Notice to start with, that if $||T|| < \infty$, and if $0 \neq x \in V$, then

$$||Tx|| \;=\; ||x|| \cdot ||T(\frac{1}{||x||}x)|| \;\leq\; ||x|| \cdot ||T|| \;,$$

and hence

$$\begin{aligned} x_n \to 0 \;&\Rightarrow\; ||x_n|| \to 0 \\ &\Rightarrow\; ||Tx_n|| \;\leq\; ||T|| \; ||x_n|| \to 0 \end{aligned}$$

and the proposition is proved.

□

Remark 1.3.3 The collection of continuous linear transformations from a normed space V into a normed space W will always be denoted, in these notes, by the symbol $\mathcal{L}(V, W)$. As usual, we shall use the abbreviation $\mathcal{L}(V)$ for $\mathcal{L}(V, V)$. In view of condition (iii) of Proposition 1.3.2, an element of $\mathcal{L}(V, W)$ is also described by the adjective 'bounded'. In the sequel, we shall use the expression **bounded operator** to mean a continuous (equivalently, bounded) linear transformation between normed spaces. □

We relegate some standard properties of the assignment $T \mapsto ||T||$ to the exercises.

Exercise 1.3.4 *Let V, W be non-zero normed vector spaces, and let $T \in \mathcal{L}(V, W)$.*

(1) Show that

$$\begin{aligned} ||T|| \;&=\; \sup\{||Tx|| : x \in V, ||x|| \leq 1\} \\ &=\; \sup\{||Tx|| : x \in V, ||x|| = 1\} \\ &=\; \sup\{\frac{||Tx||}{||x||} : x \in V, x \neq 0\} \\ &=\; \inf\{K > 0 : ||Tx|| \leq K||x|| \; \forall \; x \in V\} \;; \end{aligned}$$

(2) $\mathcal{L}(V, W)$ is a vector space, and the assignment $T \mapsto ||T||$ defines a norm on $\mathcal{L}(V, W)$;

(3) If $S \in \mathcal{L}(W, X)$, where X is another normed space, show that $ST \; (= S \circ T) \; \in \mathcal{L}(V, X)$ and $||ST|| \leq ||S|| \cdot ||T||$.

(4) If V is a finite-dimensional normed space, show that $L(V, W) = \mathcal{L}(V, W)$; what if it is only assumed that W (and not necessarily V) is finite-dimensional. (Hint : let V be the space $\mathbb{P}$ of all polynomials, viewed as a subspace of the normed space $C[0,1]$ (equipped with the 'sup'-norm as in Example 1.2.2 (2)), and consider the mapping $\phi : \mathbb{P} \to \mathbb{C}$ defined by $\phi(f) = f'(0)$.)

(5) Let $\lambda_i \in \mathbb{C}, 1 \leq i \leq n,\ 1 \leq p \leq \infty$; consider the operator $T \in \mathcal{L}(\ell_n^p)$ defined by

$$T(x_1, \cdots, x_n) = (\lambda_1 x_1, \cdots, \lambda_n x_n) ,$$

and show that $||T|| = max\{|\lambda_i| : 1 \leq i \leq n\}$. (You may assume that ℓ_n^p is a normed space.)

(6) Show that the equation $U(T) = T(1)$ defines a linear isomorphism $U : \mathcal{L}(\mathbb{C}, V) \to V$ which is **isometric** *in the sense that $||U(T)|| = ||T||$ for all $T \in \mathcal{L}(\mathbb{C}, V)$.*

1.4 The Hahn-Banach theorem

This section is devoted to one version of the celebrated **Hahn-Banach theorem**.

We begin with a lemma.

LEMMA 1.4.1 *Let M be a linear subspace of a real vector space X. Suppose $p : X \to \mathbb{R}$ is a mapping such that*
(i) $p(x + y) \leq p(x) + p(y)\ \forall\ x, y \in X$; and
(ii) $p(tx) = tp(x)\ \forall x \in X,\ 0 \leq t \in \mathbb{R}$.
Suppose $\phi_0 : M \to \mathbb{R}$ is a linear map such that $\phi_0(x) \leq p(x)$ for all $x \in M$. Then, there exists a linear map $\phi : X \to \mathbb{R}$ such that $\phi|_M = \phi_0$ and $\phi(x) \leq p(x)\ \forall\ x \in X$.

Proof : The proof is an instance of the use of Zorn's lemma. (See §A.2, if you have not yet got acquainted with Zorn's lemma.) Consider the partially ordered set $\mathcal{P}$, whose typical member is a pair (Y, ψ), where (i) Y is a linear subspace of X which contains X_0; and (ii) $\psi : Y \to \mathbb{R}$ is a linear map which is an extension of ϕ_0 and satisfies $\psi(x) \leq p(x)\ \forall\ x \in Y$; the partial order on $\mathcal{P}$ is defined by setting $(Y_1, \psi_1) \leq (Y_2, \psi_2)$ precisely when (a) $Y_1 \subset Y_2$, and (b) $\psi_2|_{Y_1} = \psi_1$. (Verify that this prescription indeed defines a partial order on $\mathcal{P}$.)

Furthermore, if $\mathcal{C} = \{(Y_i, \psi_i) : i \in I\}$ is any totally ordered set in $\mathcal{P}$, an easy verification shows that an upper bound for the family $\mathcal{C}$ is given by (Y, ψ), where $Y = \cup_{i \in I} Y_i$ and $\psi : Y \to \mathbb{R}$ is the unique (necessarily linear) map satisfying $\psi|_{Y_i} = \psi_i$ for all i.

Hence, by Zorn's lemma. the partially ordered set $\mathcal{P}$ has a maximal element, call it (Y, ψ). The proof of the lemma will be completed once we have shown that $Y = X$.

Suppose $Y \neq X$; fix $x_0 \in X - Y$, and let $Y_1 = Y + \mathbb{R}x_0 = \{y + tx_0 : y \in Y,\ t \in \mathbb{R}\}$. The definitions ensure that Y_1 is a subspace of X which properly contains Y. Also, notice that any linear map $\psi_1 : Y_1 \to \mathbb{R}$ which extends ψ is prescribed uniquely by the number $t_0 = \psi_1(x_0)$ (and the equation $\psi_1(y + tx_0) = \psi(y) + tt_0$).

We assert that it is possible to find a number $t_0 \in \mathbb{R}$ such that the associated map ψ_1 would - in addition to extending ψ - also satisfy $\psi_1 \leq p$. This would then establish the inequality $(Y, \psi) \leq (Y_1, \psi_1)$, contradicting the maximality of (Y, ψ); this contradiction would then imply that we must have had $Y = X$ in the first place, and the proof would be complete.

First observe that if $y_1, y_2 \in Y$ are arbitrary, then,

$$\begin{aligned}
\psi(y_1) + \psi(y_2) &= \psi(y_1 + y_2) \\
&\leq p(y_1 + y_2) \\
&\leq p(y_1 - x_0) + p(y_2 + x_0) \ ;
\end{aligned}$$

and consequently,

$$\sup_{y_1 \in Y} [\, \psi(y_1) - p(y_1 - x_0) \,] \ \leq \ \inf_{y_2 \in Y} [\, p(y_2 + x_0) \ - \ \psi(y_2) \,] \ . \quad (1.4.14)$$

Let t_0 be any real number which lies between the supremum and the infimum appearing in equation 1.4.14. We now verify that this t_0 does the job.

Indeed, if $t > 0$, and if $y \in Y$, then, since the definition of t_0 ensures that $\psi(y_2) + t_0 \leq p(y_2 + x_0)\ \forall\ y_2 \in Y$, we find that:

$$\begin{aligned}
\psi_1(y + tx_0) &= \psi(y) + tt_0 \\
&= t\ [\psi(\frac{y}{t}) + t_0] \\
&\leq t\ p(\frac{y}{t} + x_0) \\
&= p(y + tx_0) \ .
\end{aligned}$$

Similarly, if $t < 0$, then, since the definition of t_0 also ensures that $\psi(y_1) - t_0 \leq p(y_1 - x_0) \ \forall \ y_1 \in Y$, we find that:

$$\begin{aligned}
\psi_1(y + tx_0) &= \psi(y) + tt_0 \\
&= -t \ [\psi(\frac{y}{-t}) - t_0] \\
&\leq -t \ p(\frac{y}{-t} - x_0) \\
&= p(y + tx_0) \ .
\end{aligned}$$

Thus, $\psi_1(y + tx_0) \leq p(y + tx_0) \ \forall \ y \in Y, t \in \mathbb{R}$, and the proof of the lemma is complete.

$\square$

It is customary to use the notation V^* for the space $\mathcal{L}(V, \mathbb{C})$ of continuous (equivalently, bounded) linear functionals on the normed space V.

We are ready now for the Hahn-Banach theorem, which guarantees the existence of 'sufficiently many' continuous linear functionals on any normed space. (Before proceeding to this theorem, the reader should spend a little time thinking about why $V^* \neq \{0\}$ for a normed space $V \neq \{0\}$.)

THEOREM 1.4.2 *(***Hahn-Banach theorem***)*

Let V be a normed space and let V_0 be a subspace of V. Suppose $\phi_0 \in V_0^*$; then there exists a $\phi \in V^*$ such that

(i) $\phi|_{V_0} = \phi_0$; and

(ii) $||\phi|| = ||\phi_0||$.

Proof: We first consider the case when V is a 'real normed space' and V_0 is a (real) linear subspace of V. In this case, apply the preceding lemma with $X = V, M = V_0$ and $p(x) = ||\phi_0|| \cdot ||x||$, to find that the desired conclusion follows immediately.

Next consider the case of complex scalars. Define $\psi_0(x) = Re \ \phi_0(x)$ and $\chi_0(x) = Im \ \phi_0(x)$, $\forall \ x \in V_0$, and note that if $x \in X_0$, then

$$\begin{aligned}
\psi_0(ix) + i\chi_0(ix) &= \phi_0(ix) \\
&= i\phi_0(x) \\
&= -\chi_0(x) + i\psi_0(x) \ ,
\end{aligned}$$

and hence, $\chi_0(x) = -\psi_0(ix) \;\forall\; x \in X_0$.

Observe now that $\psi_0 : X_0 \to \mathbb{R}$ is a 'real' linear functional of the real normed linear space X_0 , which satisfies :

$$|\psi_0(x)| \;\leq\; |\phi_0(x)| \;\leq\; ||\phi_0|| \cdot ||x|| \;\forall\; x \in V_0 \;;$$

deduce from the already established real case that ψ_0 extends to a real linear functional - call it ψ - of the real normed space V such that $|\psi(x)| \;\leq\; ||\phi_0|| \cdot ||x|| \;\forall\; x \in V$. Now define $\phi : V \to \mathbb{C}$ by $\phi(x) \;=\; \psi(x) - i\psi(ix)$, and note that ϕ indeed extends ϕ_0. Finally, if $x \in V$ and if $\phi(x) = re^{i\theta}$, with $r > 0, \theta \in \mathbb{R}$, then,

$$\begin{aligned} |\phi(x)| &= e^{-i\theta}\phi(x) \\ &= \phi(e^{-i\theta}x) \\ &= \psi(e^{-i\theta}x) \\ &\leq ||\phi_0|| \cdot ||e^{-i\theta}x|| \\ &= ||\phi_0|| \cdot ||x|| \;, \end{aligned}$$

and the proof of the theorem is complete. □

We shall postpone the deduction of some easy corollaries of this theorem until we have introduced some more terminology in the next section. Also, we postpone the discussion of further variations of the Hahn-Banach theorem - which are best described as Hahn-Banach separation results, and are stated most easily using the language of 'topological vector spaces' - to the last section of this chapter, which is where topological vector spaces are discussed.

1.5 Completeness

A fundamental notion in the study of metric spaces is that of *completeness*. It would be fair to say that the best - and most useful - metric spaces are the ones that are complete.

To see what this notion is, we start with an elementary observation. Suppose (X, d) is a metric space, and $\{x_n : n \in \mathbb{N}\}$ is a sequence in X; recall that to say that this sequence converges is to say that there exists a point $x \in X$ such that $x_n \to x$ - or equivalently, in the language of epsilons, this means that for each $\epsilon > 0$, it is possible to find a natural number N such that $d(x_n, x) < \epsilon$ whenever $n \geq N$. This is easily seen to imply the following condition:

Cauchy criterion: *For each $\epsilon > 0$, there exists a natural number N such that $d(x_m, x_n) < \epsilon$ whenever $n, m \geq N$.*

A sequence which satisfies the preceding condition is known as a **Cauchy sequence**. Thus, the content of our preceding discussion is that any convergent sequence is a Cauchy sequence.

DEFINITION 1.5.1 *(a) A metric space is said to be* **complete** *if every Cauchy sequence in it is convergent.*

(b) A normed space which is complete with respect to the metric defined by the norm is called a **Banach space**.

The advantage with the notion of a Cauchy sequence is that it is a condition on only the members of the sequence; on the other hand, to check that a sequence is convergent, it is necessary to have the prescience to find a possible x to which the sequence wants to converge. The following exercise will illustrate the use of this notion.

EXERCISE 1.5.2 *(1) Let $X = C[0,1]$ be endowed with the sup-norm $||\cdot||_\infty$, as in Example 1.2.3 (2).*

(a) Verify that a sequence $\{f_n\}$ converges to $f \in X$ precisely when the sequence $\{f_n\}$ of functions converges **uniformly** *to the function f.*

(b) Show that $\mathbb{C}$ is a complete metric space with respect to $d(z, w) = |z - w|$, and use the fact that the limit of a uniformly convergent sequence of continuous functions is continuous, to prove that X is complete (in the metric coming from the sup-norm), and cosequently, a Banach space.

(c) Show, more generally, that the space $C_0(Z)$, where Z is a locally compact Hausdorff space, - see Example 1.2.3 (2) - is a Banach space.

(2) Suppose $(X, ||\cdot||)$ is a Banach space; a series $\sum_{n=1}^{\infty} x_n$ is said to converge in X if the sequence $\{s_n\}$ of partial sums defined by $s_n = \sum_{k=1}^{n} x_k$, is a convergent sequence.

Show that an 'absolutely summable' series is convergent, provided the ambient normed space is complete - i.e., show that if $\{x_n\}$ *is a sequence in* X *such that the series* $\sum_{n=1}^{\infty} ||x_n||$ *of non-negative numbers is convergent (in* $\mathbb{R}$*), then the series* $\sum_{n=1}^{\infty} x_n$ *converges in* X.

Show, conversely, that if every absolutely summable series in a normed vector space X *is convergent, then* X *is necessarily complete.*

(3) Show that the series $\sum_{n=1}^{\infty} \frac{1}{n^2} sin\ nx$ *converges to a continuous function in [0,1]. (Try to think of how you might prove this assertion using the definition of a convergent sequence, without using the notion of completeness and the fact that* $C[0,1]$ *is a Banach space.)*

(4) Adapt the proof of (1)(b) to show that if W *is a Banach space, and if* V *is any normed space, then also* $\mathcal{L}(V, W)$ *is a Banach space.*

(5) Show that the normed space ℓ^p *is a Banach space, for* $1 \leq p \leq \infty$.

As a special case of Exercise 1.5.2(4) above, we see that even if V is not necessarily complete, the space V^* is always a Banach space. It is customary to refer to V^* as the **dual space** of the normed space V. The justification, for the use of the term 'dual' in the last sentence, lies in the Hahn-Banach theorem. To see this, we first list various consequences of this very important theorem as exercises.

EXERCISE 1.5.3 *Let* X *be a normed space, with dual space* X^*.

(1) If $0 \neq x_0 \in X$, *show that there exists a* $\phi \in X^*$ *such that* $||\phi|| = 1$ *and* $\phi(x_0) = ||x_0||$. *(Hint: set* $X_0 = \mathbb{C}x_0 = \{\alpha x_0 : \alpha \in \mathbb{C}\}$, *consider the linear functional* $\phi_0 \in X_0^*$ *defined by* $\phi_0(\lambda x_0) = \lambda ||x_0||$, *and appeal to the Hahn-Banach theorem.)*

(2) Show that the equation

$$(\ j(x)\ )(\phi) \ = \ \phi(x) \tag{1.5.15}$$

defines an isometric mapping $j : X \to (X^*)^*$. *(It is customary to write* X^{**} *for* $(X^*)^*$.

(3) If X_0 is a closed subspace of X - i.e., X_0 is a vector subspace which is closed as a subset of the metric space X - let X/X_0 denote the quotient vector space; (thus, X/X_0 is the set of equivalence classes with respect to the equivalence relation $x \sim_{X_0} y \Leftrightarrow (x-y) \in X_0$; it is easy to see that X/X_0 is a vector space with respect to the natural vector operations; note that a typical element of X/X_0 may - and will - be denoted by $x + X_0,\ x \in X$).

Define

$$||x+X_0|| = \inf\{||x-x_0|| : x_0 \in X_0\} = dist(x, X_0) \ . \quad (1.5.16)$$

Show that

(i) X/X_0 is a normed vector space with respect to the above definition of the norm, and that this **quotient space** *is complete if X is;*

(ii) if $x \in X$, then $x \notin X_0$ if and only if there exists a non-zero linear functional $\phi \in X^*$ such that $\phi(x) \neq 0$ and $\phi(y) = 0 \ \forall \ y \in X_0$. (Hint: Apply Exercise (1) above to the space X/X_0.)

Remark 1.5.4 A Banach space X is said to be **reflexive** if the mapping j of Exercise 1.5.3(2) is surjective - i.e., $j(X) = X^{**}$. (Since any dual space is complete, a reflexive space is necessarily complete, and this is why we have - without loss of generality - restricted ourselves to Banach spaces in defining this notion.

It is a fact that ℓ^p - see Example 1.2.2(4) - is reflexive if and only if $1 < p < \infty$. $\square$

In the next sequence of exercises, we outline one procedure for showing how to 'complete' a normed space.

Exercise 1.5.5 *(1) Suppose Y is a normed space and Y_0 is a (not necessarily closed) subspace in Y. Show that:*

(a) if Y_1 denotes the **closure** *of Y_0 - i.e., if $Y_1 = \{y \in Y : \exists \ \{y_n\}_{n=1}^\infty \subset Y_0$ such that $y_n \to y\}$ - then Y_1 is a closed subspace of Y; and*

(b) if Z is a Banach space, and if $T \in \mathcal{L}(Y_0, Z)$, then there exists a unique operator $\tilde{T} \in \mathcal{L}(Y_1, Z)$ such that $\tilde{T}|_{Y_0} = T$; further, if T is an isometry, show that so also is $\tilde{T}$.

(2) Suppose X is a normed vector space. Show that there exists a Banach space $\overline{X}$ which admits an isometric embedding of X onto a dense subspace X_0; i.e., there exists an isometry $T \in \mathcal{L}(X, \overline{X})$ such that $X_0 = T(X)$ is dense in $\overline{X}$ (meaning that the closure of X_0 is all of $\overline{X}$). (Hint: Consider the map j of Exercise 1.5.3(2), and choose $\overline{X}$ to be the closure of $j(X)$.)

(3) Use Exercise (1) above to conclude that if Z is another Banach space such that there exists an isometry $S \in \mathcal{L}(X, Z)$ whose range is dense in Z, and if $\overline{X}$ and T are as in Exercise (2) above, then show that there exists a unique isometry $U \in \mathcal{L}(\overline{X}, Z)$ such that $U \circ T = S$ and U is onto. Hence the space $\overline{X}$ is essentially uniquely determined by the requirement that it is complete and contains an isometric copy of X as a dense subspace; such a Banach space is called a **completion** *of the normed space X.*

We devote the rest of this section to a discussion of some important consequences of completeness in the theory of normed spaces. We begin with two results pertaining to arbitrary metric (and not necessarily vector) spaces, namely the celebrated 'Cantor intersection theorem' and the 'Baire category theorem'.

THEOREM 1.5.6 *(***Cantor intersection theorem***)*
Suppose $C_1 \supset C_2 \supset \cdots \supset C_n \supset \cdots$ is a non-increasing sequence of non-empty closed sets in a complete metric space X. Assume further that

$$diam(C_n) = \sup\{d(x,y) : x, y \in C_n\} \rightarrow 0 \text{ as } n \rightarrow \infty .$$

Then $\cap_{n=1}^{\infty} C_n$ is a singleton set.

Proof: Pick $x_n \in C_n$ - this is possible since each C_n is assumed to be non-empty. The hypothesis on the diameters of the C_n's shows that $\{x_n\}$ is a Cauchy sequence in X, and hence the sequence converges to a limit, say x.

The assumption that the C_n's are nested and closed are easily seen to imply that $x \in \cap_n C_n$. If also $y \in \cap_n C_n$, then, for each n, we have $d(x,y) \leq diam(C_n)$. This clearly forces $x = y$, thereby completing the proof.

□

EXERCISE 1.5.7 *Consider the four hypotheses on the sequence $\{C_n\}$ in the preceding theorem - viz., non-emptiness, decreasing property, closedness, shrinking of diameters to 0 - and show, by example, that the theorem is false if any one of these hypotheses is dropped.*

Before we ge to the next theorem, we need to introduce some notions. Call a set (in a metric space) **nowhere dense** if its closure has empty interior; thus a set A is nowhere dense precisely when every non-empty open set contains a non-empty open subset which is disjoint from A. (Verify this last statement!) A set is said to be of **first category** if it is expressible as a countable union of nowhere dense sets.

THEOREM 1.5.8 *(**Baire Category Theorem**) No non-empty open set in a complete metric space is of the first category; or, equivalently, a countable intersection of dense open sets in a complete metric space is also dense.*

Proof: The equivalence of the two assertions is an easy exercise in taking complements and invoking de Morgan's laws, etc. We content ourselves with proving the second of the two formulations of the theorem. For this, we shall repeatedly use the following assertion, whose easy verification is left as an exercise to the reader:

If B is a closed ball of positive diameter - say δ - and if U is a dense open set in X, then there exists a closed ball B_0 which is contained in $B \cap U$ and has a positive diameter which is less than $\frac{1}{2}\delta$.

Suppose $\{U_n\}_{n=1}^{\infty}$ is a sequence of dense open sets in the complete metric space X. We need to verify that if U is any non-empty open set, then $U \cap (\cap_{n=1}^{\infty} U_n)$ is not empty. Given such a U, we may clearly find a closed ball B_0 of positive diameter - say δ - such that $B_0 \subset U$.

Repeated application of the preceding italicised assertion and an easy induction argument yield a sequence $\{B_n\}_{n=1}^{\infty}$ of closed balls of positive diameter such that:

(i) $B_n \subset B_{n-1} \cap U_n \ \forall \ n \geq 1$;

(ii) $diam \ B_n < (\frac{1}{2})^n \delta \ \forall \ n \geq 1$.

An appeal to Cantor's intersection theorem completes the proof.

□

REMARK 1.5.9 In the foregoing proof, we used the expression 'diameter of a ball' to mean the diameter of the set as defined, for instance, in the statement of the Cantor intersection theorem. In particular, the reader should note that if $B = \{y \in X : d(x,y) < \delta\}$ is the typical open ball in a general metric space, then the diamter of B might not be equal to 2δ in general. (For instance, consider $X = \mathbb{Z}, x = 0, \delta = 1$, in which case the 'ball' B in question is actually a singleton and hence has diameter 0.) It is in this sense that the word diameter is used in the preceding proof. □

We give a flavour of the kind of consequence that the Baire category theorem has, in the following exercise.

EXERCISE 1.5.10 *(i) Show that the plane cannot be covered by a countable number of straight lines;*

(ii) more generally, show that a Banach space cannot be expressed as a countable union of 'translates of proper closed subspaces';

(iii) where do you need the hypothesis that the subspaces in (ii) are closed?

As in the preceding exercise, the Baire category theorem is a powerful 'existence result'. (There exists a point outside the countably many lines, etc.) The interested reader can find applications in [Rud], for instance, of the Baire Category Theorem to prove the following existence results:

(i) there exists a continuous function on [0,1] which is nowhere differentiable;

(ii) there exists a continuous function on the circle whose Fourier series does not converge anywhere.

The reader who does not understand some of these statements may safely ignore them.

We now proceed to use the Baire category theorem to prove some of the most useful results in the theory of linear operators on Banach spaces.

THEOREM 1.5.11 *(***Open mapping theorem***)*

Suppose $T \in \mathcal{L}(X, Y)$ where X and Y are Banach spaces, and suppose T maps X onto Y. Then T is an open mapping - i.e., T maps open sets to open sets.

Proof: For $Z \in \{X, Y\}$, let us write $B_r^Z = \{z \in Z : ||z|| < r\}$. Since a typical open set in a normed space is a union of translates of dilations of the unit ball, it is clearly sufficient to prove that $T(B_1^X)$ is open in Y. In fact it is sufficient to show that there exists an $r > 0$ such that $B_r^Y \subset T(B_1^X)$. (*Reason:* suppose this latter assertion is true; it is then clearly the case that $B_{r\epsilon}^Y \subset T(B_\epsilon^X)$ $\forall \epsilon > 0$. Fix $y \in T(B_1^X)$, say $y = Tx, ||x|| < 1$; choose $0 < \epsilon < 1 - ||x||$ and observe that $y + B_{r\epsilon}^Y \subset y + T(B_\epsilon^X) = T(x + B_\epsilon^X) \subset T(B_1^X)$, so we have verified that every point of $T(B_1^X)$ is an interior point.)

The assumption of surjectivity shows that $Y = \cup_n T(B_n^X)$; deduce from the Baire category theorem that (since Y is assumed to be complete) there exists some n such that $T(B_n^X)$ is not nowhere dense; this means that the closure $\overline{T(B_n^X)}$ of $T(B_n^X)$ has non-empty interior; since $\overline{T(B_n^X)} = n\overline{T(B_1^X)}$, this implies that $\overline{T(B_1^X)}$ has non-empty interior. Suppose $y + B_s^Y \subset \overline{T(B_1^X)}$; this is easily seen to imply that $B_s^Y \subset \overline{T(B_2^X)}$. Hence, setting $t = \frac{1}{2}s$, we thus see that $B_t^Y \subset \overline{T(B_1^X)}$, and hence also $B_{\epsilon t}^Y \subset \overline{T(B_\epsilon^X)}$ $\forall \epsilon > 0$.

Thus, we have produced a positive number t with the following property:

$$\delta, \epsilon > 0,\ y \in Y,\ ||y|| < \epsilon t \ \Rightarrow\ \exists x \in X \text{ such that } ||x|| < \epsilon \text{ and } ||y - Tx|| < \delta\ . \qquad (1.5.17)$$

Suppose now that $y_0 \in B_t^Y$. Apply 1.5.17 with $\epsilon = 1,\ \delta = \frac{1}{2}t,\ y = y_0$ to find $x_0 \in X$ such that $||x_0|| < 1$ and $||y_0 - Tx_0|| < \frac{1}{2}t$.

Next apply 1.5.17 with $\epsilon = \frac{1}{2},\ \delta = (\frac{1}{2})^2 t,\ y = y_1 = y_0 - Tx_0$, to find $x_1 \in X$ such that $||x_1|| < \frac{1}{2}$ and $||y_1 - Tx_1|| < (\frac{1}{2})^2 t$.

By repeating this process inductively, we find vectors $y_n \in Y,\ x_n \in X$ such that $||y_n|| < (\frac{1}{2})^n t,\ ||x_n|| < (\frac{1}{2})^n,\ y_{n+1} = y_n - Tx_n$ $\forall n \geq 0$.

Notice that the inequalities ensure that the series $\sum_{n=0}^\infty x_n$ is an 'absolutely summable' series in the Banach space X and consequently - by Exercise 1.5.2(2) - the series converges to a limit, say x, in X. On the other hand, the definition of the y_n's implies that

$$\begin{aligned} y_{n+1} &= y_n - Tx_n \\ &= y_{n-1} - Tx_{n-1} - Tx_n \\ &= \cdots \\ &= y_0 - T(\sum_{k=0}^{n} x_k) \ . \end{aligned}$$

Since $||y_{n+1}|| \to 0$, we may conclude from the continuity of T that $y_0 = Tx$. Finally, since clearly $||x|| < 2$, we have now shown that $B_t^Y \subset T(B_2^X)$, and hence $B_{\frac{1}{2}t}^Y \subset T(B_1^X)$ and the proof of the theorem is complete. □

Corollary 1.5.12 *Suppose X, Y are Banach spaces and suppose $T \in \mathcal{L}(X, Y)$. Assume that T is 1-1.*

(1) Then the following conditions on T are equivalent:

(i) ran $T = T(X)$ is a closed subspace of Y;

(ii) the operator T is **bounded below**, *meaning that there exists a constant $c > 0$ such that $||Tx|| \geq c||x|| \ \forall \ x \in X$.*

(2) In particular, the following conditions are equivalent:

(i) T is 1-1 and maps X onto Y;

(ii) there exists a bounded operator $S \in \mathcal{L}(Y, X)$ such that $ST = id_X$ and $TS = id_Y$; in other words, T is invertible in the strong sense that T has an inverse which is a bounded operator.

Proof: (1) Suppose T is bounded below and suppose $Tx_n \to y$; then $\{Tx_n\}_n$ is a Cauchy sequence; but the assumption that T is bounded below then implies that also $\{x_n\}_n$ is a Cauchy sequence which must converge to some point, whose image under T must be y in view of continuity of T.

Conversely suppose $Y_0 = T(X)$ is closed; then also Y_0 is a Banach space in its own right. An application of the open mapping theorem to $T \in \mathcal{L}(X, Y_0)$ now implies that there exists a constant $\epsilon > 0$ such that $B_\epsilon^{Y_0} \subset T(B_1^X)$ (in the notation of the proof of Theorem 1.5.11) . We now assert that $||Tx|| \geq \frac{\epsilon}{2}||x|| \ \forall \ x \in X$. Indeed, suppose $x \in X$ and $||Tx|| = r$; then $T(\frac{\epsilon}{2r}x) \in B_\epsilon^{Y_0} \subset T(B_1^X)$. Since T is 1-1, this clearly implies that $\frac{\epsilon}{2r}x \in B_1^X$, whence $||x|| < \frac{2r}{\epsilon}$; this proves our assertion.

(2) The implication $(ii) \Rightarrow (i)$ is obvious. Conversely, if T is a bijection, it admits an inverse map which is easily verified to be linear; the boundedness of the inverse is a consequence of the fact that T is bounded below (in view of (1) above). □

Before proceeding to obtain another important consequence of the open mapping theorem, we need to set up some terminology, which we do in the guise of an exercise.

EXERCISE 1.5.13 *The* **graph** *of a function $f : X \to Y$ is, by definition, the subset $G(f) \subset X \times Y$ defined by*

$$G(f) \;=\; \{(x, f(x)) : x \in X\} \;.$$

Recall that if X and Y are metric spaces, then the product $X \times Y$ can be metrised in a variety of ways - see Exercise 1.2.3(1) - in such a way that a sequence in $X \times Y$ converges precisely when each of the co-ordinate sequences converges.

(i) Show that if f is a continuous map of metric spaces, then its graph is closed (as a subset of the product) - i.e., if $\{x_n\}$ is a sequence in X and if there exists $(x, y) \in X \times Y$ such that $x_n \to x$ and $f(x_n) \to y$, then necessarily $y = f(x)$.

(ii) What about the converse? i.e., if a function between metric spaces has a closed graph, is the function necessarily continuous?

The content of the next result is that for linear mappings between Banach spaces, the requirements of continuity and having a closed graph are equivalent. In view of Exercise 1.5.13(i), we only state the non-trivial implication in the theorem below.

THEOREM 1.5.14 *(***Closed Graph Theorem***)*

Suppose X and Y are Banach spaces, and suppose $T : X \to Y$ is a linear mapping which has a closed graph. Then T is continuous, i.e., $T \in \mathcal{L}(X, Y)$.

Proof: The graph $G(T)$ is given to be a closed subspace of the Banach space $X \oplus_{\ell^1} Y$ (by which, of course, we mean the space $X \times Y$ with norm given by $||(x, y)|| = ||x|| + ||y||$). Hence, we may regard $G(T)$ as a Banach space in its own right.

Consider the mapping $S : G(T) \to X$ defined by the equation $S(x, f(x)) = x$. This is clearly a linear 1-1 map of $G(T)$ onto X. Also, it follows from the definition of our choice of the norm on $G(T)$ that S is a bounded operator with $||S|| \leq 1$. It now follows from Corollary 1.5.12(1) that the operator S is bounded below; i.e., there exists a constant $c > 0$ such that $||x|| \geq c(||x|| + ||Tx||)$. Hence $||Tx|| \leq \frac{1-c}{c}||x||$, and the proof is complete.

□

REMARK 1.5.15 Suppose $T \in \mathcal{L}(X, Y)$, where X and Y are Banach spaces, and suppose that T is only assumed to be 1-1. (An example of such an operator, with $X = Y = \ell^2$ is given by $T(x_1, x_2, \cdots, x_n, \cdots) = (x_1, \frac{1}{2}x_2, \cdots, \frac{1}{n}x_n, \cdots)$.) Let Y_0 be the (not necessarily closed, and hence Banach) subspace of Y defined by $Y_0 = T(X)$. Let $T^{-1} : Y_0 \to X$ be the (necessarily linear) inverse of T. Then T^{-1} has a closed graph, but is not continuous unless T happened to be bounded below (which is not the case in the example cited above). Hence the hypothesis of completeness of the domain cannot be relaxed in the closed graph theorem. (Note that T^{-1} is bounded precisely when T is bounded below, which, under the assumed completeness of X, happens precisely when Y_0 is complete - see Corollary 1.5.12.)

This example also shows that the 'onto' requirement in the Open Mapping Theorem is crucial, and cannot be dropped. □

We conclude with another very important and useful result on operators between Banach spaces.

THEOREM 1.5.16 *(***Uniform Boundedness Principle***)*

Suppose X and Y are Banach saces. The following conditions on an arbitrary family $\{T_i : i \in I\} \subset \mathcal{L}(X, Y)$ of operators are equivalent:

(i) $\{T_i : i \in I\}$ is uniformly bounded - i.e., $\sup_{i \in I} ||T_i|| < \infty$;

(ii) $\{T_i : i \in I\}$ is 'pointwise' bounded - i.e., for each $x \in X$, the family $\{T_i x : i \in I\}$ is bounded in Y, $\sup_{i \in I} ||T_i x|| < \infty$.

Proof: We only prove the non-trivial implication; so suppose (ii) holds. Let $A_n = \{x \in X : ||T_i x|| \leq n \ \forall \ i \in I\}$. The hypothesis implies that $X = \cup_{n=1}^{\infty} A_n$. Notice also that each A_n is clearly a closed set. Conclude from the Baire category theorem that there exists a positive constant $r > 0$, an integer n, and a vector $x_0 \in X$ such that $(x_0 + B_r^X) \subset A_n$. It follows that $B_r^X \subset A_{2n}$. Hence,

$$\begin{aligned} ||x|| \leq 1 \quad &\Rightarrow \quad \frac{r}{2} x \in B_r^X \\ &\Rightarrow \quad ||T_i(\frac{r}{2} x)|| \leq 2n \ \forall \ i \in I \\ &\Rightarrow \quad \sup_{i \in I} ||T_i x|| \leq \frac{4n}{r} \ . \end{aligned}$$

This proves the theorem. □

We conclude this section with some exercises on the uniform boundedness principle. (This 'principle' is also sometimes referred to as the **Banach-Steinhaus theorem**.)

EXERCISE 1.5.17 *(1) Suppose S is a subset of a (not necessarily complete) normed space X which is* **weakly bounded**, *meaning that $\phi(S)$ is a bounded set in $\mathbb{C}$, for each $\phi \in X^*$. Then show that S is norm bounded, meaning that $\sup\{||x|| : x \in S\} < \infty$. (Hint: Apply the uniform boundedness principle to the family $\{j(x) : x \in S\} \subset X^{**}$, in the notation of Exercise 1.5.3(2).)*

(2) Show that the completeness of Y is not needed for the validity of the uniform boundedness principle. (Hint: reduce to the complete case by considering the completion $\overline{Y}$ of Y.)

(3) Suppose $\{T_n : n = 1, 2, \cdots\} \subset \mathcal{L}(X, Y)$ is a sequence of bounded operators, where X, Y are Banach spaces, and suppose

the sequence converges 'strongly'; i.e., assume that $\{T_n x\}$ is a convergent sequence in Y, for each $x \in X$. Show that the equation $Tx = lim_n T_n x$ defines a bounded operator $T \in \mathcal{L}(X,Y)$.

1.6 Some topological considerations

This section is devoted to some elementary facts concerning 'topological vector spaces', which are more general than normed spaces - in much the same way as a topological space is more general than a metric space. The following text assumes that the reader knows at least the extent of 'general topology' that is treated in the appendix; the reader who is unfamiliar with the contents of §§A.3 and A.4, should make sure she understands that material at least by the time she finishes the contents of this section.

DEFINITION 1.6.1 *A* **topological vector space** - *henceforth abbreviated to t.v.s. - is a vector space X which is at the same time a topological space in such a way that the 'vector operations are continuous'; explicitly, we demand that the following maps are continuous:*

(i) $\mathbb{C} \times X \ni (\alpha, x) \mapsto \alpha x \in \mathbb{C}$;

(ii) $X \times X \ni (x, y) \mapsto x + y \in \mathrm{X}$.

If X, Y are topological vector spaces, the set of all continuous linear transformations $T : X \to Y$ is denoted by $\mathcal{L}(X,Y)$, and we write $\mathcal{L}(X) = \mathcal{L}(X,X)$.

We leave some simple verifications as an exercise for the reader.

EXERCISE 1.6.2 *(1) Suppose X is a t.v.s. Show that:*

(a) the map $X \ni x \mapsto T_x \in \mathcal{L}(X)$, defined by $T_x(y) = x + y$, is a homomorphism of the additive group of the vector space X into the group $Homeo(X)$ of homeomorphisms of the t.v.s. X; (recall that a homeomorphism of a topological space is a continuous bijection whose inverse is also continuous;)

(b) the map $\mathbb{C}^\times \ni \alpha \mapsto D_\alpha \in \mathcal{L}(X)$, defined by $D_\alpha x = \alpha x$, is a homomorphism from the multiplicative group of non-zero complex numbers into the multiplicative group $\mathcal{G}(\mathcal{L}(X))$ of invertible elements of the algebra $\mathcal{L}(X)$;

(c) with the preceding notation, all the translation and dilation operators T_x, D_α belong to $Homeo(X)$; in particular, the group of homeomorphisms of X acts transitively on X - i.e., if x, y are arbitrary points in X, then there exists an element L in this group such that $L(x) = y$.

(2) Show that $\mathcal{N}_{sym} = \{U : U$ is an open neighbourhood of 0 which is **symmetric** *in the sense that $U = -U = \{-x : x \in U\}\}$ is a* **neighbourhood base** *for 0, meaning that if V is any open neighbourhood of 0, then there exists a $U \in \mathcal{N}_{sym}$ such that $U \subset V$.*

The most important examples of topological vector spaces are the the so-called **locally convex** ones; these are the spaces where the topology is given by a family of *seminorms*.

DEFINITION 1.6.3 *A* **seminorm** *on a vector space X is a function $p : X \to [0, \infty)$ which satisfies all the properties of a norm except for 'positive-definiteness'; specifically, the conditions to be satisfied by a function p as above, in order to be called a seminorm are: for all $\alpha \in \mathbb{C}$, $x, y \in X$, we demand that*

$$p(x) \geq 0 \tag{1.6.18}$$
$$p(\alpha x) = |\alpha| p(x) \tag{1.6.19}$$
$$p(x+y) \leq p(x) + p(y) \tag{1.6.20}$$

EXAMPLE 1.6.4 If X is a vector space, and if $T : X \to Y$ is a linear map of X into a normed space Y, then the equation

$$p_T(x) = ||Tx||$$

defines a seminorm on X.

In particular, if $\phi : X \to \mathbb{C}$ is a linear functional, then the equation

$$p_\phi(x) = |\phi(x)| \tag{1.6.21}$$

defines a seminorm on X. □

REMARK 1.6.5 Every normed space is clearly a t.v.s. in an obvious manner. More generally, suppose $\mathcal{P} = \{p_i : i \in I\}$ is a family of seminorms on a vector space; for each $x \in X$ and $i \in I$, consider the map $f_{i,x} : X \to [0, \infty)$ defined by $f_{i,x}(y) = p_i(y - x)$. Let $\tau_{\mathcal{P}}$ denote the weak topology on X which is induced by the family $\{f_{i,x} : i \in I, x \in X\}$ of maps - see Proposition A.3.10 for the definition of weak topology. We may then conclude the following facts concerning this topology on X (from Exercise A.3.11):

(a) For each fixed $x \in X, i \in I, \epsilon > 0$, define $U_{(i,x,\epsilon)} = \{y \in X : p_i(y - x) < \epsilon\}$; then the family $\mathcal{S}_{\mathcal{P}} = \{U_{(i,x,\epsilon)} : i \in I, x \in X, \epsilon > 0\}$ defines a sub-base for the topology $\tau_{\mathcal{P}}$, and consequently the family $\mathcal{B}_{\mathcal{P}} = \{\cap_{j=1}^n U_{(i_j,x_j,\epsilon_j)} : \{i_j\}_{j=1}^n \subset I, \{x_j\}_{j=1}^n \subset X, \{\epsilon_j\}_{j=1}^n \subset (0, \infty),\ n \in \mathbb{N}\}$ is a base for the topology $\tau_{\mathcal{P}}$. (The reader should verify that the family of finite intersections of the above form where (in any particular finite intersection) all the ϵ_j's are equal, also constitutes a base for the topology $\tau_{\mathcal{P}}$.)

(b) A net - see Definition 2.2.3 - $\{x_\lambda : \lambda \in \Lambda\}$ converges to x in the topological space $(X, \tau_{\mathcal{P}})$ if and only if the net $\{p_i(x_\lambda - x) : \lambda \in \Lambda\}$ converges to 0 in $\mathbb{R}$, for each $i \in I$. (Verify this; if you do it properly, you will use the triangle inequality for seminorms, and you will see the need for the functions $f_{i,x}$ constructed above.)

It is not hard to verify - using the above criterion for convergence of nets in X - that X, equipped with the topology $\tau_{\mathcal{P}}$ is a t.v.s; further, $\tau_{\mathcal{P}}$ is the smallest topology with respect to which $(X, \tau_{\mathcal{P}})$ is a topological vector space with the property that $p_i : X \to \mathbb{R}$ is continuous for each $i \in I$. This is called the t.v.s. structure on X which is induced by the family $\mathcal{P}$ of seminorms on X.

However, this topology can be somewhat trivial; thus, in the extreme case where $\mathcal{P}$ is empty, the resulting topology on X is the indiscrete topology - see Example A.3.2(1); a vector space with the indiscrete topology certainly satisfies all the requirements we have so far imposed on a t.v.s., but it is quite uninteresting.

To get interesting t.v.s., we normally impose the further condition that the underlying topological space is a Hausdorff space.

It is not hard to verify that the topology $\tau_{\mathcal{P}}$ induced by a family of seminorms satisfies the Hausdorff separation requirement precisely when the family $\mathcal{P}$ of seminorms **separates points** in X, meaning that if $x \neq 0$, then there exists a seminorm $p_i \in \mathcal{P}$ such that $p_i(x) \neq 0$. (Verify this!)

We will henceforth encounter only t.v.s where the underlying topology is induced by a family of seminorms which separates points in the above sense. $\square$

DEFINITION 1.6.6 *Let X be a normed vector space.*

(a) The **weak topology** *on X is the topology $\tau_{\mathcal{P}}$, where $\mathcal{P} = \{p_\phi : \phi \in X^*\}$, where the p_ϕ's are defined as in Example 1.6.4, and, of course, the symbol X^* denotes the Banach dual space of X.*

(b) The **weak* topology** *on the dual space X^* is the topology $\tau_{\mathcal{P}}$, where $\mathcal{P} = \{p_{j(x)} : x \in X\}$, where the p_ϕ's are as before, and $j : X \to X^{**}$ is the canonical inclusion of X into its second dual.*

(c) In order to distinguish between the different notions, the natural norm topology (which is the same as $\tau_{\{\|\cdot\|\}}$) on X is referred to as the **strong** *topology on X.*

It should be clear that any normed space (resp., any Banach dual space) is a t.v.s. with respect to the weak (resp., weak*) topology. Further, it follows from the definitions that if X is any normed space, then:

(a) any set in X which is weakly open (resp., closed) is also strongly open (resp., closed); and

(b) any set in X^* which is weak* open (resp., closed) is also weakly open (resp., closed).

REMARK 1.6.7 (1) Some care must be exercised with the adjective 'strong'; there is a conflict between the definition given here, and the sense in which the expression 'strong convergence' is used for operators, as in §2.5, for instance; but both forms of usage have become so wide-spread that this is a (minor) conflict that one has to get used to.

(2) The above definitions also make sense for topological vector spaces; you only have to note that X^* is the space of continuous linear functionals on the given t.v.s. X. In this case,

the weak* topology on X^* should be thought of as the topology induced by the family of 'evaluation functionals' on X^* coming from X. $\Box$

Exercise 1.6.8 *(1) Let X be a normed space. Show that the weak topology on X is a Hausdorff topology, as is the weak * topology on X^*.*

(2) Verify all the facts that have been merely stated, without a proof, in this section.

Our next goal is to establish a fact which we will need in the sequel.

Theorem 1.6.9 *(**Alaoglu's theorem**)*

Let X be a normed space. Then the unit ball of X^* - i.e.,

$$ball\ X^* \ = \ \{\phi \in X^* : ||\phi|| \leq 1\}$$

*is a compact Hausdorff space with respect to the weak * topology.*

Proof : Notice the following facts to start with:

(1) if $\phi \in ball\ X^*$ and if $x \in ball\ X$, then $|\phi(x)| \leq 1$;

(2) if $f : ball\ X \to \mathbb{C}$ is a function, then there exists a $\phi \in ball\ X^*$ such that $f = \phi|_{ball\ X}$ if and only if the function f satisfies the following conditions:

(i) $|f(x)| \leq 1\ \forall x \in ball\ X$;

(ii) whenever $x, y \in ball\ X$ and $\alpha, \beta \in \mathbb{C}$ are such that $\alpha x + \beta y \in ball\ X$, then $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$.

Let $Z = \overline{\mathbb{D}}^{ball\ X}$ denote the space of functions from $ball\ X$ into $\overline{\mathbb{D}}$, where of course $\overline{\mathbb{D}} = \{z \in \mathbb{C} : |z| \leq 1\}$ (is the closed unit disc in $\mathbb{C}$); then, by Tychonoff's theorem (see Theorem A.4.15), Z is a compact Hausdorff space with respect to the product topology.

Consider the map $F : ball\ X^* \to Z$ defined by

$$F(\phi) \ = \ \phi|_{ball\ X} \ .$$

The definitions of the weak* topology on $ball\ X^*$ and the product topology on Z show easily that a net $\{\phi_i : i \in I\}$ converges with respect to the weak* topology to ϕ in $ball\ X^*$ if and only if the net

$\{F(\phi_i) : i \in I\}$ converges with respect to the product topology to $F(\phi)$ in Z. Also, since the map F is clearly 1-1, we find that F maps *ball* X^* (with the weak* topology) homeomorphically onto its image *ran* F (with the subspace topology inherited from Z).

Thus, the proof of the theorem will be complete once we show that *ran* F is closed in Z - since a closed subspace of a compact space is always compact. This is where fact (2) stated at the start of the proof is needed. Whenever $x, y \in ballX$ and $\alpha, \beta \in \mathbb{C}$ are such that $\alpha x + \beta y \in ball\ X$, define

$$K_{x,y,\alpha,\beta} \ = \ \{f \in Z : f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)\} \ ;$$

note that $K_{x,y,\alpha,\beta}$ is a closed subset of Z (when defined) and consequently compact; the content of fact (2) above is that the image *ran* F is nothing but the set $\cap_{x,y,\alpha,\beta} K_{x,y,\alpha,\beta}$, and the proof is complete, since an arbitrary intersection of closed sets is closed. □

The reason that we have restricted ourselves to discussing topological vector spaces, where the topology is induced by a family of seminorms in the above fashion, is that most 'decent' t.v.s. arise in this manner. The object of the following exercises is to pave the way for identifying topological vector spaces of the above sort with the so-called **locally convex** topological vector spaces.

EXERCISE 1.6.10 *(1) Let X be a vector space. Show that if $p : X \to \mathbb{R}$ is a semi-norm on X, and if $C = \{x \in X : p(x) < 1\}$ denotes the associated 'open unit ball', then C satisfies the following conditions:*

(a) C is convex - i.e., $x, y \in C,\ 0 \leq t \leq 1 \ \Rightarrow\ tx + (1-t)y \in C$;

(b) C is 'absorbing' - i.e., if $x \in X$ is arbitrary, then there exists $\epsilon > 0$ such that $\epsilon x \in C$; and

(c) C is 'balanced' - i.e., $x \in C, \alpha \in \mathbb{C}, |\alpha| \leq 1 \ \Rightarrow \alpha x \in C$.

(d) if $x \in X$, then the set $\{\lambda > 0 : \lambda x \in C\}$ is an open set. (Hint: this is actually just the open interval given by $(0, \frac{1}{p(x)})$, where $\frac{1}{0}$ is to be interpreted as ∞.)

(2) (i) Show that any absorbing set contains 0.

(ii) If C is a convex and absorbing set in X, define $p_C : X \to$ R *by*

$$p_C(x) \ = \ \inf\{\lambda > 0 : \lambda^{-1}x \in C\} \ , \qquad (1.6.22)$$

and show that the map p_C satisfies the conditions of Lemma 1.4.1.

(3) In the converse direction to (1), if C is any subset of X satisfying the conditions $(i)(a)-(d)$, show that there exists a semi-norm p on X such that $C = \{x \in X : p(x) < 1\}$. (Hint: Define p by equation 1.6.22, and verify that this does the job.) The function p is called the **Minkowski functional** *associated with the set C.*

The proof of the following proposition is a not very difficult use of the Minkowski functionals associated with convex, balanced, absorbing sets; we omit the proof; the reader is urged to try and prove the proposition, failing which she should look up a proof (in [Yos], for instance).

PROPOSITION 1.6.11 *The following conditions on a topological vector space X are equivalent:*

(i) the origin 0 has a neighbourhood base consisting of convex, absorbing and balanced open sets;

(ii) the topology on X is induced by a family of semi-norms - see Remark 1.6.5.

A topological vector space which satisfies these equivalent conditions is said to be **locally convex**.

We conclude this section (and chapter) with a version of the **Hahn-Banach separation theorem**.

THEOREM 1.6.12 *(a) Suppose A and B are disjoint convex sets in a topological vector space X, and suppose A is open. Then there exists a continuous linear functional $\phi : X \to \mathbb{C}$ and a real number t such that*

$$Re \ \phi(a) \ < \ t \ \leq \ Re \ \phi(b) \ \forall \ a \in A, b \in B \ .$$

(b) If X is a locally convex topological vector space, and if $A \subset X$ (resp., $B \subset X$) is a compact (resp., closed) convex set, then there exists a continuous linear functional $\phi : X \to \mathbb{C}$ and $t_1, t_2 \in \mathbb{R}$ such that

$$Re\ \phi(a) \ < \ t_1 < t_2 \ \leq \ Re\ \phi(b)\ \forall\ a \in A, b \in B\ .$$

Proof : By considering real- and imaginary- parts, as in the proof of the Hahn-Banach theorem, it suffices to treat the case of a real vector space X.

(a) Let $C = A - B = \{a - b : a \in A, b \in B\}$, and fix $x_0 = a_0 - b_0 \in C$. The hypothesis ensures that $D = C - x_0$ is a convex open neighbourhood of 0, and consequently D is an absorbing set. As in Exercise 1.6.10(1)(iii), define

$$p(x) \ = \ \inf\{\alpha > 0 : \alpha^{-1}x \in D\}\ .$$

Then $p(x) \leq 1\ \forall\ x \in D$. Conversely, suppose $p(x) < 1$ for some $x \in X$; then, there exists some $0 < \alpha < 1$ such that $\alpha^{-1}x \in D$; since D is convex, $0 \in D$ and $\alpha^{-1} > 1$, it follows that $x \in D$; thus $p(x) < 1 \ \Rightarrow x \in D$. In particular,

$$0 \notin C \ \Rightarrow (-x_0) \notin D \ \Rightarrow p(-x_0) \geq 1\ .$$

Since $p(tx) = tp(x)\ \forall x \in X\ \forall t > 0$ - verify this! - we find thus that

$$p(-tx_0) \geq t\ \forall\ t > 0\ . \tag{1.6.23}$$

Now define $\phi_0 : X_0 = \mathbb{R}x_0 \to \mathbb{R}$ by $\phi_0(tx_0) = -t$, and note that $\phi_0(x) \leq p(x)\ \forall\ x \in \mathbb{R}x_0$. (*Reason:* Let $x = tx_0$; if $t \geq 0$, then $\phi(x) = -t \leq 0 \leq p(x)$; if $t < 0$, then, $\phi(x) = -t = |t| \leq p(-|t|x_0) = p(x)$, by equation 1.6.23.) Deduce from Lemma 1.4.1 that ϕ_0 extends to a linear functional $\phi : X \to \mathbb{R}$ such that $\phi|_{X_0} = \phi_0$ and $\phi(x) \leq p(x)\ \forall x \in X$. Notice that $\phi(x) \leq 1$ whenever $x \in D \cap (-D)$; since $V = D \cap (-D)$ is an open neighbourhood of 0, it follows that ϕ is 'bounded' and hence continuous. (Reason: if $x_n \to 0$, and if $\epsilon > 0$ is arbitrary, then (ϵV is also an open neighbourhood of 0, and so) there exists n_0 such that $x_n \in \epsilon V\ \forall\ n \geq n_0$; this implies that $|\phi(x_n)| \leq \epsilon\ \forall\ n \geq n_0$.)

Finally, if $a \in A, b \in B$, then $a - b - x_0 \in D$, and so, $\phi(a) - \phi(b) + 1 \ = \ \phi(a - b - x_0) \leq 1$; i.e., $\phi(a) \leq \phi(b)$. It follows that if $t = \inf\{\phi(b) : b \in B\}$, then $\phi(a) \leq t \leq \phi(b) \ \forall \ a \in A, b \in B$. Now, if $a \in A$, since A is open, it follows that $a - \epsilon x_0 \in A$ for sufficiently small $\epsilon > 0$; hence also, $\phi(a) + \epsilon \ = \ \phi(a - \epsilon x_0) \leq t$; thus, we do indeed have $\phi(a) < t \leq \phi(b) \ \forall \ a \in A, b \in B$.

(b) For each $x \in A$, (since $0 \notin B - x$, and since addition is continuous) it follows that there exists a convex open neighbourhood V_x of 0 such that $B \cap (x + V_x + V_x) = \emptyset$. Since $\{x + V_x : x \in A\}$ is an open cover of the compact set A, there exist $x_1, \cdots, x_n \in A$ such that $A \subset \ \cup_{i=1}^n (x_i + V_{x_i})$. Let $V = \cap_{i=1}^n V_{x_i}$; then V is a convex open neighbourhood of 0, and if $U = A + V$, then U is open (since $U = \cup_{a \in A} a + V$) and convex (since A and V are); further, $A + V \ \subset \cup_{i=1}^n (x_i + V_{x_i} + V) \subset \cup_{i=1}^n (x_i + V_{x_i} + V_{x_i})$; consequently, we see that $U \cap B = \emptyset$. Then, by (a) above, we can find a continuous linear functional $\phi : X \to \mathbb{R}$ and a scalar $t \in \mathbb{R}$ such that $\phi(u) < t \leq \phi(b) \ \forall \ u \in U, b \in B$. Then, $\phi(A)$ is a compact subset of $(-\infty, t)$, and so we can find $t_1, t_2 \in \mathbb{R}$ such that $\sup \phi(A) < t_1 < t_2 < t$, and the proof is complete. $\Box$

The reason for calling the preceding result a 'separation theorem' is this: given a continuous linear functional ϕ on a topological vector space X, the set $H = \{x \in X : Re\phi(x) = c\}$ (where c is an arbitrary real number) is called a 'hyperplane' in X; thus, for instance, part (b) of the preceding result says that a compact convex set can be 'separated' from a closed convex set from which it is disjoint, by a hyperplane.

An arbitrary t.v.s. may not admit any non-zero continuous linear functionals. (Example?) In order to ensure that there are 'sufficiently many' elements in X^*, it is necessary to establish something like the Hahn-Banach theorem, and for this to be possible, we need something more than just a topological vector space; this is where local convexity comes in.

Chapter 2

Hilbert spaces

2.1 Inner Product spaces

While normed spaces permit us to study 'geometry of vector spaces', we are constrained to discussing those aspects which depend only upon the notion of 'distance between two points'. If we wish to discuss notions that depend upon the angles between two lines, we need something more - and that something more is the notion of an *inner product*.

The basic notion is best illustrated in the example of the space R^2 that we are most familiar with, where the most natural norm is what we have called $||\cdot||_2$. The basic fact from plane geometry that we need is the so-called *cosine law* which states that if A, B, C are the vertices of a triangle and if θ is the angle at the vertex C, then

$$2(AC)(BC)\ cos\ \theta \ = \ (AC)^2 + (BC)^2 - (AB)^2\ .$$

If we apply this to the case where the points A, B and C are represented by the vectors $x = (x_1, x_2),\ y = (y_1, y_2)$ and $(0, 0)$ respectively, we find that

$$\begin{aligned} 2||x|| \cdot ||y|| \cdot \ cos\ \theta \ &= \ ||x||^2 + ||y||^2 - ||x - y||^2 \\ &= \ 2\ (x_1y_1 \ + \ x_2y_2\ ). \end{aligned}$$

Thus, we find that the function of two (vector) variables given by

$$\langle x, y\rangle \ = \ x_1y_1 \ + \ x_2y_2 \tag{2.1.1}$$

simultaneously encodes the notion of angle as well as distance (and has the explicit interpretation $\langle x, y \rangle = ||x|| \, ||y|| \, cos \, \theta$). This is because the norm can be recovered from the inner product by the equation

$$||x|| = \langle x, x \rangle^{\frac{1}{2}} . \tag{2.1.2}$$

The notion of an inner product is the proper abstraction of this function of two variables.

DEFINITION 2.1.1 *(a) An* **inner product** *on a (complex) vector space V is a mapping $V \times V \ni (x, y) \mapsto \langle x, y \rangle \in \mathbb{C}$ which satisfies the following conditions, for all $x, y, z \in V$ and $\alpha \in \mathbb{C}$:*
(i) (positive definiteness) $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0 \Leftrightarrow x = 0$;
(ii) (Hermitian symmetry) $\langle x, y \rangle = \overline{\langle y, x \rangle}$;
(iii) (linearity in first variable) $\langle \alpha x + \beta z, y \rangle = \alpha \langle x, y \rangle + \beta \langle z, y \rangle$.

An **inner product space** *is a vector space equipped with a (distinguished) inner product.*

(b) An inner product space which is complete in the norm coming from the inner product is called a **Hilbert space***.*

EXAMPLE 2.1.2 (1) If $z = (z_1, \cdots, z_n), w = (w_1, \cdots, w_n) \in \mathbb{C}^n$, define

$$\langle z, w \rangle = \sum_{i=1}^{n} z_i \overline{w_i} \; ; \tag{2.1.3}$$

it is easily verified that this defines an inner product on $\mathbb{C}^n$.

(2) The equation

$$\langle f, g \rangle = \int_{[0,1]} f(x) \overline{g(x)} \, dx \tag{2.1.4}$$

is easily verified to define an inner product on $C[0, 1]$. □

As in the (real) case discussed earlier of $\mathbb{R}^2$, it is generally true that any inner product gives rise to a norm on the underlying space via equation 2.1.2. Before verifying this fact, we digress with an exercise that states some easy consequences of the definitions.

EXERCISE 2.1.3 *Suppose we are given an inner product space V; for $x \in V$, define $||x||$ as in equation 2.1.2, and verify the following identities, for all $x, y, z \in V$, $\alpha \in \mathbb{C}$:*

(1) $\langle x, y + \alpha z \rangle = \langle x, y \rangle + \overline{\alpha} \langle x, z \rangle$;

(2) $||x + y||^2 = ||x||^2 + ||y||^2 + 2 \, Re\langle x, y \rangle$;

(3) two vectors in an inner product space are said to be **orthogonal** *if their inner product is 0; deduce from (2) above and an easy induction argument that if $\{x_1, x_2, \cdots, x_n\}$ is a set of pairwise orthogonal vectors, then*

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

(4) $||x + y||^2 + ||x - y||^2 = 2 \, (||x||^2 + ||y||^2)$; *draw some diagrams and convince yourself as to why this identity is called the* **parallelogram identity***;*

(5) (Polarisation identity) $4\langle x, y \rangle = \sum_{k=0}^{3} i^k \langle x + i^k y, x + i^k y \rangle$, *where, of course,* $i = \sqrt{-1}$.

The first (and very important) step towards establishing that any inner product defines a norm via equation 2.1.2 is the following celebrated inequality.

PROPOSITION 2.1.4 *(***Cauchy-Schwarz inequality***)*

If x, y are arbitrary vectors in an inner product space V, then

$$|\langle x, y \rangle| \leq ||x|| \cdot ||y|| .$$

Further, this inequality is an equality if and only if the vectors x and y are linearly dependent.

Proof: If $y = 0$, there is nothing to prove; so we may assume, without loss of generality, that $||y|| = 1$ (since the statement of the proposition is unaffected upon scaling y by a constant).

Notice now that, for arbitrary $\alpha \in \mathbb{C}$,

$$\begin{aligned} 0 &\leq ||x - \alpha y||^2 \\ &= ||x||^2 + |\alpha|^2 - 2Re(\alpha \langle y, x \rangle) . \end{aligned}$$

A little exercise in the calculus shows that this last expression is minimised for the choice $\alpha_0 = \langle x, y \rangle$, for which choice we find, after some minor algebra, that

$$0 \leq ||x - \alpha_0 y||^2 = ||x||^2 - |\langle x, y \rangle|^2 ,$$

thereby establishing the desired inequality.

The above reasoning shows that (if $||y|| = 1$) if the inequality becomes an equality, then we should have $x = \alpha_0 y$, and the proof is complete. □

REMARK 2.1.5 (1) For future reference, we note here that the positive definite property $||x|| = 0 \Rightarrow x = 0$ was never used in the proof of the Cauchy-Schwarz inequality; in particular, the inequality is valid for any *positive definite sesquilinear form* on V; i.e., suppose $B : V \times V \to \mathbb{C}$ is a mapping which is positive-semidefinite - meaning only that $B(x, x) \geq 0 \; \forall x$ - and satisfies the conditions of Hermitian symmetry and linearity (resp., 'conjugate' linearity) in the first (resp., second) variable (see definition of inner product and Exercise 2.1.3 (1)) ; then

$$|B(x, y)|^2 \leq B(x, x) \cdot B(y, y) .$$

(2) Similarly, any sesquilinear form B on a complex vector space satisfies the polarisation identity, meaning:

$$B(x, y) = \sum_{k=0}^{3} i^k \; B(x + i^k y, x + i^k y) \; \forall \; x, y \in V .$$

□

COROLLARY 2.1.6 *Any inner product gives rise to a norm via equation 2.1.2.*

Proof: Positive-definiteness and homogeneity with respect to scalar multiplication are obvious; as for the triangle inequality,

$$\begin{aligned} ||x + y||^2 &= ||x||^2 + ||y||^2 + 2 \; Re\langle x, y\rangle \\ &\leq ||x||^2 + ||y||^2 + 2||x|| \cdot ||y|| , \end{aligned}$$

and the proof is complete. □

EXERCISE 2.1.7 *We use the notation of Examples 1.2.2(1) and 2.1.2.*

(1) Show that

$$|\sum_{i=1}^{n} z_i \overline{w_i}|^2 \leq \left(\sum_{i=1}^{n} |z_i|^2 \right) \left(\sum_{i=1}^{n} |w_i|^2 \right) , \; \forall \; z, w \in \mathbb{C}^n .$$

(2) Deduce from (1) that the series $\sum_{i=1}^{\infty} \alpha_i \overline{\beta_i}$ converges, for any $\alpha, \beta \in \ell^2$, and that

$$|\sum_{i=1}^{\infty} \alpha_i \overline{\beta_i}|^2 \ \leq \ \left(\sum_{i=1}^{\infty} |\alpha_i|^2\right) \ \left(\sum_{i=1}^{\infty} |\beta_i|^2\right) \ , \ \forall \ \alpha, \beta \in \ell^2 \ ;$$

deduce that ℓ^2 is indeed (a vector space, and in fact) an inner product space, with respect to inner product defined by

$$\langle \alpha, \beta \rangle \ = \ \sum_{i=1}^{\infty} \alpha_i \overline{\beta_i} \ . \tag{2.1.5}$$

(3) Write down what the Cauchy-Schwarz inequality translates into in the example of $C[0,1]$.

(4) Show that the inner product is continuous as a mapping from $V \times V$ into $\mathbb{C}$. (In view of Corollary 2.1.6 and the discussion in Exercise 1.2.3, this makes sense.)

2.2 Some preliminaries

This section is devoted to a study of some elementary aspects of the theory of Hilbert spaces and the bounded linear operators between them.

Recall that a Hilbert space is a complete inner product space - by which, of course, is meant a Banach space where the norm arises from an inner product. We shall customarily use the symbols $\mathcal{H}, \mathcal{K}$ and variants of these symbols (obtained using subscripts, primes, etc.) for Hilbert spaces. Our first step is to arm ourselves with a reasonably adequate supply of examples of Hilbert spaces.

Example 2.2.1 (1) $\mathbb{C}^n$ is an example of a finite-dimensional Hilbert space, and we shall soon see that these are essentially the only such examples. We shall write ℓ^2_n for this Hilbert space.

(2) ℓ^2 is an infinite-dimensional Hilbert space - see Exercises 2.1.7(2) and 1.5.2(5). Nevertheless, this Hilbert space is not 'too big', since it is at least equipped with the pleasant feature of being a **separable** Hilbert space - i.e., it is separable as a metric

space, meaning that it has a countable dense set. (Verify this assertion!)

(3) More generally, let S be an arbitrary set, and define

$$\ell^2(S) \; = \; \{x \; = \; ((x_s))_{s \in S} : \sum_{s \in S} |x_s|^2 < \infty\} \; .$$

(The possibly uncountable sum might be interpreted as follows: a typical element of $\ell^2(S)$ is a family $x = ((x_s))$ of complex numbers which is indexed by the set S, and which has the property that $x_s = 0$ except for s coming from some countable subset of S (which depends on the element x) and which is such that the possibly non-zero x_s's, when written out as a sequence in any (equivalently, some) way, constitute a square-summable sequence.)

Verify that $\ell^2(S)$ is a Hilbert space in a natural fashion.

(4) This example will make sense to the reader who is already familiar with the theory of measure and Lebesgue integration; the reader who is not, may safely skip this example; the subsequent exercise will effectively recapture this example, at least in all cases of interest. In any case, there is a section in the Appendix - see §A.5 - which is devoted to a brief introduction to the theory of measure and Lebesgue integration.

Suppose $(X, \mathcal{B}, \mu)$ is a measure space. Let $\mathcal{L}^2(X, \mathcal{B}, \mu)$ denote the space of $\mathcal{B}$-measurable complex-valued functions f on X such that $\int_X |f|^2 d\mu \; < \; \infty$. Note that $|f + g|^2 \; \leq \; 2(|f|^2 + |g|^2)$, and deduce that $\mathcal{L}^2(X, \mathcal{B}, \mu)$ is a vector space. Note next that $|f\overline{g}| \; \leq \; \frac{1}{2}(|f|^2 + |g|^2)$, and so the right-hand side of the following equation makes sense, if $f, g \in \mathcal{L}^2(X, \mathcal{B}, \mu)$:

$$\langle f, g \rangle \; = \; \int_X f\overline{g} d\mu \; . \tag{2.2.6}$$

It is easily verified that the above equation satisfies all the requirements of an inner product with the solitary possible exception of the positive-definiteness axiom: if $\langle f, f \rangle = 0$, it can only be concluded that $f = 0$ *a.e.* - meaning that $\{x : f(x) \neq 0\}$ is a set of μ-measure 0 (which might very well be non-empty).

Observe, however, that the set $N = \{f \in \mathcal{L}^2(X, \mathcal{B}, \mu) : f = 0 \; a.e.\}$ is a vector subspace of $\mathcal{L}^2(X, \mathcal{B}, \mu)$; now consider the quotient space $L^2(X, \mathcal{B}, \mu) = \mathcal{L}^2(X, \mathcal{B}, \mu)/N$, a typical element of which is thus an equivalence class of square-integrable functions, where two functions are considered to be equivalent if they agree outside a set of μ-measure 0.

For simplicity of notation, we shall just write $L^2(X)$ for $L^2(X, \mathcal{B}, \mu)$, and we shall denote an element of $L^2(X)$ simply by such symbols as f, g, etc., and think of these as actual functions with the understanding that we shall identify two functions which agree almost everywhere. The point of this exercise is that equation 2.2.6 now does define a genuine inner product on $L^2(X)$; most importantly, it is true that $L^2(X)$ is complete and is thus a Hilbert space. □

EXERCISE 2.2.2 *(1) Suppose X is an inner product space. Let $\overline{X}$ be a completion of X regarded as a normed space - see Exercise 1.5.5. Show that $\overline{X}$ is actually a Hilbert space. (Thus, every inner product space has a Hilbert space completion.)*

(2) Let $X = C[0,1]$ and define

$$\langle f, g \rangle = \int_0^1 f(x)\overline{g(x)}dx \ .$$

Verify that this defines a genuine (positive-definite) inner product on $C[0,1]$. The completion of this inner product space is a Hilbert space - see (1) above - which may be identified with what was called $L^2([0,1], \mathcal{B}, m)$ in Example 2.2.1(4), where ($\mathcal{B}$ is the σ-algebra of Borel sets in [0,1] and) m denotes the so-called Lebesgue measure on [0,1].

In the sequel, we will have to deal with 'uncountable sums' repeatedly, so it will be prudent to gather together some facts concerning such sums, which we shall soon do in some exercises. These exercises will use the notion of *nets*, which we pause to define and elucidate with some examples.

DEFINITION 2.2.3 *A* **directed set** *is a partially ordered set $(I, \leq)$ which, - in addition to the usual requirements of reflexivity ($i \leq i$), antisymmetry ($i \leq j, j \leq i \Rightarrow i = j$) and transitivity ($i \leq j, j \leq k \Rightarrow i \leq k$), which have to be satisfied by any partially ordered set - satisfies the following property:*

$$\forall\ i, j \in I,\ \exists\ k \in I \text{ such that } i \leq k, j \leq k \ .$$

A **net** *in a set X is a family $\{x_i : i \in I\}$ of elements of X which is indexed by some directed set I.*

EXAMPLE 2.2.4 (1) The motivating example for the definition of a directed set is the set $\mathbb{N}$ of natural numbers with the natural ordering, and a net indexed by this directed set is nothing but a sequence.

(2) Let S be any set and let $\mathcal{F}(S)$ denote the class of all finite subsets of S; this is a directed set with respect to the order defined by inclusion; thus, $F_1 \leq F_2 \Leftrightarrow F_1 \subseteq F_2$. This will be referred to as the directed set of finite subsets of S.

(3) Let X be any topological space, and let $x \in X$; let $\mathcal{N}(x)$ denote the family of all open neighbourhoods of the point x; then $\mathcal{N}(x)$ is directed by 'reverse inclusion'; i.e, it is a directed set if we define $U \leq V \Leftrightarrow V \subseteq U$.

(4) If I and J are directed sets, then the Cartesian product $I \times J$ is a directed set with respect to the 'co-ordinate-wise ordering' defined by $(i_1, j_1) \leq (i_2, j_2) \Leftrightarrow i_1 \leq i_2$ *and* $j_1 \leq j_2$. □

The reason that nets were introduced in the first place was to have analogues of sequences when dealing with more general topological spaces than metric spaces. In fact, if X is a general topological space, we say that a net $\{x_i : i \in I\}$ in X xonverges to a point x if, for every open neighbourhood U of x, it is possible to find an index $i_0 \in I$ with the property that $x_i \in U$ whenever $i_0 \leq i$. (As with sequences, we shall, in the sequel, write $i \geq j$ to mean $j \leq i$ in an abstract partially ordered set.)

EXERCISE 2.2.5 *(1) If $f : X \to Y$ is a map of topological spaces, and if $x \in X$, then show that f is continuous at x if and only the net $\{f(x_i) : i \in I\}$ converges to $f(x)$ in Y whenever $\{x_i : i \in I\}$ is a net which converges to x in X. (Hint: use the directed set of Example 2.2.4(3).)*

(2) Define what should be meant by saying that a net in a metric space is a 'Cauchy net', and show that every convergent net is a Cauchy net.

(3) Show that if X is a metric space which is complete - meaning, of course, that Cauchy sequences in X converge - show

that every Cauchy net in X also converges. (Hint: for each n, pick $i_n \in I$ such that $i_1 \leq i_2 \leq \cdots \leq i_n \leq \cdots$, and such that $d(x_i, x_j) < \frac{1}{n}$, whenever $i, j \geq i_n$; now show that the net should converge to the limit of the Cauchy sequence $\{x_{i_n}\}_{n \in \mathbb{N}}$.)

(4) Is the Cartesian product of two directed sets directed with respect to the 'dictionary ordering'?

We are now ready to discuss the problem of 'uncountable sums'. We do this in a series of exercises following the necessary definition.

DEFINITION 2.2.6 *Suppose X is a normed space, and suppose $\{x_s : s \in S\}$ is some indexed collection of vectors in X, whose members may not all be distinct; thus, we are looking at a function $S \ni s \mapsto x_s \in X$. We shall, however, be a little sloppy and simply call $\{x_s : s \in S\}$ a family of vectors in X - with the understanding being that we are actually talking about a function as above.*

If $F \in \mathcal{F}(S)$ define $x(F) = \sum_{s \in F} x_s$ - in the notation of Example 2.2.4(2). We shall say that the family $\{x_s : s \in S\}$ is **unconditionally summable***, and has the sum x, if the net $\{x(F) : F \in \mathcal{F}(S)\}$ converges to x in X. When this happens, we shall write*

$$x = \sum_{s \in S} x_s .$$

EXERCISE 2.2.7 *(1) If X is a Banach space, then show that a family $\{x_s : s \in S\} \subset X$ is unconditionally summable if and only if, given any $\epsilon > 0$, it is possible to find a finite subset $F_0 \subset S$ such that $||\sum_{s \in F} x_s|| < \epsilon$ whenever F is a finite subset of S which is disjoint from F_0. (Hint: use Exercise 2.2.5(3).)*

(2) If we take S to be the set $\mathbb{N}$ of natural numbers, and if the sequence $\{x_n : n \in \mathbb{N}\} \subset X$ is unconditionally summable, show that if π is any permutation of $\mathbb{N}$, then the series $\sum_{n=1}^{\infty} x_{\pi(n)}$ is convergent in X and the sum of this series is independent of the permutation π. (This is the reason for the use of the adjective 'unconditional' in the preceding definition.)

(3) Suppose $\{a_i : i \in S\} \subset [0, \infty)$; regard this family as a subset of the complex normed space $\mathbb{C}$; show that the following conditions are equivalent:

(i) this family is unconditionally summable;

(ii) $A = sup\{\sum_{s\in F} a_s : F \in \mathcal{F}(S)\} < \infty$, *(in the notation of Definition 2.2.6).*

When these equivalent conditions are satisfied, show that;

(a) $a_s \neq 0$ *for at most countably many* $s \in S$*; and*

(b) $\sum_{s\in S} a_s \;=\; A$.

2.3 Orthonormal bases

DEFINITION 2.3.1 *A collection* $\{e_i : i \in I\}$ *in an inner product space is said to be* **orthonormal** *if*

$$\langle e_i, e_j\rangle \;=\; \delta_{ij} \;\; \forall \; i,j \in I \; .$$

Thus, an orthonormal set is nothing but a set of unit vectors which are pairwise orthogonal; here, and in the sequel, we say that two vectors x, y in an inner product space are **orthogonal** if $\langle x, y\rangle \;=\; 0$, and we write $x \perp y$.

EXAMPLE 2.3.2 (1) In ℓ^2_n, for $1 \leq i \leq n$, let e_i be the element whose i-th co-ordinate is 1 and all other co-ordinates are 0; then $\{e_1, \cdots, e_n\}$ is an orthonormal set in ℓ^2_n.

(2) In ℓ^2, for $1 \leq n < \infty$, let e_i be the element whose i-th co-ordinate is 1 and all other co-ordinates are 0; then $\{e_n : n = 1, 2, \cdots\}$ is an orthonormal set in ℓ^2. More generally, if S is any set, an entirely similar prescription leads to an orthonormal set $\{e_s : s \in S\}$ in $\ell^2(S)$.

(3) In the inner product space $C[0,1]$ - with inner product as described in Exercise 2.2.2 - consider the family $\{e_n : n \in \mathbb{Z}\}$ defined by $e_n(x) \;=\; exp(2\pi i n x)$, and show that this is an orthonormal set; hence this is also an orthonormal set when regarded as a subset of $L^2([0,1], m)$ - see Exercise 2.2.2(2). □

PROPOSITION 2.3.3 *Let* $\{e_1, e_2, \cdots, e_n\}$ *be an orthonormal set in an inner product space* X*, and let* $x \in X$ *be arbitrary. Then,*

(i) if $x \;=\; \sum_{i=1}^n \alpha_i e_i, \;\alpha_i \in \mathbb{C}$, *then* $\alpha_i \;=\; \langle x, e_i\rangle \;\forall i$*;*

(ii) $(x - \sum_{i=1}^n \langle x, e_i\rangle e_i \;) \;\perp\; e_j \;\forall 1 \leq j \leq n$*;*

*(iii) (***Bessel's inequality***)* $\sum_{i=1}^n |\langle x, e_i\rangle|^2 \;\leq\; ||x||^2$.

Proof: (i) If x is a linear combination of the e_j's as indicated, compute $\langle x, e_i\rangle$, and use the assumed orthonormality of the e_j's, to deduce that $\alpha_i = \langle x, e_i\rangle$.

(ii) This is an immediate consequence of (i).

(iii) Write $y = \sum_{i=1}^n \langle x, e_i\rangle e_i$, $z = x - y$, and deduce from (two applications of) Exercise 2.1.3(3) that

$$\begin{aligned} ||x||^2 &= ||y||^2 + ||z||^2 \\ &\geq ||y||^2 \\ &= \sum_{i=1}^n |\langle x, e_i\rangle|^2 . \end{aligned}$$

□

We wish to remark on a few consequences of this proposition; for one thing, (i) implies that an arbitrary orthonormal set is linearly independent; for another, if we write $\bigvee\{e_i : i \in I\}$ for the vector subspace *spanned* by $\{e_i : i \in I\}$ - i.e., this consists of the set of linear combinations of the e_i's, and is the smallest subspace containing $\{e_i : i \in I\}$ - it follows from (i) that we know how to write any element of $\bigvee\{e_i : i \in I\}$ as a linear combination of the e_i's.

We shall find the following notation convenient in the sequel: if S is a subset of an inner product space X, let $[S]$ denote the smallest closed subspace containing S; it should be clear that this could be described in either of the following equivalent ways: (a) $[S]$ is the intersection of all closed subspaces of X which contain S, and (b) $[S] = \overline{\bigvee S}$. (Verify that (a) and (b) describe the same set.)

LEMMA 2.3.4 *Suppose* $\{e_i : i \in I\}$ *is an orthonormal set in a Hilbert space* $\mathcal{H}$*. Then the following conditions on an arbitrary family* $\{\alpha_i : i \in I\}$ *of complex numbers are equivalent:*

(i) the family $\{\alpha_i e_i : i \in I\}$ *is unconditionally summable in* $\mathcal{H}$*;*

(ii) the family $\{|\alpha_i|^2 : i \in I\}$ *is unconditionally summable in* $\mathbb{C}$*;*

(iii) sup $\{\sum_{i\in F} : |\alpha_i|^2 : F \in \mathcal{F}(I)\} < \infty$, *where* $\mathcal{F}(I)$ *denote the directed set of finite subsets of the set* I*;*

(iv) there exists a vector $x \in [\{e_i : i \in I\}]$ *such that* $\langle x, e_i\rangle = \alpha_i \ \forall i \in I$.

Proof: For each $F \in \mathcal{F}(I)$, let us write $x(F) = \sum_{i \in F} \alpha_i e_i$. We find, from Exercise 2.1.3(3) that

$$||x(F)||^2 = \sum_{i \in F} |\alpha_i|^2 = A(F) \; ; (say)$$

thus we find, from Exercise 2.2.7 (1), that the family $\{\alpha_i e_i : i \in I\}$ (resp., $\{|\alpha_i|^2 : i \in I\}$) is unconditionally summable in the Hilbert space $\mathcal{H}$ (resp., $\mathbb{C}$) if and only if, for each $\epsilon > 0$, it is possible to find $F_0 \in \mathcal{F}(I)$ such that

$$||x(F)||^2 = A(F) < \epsilon \; \forall \; F \in \mathcal{F}(I - F_0) \; ;$$

this latter condition is easily seen to be equivalent to condition (iii) of the lemma; thus, the conditions (i), (ii) and (iii) of the lemma are indeed equivalent.

(i) $\Rightarrow$ *(iv)* : Let $x = \sum_{i \in I} \alpha_i e_i$; in the preceding notation, the net $\{x(F) : F \in \mathcal{F}(I)\}$ converges to x in $\mathcal{H}$. Hence also the net $\{\langle x(F), e_i \rangle : F \in \mathcal{F}(I)\}$ converges to $\langle x, e_i \rangle$, for each $i \in I$. An appeal to Proposition 2.3.3 (1) completes the proof of this implication.

(iv) $\Rightarrow$ *(iii)* : This follows immediately from Bessel's inequality. □

We are now ready to establish the fundamental proposition concerning *orthonormal bases* in a Hilbert space.

PROPOSITION 2.3.5 *The following conditions on an orthonormal set* $\{e_i : i \in I\}$ *in a Hilbert space* $\mathcal{H}$ *are equivalent:*

(i) $\{e_i : i \in I\}$ *is a maximal orthonormal set, meaning that it is not strictly contained in any other orthonormal set;*

(ii) $x \in \mathcal{H} \Rightarrow x = \sum_{i \in I} \langle x, e_i \rangle e_i$;

(iii) $x, y \in \mathcal{H} \Rightarrow \langle x, y \rangle = \sum_{i \in I} \langle x, e_i \rangle \langle e_i, y \rangle$;

(iv) $x \in \mathcal{H} \Rightarrow ||x||^2 = \sum_{i \in I} |\langle x, e_i \rangle|^2$.

Such an orthonormal set is called an **orthonormal basis** *of* $\mathcal{H}$.

Proof : *(i)* $\Rightarrow$ *(ii)* : It is a consequence of Bessel's inequality and (the equivalence *(iii)* $\Leftrightarrow$ *(iv)* of) the last lemma that there exists a vector, call it $x_0 \in \mathcal{H}$, such that $x_0 = \sum_{i \in I} \langle x, e_i \rangle e_i$. If

$x \neq x_0$, and if we set $f = \frac{1}{||x-x_0||}(x - x_0)$, then it is easy to see that $\{e_i : i \in I\} \cup \{f\}$ is an orthonormal set which contradicts the assumed maximality of the given orthonormal set.

(ii) $\Rightarrow$ *(iii)* : For $F \in \mathcal{F}(I)$, let $x(F) = \sum_{i \in F} \langle x, e_i \rangle e_i$ and $y(F) = \sum_{i \in F} \langle y, e_i \rangle e_i$, and note that, by the assumption (ii) (and the meaning of the uncountable sum), and the assumed orthonormality of the e_i's, we have

$$\begin{aligned}
\langle x, y \rangle &= \lim_F \langle x(F), y(F) \rangle \\
&= \lim_F \sum_{i \in F} \langle x, e_i \rangle \overline{\langle y, e_i \rangle} \\
&= \sum_{i \in F} \langle x, e_i \rangle \langle e_i, y \rangle \ .
\end{aligned}$$

(iii) $\Rightarrow$ *(iv)* : Put $y = x$.

(iv) $\Rightarrow$ *(i)* : Suppose $\{e_i : i \in I \cup J\}$ is an orthonormal set and suppose J is not empty; then for $j \in J$, we find, in view of (iv), that

$$||e_j||^2 = \sum_{i \in I} |\langle e_j, e_i \rangle|^2 = 0 \ ,$$

which is impossible in an orthonormal set; hence it must be that J is empty - i.e., the maximality assertion of (i) is indeed implied by (iv). □

COROLLARY 2.3.6 *Any orthonormal set in a Hilbert space can be 'extended' to an orthonormal basis - meaning that if* $\{e_i : i \in I\}$ *is any orthonormal set in a Hilbert space* $\mathcal{H}$*, then there exists an orthonormal set* $\{e_i : i \in J\}$ *such that* $I \cap J = \emptyset$ *and* $\{e_i : i \in I \cup J\}$ *is an orthonormal basis for* $\mathcal{H}$*.*

In particular, every Hilbert space admits an orthonormal basis.

Proof : This is an easy consequence of Zorn's lemma. (For the reader who is unfamiliar with Zorn's lemma, there is a small section - see §A.2 - in the appendix which discusses this very useful variant of the Axiom of Choice.) □

REMARK 2.3.7 (1) It follows from Proposition 2.3.5 (ii) that if $\{e_i : i \in I\}$ is an orthonormal basis for a Hilbert space $\mathcal{H}$, then $\mathcal{H} = [\{e_i : i \in I\}]$; conversely, it is true - and we shall soon prove this fact - that if an orthonormal set is **total** in the sense that the vector subspace spanned by the set is dense in the Hilbert space, then such an orthonormal set is necessarily an orthonormal basis.

(2) Each of the three examples of an orthonormal set that is given in Example 2.3.2, is in fact an orthonormal basis for the underlying Hilbert space. This is obvious in cases (1) and (2). As for (3), it is a consequence of the Stone-Weierstrass theorem - see Exercise A.6.10(i) - that the vector subspace of finite linear combinations of the exponential functions $\{exp(2\pi inx) : n \in \mathbb{Z}\}$ is dense in $\{f \in C[0,1] : f(0) = f(1)\}$ (with respect to the uniform norm - i.e., with respect to $|| \cdot ||_\infty$); in view of Exercise 2.2.2(2), it is not hard to conclude that this orthonormal set is total in $L^2([0,1], m)$ and hence, by remark (1) above, this is an orthonormal basis for the Hilbert space in question.

Since $exp(\pm 2\pi inx) = cos(2\pi nx) \pm i\ sin(2\pi nx)$, and since it is easily verified that $cos(2\pi mx) \perp sin(2\pi nx)\ \forall m, n = 1, 2, \cdots$, we find easily that

$$\{1 = e_0\} \cup \{\sqrt{2}cos(2\pi nx),\ \sqrt{2}sin(2\pi nx) : n = 1, 2, \cdots\}$$

is also an orthonormal basis for $L^2([0,1], m)$. (Reason: this is orthonormal, and this sequence spans the same vector subspace as is spanned by the exponential basis.) (Also, note that these are real-valued functions, and that the inner product of two real-valued functions in clearly real.) It follows, in particular, that if f is any (real-valued) continuous function defined on [0,1], then such a function admits the following **Fourier series** (with real coefficients):

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

where the meaning of this series is that we have convergence of the sequence of the partial sums to the function f with respect to the norm in $L^2[0,1]$. Of course, the coefficients a_n, b_n are given by

$$a_0 \;=\; \int_0^1 f(x)dx$$

$$\begin{aligned} a_n &= 2\int_0^1 f(x)cos(2\pi nx)dx \ , \ \forall \ n > 0, \\ b_n &= 2\int_0^1 f(x)sin(2\pi nx)dx \ , \ \forall \ n > 0 \end{aligned}$$

The theory of Fourier series was the precursor to most of modern functional analysis; it is for this reason that if $\{e_i : i \in I\}$ is any orthonormal basis of any Hilbert space, it is customary to refer to the numbers $\langle x, e_i \rangle$ as the **Fourier coefficients** of the vector x with respect to the orthonormal basis $\{e_i : i \in I\}$. □

It is a fact that any two orthonormal bases for a Hilbert space have the same cardinality, and this common cardinal number is called the **dimension** of the Hilbert space; the proof of this statement, in its full generality, requires facility with infinite cardinal numbers and arguments of a transfinite nature; rather than giving such a proof here, we adopt the following compromise: we prove the general fact in the appendix - see §A.2 - and discuss here only the cases that we shall be concerned with in these notes.

We temporarily classify Hilbert spaces into three classes of increasing complexity; the first class consists of finite-dimensional ones; the second class consists of separable Hilbert spaces - see Example 2.2.1(2) - which are not finite-dimensional; the third class consists of non-separable Hilbert spaces.

The purpose of the following result is to state a satisfying characterisation of separable Hilbert spaces.

PROPOSITION 2.3.8 *The following conditions on a Hilbert space $\mathcal{H}$ are equivalent:*
(i) $\mathcal{H}$ is separable;
(ii) $\mathcal{H}$ admits a countable orthonormal basis.

Proof : $(i) \Rightarrow (ii)$: Suppose D is a countable dense set in $\mathcal{H}$ and suppose $\{e_i : i \in I\}$ is an orthonormal basis for $\mathcal{H}$. Notice that

$$i \neq j \ \Rightarrow \ ||e_i - e_j||^2 = 2 \ . \tag{2.3.7}$$

Since D is dense in $\mathcal{H}$, we can, for each $i \in I$, find a vector $x_i \in D$ such that $||x_i - e_i|| < \frac{\sqrt{2}}{2}$. The identity 2.3.7 shows that the map $I \ni i \mapsto x_i \in D$ is necessarily 1-1; since D is countable, we may conclude that so is I.

$(ii) \Rightarrow (i)$: If I is a countable (finite or infinite) set and if $\{e_i : i \in I\}$ is an orthonormal basis for $\mathcal{H}$, let D be the set whose typical element is of the form $\sum_{j \in J} \alpha_j e_j$, where J is a finite subset of I and α_j are complex numbers whose real and imaginary parts are both rational numbers; it can then be seen that D is a countable dense set in $\mathcal{H}$. □

Thus, non-separable Hilbert spaces are those whose orthonormal bases are uncountable. It is probably fair to say that any true statement about a general non-separable Hilbert space can be established as soon as one knows that the statement is valid for separable Hilbert spaces; it is probably also fair to say that almost all useful Hilbert spaces are separable. So, the reader may safely assume that all Hilbert spaces in the sequel are separable; among these, the finite-dimensional ones are, in a sense, 'trivial', and one only need really worry about infinite-dimensional separable Hilbert spaces.

We next establish a lemma which will lead to the important result which is sometimes referred to as 'the projection theorem'.

LEMMA 2.3.9 *Let $\mathcal{M}$ be a closed subspace of a Hilbert space $\mathcal{H}$; (thus $\mathcal{M}$ may be regarded as a Hilbert space in its own right;) let $\{e_i : i \in I\}$ be any orthonormal basis for $\mathcal{M}$, and let $\{e_j : j \in J\}$ be any orthonormal set such that $\{e_i : i \in I \cup J\}$ is an orthonormal basis for $\mathcal{H}$, where we assume that the index sets I and J are disjoint. Then, the following conditions on a vector $x \in \mathcal{H}$ are equivalent:*

(i) $x \perp y \ \forall \ y \in \mathcal{M}$;

(ii) $x = \sum_{j \in J} \langle x, e_j \rangle e_j$.

Proof : The implication $(ii) \Rightarrow (i)$ is obvious. Conversely, it follows easily from Lemma 2.3.4 and Bessel's inequality that the 'series' $\sum_{i \in I} \langle x, e_i \rangle e_i$ and $\sum_{j \in J} \langle x, e_j \rangle e_j$ converge in $\mathcal{H}$. Let the sums of these 'series' be denoted by y and z respectively. Further, since $\{e_i : i \in I \cup J\}$ is an orthonormal basis for $\mathcal{H}$, it should be clear that $x = y + z$. Now, if x satisfies condition (i) of the Lemma, it should be clear that $y = 0$ and that hence, $x = z$, thereby completing the proof of the lemma. □

We now come to the basic notion of **orthogonal complement**.

DEFINITION 2.3.10 *The orthogonal complement* $S^\perp$ *of a subset* S *of a Hilbert space is defined by*

$$S^\perp = \{x \in \mathcal{H} : x \perp y \ \forall \ y \in S\} .$$

EXERCISE 2.3.11 *If* $S_0 \subset S \subset \mathcal{H}$ *are arbitrary subsets, show that*

$$S_0^\perp \supset S^\perp = \left(\bigvee S\right)^\perp = ([S])^\perp .$$

Also show that $S^\perp$ *is always a closed subspace of* $\mathcal{H}$.

We are now ready for the basic fact concerning orthogonal complements of closed subspaces.

THEOREM 2.3.12 *Let* $\mathcal{M}$ *be a closed subspace of a Hilbert space* $\mathcal{H}$. *Then,*

(1) $\mathcal{M}^\perp$ *is also a closed subspace;*

(2) $\left(\mathcal{M}^\perp\right)^\perp = \mathcal{M}$;

(3) any vector $x \in \mathcal{H}$ *can be uniquely expressed in the form* $x = y + z$, *where* $y \in \mathcal{M}$, $z \in \mathcal{M}^\perp$;

(4) if x, y, z *are as in (3) above, then the equation* $Px = y$ *defines a bounded operator* $P \in \mathcal{L}(\mathcal{H})$ *with the property that*

$$||Px||^2 = \langle Px, x\rangle = ||x||^2 - ||x - Px||^2 , \ \forall x \in \mathcal{H} .$$

Proof : (i) This is easy - see Exercise 2.3.11.

(ii) Let $I, J, \{e_i : i \in I \cup J\}$ be as in Lemma 2.3.9. We assert, to start with, that in this case, $\{e_j : j \in J\}$ is an orthonormal basis for $\mathcal{M}^\perp$. Suppose this were not true; since this is clearly an orthonormal set in $\mathcal{M}^\perp$, this would mean that $\{e_j : j \in J\}$ is not a maximal orthonormal set in $\mathcal{M}^\perp$, which implies the existence of a unit vector $x \in \mathcal{M}^\perp$ such that $\langle x, e_j\rangle = 0 \ \forall \ j \in J$; such an x would satisfy condition (i) of Lemma 2.3.9, but not condition (ii).

If we now revese the roles of $\mathcal{M}, \{e_i : i \in I\}$ and $\mathcal{M}^\perp, \{e_j : j \in J\}$, we find from the conclusion of the preceding paragraph that $\{e_i : i \in I\}$ is an orthonormal basis for $\left(\mathcal{M}^\perp\right)^\perp$, from which we may conclude the validity of (ii) of this theorem.

(iii) The existence of y and z was demonstrated in the proof of Lemma 2.3.9; as for uniqueness, note that if $x = y_1 + z_1$ is another such decomposition, then we would have

$$y - y_1 \ = \ z_1 - z \ \in \ \mathcal{M} \cap \mathcal{M}^\perp \ ;$$

but $w \in \mathcal{M} \cap \mathcal{M}^\perp \ \Rightarrow \ w \perp w \ \Rightarrow ||w||^2 = 0 \ \Rightarrow \ w = 0$.

(iv) The uniqueness of the decomposition in (iii) is easily seen to iimply that P is a linear mapping of $\mathcal{H}$ into itself; further, in the notation of (iii), we find (since $y \perp z$) that

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

this implies that $||Px|| \leq ||x|| \ \forall \ x \in \mathcal{H}$, and hence $P \in \mathcal{L}(\mathcal{H})$.

Also, since $y \perp z$, we find that

$$||Px||^2 \ = \ ||y||^2 \ = \ \langle y, y + z \rangle \ = \ \langle Px, x \rangle \ ,$$

thereby completing the proof of the theorem. □

The following corollary to the above theorem justifies the final assertion made in Remark 2.3.7(1).

Corollary 2.3.13 *The following conditions on an orthonormal set $\{e_i : i \in I\}$ in a Hilbert space $\mathcal{H}$ are equivalent:*

(i) $\{e_i : i \in I\}$ is an orthonormal basis for $\mathcal{H}$;

(ii) $\{e_i : i \in I\}$ is total in $\mathcal{H}$ - meaning, of course, that $\mathcal{H} \ = \ [\ \{e_i : i \in I\}\ ]$.

Proof : As has already been observed in Remark 2.3.7(1), the implication $(i) \Rightarrow (ii)$ follows from Proposition 2.3.5(ii).

Conversely, suppose (i) is not satisfied; then $\{e_i : i \in I\}$ is not a maximal orthonormal set in $\mathcal{H}$; hence there exists a unit vector x such that $x \perp e_i \ \forall i \in I$; if we write $\mathcal{M} \ = \ [\ \{e_i : i \in I\}\ ]$, it follows easily that $x \in \mathcal{M}^\perp$, whence $\mathcal{M}^\perp \neq \{0\}$; then, we may deduce from Theorem 2.3.12(2) that $\mathcal{M} \neq \mathcal{H}$ - i.e., (ii) is also not satisfied. □

Remark 2.3.14 The operator $P \in \mathcal{L}(\mathcal{H})$ constrcted in Theorem 2.3.12(4) is referred to as the **orthogonal projection** onto the closed subspace $\mathcal{M}$. When it is necessary to indicate the relation between the subspace $\mathcal{M}$ and the projection P, we will write

$P = P_{\mathcal{M}}$ and $\mathcal{M} = ran\ P$; (note that $\mathcal{M}$ is indeed the range of the operator P;) some other facts about closed subspaces and projections are spelt out in the following exercises.

□

EXERCISE 2.3.15 *(1) Show that* $\left(S^{\perp}\right)^{\perp} = [S]$, *for any subset* $S \subset \mathcal{H}$.

(2) Let $\mathcal{M}$ *be a closed subspace of* $\mathcal{H}$, *and let* $P = P_{\mathcal{M}}$;

(a) show that $P_{\mathcal{M}^{\perp}} = 1 - P_{\mathcal{M}}$, *where we write* 1 *for the identity operator on* $\mathcal{H}$ *(the reason being that this is the multiplicative identity of the algebra* $\mathcal{L}(\mathcal{H})$*);*

(b) Let $x \in \mathcal{H}$; *the following conditions are equivalent:*

(i) $x \in \mathcal{M}$;

(ii) $x \in\ ran\ P\ (= P\mathcal{H})$;

(iii) $Px = x$;

(iv) $||Px|| = ||x||$.

(c) show that $\mathcal{M}^{\perp} = ker\ P = \{x \in \mathcal{H} : Px = 0\}$.

(3) Let $\mathcal{M}$ *and* $\mathcal{N}$ *be closed subspaces of* $\mathcal{H}$, *and let* $P = P_{\mathcal{M}}, Q = P_{\mathcal{N}}$; *show that the following conditions are equivalent:*

(i) $\mathcal{N} \subset \mathcal{M}$;

(ii) $PQ = Q$;

(i)' $\mathcal{M}^{\perp} \subset \mathcal{N}^{\perp}$;

(ii)' $(1 - Q)(1 - P) = 1 - P$;

(iii) $QP = Q$.

(4) With $\mathcal{M}, \mathcal{N}, P, Q$ *as in (3) above, show that the following conditions are equivalent:*

(i) $\mathcal{M} \perp \mathcal{N}$ *- i.e.,* $\mathcal{N} \subset \mathcal{M}^{\perp}$;

(ii) $PQ = 0$;

(iii) $QP = 0$.

(5) When the equivalent conditions of (4) are met, show that:

(a) $[\mathcal{M} \cup \mathcal{N}] = \mathcal{M} + \mathcal{N} = \{x + y : x \in \mathcal{M}, y \in \mathcal{N}\}$; *and*

(b) $(P + Q)$ *is the projection onto the subspace* $\mathcal{M} + \mathcal{N}$.

(c) more generally, if $\{\mathcal{M}_i : 1 \leq i \leq n\}$ *is a family of closed subspaces of* $\mathcal{H}$ *which are pairwise orthogonal, show that their 'vector sum' defined by* $\sum_{i=1}^{n} \mathcal{M}_i = \{\sum_{i=1}^{n} x_i\ :\ x_i \in \mathcal{M}_i\ \forall i\}$ *is*

a closed subspace and the projection onto this subspace is given by $\sum_{i=1}^n P_{\mathcal{M}_i}$.

If $\mathcal{M}_1, \cdots, \mathcal{M}_n$ are pairwise orthogonal closed subspaces - see Exercise 2.3.15(5)(c) above - and if $\mathcal{M} = \sum_{i=1}^n \mathcal{M}_i$, we say that $\mathcal{M}$ is the **direct sum** of the closed subspaces $\mathcal{M}_i, 1 \leq i \leq n$, and we write

$$\mathcal{M} = \oplus_{n=1}^{\infty} \mathcal{M}_i ; \tag{2.3.8}$$

conversely, whenever we use the above symbol, it will always be tacitly assumed that the $\mathcal{M}_i$'s are closed subspaces which are pairwise orthogonal and that $\mathcal{M}$ is the (closed) subspace spanned by them.

2.4 The Adjoint operator

We begin this section with an identification of the bounded linear functionals - i.e., the elements of the Banach dual space $\mathcal{H}^*$ - of a Hilbert space.

THEOREM 2.4.1 *(***Riesz lemma***)*
Let $\mathcal{H}$ be a Hilbert space.
(a) If $y \in \mathcal{H}$, the equation

$$\phi_y(x) = \langle x, y \rangle \tag{2.4.9}$$

defines a bounded linear functional $\phi_y \in \mathcal{H}^*$; and furthermore, $||\phi_y||_{\mathcal{H}^*} = ||y||_{\mathcal{H}}$.
(b) Conversely, if $\phi \in \mathcal{H}^*$, there exists a unique element $y \in \mathcal{H}$ such that $\phi = \phi_y$.

Proof : (a) Linearity of the map ϕ_y is obvious, while the Cauchy-Schwarz inequality shows that ϕ_y is bounded and that $||\phi_y|| \leq ||y||$. Since $\phi_y(y) = ||y||^2$, it easily follows that we actually have equality in the preceding inequality.

(b) Suppose conversely that $\phi \in \mathcal{H}^*$. Let $\mathcal{M} = ker\ \phi$. Since $||\phi_{y_1} - \phi_{y_2}|| = ||y_1 - y_2|| \ \forall\ y_1, y_2 \in \mathcal{H}$, the uniqueness assertion is obvious; we only have to prove existence. Since existence is clear if $\phi = 0$, we may assume that $\phi \neq 0$, or i.e., that $\mathcal{M} \neq \mathcal{H}$, or equivalently that $\mathcal{M}^\perp \neq 0$.

Notice that ϕ maps $\mathcal{M}^\perp$ 1-1 into $\mathbb{C}$; since $\mathcal{M}^\perp \neq 0$, it follows that $\mathcal{M}^\perp$ is one-dimensional. Let z be a unit vector in $\mathcal{M}^\perp$. The y that we seek - assuming it exists - must clearly be an element of $\mathcal{M}^\perp$ (since $\phi(x) = 0 \ \forall x \in \mathcal{M}$). Thus, we must have $y = \alpha z$ for some uniquely determined scalar $0 \neq \alpha \in \mathbb{C}$. With y defined thus, we find that $\phi_y(z) = \overline{\alpha}$; hence we must have $\alpha = \overline{\phi(z)}$. Since any element in $\mathcal{H}$ is uniquely expressible in the form $x + \gamma z$ for some $x \in \mathcal{M}, \gamma \in \mathbb{C}$, we find easily that we do indeed have $\phi = \phi_{\overline{\phi(z)}z}$. □

It must be noted that the mapping $y \mapsto \phi_y$ is not quite an isometric isomorphism of Banach spaces; it is not a linear map, since $\phi_{\alpha y} = \overline{\alpha}\phi_y$; it is only 'conjugate-linear'. The Banach space $\mathcal{H}^*$ is actually a Hilbert space if we define

$$\langle \phi_y, \phi_z \rangle = \langle z, y \rangle \ ;$$

that this equation satisfies the requirements of an inner product are an easy consequence of the Riesz lemma (and the already stated conjugate-linearity of the mapping $y \mapsto \phi_y$); that this inner product actually gives rise to the norm on $\mathcal{H}^*$ is a consequence of the fact that $||y|| = ||\phi_y||$.

Exercise 2.4.2 *(1) Where is the completeness of* $\mathcal{H}$ *used in the proof of the Riesz lemma; more precisely, what can you say about* X^* *if you only know that* X *is a (not necessarily complete) inner product space? (Hint: Consider the completion of* X*.)*

(2) If $T \in \mathcal{L}(\mathcal{H}, \mathcal{K})$*, where* $\mathcal{H}, \mathcal{K}$ *are Hilbert spaces, prove that*

$$||T|| = \sup\{|\langle Tx, y\rangle| : x \in \mathcal{H}, y \in \mathcal{K}, \ ||x|| \leq 1, ||y|| \leq 1\} \ .$$

(3) If $T \in \mathcal{L}(\mathcal{H})$ *and if* $\langle Tx, x\rangle = 0 \ \forall \ x \in \mathcal{H}$*, show that* $T = 0$*. (Hint: Use the Polarisation identity - see Remark 2.1.5(2). Notice that this proof relies on the fact that the underlying field of scalars is* $\mathbb{C}$ *- since the Polarisation identity does. In fact, the content of this exercise is false for 'real' Hilbert spaces; for instance, consider rotation by* 90^o *in* R^2*.)*

We now come to a very useful consequence of the Riesz representation theorem; in order to state this result, we need some terminology.

Definition 2.4.3 *Let X, Y be inner product spaces.*

(a) A mapping $B : X \times Y \to \mathbb{C}$ is said to be a **sesquilinear form** *on $X \times Y$ if it satisfies the following conditions:*

(i) for each fixed $y \in Y$, the mapping $X \ni x \mapsto B(x, y) \in \mathbb{C}$ is linear; and

(ii) for each fixed $x \in X$, the mapping $Y \ni y \mapsto B(x, y) \in \mathbb{C}$ is conjugate-linear;

in other words, for arbitrary $x_1, x_2 \in X, y_1, y_2 \in Y$ and $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{C}$, we have

$$B\left(\sum_{i=1}^{2} \alpha_i x_i, \sum_{j=1}^{2} \beta_j y_j\right) = \sum_{i,j=1}^{2} \alpha_i \overline{\beta_j} B(x_i, y_j) .$$

(b) a sesquilinear form $B : X \times Y \to \mathbb{C}$ is said to be **bounded** *if there exists a finite constant $0 < K < \infty$ such that*

$$|B(x, y)| \leq K \cdot ||x|| \cdot ||y|| , \ \forall \ x \in X, y \in Y.$$

Examples of bounded sesquilinear maps are obtined from bounded linear operators, thus: if $T \in \mathcal{L}(X, Y)$, then the assignment

$$(x, y) \mapsto \langle Tx, y \rangle$$

defines a bounded sesquilinear map. The useful fact, which we now establish, is that for Hilbert spaces, this is the only way to construct bounded sesquilinear maps.

Proposition 2.4.4 *Suppose $\mathcal{H}$ and $\mathcal{K}$ are Hilbert spaces. If $B : \mathcal{H} \times \mathcal{K} \to \mathbb{C}$ is a bounded sesquilinear map, then there exists a unique bounded operator $T : \mathcal{H} \to \mathcal{K}$ such that*

$$B(x, y) = \langle Tx, y \rangle , \ \forall x \in \mathcal{H}, y \in \mathcal{K} .$$

Proof : Temporarily fix $x \in \mathcal{H}$. The hypothesis implies that the map

$$\mathcal{K} \ni y \mapsto \overline{B(x, y)}$$

defines a bounded linear functional on $\mathcal{K}$ with norm at most $K||x||$, where K and B are related as in Definition 2.4.3(b).

Deduce from the Riesz lemma that there exists a unique vector in $\mathcal{K}$ - call it Tx - such that

$$\overline{B(x,y)} = \langle y, Tx\rangle \ \forall \ y \in \mathcal{K}$$

and that, further, $||Tx|| \leq K||x||$.

The previous equation unambiguously defines a mapping $T : \mathcal{H} \to \mathcal{K}$ which is seen to be linear (as a consequence of the uniqueness assertion in the Riesz lemma). The final inequality of the previous paragraph guarantees the boundedness of T.

Finally, the uniqueness of T is a consequence of Exercise 2.4.2. $\square$

As in Exercise 2.4.2(1), the reader is urged to see to what extent the completeness of $\mathcal{H}$ and $\mathcal{K}$ are needed in the hypothesis of the preceding proposition. (In particular, she should be able to verify that the completeness of $\mathcal{H}$ is unnecessary.)

COROLLARY 2.4.5 *Let $T \in \mathcal{L}(\mathcal{H}, \mathcal{K})$, where $\mathcal{H}, \mathcal{K}$ are Hilbert spaces. Then there exists a unique bounded operator $T^* \in \mathcal{L}(\mathcal{K}, \mathcal{H})$ such that*

$$\langle T^*y, x\rangle = \langle y, Tx\rangle \ \forall \ x \in \mathcal{H}, y \in \mathcal{K} \ . \tag{2.4.10}$$

Proof : Consider the (obviously) bounded sesquilinear form $B : \mathcal{K} \times \mathcal{H} \to \mathbb{C}$ defined by $B(y,x) = \langle y, Tx\rangle$ and appeal to Proposition 2.4.4 to lay hands on the operator T^*. $\square$

DEFINITION 2.4.6 *If T, T^* are as in Corollary 2.4.5, the operator T^* is called the* **adjoint** *of the operator T.*

The next proposition lists various elementary properties of the process of adjunction - i.e., the passage from an operator to its adjoint.

PROPOSITION 2.4.7 *Let $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3$ be Hilbert spaces, and suppose $T, T_1 \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, $S \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_3)$ and suppose $\alpha \in \mathbb{C}$. Then,*

(a) $(\alpha T + T_1)^* = \overline{\alpha} T^* + T_1^*$;

(b) $(T^*)^* = T$;

(c) $1_{\mathcal{H}}^* = 1_{\mathcal{H}}$, *where $1_{\mathcal{H}}$ denotes the identity operator on $\mathcal{H}$;*

(d) $(ST)^* = T^*S^*$;

(e) $||T|| = ||T^*||$; *and*

(f) $||T||^2 = ||T^*T||$.

Proof : Most of these assertions are verified by using the fact that the adjoint operator T^* is characterised by the equation 2.4.10.

(a) For arbitrary $x \in \mathcal{H}_1, y \in \mathcal{H}_2$, note that

$$\begin{aligned}
\langle y, (\alpha T + T_1)x\rangle &= \overline{\alpha}\langle y, Tx\rangle + \langle y, T_1 x\rangle \\
&= \overline{\alpha}\langle T^* y, x\rangle + \langle T_1^* y, x\rangle \\
&= \langle (\overline{\alpha}T^* + T_1^*)y, x\rangle
\end{aligned}$$

and deduce that $(\alpha T + T_1)^* = \overline{\alpha}T^* + T_1^*$.

The proofs of (b), (c) and (d) are entirely similar to the one given above for (i) and the reader is urged to verify the details.

(e) This is an immediate consequence of equation 2.4.10 and (two applications, once to T and once to T^*, of) Exercise 2.4.2(2).

(f) If $x \in \mathcal{H}_1$, note that

$$\begin{aligned}
||Tx||^2 &= \langle Tx, Tx\rangle \\
&= \langle T^*Tx, x\rangle \\
&\leq ||T^*Tx|| \cdot ||x|| \quad (\text{ by Cauchy-Schwarz}) \\
&\leq ||T^*T|| \cdot ||x||^2 \ ;
\end{aligned}$$

since x was arbitrary, deduce that $||T||^2 \leq ||T^*T||$. The reverse inequality follows at once from (e) above and Exercise 1.3.4(3). □

In the remainder of this section, we shall discuss some important classes of bounded operators.

DEFINITION 2.4.8 *An operator $T \in \mathcal{L}(\mathcal{H})$, where $\mathcal{H}$ is a Hilbert space, is said to be* **self-adjoint** *if $T = T^*$.*

PROPOSITION 2.4.9 *Let $T \in \mathcal{L}(\mathcal{H})$, where $\mathcal{H}$ is a Hilbert space.*

(a) T is self-adjoint if and only if $\langle Tx, x\rangle \in \mathbb{R}\ \forall x \in \mathcal{H}$;

(b) there exists a unique pair of self-adjoint operators $T_i, i = 1, 2$ such that $T = T_1 + iT_2$; this decomposition is given by $T_1 = \frac{T+T^*}{2}$, $T_2 = \frac{T-T^*}{2i}$, and it is customary to refer to T_1 and T_2 as the real and imaginary parts of T and to write $Re\ T = T_1$, $Im\ T = T_2$.

Proof : (a) If $T \in \mathcal{L}(\mathcal{H})$, we see that, for arbitrary x,

$$\langle Tx, x\rangle = \langle x, T^*x\rangle = \overline{\langle T^*x, x\rangle}\ ,$$

and consequently, $\langle Tx, x\rangle \in \mathbb{R} \Leftrightarrow \langle Tx, x\rangle = \langle T^*x, x\rangle$.

Since T is determined - see Exercise 2.4.2(3) - by its quadratic form (i.e., the map $\mathcal{H} \ni x \mapsto \langle Tx, x\rangle \in \mathbb{C}$), the proof of (a) is complete.

(b) It is clear that the equations

$$Re\ T = \frac{T + T^*}{2}, \ Im\ T = \frac{T - T^*}{2i}$$

define self-adjoint operators such that $T = Re\ T + i\ Im\ T$.

Conversely, if $T = T_1 + iT_2$ is such a decomposition, then note that $T^* = T_1 - iT_2$ and conclude that $T_1 = Re\ T,\ T_2 = Im\ T$. □

Notice, in view of (a) and (b) of the preceding proposition, that the quadratic forms corresponding to the real and imaginary parts of an operator are the real and imaginary parts of the quadratic form of that operator.

Self-adjoint operators are the building blocks of all operators, and they are by far the most important subclass of all bounded operators on a Hilbert space. However, in order to see their structure and usefulness, we will have to wait until after we have proved the fundamental spectral theorem - which will allow us to handle self-adjoint operators with exactly the same facility that we have when handling real-valued functions.

Nevertheless, we have already seen one special class of self-adjoint operators, as shown by the next result.

PROPOSITION 2.4.10 *Let $P \in \mathcal{L}(\mathcal{H})$. Then the following conditions are equivalent:*

(i) $P = P_{\mathcal{M}}$ is the orthogonal projection onto some closed subspace $\mathcal{M} \subset \mathcal{H}$;

(ii) $P = P^2 = P^*$.

Proof : $(i) \Rightarrow (ii)$: If $P = P_{\mathcal{M}}$, the definition of an orthogonal projection shows that $P = P^2$; the self-adjointness of P follows from Theorem 2.3.12(4) and Proposition 2.4.9(a).

$(ii) \Rightarrow (i)$: Suppose (ii) is satisfied; let $\mathcal{M} = ran\ P$, and note that

$$\begin{aligned} x \in \mathcal{M} &\Rightarrow \exists y \in \mathcal{H} \text{ such that } x = Py \\ &\Rightarrow Px = P^2y = Py = x\ ; \end{aligned} \qquad (2.4.11)$$

on the other hand, note that

$$\begin{aligned} y \in \mathcal{M}^\perp & \Leftrightarrow \langle y, Pz \rangle = 0 \ \forall z \in \mathcal{H} \\ & \Leftrightarrow \langle Py, z \rangle = 0 \ \forall z \in \mathcal{H} \quad (\text{since } P = P^*) \\ & \Leftrightarrow Py \ = \ 0 \ ; \end{aligned} \tag{2.4.12}$$

hence, if $z \in \mathcal{H}$ and $x = P_{\mathcal{M}} z, \ y = P_{\mathcal{M}^\perp} z$, we find from equations 2.4.11 and 2.4.12 that $Pz = Px + Py = x = P_{\mathcal{M}} z$. □

The next two propositions identify two important classes of operators between Hilbert spaces.

Proposition 2.4.11 *Let $\mathcal{H}, \mathcal{K}$ be Hilbert spaces; the following conditions on an operator $U \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ are equivalent:*

(i) if $\{e_i : i \in I\}$ is any orthonormal set in $\mathcal{H}$, then also $\{Ue_i : i \in I\}$ is an orthonormal set in $\mathcal{K}$;

(ii) there exists an orthonormal basis $\{e_i : i \in I\}$ for $\mathcal{H}$ such that $\{Ue_i : i \in I\}$ is an orthonormal set in $\mathcal{K}$;

(iii) $\langle Ux, Uy \rangle \ = \ \langle x, y \rangle \ \forall \ x, y \in \mathcal{H}$;

(iv) $||Ux|| \ = \ ||x|| \ \forall \ x \in \mathcal{H}$;

(v) $U^*U \ = \ 1_{\mathcal{H}}$.

An operator satisfying these equivalent conditions is called an **isometry**.

Proof : $(i) \Rightarrow (ii)$: There exists an orthonormal basis for $\mathcal{H}$.

$(ii) \Rightarrow (iii)$: If $x, y \in \mathcal{H}$ and if $\{e_i : i \in I\}$ is as in (ii), then

$$\begin{aligned} \langle Ux, Uy \rangle &= \langle U \left(\sum_{i \in I} \langle x, e_i \rangle e_i \right) \ , \ U \left(\sum_{j \in I} \langle y, e_j \rangle e_j \right) \ \rangle \\ &= \sum_{i,j \in I} \langle x, e_i \rangle \langle e_j, y \rangle \langle Ue_i, Ue_j \rangle \\ &= \sum_{i \in I} \langle x, e_i \rangle \langle e_i, y \rangle \\ &= \langle x, y \rangle \ . \end{aligned}$$

$(iii) \Rightarrow (iv)$: Put $y = x$.

$(iv) \Rightarrow (v)$: If $x \in \mathcal{H}$, note that

$$\langle U^*Ux, x \rangle \ = \ ||Ux||^2 \ = \ ||x||^2 \ = \ \langle 1_{\mathcal{H}} x, x \rangle \ ,$$

and appeal to the fact that a bounded operator is determined by its quadratic form - see Exercise 2.4.2(3).

$(v) \Rightarrow (i)$: If $\{e_i : i \in I\}$ is any orthonormal set in $\mathcal{H}$, then

$$\langle Ue_i, Ue_j\rangle \;=\; \langle U^*Ue_i, e_j\rangle \;=\; \langle e_i, e_j\rangle \;=\; \delta_{ij} \ .$$

□

PROPOSITION 2.4.12 *The following conditions on an isometry $U \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ are equivalent:*

(i) if $\{e_i : i \in I\}$ is any orthonormal basis for $\mathcal{H}$, then $\{Ue_i : i \in I\}$ is an orthonormal basis for $\mathcal{K}$;

(ii) there exists an orthonormal set $\{e_i : i \in I\}$ in $\mathcal{H}$ such that $\{Ue_i : i \in I\}$ is an orthonormal basis for $\mathcal{K}$;

(iii) $UU^* = 1_{\mathcal{K}}$;

(iv) U is invertible;

(v) U maps $\mathcal{H}$ onto $\mathcal{K}$.

An isometry which satisfies the above equivalent conditions is said to be **unitary**.

Proof : $(i) \Rightarrow (ii)$: Obvious.

$(ii) \Rightarrow (iii)$: If $\{e_i : i \in I\}$ is as in (ii), and if $x \in \mathcal{K}$, observe that

$$\begin{aligned} UU^*x &= UU^*(\sum_{i\in I}\langle x, Ue_i\rangle Ue_i) \\ &= \sum_{i\in I}\langle x, Ue_i\rangle UU^*Ue_i \\ &= \sum_{i\in I}\langle x, Ue_i\rangle Ue_i \quad \text{(since } U \text{ is an isometry)} \\ &= x \ . \end{aligned}$$

$(iii) \Rightarrow (iv)$: The assumption that U is an isometry, in conjunction with the hypothesis (iii), says that $U^* = U^{-1}$.

$(iv) \Rightarrow (v)$: Obvious.

$(v) \Rightarrow (i)$: If $\{e_i : i \in I\}$ is an orthonormal basis for $\mathcal{H}$, then $\{Ue_i : i \in I\}$ is an orthonormal set in $\mathcal{H}$, since U is isometric. Now, if $z \in \mathcal{K}$, pick $x \in \mathcal{H}$ such that $z = Ux$, and observe that

$$||z||^2 \;=\; ||Ux||^2$$

$$\begin{aligned} &= ||x||^2 \\ &= \sum_{i \in I} |\langle x, e_i \rangle|^2 \\ &= \sum_{i \in I} |\langle z, Ue_i \rangle|^2 \ , \end{aligned}$$

and since z was arbitrary, this shows that $\{Ue_i : i \in I\}$ is an orthonormal basis for $\mathcal{K}$. □

Thus, unitary operators are the natural isomorphisms in the context of Hilbert spaces. The collection of unitary operators from $\mathcal{H}$ to $\mathcal{K}$ will be denoted by $\mathcal{U}(\mathcal{H}, \mathcal{K})$; when $\mathcal{H} = \mathcal{K}$, we shall write $\mathcal{U}(\mathcal{H}) = \mathcal{U}(\mathcal{H}, \mathcal{H})$. We list some elementary properties of unitary and isometric operators in the next exercise.

EXERCISE 2.4.13 *(1) Suppose that $\mathcal{H}$ and $\mathcal{K}$ are Hilbert spaces and suppose $\{e_i : i \in I\}$ (resp., $\{f_i : i \in I\}$) is an orthonormal basis (resp., orthonormal set) in $\mathcal{H}$ (resp., $\mathcal{K}$), for some index set I. Show that:*

(a) dim $\mathcal{H} \leq$ dim $\mathcal{K}$; and

(b) there exists a unique isometry $U \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ such that $Ue_i = f_i \ \forall i \in I$.

(2) Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces. Show that:

(a) there exists an isometry $U \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ if and only if dim $\mathcal{H} \leq$ dim $\mathcal{K}$;

(b) there exists a unitary $U \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ if and only if dim $\mathcal{H}$ = dim $\mathcal{K}$.

(3) Show that $\mathcal{U}(\mathcal{H})$ is a group under multiplication, which is a (norm-) closed subset of the Banach space $\mathcal{L}(\mathcal{H})$.

(4) Suppose $U \in \mathcal{U}(\mathcal{H}, \mathcal{K})$; show that the equation

$$\mathcal{L}(\mathcal{H}) \ni T \stackrel{ad\ U}{\mapsto} UTU^* \in \mathcal{L}(\mathcal{K}) \tag{2.4.13}$$

*defines a mapping $(ad\ U) : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{K})$ which is an 'isometric isomorphism of Banach *-algebras', meaning that:*

(a) ad U is an isometric isomorphism of Banach spaces: i.e., ad U is a linear mapping which is 1-1, onto, and is norm-preserving; (Hint: verify that it is linear and preserves norm and that an inverse is given by ad U^*.)

(b) ad U is a product-preserving map between Banach algebras; i.e., $(ad\ U)(T_1T_2) = (\ (ad\ U)(T_1)\ )(\ ad\ U)(T_2)\ )$, *for all* $T_1, T_2 \in \mathcal{L}(\mathcal{H})$*;*

*(c) ad U is a *-preserving map between C^*-algebras; i.e.,* $(\ (ad\ U)(T)\ )^* = (ad\ U)(T^*)$ *for all* $T \in \mathcal{L}(\mathcal{H})$.

(5) Show that the map $U \mapsto (ad\ U)$ *is a homomorphism from the group* $\mathcal{U}(\mathcal{H})$ *into the group Aut* $\mathcal{L}(\mathcal{H})$ *of all automorphisms (= isometric isomorphisms of the Banach *-algebra* $\mathcal{L}(\mathcal{H})$ *onto itself); further, verify that if* $U_n \to U$ *in* $\mathcal{U}(\mathcal{H}, \mathcal{K})$*, then* $(ad\ U_n)(T) \to (ad\ U)(T)$ *in* $\mathcal{L}(\mathcal{K})$ *for all* $T \in \mathcal{L}(\mathcal{H})$.

A unitary operator between Hilbert spaces should be viewed as 'implementing an inessential variation'; thus, if $U \in \mathcal{U}(\mathcal{H}, \mathcal{K})$ and if $T \in \mathcal{L}(\mathcal{H})$, then the operator $UTU^* \in \mathcal{L}(\mathcal{K})$ should be thought of as being 'essentially the same as T', except that it is probably being viewed from a different observer's perspective. All this is made precise in the following definition.

DEFINITION 2.4.14 *Two operators* $T \in \mathcal{L}(H)$ *and* $S \in \mathcal{L}(K)$ *(on two possibly different Hilbert spaces) are said to be* **unitarily equivalent** *if there exists a unitary operator* $U \in \mathcal{U}(\mathcal{H}, \mathcal{K})$ *such that* $S = UTU^*$.

We conclude this section with a discussion of some examples of isometric operators, which will illustrate the preceding notions quite nicely.

EXAMPLE 2.4.15 To start with, notice that if $\mathcal{H}$ is a finite-dimensional Hilbert space, then an isometry $U \in \mathcal{L}(\mathcal{H})$ is necessarily unitary. (Prove this!) Hence, the notion of non-unitary isometries of a Hilbert space into itself makes sense only in infinite-dimensional Hilbert spaces. We discuss some examples of a non-unitary isometry in a separable Hilbert space.

(1) Let $\mathcal{H} = \ell^2\ (= \ell^2(\mathbb{N})\ )$. Let $\{e_n : n \in \mathbb{N}\}$ denote the standard orthonormal basis of $\mathcal{H}$ (consisting of sequences with a 1 in one co-ordinate and 0 in all other co-ordinates). In view of Exercise 2.4.13(1)(b), there exists a unique isometry $S \in \mathcal{L}(H)$ such that $Se_n = e_{n+1}\ \forall n \in \mathbb{N}$; equivalently, we have

$$S(\alpha_1, \alpha_2, \cdots) = (0, \alpha_1, \alpha_2, \cdots).$$

For obvious reasons, this operator is referred to as a 'shift' operator; in order to distinguish it from a near realtive, we shall refer to it as the **unilateral shift**. It should be clear that S is an isometry whose range is the proper subspace $\mathcal{M} = \{e_1\}^{\perp}$, and consequently, S is not unitary.

A minor computation shows that the adjoint S^* is the 'backward shift':

$$S^*(\alpha_1, \alpha_2, \cdots) = (\alpha_2, \alpha_3, \cdots)$$

and that $SS^* = P_{\mathcal{M}}$ (which is another way of seeing that S is not unitary). Thus S^* is a left-inverse, but not a right-inverse, for S. (This, of course, is typical of a non-unitary isometry.)

Further - as is true for any non-unitary isometry - each power $S^n, n \geq 1$, is a non-unitary isometry.

(2) The 'near-relative' of the unilateral shift, which was referred to earlier, is the so-called **bilateral shift**, which is defined as follows: consider the Hilbert space $\mathcal{H} = \ell^2(\mathbb{Z})$ with its standard basis $\{e_n : n \in \mathbb{Z}\}$ for $\mathcal{H}$. The bilateral shift is the unique isometry B on $\mathcal{H}$ such that $Be_n = e_{n+1} \; \forall n \in \mathbb{Z}$. This time, however, since B maps the standard basis onto itself, we find that B is unitary. The reason for the terminology 'bilateral shift' is this: denote a typical element of $\mathcal{H}$ as a 'bilateral' sequence (or a sequence extending to infinity in both directions); in order to keep things straight, let us underline the 0-th co-ordinate of such a sequence; thus, if $x = \sum_{n=-\infty}^{\infty} \alpha_n e_n$, then we write $x = (\cdots, \alpha_{-1}, \underline{\alpha_0}, \alpha_1, \cdots)$; we then find that

$$B(\cdots, \alpha_{-1}, \underline{\alpha_0}, \alpha_1, \cdots) = (\cdots, \alpha_{-2}, \underline{\alpha_{-1}}, \alpha_0, \cdots) .$$

(3) Consider the Hilbert space $\mathcal{H} = L^2([0,1], m)$ (where, of course, m denotes 'Lebesgue measure') - see Remark 2.3.7(2) - and let $\{e_n : n \in \mathbb{Z}$ denote the exponential basis of this Hilbert space. Notice that $|e_n(x)|$ is identically equal to 1, and conclude that the operator defined by

$$(Wf)(x) = e_1(x)f(x) \; \forall f \in \mathcal{H}$$

is necessarily isometric; it should be clear that this is actually unitary, since its inverse is given by the operator of multiplication by e_{-1}.

It is easily seen that $We_n = e_{n+1}$ $\forall n \in \mathbb{Z}$. If $U : \ell^2(\mathbb{Z}) \to \mathcal{H}$ is the unique unitary operator such that U maps the n-th standard basis vector to e_n, for each $n \in \mathbb{Z}$, it follows easily that $W = UBU^*$. Thus, the operator W of this example is unitarily equivalent to the bilateral shift (of the previous example).

More is true; let $\mathcal{M}$ denote the closed subspace $\mathcal{M} = [\{e_n : n \geq 1\}]$; then $\mathcal{M}$ is invariant under W - meaning that $W(\mathcal{M}) \subset \mathcal{M}$; and it should be clear that the restricted operator $W|_{\mathcal{M}} \in \mathcal{L}(\mathcal{M})$ is unitarily equivalent to the unilateral shift.

(4) More generally, if $(X, \mathcal{B}, \mu)$ is any measure space - see the appendix (§A.5) - and if $\phi : X \to \mathbb{C}$ is any measurable function such that $|\phi| = 1$ $\mu - a.e.$, then the equation

$$M_\phi f \ = \ \phi f \ , f \in L^2(X, \mathcal{B}, \mu)$$

defines a unitary operator on $L^2(X, \mathcal{B}, \mu)$ (with inverse given by $M_{\overline{\phi}}$). □

2.5 Strong and weak convergence

This section is devoted to a discussion of two (vector space) topologies on the space $\mathcal{L}(\mathcal{H}, \mathcal{K})$ of bounded operators between Hilbert spaces.

DEFINITION 2.5.1 *(1) For each $x \in \mathcal{H}$, consider the seminorm p_x defined on $\mathcal{L}(\mathcal{H}, \mathcal{K})$ by $p_x(T) = ||Tx||$; the* **strong operator topology** *is the topology on $\mathcal{L}(\mathcal{H}, \mathcal{K})$ defined by the family $\{p_x : x \in \mathcal{H}\}$ of seminorms (see Remark 1.6.5).*

(2) For each $x \in \mathcal{H}, y \in \mathcal{K}$, consider the seminorm $p_{x,y}$ defined on $\mathcal{L}(\mathcal{H}, \mathcal{K})$ by $p_{x,y}(T) = |\langle Tx, y\rangle|$; the **weak operator topology** *is the topology on $\mathcal{L}(\mathcal{H}, \mathcal{K})$ defined by the family $\{p_{x,y} : x \in \mathcal{H}, y \in \mathcal{K}\}$ of seminorms (see Remark 1.6.5).*

Thus, it should be clear that if $\{T_i : i \in I\}$ is a net in $\mathcal{L}(\mathcal{H}, \mathcal{K})$, then the net converges to $T \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ with respect to the strong operator topology precisely when $||T_i x - Tx|| \to 0$ $\forall x \in \mathcal{H}$, i.e., precisely when the net $\{T_i x : i \in I\}$ converges to Tx with respect to the strong (i.e., norm) topology on $\mathcal{K}$, for every $x \in \mathcal{H}$. Likewise, the net $\{T_i\}$ converges to T with respect to the weak operator topology if and only if $\langle (T_i - T)x, y\rangle \to 0$ for all $x \in \mathcal{H}, y \in \mathcal{K}$, i.e., if and only if the net $\{T_i x\}$ converges to Tx with respect to the weak topology on $\mathcal{K}$, for every $x \in \mathcal{H}$.

At the risk of abusing our normal convention for use of the expressions 'strong convergence' and 'weak convergence' (which we adopted earlier for general Banach spaces), in spite of the fact that $\mathcal{L}(\mathcal{H}, \mathcal{K})$ is itself a Banach space, we adopt the following convention for operators: we shall say that a net $\{T_i\}$ converges to T in $\mathcal{L}(\mathcal{H}, \mathcal{K})$ (i) uniformly, (ii) strongly, or (iii) weakly, if it is the case that we have convergence with respect to the (i) norm topology, (ii) strong operator topology, or (iii) weak operator topology, respectively, on $\mathcal{L}(\mathcal{H}, \mathcal{K})$.

Before discussing some examples, we shall establish a simple and useful lemma for deciding strong or weak convergence of a net of operators. (Since the strong and weak operator topologies are vector space topologies, a net $\{T_i\}$ converges strongly or weakly to T precisely when $\{T_i - T\}$ converges strongly or weakly to 0; hence we will state our criteria for nets which converge to 0.)

LEMMA 2.5.2 *Suppose* $\{T_i : i \in I\}$ *is a net in* $\mathcal{L}(\mathcal{H}, \mathcal{K})$ *which is uniformly bounded - i.e.,* $\sup\{||T_i|| : i \in I\} = K < \infty$. *Let* $\mathcal{S}_1$ *(resp.,* $\mathcal{S}_2$*) be a* **total** *set in* $\mathcal{H}$ *(resp.,* $\mathcal{K}$*) - i.e., the set of finite linear combinations of elements in* $\mathcal{S}_1$ *(resp.,* $\mathcal{S}_2$*) is dense in* $\mathcal{H}$ *(resp.,* $\mathcal{K}$*).*

(a) $\{T_i\}$ *converges strongly to 0 if and only if* $lim_i ||T_i x|| = 0$ *for all* $x \in \mathcal{S}_1$*; and*

(b) $\{T_i\}$ *converges weakly to 0 if and only if* $lim_i \langle T_i x, y\rangle = 0$ *for all* $x \in \mathcal{S}_1, y \in \mathcal{S}_2$.

Proof : Since the proofs of the two assertions are almost identical, we only prove one of them, namely (b). Let $x \in \mathcal{H}, y \in \mathcal{K}$ be arbitrary, and suppose $\epsilon > 0$ is given. We assume, without loss of generality, that $\epsilon < 1$. Let $M = 1 + max\{||x||, ||y||\}$, and $K = 1 + sup_i ||T_i||$; (the 1 in these two definitions is for the sole reason of ensuring that K and M are not zero, and we need not worry about dividing by zero).

The assumed totality of the sets $\mathcal{S}_i, i = 1, 2$, implies the existence of vectors $x' = \sum_{k=1}^{m} \alpha_k x_k, y' = \sum_{j=1}^{n} \beta_j y_j$, with the property that $x_k \in \mathcal{S}_1, y_j \in \mathcal{S}_2 \ \forall k, j$ and such that $||x - x'|| < \frac{\epsilon}{3KM}$

and $||y - y'|| < \frac{\epsilon}{3KM}$. Since $\epsilon < 1 \leq K, M$, it follows that $||x'|| \leq ||x|| + 1 \leq M$; similarly, also $||y'|| \leq M$. Let $N = 1 + \max\{|\alpha_k|, |\beta_j| : 1 \leq k \leq m, 1 \leq j \leq n\}$.

Since I is a directed set, the assumption that $lim_i \langle T_i x_k, y_j \rangle = 0 \; \forall j, k$ implies that we can find an index $i_0 \in I$ with the property that $|\langle T_i x_k, y_j \rangle| < \frac{\epsilon}{3N^2 mn}$ whenever $i \geq i_0, 1 \leq k \leq m, 1 \leq j \leq n$. It then follows that for all $i \geq i_0$, we have:

$$\begin{aligned} |\langle T_i x, y \rangle| &\leq |\langle T_i x, y - y' \rangle| \; + \; |\langle T_i(x - x'), y' \rangle| \; + \; |\langle T_i x', y' \rangle| \\ &\leq 2KM \frac{\epsilon}{3KM} \; + \; \sum_{k=1}^{m} \sum_{j=1}^{n} |\alpha_k \beta_j| \cdot |\langle T_i x_k, y_j \rangle| \\ &\leq \frac{2\epsilon}{3} \; + \; N^2 mn \frac{\epsilon}{3N^2 mn} \\ &= \epsilon \; , \end{aligned}$$

and the proof is complete. □

EXAMPLE 2.5.3 (1) Let $S \in \mathcal{L}(\ell^2)$ be the unilateral shift - see Example 2.4.15(1). Then the sequence $\{(S^*)^n : n = 1, 2, \cdots\}$ converges strongly, and hence also weakly, to 0. (Reason: the standard basis $\{e_m : m \in \mathbb{N}\}$ is total in ℓ^2, the sequence $\{(S^*)^n : n = 1, 2, \cdots\}$ is uniformly bounded, and $(S^*)^n e_m = 0 \; \forall n > m$. On the other hand, $\{S^n = ((S^*)^n)^* : n = 1, 2, \cdots\}$ is a sequence of isometries, and hence certainly does not converge strongly to 0. Thus, the adjoint operation is not 'strongly continuous'.

(2) On the other hand, it is obvious from the definition that if $\{T_i : i \in I\}$ is a net which converges weakly to T in $\mathcal{L}(\mathcal{H}, \mathcal{K})$, then the net $\{T_i^* : i \in I\}$ converges weakly to T^* in $\mathcal{L}(\mathcal{K}, \mathcal{H})$. In particular, conclude from (1) above that both the sequences $\{(S^*)^n : n = 1, 2, \cdots\}$ and $\{S^n : n = 1, 2, \cdots\}$ converge weakly to 0, but the sequence $\{(S^*)^n S^n : n = 1, 2, \cdots\}$ (which is the constant sequence given by the identity operator) does not converge to 0; thus, multiplication is not 'jointly weakly continuous'.

(3) Multiplication is 'separately strongly (resp., weakly) continuous' - i.e., if $\{S_i\}$ is a net which converges strongly (resp., weakly) to S in $\mathcal{L}(H, \mathcal{K})$, and if $A \in \mathcal{L}(\mathcal{K}, \mathcal{K}_1), B \in \mathcal{L}(\mathcal{H}_1, \mathcal{H})$, then the net $\{AS_iB\}$ converges strongly (resp., weakly) to ASB in $\mathcal{L}(\mathcal{H}_1, \mathcal{K}_1)$. (For instance, the 'weak' assertion follows from the equation $\langle AS_iBx, y \rangle \; = \; \langle S_i(Bx), A^*y \rangle$.)

(4) Multiplication is 'jointly strongly continuous' if we restrict ourselves to uniformly bounded nets - i.e., if $\{S_i : i \in I\}$ (resp., $\{T_j : j \in J\}$) is a uniformly bounded net in $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ (resp., $\mathcal{L}(\mathcal{H}_2, \mathcal{H}_3)$ which converges strongly to S (resp., T), and if $K = I \times J$ is the directed set obtained by coordinate-wise ordering as in Example 2.2.4(4), then the net $\{T_j \circ S_i : (i,j) \in K\}$ converges strongly to $T \circ S$. (Reason: assume, by translating if necessary, that $S = 0$ and $T = 0$; (the assertion in (3) above is needed to justify this reduction); if $x \in \mathcal{H}_1$ and if $\epsilon > 0$, first pick $i_0 \in I$ such that $||S_i x|| < \frac{\epsilon}{M}$, $\forall i \geq i_0$, where $M = 1 + sup_j ||T_j||$; then pick an arbitrary $j_0 \in J$, and note that if $(i,j) \geq (i_0, j_0)$ in K, then $||T_j S_i x|| \leq M ||S_i x|| < \epsilon$.)

(5) The purpose of the following example is to show that the asserted joint strong continuity of multiplication is false if we drop the restriction of uniform boundedness. Let $\mathcal{N} = \{N \in \mathcal{L}(\mathcal{H}) : N^2 = 0\}$, where $\mathcal{H}$ is infinite-dimensional.

We assert that $\mathcal{N}$ is strongly dense in $\mathcal{L}(\mathcal{H})$. To see this, note that sets of the form $\{T \in \mathcal{L}(\mathcal{H}) : ||(T - T_0)x_i|| < \epsilon, \forall 1 \leq i \leq n\}$, where $T_0 \in \mathcal{L}(\mathcal{H})$, $\{x_1, \cdots, x_n\}$ is a linearly independent set in $\mathcal{H}$, $n \in \mathbb{N}$ and $\epsilon > 0$, constitute a base for the strong operator topology on $\mathcal{L}(\mathcal{H})$. Hence, in order to prove the assertion, we need to check that every set of the above form contains elements of $\mathcal{N}$. For this, pick vectors $y_1, \cdots, y_n$ such that $\{x_1, \cdots, x_n, y_1, \cdots, y_n\}$ is a linearly independent set in $\mathcal{H}$, and such that $||T_0 x_i - y_i|| < \epsilon \ \forall i$; now consider the operator T defined thus: $Tx_i = y_i, Ty_i = 0 \ \forall i$ and $Tz = 0$ whenever $z \perp \{x_1, \cdots, x_n, y_1, \cdots, y_n\}$; it is seen that $T \in \mathcal{N}$ and that T belongs to the open set exhibited above.

Since $\mathcal{N} \neq \mathcal{L}(\mathcal{H})$, the assertion (of the last paragraph) shows three things: (a) multiplication is not jointly strongly continuous; (b) if we choose a net in $\mathcal{N}$ which converges strongly to an operator which is not in $\mathcal{N}$, then such a net is not uniformly bounded (because of (4) above); thus, strongly convergent nets need not be uniformly bounded; and (c) since a strongly convergent sequence is necessarily uniformly bounded - see Exercise 1.5.17(3) - the strong operator topology cannot be described with 'just sequences'. □

We wish now to make certain statements concerning (arbitrary) direct sums.

PROPOSITION 2.5.4 *Suppose* $\{\mathcal{M}_i : i \in I\}$ *is a family of closed subspaces of a Hilbert space* $\mathcal{H}$*, which are pairwise orthogonal - i.e.,* $\mathcal{M}_i \perp \mathcal{M}_j$ *for* $i \neq j$*. Let* $P_i = P_{\mathcal{M}_i}$ *denote the orthogonal projection onto* $\mathcal{M}_i$*. Then the family* $\{P_i : i \in I\}$ *is unconditionally summable, with respect to the strong (and hence also the weak) operator topology, with sum given by* $P = P_{\mathcal{M}}$*, where* $\mathcal{M}$ *is the closed subspace spanned by* $\cup_{i\in I}\mathcal{M}_i$*.*

Conversely, suppose $\{\mathcal{M}_i : i \in I\}$ *is a family of closed subspaces, suppose* $P_i = P_{\mathcal{M}_i}$*, suppose the family* $\{P_i : i \in I\}$ *is unconditionally summable, with respect to the weak topology; and suppose the sum* P *is an orthogonal projection; then it is necessarily the case that* $\mathcal{M}_i \perp \mathcal{M}_j$ *for* $i \neq j$*, and we are in the situation described by the previous paragraph.*

Proof : For any finite subset $F \subset I$, let $P(F) = \sum_{i\in F} P_i$, and deduce from Exercise 2.3.15(5)(c) that $P(F)$ is the orthogonal projection onto the subspace $\mathcal{M}(F) = \sum_{i\in F}\mathcal{M}_i$.

Consider the mapping $\mathcal{F}(I) \ni F \mapsto P(F) \in \mathcal{L}(\mathcal{H})$ - where $\mathcal{F}(I)$ denotes the net of finite subsets of I, directed by inclusion, as in Example 2.2.4(2). The net $\{P(F) : F \in \mathcal{F}(I)\}$ is thus seen to be a net of projections and is hence a uniformly bounded net of operators.

Observe that $F_1 \leq F_2 \Rightarrow F_1 \subset F_2 \Rightarrow \mathcal{M}(F_1) \subset \mathcal{M}(F_2)$. Hence $\mathcal{M} = \overline{\cup_{F\in\mathcal{F}(I)}\mathcal{M}(F)}$. The set $\mathcal{S} = \cup_{F\in\mathcal{F}(I)}\mathcal{M}(F) \cup \mathcal{M}^\perp$ is clearly a total set in $\mathcal{H}$ and if $x \in \mathcal{S}$, then the net $\{P(F)x\}$ is eventually constant and equal to x or 0, according as $x \in \cup_{F\in\mathcal{F}(I)}\mathcal{M}(F)$ or $x \in \mathcal{M}^\perp$. It follows from Lemma 2.5.2 that the net $\{P(F)\}$ converges strongly to P.

In the converse direction, assume without loss of generality that all the $\mathcal{M}_i$ are non-zero subspaces, fix $i_0 \in I$ and let $x \in \mathcal{M}_{i_0}$ be a unit vector. Notice, then, that

$$\begin{aligned} 1 &\geq \langle Px, x\rangle \\ &= \sum_{i\in I}\langle P_i x, x\rangle \\ &\geq \langle P_{i_0}x, x\rangle \\ &= 1 \; ; \end{aligned}$$

conclude that $\langle P_i x, x \rangle = 0 \ \forall i \neq i_0$, i.e., $\mathcal{M}_{i_0} \perp \mathcal{M}_i$ for $i \neq i_0$, and the proof of the proposition is complete. $\square$

If $\mathcal{M}$ and $\mathcal{M}_i, i \in I$ are closed subspaces of a Hilbert space $\mathcal{H}$ as in Proposition 2.5.4, then the subspace $\mathcal{M}$ is called the (orthogonal) **direct sum** of the subspaces $\mathcal{M}_i, i \in I$, and we write

$$\mathcal{M} \ = \ \oplus_{i \in I} \ \mathcal{M}_i \ . \tag{2.5.14}$$

Conversely, if the equation 2.5.14 is ever written, it is always tacitly understood that the $\mathcal{M}_i$'s constitute a family of pairwise orthogonal closed subspaces of a Hilbert space, and that $\mathcal{M}$ is related to them as in Proposition 2.5.4.

There is an 'external' direct sum construction for an arbitrary family of Hilbert spaces, which is closely related to the notion discussed earlier, which is described in the following exercise.

EXERCISE 2.5.5 *Suppose $\{\mathcal{H}_i : i \in I\}$ is a family of Hilbert spaces; define*

$$\mathcal{H} = \{((x_i)) : x_i \in \mathcal{H}_i \ \forall i, \ \sum_{i \in I} ||x_i||^2 < \infty\} \ .$$

(a) Show that $\mathcal{H}$ is a Hilbert space with respect to the inner product defined by $\langle ((x_i)), ((y_i)) \rangle \ = \ \sum_{i \in I} \langle x_i, y_i \rangle$.

(b) Let $\mathcal{M}_j = \{((x_i)) \in \mathcal{H} : x_i = 0 \forall i \neq j\}$ for each $j \in I$; show that $\mathcal{M}_i, i \in I$ is a family of closed subspaces such that $\mathcal{H} \ = \ \oplus_{i \in I} \mathcal{M}_i$, and that $\mathcal{M}_i$ is naturally isomorphic to $\mathcal{H}_i$, for each $i \in I$.

In this situation, also, we shall write $\mathcal{H} = \oplus_{i \in I} \mathcal{H}_i$, and there should be no serious confusion as a result of using the same notation for both 'internal direct sums' (as in equation 2.5.14) and 'external direct sums' (as above).

There is one very useful fact concerning operators on a direct sum, and matrices of operators, which we hasten to point out.

PROPOSITION 2.5.6 *Suppose $\mathcal{H} = \oplus_{j \in J} \mathcal{H}_j$ and $\mathcal{K} = \oplus_{i \in I} \mathcal{K}_i$ are two direct sums of Hilbert spaces - which we regard as internal direct sums. Let P_i (resp., Q_j) denote the orthogonal projection of $\mathcal{K}$ (resp., $\mathcal{H}$) onto $\mathcal{K}_i$ (resp., $\mathcal{H}_j$).*

(1) If $T \in \mathcal{L}(\mathcal{H}, \mathcal{K})$, define $T^i_j x = P_i T x$, $\forall\ x \in \mathcal{H}_j$; then $T^i_j \in \mathcal{L}(\mathcal{H}_j, \mathcal{K}_i)\ \forall i \in I, j \in J$;

(2) The operator T is recaptured from the matrix $((T^i_j))$ of operators by the following prescription: if $x = ((x_j)) \in \mathcal{H}$, and if $Tx = ((y_i)) \in \mathcal{K}$, then $y_i = \sum_{j \in J} T^i_j x_j$, $\forall i \in I$, where the series is interpreted as the sum of an unconditionally summable family in $\mathcal{K}$ (with respect to the norm in $\mathcal{K}$); (thus, if we think of elements of $\mathcal{H}$ (resp., $\mathcal{K}$) as column vectors indexed by J (resp., I) - whose entries are themselves vectors in appropriate Hilbert spaces - then the action of the operator T is given by 'matrix multiplication' by the matrix $((T^i_j))$.)

(3) For finite subsets $I_0 \subset I, J_0 \subset J$, let $P(I_0) = \sum_{i \in I_0} P_i$ and $Q(J_0) = \sum_{j \in J_0} Q_j$. Then the net $\{P(I_0)TQ(J_0) : k = (I_0, J_0) \in K = \mathcal{F}(I) \times \mathcal{F}(J)\}$ - where K is ordered by coordinate-wise inclusion - converges strongly to T, and

$$\begin{aligned} \|T\| &= \lim_{(I_0,J_0) \in K} \|P(I_0)TQ(J_0)\| \\ &= \sup_{(I_0,J_0) \in K} \|P(I_0)TQ(J_0)\| . \end{aligned} \tag{2.5.15}$$

(4) Conversely, a matrix $((T^i_j))$, where $T^i_j \in \mathcal{L}(\mathcal{H}_j, \mathcal{K}_i)$, $\forall i \in I, j \in J$, defines a bounded operator $T \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ as in (2) above, precisely when the family $\{\|S\|\}$ - where S ranges over all those matrices which are obtained by replacing all the entries outside a finite number of rows and columns of the given matrix by 0 and leaving the remaining finite 'rectangular submatrix' of entries unchanged - is uniformly bounded.

(5) Suppose $\mathcal{M} = \oplus_{l \in L} \mathcal{M}_l$ is another (internal) direct sum of Hilbert spaces, and suppose $S \in \mathcal{L}(\mathcal{K}, \mathcal{M})$ has the matrix decomposition $((S^l_i))$ with respect to the given direct sum decompositions of $\mathcal{K}$ and $\mathcal{M}$, respectively, then the matrix decomposition for $S \circ T \in \mathcal{L}(\mathcal{H}, \mathcal{M})$ (with respect to the given direct sum decompositions of $\mathcal{H}$ and $\mathcal{M}$, respectively) is given by $(S \circ T)^l_j = \sum_{i \in I} S^l_i \circ T^i_j$, where this series is unconditionally summable with respect to the strong operator topology.

(6) The assignment of the matrix $((T^i_j))$ to the operator T is a linear one, and the matrix associated with T^* is given by $(T^*)^j_i = (T^i_j)^*$.

Proof : (1) Sice $\|P_i\| \leq 1$, it is clear theat $T^i_j \in \mathcal{L}(\mathcal{H}_j, \mathcal{K}_i)$ and that $\|T^i_j\| \leq \|T\|$ (see Exercise 1.3.4(3)).

(2) Since $\sum_{j\in J} Q_j = id_{\mathcal{H}}$ - where this series is interpreted as the sum of an unconditionally summable family, with respect to the strong operator topology in $\mathcal{L}(\mathcal{H})$ - it follows from 'separate strong continuity' of multiplication - see Example 2.5.3(3) - that

$$x \in \mathcal{H} \Rightarrow P_i T x = \sum_{j\in J} P_i T Q_j x = \sum_{j\in J} T^i_j x_j ,$$

as desired.

(3) Since the nets $\{P(I_0) : I_0 \in \mathcal{F}(I)\}$ and $\{Q(J_0) : J_0 \in \mathcal{F}(J)\}$ are uniformly bounded and converge strongly to $id_{\mathcal{K}}$ and $id_{\mathcal{H}}$ respectively, it follows from 'joint strong continuity of multiplication on uniformly bounded sets' - see Example 2.5.3(4) - that the net $\{P(I_0)TQ(J_0) : (I_0, J_0) \in \mathcal{F}(I) \times \mathcal{F}(J)\}$ converges strongly to T and that $||T|| \leq\ lim\ inf_{(I_0,J_0)}||P(I_0)TQ(J_0)||$. (The limit inferior and limit superior of a bounded net of real numbers is defined exactly as such limits of a bounded sequence of real numbers is defined - see equations A.5.16 and A.5.17, respectively, for these latter definitions.) Conversely, since $P(I_0)$ and $Q(J_0)$ are projections, it is clear that $||P(I_0)TQ(J_0)|| \leq ||T||$ for each (I_0, J_0); hence $||T|| \geq\ lim\ sup_{(I_0,J_0)}||P(I_0)TQ(J_0)||$, and the proof of (3) is complete.

(4) Under the hypothesis, for each $k = (I_0, J_0) \in K = \mathcal{F}(I) \times \mathcal{F}(J)$, define T_k to be the (clearly bounded linear) map from $\mathcal{H}$ to $\mathcal{K}$ defined by 'matrix multiplication' by the matrix which agrees with $((T^i_j))$ in rows and columns indexed by indices from I_0 and J_0 respectively, and has 0's elsewhere; then $\{T_k : k \in K\}$ is seen to be a uniformly bounded net of operators in $\mathcal{L}(\mathcal{H}, \mathcal{K})$; this net is strongly convergent. (*Reason:* $\mathcal{S} = \cup_{J_0\in\mathcal{F}(J)} Q(J_0)\mathcal{H}$ is a total set in $\mathcal{H}$, and the net $\{T_k x : k \in K\}$ is easily verified to converge whenever $x \in \mathcal{S}$; and we may appeal to Lemma 2.5.2.) If T denotes the strong limit of this net, it is easy to verify that this operator does indeed have the given matrix decomposition.

(5) This is proved exactly as was (2) - since the equation $\sum_{i\in I} P_i = id_{\mathcal{K}}$ implies that $\sum_{i\in I} SP_i T = ST$, both 'series' being interpreted in the strong operator topology.

(6) This is an easy verification, which we leave as an exercise for the reader. □

We explicitly spell out a special case of Proposition 2.5.6, because of its importance. Suppose $\mathcal{H}$ and $\mathcal{K}$ are separable Hilbert spaces, and suppose $\{e_j\}$ and $\{f_i\}$ are orthonormal bases for $\mathcal{H}$ and $\mathcal{K}$, respectively. Then, we have the setting of Proposition 2.5.6, if we set $\mathcal{H}_j = \mathbb{C}e_j$ and $\mathcal{K}_i = \mathbb{C}f_i$ respectively. Since $\mathcal{H}_j$ and $\mathcal{K}_i$ are 1-dimensional, any operator $S : \mathcal{H}_j \to \mathcal{K}_i$ determines and is determined by a uniquely defined scalar λ, via the requirement that $Se_j = \lambda f_i$. If we unravel what the proposition says in this special case, this is what we find:

For each i, j, define $t^i_j = \langle Te_j, f_i\rangle$; think of an element $x \in \mathcal{H}$ (resp., $y \in \mathcal{K}$) as a column vector $x = ((x_j = \langle x, e_j\rangle))$ (resp., $y = ((y_i = \langle y, f_i\rangle)))$; then the action of the operator is given by matrix-multiplication, thus: if $Tx = y$, then $y_i = \sum_j t^i_j x_j$ - where the series is unconditionally summable in $\mathbb{C}$.

In the special case when $\mathcal{H} = \mathcal{K}$ has finite dimension, say n, if we choose and fix one orthonormal basis $\{e_1, \cdots, e_n\}$ for $\mathcal{H}$ - thus we choose the f's to be the same as the e's - then the choice of any such (ordered) basis sets up a bijective correspondence between $\mathcal{L}(\mathcal{H})$ and $M_n(\mathbb{C})$ which is an 'isomorphism of *-algebras', where (i) we think of $\mathcal{L}(\mathcal{H})$ as being a *-algebra in the obvious fashion, and (ii) the *-algebra structure on $M_n(\mathbb{C})$ is defined thus: the vector operations, i.e., addition and scalar multiplication, are as in Example 1.1.3(2), the product of two $n \times n$ matrices is as in Exercise 1.3.1(7)(ii), and the 'adjoint' of a matrix is the 'complex conjugate transpose' - i.e., if A is the matrix with (i, j)-th entry given by a^i_j, then the adjoint is the matrix A^* with (i, j)-th entry given by $\overline{a^j_i}$.

This is why, for instance, a matrix $U \in M_n(\mathbb{C})$ is said to be unitary if $U^*U = I_n$ (the identity matrix of size n), and the set $U(n)$ of all such matrix has a natural structure of a compact group.

EXERCISE 2.5.7 *If $\mathcal{H}$ is a Hilbert space, and if $x, y \in \mathcal{H}$, define $T_{x,y} : \mathcal{H} \to \mathcal{H}$ by $T_{x,y}(z) = \langle z, y\rangle x$. Show that*

(a) $T_{x,y} \in \mathcal{L}(\mathcal{H})$ and $||T_{x,y}|| = ||x|| \cdot ||y||$.

(b) $T^*_{x,y} = T_{y,x}$, and $T_{x,y}T_{z,w} = \langle z, y\rangle T_{x,w}$.

(c) if $\mathcal{H}$ is separable, if $\{e_1, e_2, \cdots, e_n, \cdots\}$ is an orthonormal basis for $\mathcal{H}$, and if we write $E_{ij} = T_{e_i,e_j}$, show that $E_{ij}E_{kl} = \delta_{jk}\, E_{il}$, and $E^*_{ij} = E_{ji}$ for all i, j, k, l.

(d) Deduce that if $\mathcal{H}$ is at least two-dimensional, then $\mathcal{L}(\mathcal{H})$ is not commutative.

(e) Show that the matrix representing the operator E_{kl} *with respect to the orthonormal basis* $\{e_j\}_j$ *is the matrix with a 1 in the* (k,l) *entry and 0's elsewhere.*

(f) For $1 \leq i,j \leq n$, *let* $E^i_j \in M_n(\mathbb{C})$ *be the matrix which has 1 in the* (i,j) *entry and 0's elsewhere. (Thus* $\{E^i_j : 1 \leq i,j \leq n\}$ *is the 'standard basis' for* $M_n(\mathbb{C})$*.) Show that* $E^i_j E^k_l = \delta^k_j E^i_l \; \forall \; i,j,k,l$, *and that* $\sum_{i=1}^n E^i_i$ *is the* $(n \times n)$ *identity matrix* I_n.

In the infinite-dimensional separable case, almost exactly the same thing - as in the case of finite-dimensional $\mathcal{H}$ - is true, with the only distinction now being that not all infinite matrices will correspond to operators; to see this, let us re-state the description of the matrix thus: the j-th column of the matrix is the sequence of Fourier coefficients of Te_j with respect to the orthonormal basis $\{e_i : i \in \mathbb{N}\}$; in particular, every column must define an element of ℓ^2.

Let us see precisely which infinite matrices do arise as the matrix of a bounded operator. Suppose $T = ((t^i_j))_{i,j=1}^\infty$ is an infinite matrix. For each n, let T_n denote the 'n-th truncation' of T defined by

$$(T_n)^i_j \;=\; \begin{cases} t^i_j & if\ 1 \leq i,j \leq n \\ 0 & otherwise \end{cases}$$

Then it is clear that these truncations do define bounded operators on $\mathcal{H}$, the n-th operator being given by

$$T_n x \;=\; \sum_{i,j=1}^n t^i_j \langle x, e_j \rangle e_i \; .$$

It follows from Proposition 2.5.6(4) that T is the matrix of a bounded operator on $\mathcal{H}$ if and only if $\sup_n ||T_n|| < \infty$. It will follow from our discussion in the next chapter that if A_n denotes the $n \times n$ matrix with (i,j)-th entry t^i_j, then $||T_n||^2$ is nothing but the largest eigenvalue of the matrix $A_n^* A_n$ (all of whose eigenvalues are real and non-negative).

We conclude with a simple exercise concerning 'direct sums' of operators.

EXERCISE 2.5.8 *Let* $\mathcal{H} = \oplus_{i\in I}\mathcal{H}_i$ *be a(n internal) direct sum of Hilbert spaces. Let* $\oplus_{i\in I}^{\ell^\infty}\mathcal{L}(\mathcal{H}_i) = \{\mathbf{T} = ((T_i))_{i\in I} : T_i \in \mathcal{L}(\mathcal{H}_i)\ \forall i,\ \ \text{and}\ \sup_i ||T_i|| < \infty\}$.

(a) If $\mathbf{T} \in \oplus_{i\in I}^{\ell^\infty}\mathcal{L}(\mathcal{H}_i)$, *show that there exists a unique operator* $\oplus_i T_i \in \mathcal{L}(\mathcal{H})$ *with the property that* $(\oplus_i T_i)x_j = T_j x_j,\ \forall x_j \in \mathcal{H}_j, \forall j \in I$; *further,* $|| \oplus_i T_i|| = \sup_i ||T_i||$. *(The operator* $(\oplus_i T_i)$ *is called the 'direct sum' of the family* $\{T_i\}$ *of operators.)*

(b) If $\mathbf{T}, \mathbf{S} \in \oplus_{i\in I}^{\ell^\infty}\mathcal{L}(\mathcal{H}_i)$, *show that* $\oplus_i(T_i + S_i) = (\oplus_i T_i) + (\oplus_i S_i)$, $\oplus_i(T_i S_i) = (\oplus_i T_i)(\oplus_i S_i)$, *and* $(\oplus_i T_i)^* = (\oplus_i T_i^*)$.

Chapter 3

C*-algebras

3.1 Banach algebras

The aim of this chapter is to build up the necessary background from the theory of C^*-algebras to be able to prove the spectral theorem. We begin with the definition and some examples of a normed algebra.

DEFINITION 3.1.1 *A* **normed algebra** *is a normed space* $\mathcal{A}_0$ *with the following additional structure: there exists a well-defined multiplication in* $\mathcal{A}_0$*, meaning a map* $\mathcal{A}_0 \times \mathcal{A}_0 \to \mathcal{A}_0$*, denoted by* $(x, y) \to xy$*, which satisfies the following conditions, for all* $x, y, z \in \mathcal{A}_0, \ \alpha \in \mathbb{C}$*:*
(i) (associativity) $(xy)z = x(yz)$*;*
(ii) (distributivity) $(\alpha x+y)z = \alpha xz+yz$, $z(\alpha x+y) = \alpha zx+zy$*;*
(iii) (sub-multiplicativity of norm) $||xy|| \leq ||x|| \cdot ||y||$*.*

A **Banach algebra** *is a normed algebra which is complete as a normed space. A normed (or Banach) algebra is said to be* **unital** *if it has a multiplicative identity - i.e., if there exists an element, which we shall denote simply by 1, such that* $1x = x1 = x \ \forall x$*.*

EXAMPLE 3.1.2 (1) The space $B(X)$ of all bounded complex-valued functions on a set X - see Example 1.2.2(2) - is an example of a Banach algebra, with respect to pointwise product of functions - i.e., $(fg)(x) = f(x)g(x)$. Notice that this is a commutative algebra - i.e., $xy = yx \ \forall x, y$.

(2) If $\mathcal{A}_0$ is a normed algebra, and if $\mathcal{B}_0$ is a vector subspace which is closed under multiplication, then $\mathcal{B}_0$ is a normed algebra in its own right (with the structure induced by the one on $\mathcal{A}_0$) and is called a normed subalgebra of $\mathcal{A}_0$; a normed subalgebra $\mathcal{B}$ of a Banach algebra $\mathcal{A}$ is a Banach algebra (in the induced structure) if and only if $\mathcal{B}$ is a closed subspace of $\mathcal{A}$. (As in this example, we shall use the notational convention of using the subscript 0 for not necessarily complete normed algebras, and dropping the 0 subscript if we are dealing with Banach algebras.)

(3) Suppose $\mathcal{A}_0$ is a normed algebra, and $\mathcal{A}$ denotes its completion as a Banach space. Then $\mathcal{A}$ acquires a natural Banach algebra structure as follows: if $x, y \in \mathcal{A}$, pick sequences $\{x_n\}, \{y_n\}$ in $\mathcal{A}_0$ such that $x_n \to x,\ y_n \to y$. Then, pick a constant K such that $||x_n||, ||y_n|| \leq K\ \forall n$, note that

$$\begin{aligned} ||x_n y_n - x_m y_m|| &= ||x_n(y_n - y_m) + (x_n - x_m)y_m|| \\ &\leq ||x_n(y_n - y_m)|| + ||(x_n - x_m)y_m|| \\ &\leq ||x_n|| \cdot ||y_n - y_m|| + ||x_n - x_m|| \cdot ||y_m|| \\ &\leq K(||y_n - y_m|| + ||x_n - x_m||) \ , \end{aligned}$$

and conclude that $\{x_n y_n\}$ is a Cauchy sequence and hence convergent; define $xy = lim_n x_n y_n$. To see that the 'product' we have proposed is unambiguously defined, we must verify that the above limit is independent of the choice of the approximating sequences, and depends only on x and y. To see that this is indeed the case, suppose $x_{i,n} \to x, y_{i,n} \to y,\ i = 1, 2$. Define $x_{2n-1} = x_{1,n}, x_{2n} = x_{2,n}, y_{2n-1} = y_{1,n}, y_{2n} = y_{2,n}$, and apply the preceding reasoning to conclude that $\{x_n y_n\}$ is a convergent sequence, with limit z, say; then also $\{x_{1,n} y_{1,n} = x_{2n-1} y_{2n-1}\}$ converges to z, since a subsequence of a convergent sequence converges to the same limit; similarly, also $x_{2,n} y_{2,n} \to z$.

Thus, we have unambiguously defined the product of any two elements of $\mathcal{A}$. We leave it to the reader to verify that (a) with the product so defined, $\mathcal{A}$ indeed satisfies the requirements of a Banach algebra, and (b) the multiplication, when restricted to elements of $\mathcal{A}_0$, coincides with the initial multiplication in $\mathcal{A}_0$, and hence $\mathcal{A}_0$ is a dense normed subalgebra of $\mathcal{A}$.

(4) The set $C_c(\mathbb{R})$, consisting of continuous functions on $\mathbb{R}$

which vanish outside a bounded interval (which interval will typically depend on the particular function in question), is a normed subalgebra of $B(\mathbb{R})$, and the set $C_0(\mathbb{R})$, which may be defined as the closure in $B(\mathbb{R})$ of $C_c(\mathbb{R})$ is a Banach algebra. More generally, if X is any locally compact Hausdorff space, then the set $C_c(X)$ of continuous functions on X which vanish outside a compact set in X is a normed subalgebra of $B(X)$, and its closure in $B(X)$ is the Banach algebra $C_0(X)$ consisting of continuous functions on X which vanish at ∞.

In particular, if X is a compact Hausdorff space, then the space $C(X)$ of all continuous functions on X is a Banach algebra.

(5) The Banach space $\ell^1(\mathbb{Z})$ is a Banach algebra, if multiplication is defined thus: if we think of elements of ℓ^1 as functions $f : \mathbb{Z} \to \mathbb{C}$ with the property that $||f||_1 = \sum_{n\in\mathbb{Z}} |f(n)| < \infty$, then the (so-called **convolution**) product is defined by

$$(f * g)(n) = \sum_{m\in\mathbb{Z}} f(m)g(n-m) . \tag{3.1.1}$$

To see this, note - using the change of variables $l = n - m$ - that

$$\sum_{m,n} |f(m)g(n-m)| = \sum_{m,l} |f(m)g(l)| = ||f||_1 \cdot ||g||_1 < \infty ,$$

and conclude that if $f, g \in \ell^1(\mathbb{Z})$, then, indeed the series defining $(f * g)(n)$ in equation 3.1.1 is (absolutely) convergent. The argument given above actually shows that also $||f*g||_1 \leq ||f||_1 \cdot ||g||_1$.

An alternative way of doing this would have been to verify that the set $c_c(\mathbb{Z})$, consisting of functions on $\mathbb{Z}$ which vanish outside a finite set, is a normed algebra with respect to convolution product, and then saying that $\ell^1(\mathbb{Z})$ is the Banach algebra obtained by completing this normed algebra, as discussed in (3) above.

There is a continuous analogue of this, as follows: define a convolution product in $C_c(\mathbb{R})$ - see (4) above - as follows:

$$(f * g)(x) = \int_{-\infty}^{\infty} f(y)g(x-y)dy ; \tag{3.1.2}$$

a 'change of variable argument' entirely analogous to the discrete case discussed above, shows that if we define

$$||f||_1 \;=\; \int_{-\infty}^{\infty} |f(x)|dx \;, \qquad (3.1.3)$$

then $C_c(\mathbb{R})$ becomes a normed algebra when equipped with the norm $||\cdot||_1$ and convolution product; the Banach algebra completion of this normed algebra - at least one manifestation of it - is the space $L^1(\mathbb{R}, m)$, where m denotes Lebesgue measure.

The preceding constructions have analogues in any 'locally compact topological group'; thus, suppose G is a group which is equipped with a topology with respect to which mulitplication and inversion are continuous, and suppose that G is a locally compact Hausdorff space. For $g \in G$, let L_g be the homeomorphism of G given by left multiplication by g; it is easy to see (and useful to bear in mind) that G is homogeneous as a topological space, since the group of homeomorphisms acts transitively on G (meaning that given any two points $g, h \in G$, there is a homeomorphism - namely $L_{hg^{-1}}$ - which maps g to h).

Useful examples of topological groups to bear in mind are: (i) the group $\mathcal{G}(\mathcal{A})$ of invertible elements in a unital Banach algebra $\mathcal{A}$ - of which more will be said soon - (with respect to the norm and the product in $\mathcal{A}$); (ii) the group $\mathcal{U}(\mathcal{H})$ of unitary operators on a Hilbert space $\mathcal{H}$ (with respect to the norm and the product in $\mathcal{L}(\mathcal{H})$); and (iii) the finite-dimensional (or matricial) special cases of the preceding examples $GL_n(\mathbb{C})$ and the compact group $U(n, \mathbb{C}) \;=\; \{U \in M_n(\mathbb{C}) : U^*U = I_n\}$ (where U^* denotes the complex conjugate transpose of the matrix U). (When $\mathcal{A}$ and $\mathcal{H}$ are one-dimensional, we recover the groups $\mathbb{C}^\times$ (of non-zero-complex numbers) and $\mathbb{T}$ (of complex numbers of unit modulus; also note that in case $\mathcal{A}$ (resp., $\mathcal{H}$) is finite-dimensional, then the example in (i) (resp., (ii)) above are instances of locally compact groups.)

Let $C_c(G)$ be the space of compactly supported continuous functions on G, regarded as being normed by $||\cdot||_\infty$; the map L_g of G induces a map, via composition, on $C_c(G)$ by $\lambda_g(f) \;=\; f \circ L_{g^{-1}}$. (The reason for the inverse is 'functorial', meaning that it is only this way that we can ensure that $\lambda_{gh} = \lambda_g \circ \lambda_h$; thus $g \mapsto \lambda_g$ defines a 'representation' of G on $C_c(G)$ (meaning a group homomorphism $\lambda : G \to \mathcal{G}(\ \mathcal{L}(C_c(G)))$, in the notation of example (i) of the preceding paragraph). The map λ is 'strongly continuous', meaning that if $g_i \to g$ (is a convergent net in G), then $||\lambda_{g_i} f - \lambda_g f||_\infty \to 0 \ \forall f \in C_c(G)$.

Analogous to the case of $\mathbb{R}$ or $\mathbb{Z}$, it is a remarkable fact that there is always a bounded linear functional $m_G \in C_c(G)^*$ which is '(left-) translation invariant' in the sense that $m_G \circ \lambda_g = m_G \ \forall g \in G$. It is then a consequence of the Riesz Representation theorem (see the Appendix - §A.7) that there exists a unique measure - which we shall denote by μ_G defined on the Borel sets of G such that $m_g(f) = \int_G f \ d\mu_G$. This measure μ_G inherits translational invariance from m_G - i.e., $\mu_G \circ L_g = \mu_G$.

It is a fact that the functional m_G, and consequently the measure μ_G, is (essentially) uniquely determined by the translational invariance condition (up to scaling by a positive scalar); this unique measure is referred to as **(left) Haar measure** on G. It is then the case that $C_c(G)$ can alternatively be normed by

$$||f||_1 = \int_G |f| \ d\mu_G ,$$

and that it now becomes a normed algebra with respect to convolution product

$$(f * g)(t) = \int_G f(s)g(s^{-1}t) \ d\mu_G(s) . \tag{3.1.4}$$

The completion of this normed algebra - at least one version of it - is given by $L^1(G, \mu_G)$.

(6) Most examples presented so far have been commutative algebras - meaning that $xy = yx \ \forall x, y$. (It is a fact that $L^1(G)$ is commutative precisely when G is commutative, i.e., abelian.) The following one is not. If X is any normed space and X is not 1-dimensional, then $\mathcal{L}(X)$ is a non-commutative normed algebra; (prove this fact!); further, $\mathcal{L}(X)$ is a Banach algebra if and only if X is a Banach space. (The proof of one implication goes along the lines of the proof of the fact that $C[0, 1]$ is complete; for the other, consider the embedding of X into $\mathcal{L}(X)$ given by $\mathcal{L}_y(x) = \phi(x)y$, where ϕ is a linear functional of norm 1.)

In particular, - see Exercise 2.5.7 - $M_n(\mathbb{C})$ is an example of a non-commutative Banach algebra for each $n = 2, 3, \cdots$. $\square$

Of the preceding examples, the unital algebras are $B(X)$ (X any set), $C(X)$ (X a compact Hausdorff space), $\ell^1(\mathbb{Z})$ (and more generally, $L^1(G)$ in case the group G is (equipped with the) discrete (topology), and $\mathcal{L}(X)$ (X any normed space). Verify this assertion by identifying the multiplicative identity in these examples, and showing that the other examples do not have any identity element.

Just as it is possible to complete a normed algebra and manufacture a Banach algebra out of an incomplete normed algebra, there is a (and in fact, more than one) way to start with a normed algebra without identity and manufacture a near relative which is unital. This, together with some other elementary facts concerning algebras, is dealt with in the following exercises.

EXERCISE 3.1.3 *(1) If $\mathcal{A}_0$ is a normed algebra, consider the algebra defined by $\mathcal{A}_0^+ = \mathcal{A}_0 \oplus_{\ell^1} \mathbb{C}$; thus, $\mathcal{A}_0^+ = \{(x, \alpha) : x \in X, \alpha \in \mathbb{C}\}$ as a vector space, and the product and norm are defined by $(x, \alpha)(y, \beta) = (xy+\alpha y+\beta x, \alpha\beta)$ and $||(x, \alpha)|| = ||x||+|\alpha|$. Show that with these definitions,*

(a) $\mathcal{A}_0^+$ is a unital normed algebra;

(b) the map $\mathcal{A}_0 \ni x \mapsto (x, 0) \in \mathcal{A}_0^+$ is an isometric isomorphism of the algebra $\mathcal{A}_0$ onto the subalgebra $\{(x, 0) : x \in \mathcal{A}_0\}$ of $\mathcal{A}_0^+$;

(c) $\mathcal{A}_0$ is a Banach algebra if and only if $\mathcal{A}_0^+$ is a Banach algebra.

(2) Show that multiplication is jointly continuous in a normed algebra, meaning that $x_n \to x, y_n \to y \Rightarrow x_n y_n \to xy$.

(3) If $\mathcal{A}_0$ is a normed algebra, show that a multiplicative identity, should one exist, is necessarily unique; what can you say about the norm of the identity element?

(4) (a) Suppose $\mathcal{I}$ is a closed ideal in a normed algebra $\mathcal{A}_0$ - i.e., $\mathcal{I}$ is a (norm-) closed subspace of $\mathcal{A}_0$, such that whenever $x, y \in \mathcal{I}, z \in \mathcal{A}_0, \alpha \in \mathbb{C}$, we have $\alpha x + y, xz, zx \in \mathcal{I}$; then show that the quotient space $\mathcal{A}_0/\mathcal{I}$ - with the normed space structure discussed in Exercise 1.5.3(3) - is a normed algebra with respect to multiplication defined (in the obvious manner) as: $(z+\mathcal{I})(z'+\mathcal{I}) = (zz'+\mathcal{I})$.

(b) if $\mathcal{I}$ is a closed ideal in a Banach algebra $\mathcal{A}$, deduce that $\mathcal{A}/\mathcal{I}$ is a Banach algebra.

(c) in the notation of Exercise (1) above, show that $\{(x, 0) : x \in \mathcal{A}_0\}$ is a closed ideal - call it $\mathcal{I}$ - of $\mathcal{A}_0^+$ such that $\mathcal{A}_0^+/\mathcal{I} \cong \mathbb{C}$.

We assume, for the rest of this section, that $\mathcal{A}$ is a unital Banach algebra with (multiplicative) identity 1; we shall assume that $\mathcal{A} \neq \{0\}$, or equivalently, that $1 \neq 0$. As with 1, we shall simply write λ for $\lambda 1$. We shall call an element $x \in \mathcal{A}$ *invertible* (in $\mathcal{A}$) if there exists an element - which is necessarily unique and which we shall always denote by x^{-1} - such that $xx^{-1} = x^{-1}x = 1$, and we shall denote the collection of all such invertible elements of $\mathcal{A}$ as $\mathcal{G}(\mathcal{A})$. (This is obviously a group with repect to multiplication, and is sometimes referred to as the 'group of units' in $\mathcal{A}$.)

We gather together some basic facts concerning units in a Banach algebra in the following proposition.

PROPOSITION 3.1.4 *(1) If $x \in \mathcal{G}(\mathcal{A})$, and if we define L_x (to be the map given by left-multiplication by x) as $L_x(y) = xy, \forall y \in \mathcal{A}$ then $L_x \in \mathcal{L}(\mathcal{A})$ and further, L_x is an invertible operator, with $(L_x)^{-1} = L_{x^{-1}}$;*

(2) If $x \in \mathcal{G}(\mathcal{A})$ and $y \in \mathcal{A}$, then $xy \in \mathcal{G}(\mathcal{A}) \Leftrightarrow yx \in \mathcal{G}(\mathcal{A}) \Leftrightarrow y \in \mathcal{G}(\mathcal{A})$.

(3) If $\{x_1, x_2, \cdots, x_n\} \subset \mathcal{A}$ is a set of pairwise commuting elements - i.e., $x_i x_j = x_j x_i \; \forall i, j$ - then the product $x_1 x_2 \cdots x_n$ is invertible if and only if each x_i is invertible.

(4) If $x \in \mathcal{A}$ and $||1 - x|| < 1$, then $x \in \mathcal{G}(\mathcal{A})$ and

$$x^{-1} = \sum_{n=0}^{\infty} (1 - x)^n , \tag{3.1.5}$$

where we use the convention that $y^0 = 1$.

In particular, if $\lambda \in \mathbb{C}$ and $|\lambda| > ||x||$, then $(x - \lambda)$ is invertible and

$$(x - \lambda)^{-1} = - \sum_{n=0}^{\infty} \frac{x^n}{\lambda^{n+1}} . \tag{3.1.6}$$

(5) $\mathcal{G}(\mathcal{A})$ is an open set in $\mathcal{A}$ and the mapping $x \mapsto x^{-1}$ is a homeomorphism of $\mathcal{G}(\mathcal{A})$ onto itself.

Proof : (1) Obvious.

(2) Clearly if y is invertible, so is xy (resp., yx); conversely, if xy (resp., yx) is invertible, so is $y = x^{-1}(xy)$ (resp., $y = (yx)x^{-1}$).

(3) One implication follows from the fact that $\mathcal{G}(\mathcal{A})$ is closed under multiplication. Suppose, conversely, that $x = x_1 x_2 \cdots x_n$ is invertible; it is clear that x_i commutes with x, for each i; it follows easily that each x_i also commutes with x^{-1}, and that the inverse of x_i is given by the product of x^{-1} with the product of all the x_j's with $j \neq 1$. (Since all these elements commute, it does not matter in what order they are multiplied.)

(4) The series on the right side of equation 3.1.5 is (absolutely, and hence) summable (since the appropriate geometric series converges under the hypothesis); let $\{s_n = \sum_{i=0}^{n}(1-x)^i\}$ be the sequence of partial sums, and let s denote the limit; clearly each s_n commutes with $(1-x)$, and also, $(1-x)s_n = s_n(1-x) = s_{n+1} - 1$; hence, in the limit, $(1-x)s = s(1-x) = s - 1$, i.e., $xs = sx = 1$.

As for the second equation in (4), we may apply equation 3.1.5 to $(1 - \frac{x}{\lambda})$ in place of x, and justify the steps

$$\begin{aligned}(x-\lambda)^{-1} &= -\lambda^{-1}\,(1-\frac{x}{\lambda})^{-1} \\ &= -\lambda^{-1}\sum_{n=0}^{\infty}\frac{x^n}{\lambda^n} \\ &= -\sum_{n=0}^{\infty}\frac{x^n}{\lambda^{n+1}}\ .\end{aligned}$$

(5) The assertions in the first half of (4) imply that 1 is in the interior of $\mathcal{G}(\mathcal{A})$ and is a point of continuity of the inversion map; by the already established (1) and (2) of this proposition, we may 'transport this local picture at 1' via the linear homeomorphism L_x (which maps $\mathcal{G}(\mathcal{A})$ onto itself) to a 'corresponding local picture at x', and deduce that any element of $\mathcal{G}(\mathcal{A})$ is an interior point and is a point of continuity of the inversion map. □

We now come to the fundamental notion in the theory.

DEFINITION 3.1.5 *Let $\mathcal{A}$ be a unital Banach algebra, and let $x \in \mathcal{A}$. Then the* **spectrum** *of x is the set, denoted by $\sigma(x)$,* *defined by*

$$\sigma(x) \;=\; \{\lambda \in \mathbb{C} : (x - \lambda) \textit{ is not invertible}\} \;, \qquad (3.1.7)$$

and the **spectral radius** *of x is defined by*

$$r(x) \;=\; \sup\{|\lambda| : \lambda \in \sigma(x)\} \;. \qquad (3.1.8)$$

(This is thus the radius of the smallest disc centered at 0 in $\mathbb{C}$ which contains the spectrum. Strictly speaking, the last sentence makes sense only if the spectrum is a bounded set; we shall soon see that this is indeed the case.)

(b) The **resolvent set** *of x is the complementary set $\rho(x)$ defined by*

$$\rho(x) \;=\; \mathbb{C} - \sigma(x) \;=\; \{\lambda \in \mathbb{C} : (x - \lambda) \textit{ is invertible}\} \;, \quad (3.1.9)$$

and the map

$$\rho(x) \ni \lambda \stackrel{R_x}{\mapsto} (x - \lambda)^{-1} \;\in\; \mathcal{G}(\mathcal{A}) \qquad (3.1.10)$$

is called the **resolvent function** *of x.*

To start with, we observe - from the second half of Proposition 3.1.4(4) - that

$$r(x) \;\leq\; ||x|| \;. \qquad (3.1.11)$$

Next, we deduce - from Proposition 3.1.4(5) - that $\rho(x)$ is an open set in $\mathbb{C}$ and that the resolvent function of x is continuous. It follows immediately that $\sigma(x)$ is a compact set.

In the proof of the following proposition, we shall use some elementary facts from complex function theory; the reader who is unfamiliar with the notions required is urged to fill in the necessary background from any standard book on the subject (such as [Ahl], for instance).

PROPOSITION 3.1.6 *Let $x \in \mathcal{A}$; then,*

(a) $\lim_{|\lambda| \to \infty} \; ||R_x(\lambda)|| \;=\; 0$*; and*

*(b) (***Resolvent equation***)*

$$R_x(\lambda) - R_x(\mu) \;=\; (\lambda - \mu) R_x(\lambda) R_x(\mu) \;\forall\; \lambda, \mu \in \rho(x) \;;$$

(c) the resolvent function is 'weakly analytic', meaning that if $\phi \in \mathcal{A}^*$, then the map $\phi \circ R_x$ is an analytic function (defined on the open set $\rho(x) \subset \mathbb{C}$), such that

$$\lim_{|\lambda| \to \infty} \phi \circ R_x(\lambda) \ = \ 0 \ .$$

Proof : (a) It is an immediate consequence of equation 3.1.6 that there exists a constant $C > 0$ such that

$$||R_x(\lambda)|| \ \leq \frac{C}{|\lambda|} \ \forall |\lambda| > ||x||,$$

and assertion (a) follows.

(b) If $\lambda, \mu \in \rho(x)$, then

$$\begin{aligned}
(\lambda - \mu) R_x(\lambda) R_x(\mu) &= (\lambda - \mu)(x - \lambda)^{-1}(x - \mu)^{-1} \\
&= R_x(\lambda) \ (\ (x - \mu) - (x - \lambda) \ ) \ R_x(\mu) \\
&= R_x(\lambda) - R_x(\mu) \ .
\end{aligned}$$

(c) If $\phi \in \mathcal{A}^*$ is any continuous linear functional on $\mathcal{A}$, it follows from Proposition 3.1.4(5) and the resolvent equation above, that if $\mu \in \rho(x)$, and if λ is sufficiently close to μ, then $\lambda \in \rho(x)$, and

$$\begin{aligned}
\lim_{\lambda \to \mu} \left(\frac{\phi \circ R_x(\lambda) - \phi \circ R_x(\mu)}{\lambda - \mu} \right) &= \lim_{\lambda \to \mu} (\phi(\ R_x(\lambda) \cdot R_x(\mu) \ )) \\
&= \phi(R_x(\mu)^2) \ ,
\end{aligned}$$

thereby establishing (complex) differentiability, i.e., analyticity, of $\phi \circ R_x$ at the point μ.

It is an immediate consequence of (a) above and the boundedness of the linear functional ϕ that

$$\lim_{\lambda \to \mu} \phi \circ R_x(\lambda) \ ,$$

and the proof is complete. $\Box$

THEOREM 3.1.7 *If $\mathcal{A}$ is any Banach algebra and $x \in \mathcal{A}$, then $\sigma(x)$ is a non-empty compact subset of $\mathbb{C}$.*

Proof : Assume the contrary, which means that $\rho(x) = \mathbb{C}$. Let $\phi \in \mathcal{A}^*$ be arbitrary. Suppose $\rho(x) = \mathbb{C}$; in view of Proposition 3.1.6(c), this means that $\phi \circ R_x$ is an 'entire' function - meaning a function which is 'complex-differentiable' throughout $\mathbb{C}$ - which vanishes at infinity; it then follows from Liouville's theorem that $\phi \circ R_x(\lambda) = 0 \ \forall \lambda \in \mathbb{C}$. (For instance, see [Ahl], for Liouville's theorem.) Since ϕ was arbitrary, and since $\mathcal{A}^*$ 'separates points of $\mathcal{A}$ - see Exercise 1.5.3(1) - we may conclude that $R_x(\lambda) = 0 \ \forall \ \lambda \in \mathbb{C}$; but since $R_x(\lambda) \in \mathcal{G}(\mathcal{A})$, this is absurd, and the contradiction completes the proof. □

REMARK 3.1.8 Suppose $\mathcal{A} = M_n(\mathbb{C})$. It is then a consequence of basic properties of the determinant mapping - see the Appendix (§A.1) - that if $T \in M_n(\mathbb{C})$, then $\lambda \in \sigma(T)$ if and only if λ is an eigenvalue of the matrix T, i.e, $p_T(\lambda) = 0$ where $p_T(z) = det(T - z)$ denotes the characteristsc polynomial of the matrix T. On the other hand, it is true - see §A.1 - that any polynomial of degree n, with leading coefficient equal to $(-1)^n$, is the characteristic polynomial of some matrix in $M_n(\mathbb{C})$. Hence the statement that $\sigma(T)$ is a non-empty set, for every $T \in M_n(\mathbb{C})$, is equivalent to the statement that every complex polynomial has a complex root, viz., the so-called Fundamental Theorem of Algebra. Actually, the fundamental theorem is the statement that every complex polynomial of degree N is expressible as a product of N polynomials of degree 1; this version is easily derived, by an induction argument, from the slightly weaker statement earlier referred to as the fundamental theorem. □

PROPOSITION 3.1.9 *If* $p(z) = \sum_{n=0}^{N} a_n z^n$ *is a polynomial with complex coefficients, and if we define* $p(x) = a_0 \cdot 1 + \sum_{n=1}^{N} a_n x^n$ *for each* $x \in \mathcal{A}$, *then*

(i) $p(z) \mapsto p(x)$ *is a homomorphism from the algebra* $\mathbb{C}[z]$ *of complex polynomials onto the subalgebra of* $\mathcal{A}$ *which is generated by* $\{1, x\}$*;*

*(ii) (***spectral mapping theorem***) if* p *is any polynomial as above, then*

$$\sigma(\ p(x)\ ) \ = \ p(\ \sigma(x)\ ) \ = \ \{p(\lambda) : \lambda \in \sigma(x)\} \ .$$

Proof : The first assertion is easily verified. As for (ii), temporarily fix $\lambda \in \mathbb{C}$; if $p(z) = \sum_{n=0}^{N} a_n z^n$, assume without loss of generality that $a_N \neq 0$, and (appeal to the fundamental theorem of algebra to) conclude that there exist $\lambda_1, \cdots, \lambda_N$ such that

$$p(z) - \lambda = a_N \prod_{n=1}^{N} (z - \lambda_i) .$$

(Thus the λ_i are all the zeros, counted according to multiplicity, of the polynomial $p(z) - \lambda$.)

Deduce from part (i) of this Proposition that

$$p(x) - \lambda = a_N \prod_{n=1}^{N} (x - \lambda_i) .$$

Conclude now, from Proposition 3.1.4(3), that

$$\begin{aligned} \lambda \notin \sigma(p(x)) &\Leftrightarrow \lambda_i \notin \sigma(x) \ \forall \ 1 \leq i \leq n \\ &\Leftrightarrow \lambda \notin p(\sigma(x)) \end{aligned}$$

and the proof is complete. □

The next result is a quantitative strengthening of the assertion that the spectrum is non-empty.

THEOREM 3.1.10 *(***spectral radius formula***)*
If $\mathcal{A}$ is a Banach algebra and if $x \in \mathcal{A}$, then

$$r(x) = \lim_{n \to \infty} ||x^n||^{\frac{1}{n}} . \tag{3.1.12}$$

Proof : To start with, fix an arbitrary $\phi \in \mathcal{A}^*$, and let $F = \phi \circ R_x$; by definition, this is analytic in the exterior of the disc $\{\lambda \in \mathbb{C} : |\lambda| \leq r(x)\}$; on the other hand, we may conclude from equation 3.1.6 that F has the Laurent series expansion

$$F(\lambda) = - \sum_{n=0}^{\infty} \frac{\phi(x^n)}{\lambda^{n+1}} ,$$

which is valid in the exterior of the disc $\{\lambda \in \mathbb{C} : |\lambda| \leq ||x||\}$. Since F vanishes at infinity, the function F is actually analytic at the point at infinity, and consequently the Laurent expansion above is actually valid in the larger region $|\lambda| > r(x)$.

So if we temporarily fix a λ with $|\lambda| > r(x)$, then we find, in particular, that

$$lim_{n\to\infty} \frac{\phi(x^n)}{\lambda^n} = 0 .$$

Since ϕ was arbitrary, it follows from the uniform boundedness principle - see Exercise 1.5.17(1) - that there exists a constant $K > 0$ such that

$$||x^n|| \leq K |\lambda|^n \ \forall n .$$

Hence, $||x^n||^{\frac{1}{n}} \leq K^{\frac{1}{n}} |\lambda|$. By allowing $|\lambda|$ to decrease to $r(x)$, we may conclude that

$$\lim \sup_n ||x^n||^{\frac{1}{n}} \leq r(x) . \tag{3.1.13}$$

On the other hand, deduce from Proposition 3.1.9(2) and equation 3.1.11 that $r(x) = r(x^n)^{\frac{1}{n}} \leq ||x^n||^{\frac{1}{n}}$, whence we have

$$r(x) \leq \lim \inf_n ||x^n||^{\frac{1}{n}} , \tag{3.1.14}$$

and the theorem is proved. □

Before proceeding further, we wish to draw the reader's attention to one point, as a sort of note of caution; and this pertains to the possible dependence of the notion of the spectrum on the ambient Banach algebra.

Thus, suppose $\mathcal{A}$ is a unital Banach algebra, and suppose $\mathcal{B}$ is a unital Banach subalgebra of $\mathcal{A}$; suppose now that $x \in \mathcal{B}$; then, for any intermediate Banach subalgebra $\mathcal{B} \subset \mathcal{C} \subset \mathcal{A}$, we can talk of

$$\rho_{\mathcal{C}}(x) = \{\lambda \in \mathbb{C} : \exists z \in \mathcal{C} \ such\ that\ z(x-\lambda) = (x-\lambda)z = 1\} ,$$

and $\sigma_{\mathcal{C}}(x) = \mathbb{C} - \rho_{\mathcal{C}}(x)$. The following result describes the relations between the different notions of spectra.

Proposition 3.1.11 *Let* $1, x \in \mathcal{B} \subset \mathcal{A}$*, as above. Then,*

(a) $\sigma_{\mathcal{B}}(x) \supset \sigma_{\mathcal{A}}(x)$;

(b) $\partial(\sigma_{\mathcal{B}}(x)) \subset \partial(\sigma_{\mathcal{A}}(x))$, *where* $\partial(\Sigma)$ *denotes the 'topological boundary' of the subset* $\Sigma \subset \mathbb{C}$. *(This is the set defined by* $\partial\Sigma = \overline{\Sigma} \cap \overline{\mathbb{C} - \Sigma}$*; thus,* $\lambda \in \partial\Sigma$ *if and only if there exists sequences* $\{\lambda_n\} \subset \Sigma$ *and* $\{z_n\} \subset \mathbb{C} - \Sigma$ *such that* $\lambda = lim_n \lambda_n = lim_n z_n$*.)*

Proof : (a) We need to show that $\rho_{\mathcal{B}}(x) \subset \rho_{\mathcal{A}}(x)$. So suppose $\lambda \in \rho_{\mathcal{B}}(x)$; this means that there exists $z \in \mathcal{B} \subset \mathcal{A}$ such that $z(x-\lambda) = (x-\lambda)z = 1$; hence, $\lambda \in \rho_{\mathcal{A}}(x)$.

(b) Suppose $\lambda \in \partial(\ \sigma_{\mathcal{B}}(x)\ )$; since the spectrum is closed, this means that $\lambda \in \sigma_{\mathcal{B}}(x)$ and that there exists a sequence $\{\lambda_n\} \subset \rho_{\mathcal{B}}(x)$ such that $\lambda_n \to \lambda$. Since (by (a)) $\lambda_n \in \rho_{\mathcal{A}}(x)\ \forall n$, we only need to show that $\lambda \in \sigma_{\mathcal{A}}(x)$; if this were not the case, then $(x-\lambda_n) \to (x-\lambda) \in \mathcal{G}(\mathcal{A})$, which would imply - by Proposition 3.1.4(5) - that $(x-\lambda_n)^{-1} \to (x-\lambda)^{-1}$; but the assumptions that $\lambda_n \in \rho_{\mathcal{B}}(x)$ and that $\mathcal{B}$ is a Banach subalgebra of $\mathcal{A}$ would then force the conclusion $(x-\lambda)^{-1} \in \mathcal{B}$, contradicting the assumption that $\lambda \in \sigma_{\mathcal{B}}(x)$. $\square$

The purpose of the next example is to show that it is possible to have strict inclusion in Proposition 3.1.11(a).

EXAMPLE 3.1.12 In this example, and elsewhere in this book, we shall use the symbols $\mathbb{D}, \overline{\mathbb{D}}$ and $\mathbb{T}$ to denote the open unit disc, the closed unit disc and the unit circle in $\mathbb{C}$, respectively. (Thus, if $z \in \mathbb{C}$, then $z \in \mathbb{D}$ (resp., $\overline{\mathbb{D}}$, *resp.*, $\mathbb{T}$) if and only if the absolute value of z is less than (resp., not greater than, resp., equal to) 1.

The **disc algebra** is, by definition, the class of all functions which are analytic in $\mathbb{D}$ and have continuous extensions to $\overline{\mathbb{D}}$; thus,

$$A(\mathbb{D}) = \{f \in C(\overline{\mathbb{D}}) : f|_{\mathbb{D}} \text{ is analytic}\} .$$

It is easily seen that $A(\mathbb{D})$ is a (closed subalgebra of $C(\overline{\mathbb{D}})$ and consequently a) Banach algebra. (Verify this!) Further, if we let $T : A(\mathbb{D}) \to C(\mathbb{T})$ denote the 'restriction map', it follows from the maximum modulus principle that T is an isometric algebra isomorphism of $A(\mathbb{D})$ into $C(\mathbb{T})$. Hence, we may - and do - regard $\mathcal{B} = A(\mathbb{D})$ as a Banach subalgebra of $\mathcal{A} = C(\mathbb{T})$. It is another easy exercise - which the reader is urged to carry out - to verify that if $f \in \mathcal{B}$, then $\sigma_{\mathcal{B}}(f) = f(\overline{\mathbb{D}})$, while $\sigma_{\mathcal{A}}(f) = f(\mathbb{T})$. Thus, for example, if we consider $f_0(z) = z$, then we see that $\sigma_{\mathcal{B}}(f_0) = \overline{\mathbb{D}}$, while $\sigma_{\mathcal{A}}(f_0) = \mathbb{T}$.

The preceding example more or less describes how two compact subsets in $\mathbb{C}$ must look like, if they are to satisfy the two conditions stipulated by (a) and (b) of Proposition 3.1.11. Thus, in general, if $\Sigma_0, \Sigma \subset \mathbb{C}$ are compact sets such that $\Sigma \subset \Sigma_0$ and $\partial\Sigma_0 \subset \partial\Sigma$, then it is the case that Σ_0 is obtained by 'filling in some of the holes in Σ'. (The reader is urged to make sense of this statement and to then try and prove it.) $\square$

3.2 Gelfand-Naimark theory

DEFINITION 3.2.1 *(a) Recall - see Exercise 3.1.3 (4) - that a subset $\mathcal{I}$ of a normed algebra $\mathcal{A}_0$ is said to be an* **ideal** *if the following conditions are satisfied, for all choices of $x, y \in \mathcal{I}, z \in \mathcal{A}_0$, and $\alpha \in \mathbb{C}$:*

$$\alpha x + y,\ xz,\ zx\ \in \mathcal{I}\ .$$

(b) A **proper ideal** *is an ideal $\mathcal{I}$ which is distinct from the trivial ideals $\{0\}$ and $\mathcal{A}_0$.*

(c) A **maximal ideal** *is a proper ideal which is not strictly contained in any larger ideal.*

REMARK 3.2.2 According to our definitions, $\{0\}$ is not a maximal ideal; in order for some of our future statements to be applicable in all cases, we shall find it convenient to adopt the convention that in the single exceptional case when $\mathcal{A}_0 = \mathbb{C}$, we shall consider $\{0\}$ as a maximal ideal. $\square$

EXERCISE 3.2.3 *(1) Show that the (norm-) closure of an ideal in a normed algebra is also an ideal.*

(2) If $\mathcal{I}$ is a closed ideal in a normed algebra $\mathcal{A}_0$, show that the quotient normed space $\mathcal{A}_0/\mathcal{I}$ - see Exercise 1.5.3(3) - is a normed algebra with respect to the natural definition $(x+\mathcal{I})(y+\mathcal{I}) = (xy+\mathcal{I})$, and is a Banach algebra if $\mathcal{A}_0$ is.

(3) Show that if $\mathcal{A}$ is a unital normed algebra, then the following conditions on a non-zero ideal $\mathcal{I}$ (i.e., $\mathcal{I} \neq \{0\}$) are equivalent:

(i) $\mathcal{I}$ is a proper ideal;

(ii) $\mathcal{I} \cap \mathcal{G}(\mathcal{A}) = \emptyset$;

(iii) $1 \notin \mathcal{I}$.

(4) Deduce from (3) above that the closure of a proper ideal in a unital Banach algebra is also a proper ideal, and hence, conclude that any maximal ideal in a unital Banach algebra is closed.

(5) Consider the unital Banach algebra $\mathcal{A} = C(X)$, *where* X *is a compact Hausdorff space.*

(a) For a subset $S \subset X$, *define* $\mathcal{I}(S) = \{f \in \mathcal{A} : f(x) = 0 \ \forall x \in S\}$; *show that* $\mathcal{I}(S)$ *is a closed ideal in* $\mathcal{A}$ *and that* $\mathcal{I}(S) = \mathcal{I}(\overline{S})$ *- where, of course, the symbol* $\overline{S}$ *denotes the closure of* S.

(b) For $S \subset X$, *define* $\mathcal{I}^0(S) = \{f \in \mathcal{A} : f$ *vanishes in some open neighbourhood of the set* S *(which might depend on* f*)*$\}$; *show that* $\mathcal{I}^0(S)$ *is an ideal in* $\mathcal{A}$, *and try to determine what subsets* S *would have the property that* $\mathcal{I}^0(S)$ *is a closed ideal.*

(c) Show that every closed ideal in $\mathcal{A}$ *has the form* $\mathcal{I}(F)$ *for some closed subset* $F \subset X$, *and that the closed set* F *is uniquely determined by the ideal. (Hint: Let* $\mathcal{I}$ *be a closed ideal in* $C(X)$. *Define* $F = \{x \in X : f(x) = 0 \forall f \in \mathcal{I}\}$. *Then clearly* $\mathcal{I} \subset \mathcal{I}(F)$. *Conversely, let* $f \in \mathcal{I}(F)$ *and let* $U = \{x : |f(x)| < \epsilon\}$, $V = \{x : |f(x)| < \frac{\epsilon}{2}\}$, *so* U *and* V *are open sets such that* $F \subset V \subset \overline{V} \subset U$. *First deduce from Urysohn's lemma that there exists* $h \in C(X)$ *such that* $h(\overline{V}) = \{0\}$ *and* $h(X - U) = \{1\}$. *Set* $f_1 = fh$, *and note that* $||f_1 - f|| < \epsilon$. *For each* $x \notin V$, *pick* $f_x \in \mathcal{I}$ *such that* $f_x(x) = 1$; *appeal to compactness of* $X - V$ *to find finitely many points* $x_1, \cdots, x_n$ *such that* $X - V \subset \cup_{i=1}^n \{y \in X : |f_{x_i}(y)| > \frac{1}{2}\}$; *set* $g = 4\sum_{i=1}^n |f_{x_i}|^2$, *and note that* $g \in \mathcal{I}$ *- since* $|f_i|^2 = f_i \overline{f_i}$ *- and* $|g(y)| > 1 \ \forall y \in X - V$; *conclude from Tietze's extension theorem - see Theorem A.4.24 - that there exists an* $h_1 \in C(X)$ *such that* $h_1(y) = \frac{1}{g(y)} \ \forall \ y \notin V$. *Notice now that* $f_1 g h_1 \in \mathcal{I}$ *and that* $f_1 g h_1 = f_1$, *since* f_1 *is supported in* $X - V$. *Since* $\epsilon > 0$ *was arbitrary, and since* $\mathcal{I}$ *was assumed to be closed, this shows that* $f \in \mathcal{I}$.*)*

(d) Show that if $F_i, i = 1, 2$ *are closed subsets of* X, *then* $I(F_1) \subset I(F_2) \Leftrightarrow F_1 \supset F_2$, *and hence deduce that the maximal ideals in* $C(X)$ *are in bijective correspondence with the points in* X.

(e) If $S \subset X$, *and if the closure of* $I^0(S)$ *is* $I(F)$, *what is the relationship between* S *and* F.

(6) Can you show that there are no proper ideals in $M_n(\mathbb{C})$?

(7) If $\mathcal{I}$ is an ideal in an algebra $\mathcal{A}$, show that:
(i) the vector space $\mathcal{A}/\mathcal{I}$ is an algebra with the natural structure;
(ii) there exists a 1-1 correspondence between ideals in $\mathcal{A}/\mathcal{I}$ and ideals in $\mathcal{A}$ which contain $\mathcal{I}$. (Hint: Consider the correspondence $\mathcal{J} \to \pi^{-1}(\mathcal{J})$, where $\pi : \mathcal{A} \to \mathcal{A}/\mathcal{I}$ is the quotient map.)

Throughout this section, we shall assume that $\mathcal{A}$ is a commutative Banach algebra, and unless it is explicitly stated to the contrary, we will assume that $\mathcal{A}$ is a unital algebra. The key to the Gelfand-Naimark theory is the fact there are always lots of maximal ideals in such an algebra; the content of the Gelfand-Naimark theorem is that the collection of maximal ideals contains a lot of information about $\mathcal{A}$, and that in certain cases, this collection 'recaptures' the algebra. (See the example discussed in Exercise 3.2.3(5), for an example where this is the case.)

PROPOSITION 3.2.4 *If $x \in \mathcal{A}$, then the following conditions are equivalent:*
(i) x is not invertible;
(ii) there exists a maximal ideal $\mathcal{I}$ in $\mathcal{A}$ such that $x \in \mathcal{I}$.

Proof : The implication $(ii) \Rightarrow (i)$ is immediate - see Exercise 3.2.3(3).

Conversely, suppose x is not invertible. If $x = 0$, there is nothing to prove, so assume $x \neq 0$. Then, $\mathcal{I}_0 = \{ax : a \in \mathcal{A}\}$ is an ideal in $\mathcal{A}$. Notice that $\mathcal{I}_0 \neq \{0\}$ (since it contains $x = 1x$) and $\mathcal{I}_0 \neq \mathcal{A}$ (since it does not contain 1); thus, $\mathcal{I}_0$ is a proper ideal. An easy application of Zorn's lemma now shows that there exists a maximal ideal $\mathcal{I}$ which contains $\mathcal{I}_0$, and the proof is complete. □

The proof of the preceding proposition shows that any commutative Banach algebra, which contains a non-zero element which is not invertible, must contain non-trivial (maximal) ideals. The next result disposes of commutative Banach algebras which do not satisfy this requirement.

THEOREM 3.2.5 *(***Gelfand-Mazur theorem***)*

The following conditions on a unital commutative Banach algebra $\mathcal{A}$ are equivalent:

(i) $\mathcal{A}$ is a division algebra - i.e., every non-zero element is invertible;

(ii) $\mathcal{A}$ is simple - i.e., there exist no proper ideals in $\mathcal{A}$; and

(iii) $\mathcal{A} = \mathbb{C}1$.

Proof : $(i) \Rightarrow (ii)$: If $\mathcal{A}_0$ contains a proper ideal, then (by Zorn) it contains maximal ideals, whence (by the previous proposition) it contains non-zero non-invertible elements.

$(ii) \Rightarrow (iii)$: Let $x \in \mathcal{A}$; pick $\lambda \in \sigma(x)$; this is possible, in view of the already established fact that the spectrum is always non-empty; then $(x - \lambda)$ is not invertible and is consequently contained in the ideal $\mathcal{A}(x - \lambda) \neq \mathcal{A}$; deduce from the hypothesis (ii) that $\mathcal{A}(x - \lambda) = \{0\}$, i.e., $x = \lambda 1$.

$(iii) \Rightarrow (i)$: Obvious. $\square$

The conclusion of the Gelfand-Naimark theorem is, loosely speaking, that commutative unital Banach algebras are 'like' the algebra of continuous functions on a compact Hausdorff space. The first step in this direction is the identification of the points in the underlying compact space.

DEFINITION 3.2.6 *(a) A* **complex homomorphism** *on a commutative Banach algebra $\mathcal{A}$ is a mapping $\phi : \mathcal{A} \to \mathbb{C}$ which is a non-zero algebra homomorphism - i.e., ϕ satisfies the following conditions:*

(i) $\phi(\alpha x + y) = \alpha\phi(x) + \phi(y)$, $\phi(xy) = \phi(x)\phi(y)$ $\forall\ x, y \in \mathcal{A}, \alpha \in \mathbb{C}$;

(ii) ϕ is not identically equal to 0.

The collection of all complex homomorphisms on $\mathcal{A}$ is called the **spectrum** *of $\mathcal{A}$ - for reasons that will become clear later - and is denoted by $\hat{\mathcal{A}}$.*

(b) For $x \in \mathcal{A}$, we shall write $\hat{x}$ for the function $\hat{x} : \hat{\mathcal{A}} \to \mathbb{C}$, defined by $\hat{x}(\phi) = \phi(x)$.

Remark 3.2.7 (a) It should be observed that for unital Banach algebras, condition (ii) in Definition 3.2.6 is equivalent to the requirement that $\phi(1) = 1$. (Verify this!)

(b) Deduce, from Exercise 3.2.3(5), that in case $\mathcal{A} = C(X)$, with X a compact Hausdorff space, then there is an identification of $\hat{\mathcal{A}}$ with X in such a way that $\hat{f}$ is identified with f, for all $f \in \mathcal{A}$. □

Lemma 3.2.8 *Let $\mathcal{A}$ denote a unital commutative Banach algebra.*

(a) the mapping $\phi \to$ ker ϕ sets up a bijection between $\hat{\mathcal{A}}$ and the collection of all maximal ideals in $\mathcal{A}$;

(b) $\hat{x}(\hat{\mathcal{A}}) = \sigma(x)$;

(c) if it is further true that $||1|| = 1$, then, the following conditions on a map $\phi : \mathcal{A} \to \mathbb{C}$ are equivalent:

(i) $\phi \in \hat{\mathcal{A}}$;

(ii) $\phi \in \mathcal{A}^*$, $||\phi|| = 1$, and $\phi(xy) = \phi(x)\phi(y) \; \forall \; x, y \in \mathcal{A}$.

Proof : (a) Let $\phi \in \hat{\mathcal{A}}$ and define $\mathcal{I} = ker \; \phi$. Since ϕ is an algebra homomorphism of $\mathcal{A}$ onto $\mathbb{C}$, it follows - see Exercise 3.2.3(7) - that $\mathcal{I}$ is a maximal ideal in $\mathcal{A}$. Conversely, if $\mathcal{I}$ is a maximal ideal in $\mathcal{A}$, it follows from Exercise 3.2.3(7) that $\mathcal{A}/\mathcal{I}$ is simple, and hence, by the Gelfand-Mazur Theorem, we see that $\mathcal{A}/\mathcal{I} = \mathbb{C}1$; consequently, the quotient map $\mathcal{A} \to \mathcal{A}/\mathcal{I} = \mathbb{C}$ gives rise to a complex homomorphism $\phi_{\mathcal{I}}$ which clearly has the property that $ker \; \phi_{\mathcal{I}} = \mathcal{I}$. Finally, the desired conclusion follows from the following facts: (i) two linear functionals on a vector space with the same kernel are multiples of one another, and (ii) if $\phi \in \hat{\mathcal{A}}$, then $\mathcal{A} = ker \; \phi \oplus \mathbb{C}1$ (as vector spaces), and $\phi(1) = 1$.

(b) If $x \in \mathcal{A}, \phi \in \hat{\mathcal{A}}$, then $(x - \phi(x)1) \in ker \; \phi$, and hence, by Proposition 3.2.4, we see that $\phi(x) \in \sigma(x)$. Conversely, it follows from Proposition 3.2.4 and (a) of this Lemma, that $\sigma(x) \subset \hat{x}(\hat{\mathcal{A}})$.

(c) Deduce from (b) and equation 3.1.11 that if $\phi \in \hat{\mathcal{A}}$, then, for any $x \in \mathcal{A}$, we have

$$|\phi(x)| \;\leq\; r(x) \;\leq\; ||x|| \;,$$

and hence $||\phi|| \leq 1$. On the other hand, since $\phi(1) = 1$, we must have $||\phi|| = 1$, and hence $(i) \Rightarrow (ii)$; the other implication is obvious. □

COROLLARY 3.2.9 *Let $\mathcal{A}$ be a non-unital commutative Banach algebra; then $\phi \in \hat{\mathcal{A}} \Rightarrow \phi \in \mathcal{A}^*$ and $||\phi|| \leq 1$.*

Proof : Let $\mathcal{A}^+$ denote the 'unitisation' of $\mathcal{A}$ as in Exercise 3.1.3(1). Define $\phi^+ : \mathcal{A}^+ \to \mathbb{C}$ by $\phi^+(x, \alpha) = \phi(x) + \alpha$; it is easily verified that $\phi^+ \in \hat{\mathcal{A}^+}$; hence, by Lemma 3.2.8(c), we see that $\phi^+ \in (\mathcal{A}^+)^*$ and $||\phi^+|| \leq 1$, and the desired conclusion follows easily. □

In the sequel, given a commutative Banach algebra $\mathcal{A}$, we shall regard $\hat{\mathcal{A}}$ as a topological space with respect to the subspace topology it inherits from *ball* $\mathcal{A}^* = \{\phi \in \mathcal{A}^* : ||\phi|| \leq 1\}$, where this unit ball is considered as a compact Hausdorff space with respect to the weak*-topology - see Alaoglu's theorem (Theorem 1.6.9).

THEOREM 3.2.10 *Let $\mathcal{A}$ be a commutative Banach algebra, with spectrum $\hat{\mathcal{A}}$. Then,*

(1) $\hat{\mathcal{A}}$ is a locally compact Hausdorff space (with respect to the weak-topology), which is compact in case the Banach algebra $\mathcal{A}$ contains an identity; and*

(2) the map $x \to \hat{x}$ defines a mapping $\Gamma : \mathcal{A} \to C_0(\hat{\mathcal{A}})$ which is a contractive homomorphism of Banach algebras (where, of course, $C_0(\hat{\mathcal{A}})$ is regarded as a Banach algebra as in Example 3.1.2(4)) - meaning that the following relations hold, for all $x, y \in \mathcal{A}, \alpha \in \mathbb{C}$:

(i) $||\Gamma(x)|| \leq ||x||$;

(ii) $\Gamma(\alpha x + y) = \alpha\Gamma(x) + \Gamma(y)$; and

(iii) $\Gamma(xy) = \Gamma(x)\Gamma(y)$.

The map Γ is referred to as the **Gelfand transform** *of $\mathcal{A}$.*

Proof : (1) Consider the following sets: for fixed $x, y \in \mathcal{A}$, let $K_{x,y} = \{\phi \in ball\ \mathcal{A}^* : \phi(xy) = \phi(x)\phi(y)\}$; and let $V = \{\phi \in ball\ \mathcal{A}^* : \phi \neq 0\}$. The definition of the weak* topology implies that $K_{x,y}$ (resp., V) is a closed (resp., open) subset of *ball* $\mathcal{A}^*$; hence $K = \cap_{x,y \in \mathcal{A}} K_{x,y}$ is also a closed, hence compact, set in the weak* topology. Notice now that $\hat{\mathcal{A}} = K \cap V$, and since an open subset of a (locally) compact Hausdorff space is locally compact - see Proposition A.6.2(3) - and Hausdorff, we see that

$\hat{\mathcal{A}}$ is indeed a locally compact Hausdorff space. In case $\mathcal{A}$ has an identity, then $F = \{\phi \in ball\ \mathcal{A}^* : \phi(1) = 1\}$ is weak* closed, and $\hat{\mathcal{A}} = K \cap F$ is a closed subset of *ball* $\mathcal{A}^*$, and the proof of (1) is complete.

(2) The definition of the weak* topology guarantees that $\hat{x}$ is a continuous function on $\hat{\mathcal{A}}$, for every $x \in \mathcal{A}$; it is also obvious that the mapping $\Gamma : x \mapsto \hat{x}$ is linear and preserves products; (for example, if $x, y \in \mathcal{A}$ and if $\phi \in \hat{\mathcal{A}}$ is arbitrary, then (by the definition of a complex homomorphism) we have: $\widehat{xy}(\phi) = \phi(xy) = \phi(x)\phi(y) = \hat{x}(\phi)\hat{y}(\phi)$, whence $\widehat{xy} = \hat{x}\hat{y}$).

To complete the proof, we need to verify that Γ maps $\mathcal{A}$ into $C_0(\hat{\mathcal{A}})$, and that $||\Gamma(x)|| \leq ||x||$.

First consider the case when $\mathcal{A}$ has an identity. In this case, $C_0(\hat{\mathcal{A}}) = C(\hat{\mathcal{A}})$, while, by Lemma 3.2.8(b), we find that

$$\begin{aligned} ||\Gamma(x)|| &= \sup\{|\hat{x}(\phi)| : \phi \in \hat{\mathcal{A}}\} \\ &= \sup\{|\lambda| : \lambda \in \sigma(x)\} \\ &= r(x) \\ &\leq ||x|| \ , \end{aligned}$$

and the proof of the unital case is complete.

Now suppose $\mathcal{A}$ does not have an identity. To start with, deduce, from Corollary 3.2.9, that Γ is indeed contractive, so we only need to verify that functions on $(\mathcal{A}^+)\hat{}$ of the form $\hat{x}$ do indeed vanish at infinity.

Let $\mathcal{A}^+$ be the unitisation of $\mathcal{A}$, and let $\hat{\mathcal{A}} \ni \phi \stackrel{f}{\mapsto} \phi^+ \in (\mathcal{A}^+)\hat{}$ be as in the proof of Corollary 3.2.9. Notice now that $(\mathcal{A}^+)\hat{} = f(\hat{\mathcal{A}}) \cup \{\phi_0\}$, where $\phi_0(x, \alpha) = \alpha$, since ϕ_0 is the unique complex homomorphism of $\mathcal{A}^+$ whose restriction to $\mathcal{A}$ is identically equal to 0. On the other hand, (the definitions of) the topologies on $\hat{\mathcal{A}}$ and $(\mathcal{A}^+)\hat{}$ show that a net $\{\phi_i : i \in I\}$ converges to ϕ in $\hat{\mathcal{A}}$ if and only if $\{\phi_i^+ : i \in I\}$ converges to ϕ^+ in $\hat{\mathcal{A}}$; in other words, the function f maps $\hat{\mathcal{A}}$ homeomorphically onto $f(\hat{\mathcal{A}})$.

Thus, we find - from the compactness of $(\mathcal{A}^+)\hat{}$, and from the nature of the topologies on the spaces concerned - that we may identify $(\mathcal{A}^+)\hat{}$ with the one-point compactification $(\hat{\mathcal{A}})^+$ of the locally compact Hausdorff space $\hat{\mathcal{A}}$, with the element ϕ_0 playing the role of the point at infinity; notice now that if we identify $\hat{\mathcal{A}}$ as a subspace of $(\mathcal{A}^+)\hat{}$ (via f), then the mapping $\widehat{(x, 0)}$ is a continuous functon on $(\mathcal{A}^+)\hat{}$ which extends $\hat{x}$ and which vanishes at '∞' (since $\phi_0(x,0) = 0$); thus, (by Proposition A.6.7(a), for instance) we find that indeed $\hat{x} \in C_0((\mathcal{A}^+)\hat{}), \ \forall x \in \mathcal{A}$, and the proof is complete. □

COROLLARY 3.2.11 *Let Γ be the Gelfand transform of a commutative Banach algebra $\mathcal{A}$. The following conditions on an element $x \in \mathcal{A}$ are equivalent:*
(i) $x \in ker\ \Gamma$ – i.e., $\Gamma(x) = 0$;
(ii) $\lim_n ||x^n||^{\frac{1}{n}} = 0$.

In case $\mathcal{A}$ has an identity, these conditions are equivalent to the following two conditions also:
(iii) $x \in \mathcal{I}$, for every maximal ideal $\mathcal{I}$ of $\mathcal{A}$;
(iv) $\sigma(x) = \{0\}$.

Proof : Suppose first that $\mathcal{A}$ has an identity. Then, since $||\Gamma(x)|| = r(x)$, the equivalence of the four conditions follows easily from Proposition 3.2.4, Lemma 3.2.8 and the spectral radius formula.

For non-unital $\mathcal{A}$, apply the already established unital case of the corollary to the element $(x,0)$ of the unitised algebra $\mathcal{A}^+$. □

DEFINITION 3.2.12 *An element of a commutative Banach algebra $\mathcal{A}$ is said to be* **quasi-nilpotent** *if it satisfies the equivalent conditions of Corollary 3.2.11. The* **radical** *of $\mathcal{A}$ is defined to be the ideal consisting of all quasinilpotent elements of $\mathcal{A}$. A commutative Banach algebra is said to be* **semi-simple** *if it has trivial radical, (or equivalently, it has no quasinilpotent elements, which is the same as saying that the Gelfand transform of $\mathcal{A}$ is injective).*

A few simple facts concerning the preceding definitions are contained in the following exercises.

EXERCISE 3.2.13 *(1) Consider the matrix $N \in M_n(\mathbb{C})$ defined by $N = ((n^i_j))$, where $n^i_j = 1$ if $i = j - 1$, and $n^i_j = 0$ otherwise. (Thus, the matrix has 1's on the 'first super-diagonal' and has 0's elsewhere.) Show that $N^k = 0 \Leftrightarrow k \geq n$, and hence* *deduce that if* $\mathcal{A} = \{\sum_{i=1}^{n-1} \alpha_i N^i : \alpha_i \in \mathbb{C}\}$ *(is the subalgebra of* $M_n(\mathbb{C})$ *which is generated by* $\{N\}$*), then* $\mathcal{A}$ *is a commutative non-unital Banach algebra, all of whose elements are (nilpotent, and consequently) quasi-nilpotent.*

(2) If $\mathcal{A}$ *is as in (1) above, show that* $\hat{\mathcal{A}}$ *is empty, and consequently the requirement that* $\hat{\mathcal{A}}$ *is compact does not imply that* $\mathcal{A}$ *must have an identity.*

(3) Show that every nilpotent element (i.e., an element with the property that some power of it is 0) in a commutative Banach algebra is necessarily quasi-nilpotent, but that a quasi-nilpotent element may not be nilpotent. (Hint: For the second part, let N_n *denote the operator considered in (1) above, and consider the operator given by* $T = \oplus \frac{1}{n} N_n$ *acting on* $\oplus_{n=1}^{\infty} \mathbb{C}^n$*.)*

We now look at a couple of the more important applications of the Gelfand-Naimark theorem in the following examples.

EXAMPLE 3.2.14 (1) Suppose $\mathcal{A} = C(X)$, where X is a compact Hausdorff space. It follows from Exercise 3.2.3(5)(d) that the typical complex homomorphism of $\mathcal{A}$ is of the form $\phi_x(f) = f(x)$, for some $x \in X$. Hence the map $X \ni x \to \phi_x \in \hat{\mathcal{A}}$ is a bijective map which is easily seen to be continuous (since $x_i \to x \Rightarrow f(x_i) \to f(x)$ for all $f \in \mathcal{A}$); since X is a compact Hausdorff space, it follows that the above map is actually a homeomorphism, and if $\hat{\mathcal{A}}$ is identified with X via this map, then the Gelfand transform gets identified with the identity map on $\mathcal{A}$.

(2) Let $\mathcal{A} = \ell^1(\mathbb{Z})$, as in Example 3.1.2(5). Observe that $\mathcal{A}$ is 'generated, as a Banach algebra' by the element e_1, which is the basis vector indexed by 1; in fact, $e_1^n = e_n \ \forall n \in \mathbb{Z}$, and $x = \sum_{n \in \mathbb{Z}} x(n) e_n$, the series converging (unconditionally) in the norm. If $\phi \in \hat{\mathcal{A}}$ and $\phi(e_1) = z$, then note that for all $n \in \mathbb{Z}$, we have $|z^n| = |\phi(e_n)| \leq 1$, since $||e_n|| = 1 \ \forall n$; this clearly forces $|z| = 1$. Conversely, it is easy to show that if $|z| = 1$, then there exists a unique element $\phi_z \in \hat{\mathcal{A}}$ such that $\phi_z(e_1) = z$. (Simply define $\phi_z(x) = \sum_{n \in \mathbb{Z}} x(n) z^n$, and verify that this defines a complex homomorphism.) Thus we find that $\hat{\mathcal{A}}$ may be identified with the unit circle $\mathbb{T}$; notice, incidentally, that since $\mathbb{Z}$ is the infinite cyclic group with generator 1, we also have

a natural identification of $\mathbb{T}$ with the set of all homomorphisms of the group $\mathbb{Z}$ into the multiplicative group $\mathbb{T}$; and if $\mathbb{Z}$ is viewed as a locally compact group with respect to the discrete topology, then all homomorphisms on $\mathbb{Z}$ are continuous.

More generally, the preceding discussion has a beautiful analogue for general locally compact Hausdorff abelian groups. If G is such a group, let $\hat{G}$ denote the set of continuous homomorphisms from G into $\mathbb{T}$; for $\gamma \in \hat{G}$, consider the equation $\phi_\gamma(f) = \int_G f(t)\overline{\gamma(t)}dm_G(t)$, where m_G denotes (some choice, which is unique up to normalisation of) a (left = right) Haar measure on G - see Example 3.1.2(5). (The complex conjugate appears in this equation for historic reasons, which ought to become clearer at the end of this discussion.) It is then not hard to see that ϕ_γ defines a complex homomorphism of the Banach algebra $L^1(G, m_G)$. The pleasant fact is that every complex homomorphism of $L^1(G, m_G)$ arises in this manner, and that the map $\hat{G} \ni \gamma \mapsto \phi_\gamma \in (L^1(G, m_G))\hat{}$ is a bijective correspondence.

The (so-called **Pontrjagin duality**) theory goes on to show that if the (locally compact Hausdorff) topology on $L^1(\widehat{G, m_G})$ is transported to $\hat{G}$, then $\hat{G}$ acquires the structure of a locally compact Hausdorff group, which is referred to as the **dual group** of G. (What this means is that (a) $\hat{G}$ is a group with respect to pointwise operations - i.e., $(\gamma_1 \cdot \gamma_2)(t) = \gamma_1(t)\gamma_2(t)$, etc; and (b) this group structure is 'compatible' with the topology inherited from $\hat{\mathcal{A}}$, in the sense that the group operations are continuous.) The elements of $\hat{G}$ are called **characters** on G, and $\hat{G}$ is also sometimes referred to as the **character group** of G.

The reason for calling $\hat{G}$ the 'dual' of G is this: each element of G defines a character on $\hat{G}$ via evaluation, thus: if we define $\chi_t(\gamma) = \gamma(t)$, then we have a mapping $G \ni t \mapsto \chi_t \in \hat{\hat{G}} = (\hat{G})\hat{}$; the theory goes on to show that the above mapping is an isomorphism of groups which is a homeomorphism of topological spaces; thus we have an identification $G \cong \hat{\hat{G}}$ of topological groups.

Under the identification of $(L^1(G, m_G))\hat{}$ with $\hat{G}$, the Gelfand transform becomes identified with the mapping $L^1(G, m_G) \ni f \mapsto \hat{f} \in C_0(\hat{G})$ defined by

$$\hat{f}(\gamma) = \int_G f(t)\overline{\gamma(t)}dm_G(t) \; ; \qquad (3.2.15)$$

the mapping $\hat{f}$ defined by this equation is referred to, normally, as the **Fourier transform** of the function f - for reasons stemming from specialisations of this theory to the case when $G \in \{\mathbb{R}, \mathbb{Z}, \mathbb{T}\}$.

Thus, our earlier discussion shows that $\hat{\mathbb{Z}} = \mathbb{T}$ (so that, also $\hat{\mathbb{T}} = \mathbb{Z}$). (It is a fact that, in general, G is discrete if and only if $\hat{G}$ is compact.) It is also true that the typical element of $\hat{\mathbb{R}}$ is of the form $\gamma_t(s) = exp(its)$, so that $\hat{\mathbb{R}}$ may be identified with $\mathbb{R}$ as a topological group. Further, if we choose the right normalisations of Haar measure, we find the equations:

$$\begin{aligned}\hat{f}(z) &= \sum_{n \in \mathbb{Z}} f(n)\overline{z}^n, \ \forall\ f \in \ell^1(\mathbb{Z}) \\ \hat{f}(n) &= \frac{1}{2\pi}\int_{[0,2\pi]} f(e^{i\theta})e^{-in\theta}d\theta\ ,\ \forall\ f \in L^1(\mathbb{T}, m_{\mathbb{T}}) \\ \hat{f}(x) &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f(y)e^{-ixy}dy\ ,\ \forall\ f \in L^1(\mathbb{R}, m_{\mathbb{R}})\end{aligned}$$

which are the equations defining the classical Fourier transforms. It is because of the occurrence of the complex conjugates in these equations, that we also introduced a complex conjugate earlier, so that it could be seen that the Gelfand transform specialised, in this special case, to the Fourier transform. □

3.3 Commutative C*-algebras

By definition, the Gelfand transform of a commutative Banach algebra is injective precisely when the algebra in question is semi-simple. The best possible situation would, of course, arise when the Gelfand transform turned out to be an isometric algebra isomorphism onto the appropriate function algebra. For this to be true, the Banach algebra would have to behave 'like' the algebra $C_0(X)$; in particular, in the terminology of the following definition, such a Banach algebra would have to possess the structure of a commutative C^*-algebra.

DEFINITION 3.3.1 *A* **C*-algebra** *is, by definition, a Banach algebra* $\mathcal{A}$*, which is equipped with an 'involution'* $\mathcal{A} \ni x \mapsto x^* \in \mathcal{A}$ *which satisfies the following conditions, for all* $x, y \in \mathcal{A}, \alpha \in \mathbb{C}$*:*

(i) (conjugate-linearity) $(\alpha x + y)^* = \overline{\alpha} x^* + y^*$*;*
(ii) (product-reversal) $(xy)^* = y^* x^*$*;*
(iii) (order two) $(x^*)^* = x$*; and*
(iv) (C^*-identity) $||x^* x|| = ||x||^2$*.*

The element x^* is called the **adjoint** *of the element x, and the mapping $x \mapsto x^*$ is referred to as the 'adjoint-map' or as 'adjunction'.*

EXAMPLE 3.3.2 (1) An example of a commutative C^*-algebra is furnished by $C_0(X)$, where X is a locally compact Hausdorff space; it is clear that this algebra has an identity precisely when X is compact.

(2) An example of a non-commutative C^*-algebra is provided by $\mathcal{L}(\mathcal{H})$, where $\mathcal{H}$ is a Hilbert space of dimension at least two. (Why two?)

(3) If $\mathcal{A}$ is a C^*-algebra, and if $\mathcal{B}$ is a subalgebra which is closed (with respect to the norm) and is 'self-adjoint' - meaning that it is closed under the mapping $x \mapsto x^*$ - then $\mathcal{B}$ has the structure of a C^*-algebra in its own right. In this case, we shall call $\mathcal{B}$ a C^*-subalgebra of $\mathcal{A}$.

(4) This is a non-example. If G is a locally compact abelian group - or more generally, if G is a locally compact group whose left Haar measure is also invariant under right translations - define an involution on $L^1(G, m_G)$ by $f^*(t) = \overline{f(t^{-1})}$; then this is an involution of the Banach algebra which satisfies the first three algebraic requirements of Definition 3.3.1, but does NOT satisfy the crucial fourth condition; that is, the norm is not related to the involution by the C^*-identity; nevertheless the involution is still not too badly behaved with respect to the norm, since it is an isometric involution. (We ignore the trivial case of a group with only one element, which is the only case when the C^*-identity is satisfied.) $\Box$

So as to cement these notions and get a few simple facts enunciated, we pause for an exercise.

EXERCISE 3.3.3 *(1) Let $\mathcal{A}$ be a Banach algebra equipped with an involution $x \mapsto x^*$ which satisfies the first three conditions of Definition 3.3.1.*

(a) Show that if $\mathcal{A}$ is a C^*-algebra, then $||x^*|| = ||x|| \; \forall \, x \in \mathcal{A}$. (Hint: Use the submultiplicativity of the norm in any Banach algebra, together with the C^*-identity.)

(b) Show that $\mathcal{A}$ is a C^*-algebra if and only if $||x||^2 \leq ||x^*x||$ for all $x \in \mathcal{A}$. (Hint: Use the submultiplicativity of the norm in any Banach algebra, together with (a) above.)

(2) An element x of a C^*-algebra is said to be **self-adjoint** *if $x = x^*$.*

(a) Show that any element x of a C^*-algebra is uniquely expressible in the form $x = x_1 + ix_2$, where x_1 and x_2 are self-adjoint elements. (Hint: Imitate the proof of Proposition 2.4.9.)

(b) If $\mathcal{A}$ is a C^*-algebra which has an identity 1, show that 1 is necessarily self-adjoint.

(3) Let $\mathcal{A}$ be a C^*-algebra, and let $\mathcal{S}$ be an arbitrary subset of $\mathcal{A}$. (This exercise gives an existential, as well as a constructive, description of the 'C^*-subalgebra generated by $\mathcal{S}$'.)

(a) Show that $\cap\{\mathcal{B} : \mathcal{S} \subset \mathcal{B}, \mathcal{B}$ is a C^*-subalgebra of $\mathcal{A}\}$ is a C^*-subalgebra of $\mathcal{A}$ that contains $\mathcal{S}$, which is minimal with respect to this property, in the sense that it is contained in any other C^*-subalgebra of $\mathcal{A}$ which contains $\mathcal{S}$; this is called the C^*-subalgebra generated by $\mathcal{S}$ and is denoted by $C^*(\mathcal{S})$.

(b) Let $\mathcal{S}_1$ denote the set of all 'words in $\mathcal{S} \cup \mathcal{S}^*$'; i.e., $\mathcal{S}_1$ consists of all elements of the form $w = a_1 a_2 \cdots a_n$, where n is any positive integer and either $a_i \in \mathcal{S}$ or $a_i^* \in \mathcal{S}$ for all i; show that $C^*(\mathcal{S})$ is the closure of the set of all finite linear combinations of elements of $\mathcal{S}_1$.

In order to reduce assertions about non-unital C^*-algebras to those about unital C^*-algebras, we have to know that we can embed any non-unital C^*-algebra as a maximal ideal in a unital C^*-algebra (just as we did for Banach algebras). The problem with our earlier construction is that the norm we considered earlier (in Exercise 3.1.3(1)) will usually not satisfy the crucial C^*-identity. We get around this difficulty as follows.

PROPOSITION 3.3.4 *Suppose $\mathcal{A}$ is a C^*-algebra without identity. Then there exists a C^*-algebra $\mathcal{A}^+$ with identity which contains a maximal ideal which is isomorphic to $\mathcal{A}$ as a C^*-algebra.*

Proof : Consider the mapping $\mathcal{A} \ni x \mapsto L_x \in \mathcal{L}(\mathcal{A})$ defined by

$$L_x(y) = xy \; \forall \; y \in \mathcal{A} \; .$$

It is a consequence of sub-multiplicativity of the norm (with respect to products) in any Banach algebra, and the C^*-identity that this map is an isometric algebra homomorphism of $\mathcal{A}$ into $\mathcal{L}(\mathcal{A})$.

Let $\mathcal{A}^+ = \{L_x + \alpha 1 : x \in \mathcal{A}, \alpha \in \mathbb{C}\}$, where we write 1 for the identity operator on $\mathcal{A}$; as usual, we shall simply write α for $\alpha 1$. It is elementary to verify that $\mathcal{A}^+$ is a unital subalgebra of $\mathcal{L}(\mathcal{A})$; since this subalgebra is expressed as the vector sum of the (complete, and hence) closed subspace $\mathcal{I} = \{L_x : x \in \mathcal{A}\}$ and the one-dimensional subspace $\mathbb{C}1\}$, it follows from Exercise 3.3.5(1) below that $\mathcal{A}^+$ is a closed subspace of the complete space $\mathcal{L}(\mathcal{A})$ and thus $\mathcal{A}^+$ is indeed a Banach algebra. (Also notice that $\mathcal{I}$ is a maximal ideal in $\mathcal{A}^+$ and that the mapping $x \mapsto L_x$ is an isometric algebra isomorphism of $\mathcal{A}$ onto the Banach algebra $\mathcal{I}$.)

Define the obvious involution on $\mathcal{A}^+$ by demanding that $(L_x + \alpha)^* = L_{x^*} + \overline{\alpha}$. It is easily verified that this involution satisfies the conditions $(i) - (iii)$ of Definition 3.3.1.

In view of Exercise 3.3.3(1)(a), it only remains to verify that $||z^*z|| \geq ||z||^2$ for all $z \in \mathcal{A}^+$. We may, and do, assume that $z \neq 0$, since the other case is trivial. Suppose $z = L_x + \alpha$; then $z^*z = L_{x^*x + \overline{\alpha}x + \alpha x^*} + |\alpha|^2$. Note that for arbitrary $y \in \mathcal{A}$, we have

$$\begin{aligned} ||z(y)||^2 &= ||xy + \alpha y||^2 \\ &= ||(y^*x^* + \overline{\alpha}y^*)(xy + \alpha y)|| \\ &= ||y^* \cdot \; (z^*z)(y)|| \\ &\leq ||z^*z|| \cdot ||y||^2 \; , \end{aligned}$$

where we have used the C^*-identity (which is valid in $\mathcal{A}$) in the second line, and the submultiplicativity of the norm (with respect to products), the definition of the norm on $\mathcal{A}^+$, and Exercise 3.3.3(1)(b) in the last line. Since y was arbitrary, this shows that, indeed, we have $||z||^2 \leq ||z^*z||$, and the proof is complete. $\square$

EXERCISE 3.3.5 *(1) Show that if Y is a closed subspace of a Banach space X, and if F is a finite-dimensional subspace of X, then also the vector sum $Y + F$ is a closed subspace of X. (Hint: Consider the quotient map $\pi : X \to X/Y$, note that $F + Y = \pi^{-1}(\pi(F))$, and appeal to Exercise A.6.5(1)(c) and the continuity of π.)*

(2) Show that if, in (1) above, the requirement that F is finite-dimensional is relaxed to just requiring that F is closed, then the conclusion is no longer valid in general. (Hint: Let $T \in \mathcal{L}(\mathcal{H})$ be an operator which is 1-1, but whose range is not closed - (see Remark 1.5.15, for instance); let $X = \mathcal{H} \oplus \mathcal{H}, Y = \mathcal{H} \oplus \{0\}$, and $F = G(T)$ (the graph of T).)

(3) If $\mathcal{A}$ is any (not necessarily commutative) unital Banach algebra, and if $x \in \mathcal{A}$, show that the equation

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \tag{3.3.16}$$

defines a mapping $\mathcal{A} \ni x \stackrel{exp}{\mapsto} e^x \in \mathcal{G}(\mathcal{A})$ with the property that

(i) $e^{x+y} = e^x e^y$, whenever x and y are two elements which commute with one another;

(ii) for each fixed $x \in \mathcal{A}$, the map $\mathbb{R} \ni t \mapsto e^{tx} \in \mathcal{G}(\mathcal{A})$ defines a (continuous) homomorphism from the additive group $\mathbb{R}$ into the multiplicative (topological) group $\mathcal{G}(\mathcal{A})$.

We are now ready to show that in order for the Gelfand transform of a commutative Banach algebra to be an isometric isomorphism onto the algebra of all continuous functions on its spectrum which vanish at infinity, it is necessary and sufficient that the algebra have the structure of a commutative C^*-algebra. Since the necessity of the condition is obvious, we only state the sufficiency of the condition in the following formulation of the **Gelfand-Naimark theorem for commutative C^*-algebras**.

THEOREM 3.3.6 *Suppose $\mathcal{A}$ is a commutative C^*-algebra; then the Gelfand transform is an isometric *-algebra isomorphism of $\mathcal{A}$ onto $C_0(\hat{\mathcal{A}})$; further, $\hat{\mathcal{A}}$ is compact if and only if $\mathcal{A}$ has an identity.*

Proof : *Case (a):* $\mathcal{A}$ has an identity.

In view of Theorem 3.2.10, we only need to show that (i) if $x \in \mathcal{A}$ is arbitrary, then $\Gamma(x^*) = \Gamma(x)^*$ and $||\Gamma(x)|| = ||x||$; (ii) $\hat{\mathcal{A}}$ is compact; and (iii) Γ maps onto $C(\hat{\mathcal{A}})$.

(i) We first consider the case of a self-adjoint element $x = x^* \in \mathcal{A}$. Define $u_t = e^{itx}$, $\forall t \in \mathbb{R}$; then, notice that the self-adjointness assumption, and the continuity of adjunction, imply that

$$\begin{aligned} u_t^* &= \sum_{n=0}^{\infty} \left(\frac{(itx)^n}{n!} \right)^* \\ &= \sum_{n=0}^{\infty} \left(\frac{(-itx)^n}{n!} \right) \\ &= u_{-t} \ ; \end{aligned}$$

hence, by the C^*-identity and Exercise 3.3.5(3), we find that for any $t \in \mathbb{R}$,

$$||u_t||^2 \ = \ ||u_{-t}u_t|| \ = \ 1 \ .$$

Now, if $\phi \in \hat{\mathcal{A}}$, since $||\phi|| = 1$, it follows that $|\phi(u_t)| \leq 1 \ \forall t \in \mathbb{R}$; on the other hand, since ϕ is (continuous and) multiplicative, it is seen that $\phi(u_t) = e^{it\phi(x)}$; in other words, $|e^{it\phi(x)}| \leq 1 \ \forall t \in \mathbb{R}$; this clearly implies that $\phi(x) \in \mathbb{R}$. Thus, we have shown that $\hat{x}$ is a real-valued function, for every self-adjoint element $x \in \mathcal{A}$. Thanks to the Cartesian decomposition (of an element as a sum of a self-adjoint and a 'skew-adjoint' element), we may now conclude that Γ is indeed a homomorphism of *-algebras.

Again, if $x = x^* \in \mathcal{A}$, observe that

$$||x||^2 \ = \ ||x^*x|| \ = \ ||x^2|| \ ,$$

and conclude, by an easy induction, that $||x||^{2^n} = ||x^{2^n}||$ for every positive integer n; it follows from the spectral radius formula that

$$r(x) \ = \ \lim_n ||x^{2^n}||^{\frac{1}{2^n}} \ = \ ||x|| \ ;$$

on the other hand, we know that $||\Gamma(x)|| = r(x)$ - see Lemma 3.2.8(a); thus, for self-adjoint x, we find that indeed $||\Gamma(x)|| = ||x||$; the case of general x follows from this, the fact that Γ is a *-homomorphism, and the C^*-identity (in both $\mathcal{A}$ and $C(\hat{\mathcal{A}})$), thus :

$$\begin{aligned} ||\Gamma(x)||^2 &= ||\Gamma(x)^*\Gamma(x)|| \\ &= ||\Gamma(x^*x)|| \\ &= ||x^*x|| \\ &= ||x||^2 , \end{aligned}$$

thereby proving (i).

(ii) This follows from Theorem 3.2.10 (1).

(iii) It follows from the already established (i) that the $\Gamma(\mathcal{A})$ is a norm-closed self-adjoint subalgebra of $C(\hat{\mathcal{A}})$ which is easily seen to contain the constants and to separate points of $\hat{\mathcal{A}}$; according to the Stone-Weierstrass theorem, the only such subalgebra is $C(\hat{\mathcal{A}})$.

Case (b): $\mathcal{A}$ does not have a unit.

Let $\mathcal{A}^+$ and $\mathcal{I}$ be as in the proof of Proposition 3.3.4. By the already established Case (a) above, and Exercise 3.2.3(5)(d), we know that $\Gamma_{\mathcal{A}^+}$ is an isometric *-isomorphism of $\mathcal{A}^+$ onto $C(\hat{\mathcal{A}})$, and that there exists a point $\phi_0 \in (\mathcal{A}^+)\hat{}$ such that $\Gamma_{\mathcal{A}^+}(\mathcal{I}) = \{f \in C(\hat{\mathcal{A}}) : f(\phi_0) = 0\}$. As in the proof of Theorem 3.2.10, note that if $\pi : \mathcal{A} \to \mathcal{I}$ is the natural isometric *-isomorphism of $\mathcal{A}$ onto $\mathcal{I}$, then the map $\phi \mapsto \phi \circ \pi$ defines a bijective correspondence between $(\mathcal{A}^+)\hat{}-\{\phi_0\}$ and $\hat{\mathcal{A}}$ which is a homeomorphism (from the domain, with the subspace topology inherited from $(\mathcal{A}^+)\hat{}$, and the range); finally, it is clear that if $x \in \mathcal{A}$, and if $\phi_0 \neq \phi \in (\mathcal{A}^+)\hat{}$, then $(\Gamma_{\mathcal{A}^+}(\pi(x)))\,(\phi) = (\Gamma_{\mathcal{A}}(x))\,(\phi\circ\pi)$, and it is thus seen that even in the non-unital case, the Gelfand transform is an isometric *-isomorphism of $\mathcal{A}$ onto $C_0(\hat{\mathcal{A}})$.

Finally, if X is a locally compact Hausdorff space, then since $C_0(X)$ contains an identity precisely when X is compact, the proof of the theorem is complete. □

Before reaping some consequences of the powerful Theorem 3.3.6, we wish to first establish the fact that the notion of the spectrum of an element of a unital C^*-algebra is an 'intrinsic' one, meaning that it is independent of the ambient algebra, unlike the case of general Banach algebras (see Proposition 3.1.11 and Example 3.1.12).

PROPOSITION 3.3.7 *Let $\mathcal{A}$ be a unital C^*-algebra, let $x \in \mathcal{A}$, and suppose $\mathcal{B}$ is a C^*-subalgebra of $\mathcal{A}$ such that $1, x \in \mathcal{B}$. Then, $\sigma_{\mathcal{A}}(x) = \sigma_{\mathcal{B}}(x)$; in particular, or equivalently, if we let $\mathcal{A}_0 = C^*(\{1, x\})$, then, $\sigma_{\mathcal{A}}(x) = \sigma_{\mathcal{A}_0}(x)$.*

Proof : We already know that $\sigma_{\mathcal{B}}(x) \supset \sigma_{\mathcal{A}}(x)$ - see Proposition 3.1.11; to prove the reverse inclusion, we need to show that $\rho_{\mathcal{A}}(x) \subset \rho_{\mathcal{B}}(x)$; i.e., we need to show that if $\lambda \in \mathbb{C}$ is such that $(x - \lambda)$ admits an inverse in $\mathcal{A}$, then that inverse should actually belong to $\mathcal{B}$; in view of the spectral mapping theorem, we may assume, without loss of generality, that $\lambda = 0$; thus we have to prove the following assertion:

If $1, x \in \mathcal{B} \subset \mathcal{A}$, and if $x \in \mathcal{G}(\mathcal{A})$, then $x^{-1} \in \mathcal{B}$.

We prove this in two steps.

Case (i) : $x = x^*$.

In this case, we know from Theorem 3.3.6 - applied to the commutative unital C^*-algebra $\mathcal{A}_0 = C^*(\{1, x\})$, that $\sigma_{\mathcal{A}_0}(x) \subset \mathbb{R}$; since any closed subset of the real line is, when viewed as a subset of the complex plane, its own boundary, we may deduce from Proposition 3.1.11 (applied to the inclusions $\mathcal{A}_0 \subset \mathcal{B} \subset \mathcal{A}$) that

$$\sigma_{\mathcal{A}}(x) \subset \sigma_{\mathcal{B}}(x) \subset \sigma_{\mathcal{A}_0}(x) = \partial(\sigma_{\mathcal{A}_0}(x)) \subset \partial(\sigma_{\mathcal{A}}(x)) \subset \sigma_{\mathcal{A}}(x) \ ;$$

consequently, all the inclusions in the previous line are equalities, and the proposition is proved in this case.

Case (ii) : Suppose $x \in \mathcal{G}(\mathcal{A})$ is arbitrary; then also $x^* \in \mathcal{G}(\mathcal{A})$ and consequently, $x^*x \in \mathcal{G}(\mathcal{A})$; by Case (i), this implies that $(x^*x)^{-1} \in \mathcal{B}$; then $(x^*x)^{-1}x^*$ is an element in $\mathcal{B}$ which is left-inverse of x; since x is an invertible element in $\mathcal{A}$, deduce that $x^{-1} = (x^*x)^{-1}x^* \in \mathcal{B}$, and the proposition is completely proved. □

Thus, in the sequel, when we discuss C^*-algebras, we may talk unambiguously of the spectrum of an element of the algebra; in particular, if $T \in \mathcal{L}(\mathcal{H})$, and if $\mathcal{A}$ is any unital C^*-subalgebra of $\mathcal{L}(\mathcal{H})$ which contains T, then the spectrum of T, regarded as an element of $\mathcal{A}$, is nothing but $\sigma_{\mathcal{L}(\mathcal{H})}(T)$, which, of course, is given by the set $\{\lambda \in \mathbb{C} : T - \lambda \text{ is either not 1-1 or not onto } \}$.

Further, even if we have a non-unital C^*-algebra $\mathcal{A}$, we will talk of the spectrum of an element $x \in \mathcal{A}$, by which we will mean the set $\sigma_{\mathcal{A}^+}(x)$, where $\mathcal{A}^+$ is any unital C^*-algebra containing $\mathcal{A}$ (or equivalently, where $\mathcal{A}^+$ is **the** unitisation of $\mathcal{A}$, as in Proposition 3.3.4); it should be noticed that if x belongs to some C^*-algebra without identity, then its spectrum must necessarily contain 0.

Since we wish to obtain various consequences, for a general not necessarily commutative C^*-algebra, of Theorem 3.3.6 (which is a result about commutative C^*-algebras), we introduce a definition which permits us to regard appropriate commutative C^*-subalgebras of a general C^*-algebra.

DEFINITION 3.3.8 *An element x of a C^*-algebra is said to be* **normal** *if $x^*x = xx^*$.*

EXERCISE 3.3.9 *Let $\mathcal{A}$ be a C^*-algebra and let $x \in \mathcal{A}$. Show that the following conditions are equivalent:*

(i) x is normal;

(ii) $C^*(\{x\})$ is a commutative C^*-algebra;

(iii) if $x = x_1 + ix_2$ denotes the Cartesian decomposition of x, then $x_1x_2 = x_2x_1$.

We now come to the first of our powerful corollaries of Theorem 3.3.6; this gives us a **continuous functional calculus** for normal elements of a C^*-algebra.

PROPOSITION 3.3.10 *Let x be a normal element of a C^*-algebra $\mathcal{A}$, and define*

$$\mathcal{A}_0 = \begin{cases} C^*(\{1, x\}) & \textit{if } \mathcal{A} \textit{ has an identity } 1 \\ C^*(\{x\}) & \textit{otherwise} \end{cases} .$$

Then,

(a) if $\mathcal{A}$ has an identity, there exists a unique isometric isomorphism of C^*-algebras denoted by $C(\sigma(x)) \ni f \mapsto f(x) \in \mathcal{A}_0$ with the property that $f_1(x) = x$, where $f_1 : \sigma(x)\ (\subset \mathbb{C}) \to \mathbb{C}$ is the identity function defined by $f_1(z) = z$; and

(b) if $\mathcal{A}$ does not have an identity, and if we agree to use the (not incorrect) notation $C_0(\sigma(x)) = \{f \in C(\sigma(x)) : f(0) = 0\}$,

then there exists a unique isometric isomorphism of C^*-algebras denoted by $C_0(\sigma(x)) \ni f \mapsto f(x) \in \mathcal{A}_0$ with the property that $f_1(x) = x$, where f_1 is as in (a) above.

Proof : (a) Let $\Sigma = \sigma(x)$. Notice first that $\hat{x}$ $(= \Gamma_{\mathcal{A}_0}(x))$: $\hat{\mathcal{A}}_0 \to \Sigma$ is a surjective continuous mapping. We assert that this map is actually a homeomorhism; in view of the compactness of $\hat{\mathcal{A}}_0$ (and the fact that $\mathbb{C}$ is a Hausdorff space), we only need to show that $\hat{x}$ is 1-1; suppose $\hat{x}(\phi_1) = \hat{x}(\phi_2)$, $\phi_j \in \hat{\mathcal{A}}_0$; it follows that $\phi_1|_{\mathcal{D}} = \phi_2|_{\mathcal{D}}$, where $\mathcal{D} = \{\sum_{i,j=0}^n \alpha_{i,j} x^i (x^*)^j : n \in \mathbb{N}, \alpha_{i,j} \in \mathbb{C}\}$; on the other hand, the hypothesis (together with Exercise 3.3.3(3)(b)) implies that $\mathcal{D}$ is dense in $\mathcal{A}_0$, and the continuity of the ϕ_j's implies that $\phi_1 = \phi_2$, as asserted.

Hence $\hat{x}$ is indeed a homeomorphism of $\hat{\mathcal{A}}_0$ onto Σ. Consider the map given by

$$C(\Sigma) \ni f \ \mapsto \ \Gamma_{\mathcal{A}_0}^{-1}(f \circ \hat{x}) \ ;$$

it is easily deduced that this is indeed an isometric isomorphism of the C^*-algebra $C(\Sigma)$ onto $\mathcal{A}_0$ with the desired properties.

On the other hand, it follows from the Stone-Weierstrass theorem - see Exercise A.6.10(iii) - that the set $\mathcal{D}' = \{f : f(z) = \sum_{i,j=0}^n \alpha_{i,j} z^i \overline{z}^j \ , \ n \in \mathbb{N}, \alpha_{i,j} \in \mathbb{C}\}$ is dense in $C(\Sigma)$; so, any continuous *-homomorphism of $C(\Sigma)$ is determined by its values on the set $\mathcal{D}'$, and hence by its values on the set $\{f_0, f_1\}$, where $f_j(z) = z^j, j = 0, 1$; and the proof of the corollary is complete.

(b) Suppose $\mathcal{A}_0$ does not have an identity; again let $\Sigma = \sigma(x)$, and observe that $0 \in \Sigma$ - see the few paragraphs preceding Definition 3.3.8. Thus, if $\mathcal{A}^+$ denotes the unitisation of $\mathcal{A}$ as in Proposition 3.3.4, and if we regard $\mathcal{A}$ as a maximal ideal in $\mathcal{A}^+$ (by identifying $x \in \mathcal{A}$ with $L_x \in \mathcal{A}^+$), we see that $\mathcal{A}_0$ gets identified with the maximal ideal $\mathcal{I}_0 = \{L_y : y \in \mathcal{A}_0\}$ of $\mathcal{A}_0^+ = \{L_y + \alpha : y \in \mathcal{A}_0, \alpha \in \mathbb{C}\}$. Under the isomorphism of $C(\Sigma)$ with $\mathcal{A}_0^+$ that is guaranteed by applying (a) to the element $L_x \in \mathcal{A}_0^+$, it is easy to see that what we have called $C_0(\Sigma)$ gets mapped onto $\mathcal{A}_0$, and this proves the existence of an isomorphism with the desired properties. Uniqueness is established along exactly the same lines as in (a). □

The moral to be drawn from the preceding proposition is this: if something is true for continuous functions defined on a compact subset of $\mathbb{C}$ (resp., $\mathbb{R}$), then it is also true for normal (resp., self-adjoint) elements of a C^*-algebra. We make precise the sort of thing we mean by the foregoing 'moral' in the following proposition.

PROPOSITION 3.3.11 *Let $\mathcal{A}$ be a C^*-algebra.*

(a) The following conditions on an element $x \in \mathcal{A}$ are equivalent:
(i) x is normal, and $\sigma(x) \subset \mathbb{R}$;
(ii) x is self-adjoint.

(b) The following conditions on an element $u \in \mathcal{A}$ are equivalent:
(i) u is normal, and $\sigma(u) \subset \mathbb{T}$;
(ii) $\mathcal{A}$ has an identity, and u is **unitary**, *i.e., $u^*u = uu^* = 1$.*

(c) The following conditions on an element $p \in \mathcal{A}$ are equivalent:
(i) p is normal, and $\sigma(p) \subset \{0, 1\}$;
(ii) p is a **projection**, *i.e., $p = p^2 = p^*$.*

(d) The following conditions on an element $x \in \mathcal{A}$ are equivalent:
(i) x is normal, and $\sigma(x) \subset [0, \infty)$;
(ii) x is **positive**, *i.e., there exists a self-adjoint element $y \in \mathcal{A}$ such that $x = y^2$.*

(e) If $x \in \mathcal{A}$ is positive (as in (d) above), then there exists a unique positive element $y \in \mathcal{A}$ such that $x = y^2$; this y actually belongs to $C^*(\{x\})$ and is given by $y = f(x)$ where $f(t) = t^{\frac{1}{2}}$; this unique element y is called **the positive square root** *of x and is denoted by $y = x^{\frac{1}{2}}$.*

(f) Let x be a self-adjoint element in $\mathcal{A}$. Then there exists a unique decomposition $x = x_+ - x_-$, where x_+, x_- are positive elements of $\mathcal{A}$ (in the sense of (d) above) which satisfy the condition $x_+x_- = 0$.

Proof : *(a):* $(i) \Rightarrow (ii)$: By Proposition 3.3.10, we have an isomorphism $C^*(\{x\}) \cong C(\sigma(x))$ in which x corresponds to the identity function f_1 on $\sigma(x)$; if $\sigma(x) \subset \mathbb{R}$, then f_1, and consequently, also x, is self-adjoint.

$(ii) \Rightarrow (i)$: This also follows easily from Proposition 3.3.10.

The proofs of (b) and (c) are entirely similar. (For instance, for $(i) \Rightarrow (ii)$ in (b), you would use the fact that if $z \in \mathbb{T}$, then $|z|^2 = 1$, so that the identity function f_1 on $\sigma(x)$ would satisfy $f_1^* f_1 = f_1 f_1^* = 1$.)

(d) For $(i) \Rightarrow (ii)$, put $y = f(x)$, where $f \in C(\sigma(x))$ is defined by $f(t) = t^{\frac{1}{2}}$, where $t^{\frac{1}{2}}$ denotes the non-negative square root of the non-negative number t. The implication $(ii) \Rightarrow (i)$ follows from Proposition 3.3.10 and the following two facts : the square of a real-valued function is a function with non-negative values; and, if $f \in C(X)$, with X a compact Hausdorff space, then $\sigma(f) = f(X)$. (Prove this last fact!)

(e) If f is as in (d), and if we define $y = f(x)$, it then follows (since $f(0) = 0$) that $y \in C^*(\{x\})$ and that y is positive. Suppose now that z is some other positive element of $\mathcal{A}$ such that $x = z^2$. It follows that if we write $\mathcal{A}_0 = C^*(\{z\})$, then $x \in \mathcal{A}_0$, and consequently, also $y \in C^*(\{x\}) \subset \mathcal{A}_0$; but now, by Theorem 3.3.6, there is an isomorphism of $\mathcal{A}_0$ with $C(X)$ where $X = \sigma(z) \subset \mathbb{R}$, and under this isomorphism, both z and y correspond to non-negative continuous functions on X, such that the squares of these two functions are identical; since the non-negative square root of a non-negative number is unique, we may conclude that $z = y$.

(f) For existence of the decomposition, define $x_\pm = f_\pm(x)$, where $f_\pm$ is the continuous function on $\mathbb{R}$ (and hence on any subset of it) defined by $f_\pm(t) = max\{0, \pm t\}$, and note that $f_\pm(t) \geq 0$, $f_+(t) - f_-(t) = t$ and $f_+(t) f_-(t) = 0$ for all $t \in \mathbb{R}$. Also, note that $f_\pm(0) = 0$, so that $f_\pm(x) \in C^*(\{x\})$.

For uniqueness, suppose $x = x_+ - x_-$ is a decomposition of the desired sort. Then note that $x_- x_+ = x_-^* x_+^* = (x_+ x_-)^* = 0$, so that x_+ and x_- are commuting positive elements of $\mathcal{A}$. It follows that $\mathcal{A}_0 = C^*(\{x_+, x_-\})$ is a commutative C^*-algebra which contains $\{x_+, x_-, x\}$ and consequently also $f_\pm(x) \in C^*(\{x\}) \subset \mathcal{A}_0$. Now appeal to Theorem 3.3.6 (applied to $\mathcal{A}_0$), and the fact (which you should prove for yourself) that if X is a compact Hausdorff space, if f is a real-valued continuous function on X, and if $g_j, j = 1, 2$ are non-negative continuous functions on X such that $f = g_1 - g_2$ and $g_1 g_2 = 0$, then necessarily $g_1(x) = max\{f(x), 0\}$ and $g_2(x) = max\{-f(x), 0\}$, and conclude the proof of the proposition. $\square$

Given an element x of a C^*-algebra which is positive in the sense of Proposition 3.3.11(d), we shall write $x \geq 0$ or $0 \leq x$. We wish to show that the set of positive elements of a C^*-algebra form a *positive cone* - meaning that if x and y are positive elements of a C^*-algebra, and if a, b are non-negative real numbers, then also $ax + by \geq 0$; since clearly $ax, by \geq 0$, we only need to show that a sum of positive elements is positive; we proceed towards this goal through a lemma.

LEMMA 3.3.12 *Let $\mathcal{A}$ be a unital Banach algebra, and suppose $x, y \in \mathcal{A}$. Then, $\sigma(xy) \cup \{0\} = \sigma(yx) \cup \{0\}$.*

Proof : We wish to show that if $\lambda \neq 0$, then $(\lambda - xy) \in \mathcal{G}(\mathcal{A}) \Leftrightarrow (\lambda - yx) \in \mathcal{G}(\mathcal{A})$; by replacing x by $\frac{x}{\lambda}$ if necessary, it is sufficient to consider the case $\lambda = 1$. By the symmetry of the problem, it clearly suffices to show that if $(1 - yx)$ is invertible, then so is $(1 - xy)$.

We first present the heuristic and non-rigorous motivation for the proof. Write $(1 - yx)^{-1} = 1 + yx + yxyx + yxyxyx + \cdots$; then, $(1 - xy)^{-1} = 1 + xy + xyxy + xyxyxy + \cdots = 1 + x(1 - yx)^{-1}y$.

Coming back to the (rigorous) proof, suppose $u = (1 - yx)^{-1}$; thus, $u - uyx = u - yxu = 1$. Set $v = 1 + xuy$, and note that

$$\begin{aligned}
v(1 - xy) &= (1 + xuy)(1 - xy) \\
&= 1 + xuy - xy - xuyxy \\
&= 1 + x(u - 1 - uyx)y \\
&= 1\ ,
\end{aligned}$$

and an entirely similar computation shows that also $(1 - xy)v = 1$, thus showing that $(1 - xy)$ is indeed invertible (with inverse given by v). $\square$

PROPOSITION 3.3.13 *If $\mathcal{A}$ is a C^*-algebra, if $x, y \in \mathcal{A}$, and if $x \geq 0$ and $y \geq 0$, then also $(x + y) \geq 0$.*

Proof : To start with, (by embedding $\mathcal{A}$ in a larger unital C^*-algebra, if necessary), we may assume that $\mathcal{A}$ is itself a unital C^*-algebra. Next, we may (by scaling both x and y down by the same small positive scalar, if necessary), assume (without loss of generality) that $||x||, ||y|| \leq 1$. Thus $r(x) \leq 1$ and we may conclude that $\sigma(x) \subset [0,1]$, and consequently deduce that $\sigma(1-x) \subset [0,1]$, and hence that $(1-x) \geq 0$ and $||1-x|| = r(1-x) \leq 1$. Similarly, also $||1-y|| \leq 1$. Then, $||1 - \frac{x+y}{2}|| = \frac{1}{2}||(1-x)+(1-y)|| \leq 1$. Since $\frac{x+y}{2}$ is clearly self-adjoint, this means that $\sigma(\frac{x+y}{2}) \subset [0,2]$, whence $\sigma(x+y) \subset [0,4]$ (by the spectral mapping theorem), and the proof of lemma is complete. □

COROLLARY 3.3.14 *If $\mathcal{A}$ is a C^*-algebra, let $\mathcal{A}_{sa}$ denote the set of self-adjoint elements of $\mathcal{A}$. If $x, y \in \mathcal{A}_{sa}$, say $x \geq y$ (or equivalently, $y \leq x$) if it is the case that $(x-y) \geq 0$. Then this defines a partial order on the set $\mathcal{A}_{sa}$.*

Proof : Reflexivity follows from $0 \geq 0$; transitivity follows from Proposition 3.3.13; anti-symmetry amounts to the statement that $x = x^*, x \geq 0$ and $-x \geq 0$ imply that $x = 0$; this is because such an x should have the property that $\sigma(x) \subset [0,\infty)$ and $\sigma(x) \subset (-\infty, 0]$; this would mean that $\sigma(x) = \{0\}$, and since $x = x^*$, it would follow that $||x|| = r(x) = 0$. □

LEMMA 3.3.15 *Suppose z is an element of a C^*-algebra and that $z^*z \leq 0$; then $z = 0$.*

Proof : Deduce from Lemma 3.3.12 that the hypothesis implies that also $zz^* \leq 0$; hence, we may deduce from Proposition 3.3.13 that $z^*z + zz^* \leq 0$. However, if $z = u+iv$ is the Cartesian decomposition of z, note then that $z^*z+zz^* = 2(u^2+v^2)$; hence, we may deduce, from Proposition 3.3.13, that $z^*z + zz^* \geq 0$. Thus, we find that $z^*z+zz^* = 0$. This means that $u^2 = -v^2$; arguing exactly as before, we find that $u^2 \geq 0$ and $u^2 \leq 0$, whence $u^2 = 0$, and so $u = 0$ (since $||u|| = ||u^2||^{\frac{1}{2}}$ for self-adjoint u). Hence also $v = 0$ and the proof of the lemma is complete. □

PROPOSITION 3.3.16 *The following conditions on an element x in a C^*-algebra are equivalent:*

(i) $x \geq 0$;

(ii) there exists an element $z \in A$ such that $x = z^*z$.

Proof : The implication $(i) \Rightarrow (ii)$ follows upon setting $z = x^{\frac{1}{2}}$ - see Proposition 3.3.11(e).

As for $(ii) \Rightarrow (i)$, let $x = x_+ - x_-$ be the canonical decomposition of the self-adjoint element x into its positive and negative parts; we then find (since $x_- x_+ = 0$) that $x_- x x_- = -x_-^3 \leq 0$; but $x_- x x_- = (zx_-)^*(zx_-)$, and we may conclude from Lemma 3.3.15 that $x_-^3 = 0$ whence $z^* z = x = x_+ \geq 0$, as desired. □

3.4 Representations of C*-algebras

We will be interested in representing an abstract C^*-algebra as operators on Hilbert space; i.e., we will be looking at unital *-homomorphisms from abstract unital C^*-algebras into the concrete unital C^*-algebra $\mathcal{L}(\mathcal{H})$ of operators on Hilbert space. Since we wish to use lower case Roman letters (such as x, y, z, etc.) to denote elements of our abstract C^*-algebras, we shall, for the sake of clarity of exposition, use Greek letters (such as ξ, η, ζ, etc.) to denote vectors in Hilbert spaces, in the rest of this chapter. Consistent with this change in notation (from Chapter 2, where we used symbols such as A, T etc., for operators on Hilbert space), we shall adopt the following further notational conventions in the rest of this chapter: upper case Roman letters (such as A, B, M, etc.) will be used to denote C^*-algebras as well as subsets of C^*-algebras, while calligraphic symbols (such as $\mathcal{S}, \mathcal{M}$, etc.) will be reserved for subsets of Hilbert spaces.

We now come to the central notion of this section, and some of the related notions.

DEFINITION 3.4.1 *(a) A* **representation** *of a unital C^*-algebra A is a *-homomorphism $\pi : A \to \mathcal{L}(\mathcal{H})$ (where $\mathcal{H}$ is some Hilbert space), which will always be assumed to be a 'unital homomorphism', meaning that $\pi(1) = 1$ - where the symbol 1 on the left (resp., right) denotes the identity of the C^*-algebra A (resp., $\mathcal{L}(\mathcal{H})$).*

(b) Two representations $\pi_i : A \to \mathcal{L}(\mathcal{H}_i)$, $i = 1, 2$, are said to be **equivalent** *if there exists a unitary operator $U : \mathcal{H}_1 \to \mathcal{H}_2$ with the property that $\pi_2(x) = U\pi_1(x)U^* \ \forall\ x \in A$.*

(c) A representation $\pi : A \to \mathcal{L}(\mathcal{H})$ is said to be **cyclic** *if there exists a vector $\xi \in \mathcal{H}$ such that $\{\pi(x)(\xi) : x \in A\}$ is dense in $\mathcal{H}$. (In this case, the vector ξ is said to be a* **cyclic vector** *for the representation π.*

We commence with a useful fact about *-homomorphisms between unital C^*-algebras.

LEMMA 3.4.2 *Suppose $\pi : A \to B$ is a unital *-homomorphism between unital C^*-algebras; then,*
(a) $x \in A \Rightarrow \sigma(\pi(x)) \subset \sigma(x)$; and
(b) $||\pi(x)|| \leq ||x|| \ \forall x \in A$.

Proof : (a) Since π is a unital algebra homomorphism, it follows that π maps invertible elements to invertible elements; this clearly implies the asserted spectral inclusion.

(b) If $x = x^* \in A$, then also $\pi(x) = \pi(x^*) = \pi(x)^* \in B$, and since the norm of a self-adjoint element in a C^*-algebra is equal to its spectral radius, we find (from (a)) that

$$||\pi(x)|| = r(\pi(x)) \leq r(x) = ||x|| ;$$

for general $x \in A$, deduce that

$$||\pi(x)||^2 = ||\pi(x)^*\pi(x)|| = ||\pi(x^*x)|| \leq ||x^*x|| = ||x||^2 ,$$

and the proof of the lemma is complete. □

EXERCISE 3.4.3 *Let $\{\pi_i : A \to \mathcal{L}(\mathcal{H}_i)\}_{i \in I}$ be an arbitrary family of representations of (an arbitrary unital C^*-algebra) A; show that there exists a unique (unital) representation $\pi : A \to \mathcal{L}(\mathcal{H})$, where $\mathcal{H} = \oplus_{i \in I} \mathcal{H}_i$, such that $\pi(x) = \oplus_{i \in I} \pi_i(x)$. (See Exercise 2.5.8 for the definition of a direct sum of an arbitrary family of operators.) The representation π is called the direct sum of the representations $\{\pi_i : i \in I\}$, and we write $\pi = \oplus_{i \in I} \pi_i$. (Hint: Use Exercise 2.5.8 and Lemma 3.4.2.)*

DEFINITION 3.4.4 *If $\mathcal{H}$ is a Hilbert space, and if $S \subset \mathcal{L}(\mathcal{H})$ is any set of operators, then the* **commutant** *of S is denoted by the symbol S', and is defined by*

$$S' = \{x' \in \mathcal{L}(\mathcal{H}) : x'x = xx' \ \forall x \in S\} .$$

PROPOSITION 3.4.5 *(a) If $S \subset \mathcal{L}(\mathcal{H})$ is arbitrary, then S' is a unital subalgebra of $\mathcal{L}(\mathcal{H})$ which is closed in the weak-operator topology (and consequently also in the strong operator and norm topologies) on $\mathcal{L}(\mathcal{H})$.*

(b) If S is closed under formation of adjoints - i.e., if $S = S^*$, where of course $S^* = \{x^* : x \in S\}$ - then S' is a C^*-subalgebra of $\mathcal{L}(\mathcal{H})$.

(c) If $S \subset T \subset \mathcal{L}(\mathcal{H})$, then

(i) $S' \supset T'$;

(ii) if we inductively define $S'^{(n+1)} = (S'^{(n)})'$ and $S'^{(1)} = S'$, then

$$S' = S'^{(2n+1)} \ \forall n \geq 0$$

and

$$S \subset S'' = S'^{(2)} = S'^{(2n+2)} \ \forall n \geq 0 \ .$$

(d) Let $\mathcal{M}$ be a closed subspace of $\mathcal{H}$ and let p denote the orthogonal projection onto $\mathcal{M}$; and let $S \subset \mathcal{L}(\mathcal{H})$; then the following conditions are equivalent:

(i) $x(\mathcal{M}) \subset \mathcal{M}, \ \forall x \in S$;

(ii) $xp = pxp \ \forall x \in S$.

If $S = S^*$ is 'self-adjoint', then the above conditions are also equivalent to

(iii) $p \in S'$.

Proof : The topological assertion in (a) is a consequence of the fact that multiplication is 'separately weakly continuous' - see Example 2.5.3(3); the algebraic assertions are obvious.

(b) Note that, in general,

$$\begin{aligned} y \in (S^*)' &\Leftrightarrow yx^* = x^*y \ \forall x \in S \\ &\Leftrightarrow xy^* = y^*x \ \forall x \in S \\ &\Leftrightarrow y^* \in S' \end{aligned}$$

so that $(S^*)' = (S')^*$, for any $S \subset \mathcal{L}(\mathcal{H})$.

(c) (i) is an immediate consequence of the definition, as is the fact that

$$S \subset S'' \ ; \tag{3.4.17}$$

applying (i) to the inclusion 3.4.17, we find that $S' \supset S'''$; on the other hand, if we replace S by S' in the inclusion 3.4.17, we see that $S' \subset S'''$; the proof of (c)(ii) is complete.

(d) For one operator $x \in \mathcal{L}(\mathcal{H})$, the condition that $x(\mathcal{M}) \subset \mathcal{M}$, and that $xp = pxp$, are both seen to be equivalent to the requirement that if $((x^i_j))_{1 \leq i,j \leq 2}$ is the 'matrix of x with respect to the direct sum decomposition $\mathcal{H} = \mathcal{M} \oplus \mathcal{M}^{\perp}$' (as discussed in Proposition 2.5.6), then $x_1^2 = 0$; i.e., the 'matrix of x' has the form

$$[x] \;=\; \begin{bmatrix} a & b \\ 0 & c \end{bmatrix} ,$$

where, of course, $a \in \mathcal{L}(\mathcal{M})), b \in \mathcal{L}(\mathcal{M}^{\perp}, \mathcal{M}))$ and $c \in \mathcal{L}(\mathcal{M}^{\perp}))$ are the appropriate 'compressions' of x - where a compression of an operator $x \in \mathcal{L}(\mathcal{H})$ is an operator of the form $z = P_{\mathcal{M}} \circ x|_{\mathcal{N}}$ for some subspaces $\mathcal{M}, \mathcal{N} \subset \mathcal{H}$.

If S is self-adjoint, and if (i) and (ii) are satisfied, then we find that $xp = pxp$ and $x^*p = px^*p$ for all $x \in S$; taking adjoints in the second equation, we find that $px = pxp$; combining with the first equation, we thus find that $px = xp \; \forall x \in S$; i.e., $p \in S'$.

Conversely, (iii) clearly implies (ii), since if $px = xp$, then $pxp \;=\; xp^2 = xp$. □

We are now ready to establish a fundamental result concerning self-adjoint subalgebras of $\mathcal{L}(\mathcal{H})$.

THEOREM 3.4.6 *(***von Neumann's density theorem***)*
*Let A be a unital *-subalgebra of $\mathcal{L}(\mathcal{H})$. Then A'' is the closure of A in the strong operator topology.*

Proof : Since $A \subset A''$, and since A'' is closed in the strong operator topology, we only need to show that A is strongly dense in A''.

By the definition of the strong topology, we need to prove the following:

Assertion: Let $z \in A''$; then for any $n \in \mathbb{N}, \xi_1, \cdots, \xi_n \in \mathcal{H}$ and $\epsilon > 0$, there exists an $x \in A$ such that $||(x - z)\xi_i|| < \epsilon$ for $1 \leq i \leq n$.

Case (i): We first prove the assertion when $n = 1$.

Let us write $\xi = \xi_1$, and let $\mathcal{M}$ denote the closure of the set $A\xi \;=\; \{x\xi : x \in A\}$. Note that $A\xi$ is a vector space containing ξ (since A is an algebra containing 1), and hence $\mathcal{M}$ is a closed subspace of $\mathcal{H}$ which is obviously stable under A, meaning that $x\mathcal{M} \subset \mathcal{M} \; \forall x \in A$. Hence, if p is the projection onto $\mathcal{M}$, we may deduce (from Proposition 3.4.5(d)) that $p \in A'$ and that $p\xi = \xi$.

Since $z \in A''$, we find that $zp = pz$, whence $z\mathcal{M} \subset \mathcal{M}$; in particular, $z\xi \in \mathcal{M}$; by definition of $\mathcal{M}$, this means there exists $x \in A$ such that $||(z\xi - x\xi|| < \epsilon$.

Case (ii) : $n \in \mathbb{N}$ arbitrary.

Let $\mathcal{H}^n = \mathcal{H} \oplus \overset{n \ terms}{\cdots} \oplus \mathcal{H}$; as in Proposition 2.5.6, we shall identify $\mathcal{L}(\mathcal{H}^n)$ with $M_n(\mathcal{L}(\mathcal{H}))$.

Let us adopt the following notation: if $a \in \mathcal{L}(\mathcal{H})$, let us write $a^{(n)}$ for the element of $M_n(\mathcal{L}(\mathcal{H}))$ with (i,j)-th entry being given by $\delta^i_j a$; thus,

$$a^{(n)} = \begin{bmatrix} a & 0 & 0 & \cdots & 0 \\ 0 & a & 0 & \cdots & 0 \\ 0 & 0 & a & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a \end{bmatrix} .$$

With this notation, define

$$A^{(n)} = \{a^{(n)} : a \in A\} ,$$

and observe that $A^{(n)}$ is a unital *-subalgebra of $\mathcal{L}(\mathcal{H}^{(n)})$.

We claim that

$$(A^{(n)})' = \{((b^i_j)) : \ b^i_j \in A' \ \forall \ i,j\} \tag{3.4.18}$$

and that

$$\begin{aligned} (A^{(n)})'' &= (A'')^{(n)} \\ &= \{z^{(n)} : z \in A''\} . \end{aligned} \tag{3.4.19}$$

Let $b = ((b^i_j))$. Then $ba^{(n)} = ((b^i_j a))$, while $a^{(n)}b = ((ab^i_j))$; the validity of equation 3.4.18 follows.

As for equation 3.4.19, begin by observing that (in view of equation 3.4.18), we have $e^i_j \in (A^{(n)})'$, where e^i_j is the matrix which has entry $1 = id_{\mathcal{H}}$ in the (i,j)-th place and 0's elsewhere; hence, if $y \in (A^{(n)})''$, we should, in particular, have $ye^i_j = e^i_j y$ for all $1 \leq i,j \leq n$. This is seen, fairly easily, to imply that there must exist some $z \in \mathcal{L}(\mathcal{H})$ such that $y = z^{(n)}$; another routine verification shows that $z^{(n)}$ will commute with $((b^i_j))$ for arbitrary $b^i_j \in A'$ if and only if $z \in A''$, thus completing the proof of equation 3.4.19.

Coming back to the assertion, if $z \in A''$, consider the element $z^{(n)} \in (A^{(n)})''$ and the vector $\xi = (\xi_1, \xi_2, \cdots, \xi_n) \in \mathcal{H}^n$, and appeal to the already established Case (i) to find an element of $A^{(n)}$ - i.e., an element of the form $a^{(n)}$, with $a \in A$ - such that $||(z^{(n)} - a^{(n)})\xi|| < \epsilon$; this implies that $||(z-a)\xi_i|| < \epsilon$ for $1 \leq i \leq n$. □

COROLLARY 3.4.7 (**Double commutant theorem**)

*The following conditions on a unital *-subalgebra $M \subset \mathcal{L}(\mathcal{K})$ are equivalent:*

(i) M is weakly closed;

(ii) M is strongly closed;

(iii) $M = M''$.

A subalgebra M as above is called a **von Neumann algebra**.

Proof : $(i) \Rightarrow (ii)$ Obvious.

$(ii) \Rightarrow (iii)$ This is an immediate consequence of the density theorem above.

$(iii) \Rightarrow (i)$ This follows from Proposition 3.4.5(a). □

REMARK 3.4.8 Suppose $\pi : A \to \mathcal{L}(\mathcal{H})$ is a representation; a closed subspace $\mathcal{M} \subset \mathcal{H}$ is said to be π-**stable** if $\pi(x)(\mathcal{M}) \subset \mathcal{M}, \ \forall x \in A$. In view of Proposition 3.4.5(d), this is equivalent to the condition that $p \in \pi(A)'$, where p is the projection of $\mathcal{H}$ onto $\mathcal{M}$. Thus π-stable subspaces of $\mathcal{H}$ are in bijective correspondence with projections in the von Neumann algebra $\pi(A)'$. In particular - since $p \in M \Rightarrow 1 - p \in M$ - we see that the orthogonal complement of a π-stable subspace is also π-stable.

Every π-stable subspace yields a new representation - call it $\pi|_{\mathcal{M}}$ - by restriction: thus $\pi|_{\mathcal{M}} : A \to \mathcal{L}(\mathcal{M})$ is defined by $\pi|_{\mathcal{M}}(x)(\xi) = \pi(x)(\xi), \ \forall x \in A, \ \xi \in \mathcal{M}$. The representation $\pi|_{\mathcal{M}}$ is called a **sub-representation** of π.

LEMMA 3.4.9 *Any representation is equivalent to a direct sum of cyclic (sub-)representations; any separable representation is equivalent to a countable direct sum of cyclic representations.*

Proof : Notice, to start with, that if $\pi : A \to \mathcal{L}(\mathcal{H})$ is a representation and if $\xi \in \mathcal{H}$ is arbitrary, then the subspace $\mathcal{M} = \overline{\{\pi(x)\xi : x \in A\}}$ is a closed subspace which is π-stable, and, by definition, the sub-representation $\pi|_{\mathcal{M}}$ is cyclic (with cyclic vector ξ). Consider the non-empty collection $\mathcal{P}$ whose typical element is a non-empty collection $\mathcal{S} = \{\mathcal{M}_i : i \in I\}$ of pairwise orthogonal non-zero π-stable subspaces which are cyclic (in the sense that the sub-representation afforded by each of them is a cyclic representation). It is clear that the set $\mathcal{P}$ is partially ordered by inclusion, and that if $\mathcal{C} = \{\mathcal{S}_\lambda : \lambda \in \Lambda\}$ is any totally ordered subset of $\mathcal{P}$, then $\mathcal{S} = \cup_{\lambda\in\Lambda}\mathcal{S}_\lambda$ is an element of $\mathcal{P}$; thus every totally ordered set in $\mathcal{P}$ admits an upper bound. Hence, Zorn's lemma implies the existence of a maximal element $\mathcal{S} = \{\mathcal{M}_i : i \in I\}$ of $\mathcal{P}$.

Then $\mathcal{M} = \oplus_{i\in I}\mathcal{M}_i$ is clearly a π-stable subspace of $\mathcal{H}$, and so also is $\mathcal{M}^\perp$. If $\mathcal{M}^\perp \neq \{0\}$, pick a non-zero $\xi \in \mathcal{M}^\perp$, let $\mathcal{M}_0 = \overline{\{\pi(x)\xi : x \in A\}}$, and observe that $\mathcal{S} \cup \{\mathcal{M}_0\}$ is a member of $\mathcal{P}$ which contradicts the maximality of $\mathcal{S}$. Thus, it should be the case that $\mathcal{M}^\perp = \{0\}$; i.e., $\mathcal{H} = \oplus_{i\in I}\mathcal{M}_i$, and the proof of the first assertion is complete.

As for the second, if $\mathcal{H} = \oplus_{i\in I}\mathcal{M}_i$ is an orthogonal decomposition of $\mathcal{H}$ as a direct sum of non-zero cyclic π-stable subspaces as above, let ξ_i be any unit vector in $\mathcal{M}_i$. Then $\{\xi_i : i \in I\}$ is an orthonormal set in $\mathcal{H}$, and the assumed separability of $\mathcal{H}$ implies that I is necessarily countable. □

Before beginning the search for cyclic representations, a definition is in order.

DEFINITION 3.4.10 *A* **state** *on a unital C^*-algebra A is a linear functional $\phi : A \to \mathbb{C}$ which satisfies the following two conditions:*

(i) ϕ is **positive**, *meaning that $\phi(x^*x) \geq 0\ \forall x \in A$ - i.e., ϕ assumes non-negative values on positive elements of A; and*

(ii) $\phi(1) = 1$.

If $\pi : A \to \mathcal{L}(\mathcal{H})$ is a representation, and ξ is a unit vector in $\mathcal{H}$, the functional ϕ, defined by $\phi(x) = \langle \pi(x)\xi, \xi\rangle$ for all $x \in A$, yields an example of a state. It follows from Lemma 3.4.2 that in fact ϕ is a bounded linear functional, with $||\phi|| = \phi(1) = 1$.

The content of the following Proposition is that this property caharacterises a state.

Proposition 3.4.11 *The following conditions on a linear functional $\phi : A \to \mathbb{C}$ are equivalent:*
(i) ϕ is a state;
(ii) $\phi \in A^*$ and $||\phi|| = \phi(1) = 1$.

Proof : $(i) \Rightarrow (ii)$: Notice, to begin with, that since any self-adjoint element is expressible as a difference of two positive elements - see Proposition 3.3.11(f) - that $\phi(x) \in \mathbb{R}$ whenever $x = x^*$; it follows now from the Cartesian decomposition, that $\phi(x^*) = \overline{\phi(x)}$ for all $x \in A$.

Now, consider the sesquilinear form $B_\phi : A \times A \to \mathbb{C}$ defined by

$$B_\phi(x, y) \;=\; \phi(y^* x) \;,\; \forall\, x, y \in A\;. \tag{3.4.20}$$

It follows from the assumed positivity of ϕ (and the previous paragraph) that B_ϕ is a positive-semidefinite sesquilinear form on A - as in Remark 2.1.5; in fact, that remark now shows that

$$|\phi(y^* x)|^2 \;\leq\; \phi(x^* x) \;\cdot\; \phi(y^* y)\;\forall\, x, y \in A\;. \tag{3.4.21}$$

In particular, setting $y = 1$ in equation 3.4.21, and using the obvious inequality $x^* x \leq ||x||^2\, 1$ and positivity of ϕ, we find that

$$|\phi(x)|^2 \leq \phi(x^* x)\phi(1) \leq ||x||^2 \phi(1)^2 \;=\; ||x||^2$$

so that, indeed $\phi \in A^*$ and $||\phi|| \leq 1$; since $||1|| = 1 = \phi(1)$, we find that $||\phi|| = \phi(1) = 1$, as desired.

$(ii) \Rightarrow (i)$: We shall show that if $x = x^*$, and if $\sigma(x) \subset [a, b]$, then $\phi(x) \in [a, b]$. (This will imply that ϕ is positive and that $\phi(1) = 1$.) Set $c = \frac{a+b}{2}$, $r = \frac{b-a}{2}$; and note that $\sigma(x - c) \subset [-r, r]$ and consequently, $||x - c|| \leq r$; hence it follows (from the assumptions (ii)) that

$$|\phi(x) - c| \;=\; |\phi(x - c)| \;\leq\; ||x - c|| \;\leq r\;;$$

in other words, we indeed have $\phi(x) \in [a, b]$ as asserted, and the proof is complete. □

We are almost ready for the fundamental construction due to Gelfand, Naimark and Segal, which associates cyclic representations to states. We will need a simple fact in the proof of this basic construction, which we isolate as an exercise below, before proceeding to the construction.

EXERCISE 3.4.12 *Suppose* $\mathcal{D}^{(k)} = \{x_i^{(k)} : i \in I\}$ *is a dense subspace of* $\mathcal{H}_k$*, for* $k = 1, 2$*, so that* $\langle x_i^{(1)}, x_j^{(1)} \rangle_{\mathcal{H}_1} = \langle x_i^{(2)}, x_j^{(2)} \rangle_{\mathcal{H}_2}$ *for all* $i, j \in I$*. Show that there exists a unique unitary operator* $U : \mathcal{H}_1 \to \mathcal{H}_2$ *such that* $U x_i^{(1)} = x_i^{(2)} \ \forall \ i \in I$*.*

THEOREM 3.4.13 *(***GNS construction***)*

Let ϕ *be a state on a unital* C^**-algebra* A*; then there exists a cyclic representation* $\pi_\phi : A \to \mathcal{L}(\mathcal{H}_\phi)$ *with a cyclic vector* ξ_ϕ *of unit norm such that*

$$\phi(x) = \langle \pi_\phi(x)\xi_\phi, \xi_\phi \rangle \ \forall \ x \in A \ . \tag{3.4.22}$$

The triple $(\mathcal{H}_\phi, \pi_\phi, \xi_\phi)$ *is unique in the sense that if* $\pi : A \to \mathcal{H}$ *is a cyclic representation with cyclic vector* ξ *such that* $\phi(x) = \langle \pi(x)\xi, \xi \rangle \ \forall \ x \in A$*, then there exists a unique unitary operator* $U : \mathcal{H} \to \mathcal{H}_\phi$ *such that* $U\xi = \xi_\phi$ *and* $U\pi(x)U^* = \pi_\phi(x) \ \forall x \in A$*.*

Proof : Let B_ϕ be the positive-semidefinite sesquilinear form defined on A using the state ϕ as in equation 3.4.20. Let $N_\phi = \{x \in A : B_\phi(x, x) = 0\}$. It follows from equation 3.4.21 that $x \in N_\phi$ if and only if $\phi(y^*x) = 0 \ \forall y \in A$; this implies that N_ϕ is a vector subspace of A which is in fact a left-ideal (i.e., $x \in N_\phi \ \Rightarrow zx \in N_\phi \ \forall \ z \in A$).

Deduce now that the equation

$$\langle x + N_\phi, y + N_\phi \rangle = \phi(y^*x)$$

defines a genuine inner product on the quotient space $V = A/N_\phi$. For notational convenience, let us write $\eta(x) = x + N_\phi$ so that $\eta : A \to V$; since N_ϕ is a left-ideal in A, it follows that each $x \in A$ unambiguously defines a linear map $L_x : V \to V$ by the prescription: $L_x\eta(y) = \eta(xy)$.

We claim now that each L_x is a bounded operator on the inner product space V and that $||L_x||_{\mathcal{L}(V)} \leq ||x||_A$. This amounts to the assertion that

$$\phi(y^*x^*xy) = ||L_x\eta(y)||^2 \leq ||x||^2||\eta(y)||^2 = ||x||^2\phi(y^*y)$$

for all $x, y \in A$. Notice now that, for each fixed $y \in A$, if we consider the functional $\psi(z) = \phi(y^*zy)$, then ψ is a positive linear functional; consequently, we find from Proposition 3.4.11 that $||\psi|| = \psi(1) = \phi(y^*y)$; in particular, we find that for arbitrary $x, y \in A$, we must have $\phi(y^*x^*xy) = \psi(x^*x) \leq ||\psi|| \cdot ||x^*x||$; in other words, $\phi(y^*x^*xy) \leq ||x||^2\phi(y^*y)$, as asserted.

Since V is a genuine inner product space, we may form its completion - call it $\mathcal{H}_\phi$ - where we think of V as a dense subspace of $\mathcal{H}_\phi$. We may deduce from the previous paragraph that each L_x extends uniquely to a bounded operator on $\mathcal{H}_\phi$, which we will denote by $\pi_\phi(x)$; the operator $\pi_\phi(x)$ is defined by the requirement that $\pi_\phi(x)\eta(y) = \eta(xy)$; this immediately implies that π_ϕ is an unital algebra homomorphism of A into $\mathcal{L}(\mathcal{H}_\phi)$. To see that π_ϕ preserves adjoints, note that if $x, y, z \in A$ are arbitrary, then

$$\begin{aligned} \langle \pi_\phi(x)\eta(y), \eta(z)\rangle &= \phi(z^*(xy)) \\ &= \phi((x^*z)^*y) \\ &= \langle \eta(y), \pi_\phi(x^*)\eta(z)\rangle , \end{aligned}$$

which implies, in view of the density of $\eta(A)$ in $\mathcal{H}_\phi$, that $\pi_\phi(x)^* = \pi_\phi(x^*)$, so that π_ϕ is indeed a representation of A on $\mathcal{H}_\phi$. Finally, it should be obvious that $\xi_\phi = \eta(1)$ is a cyclic vector for this representation.

Conversely, if $(\mathcal{H}, \pi, \xi)$ is another triple which also 'works' for ϕ as asserted in the statement of the second half of Theorem 3.4.13, observe that for arbitrary $x, y \in A$, we have

$$\langle \pi(x)\xi, \pi(y)\xi\rangle_{\mathcal{H}} = \phi(y^*x) = \langle \pi_\phi(x)\xi_\phi, \pi_\phi(y)\xi_\phi\rangle_{\mathcal{H}_\phi}$$

for all $x, y \in A$; the assumptions that ξ and ξ_ϕ are cyclic vectors for the representations π and π_ϕ respectively imply, via Exercise 3.4.12, that there exists a unique unitary operator $U : \mathcal{H} \to \mathcal{H}_\phi$ with the property that $U(\pi(x)\xi) = \pi_\phi(x)\xi_\phi$ for all $x \in A$; it is clear that U has the properties asserted in Theorem 3.4.13. □

We now wish to establish that there exist 'sufficiently many' representations of any C^*-algebra.

LEMMA 3.4.14 *Let x be a self-adjoint element of a C^*-algebra A. Then there exists a cyclic representation π of A such that $||\pi(x)|| = ||x||$.*

Proof : Let $A_0 = C^*(\{1, x\})$ be the commutative unital C^*-subalgebra generated by x. Since $A_0 \cong C(\sigma(x))$, there exists - see Exercise 3.2.3(5)(d) - a complex homomorphism $\phi_0 \in \hat{A}_0$ such that $|\phi_0(x)| = ||x||$. Notice that $||\phi_0|| = 1 = \phi_0(1)$.

By the Hahn-Banach theorem, we can find a $\phi \in A^*$ such that $\phi|_{A_0} = \phi_0$ and $||\phi|| = ||\phi_0||$. It follows then that $||\phi|| = 1 = \phi(1)$; hence ϕ is a state on A, by Proposition 3.4.11.

If π_ϕ is the (cyclic) GNS-representation afforded by the state ϕ as in Theorem 3.4.13, we then see that

$$||x|| = |\phi_0(x)| = |\phi(x)| = |\langle \pi_\phi(x)\xi_\phi, \xi_\phi \rangle| \leq ||\pi_\phi(x)|| \ ;$$

in view of Lemma 3.4.2 (b), the proof of the lemma is complete. □

THEOREM 3.4.15 *If A is any C^*-algebra, there exists an isometric representation $\pi : A \to \mathcal{L}(\mathcal{H})$.*

If A is separable, then we can choose the Hilbert space $\mathcal{H}$ above to be separable.

Proof : Let $\{x_i : i \in I\}$ be a dense set in A. For each $i \in I$, pick a cyclic representation $\pi_i : A \to \mathcal{L}(\mathcal{H}_i)$ such that $||\pi_i(x_i^* x_i)|| = ||x_i^* x_i||$ - which is possible, by the preceding lemma. Note that the C^*-identity shows that we have $||\pi_i(x_i)|| = ||x_i||$ for all $i \in I$.

Let $\pi = \oplus_{i \in I} \pi_i$; deduce from Lemma 3.4.2 and the previous paragraph that, for arbitrary $i, j \in I$, we have $||\pi_j(x_i)|| \leq ||x_i|| = ||\pi_i(x_i)||$, and hence we see that $||\pi(x_i)|| = ||x_i|| \ \forall i \in I$; since the set $\{x_i : i \in I\}$ is dense in A, we conclude that the representation π is necessarily isometric.

Suppose now that A is separable. Then, in the above notation, we may assume that I is countable. Further, note that if $\xi_i \in \mathcal{H}_i$ is a cyclic vector for the cyclic representation π_i, then it follows that $\{\pi_i(x_j)\xi_i : j \in I\}$ is a countable dense set in $\mathcal{H}_i$; it follows that each $\mathcal{H}_i$ is separable, and since I is countable, also $\mathcal{H}$ must be separable. □

Hence, when proving many statements regarding general C^*-algebras, it would suffice to consider the case when the algebra in question is concretely realised as a C^*-algebra of operators on some Hilbert space. The next exercise relates the notion of positivity that we have for elements of abstract C^*-algebras to what is customarily referred to as 'positive-definiteness' (or 'positive-semidefiniteness') in the context of matrices.

EXERCISE 3.4.16 *Show that the following conditions on an operator $T \in \mathcal{L}(\mathcal{H})$ are equivalent:*

(i) T is positive, when regarded as an element of the C^*-algebra $\mathcal{L}(\mathcal{H})$ (i.e., $T = T^*$ and $\sigma(T) \subset [0, \infty)$, or $T = S^2$ for some self-adjoint element $S \in \mathcal{L}(\mathcal{H})$, etc.);

(ii) there exists an operator $S \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ such that $T = S^*S$, where $\mathcal{K}$ is some Hilbert space;

(iii) $\langle Tx, x\rangle \geq 0 \ \forall \ x \in \mathcal{H}$.

(Hint: for (i) $\Rightarrow$ (ii), set $\mathcal{K} = \mathcal{H}, S = T^{\frac{1}{2}}$; the implication (ii) $\Rightarrow$ (iii) is obvious; for (iii) $\Rightarrow$ (i), first deduce from Proposition 2.4.9(a) that T must be self-adjoint; let $T = T_+ - T_-$ be its canonical decomposition as a difference of positive operators; use the validity of the implication (i) $\Rightarrow$ (iii) of this exercise to the operator T_- to deduce that it must be the case that $\langle T_- x, x\rangle = 0 \ \forall \ x \in \mathcal{H}$; since T_- is 'determined by its quadratic form', conclude that $T_- = 0$, or equivalently, that $T = T_+ \geq 0$.

3.5 The Hahn-Hellinger theorem

This section is devoted to the classification of separable representations of a separable commutative C^*-algebra. Hence, throughout this section, we will be concerned with the commutative unital C^*-algebra $C(X)$, where X is a compact Hausdorff space which is metrisable. (The reason for this is that for a compact Hausdorff space X, the separability of $C(X)$ is equivalent to metrisability of X - i.e., the topology on X coming from some metric on X.)

It is an immediate consequence of the Riesz representation theorem - see §A.7 in the Appendix - that there is a bijective correspondence between states ϕ on (the commutative unital C^*-algebra) $C(X)$ on the one hand, and probability measures on $(X, \mathcal{B}_X)$, on the other, given by integration, thus:

$$\phi(f) \ = \ \int f d\mu \ . \qquad (3.5.23)$$

The next thing to notice is that the 'GNS'- triple $(\mathcal{H}_\mu, \pi_\mu, \xi_\mu)$ that is associated with this state (as in Theorem 3.4.13) may be easily seen to be given as follows: $\mathcal{H}_\mu = L^2(X, \mathcal{B}_X, \mu)$; ξ_μ is the constant function which is identically equal to 1; and π_μ is the 'multiplication representation' defined by $\pi_\mu(f)\xi = f\xi \ \forall \ f \in C(X), \ \xi \in L^2(X, \mu)$.

Thus, we may, in view of this bijection between states on $C(X)$ and probability measures on $(X, \mathcal{B}_X)$, deduce the following specialisation of the general theory developed in the last section - see Lemma 3.4.9 and Theorem 3.4.13 .

Proposition 3.5.1 *If $\pi : C(X) \to \mathcal{L}(\mathcal{H})$ is any separable representation of $C(X)$ - where X is a compact metric space - then there exists a (finite or infinite) countable collection $\{\mu_n\}_n$ of probability measures on $(X, \mathcal{B}_X)$ such that π is equivalent to $\oplus_n \pi_{\mu_n}$.*

The problem with the preceding proposition is the lack of 'canonical'ness in the construction of the sequence of probability measures μ_n. The rest of this section is devoted, in a sense, to establishing the exact amount of the 'canonicalness' of this construction.

The first step is to identify at least some situations where two different measures can yield equivalent representations.

Lemma 3.5.2 *(a) If μ and ν are two probability measures defined on $(X, \mathcal{B}_X)$ which are mutually absolutely continuous - see §A.5 - then the representations π_μ and π_ν are equivalent.*

(b) If μ and ν are two finite positive measures defined on $(X, \mathcal{B}_X)$ which are mutually singular - see §A.5 - and if we let $\lambda = \mu + \nu$, then $\pi_\lambda \cong \pi_\mu \oplus \pi_\nu$.

Proof : (a) Let $\phi = (\frac{d\nu}{d\mu})^{\frac{1}{2}}$, and note that (by the defining property of the Radon-Nikodym derivative) the equation $(U\xi) = \phi\xi$, $\xi \in \mathcal{H}_\nu$ defines an isometric linear operator $U : \mathcal{H}_\nu \to \mathcal{H}_\mu$ - where, of course, we write $\mathcal{H}_\lambda = L^2(X, \mathcal{B}, \lambda)$. (Reason: if $\xi \in \mathcal{H}_\nu$, then

$$||U\xi||^2 = \int |\xi|^2 \phi^2 d\mu = \int |\xi|^2 \frac{d\nu}{d\mu} d\mu = \int |\xi|^2 d\nu = ||\xi||^2 .)$$

In an identical fashion, if we set $\psi = (\frac{d\mu}{d\nu})^{\frac{1}{2}}$, then the equation $V\eta = \psi\eta$ defines an isometric operator $V : \mathcal{H}_\mu \to \mathcal{H}_\nu$; but the uniqueness of the Radon-Nikodym derivative implies that $\psi = \phi^{-1}$, and that V and U are inverses of one another.

Finally, it is easy to deduce from the definitions that

$$U\pi_\nu(f)\xi = \phi f\xi = \pi_\mu(f)U\xi$$

whenever $\xi \in \mathcal{H}_\nu, \ f \in C(X)$; in other words, U is a unitary operator which implements the equivalence of the representations π_ν and π_μ.

(b) By hypothesis, there exists a Borel partition $X = A \coprod B$, such that $\mu = \mu|_A$ and $\nu = \nu|_B$ - where, as in §A.5, we use the notation $\mu|_E$ to denote the measure defined by $\mu|_E(A) = \mu(E \cap A)$. It is easily seen then that also $\mu = \lambda|_A$ and $\nu = \lambda|_B$; the mapping $\mathcal{H}_\lambda \ni f \mapsto (1_A f, 1_B f) \in \mathcal{H}_\mu \oplus \mathcal{H}_\nu$ is a unitary operator that establishes the desired equivalence of representations. $\square$

Our goal, in this section, is to prove the following classification, up to equivalence, of separable representations of $C(X)$, with X a compact metric space.

Before stating our main result, we shall fix some notation. Given a representation π, and $1 \leq n \leq \aleph_0$, we shall write π^n to denote the direct sum of n copies of π. (Here, the symbol $\aleph_0$ refers, of course, to the 'countable infinity'.)

THEOREM 3.5.3 (**Hahn-Hellinger theorem**)

Let X be a compact metric space.

(1) If π is a separable representation of $C(X)$, then there exists a probability measure μ defined on $(X, \mathcal{B}_X)$ and a family $\{E_n : 1 \leq n \leq \aleph_0\}$ of pairwise disjoint Borel subsets of X such that

(i) $\pi \cong \oplus_{1\leq n\leq\aleph_0} \pi^n_{\mu|_{E_n}}$; and

(ii) μ is supported on $\cup_{1\leq n\leq\aleph_0} E_n$ (meaning that the complement of this set has μ-measure zero).

(2) Suppose $\mu_i, i = 1, 2$ are two probability measures, and suppose that for each $i = 1, 2$ we have a family $\{E_n^{(i)} : 1 \leq n \leq \aleph_0\}$ of pairwise disjoint Borel subsets of X such that μ_i

is supported on $\cup_{1\leq n\leq\aleph_0} E_n^{(i)}$. *Then the following conditions are equivalent:*

(i) $\oplus_{1\leq n\leq\aleph_0} \pi^n_{\mu_1|_{E_n^{(1)}}} \cong \oplus_{1\leq n\leq\aleph_0} \pi^n_{\mu_2|_{E_n^{(2)}}}$;

(ii) the measures μ_i *are mutually absolutely continuous, and further,* $\mu_i(E_n^{(1)} \Delta E_n^{(2)}) = 0$ *for all* n *and for* $i = 1, 2$ *- where, of course, the symbol* Δ *has been used to signify the 'symmetric difference' of sets.*

Proof of (1) : According to Proposition 3.5.1, we can find a countable family - say $\{\mu_n : n \in N\}$ - of probability measures on $(X, \mathcal{B}_X)$ such that

$$\pi \;\cong\; \oplus_{n\in N}\, \pi_{\mu_n}. \tag{3.5.24}$$

Let $\{\epsilon_n : n \in N\}$ be any set of strictly positive numbers such that $\sum_{n\in N} \epsilon_n = 1$, and define $\mu = \sum_{n\in N} \epsilon_n \mu_n$. The definitions imply the following facts:

(i) μ is a probability measure; and

(ii) if $E \in \mathcal{B}_X$, then $\mu(E) = 0 \;\Leftrightarrow\; \mu_n(E) = 0 \;\forall\; n \in N$.

In particular, it follows from (ii) that each μ_n is absolutely continuous with respect to μ; hence, if we set $A_n = \{x \in X : \left(\frac{d\mu_n}{d\mu}\right)(x) > 0\}$, it then follows from Exercise A.5.22(4) that μ_n and $\mu|_{A_n}$ are mutually absolutely continuous; also, it follows from (ii) above that μ is supported on $\cup_{n\in N} A_n$.

We thus find now that

$$\pi \;\cong\; \oplus_{n\in N}\, \pi_{\mu|_{A_n}}. \tag{3.5.25}$$

We will find it convenient to make the following assumption (which we clearly may): $N = \{1, 2, \cdots, n\}$ for some $n \in \mathbb{N}$ (in case N is finite), or $N = \mathbb{N}$, (in case N is infinite). We will also find the introduction of certain sets very useful for purposes of 'counting'.

Let $K = \{1, 2, \cdots, |N|\}$ (where $|N|$ denotes the cardinality of N), and for each $k \in K$, define

$$E_k \;=\; \{x \in X : \sum_{n\in N} 1_{A_n}(x) \;=\; k\}\ . \tag{3.5.26}$$

Thus, $x \in E_k$ precisely when x belongs to exactly k of the A_i's; in particular, $\{E_k : k \in K\}$ is easily seen to be a Borel partition of $\cup_{n \in N} A_n$. (Note that we have deliberately omitted 0 in our definition of the set K.) Thus we may now conclude - thanks to Lemma 3.5.2(b) - that

$$\pi \cong \oplus_{n \in N} \oplus_{k \in K} \pi_{\mu|_{(A_n \cap E_k)}} . \tag{3.5.27}$$

Next, for each $n \in N, k \in K, l \in \mathbb{N}$ such that $1 \leq l \leq k$, define

$$A_{n,k,l} = \{x \in A_n \cap E_k : \sum_{j=1}^{n} 1_{A_j}(x) = l\} . \tag{3.5.28}$$

(Thus, $x \in A_{n,k,l}$ if and only if $x \in E_k$ and further, n is exactly the l-th index j for which $x \in A_j$; i.e., $x \in A_n \cap E_k$ and there are exactly l values of j for which $1 \leq j \leq n$ and $x \in A_j$.) A moment's reflection on the definitions shows that

$$\begin{aligned} A_n \cap E_k &= \coprod_{l \in \mathbb{N}, 1 \leq l \leq k} A_{n,k,l} , \quad \forall\, n \in N, k \in K \\ E_k &= \coprod_{n \in N} A_{n,k,l} , \quad \forall\, k \in K,\ 1 \leq l \leq k. \end{aligned}$$

We may once again appeal to Lemma 3.5.2(b) and deduce from the preceding equations that

$$\begin{aligned} \pi &\cong \oplus_{n \in N} \oplus_{k \in K} \pi_{\mu|_{(A_n \cap E_k)}} \\ &\cong \oplus_{n \in N} \oplus_{k \in K} \oplus_{l \in \mathbb{N}, 1 \leq l \leq k} \pi_{\mu|_{A_{n,k,l}}} \\ &\cong \oplus_{k \in K} \oplus_{l \in \mathbb{N}, 1 \leq l \leq k} \pi_{\mu|_{E_k}} \\ &\cong \oplus_{k \in K} \pi^k_{\mu|_{E_k}} , \end{aligned}$$

and the proof of (1) of the theorem is complete. □

We will have to proceed through a sequence of lemmas before we can complete the proof of (2) of the Hahn-Hellinger theorem. Thus, we have established the 'existence of the canonical decomposition' of an arbitrary separable representation of $C(X)$; we shall in fact use this existence half fairly strongly in the proof of the 'uniqueness' of the decomposition.

In the rest of this section, we assume that X is a compact metric space, and that all Hilbert spaces (and consequently all representations, as well) which we deal with are separable; further, all measures we deal with will be finite positive measures defined on $(X, \mathcal{B}_X)$.

LEMMA 3.5.4 *Let $\phi \in L^\infty(X, \mathcal{B}_X, \mu)$; then there exists a sequence $\{f_n\}_n \subset C(X)$ such that*

(i) $sup_n ||f_n||_{C(X)} < \infty$; and

(ii) the sequence $\{f_n(x)\}_n$ converges to $\phi(x)$, for μ-almost every $x \in X$.

Proof : Since $C(X)$ is dense in $L^2(X, \mathcal{B}_X, \mu)$, we may find a sequence $\{h_n\}_n$ in $C(X)$ such that $||h_n - \phi||_{L^2(X,\mathcal{B}_X,\mu)} \to 0$; since every null convergent sequence in $L^2(X, \mathcal{B}_x, \mu)$ is known - see [Hal1], for instance - to contain a subsequence which converges μ-a.e. to 0, we can find a subsequence - say $\{g_n\}_n$ - of $\{h_n\}_n$ such that $g_n(x) \to \phi(x)$ a.e.

Let $||\phi||_{L^\infty} = K$, and define a continuous 'retraction' of $\mathbb{C}$ onto the disc of radius $K + 1$ as follows:

$$r(z) = \begin{cases} z & if\ |z| \leq (K+1) \\ \left(\frac{K+1}{|z|}\right) z & if\ |z| \geq (K+1) \end{cases}$$

Finally, consider the continuous functions defined by $f_n = r \circ g_n$; it should be clear that $||f_n|| \leq (K+1)$ for all n, and that (by the assumed a.e. convergence of the sequence $\{g_n\}_n$ to f) also $f_n(x) \to f(x)\ \mu - a.e.$ □

LEMMA 3.5.5 *(a) Let $\pi : C(X) \to \mathcal{L}(\mathcal{H})$ be a representation; then there exists a probability measure μ defined on $\mathcal{B}_X$, and a representation $\tilde{\pi} : L^\infty(X, \mathcal{B}_X, \mu) \to \mathcal{L}(\mathcal{H})$ such that:*

(i) $\tilde{\pi}$ is isometric;

(ii) $\tilde{\pi}$ 'extends' π in the sense that $\tilde{\pi}(f) = \pi(f)$ whenever $f \in C(X)$; and

(iii) $\tilde{\pi}$ 'respects bounded convergence' - meaning that whenever ϕ is the μ-a.e. limit of a sequence $\{\phi_n\}_n \subset L^\infty(X, \mu)$ which is uniformly bounded (i.e., $\sup_n ||\phi_n||_{L^\infty(\mu)} < \infty$), then it is the case that the sequence $\{\tilde{\pi}(\phi_n)\}_n$ converges in the strong operator topology to $\tilde{\pi}(\phi)$.

(b) Suppose that, for $i = 1, 2$, $\widetilde{\pi_i} : L^\infty(X, \mathcal{B}_X, \mu_i) \to \mathcal{L}(\mathcal{H}_i)$ is a representation which is related to a representation $\pi_i : C(X) \to \mathcal{L}(\mathcal{H}_i)$ as in (a) above; suppose $\pi_1 \cong \pi_2$; then the measures μ_1 and μ_2 are mutually absolutely continuous.

Proof : (a) First consider the case when π is a cyclic representation; then there exists a probability measure μ such that $\pi \cong \pi_\mu$; take this μ as the measure in the statement of (a) and define $\tilde{\pi}$ to be the natural 'multiplication representation' defined by $\tilde{\pi}(\phi)\xi = \phi\xi, \ \forall \ \phi \in L^\infty(\mu), \xi \in L^2(\mu)$; then we see that statement (ii) is obvious, while (iii) is an immediate consequence of the dominated convergence theorem - see Proposition A.5.16(3). As for (i), suppose $\phi \in L^\infty(\mu)$ and $\epsilon > 0$; if $E = \{x : |\phi(x)| \geq ||\phi|| - \epsilon\}$, then $\mu(E) > 0$, by the definition of the norm on $L^\infty(X, \mathcal{B}_X, \mu)$; hence if $\xi = \mu(E)^{-\frac{1}{2}} 1_E$, we find that ξ is a unit vector in $L^2(\mu)$ such that $||\tilde{\pi}(\phi)\xi|| \ \geq \ ||\phi|| - \epsilon$; the validity of (a) - at least in the cyclic case under consideration - follows now from the arbitrariness of ϵ and from Lemma 3.4.2 (applied to the C^*-algebra $L^\infty(\mu)$).

For general (separable) π, we may (by the the already established part (1) of the Hahn-Hellinger theorem) assume without loss of generality that $\pi = \oplus_n \pi^n_{\mu|_{E_n}}$; then define $\tilde{\pi} = \oplus_n (\widetilde{\pi_{\mu|_{E_n}}})^n$, where the summands are defined as in the last paragraph. We assert that this $\tilde{\pi}$ does the job. Since μ is supported on $\cup_n E_n$ (by (1)(ii) of Theorem 3.5.3), and since $\widetilde{\pi_{\mu|_{E_n}}}$ is an isometric map of $L^\infty(E_n, \mu|_{E_n})$, it is fairly easy to deduce that $\tilde{\pi}$, as we have defined it, is indeed an isometric *-homomorphism of $L^\infty(X, \mu)$, thus establishing (a)(i). The assertion (ii) follows immediately from the definition and the already established cyclic case.

As for (iii), notice that the Hilbert space underlying the representation $\tilde{\pi}$ is of the form $\mathcal{H} = \oplus_{1 \leq n \leq \aleph_0} \oplus_{1 \leq m \leq n} \mathcal{H}_{n,m}$, where $\mathcal{H}_{n,m} = L^2(E_n, \mu|_{E_n}) \ \forall 1 \leq m \leq n \leq \aleph_0$. So, if $\{\phi_k\}_k, \phi$ are as in (ii), we see that $\pi(\phi_k)$ has the form $x^{(k)} = \oplus_{1 \leq m \leq n \leq \aleph_0} x^{(k)}_{n,m}$, where $x^{(k)}_{n,m} = \widetilde{\pi_{\mu|_{E_n}}}(\phi_k)$ for $1 \leq m \leq n \leq \aleph_0$. Similarly $\tilde{\pi}(\phi)$ has the form $x = \oplus_{1 \leq m \leq n \leq \aleph_0} x_{n,m}$, where $x_{n,m} = \widetilde{\pi_{\mu|_{E_n}}}(\phi)$ for $1 \leq m \leq n \leq \aleph_0$. The assumption that $\sup_k ||\phi_k|| < \infty$ shows that the sequence $\{x^{(k)}\}_k$ is uniformly bounded in norm; further, if we let $\mathcal{S} = \cup_{1 \leq m \leq n \leq \aleph_0} \mathcal{H}_{n,m}$ - where we naturally regard the $\mathcal{H}_{n,m}$'s as subspaces of the Hilbert space $\mathcal{H}$ - then we find that $\mathcal{S}$ is a total set in $\mathcal{H}$ and that $x^{(k)}\xi \to x\xi$ whenever $\xi \in \mathcal{S}$ (by the already established cyclic case); thus we may deduce - from Lemma 2.5.2 - that the sequence $\{x^{(k)}\}_k$ does indeed converge strongly to x, and the proof of (a) is complete.

(b) Suppose $U : \mathcal{H}_1 \to \mathcal{H}_2$ is a unitary operator such that

$U\pi_1(f)U^* = \pi_2(f)$ for all $f \in C(X)$. Define $\mu = \frac{1}{2}(\mu_1 + \mu_2)$, and note that both μ_1 and μ_2 are absolutely continuous with respect to (the probability measure) μ. Let ϕ be any bounded measurable function on X. By Lemma 3.5.4, we may find a sequence $\{f_k\}_k \subset C(X)$ such that $\sup_k ||f_k|| < \infty$ and such that $f_k(x) \to \phi(x)$ for all $x \in X - N$ where $\mu(N) = 0$; then also $\mu_i(N) = 0, i = 1, 2$; hence by the assumed property of the $\widetilde{\pi_i}$'s, we find that $\{\pi_i(f_k)\}_k$ converges in the strong operator topology to $\widetilde{\pi_i}(\phi)$; from the assumed intertwining nature of the unitary operator U, we may deduce that $U\widetilde{\pi_1}(\phi)U^* = \widetilde{\pi_2}(\phi)$ for every bounded measurable function $\phi : X \to \mathbb{C}$. In particular, we see that

$$U\widetilde{\pi_1}(1_E)U^* \;=\; \widetilde{\pi_2}(1_E) \;\forall\; E \in \mathcal{B}_X \;. \tag{3.5.29}$$

Since the $\widetilde{\pi_i}$'s are isometric, we find that, for $E \in \mathcal{B}_X$,

$$\begin{aligned} \mu_1(E) = 0 \;&\Leftrightarrow\; \widetilde{\pi_1}(1_E) = 0 \\ &\Leftrightarrow\; \widetilde{\pi_2}(1_E) \;\; \text{(by 3.5.29)} \\ &\Leftrightarrow\; \mu_2(E) = 0 \;, \end{aligned}$$

thereby completing the proof of the lemma. □

Remark 3.5.6 Let $\pi : C(X) \to \mathcal{L}(\mathcal{H})$ be a separable representation; first notice from Lemma 3.5.5(b) that the measure μ of Lemma 3.5.5(a) is uniquely determined up to mutual absolute continuity; also, the proof of Lemma 3.5.5(b) shows that if the representations π_1 and π_2 of $C(X)$ are equivalent, then the representations $\widetilde{\pi_1}$ and $\widetilde{\pi_2}$ of $L^\infty(X, \mu)$ are equivalent (with the two equivalences being implemented by the same unitary operator). Furthermore, it is a consequence of Lemma 3.5.4 and Lemma 3.5.5(a)(iii) that the *-homomorphism $\tilde{\pi}$ is also uniquely determined by the condition that it satisfies (a) (i)-(iii). Thus, the representation π uniquely determines the C^*-algebra $L^\infty(X, \mu)$ and the representation $\tilde{\pi}$ as per Lemma 3.5.5(a)(i)-(iii). □

Lemma 3.5.7 *If $\pi, \mu, \tilde{\pi}$ are as in Lemma 3.5.5(a), then*

$$(\pi(C(X))'' = \tilde{\pi}(L^\infty(X, \mu)) \;.$$

Thus, $\tilde{\pi}(L^\infty(X, \mu)$ is the von Neumann algebra generated by $\pi(C(X))$.

Proof : Let us write $A = \pi(C(X))$, $M = \tilde{\pi}(L^\infty(X,\mu))$. Thus, we need to show that $M = A''$.

Since A is clearly commutative, it follows that $A \subset A'$; since A is a unital *-subalgebra of $\mathcal{L}(\mathcal{H})$, and since A' is closed in the strong operator topology, we may conclude, from the preceding inclusion and von Neumann's density theorem - see Theorem 3.4.6 - that we must have $A'' \subset A'$.

Next, note that, in view of Lemma 3.5.4, Lemma 3.5.5(a)(iii) and Theorem 3.4.6, we necessarily have $M \subset A''$. Thus, we find that we always have

$$M \subset A'' \subset A' \ . \tag{3.5.30}$$

Case (i) : π is cyclic. In this case, we assert that we actually have $A' = A'' = M$.

In the case at hand, we may assume that the underlying Hilbert space is $\mathcal{H} = L^2(X,\mu)$, and that $\tilde{\pi}$ is the multiplication representation of $L^\infty(X,\mu)$. In view of 3.5.30, we need to show that $A' \subset M$.

So suppose $x \in A'$; we wish to conclude that $x = \tilde{\pi}(\phi)$, for some $\phi \in L^\infty(X,\mu)$. Notice that if this were true, it must be the case that $\phi = x\xi_\mu$ - where ξ_μ denotes the constant function 1, as in the notation of Theorem 3.4.13. So, define $\phi = x\xi_\mu$; we need to show that $\phi \in L^\infty(X,\mu)$ and that $x = \tilde{\pi}(\phi)$. For this, begin by deducing from the inclusion 3.5.30 that if $\psi \in L^\infty(X,\mu)$ is arbitrary, then

$$x\psi \ = \ x\tilde{\pi}(\psi)\xi_\mu \ = \ \tilde{\pi}(\psi)x\xi_\mu \ = \ \tilde{\pi}(\psi)\phi \ = \ \phi\psi \ .$$

Further, if we set $E_r = \{s \in X : |\phi(s)| > r\}$, note that $r1_{E_r} \leq |\phi|1_{E_r}$, and consequently,

$$\begin{aligned} \mu(E_r)^{\frac{1}{2}} &= ||1_{E_r}||_{\mathcal{H}} \\ &\leq \frac{1}{r}||\phi 1_{E_r}||_{\mathcal{H}} \\ &= \frac{1}{r}||x1_{E_r}||_{\mathcal{H}} \\ &\leq \frac{1}{r}||x||_{\mathcal{L}(\mathcal{H})} \cdot ||1_{E_r}||_{\mathcal{H}} \ , \end{aligned}$$

which clearly implies that $\mu(E_r) = 0 \ \forall \ r > ||x||_{\mathcal{L}(\mathcal{H})}$; in other words, $\phi \in L^\infty(X, \mu)$; but then x and $\tilde{\pi}(\phi)$ are two bounded operators on $\mathcal{H}$ which agree on the dense set $L^\infty(X, \mu)$, and hence $x = \tilde{\pi}(\phi)$, as desired.

Case (ii) : Suppose $\pi = \pi_\mu^n$, for some $1 \leq n \leq \aleph_0$. Thus, in this case, we may assume that $\mathcal{H} = \mathcal{H}_\mu^n$ is the (Hilbert space) direct sum of n copies of $\mathcal{H}_\mu = L^2(X, \mu)$. We may - and shall - identify an operator x on $\mathcal{H}_\mu^n$ with its representing matrix $((x(i,j)))$, where $x(i,j) \in \mathcal{L}(\mathcal{H}_\mu) \ \forall i,j \in \mathbb{N}$ such that $1 \leq i,j \leq n$ (as in Proposition 2.5.6). In the notation of the proof of Theorem 3.4.6, we find that $A = \pi_\mu(C(X))^{(n)} = \{x^{(n)} : x \in \pi_\mu(C(X))\}$ (where $x^{(n)}$ is the matrix with (i,j)-th entry being given by x or 0 according as $i = j$ or $i \neq j$; arguing exactly as in the proof of Theorem 3.4.6, (and using the fact, proved in Case (i) above, that $\pi_\mu(C(X))' = \widetilde{\pi_\mu}(L^\infty(X, \mu))$), we find that $y \in A'$ precisely when the entries of the representing matrix satisfy $y(i,j) = \widetilde{\pi_\mu}(\phi_{i,j}) \ \forall i,j$, for some family $\{\phi_{i,j} : 1 \leq i,j \leq n\} \subset L^\infty(X, \mu)$. Again, arguing as in the proof of Theorem 3.4.6, and using the fact, proved in Case (i) above, that $\widetilde{\pi_\mu}(L^\infty(X, \mu))' = \widetilde{\pi_\mu}(L^\infty(X, \mu))$, we then find that $A'' = \{z^{(n)} : z \in \widetilde{\pi_\mu}(L^\infty(X, \mu))\}$; in other words, this says exactly that $A'' = \widetilde{\pi_\mu^n}(L^\infty(X, \mu)) = M$, as desired.

Case (iii): π arbitrary.

By the already proved Theorem 3.5.3(1), we may assume, as in the notation of that theorem, that $\pi = \oplus_{1 \leq n \leq \aleph_0} \pi^n_{\mu|_{E_n}}$. Let $p_n = \tilde{\pi}(1_{E_n})$, for $1 \leq n \leq \aleph_0$, and let $\mathcal{H}_n = ran \ p_n$, so that the Hilbert space underlying the representation π (as well as $\tilde{\pi}$) is given by $\mathcal{H} = \oplus_{1 \leq n \leq \aleph_0} \mathcal{H}_n$.

By 3.5.30, we see that $p_n \in A'' \subset A'$. This means that any operator in A' has the form $\oplus_n x_n$, with $x_n \in \mathcal{L}(\mathcal{H}_n)$; in fact, it is not hard to see that in order that such a direct sum operator belongs to A', it is necessary and sufficient that $x_n \in \pi^n_{\mu|_{E_n}}(C(X))' \ \forall \ 1 \leq n \leq \aleph_0$. From this, it follows that an operator belongs to A'' if and only if it has the form $\oplus_n z_n$, where $z_n \in \pi^n_{\mu|_{E_n}}(C(X))'' \ \forall \ 1 \leq n \leq \aleph_0$; deduce now from (the already established) Case (ii) that there must exist $\phi_n \in L^\infty(E_n, \mu|_{E_n})$ such that $z_n = \widetilde{\pi^n_{\mu|_{E_n}}}(\phi_n) \ \forall n$; in order for $\oplus_n z_n$

to be bounded, it must be the case that the family $\{||z_n|| = ||\phi_n||_{L^\infty(E_n, \mu|_{E_n})}\}_m$ be uniformly bounded, or equivalently, that the equation $\phi = \sum_n 1_{E_n} \phi_n$ define an element of $L^\infty(X, \mu)$; in other words, we have shown that an operator belongs to A'' if and only if it has the form $z = \tilde{\pi}(\phi)$ for some $\phi \in L^\infty(X, \mu)$, i.e., $A'' \subset M$ and the proof of the lemma is complete. □

Remark 3.5.8 The reader might have observed our use of the phrase 'the von Neumann subalgebra generated' by a family of operators. Analogous to our notation in the C^*-case, we shall write $W^*(S)$ for the von Neumann subalgebra generated by a subset $S \subset \mathcal{L}(\mathcal{H})$. It must be clear that for any such set S, $W^*(S)$ is the smallest von Neumann algebra contained in $\mathcal{L}(\mathcal{H})$ and containing S. For a more constructive description of $W^*(S)$, let $\mathcal{A}$ denote the set of linear combinations of 'words in $S \cup S^*$'; then $\mathcal{A}$ is the smallest *-subalgebra containing S, $\overline{(\mathcal{A})} = C^*(S)$, and $W^*(S)$ is the weak (equivalently, strong) operator closure of $\mathcal{A}$. Also, $W^*(S) = (S \cup S^*)''$. □

We need just one more ingredient in order to complete the proof of the Hahn-Hellinger theorem, and that is provided by the following lemma.

Lemma 3.5.9 *Let $\pi = \pi_\mu^n$ and $\mathcal{H} = \mathcal{H}_\mu^n$ denote the (Hilbert space) direct sum of n copies of $\mathcal{H}_\mu = L^2(X, \mu)$, where $1 \leq n \leq \aleph_0$. Consider the following conditions on a family $\{p_i : i \in I\} \subset \mathcal{L}(\mathcal{H})$:*

(i) $\{p_i : i \in I\}$ is a family of pairwise orthogonal projections in $\pi(C(X))'$; and

(ii) if $F \in \mathcal{B}_X$ is any Borel set such that $\mu(F) > 0$, and if $\tilde{\pi}$ is associated to π as in Lemma 3.5.5 (also see Remark 3.5.6), then $p_i \tilde{\pi}(1_F) \neq 0 \ \forall i \in I$.

(a) there exists a family $\{p_i : i \in I\}$ which satisfies conditions (i) and (ii) above, such that $|I| = n$;

(b) if $n < \infty$ and if $\{p_i : i \in I\}$ is any family satisfying (i) and (ii) above, then $|I| \leq n$.

Proof : To start with, note that since $\mathcal{H} = \mathcal{H}_\mu^n$, we may identify operators on $\mathcal{H}$ with their representing matrices $((x(i,j)))$, (where, of course, $x(i,j) \in \mathcal{L}(\mathcal{H}_\mu) \ \forall \ 1 \leq i,j \leq n$). Further, $A = \pi(C(X))$ consists of those operators x whose matrices satisfy $x(i,j) = \delta_{i,j}\pi_\mu(f)$ for some $f \in C(X)$; and our analysis of Case (ii) in the proof of Lemma 3.5.7 shows that A' consists of precisely those operators $y \in \mathcal{L}(\mathcal{H})$ whose matrix entries satisfy $y(i,j) \in \pi_\mu(C(X))' \ = \ \widetilde{\pi_\mu}(L^\infty(X,\mu))$.

(a) For $1 \leq m \leq n$, define $p_m(i,j) \ = \ \delta_{i,j}\delta_{i,m}1$ (where the 1 denotes the identity operator on $\mathcal{H}_\mu$); the concluding remark of the last paragraph shows that $\{p_m : 1 \leq m \leq n\}$ is a family of pairwise orthogonal projections in A' - i.e., this family satisfies condition (i) above; if F is as in (ii), note that $\tilde{\pi}(1_F)$ is the projection $q \in \mathcal{L}(\mathcal{H})$ whose matrix is given by $q(i,j) = \delta_{i,j}\tilde{\pi_\mu}(1_F)$, and it follows that for any $1 \leq m \leq n$, the operator $p_m q$ is not zero, since it has the non-zero entry $\widetilde{\pi_\mu}(1_F)$ in the (m,m)-th place; and (a) is proved.

(b) In order to prove that $|I| \leq n(< \infty)$, it is enough to show that $|I_0| \leq n$ for every finite subset $I_0 \subset I$; hence, we may assume, without loss of generality, that I is a finite set.

Suppose $\{p_m : m \in I\}$ is a family of pairwise orthogonal projections in A' which satisfies condition (ii) above. In view of the concluding remarks of the first paragraph of this proof, we can find a family $\{\phi_m(i,j) : m \in I, 1 \leq i,j \leq n\} \subset L^\infty(X,\mu)$ such that the matrix of the projection p_m is given by $p_m(i,j) \ = \ \widetilde{\pi_\mu}(\phi_m(i,j))$.

Notive now that

$$p_m = p_m p_m^* \ \Rightarrow \ p_m(i,i) = \sum_{j=1}^{n} p_m(i,j)p_m(i,j)^* \ , \ \forall i \in I;$$

in view of the definition of the $\phi_m(i,j)$'s and the fact that $\tilde{\pi}$ is a *-homomorphism, it is fairly easy to deduce now, that we must have

$$\phi_m(i,i) \ = \ \sum_{j=1}^{n} |\phi_m(i,j)|^2 \ a.e., \ \forall 1 \leq i \leq n, \ \forall m \in I \ . \quad (3.5.31)$$

In a similar fashion, if $m,k \in I$ and if $m \neq k$, then notice that

$$p_m p_k = 0 \ \Rightarrow \ p_m p_k^* = 0 \ \Rightarrow \ \sum_{j=1}^{n} p_m(i,j)p_k(l,j)^* = 0, \ \forall i,l \in I,$$

which is seen to imply, as before, that

$$\sum_{j=1}^{n} \phi_m(i,j)\overline{\phi_k(l,j)} \;=\; 0 \text{ a.e.}, \; \forall 1 \le i,l \le n, \; \forall m \ne k \in I \;. \tag{3.5.32}$$

Since I is finite, as is n, and since the union of a finite number of sets of measure zero is also a set of measure zero, we may, assume that the functions $\phi_m(i,j)$ (have been re-defined on some set of measure zero, if necessary, so that they now) satisfy the conditions expressed in equations 3.5.31 and 3.5.32 not just a.e., but in fact, everywhere; thus, we assume that the following equations hold pointwise at every point of X:

$$\begin{aligned} \sum_{j=1}^{n} |\phi_m(i,j)|^2 &= \phi_m(i,i) \; \forall 1 \le i \le n, \; \forall m \in I \\ \sum_{j=1}^{n} \phi_m(i,j)\overline{\phi_k(l,j)} &= 0 \; \forall 1 \le i,l \le n, \; \forall m \ne k \in I \end{aligned} \tag{3.5.33}$$

Now, for each $m \in I$, define $F_m = \{s \in X : (\phi_m(i,i))(s) = 0 \, \forall 1 \le i \le n\}$. Then, if $s \notin F_m$, we can pick an index $1 \le i_m \le n$ such that $(\phi_m(i_m,i_m))(s) \ne 0$; so, if it is possible to pick a point s which does not belong to F_m for all $m \in I$, then we could find vectors

$$v_m \;=\; \begin{bmatrix} (\phi_m(i_m,1))(s) \\ (\phi_m(i_m,2))(s) \\ \vdots \\ (\phi_m(i_m,n))(s) \end{bmatrix} , m \in I$$

which would constitute a set of non-zero vectors in $\mathbb{C}^n$ which are pairwise orthogonal in view of equations 3.5.33; this would show that indeed $|I| \le n$.

So, our proof will be complete once we know that $X \ne \cup_{m \in I} F_m$; we make the (obviously) even stronger assertion that $\mu(F_m) = 0 \; \forall m$. Indeed, note that if $s \in F_m$, then $\phi_m(i,i)(s) = 0 \; \forall 1 \le i \le n$; but then, by equation 3.5.31, we find that $\phi_m(i,j)(s) = 0 \; \forall 1 \le i,j \le n$; this implies that $\phi_m(i,j)1_F = 0$ *a.e.* $\forall \; 1 \le i,j \le n$; this implies that (every entry of the matrix representing the operator) $p_m\tilde{\pi}(1_F)$ is 0; but by the assumption that the family $\{p_m : m \in I\}$ satisfies condition (ii), and so, this can happen only if $\mu(F_m) = 0$. □

We state, as a corollary, the form in which we will need to use Lemma 3.5.9.

COROLLARY 3.5.10 *Let* $\pi = \oplus_{1 \leq n \leq \aleph_0} \pi^n_{\mu|_{E_n}}$. *Suppose* $A \in \mathcal{B}_X$ *is such that* $\mu(A) > 0$, *and let* $1 \leq n < \infty$; *then the following conditions on* A *are equivalent:*
(a) there exists a family $\{p_i : 1 \leq i \leq n\}$ *of pairwise orthogonal projections in* $\pi(C(X))'$ *such that*

(i) $p_i = p_i \tilde{\pi}(1_A)$ $\forall 1 \leq i \leq n$; *and*

(ii) $F \in \mathcal{B}_X, \mu(A \cap F) > 0 \;\Rightarrow\; p_i \tilde{\pi}(1_F) \neq 0$ $\forall 1 \leq i \leq n$.

(b) $A \subset \coprod_{n \leq m \leq \aleph_0} E_m$ *(mod* μ*) - or equivalently,* $\mu(A \cap E_k) = 0$ $\forall$ $1 \leq k < n$.

Proof : If we write $e_n = \tilde{\pi}(1_{E_n})$, it is seen - from the proof of Case (iii) of Lemma 3.5.7, for instance - that any projection $p \in \pi(C(X))'$ has the form $p = \sum_n q_n$, where $q_n = pe_n$ is, when thought of as an operator on $e_n\mathcal{H}$, an element of $\pi^n_{\mu|_{E_n}}(C(X))'$.

$(a) \Rightarrow (b)$: Fix $1 \leq k < \infty$, and suppose $\mu(A \cap E_k) \neq 0$; set $q_i = p_i e_k$ and note that we can apply Lemma 3.5.9(b) to the representation $\pi^k_{\mu|_{A \cap E_k}}$ and the family $\{q_i : 1 \leq i \leq n\}$, to arrive at the concluion $n \leq k$; in other words, if $1 \leq k < n$, then $\mu(A \cap E_k) = 0$.

$(b) \Rightarrow (a)$: Fix any $m \geq n$ such that $\mu(A \cap E_m) > 0$. Then, apply Lemma 3.5.9(a) to the representation $\pi^m_{\mu|_{A \cap E_m}}$ to conclude the existence of a family $\{q_i^{(m)} : 1 \leq i \leq m\}$ of pairwise orthogonal projections in $\pi^m_{\mu|_{A \cap E_m}}(C(X))'$ with the property that $q_i^{(m)} \widetilde{\pi^m_{\mu|_{A \cap E_m}}}(1_F) \neq 0$ whenever F is a Borel subset of $A \cap E_m$ such that $\mu(F) > 0$. We may, and will, regard each $q_i^{(m)}$ as a projection in the big ambient Hilbert space $\mathcal{H}$, such that $q_i^{(m)} = q_i^{(m)} \tilde{\pi}(1_{A \cap E_m})$, so that $q_i^{(m)} \in \pi(C(X))'$. Now, for $1 \leq i \leq n$, if we define

$$p_i = \sum_{\substack{n \leq m \leq \aleph_0 \\ \mu(A \cap E_m) > 0}} q_i^{(m)} ,$$

it is clear that $\{p_i : 1 \leq i \leq n\}$ is a family of pairwise orthogonal projections in $\pi(C(X))'$, and that $p_i = p_i \tilde{\pi}(1_A)$ $\forall 1 \leq i \leq n$; further, if $F \in \mathcal{B}_X$ satisfies $\mu(A \cap F) > 0$, then (since $\mu(A \cap$

$E_k) = 0 \ \forall 1 \leq k < n)$ there must exist an $m \geq n$ such that $\mu(A \cap F \cap E_m) > 0$; in this case, note that $(p_i \tilde{\pi}(1_F))\tilde{\pi}(1_{A \cap E_m}) = q_i^{(m)} \tilde{\pi}(1_{A \cap F \cap E_m})$ which is non-zero by the way in which the $q_i^{(m)}$'s were chosen. Hence it must be that $p_i \tilde{\pi}(1_F) \neq 0$; thus, (a) is verified.

□

We now complete the proof of the Hahn-Hellinger theorem.

Proof of Theorem 3.5.3(2) : Suppose that for $i = 1, 2$, $\mu^{(i)}$ is a probability measure on $(X, \mathcal{B}_X)$, and $\{E_n^{(i)} : 1 \leq i \leq \aleph_0\}$ is a collection of pairwise disjoint Borel sets such that $\mu^{(i)}$ is supported on $\coprod_{1 \leq n \leq \aleph_0} E_n^{(i)}$; and suppose that

$$\oplus_{1 \leq n \leq \aleph_0} \pi^n_{\mu^{(1)}|_{E_n^{(1)}}} \cong \oplus_{1 \leq n \leq \aleph_0} \pi^n_{\mu^{(2)}|_{E_n^{(2)}}}$$

It follows from the proof of Lemma 3.5.5 that a choice for the measure associated to the representation $\oplus_{1 \leq n \leq \aleph_0} \pi^n_{\mu^{(i)}|_{E_n^{(i)}}}$ is given by $\mu^{(i)}$, for $i = 1, 2$; we may conclude from Lemma 3.5.5(b) that the measures $\mu^{(1)}$ and $\mu^{(2)}$ are mutually absolutely continuous; by Lemma 3.5.2(a), it is seen that $\pi_{\mu^{(1)}|_F}$ is equivalent to $\pi_{\mu^{(2)}|_F}$ for any $F \in \mathcal{B}_X$. Hence, we may assume, without loss of generality, that $\mu^{(1)} = \mu^{(2)} = \mu$ (say).

Hence we need to show that if $\{E_n^{(i)} : 1 \leq n \leq \aleph_0\}, i = 1, 2$, are two collections of pairwise disjoint Borel sets such that

$$\mu(X - \coprod_{1 \leq n \leq \aleph_0} E_n^{(i)}) = 0 \ , i = 1, 2, \tag{3.5.34}$$

and

$$\pi_1 \ = \ \oplus_{1 \leq n \leq \aleph_0} \pi^n_{\mu|_{E_n^{(1)}}} \cong \oplus_{1 \leq n \leq \aleph_0} \pi^n_{\mu|_{E_n^{(2)}}} \ = \ \pi_2 \ , \tag{3.5.35}$$

then it is the case that $\mu(E_n^{(1)} \Delta E_n^{(2)}) = 0 \ \forall \ 1 \leq n \leq \aleph_0$.

Let $\mathcal{H}_i$ be the Hilbert space underlying the representation π_i, and let U be a unitary operator $U : \mathcal{H}_1 \to \mathcal{H}_2$ such that $U\pi_1(f) = \pi_2(f)U \ , \ \forall \ f \in C(X)$; then, - see Remark 3.5.6 - also $U\widetilde{\pi_1}(\phi) = \widetilde{\pi_2}(\phi)U \ , \ \forall \phi \in L^\infty(X, \mu)$. In particular, we see that

$$U\widetilde{\pi_1}(1_A)U^* = \widetilde{\pi_2}(1_A) \ \forall \ A \in \mathcal{B}_X \ . \tag{3.5.36}$$

Thus, we need to to show that if $1 \leq k \neq m \leq \aleph_0$, then $\mu(E_m^{(1)} \cap E_k^{(2)}) = 0$. (By equation 3.5.34, this will imply that $E_m^{(1)} \subset E_m^{(2)}$ $(mod\ \mu)$.) By the symmetry of the problem, we may assume, without loss of generality, that $1 \leq k < m \leq \aleph_0$. So fix k, m as above, and choose a finite integer n such that $k < n \leq m$.

Now apply (the implication $(b) \Rightarrow (a)$ of) Corollary 3.5.10 to the representation π_1 and the set $E_m^{(1)}$ to deduce the existence of a family $\{p_i : 1 \leq i \leq n\}$ of pairwise orthogonal projections in $\pi_1(C(X))'$ such that

$$p_i = p_i \widetilde{\pi_1}(1_{E_m^{(1)}}) \ , \ \forall 1 \leq i \leq n \tag{3.5.37}$$

and

$$F \in \mathcal{B}_X, \ \mu(F \cap E_m^{(1)}) > 0 \ \Rightarrow \ p_i \widetilde{\pi_1}(1_F) \neq 0 \ , \ \forall 1 \leq i \leq n \ . \tag{3.5.38}$$

If we now set $q_i = U p_i U^*$, it is easy to deduce - in view of equation 3.5.36 and the consequent fact that $U\pi_1(C(X))'U^* = \pi_2(C(X))'$ - that equations 3.5.37 and 3.5.38 translate into the fact that $\{q_i : 1 \leq i \leq n\}$ is a family of pairwise orthogonal projections in $\pi_2(C(X))'$ such that

$$q_i = q_i \widetilde{\pi_2}(1_{E_m^{(1)}}) \ , \ \forall 1 \leq i \leq n \tag{3.5.39}$$

and

$$F \in \mathcal{B}_X, \ \mu(F \cap E_m^{(1)}) > 0 \ \Rightarrow \ p_i \widetilde{\pi_2}(1_F) \neq 0 \ , \ \forall 1 \leq i \leq n \ . \tag{3.5.40}$$

We may now conclude from (the implication $(a) \Rightarrow (b)$ of) Corollary 3.5.10 - when applied to the representation π_2 and the set $E_m^{(1)}$ - that $\mu(E_m^{(1)} \cap E_k^{(2)}) = 0$, and the proof of the theorem is finally complete. $\square$

In the remainder of this section, we shall re-phrase the results obtained so far in this section using the language of **spectral measures**.

Given a separable representation $\pi : C(X) \to \mathcal{L}(\mathcal{H})$, consider the mapping $\mathcal{B}_X \ni E \mapsto \tilde{\pi}(1_E)$ that is canonically associated to this representation (see Remark 3.5.6). Let us, for convenience, write $P(E) = \tilde{\pi}(1_E)$; thus the assignment $E \mapsto P(E)$ associates a projection in $\mathcal{L}(\mathcal{H})$ to each Borel set E. Further, it follows from the key continuity property of the representation $\tilde{\pi}$ - *viz.* Lemma 3.5.5(a)(iii) - that if $E = \coprod_{n=1}^{\infty} E_n$ is a partition of the Borel set E into countably many Borel sets, then $P(E) = \sum_{n=1}^{\infty} P(E_n)$, meaning that the family $\{P(E_n)\}_{n=1}^{\infty}$ is unconditionally summable with respect to the strong operator topology and has sum $P(E)$. Thus, this is an instance of a spectral measure in the sense of the next definition.

DEFINITION 3.5.11 *A spectral measure on the measurable space $(X, \mathcal{B}_X)$ is a mapping $E \mapsto P(E)$ from $\mathcal{B}_X$, into the set of projections of some Hilbert space $\mathcal{H}$, which satisfies the following conditions:*

(i) $P(\emptyset) = 0$, $P(X) = 1$ (where, of course, 1 denotes the identity operator on $\mathcal{H}$); and

(ii) P is countably additive, meaning that if $\{E_n\}_n$ is a countable collection of pairwise disjoint Borel sets in X with $E = \coprod_n E_n$, then $P(E) = \sum_n P(E_n)$, where the series is interpreted in the strong operator topology on $\mathcal{H}$.

Observe that the strong convergence of the series in (ii) above is equivalent to the weak convergence of the series - see Proposition 2.5.4 - and they are both equivalent to requiring that if $\{E_n\}_n$ is as in (ii) above, then the projections $\{P(E_n)\}_n$ have pairwise orthogonal ranges and the range of $P(E)$ is the direct sum of the ranges of the $P(E_n)$'s.

Given a spectral measure as above, fix $\xi, \eta \in \mathcal{H}$, and notice that the equation

$$\mu_{\xi,\eta}(E) \;=\; \langle P(E)\xi, \eta\rangle \qquad (3.5.41)$$

defines a complex measure $\mu_{\xi,\eta}$ on $(X, \mathcal{B}_X)$. The Cauchy-Schwarz inequality implies that the complex measure $\mu_{\xi,\eta}$ has total variation norm bounded by $||\xi|| \cdot ||\eta||$.

Then it is fairly straightforward to verify that for fixed $f \in C(X)$, the equation

$$B_f(\xi, \eta) \;=\; \int_X f d\mu_{\xi,\eta} \qquad (3.5.42)$$

defines a sesquilinear form on $\mathcal{H}$; further, in view of the statement above concerning the total variation norm of $\mu_{\xi,\eta}$, we find that

$$|B_f(\xi, \eta)| \;\leq\; ||f||_{C(X)} ||\xi|| \; ||\eta|| \; ;$$

hence, by Proposition 2.4.4, we may deduce the existence of a unique bounded operator - call it $\pi(f)$ - on $\mathcal{H}$ such that

$$\langle \pi(f)\xi, \eta \rangle \ = \ \int_X f d\mu_{\xi,\eta} \ .$$

After a little more work, it can be verified that the passage $f \mapsto \pi(f)$ actually defines a representation of $\pi : C(X) \to \mathcal{L}(\mathcal{H})$. It is customary to use such expressions as

$$\pi(f) \ = \ \int_X f dP = \int_X f(x) dP(x)$$

to denote the operator thus obtained from a spectral measure.

Thus, one sees that there is a bijective correspondence between separable representations of $C(X)$ and spectral measures defined on $(X, \mathcal{B}_X)$ and taking values in projection operators in a separable Hilbert space. Thus, for instance, one possible choice for the measure μ that is associated to the representation π (which is, after all, only determined up to mutual absolute continuity) is given, in terms of the spectral measure $P(\cdot)$ by

$$\mu(E) \ = \ \sum_n \langle P(E)\xi_n, \xi_n \rangle \ = \ \sum_n ||P(E)\xi_n||^2 \ ,$$

where $\{\xi_n\}_n$ is an orthonormal basis for $\mathcal{H}$.

Further, the representation which we have, so far in this section, denoted by $\tilde{\pi} : L^\infty(X, \mu) \to \mathcal{L}(\mathcal{H})$ can be expressed, in terms of the spectral measure $P(\cdot)$ thus:

$$\langle \tilde{\pi}(\phi)\xi, \eta \rangle \ = \ \int_X \phi d\mu_{\xi,\eta} \ ,$$

or, equivalently,

$$\tilde{\pi}(\phi) \ = \ \int_X \phi dP \ = \ \int_X \phi(\lambda) dP(\lambda) \ .$$

Hence, the Hahn-Hellinger theorem can be regarded as a classification, up to a natural notion of equivalence, of separable spectral measures on $(X, \mathcal{B}_X)$. We summarise this re-formulation as follows.

Given a probability measure μ defined on $(X, \mathcal{B})$, let $P_\mu : \mathcal{B} \to \mathcal{L}(L^2(X, \mathcal{B}, \mu))$ be the spectral measure defined by

$$\langle P_\mu(E) \ f, \ g \rangle \ = \ \int_E f(x)\overline{g(x)} d\mu(x) \ .$$

For $1 \leq n \leq \aleph_0$, we have a natural spectral measure $P_\mu^n : \mathcal{B} \to \mathcal{L}((L^2(X, \mathcal{B}, \mu))^n)$ obtained by defining $P_\mu^n(E) = \oplus_{k \in I_n} P_\mu(E)$, where $\mathcal{H}^n$ denotes the direct sum of n copies of $\mathcal{H}$ and I_n is any set with cardinality n.

In this notation, we may re-phrase the Hahn-Hellinger theorem thus:

THEOREM 3.5.12 *If X is a compact Hausdorff space and if $P : \mathcal{B} \to \mathcal{L}(\mathcal{H})$ is a 'separable' spectral measure (meaning that $\mathcal{H}$ is separable), then there exists a probability measure $\mu : \mathcal{B} \to [0, 1]$ and a partition $X = \coprod_{0 \leq n \leq \aleph_0} E_n$ of X into measurable sets such that μ is supported on $X - E_0$ and*

$$P \cong \oplus_{1 \leq n \leq \aleph_0} P^n_{\mu|_{E_n}} .$$

If $\tilde{P} : \mathcal{B} \to \mathcal{L}(\tilde{\mathcal{H}})$ is another spectral measure, with associated probability measure $\tilde{\mu}$ and partition $X = \coprod_{0 \leq n \leq \aleph_0} \tilde{E}_n$, then μ and $\tilde{\mu}$ are mutually absolutely continuous, and $\mu(E_n \Delta \tilde{E}_n) = 0 \; \forall \; 0 \leq n \leq \aleph_0$.

REMARK 3.5.13 Let X be a locally compact Hausdorff space, and let $\hat{X}$ denote the one-point compactification of X. Then any spectral measure $P : \mathcal{B}_X \to \mathcal{L}(\mathcal{H})$ may be viewed as a spectral measure P_1 defined on $\mathcal{B}_{\hat{X}}$ with the understanding that $P_1(\{\infty\}) = 0$, or equivalently, $P_1(E) = P(E - \{\infty\})$.

Hence we may deduce that the above formulation of the Hahn-Hellinger theorem is just as valid under the assumption that X is a locally compact Hausdorff space; in particular, we shall use this remark, in case $X = \mathbb{R}$. $\square$

Chapter 4

Some Operator theory

4.1 The spectral theorem

This section is devoted to specialising several results of the preceding chapter to the case of a 'singly generated' commutative C^*-algebra, or equivalently to the case of $C(\Sigma)$ where Σ is a compact subset of the complex plane, or equivalently, to facts concerning a single normal operator on a separable Hilbert space.

We should also remark that we are going to revert to our initial notational convention, whereby x, y, z etc., denote vectors in Hilbert spaces, and A, B, S, T, U etc., denote operators between Hilbert spaces. (We should perhaps apologise to the reader for this 'jumping around'; on the other hand, this was caused by the author's experience with the discomfort faced by beginning graduate students with vectors being ξ, η etc., and operators being x, y, z etc.)

The starting point is the following simple consequence of Proposition 3.3.10.

LEMMA 4.1.1 *The following conditions on a commutative unital C^*-algebra A are equivalent:*

(a) there exists an element $x \in A$ such that $A = C^*(\{1, x\})$ and $\sigma(x) = \Sigma$;

(b) $A \cong C(\Sigma)$.

The next step is the following consequence of the conclusions drawn in §3.5.

THEOREM 4.1.2 (**Spectral theorem** *(bounded case)*

(a) Let $\Sigma \subset \mathbb{C}$ be compact and let $\mathcal{H}$ denote a separable Hilbert space. Then there exists a 1-1 correspondence between (i) normal operators $T \in \mathcal{L}(\mathcal{H})$, such that $\sigma(T) \subset \Sigma$, and (ii) spectral measures $\mathcal{B}_\Sigma \ni E \mapsto P(E) \in \mathcal{P}(\mathcal{H})$ (where we write $\mathcal{P}(\mathcal{H})$ for the set of projection operators onto closed subspaces of $\mathcal{H}$).

This correspondence is given by the equation

$$\langle Tx, y\rangle \;=\; \int_\Sigma \lambda \; d\mu_{x,y}(\lambda) \ ,$$

where $\mu_{x,y}(E) = \langle P(E)x, y\rangle$.

(b) If $T, P(E)$ are as above, then the spectrum of T is the 'support of the spectral measure $P(\cdot)$', meaning that $\lambda \in \sigma(T)$ if and only if $P(U) \neq 0$ for every open neighbourhood of λ.

Proof : (a) If $T \in \mathcal{L}(\mathcal{H})$ is a normal operator such that $\sigma(T) = \Sigma_0 \subset \Sigma$, then according to Proposition 3.3.10, there exists a unique representation $C(\Sigma_0) \ni f \mapsto f(T) \in C^*(\{1, T\}) \subset \mathcal{L}(\mathcal{H})$ such that $f_j(T) = T^j$ if $f_j(\lambda) = \lambda^j, \ j = 0, 1$; from the concluding discussion in §3.5, this representation gives rise to a unique spectral measure $\mathcal{B}_{\Sigma_0} \ni E \mapsto P(E) \in \mathcal{P}(\mathcal{H})$ such that

$$\langle f(T)x, y\rangle \;=\; \int_{\Sigma_0} f d\mu_{x,y} \; \forall f, x, y \ . \tag{4.1.1}$$

We may think of P as a spectral measure being defined for all Borel subsets of Σ (in fact, even for all Borel subsets of $\mathbb{C}$) by the rule $P(E) = P(E \cap \Sigma_0)$.

Conversely, every spectral measure as in (a)(ii) clearly gives rise to a (representation of $C(\Sigma)$ into $\mathcal{H}$ and consequently a) normal operator $T \in \mathcal{L}(\mathcal{H})$ such that equation 4.1.1 is valid for ($f(\lambda) = \lambda^j, j = 0, 1$ and hence for all) $f \in C(\Sigma)$. Since homomorphisms of C^*-algebras 'shrink spectra' - see Lemma 3.4.2(a) - we find that $\sigma(T) \subset \Sigma$.

(b) Let $T = \int_\Sigma \lambda dP(\lambda)$ as in (a) above. Define the measure μ by $\mu(E) = \sum_n 2^{-n}||P(E)e_n||^2$, where $\{e_n\}_n$ is an orthonormal basis for $\mathcal{H}$. Then, by our discussion in §3.5 on spectral measures, we have an isometric *-isomorphism $L^\infty(\Sigma_0, \mu) \ni \phi \mapsto \int_{\Sigma_0} \phi dP \ \in \mathcal{L}(\mathcal{H})$, where $\Sigma_0 = \sigma(T)$. Under this isomorphism, the function $f_1(\lambda) = \lambda, \lambda \in \Sigma_0$ corresponds to T; but it is easy to see that $\lambda_0 \in \sigma_{L^\infty(\Sigma_0,\mu)}(f_1)$ if and only if $\mu(\{\lambda : |\lambda - \lambda_0| < \epsilon\}) > 0$ for every $\epsilon > 0$. (Reason: the negation of this latter condition is clearly equivalent to the statement that the function $\lambda \mapsto \frac{1}{\lambda - \lambda_0}$ belongs to $L^\infty(\Sigma_0, \mu)$.) Since $\mu(E) = 0 \Leftrightarrow P(E) = 0$, the proof of (b) is complete. $\square$

COROLLARY 4.1.3 *A complex number λ_0 belongs to the spectrum of a normal operator $T \in \mathcal{L}(\mathcal{H})$ if and only if λ_0 is an 'approximate eigenvalue for T', meaning that there exists a sequence $\{x_n : n \in \mathbb{N}\}$ of unit vectors in $\mathcal{H}$ such that $||Tx_n - \lambda_0 x_n|| \to 0$.*

Proof : An approximate eigenvalue of any (not necessarily normal) operator must necessarily belong to the spectrum of that operator, since an invertible operator is necessarily bounded below - see Remark 1.5.15.

On the other hand, if T is normal, if $P(\cdot)$ is the associated spectral measure, and if $\lambda_0 \in \sigma(T)$, then $P(U) \neq 0$ for every neighbourhood U of λ_0. In particular, we can find a unit vector x_n belonging to the range of the projection $P(U_n)$, where $U_n = \{\lambda \in \mathbb{C} : |\lambda - \lambda_0| < \frac{1}{n}\}$. Since $|(\lambda - \lambda_0)1_{U_n}(\lambda)| < \frac{1}{n} \ \forall \ \lambda$, we find that $||(T - \lambda_0)P(U_n)|| \leq \frac{1}{n}$, and in particular, $||(T - \lambda_0)x_n|| \leq \frac{1}{n} \ \forall \ n$. $\square$

REMARK 4.1.4 Note that the assumption of normality is crucial in the above proof that every spectral value of a normal operator is an approximate eigenvalue; indeed, if U is a non-unitary isometry - such as the unilateral shift, for instance - then U is not invertible, i.e., $0 \in \sigma(U)$, but 0 is clearly not an approximate eigenvalue of any isometry. $\square$

We now make explicit, something which we have been using all along; and this is the fact that whereas we formerly only had a 'continuous functional calculus' for normal elements of abstract C^*-algebras (and which we have already used with good effect, in Proposition 3.3.11, for instance), we now have a **measurable functional calculus** for normal operators on a separable Hilbert space. We shall only make a few small remarks in lieu of the proof of the following Proposition, which is essentially just a re-statement of results proved in §3.5.

PROPOSITION 4.1.5 *Let T be a normal operator on a separable Hilbert space $\mathcal{H}$. Let $\Sigma = \sigma(T)$ and let $P(\cdot)$ be the unique spectral measure associated to T and let $\mu(E) = \sum_n 2^{-n}||P(E)e_n||^2$, where $\{e_n\}_n$ is any orthonormal basis for $\mathcal{H}$. Then the assignment*

$$\phi \mapsto \phi(T) = \int_\Sigma \phi \, dP$$

*defines an isometric *-isomorphism of $L^\infty(\Sigma, \mu)$ onto the von Neumann subalgebra $W^*(\{1, T\})$ of $\mathcal{L}(\mathcal{H})$ generated by $\{1, T\}$, such that*

(i) $\phi_j(T) = T^j$, where $\phi_j(\lambda) = \lambda^j$, $j = 0, 1$; and

*(ii) the representation is 'weakly continuous' meaning that if $\{\phi_n\}_n$ is a sequence in $L^\infty(X, \mu)$ which converges to ϕ with respect to the 'weak * topology' (see Remark A.5.15), then $\phi_n(T) \to \phi(T)$ in the weak operator topology.*

Further, the representation is uniquely determined by conditions (i) and (ii), and is called the 'measurable functional calculus' for T.

Remarks on the proof : Note that if $\pi : C(\Sigma) \to \mathcal{L}(\mathcal{H})$ is the 'continuous functional calculus' for T, and if $\tilde{\pi}$ and μ are associated with π as in Proposition 3.5.5, then

$$\tilde{\pi}(\phi) = \int_\Sigma \phi dP$$

where $P(E) = \tilde{\pi}(1_E)$ (and we may as well assume that μ is given as in the statement of this theorem).

The content of Lemma 3.5.7 is that the image of this representation of $L^\infty(\Sigma, \mu)$ is precisely the von Neumann algebra generated by $\pi(C(\Sigma)) = C^*(\{1, T\})$.

It is a fairly simple matter (which only involves repeated applications of the Cauchy-Schwarz inequality) to verify that if $\{\phi_n\}_n$ is a sequence in $L^\infty(\Sigma, \mu)$, then to say that the sequence $\{\tilde{\pi}(\phi_n)\}_n$ converges to $\tilde{\pi}(\phi)$ with respect to the weak operator topology is exactly equivalent to requiring that $\int_\Sigma (\phi_n - \phi) f d\mu \to 0$ for every $f \in L^1(\Sigma, \mu)$, and this proves (ii).

The final statement is essentially contained in Remark 3.5.6.

□

We now discuss some simple consequences of the 'measurable functional calculus' for a normal operator.

COROLLARY 4.1.6 *(a) Let $U \in \mathcal{L}(\mathcal{H})$ be a unitary operator. Then there exists a self-adjoint operator $A \in \mathcal{L}(\mathcal{H})$ such that $U = e^{iA}$, where the right hand side is interpreted as the result of the continuous functional calculus for A; further, given any $a \in \mathbb{R}$, we may choose A to satisfy $\sigma(A) \subset [a, a + 2\pi]$.*

(b) If $T \in \mathcal{L}(\mathcal{H})$ is a normal operator, and if $n \in \mathbb{N}$, then there exists a normal operator $A \in \mathcal{L}(\mathcal{H})$ such that $T = A^n$.

Proof : (a) Let $\phi : \mathbb{C} \to \{z \in \mathbb{C} : \ Im \ z \ \in \ [a, a + 2\pi)\}$ be any (measurable) branch of the logarithm - for instance, we might set $\phi(z) = log|z| + i\theta$, if $z = |z|e^{i\theta}$, $a \leq \theta < a + 2\pi$. Setting $A = \phi(U)$, we find - since $e^{\phi(z)} = z$ - that $U = e^{iA}$.

(b) This is proved like (a), by taking some measurable branch of the logarithm defined everywhere in $\mathbb{C}$ and choosing the n-th root as the complex function defined using this choice of logarithm. □

EXERCISE 4.1.7 *Show that if $M \subset \mathcal{L}(\mathcal{H})$ is a von Neumann algebra, then M is generated as a Banach space by the set $\mathcal{P}(M) = \{p \in M : p = p^* = p^2\}$. (Hint: Thanks to the Cartesian decomposition, it is enough to be able to express any self-adjoint element $x = x^* \in M$ as a norm-limit of finite linear combinations of projections in M; in view of Proposition 3.3.11(f), we may even assume that $x \geq 0$; by Proposition A.5.9(3), we can find a sequence $\{\phi_n\}_n$ of simple functions such that $\{\phi_n(t)\}_n$ is a non-decreasing sequence which converges to t, for all $t \in \sigma(x)$, and such that the convergence is uniform on $\sigma(x)$; deduce that $||\phi_n(x) - x|| \to 0$.)*

4.2 Polar decomposition

In this section, we establish the very useful **polar decomposition** for bounded operators on Hilbert space. We begin with a few simple observations and then introduce the crucial notion of a **partial isometry**.

LEMMA 4.2.1 *Let $T \in \mathcal{L}(\mathcal{H}, \mathcal{K})$. Then,*

$$ker\ T \ = \ ker\ (T^*T) \ = \ ker\ (T^*T)^{\frac{1}{2}} \ = \ ran^{\perp}T^* \ . \qquad (4.2.2)$$

In particular, also

$$ker^{\perp}T \ = \ \overline{ran\ T^*} \ .$$

(In the equations above, we have used the notation $ran^{\perp}T^*$ and $ker^{\perp}T$, for $(ran\ T^*)^{\perp}$ and $(ker\ T)^{\perp}$, respectively.)

Proof : First observe that, for arbitrary $x \in \mathcal{H}$, we have

$$||Tx||^2 \ = \ \langle T^*Tx, x\rangle \ = \ \langle (T^*T)^{\frac{1}{2}}x, (T^*T)^{\frac{1}{2}}x\rangle \ = \ ||(T^*T)^{\frac{1}{2}}x||^2 \ , \qquad (4.2.3)$$

whence it follows that $ker\ T \ = \ ker(T^*T)^{\frac{1}{2}}$.

Notice next that

$$\begin{aligned} x \in ran^{\perp}T^* \ &\Leftrightarrow \ \langle x, T^*y\rangle = 0 \ \forall\ y \in \mathcal{K} \\ &\Leftrightarrow \ \langle Tx, y\rangle = 0 \ \forall\ y \in \mathcal{K} \\ &\Leftrightarrow \ Tx = 0 \end{aligned}$$

and hence $ran^{\perp}T^* \ = \ ker\ T$. 'Taking perps' once again, we find - because of the fact that $V^{\perp\perp} = \overline{V}$ for any linear subspace $V \subset \mathcal{K}$ - that the last statement of the Lemma is indeed valid.

Finally, if $\{p_n\}_n$ is any sequence of polynomials with the property that $p_n(0) = 0 \ \forall\ n$ and such that $\{p_n(t)\}$ converges uniformly to $\sqrt{t}$ on $\sigma(T^*T)$, it follows that $||p_n(T^*T) - (T^*T)^{\frac{1}{2}}|| \rightarrow 0$, and hence,

$$x \in \ ker(T^*T) \ \Rightarrow \ p_n(T^*T)x = 0 \ \forall n \ \Rightarrow (T^*T)^{\frac{1}{2}}x = 0$$

and hence we see that also $ker(T^*T) \subset ker(T^*T)^{\frac{1}{2}}$; since the reverse inclusion is clear, the proof of the lemma is complete. □

PROPOSITION 4.2.2 *Let $\mathcal{H}, \mathcal{K}$ be Hilbert spaces; then the following conditions on an operator $U \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ are equivalent:*
(i) $U = UU^*U$;
(ii) $P = U^*U$ is a projection;
(iii) $U|_{ker^{\perp}U}$ is an isometry.

An operator which satisfies the equivalent conditions (i)-(iii) is called a **partial isometry**.

Proof : $(i) \Rightarrow (ii)$: The assumption (i) clearly implies that $P = P^*$, and that $P^2 = U^*UU^*U = U^*U = P$.

$(ii) \Rightarrow (iii)$: Let $\mathcal{M} = ran\ P$. Then notice that, for arbitrary $x \in \mathcal{H}$, we have: $||Px||^2 = \langle Px, x\rangle = \langle U^*Ux, x\rangle = ||Ux||^2$; this clearly implies that $ker\ U = ker\ P = \mathcal{M}^\perp$, and that U is isometric on $\mathcal{M}$ (since P is identity on $\mathcal{M}$).

$(iii) \Rightarrow (ii)$: Let $\mathcal{M} = ker^\perp U$. For $i = 1, 2$, suppose $z_i \in \mathcal{H}$, and $x_i \in \mathcal{M}, y_i \in \mathcal{M}^\perp$ are such that $z_i = x_i + y_i$; then note that

$$\begin{aligned} \langle U^*Uz_1, z_2\rangle &= \langle Uz_1, Uz_2\rangle \\ &= \langle Ux_1, Ux_2\rangle \\ &= \langle x_1, x_2\rangle \ \ (since\ U|_{\mathcal{M}}\ is\ isometric) \\ &= \langle x_1, z_2\rangle\ , \end{aligned}$$

and hence U^*U is the projection onto $\mathcal{M}$.

$(ii) \Rightarrow (i)$: Let $\mathcal{M} = ran\ U^*U$; then (by Lemma 4.2.1) $\mathcal{M}^\perp = ker\ U^*U = ker\ U$, and so, if $x \in \mathcal{M}, y \in \mathcal{M}^\perp$, are arbitrary, and if $z = x + y$, then observe that $Uz = Ux + Uy = Ux = U(U^*Uz)$. □

Remark 4.2.3 Suppose $U \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ is a partial isometrry. Setting $\mathcal{M} = ker^\perp U$ and $\mathcal{N} = ran\ U (= \overline{ran\ U})$, we find that U is identically 0 on $\mathcal{M}^\perp$, and U maps $\mathcal{M}$ isometrically onto $\mathcal{N}$. It is customary to refer to $\mathcal{M}$ as the **initial space**, and to $\mathcal{N}$ as the **final space**, of the partial isometry U.

On the other hand, upon taking adjoints in condition (ii) of Proposition 4.2.2, it is seen that $U^* \in \mathcal{L}(\mathcal{K}, \mathcal{H})$ is also a partial isometry. In view of the preceding lemma, we find that $ker\ U^* = \mathcal{N}^\perp$ and that $ran\ U^* = \mathcal{M}$; thus $\mathcal{N}$ is the inital space of U^* and $\mathcal{M}$ is the final space of U^*.

Finally, it follows from Proposition 4.2.2(ii) (and the proof of that proposition) that U^*U is the projection (of $\mathcal{H}$) onto $\mathcal{M}$ while UU^* is the projection (of $\mathcal{K}$) onto $\mathcal{N}$. □

Exercise 4.2.4 *If $U \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ is a partial isometry with initial space $\mathcal{M}$ and final space $\mathcal{N}$, show that if $y \in \mathcal{N}$, then U^*y is the unique element $x \in \mathcal{M}$ such that $Ux = y$.*

Before stating the polar decomposition theorem, we introduce a convenient bit of notation: if $T \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ is a bounded operator between Hilbert spaces, we shall always use the symbol $|T|$ to denote the unique positive square root of the positive operator $|T|^2 = T^*T \in \mathcal{L}(\mathcal{H})$; thus, $|T| = (T^*T)^{\frac{1}{2}}$.

THEOREM 4.2.5 *(***Polar Decomposition***)*

(a) Any operator $T \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ admits a decomposition $T = UA$ such that

(i) $U \in \mathcal{L}(H, \mathcal{K})$ is a partial isomertry;

(ii) $A \in \mathcal{L}(\mathcal{H})$ is a positive operator; and

(iii) $ker\ T = ker\ U = ker\ A$.

(b) Further, if $T = VB$ is another decomposition of T as a product of a partial isometry V and a positive operator B such that $kerV = kerB$, then necessarily $U = V$ and $B = A = |T|$. This unique decomposition is called the polar decomposition of T.

(c) If $T = U|T|$ is the polar decomposition of T, then $|T| = U^*T$.

Proof : (a) If $x, y \in \mathcal{H}$ are arbitrary, then,

$$\langle Tx, Ty\rangle \;=\; \langle T^*Tx, y\rangle \;=\; \langle |T|^2 x, y\rangle \;=\langle |T|x, |T|y\rangle \;,$$

whence it follows - see Exercise 3.4.12 - that there exists a unique unitary operator $U_0 : \overline{ran\ |T|} \to \overline{ran\ T}$ such that $U_0(|T|x) = Tx\ \forall\ x \in \mathcal{H}$. Let $\mathcal{M} = \overline{ran\ |T|}$ and let $P = P_{\mathcal{M}}$ denote the orthogonal projection onto $\mathcal{M}$. Then the operator $U = U_0 P$ clearly defines a partial isometry with initial space $\mathcal{M}$ and final space $\mathcal{N} = \overline{ran\ T}$ which further satisfies $T = U|T|$ (by definition). It follows from Lemma 4.2.1 that $kerU = ker|T| = kerT$.

(b) Suppose $T = VB$ as in (b). Then V^*V is the projection onto $ker^{\perp}V \;=\; ker^{\perp}B \;=\; \overline{ran\ B}$, which clearly implies that $B = V^*VB$; hence, we see that $T^*T = BV^*VB = B^2$; thus B is a, and hence the, positive square root of $|T|^2$, i.e., $B = |T|$. It then follows that $V(|T|x) = Tx = U(|T|x)\ \forall x$; by continuity, we see that V agrees with U on $\overline{ran\ |T|}$, but since this is precisely the initial space of both partial isometries U and V, we see that we must have $U = V$.

(c) This is an immediate consequence of the definition of U and Exercise 4.2.4. □

EXERCISE 4.2.6 *(1) Prove the 'dual' polar decomposition theorem; i.e., each $T \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ can be uniquely expressed in the form $T = BV$ where $V \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ is a partial isometry, $B \in \mathcal{L}(\mathcal{K})$ is a positive operator and $ker B = ker V^* = ker T^*$. (Hint: Consider the usual polar decomposition of T^*, and take adjoints.)*

(2) Show that if $T = U|T|$ is the (usual) polar decomposition of T, then $U|_{ker^\perp T}$ implements a unitary equivalence between $|T|\ |_{ker^\perp |T|}$ and $|T^*|\ |_{ker^\perp |T^*|}$. (Hint: Write $\mathcal{M} = ker^\perp T$, $\mathcal{N} = ker^\perp T^*$, $W = U|_\mathcal{M}$; then $W \in \mathcal{L}(\mathcal{M}, \mathcal{N})$ is unitary; further $|T^*|^2 = TT^* = U|T|^2U^*$; deduce that if A (resp., B) denotes the restriction of $|T|$ (resp., $|T^*|$) to $\mathcal{M}$ (resp., $\mathcal{N}$), then $B^2 = WA^2W^*$; now deduce, from the uniqueness of the positive square root, that $B = WAW^*$.)

(3) Apply (2) above to the case when $\mathcal{H}$ and $\mathcal{K}$ are finite-dimensional, and prove that if $T \in L(V, W)$ is a linear map of vector spaces (over $\mathbb{C}$), then dim $V = rank(T) + nullity(T)$, where $rank(T)$ and $nullity(T)$ denote the dimensions of the range and null-space, respectively, of the map T.

(4) Show that an operator $T \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ can be expressed in the form $T = WA$, where $A \in \mathcal{L}(\mathcal{H})$ is a positive operator and $W \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ is unitary if and only if $dim(ker\ T) = dim(ker\ T^*)$. (Hint: In order for such a decomposition to exist, show that it must be the case that $A = |T|$ and that the W should agree, on $ker^\perp T$, with the U of the polar decomposition, so that W must map $ker\ T$ isometrically onto $ker\ T^*$.)

(5) In particular, deduce from (4) that in case $\mathcal{H}$ is a finite-dimensional inner product space, then any operator $T \in \mathcal{L}(\mathcal{H})$ admits a decomposition as the product of a unitary operator and a positive operator. (In view of Proposition 3.3.11(b) and (d), note that when $\mathcal{H} = \mathbb{C}$, this boils down to the usual polar decomposition of a complex number.)

Several problems concerning a general bounded operator between Hilbert spaces can be solved in two stages: in the first step, the problem is 'reduced', using the polar decomposition theorem, to a problem concerning positive operators on a Hilbert space; and in the next step, the positive case is settled using the spectral theorem. This is illustrated, for instance, in exercise 4.2.7(2).

EXERCISE 4.2.7 *(1) Recall that a subset Δ of a (real or complex) vector space V is said to be* **convex** *if it contains the 'line segment joining any two of its points'; i.e., Δ is convex if $x, y \in \Delta, 0 \leq t \leq 1 \Rightarrow tx + (1-t)y \in \Delta$.*

(a) If V is a normed (or simply a topological) vector space, and if Δ is a closed subset of V, show that Δ is convex if and only if it contains the mid-point of any two of its points - i.e., Δ is convex if and only if $x, y \in \Delta \Rightarrow \frac{1}{2}(x+y) \in \Delta$. (Hint: The set of dyadic rationals, i.e., numbers of the form $\frac{k}{2^n}$ is dense in $\mathbb{R}$.)

(b) If $\mathcal{S} \subset V$ is a subset of a vector space, show that there exists a smallest convex subset of V which contains $\mathcal{S}$; this set is called the **convex hull** *of the set $\mathcal{S}$ and we shall denote it by the symbol $co(\mathcal{S})$. Show that $co(\mathcal{S}) = \{\sum_{i=1}^n \theta_i x_i : n \in \mathbb{N}, \theta_i \geq 0, \sum_{i=1}^n \theta_i = 1\}$.*

(c) Let Δ be a convex subset of a vector space; show that the following conditions on a point $x \in \Delta$ are equivalent:

(i) $x = \frac{1}{2}(y+z),\ y, z \in \Delta \Rightarrow x = y = z$;

(ii) $x = ty + (1-t)z, 0 < t < 1,\ y, z \in \Delta \Rightarrow x = y = z$.

The point x is called an **extreme point** *of a convex set Δ if $x \in \Delta$ and if x satisfies the equivalent conditions (i) and (ii) above.*

(d) It is a fact, called the Krein-Milman theorem - see [Yos], for instance - that if K is a compact convex subset of a Banach space (or more generally, of a locally convex topological vector space which satisfies appropriate 'completeness conditions'), then $K = \overline{co(\partial_e K)}$, where $\partial_e K$ denotes the set of extreme points of K. Verify the above fact in case $K = ball(\mathcal{H}) = \{x \in \mathcal{H} : ||x|| \leq 1\}$, where $\mathcal{H}$ is a Hilbert space, by showing that $\partial_e(ball\ \mathcal{H}) = \{x \in \mathcal{H} : ||x|| = 1\}$. (Hint: Use the parallelogram law - see Exercise 2.1.3(4).)

(e) Show that $\partial_e(ball\ X) \neq \{x \in X : ||x|| = 1\}$, when $X = \ell^1_n,\ n > 1$. (Thus, not every point on the unit sphere of a normed space need be an extreme point of the unit ball.)

(2) Let $\mathcal{H}$ and $\mathcal{K}$ denote (separable) Hilbert spaces, and let $\mathbb{B} = \{A \in \mathcal{L}(\mathcal{H}, \mathcal{K}) : ||A|| \leq 1\}$ denote the unit ball of $\mathcal{L}(\mathcal{H}, \mathcal{K})$. The aim of the following exercise is to show that an operator $T \in \mathbb{B}$ is an extreme point of $\mathbb{B}$ if and only if either T or T^* is

an isometry. (See (1)(c) above, for the definition of an extreme point.)

(a) Let $\mathbb{B}_+ = \{T \in \mathcal{L}(\mathcal{H}) : T \geq 0, \ ||T|| \leq 1\}$*. Show that* $T \in \partial_e \mathbb{B}_+ \Leftrightarrow T$ *is a projection. (Hint: suppose* P *is a projection and* $P = \frac{1}{2}(A + B), A, B \in \mathbb{B}_+$*; then for arbitrary* $x \in ball(\mathcal{H})$*, note that* $0 \leq \frac{1}{2}(\langle Ax, x\rangle + \langle Bx, x\rangle) \leq 1$*; since* $\partial_e[0, 1] = \{0, 1\}$*, deduce that* $\langle Ax, x\rangle = \langle Bx, x\rangle = \langle Px, x\rangle \ \forall \ x \in (ker \ P \cup \ ran \ P)$*; but* $A \geq 0$ *and ker* $P \subset$ *ker* A *imply that* $A(ran \ P) \subset \ ran \ P$*; similarly also* $B(ran \ P) \subset \ ran \ P$*; conclude (from Exercise 2.4.2) that* $A = B = P$*. Conversely, if* $T \in \mathbb{B}_+$ *and* T *is not a projection, then it must be the case - see Proposition 3.3.11(c) - that there exists* $\lambda \in \sigma(T)$ *such that* $0 < \lambda < 1$*; fix* $\epsilon > 0$ *such that* $(\lambda - 2\epsilon, \lambda + 2\epsilon) \subset (0, 1)$*; since* $\lambda \in \sigma(T)$*, deduce that* $P \neq 0$ *where* $P = 1_{(\lambda-\epsilon,\lambda+\epsilon)}(T)$*; notice now that if we set* $A = T - \epsilon P, B = T + \epsilon P$*, then the choices ensure that* $A, B \in \mathbb{B}_+, T = \frac{1}{2}(A + B)$*, but* $A \neq T \neq B$*, whence* $T \notin \partial_e \mathbb{B}_+$*.)*

(b) Show that the only extreme point of ball $\mathcal{L}(\mathcal{H}) = \{T \in \mathcal{L}(\mathcal{H}) : ||T|| \leq 1\}$ *which is a positive operator is* 1*, the identity operator on* $\mathcal{H}$*. (Hint: Prove that 1 is an extreme point of ball* $\mathcal{L}(\mathcal{H})$ *by using the fact that 1 is an extreme point of the unit disc in the complex plane; for the other implication, by (a) above, it is enough to show that if* P *is a projection which is not equal to 1, then* P *is not an extreme point in ball* $\mathcal{L}(\mathcal{H})$*; if* $P \neq 1$*, note that* $P = \frac{1}{2}(U_+ + U_-)$*, where* $U_\pm = P \pm (1 - P)$*.)*

(c) Suppose $T \in \partial_e \mathbb{B}$*; if* $T = U|T|$ *is the polar decomposition of* T*, show that* $|T| \, |_{\mathcal{M}}$ *is an extreme point of the set* $\{A \in \mathcal{L}(\mathcal{M}) : ||A|| \leq 1\}$*, where* $\mathcal{M} = ker^\perp |T|$*, and hence deduce, from (b) above, that* $T = U$*. (Hint: if* $|T| = \frac{1}{2}(C + D)$*, with* $C, D \in$ *ball* $\mathcal{L}(\mathcal{M})$ *and* $C \neq |T| \neq D$*, note that* $T = \frac{1}{2}(A + B)$*, where* $A = UC, B = UD$*, and* $A \neq T \neq B$*.)*

(d) Show that $T \in \partial_e \mathbb{B}$ *if and only if* T *or* T^* *is an isometry. (Hint: suppose* T *is an isometry; suppose* $T = \frac{1}{2}(A + B)$*, with* $A, B \in \mathbb{B}$*; deduce from (1)(d) that* $Tx = Ax = Bx \ \forall x \in \mathcal{H}$*; thus* $T \in \partial_e \mathbb{B}$*; similarly, if* T^* *is an isometry, then* $T^* \in \partial_e \mathbb{B}$*. Conversely, if* $T \in \partial_e \mathbb{B}$*, deduce from (c) that* T *is a partial isometry; suppose it is possible to find unit vectors* $x \in kerT, \quad y \in kerT^*$*; define* $U_\pm z = Tz \pm \langle z, x\rangle y$*, and note that* $U_\pm$ *are partial isometries which are distinct from* T *and that* $T = \frac{1}{2}(U_+ + U_-)$*.)*

4.3 Compact operators

DEFINITION 4.3.1 *A linear map $T : X \to Y$ between Banach spaces is said to be* **compact** *if it satisfies the following condition: for every bounded sequence $\{x_n\}_n \subset X$, the sequence $\{Tx_n\}_n$ has a subsequence which converges with respect to the norm in Y.*

The collection of compact operators from X to Y is denoted by $\mathcal{K}(X,Y)$ (or simply $\mathcal{K}(X)$ if $X = Y$).

Thus, a linear map is compact precisely when it maps the unit ball of X into a set whose closure is compact - or equivalently, if it maps bounded sets into totally bounded sets; in particular, every compact operator is bounded.

Although we have given the definition of a compact operator in the context of general Banach spaces, we shall really only be interested in the case of Hilbert spaces. Nevertheless, we state our first result for general Banach spaces, after which we shall specialise to the case of Hilbert spaces.

PROPOSITION 4.3.2 *Let X, Y, Z denote Banach spaces.*

(a) $\mathcal{K}(X,Y)$ is a norm-closed subspace of $\mathcal{L}(X,Y)$.

(b) if $A \in \mathcal{L}(Y,Z), B \in \mathcal{L}(X,Y)$, and if either A or B is compact, then AB is also compact.

(c) In particular, $\mathcal{K}(X)$ is a closed two-sided ideal in the Banach algebra $\mathcal{L}(X)$.

Proof : (a) Suppose $A, B \in \mathcal{K}(X,Y)$ and $\alpha \in \mathbb{C}$, and suppose $\{x_n\}$ is a bounded sequence in X; since A is compact, there exists a subsequence - call it $\{y_n\}$ of $\{x_n\}$ - such that $\{Ay_n\}$ is a norm-convergent sequence; since $\{y_n\}$ is a bounded sequence and B is compact, we may extract a further subsequence - call it $\{z_n\}$ - with the property that $\{Bz_n\}$ is norm-convergent. It is clear then that $\{(\alpha A + B)z_n\}$ is a norm-convergent sequence; thus $(\alpha A + B)$ is compact; in other words, $\mathcal{K}(X,Y)$ is a subspace of $\mathcal{L}(X,Y)$.

Suppose now that $\{A_n\}$ is a sequence in $\mathcal{K}(X,Y)$ and that $A \in \mathcal{L}(X,Y)$ is such that $||A_n - A|| \to 0$. We wish to prove that A is compact. We will do this by a typical instance of the 'diagonal argument' described in Remark A.4.9. Thus, suppose

$S_0 = \{x_n\}_n$ is a bounded sequence in X. Since A_1 is compact, we can extract a subsequence $S_1 = \{x_n^{(1)}\}_n$ of S_0 such that $\{Ax_n^{(1)}\}_n$ is convergent in Y. Since A_2 is compact, we can extract a subsequence $S_2 = \{x_n^{(2)}\}_n$ of S_1 such that $\{Ax_n^{(2)}\}_n$ is convergent in Y. Proceeding in this fashion, we can find a sequence $\{S_k\}$ such that $S_k = \{x_n^{(k)}\}_n$ is a subsequence of S_{k-1} and $\{A_k x_n^{(k)}\}_n$ is convergent in Y, for each $k \geq 1$. Let us write $z_n = x_n^{(n)}$; since $\{z_n : n \geq k\}$ is a subsequence of S_k, note that $\{A_k z_n\}_n$ is a convergent sequence in Y, for every $k \geq 1$.

The proof of (a) will be completed once we establish that $\{Az_n\}_n$ is a Cauchy sequence in Y. Indeed, suppose $\epsilon > 0$ is given; let $K = 1 + sup_n ||z_n||$; first pick an integer N such that $||A_N - A|| < \frac{\epsilon}{3K}$; next, choose an integer n_0 such that $||A_N z_n - A_N z_m|| < \frac{\epsilon}{3}$ $\forall\ n, m \geq n_0$; then observe that if $n, m \geq n_0$, we have:

$$\begin{aligned} ||Az_n - Az_m|| &\leq ||(A - A_N)z_n|| + ||A_N z_n - A_N z_m|| \\ &\quad + ||(A_N - A)z_m|| \\ &\leq \frac{\epsilon}{3K}K + \frac{\epsilon}{3} + \frac{\epsilon}{3K}K \\ &= \epsilon\ . \end{aligned}$$

(b) Let $\mathbb{B}$ denote the unit ball in X; we need to show that $(AB)(\mathbb{B})$ is totally bounded (see Definition A.4.6); this is true in case (i) A is compact, since then $B(\mathbb{B})$ is bounded, and A maps bounded sets to totally bounded sets, and (ii) B is compact, since then $B(\mathbb{B})$ is totally bounded, and A (being bounded) maps totally bounded sets to totally bounded sets. □

COROLLARY 4.3.3 *Let $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, where $\mathcal{H}_i$ are Hilbert spaces. Then*

(a) T is compact if and only if $|T|$ $(= (T^*T)^{\frac{1}{2}}$) is compact;

(b) in particular, T is compact if and only if T^* is compact.

Proof : If $T = U|T|$ is the polar decomposition of T, then also $U^*T = |T|$ - see Theorem 4.2.5; so each of T and $|T|$ is a multiple of the other. Now appeal to Proposition 4.3.2(b) to deduce (a) above. Also, since $T^* = |T|U^*$, we see that the compactness of T implies that of T^*; and (b) follows from the fact that we may interchange the roles of T and T^*. □

Recall that if $T \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ and if $\mathcal{M}$ is a subspace of $\mathcal{H}$, then T is said to be 'bounded below' on $\mathcal{M}$ if there exists an $\epsilon > 0$ such that $||Tx|| \geq \epsilon ||x|| \ \forall \ x \in \mathcal{M}$.

LEMMA 4.3.4 *If $T \in \mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$ and if T is bounded below on a subspace $\mathcal{M}$ of $\mathcal{H}_1$, then $\mathcal{M}$ is finite-dimensional.*

In particular, if $\mathcal{N}$ is a closed subspace of $\mathcal{H}_2$ such that $\mathcal{N}$ is contained in the range of T, then $\mathcal{N}$ is finite-dimensional.

Proof : If T is bounded below on $\mathcal{M}$, then T is also bounded below (by the same constant) on $\overline{\mathcal{M}}$; we may therefore assume, without loss of generality, that $\mathcal{M}$ is closed. If $\mathcal{M}$ contains an infinite orthonormal set, say $\{e_n : n \in \mathbb{N}\}$, and if T is bounded below by ϵ on $\mathcal{M}$, then note that $||Te_n - Te_m|| \geq \epsilon\sqrt{2} \ \forall n \neq m$; then $\{e_n\}_n$ would be a bounded sequence in $\mathcal{H}$ such that $\{Te_n\}_n$ would have no Cauchy subsequence, thus contradicting the assumed compactness of T; hence $\mathcal{M}$ must be finite-dimensional.

As for the second assertion, let $\mathcal{M} = T^{-1}(\mathcal{N}) \cap (ker^{\perp}T)$; note that T maps $\mathcal{M}$ 1-1 onto $\mathcal{N}$; by the open mapping theorem, T must be bounded below on $\mathcal{M}$; hence by the first assertion of this Lemma, $\mathcal{M}$ is finite-dimensional, and so also is $\mathcal{N}$. $\square$

The purpose of the next exercise is to convince the reader of the fact that compactness is an essentially 'separable phenomenon', so that we may - and will - restrict ourselves in the rest of this section to separable Hilbert spaces.

EXERCISE 4.3.5 *(a) Let $T \in \mathcal{L}(\mathcal{H})$ be a positive operator on a (possibly non-separable) Hilbert space $\mathcal{H}$. Let $\epsilon > 0$ and let $\mathcal{S}_\epsilon = \{f(T)x : f \in C(\sigma(T)),\ f(t) = 0 \ \forall t \in [0, \epsilon]\}$. If $\mathcal{M}_\epsilon = [\mathcal{S}_\epsilon]$ denotes the closed subspace generated by $\mathcal{S}_\epsilon$, then show that $\mathcal{M}_\epsilon \subset$ ran T. (Hint: let $g \in C(\sigma(T))$ be any continuous function such that $g(t) = t^{-1} \ \forall \ t \geq \frac{\epsilon}{2}$; for instance, you could take*

$$g(t) \ = \ \begin{cases} \frac{1}{t} & if \ t \geq \frac{\epsilon}{2} \\ \frac{4t}{\epsilon^2} & if \ 0 \leq t \leq \frac{\epsilon}{2} \end{cases} \quad ;$$

then notice that if $f \in C(\sigma(T))$ satisfies $f(t) = 0 \ \forall t \leq \epsilon$, then $f(t) = tg(t)f(t) \ \forall t$; deduce that $\mathcal{S}_\epsilon$ is a subset of $\mathcal{N} = \{z \in \mathcal{H} : z = Tg(T)z\}$; but $\mathcal{N}$ is a closed subspace of $\mathcal{H}$ which is contained in ran T.)

(b) Let $T \in \mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$, where $\mathcal{H}_1, \mathcal{H}_2$ are arbitrary (possibly non-separable) Hilbert spaces. Show that $ker^\perp T$ and $\overline{ran\ T}$ are separable Hilbert spaces. (Hint: Let $T = U|T|$ be the polar decomposition, and let $\mathcal{M}_\epsilon$ be associated to $|T|$ as in (a) above; show that $U(\mathcal{M}_\epsilon)$ is a closed subspace of ran T and deduce from Lemma 4.3.4 that $\mathcal{M}_\epsilon$ is finite-dimensional; note that $ker^\perp T = ker^\perp |T|$ is the closure of $\cup_{n=1}^\infty \mathcal{M}_{\frac{1}{n}}$, and that $\overline{ran\ T} = U(ker^\perp T)$.)

In the following discussion of compact operators between Hilbert spaces, we shall make the standing assumption that all Hilbert spaces considered here are separable. This is for two reasons: (a) by Exercise 4.3.5, this is no real loss of generality; and (b) we can feel free to use the measurable functional calculus for normal operators, whose validity was established - in the last chapter - only under the standing assumption of separability.

PROPOSITION 4.3.6 *The following conditions on an operator $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ are equivalent:*

(a) T is compact;

(b) $|T|$ is compact;

(c) if $\mathcal{M}_\epsilon = ran\ 1_{[\epsilon,\infty)}(|T|)$, then $\mathcal{M}_\epsilon$ is finite-dimensional, for every $\epsilon > 0$;

(d) there exists a sequence $\{T_n\}_{n=1}^\infty \subset \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ such that (i) $||T_n - T|| \to 0$, and (ii) ran T_n is finite-dimensional, for each n;

(e) ran T does not contain any infinite-dimensional closed subspace of $\mathcal{H}_2$.

Proof : For $\epsilon > 0$, let us use the notation $1_\epsilon = 1_{[\epsilon,\infty)}$ and $P_\epsilon = 1_\epsilon(|T|)$.

$(a) \Rightarrow (b)$: See Corollary 4.3.3.

$(b) \Rightarrow (c)$: Since $t \geq \epsilon 1_\epsilon(t)\ \forall\ t \geq 0$, we find easily that $|T|$ is bounded below (by ϵ) on *ran* P_ϵ, and (c) follows from Lemma 4.3.4.

$(c) \Rightarrow (d)$: Define $T_n = TP_{\frac{1}{n}}$; notice that $0 \leq t(1 - 1_{\frac{1}{n}}(t)) \leq \frac{1}{n}\ \forall\ t \geq 0$; conclude that $||\ |T|(1 - 1_{\frac{1}{n}}(|T|))|| \leq \frac{1}{n}$; if $T = U|T|$ is the polar decomposition of T, deduce that $||T - T_n|| \leq \frac{1}{n}$; finally, the condition (c) clearly implies that each ($P_{\frac{1}{n}}$ and consequently) T_n has finite-dimensional range.

$(d) \Rightarrow (a)$: In view of Proposition 4.3.2(a), it suffices to show that each T_n is a compact operator; but any bounded operator with finite-dimensional range is necessarily compact, since any bounded set in a finite-dimensional space is totally bounded.

$(a) \Rightarrow (e)$: See Lemma 4.3.4.

$(e) \Rightarrow (c)$: Pick any bounded measurable function g such that $g(t) = \frac{1}{t}$, $\forall t \geq \epsilon$; then $tg(t) = 1 \, \forall t \geq 1$; deduce that $|T|g(|T|)x = x$, $\forall \, x \in \mathcal{M}_\epsilon$, and hence that $\mathcal{M}_\epsilon = |T|(\mathcal{M}_\epsilon)$ is a closed subspace of (ran $|T|$, and consequently of) the initial space of the partial isometry U; deduce that $T(\mathcal{M}_\epsilon) = U(\mathcal{M}_\epsilon)$ is a closed subspace of ran T; by condition (e), this implies that $T(\mathcal{M}_\epsilon)$ is finite-dimensional. But $|T|$ and consequently T is bounded below (by ϵ) on $\mathcal{M}_\epsilon$; in particular, T maps $\mathcal{M}_\epsilon$ 1-1 onto $T(\mathcal{M}_\epsilon)$; hence $\mathcal{M}_\epsilon$ is finite-dimensional, as desired. □

We now discuss normal compact operators.

PROPOSITION 4.3.7 *Let $T \in \mathcal{K}(\mathcal{H})$ be a normal (compact) operator on a separable Hilbert space, and let $E \mapsto P(E) = 1_E(T)$ be the associated spectral measure.*

(a) If $\epsilon > 0$, let $P_\epsilon = P(\{\lambda \in \mathbb{C} : |\lambda| \geq \epsilon\})$ denote the spectral projection associated to the complement of an ϵ-neighbourhood of 0. Then ran P_ϵ is finite-dimensional.

(b) If $\Sigma = \sigma(T) - \{0\}$, then

(i) Σ is a countable set;

(ii) $\lambda \in \Sigma \Rightarrow \lambda$ is an eigenvalue of finite multiplicity; i.e., $0 < dim \; ker(T - \lambda) < \infty$;

(iii) the only possible accumulation point of Σ is 0; and

(iv) there exist scalars $\lambda_n \in \Sigma, n \in N$ and an orthonormal basis $\{x_n : n \in N\}$ of ran $P(\Sigma)$ such that

$$Tx = \sum_{n \in N} \lambda_n \, \langle x, x_n \rangle \, x_n \; , \; \forall x \in \mathcal{H} \; .$$

Proof : (a) Note that the function defined by the equation

$$g(\lambda) = \begin{cases} \frac{1}{\lambda} & if \; |\lambda| \geq \epsilon \\ 0 & \text{otherwise} \end{cases}$$

is a bounded measurable function on $\sigma(T)$ such that $g(\lambda)\lambda = 1_{F_\epsilon}(\lambda)$, where $F_\epsilon = \{z \in \mathbb{C} : |z| \geq \epsilon\}$. It follows that $g(T)T =$

$Tg(T) = P_\epsilon$. Hence *ran* P_ϵ is contained in *ran* T, and the desired conclusion follows from Proposition 4.3.6(e).

(b) Let μ be the measure defined by $\mu(E) = \sum_n ||P(E)e_n||^2$, where $\{e_n\}_n$ is some orthonormal basis for $\mathcal{H}$. Then the measurable functional calculus yields an embedding of $L^\infty(F_\epsilon, \mathcal{B}_{F_\epsilon}, \mu|_{F_\epsilon})$ into $\mathcal{L}(ran\ P_\epsilon)$; it follows from (a) above that $L^\infty(F_\epsilon, \mathcal{B}_{F_\epsilon}, \mu|_{F_\epsilon})$ is finite-dimensional, where F_ϵ is as in the last paragraph. This implies - see Exercise 4.3.8, below - that there exists a finite set $\{\lambda_1, \cdots, \lambda_n\} \subset F_\epsilon$ such that $\mu|_{F_\epsilon} = \sum_{i=1}^n \mu_{\{\lambda_i\}}$; assume, without loss of generality, that $\mu(\{\lambda_i\}) > 0\ \forall i$; then $P_\epsilon = \sum_{i=1}^n P_i$ is a decomposition of P_ϵ as a finite sum of non-zero (pairwise orthogonal) projections, where $P_i = P(\{\lambda_i\})$; since P_i is the projection onto $ker(T - \lambda_i)$ (see Exercise 4.3.8), we thus find that (i) $\{\lambda_i : 1 \leq i \leq n\} = \Sigma \cap F_\epsilon$; and (ii) each λ_i is an eigenvalue of T which has finite multiplicity.

By allowing ϵ to decrease to 0 through a countable sequence of values, we thus find that we have proved (i) - (iii) (since a countable union of finite sets is countable). For (iv), observe that if $\Sigma = \{\lambda_n : n \in N\}$, then $\lambda = \sum_{n \in N} \lambda_n 1_{\{\lambda_n\}}(\lambda)\ \mu -$ *a.e.*.; hence $T = \sum_{\lambda \in \Sigma} \lambda P(\{\lambda\})$; finally, if $\{x_n(\lambda) : n \in N_\lambda\}$ is an orthonormal basis for *ran* $P(\{\lambda\})$, for each $\lambda \in \Sigma$, then $P(\{\lambda\})x = \sum_{n \in N_\lambda} \langle x, x_n(\lambda) \rangle\ x_n(\lambda)\ \forall\ x \in \mathcal{H}$; letting $\{x_n\}_{n \in N}$ be an enumeration of $\cup_{\lambda \in \Sigma} \{x_n(\lambda) : n \in N_\lambda\}$, we find that we do indeed have the desired decomposition of T. $\square$

EXERCISE 4.3.8 *Let X be a compact Hausdorff space and let $\mathcal{B}_X \ni E \mapsto P(E)$ be a spectral measure; let μ be a measure which is 'mutually absolutely continuous' with respect to P - thus, for instance, we may take $\mu(E) = \sum ||P(E)e_n||^2$, where $\{e_n\}_n$ is some orthonormal basis for the underlying Hilbert space $\mathcal{H}$ - and let $\pi : C(X) \to \mathcal{L}(\mathcal{H})$ be the associated representation.*

(a) Show that the following conditions are equivalent:

(i) $\mathcal{H}$ is finite-dimensional;

(ii) there exists a finite set $F \subset X$ such that $\mu = \mu|_F$, and such that ran $P(\{x\})$ is finite-dimensional, for each $x \in F$.

(b) If $x_0 \in X$, show that the following conditions on a vector $\xi \in \mathcal{H}$ are equivalent:

(i) $\pi(f)\xi = f(x_0)\xi\ \forall\ f \in C(X)$;

(ii) $\xi \in$ ran $P(\{x_0\})$.

REMARK 4.3.9 In the notation of the Proposition 4.3.7, note that if $\lambda \in \Sigma$, then λ features (as some λ_n) in the expression for T that is displayed in Proposition 4.3.7(iv); in fact, the number of n for which $\lambda = \lambda_n$ is exactly the dimension of (the eigenspace corresponding to λ, which is) $ker(T - \lambda)$. Furthermore, there is nothing canonical about the choice of the sequence $\{x_n\}$; all that can be said about the sequence $\{x_n\}_n$ is that if $\lambda \in \Sigma$ is fixed, then $\{x_n : \lambda_n = \lambda\}$ is an orthonormal basis for $ker(T - \lambda)$. In fact, the 'canonical formulation' of Proposition 4.3.7(iv) is this: $T = \sum_{\lambda \in \Sigma} \lambda P_\lambda$, where $P_\lambda = 1_{\{\lambda\}}(T)$.

The λ_n's admit a 'canonical choice' in some special cases; thus, for instance, if T is positive, then Σ is a subset of $(0, \infty)$ which has no limit points except possibly for 0; so the members of Σ may be arranged in non-increasing order; further,in the expression in Proposition 4.3.7(iv), we may assume that $N = \{1, 2, \cdots, n\}$ or $N = \{1, 2, \cdots\}$, according as $dim\ ran\ |T| = n$ or $\aleph_0$, and also $\lambda_1 \geq \lambda_2 \geq \cdots$. In this case, it must be clear that if λ_1 is an eigenvalue of 'multiplicity m_1' - meaning that $dim\ ker(T - \lambda_1) = m_1$, then $\lambda_n = \lambda_1$ for $1 \leq n \leq m_1$; and similar statements are valid for the 'multiplicities' of the remaining eigenvalues of T. □

We leave the proof of the following Proposition to the reader, since it is obtained by simply piecing together our description (see Proposition 4.3.7 and Remark 4.3.9) of normal compact operators, and the polar decomposition to arrive at a canonical 'singular value decomposition' of a general compact operator.

PROPOSITION 4.3.10 *Let $T \in \mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$. Then T admits the following decomposition:*

$$Tx = \sum_{n \in N} \lambda_n \left(\sum_{k \in I_n} \langle x, x_k^{(n)} \rangle y_k^{(n)} \right) , \tag{4.3.4}$$

where

(i) $\sigma(|T|) - \{0\} = \{\lambda_n : n \in N\}$, where N is some countable set; and

(ii) $\{x_k^{(n)} : k \in I_n\}$ (resp., $\{y_k^{(n)} : k \in I_n\}$) is an orthonormal basis for $ker(|T| - \lambda_n)$ (resp., $T(ker(|T| - \lambda_n)) = ker(|T^*| - \lambda_n)$).

In particular, we may assume that $N = \{1, 2, \cdots, n\}$ *or* $\{1, 2, \cdots\}$ *according as dim ran* $|T| < \infty$ *or* $\aleph_0$*, and that* $\lambda_1 \geq \lambda_2 \geq \cdots$*; if the sequence* $\{s_n = s_n(T)\}$ *- which is indexed by* $\{n \in \mathbb{N} : 1 \leq n \leq dim\ \mathcal{H}_1\}$ *in case* $\mathcal{H}_1$ *is finite-dimensional, and by* $\mathbb{N}$ *if* $\mathcal{H}_1$ *is infinite-dimensional - is defined by*

$$s_n = \begin{cases} \lambda_1 & if\ 0 < n \leq card(I_1) \\ \lambda_2 & if\ card(I_1) < n \leq (card(I_1) + card(I_2)) \\ \cdots & \\ \lambda_m & if\ \sum_{1 \leq k < m} card(I_k) < n \leq \sum_{1 \leq k \leq m} card(I_k) \\ 0 & if\ \sum_{k \in N} card(I_k) < n \end{cases} \tag{4.3.5}$$

then we obtain a non-increasing sequence

$$s_1(T) \geq s_2(T) \geq \cdots \geq s_n(T) \geq \cdots$$

of (uniquely defined) non-negative real numbers, called the sequence of **singular values** *of the compact operator* T.

Some more useful properties of singular values are listed in the following exercises.

Exercise 4.3.11 *(1) Let* $T \in \mathcal{L}(\mathcal{H})$ *be a positive compact operator on a Hilbert space. In this case, we write* $\lambda_n = s_n(T)$*, since* ($T = |T|$ *and consequently) each* λ_n *is then an eigenvalue of* T*. Show that*

$$\lambda_n = \max_{dim \mathcal{M} \leq n} \min\{\langle Tx, x\rangle : x \in \mathcal{M}, ||x|| = 1\} , \tag{4.3.6}$$

where the the maximum is to be interpreted as a supremum, over the collection of all subspaces $\mathcal{M} \subset \mathcal{H}$ *with appropriate dimension, and part of the assertion of the exercise is that this supremum is actually attained (and is consequently a maximum); in a similar fashion, the minimum is to be interpreted as an infimum which is attained. (This identity is called the* **max-min principle** *and is also referred to as the 'Rayleigh-Ritz principle'.) (Hint: To start with, note that if* $\mathcal{M}$ *is a finite-dimensional subspace of* $\mathcal{H}$*, then the minimum in equation 4.3.6 is attained since the unit sphere of* $\mathcal{M}$ *is compact and* T *is continuous. Let* $\{x_n : n \in \mathbb{N}\}$ *be an orthonormal set in* $\mathcal{H}$ *such that* $Tx =$

$\sum_{n \in \mathbb{N}} \lambda_n \langle x, x_n \rangle x_n$, *- see Proposition 4.3.7 and Remark 4.3.9. Define* $\mathcal{M}_n = [\{x_j : 1 \leq j \leq n\}]$*; observe that* $\lambda_n = \min\{\langle Tx, x\rangle : x \in \mathcal{M}_n, ||x|| = 1\}$*; this proves the inequality* $\leq$ *in 4.3.6. Conversely, if dim* $\mathcal{M} \leq n$*, argue that there must exist a unit vector* $x_0 \in \mathcal{M} \cap \mathcal{M}_{n-1}^{\perp}$ *(since the projection onto* $\mathcal{M}_{n-1}$ *cannot be injective on* $\mathcal{M}$ *), to conclude that* $\min\{\langle Tx, x\rangle : x \in \mathcal{M}, ||x|| = 1\} \leq \lambda_n$*.)*

(2) If $T : \mathcal{H}_1 \to \mathcal{H}_2$ *is a compact operator between Hilbert spaces, show that*

$$s_n(T) \;=\; \max_{dim \mathcal{M} \leq n} \min\{||Tx|| : x \in \mathcal{M}, ||x|| = 1\} \; . \qquad (4.3.7)$$

(Hint: Note that $||Tx||^2 = \langle |T|^2 x, x\rangle$*, apply (1) above to* $|T|^2$*, and note that* $s_n(|T|^2) = s_n(|T|)^2$*.)*

(3) If $T \in \mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$ *and if* $s_n(T) > 0$*, then show that* $n \leq$ *dim* $\mathcal{H}_2$ *(so that* $s_n(T^*)$ *is defined) and* $s_n(T^*) = s_n(T)$*. (Hint: Use Exercise 4.2.6(2).)*

We conclude this discussion with a discussion of an important class of compact operators.

LEMMA 4.3.12 *The following conditions on a linear operator* $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ *are equivalent:*

(i) $\sum_n ||Te_n||^2 < \infty$ *, for some orthonormal basis* $\{e_n\}_n$ *of* $\mathcal{H}_1$*;*

(ii) $\sum_m ||T^* f_m||^2 < \infty$ *, for every orthonormal basis* $\{f_m\}_m$ *of* $\mathcal{H}_2$*.*

(iii) $\sum_n ||Te_n||^2 < \infty$ *, for some orthonormal basis* $\{e_n\}_n$ *of* $\mathcal{H}_1$*.*

If these equivalent conditions are satisfied, then the sums of the series in (ii) and (iii) are independent of the choice of the orthonormal bases and are all equal to one another.

Proof: If $\{e_n\}_n$ (resp., $\{f_m\}_m$) is any orthonormal basis for $\mathcal{H}_1$ (resp., $\mathcal{H}_2$), then note that

$$\sum_n ||Te_n||^2 \;=\; \sum_n \sum_m |\langle Te_n, f_m\rangle|^2$$

$$= \sum_m \sum_n |\langle T^* f_m, e_n \rangle|^2$$
$$= \sum_m ||T^* f_m||^2 ,$$

and all the assertions of the proposition are seen to follow. □

DEFINITION 4.3.13 *An operator $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ is said to be a* **Hilbert-Schmidt** *operator if it satisfies the equivalent conditions of Lemma 4.3.12, and the* **Hilbert-Schmidt norm** *of such an operator is defined to be*

$$||T||_2 = \left(\sum_n ||Te_n||^2 \right)^{\frac{1}{2}} , \tag{4.3.8}$$

where $\{e_n\}_n$ is any orthonormal basis for $\mathcal{H}_1$. The collection of all Hilbert-Schmidt operators from $\mathcal{H}_1$ to $\mathcal{H}_2$ will be denoted by $\mathcal{L}^2(\mathcal{H}_1, \mathcal{H}_2)$.

Some elementary properties of the class of Hilbert-Schmidt operators are contained in the following proposition.

PROPOSITION 4.3.14 *Suppose $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, $S \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_3)$, where $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3$ are Hilbert spaces.*

(a) $T \in \mathcal{L}^2(\mathcal{H}_1, \mathcal{H}_2) \Rightarrow T^* \in \mathcal{L}^2(\mathcal{H}_2, \mathcal{H}_1)$; and furthermore, $||T^*||_2 = ||T||_2 \geq ||T||_\infty$, where we write $|| \cdot ||_\infty$ to denote the usual operator norm;

(b) if either S or T is a Hilbert-Schmidt operator, so is ST, and

$$||ST||_2 \leq \begin{cases} ||S||_2 ||T||_\infty & if\ S \in \mathcal{L}^2(\mathcal{H}_2, \mathcal{H}_3) \\ ||S||_\infty ||T||_2 & if\ T \in \mathcal{L}^2(\mathcal{H}_1, \mathcal{H}_2) \ ; \end{cases} \tag{4.3.9}$$

(c) $\mathcal{L}^2(\mathcal{H}_1, \mathcal{H}_2) \subset \mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$; and conversely,

(d) if $T \in \mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$, then T is a Hilbert-Schmidt operator if and only if $\sum_n s_n(T)^2 < \infty$; in fact,

$$||T||_2^2 = \sum_n s_n(T)^2 .$$

Proof : (a) The equality $||T||_2 = ||T^*||_2$ was proved in Lemma 4.3.12. If x is any unit vector in $\mathcal{H}_1$, pick an orthonormal basis $\{e_n\}_n$ for $\mathcal{H}_1$ such that $e_1 = x$, and note that

$$||T||_2 = \left(\sum_n ||Te_n||^2 \right)^{\frac{1}{2}} \geq ||Tx|| ;$$

since x was an arbitrary unit vector in $\mathcal{H}_1$, deduce that $||T||_2 \geq ||T||_\infty$, as desired.

(b) Suppose T is a Hilbert-Schmidt operator; then, for an arbitrary orthonormal basis $\{e_n\}_n$ of $\mathcal{H}_1$, we find that

$$\sum_n ||STe_n||^2 \leq ||S||_\infty^2 \sum_n ||Te_n||^2 ,$$

whence we find that ST is also a Hilbert-Schmidt operator and that $||ST||_2 \leq ||S||_\infty ||T||_2$; if T is a Hilbert-Schmidt operator, then, so is T^*, and by the already proved case, also S^*T^* is a Hilbert-Schmidt operator, and

$$||TS||_2 = ||(TS)^*||_2 \leq ||S^*||_\infty ||T^*||_2 = ||S||_\infty ||T||_2 .$$

(c) Let $\mathcal{M}_\epsilon = ran\ 1_{[\epsilon,\infty)}(|T|)$; then $\mathcal{M}_\epsilon$ is a closed subspace of $\mathcal{H}_1$ on which T is bounded below, by ϵ; so, if $\{e_1, \cdots, e_N\}$ is any orthonormal set in $\mathcal{M}_\epsilon$, we find that $N\epsilon^2 \leq \sum_{n=1}^N ||Te_n||^2 \leq ||T||_2^2$, which clearly implies that $dim\ \mathcal{M}_\epsilon$ is finite (and can not be greater than $\left(\frac{||T||_2}{\epsilon}\right)^2$). We may now infer from Proposition 4.3.6 that T is necessarily compact.

(d) Let $Tx = \sum_n s_n(T)\langle x, x_n\rangle y_n$ for all $x \in \mathcal{H}_1$, as in Proposition 4.3.10, for an appropriate orthonormal (finite or infinite) sequence $\{x_n\}_n$ (resp., $\{y_n\}_n$) in $\mathcal{H}_1$ (resp., in $\mathcal{H}_2$). Then notice that $||Tx_n|| = s_n(T)$ and that $Tx = 0$ if $x \perp x_n\ \forall n$. If we compute the Hilbert-Schmidt norm of T with respect to an orthonormal basis obtained by extending the orthonormal set $\{x_n\}_n$, we find that $||T||_2^2 = \sum_n s_n(T)^2$, as desired. □

Probably the most useful fact concerning Hilbert-Schmidt operators is their connection with integral operators. (Recall that a measure space $(Z, \mathcal{B}_Z, \lambda)$ is said to be σ-finite if there exists a partition $Z = \coprod_{n=1}^\infty E_n$, such that $E_n \in \mathcal{B}_Z, \mu(E_n) < \infty\ \forall n$. The reason for our restricting ourselves to σ-finite measure spaces is that it is only in the presence of some such hypothesis that Fubini's theorem - see Proposition A.5.18(d)) regarding product measures is applicable.)

PROPOSITION 4.3.15 *Let* $(X, \mathcal{B}_X, \mu)$ *and* $(Y, \mathcal{B}_Y, \nu)$ *be* σ*-finite measure spaces. Let* $\mathcal{H} = L^2(X, \mathcal{B}_X, \mu)$ *and* $\mathcal{K} = L^2(Y, \mathcal{B}_Y, \nu)$*. Then the following conditions on an operator* $T \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ *are equivalent:*

(i) $T \in \mathcal{L}^2(\mathcal{K}, \mathcal{H})$*;*

(ii) there exists $k \in L^2(X \times Y, \mathcal{B}_X \otimes \mathcal{B}_Y, \mu \times \nu)$ *such that*

$$(Tg)(x) \ = \ \int_Y k(x,y)g(y)\ d\nu(y)\ \nu - a.e.\ \forall\ g \in \mathcal{K}\ . \quad (4.3.10)$$

If these equivalent conditions are satisfied, then,

$$||T||_{\mathcal{L}^2(\mathcal{K},\mathcal{H})} \ = \ ||k||_{L^2(\mu\times\nu)}\ .$$

Proof : $(ii) \Rightarrow (i)$: Suppose $k \in L^2(\mu \times \nu)$; then, by Tonelli's theorem - see Proposition A.5.18(c) - we can find a set $A \in \mathcal{B}_X$ such that $\mu(A) = 0$ and such that $x \notin A \ \Rightarrow \ k^x\ (= k(x,\cdot)\ ) \in L^2(\nu)$, and further,

$$||k||^2_{L^2(\mu\times\nu)} \ = \ \int_{X-A} ||k^x||^2_{L^2(\nu)} d\mu(x)\ .$$

It follows from the Cauchy-Schwarz inequality that if $g \in L^2(\nu)$, then $k^x g \in L^1(\nu)\ \forall\ x \notin A$; hence equation 4.3.10 does indeed meaningfully define a function Tg on $X - A$, so that Tg is defined almost everywhere; another application of the Cauchy-Schwarz inequality shows that

$$\begin{aligned} ||Tg||^2_{L^2(\mu)} \ &= \ \int_X |\int_Y k(x,y)g(y)\ d\nu(y)|^2\ d\mu(x) \\ &= \ \int_{X-A} |\langle k^x, \overline{g}\rangle_{\mathcal{K}}|^2\ d\mu(x) \\ &\leq \ \int_{X-A} ||k^x||^2_{L^2(\nu)} ||g||^2_{L^2(\nu)}\ d\mu(x) \\ &= \ ||k||^2_{L^2(\mu\times\nu)} ||g||^2_{L^2(\nu)}\ , \end{aligned}$$

and we thus find that equation 4.3.10 indeed defines a bounded operator $T \in \mathcal{L}(\mathcal{K}, \mathcal{H})$.

Before proceeding further, note that if $g \in \mathcal{K}$ and $f \in \mathcal{H}$ are arbitrary, then, (by Fubini's theorem), we find that

$$\begin{aligned}
\langle Tg, f \rangle &= \int_X (Tg)(x)\overline{f(x)}d\mu(x) \\
&= \int_X \left(\int_Y k(x,y)g(y)d\nu(y) \right) \overline{f(x)}d\mu(x) \\
&= \langle k, f \otimes \overline{g} \rangle_{L^2(\mu\times\nu)} \ , \qquad (4.3.11)
\end{aligned}$$

where we have used the notation $(f \otimes \overline{g})$ to denote the function on $X \times Y$ defined by $(f \otimes \overline{g})(x,y) = f(x)\overline{g(y)}$.

Suppose now that $\{e_n : n \in N\}$ and $\{g_m : m \in M\}$ are orthonormal bases for $\mathcal{H}$ and $\mathcal{K}$ respectively; then, notice that also $\{\overline{g}_m : m \in M\}$ is an orthonormal basis for $\mathcal{K}$; deduce from equation 4.3.11 above and Exercise A.5.19 that

$$\begin{aligned}
\sum_{m\in M, n\in N} |\langle Tg_m, e_n\rangle_{\mathcal{H}}|^2 &= \sum_{m\in M, n\in N} |\langle k, e_n \otimes \overline{g}_m\rangle_{L^2(\mu\times\nu)}|^2 \\
&= ||k||^2_{L^2(\mu\times\nu)} \ ;
\end{aligned}$$

thus T is a Hilbert-Schmidt operator with Hilbert-Schmidt norm agreeing with the norm of k as an element of $L^2(\mu \times \nu)$.

$(i) \Rightarrow (ii)$: If $T : \mathcal{K} \to \mathcal{H}$ is a Hilbert-Schmidt operator, then, in particular - see Proposition 4.3.14(c) - T is compact; let

$$Tg = \sum_n \lambda_n \langle g, g_n\rangle f_n$$

be the singular value decomposition of T (see Proposition 4.3.10). Thus $\{g_n\}_n$ (resp., $\{f_n\}_n$) is an orthonormal sequence in $\mathcal{K}$ (resp., $\mathcal{H}$) and $\lambda_n = s_n(T)$. It follows from Proposition 4.3.14(d) that $\sum_n \lambda_n^2 < \infty$, and hence we find that the equation

$$k = \sum_n \lambda_n \ f_n \otimes \overline{g}_n$$

defines a unique element $k \in L^2(\mu\times\nu)$; if $\tilde{T}$ denotes the 'integral operator' associated to the 'kernel function' k as in equation 4.3.10, we find from equation 4.3.11 that for arbitrary $g \in \mathcal{K}, f \in \mathcal{H}$, we have

$$\langle \tilde{T}g, f\rangle_{\mathcal{H}} = \langle k, f \otimes \overline{g}\rangle_{L^2(\mu\times\nu)}$$

$$\begin{aligned}
&= \sum_n \lambda_n \langle f_n \otimes \overline{g}_n, f \otimes \overline{g} \rangle_{L^2(\mu \times \nu)} \\
&= \sum_n \lambda_n \langle f_n, f \rangle_{\mathcal{H}} \langle \overline{g}_n, \overline{g} \rangle_{\mathcal{K}} \\
&= \sum_n \lambda_n \langle f_n, f \rangle_{\mathcal{H}} \langle g, g_n \rangle_{\mathcal{K}} \\
&= \langle Tg, f \rangle_{\mathcal{H}} \ ,
\end{aligned}$$

whence we find that $T = \tilde{T}$ and so T is, indeed, the integral operator induced by the kernel function k. □

EXERCISE 4.3.16 *If T and k are related as in equation 4.3.10, we say that T is the integral operator induced by the kernel k, and we shall write $T = Int\ k$.*

For $i = 1, 2, 3$, let $\mathcal{H}_i = L^2(X_i, \mathcal{B}_i, \mu_i)$, where $(X_i, \mathcal{B}_i, \mu_i)$ is a σ-finite measure space.

(a) Let $h \in L^2(X_2 \times X_3, \mathcal{B}_2 \otimes \mathcal{B}_3, \mu_2 \times \mu_3), k, k_1 \in L^2(X_1 \times X_2, \mathcal{B}_1 \otimes \mathcal{B}_2, \mu_1 \times \mu_2)$, and let $S = Int\ h \ \in \mathcal{L}^2(\mathcal{H}_3, \mathcal{H}_2)$, $T = Int\ k, T_1 = Int\ k_1 \ \in \mathcal{L}^2(\mathcal{H}_2, \mathcal{H}_1)$; show that

(i) if $\alpha \in \mathbb{C}$, then $T + \alpha T_1 \ = \ Int\ (k + \alpha k_1)$;

(ii) if we define $k^*(x_2, x_1) = \overline{k(x_1, x_2)}$, then $k^* \in L^2(X_2 \times X_1, \mathcal{B}_2 \otimes \mathcal{B}_1, \mu_2 \times \mu_1)$ and $T^* = Int\ k^*$;

(iii) $TS \in \mathcal{L}^2(\mathcal{H}_3, \mathcal{H}_1)$ and $TS = Int\ (k * h)$, where

$$(k * h)(x_1, x_3) \ = \ \int_{X_2} k(x_1, x_2) h(x_2, x_3) \ d\mu_2(x_2)$$

for $(\mu_1 \times \mu_3)$-almost all $(x_1, x_3) \in X \times X$.
(Hint: for (ii), note that k^* is a square-integrable kernel, and use equation 4.3.11 to show that $Int\ k^* \ = \ (Int\ k)^*$; for (iii), note that $|(k * h)(x_1, x_3)| \leq ||k^{x_1}||_{L^2(\mu_2)} ||h_{x_3}||_{L^2(\mu_2)}$ to conclude that $k * h \in L^2(\mu_1 \times \mu_3)$; again use Fubini's theorem to justify interchanging the orders of integration in the verification that $Int(k * h) = (Int\ k)(Int\ h)$.)

4.4 Fredholm operators and index

We begin with a fact which is a simple consequence of results in the last section.

LEMMA 4.4.1 *Suppose $\mathcal{I} \subset \mathcal{L}(\mathcal{H})$ is an ideal in the C^*-algebra $\mathcal{L}(\mathcal{H})$.*

(a) If $\mathcal{I} \neq \{0\}$, then $\mathcal{I}$ contains every 'finite-rank operator' - i.e., if $K \in \mathcal{L}(\mathcal{H})$ and if ran K is finite-dimensional, then $K \in \mathcal{I}$.

(b) If $\mathcal{H}$ is separable and if $\mathcal{I}$ is not contained in $\mathcal{K}(\mathcal{H})$, then $\mathcal{I} = \mathcal{L}(\mathcal{H})$.

Proof : (a) If K is as in (a), then (by the singular value decomposition) there exists an orthonormal basis $\{y_n : 1 \leq n \leq d\}$ for $ker^\perp K$ and an orthonormal basis $\{z_n : 1 \leq n \leq d\}$ for $ranT$ such that $Kv = \sum_{n=1}^d s_n(K)\langle v, y_n\rangle z_n \;\forall\; v \in \mathcal{H}$.

If $\mathcal{I} \neq \{0\}$, we can find $0 \neq T \in \mathcal{I}$, and a unit vector $x \in \mathcal{H}$ such that $Tx \neq 0$; define $A_n, B_n \in \mathcal{L}(\mathcal{H}),\ 1 \leq n \leq d$ by

$$B_n v = \langle v, y_n\rangle x \ , A_n v = \frac{1}{||Tx||^2}\langle v, Tx\rangle z_n$$

and note that

$$A_n T B_n v = \langle v, y_n\rangle z_n \;\forall\; v \in \mathcal{H} \ ;$$

in particular, $K = \sum_{n=1}^d s_n(K) A_n T B_n \ \in \mathcal{I}$.

(b) Suppose there exists a non-compact operator $T \in \mathcal{I}$; then, by Proposition 4.3.6(e), there exists an infinite-dimensional closed subspace $\mathcal{N} \subset \ ran\,T$. If $\mathcal{M} = T^{-1}(\mathcal{N}) \cap ker^\perp T$, it follows from the open mapping theorem that there exists a bounded operator $S_0 \in \mathcal{L}(\mathcal{N}, \mathcal{M})$ such that $S_0 T|_{\mathcal{M}} = id_{\mathcal{M}}$; hence if we write $S = S_0 P_{\mathcal{N}}$, then $S \in \mathcal{L}(\mathcal{H})$ and $STP_{\mathcal{M}} = P_{\mathcal{M}}$; thus, $P_{\mathcal{M}} \in \mathcal{I}$; since $\mathcal{M}$ is an infinite-dimensional subspace of the separable Hilbert space $\mathcal{H}$, we can find an isometry $U \in \mathcal{L}(\mathcal{H})$ such that $UU^* = P_{\mathcal{M}}$; but then, $id_{\mathcal{H}} = UP_{\mathcal{M}}U^* \in \mathcal{I}$, and hence $\mathcal{I} = \mathcal{L}(\mathcal{H})$. □

COROLLARY 4.4.2 *If $\mathcal{H}$ is an infinite-dimensional, separable Hilbert space, then $\mathcal{K}(\mathcal{H})$ is the unique non-trivial (norm-)closed ideal in $\mathcal{L}(\mathcal{H})$.*

Proof : Suppose $\mathcal{I}$ is a closed non-trivial ideal in $\mathcal{L}(\mathcal{H})$; since the set of finite-rank operators is dense in $\mathcal{K}(\mathcal{H})$ (by Proposition 4.3.6(d)), it follows from Lemma 4.4.1(a) and the assumption that $\mathcal{I}$ is a non-zero closed ideal, that $\mathcal{I} \supset \mathcal{K}(\mathcal{H})$; the assumption that $\mathcal{I} \neq \mathcal{L}(\mathcal{H})$ and Lemma 4.4.1(b) ensure that $\mathcal{I} \subset \mathcal{K}(\mathcal{H})$. □

REMARK 4.4.3 Suppose $\mathcal{H}$ is a non-separable Hilbert space; let $\{e_i : i \in I\}$ denote any one fixed orthonormal basis for $\mathcal{H}$. There is a natural equivalence relation on the collection 2^I of subsets of I given by $I_1 \sim I_2$ if and only if $|I_1| = |I_2|$ (i.e., there exists a bijective map $f : I_1 \to I_2$); we shall think of every equivalence class as determining a 'cardinal number α such that $\alpha \leq \; dim\, \mathcal{H}$'. Let us write J for the set of 'infinite cardinal numbers' so obtained; thus,

$$\begin{aligned} J \quad &= \quad \{\{I_1 \subset I : I_1 \sim I_0\} : I_0 \subset I, \; I_0 \; is \; infinite\} \\ \text{``}= \quad &\{\aleph_0, \cdots, \alpha, \cdots, dim\mathcal{H}\} \; ''. \end{aligned}$$

Thus we say $\alpha \in J$ if there exists an infinite subset $I_0 \subset I$ (which is uniquely determined up to $\sim$) such that α "$= |I_0|$".

For each $\alpha \in J$, consider the collection $\mathcal{K}_\alpha(\mathcal{H})$ of operators on $\mathcal{H}$ with the property that if $\mathcal{N}$ is any closed subspace contained in *ran* T, then *dim* $\mathcal{N} \; < \alpha$. It is a fact that $\{\mathcal{K}_\alpha(\mathcal{H}) : \alpha \in J\}$ is precisely the collection of all closed ideals of $\mathcal{L}(\mathcal{H})$.

Most discussions of non-separable Hilbert spaces degenerate, as in this remark, to an exercise in transfinite considerations invovling infinite cardinals; almost all 'operator theory' is, in a sense, contained in the separable case; and this is why we can safely restrict ourselves to separable Hilbert spaces in general.

□

PROPOSITION 4.4.4 *(**Atkinson's theorem**) If $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, then the following conditions are equivalent:*

(a) there exist operators $S_1, S_2 \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_1)$ and compact operators $K_i \in \mathcal{L}(\mathcal{H}_i), i = 1, 2,$ such that

$$S_1 T = 1_{\mathcal{H}_1} + K_1 \;\; and \;\; TS_2 = 1_{\mathcal{H}_2} + K_2 \; .$$

(b) T satisfies the following conditions:

(i) ran T is closed; and

(ii) ker T and ker T^* are both finite-dimensional.

(c) There exists $S \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_1)$ and finite-dimensional subspaces $\mathcal{N}_i \subset \mathcal{H}_i, i = 1, 2,$ such that

$$ST = 1_{\mathcal{H}_1} - P_{\mathcal{N}_1} \;\; and \;\; TS = 1_{\mathcal{H}_2} - P_{\mathcal{N}_2} \; .$$

Proof : $(a) \Rightarrow (b)$: Begin by fixing a finite-rank operator F such that $||K_1 - F|| < \frac{1}{2}$ (see Proposition 4.3.6(d)); set $\mathcal{M} = ker\ F$ and note that if $x \in \mathcal{M}$, then

$$||S_1|| \cdot ||Tx|| \geq ||S_1Tx|| = ||x + K_1x|| = ||x + (K_1 - F)x|| \geq \frac{1}{2}||x||,$$

which shows that T is bounded below on $\mathcal{M}$; it follows that $T(\mathcal{M})$ is a closed subspace of $\mathcal{H}_2$; note, however, that $\mathcal{M}^\perp$ is finite-dimensional (since F maps this space injectively onto its finite-dimensional range); hence T satisfies condition (i) thanks to Exercise A.6.5(3) and the obvious identity $ran\ T = T(\mathcal{M}) + T(\mathcal{M}^\perp)$.

As for (ii), since $S_1T = 1_{\mathcal{H}_1} + K_1$, note that $K_1x = -x$ for all $x \in ker\ T$; this means that $ker\ T$ is a closed subspace which is contained in $ran\ K_1$ and the compactness of K_1 now demands the finite-dimensionality of $ker\ T$. Similarly, $ker\ T^* \subset ran\ K_2^*$ and condition (ii) is verified.

$(b) \Rightarrow (c)$: Let $\mathcal{N}_1 = ker\ T$, $\mathcal{N}_2 = ker\ T^*$ $(= ran^\perp T)$; thus T maps $\mathcal{N}_1^\perp$ 1-1 onto $ran\ T$; the condition (b) and the open mapping theorem imply the existence of a bounded operator $S_0 \in \mathcal{L}(\mathcal{N}_2^\perp, \mathcal{N}_1^\perp)$ such that S_0 is the inverse of the restricted operator $T|_{\mathcal{N}_1^\perp}$; if we set $S = S_0P_{\mathcal{N}_2^\perp}$, then $S \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_1)$ and by definition, we have $ST = 1_{\mathcal{H}_1} - P_{\mathcal{N}_1}$ and $TS = 1_{\mathcal{H}_2} - P_{\mathcal{N}_2}$; by condition (ii), both subspaces $\mathcal{N}_i$ are finite-dimensional.

$(c) \Rightarrow (a)$: Obvious. □

REMARK 4.4.5 (1) An operator which satisfies the equivalent conditions of Atkinson's theorem is called a **Fredholm operator**, and the collection of Fredholm operators from $\mathcal{H}_1$ to $\mathcal{H}_2$ is denoted by $\mathcal{F}(\mathcal{H}_1, \mathcal{H}_2)$, and as usual, we shall write $\mathcal{F}(\mathcal{H}) = \mathcal{F}(\mathcal{H}, \mathcal{H})$. It must be observed - as a consequence of Atkinson's theorem, for instance - that a necessary and sufficient condition for $\mathcal{F}(\mathcal{H}_1, \mathcal{H}_2)$ to be non-empty is that either (i) $\mathcal{H}_1$ and $\mathcal{H}_2$ are both finite-dimensional, in which case $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2) = \mathcal{F}(\mathcal{H}_1, \mathcal{H}_2)$, or (ii) neither $\mathcal{H}_1$ nor $\mathcal{H}_2$ is finite-dimensional, and $dim\ \mathcal{H}_1 = dim\ \mathcal{H}_2$.

(2) Suppose $\mathcal{H}$ is a separable infinite-dimensional Hilbert space. Then the quotient $\mathcal{Q}(\mathcal{H}) = \mathcal{L}(\mathcal{H})/\mathcal{K}(\mathcal{H})$ (of $\mathcal{L}(\mathcal{H})$ by the ideal $\mathcal{K}(\mathcal{H})$) is a Banach algebra, which is called the **Calkin**

algebra. If we write $\pi_{\mathcal{K}} : \mathcal{L}(\mathcal{H}) \to \mathcal{Q}(\mathcal{H})$ for the quotient mapping, then we find that an operator $T \in \mathcal{L}(\mathcal{H})$ is a Fredholm operator precisely when $\pi_{\mathcal{K}}(T)$ is invertible in the Calkin algebra; thus, $\mathcal{F}(\mathcal{H}) = \pi_{\mathcal{K}}^{-1}(\mathcal{G}(\mathcal{Q}(\mathcal{H})))$. (It is a fact, which we shall not need and consequently go into here, that the Calkin algebra is in fact a C^*-algebra - as is the quotient of any C^*-algebra by a norm-closed *-ideal.)

(3) It is customary to use the adjective 'essential' to describe a property of an operator $T \in \mathcal{L}(\mathcal{H})$ which is actually a property of the corresponding element $\pi_{\mathcal{K}}(T)$ of the Calkin algebra, thus, for instance, the **essential spectrum** of T is defined to be

$$\sigma_{ess}(T) \ = \ \sigma_{\mathcal{Q}(\mathcal{H})}(\pi_{\mathcal{K}}(T)) \ = \ \{\lambda \in \mathbb{C} : \ (T - \lambda) \notin \mathcal{F}(\mathcal{H})\} \ . \tag{4.4.12}$$

□

The next exercise is devoted to understanding the notions of Fredholm operator and essential spectrum at least in the case of normal operators.

EXERCISE 4.4.6 *(1) Let $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ have polar decomposition $T = U|T|$. Then show that*

(a) $T \in \mathcal{F}(\mathcal{H}) \ \Leftrightarrow \ U \in \mathcal{F}(\mathcal{H}_1, \mathcal{H}_2)$ and $|T| \in \mathcal{F}(\mathcal{H}_1)$.

(b) A partial isometry is a Fredholm operator if and only if both its initial and final spaces have finite co-dimension (i.e., have finite-dimensional orthogonal complements).
(Hint: for both parts, use the characterisation of a Fredholm operator which is given by Proposition 4.4.4(b).)

(2) If $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}$, consider the following conditions on an operator $T \in \mathcal{L}(\mathcal{H})$:

(i) T is normal;

(ii) U and $|T|$ commute.

Show that (i) $\Rightarrow$ (ii), and find an example to show that the reverse implication is not valid in general.
(Hint: if T is normal, then note that

$$|T|^2 U = T^*TU = TT^*U = U|T|^2U^*U = U|T|^2 \ ;$$

thus U commutes with $|T|^2$; deduce that in the decomposition $\mathcal{H} = kerT \oplus ker^{\perp}T$, we have $U = 0 \oplus U_0$, $|T| = 0 \oplus A$, where

U_0 (resp., A) is a unitary (resp., positive injective) operator of $\ker^\perp T$ onto (resp., into) itself; and infer that U_0 and A^2 commute; since U_0 is unitary, deduce from the uniqueness of positive square roots that U_0 commutes with A, and finally that U and $|T|$ commute; for the 'reverse implication', let T denote the unilateral shift, and note that $U = T$ and $|T| = 1$.)

(3) Suppose $T = U|T|$ is a normal operator as in (2) above. Then show that the following conditions on T are equivalent:

(i) T is a Fredholm operator;

(ii) there exists an orthogonal direct-sum decomposition $\mathcal{H} = \mathcal{M} \oplus \mathcal{N}$, where dim $\mathcal{N} < \infty$, with respect to which T has the form $T = T_1 \oplus 0$, where T_1 is an invertible normal operator on $\mathcal{M}$;

(iii) there exists an $\epsilon > 0$ such that $1_{\mathbb{D}_\epsilon}(T) = 1_{\{0\}}(T) = P_0$, where (a) $E \mapsto 1_E(T)$ denotes the measurable functional calculus for T, (b) $\mathbb{D}_\epsilon = \{z \in \mathbb{C} : |z| < \epsilon\}$ is the ϵ-disc around the origin, and (c) P_0 is some finite-rank projection.

(Hint: For (i) $\Rightarrow$ (ii), note, as in the hint for exercise (2) above, that we have decompositions $U = U_0 \oplus 0$, $|T| = A \oplus 0$ - with respect to $\mathcal{H} = \mathcal{M} \oplus \mathcal{N}$, where $\mathcal{M} = \ker^\perp T$ and $\mathcal{N} = \ker T$ (is finite-dimensional under the assumption (i))- where U_0 is unitary, A is positive and 1-1, and U_0 and A commute; deduce from the Fredholm condition that $\mathcal{N}$ is finite-dimensional and that A is invertible; conclude that in this decomposition, $T = U_0 A \oplus 0$ and $U_0 A$ is normal and invertible. For (ii) $\Rightarrow$ (iii), if $T = T_1 \oplus 0$ has polar decomposition $T = U|T|$, then $|T| = |T_1| \oplus 0$ and $U = U_0 \oplus 0$ with U_0 unitary and $|T_1|$ positive and invertible; then if $\epsilon > 0$ is such that T_1 is bounded below by ϵ, then argue that $1_{\mathbb{D}_\epsilon}(T) = 1_{[0,\epsilon)}(|T|) = 1_{\{0\}}(|T|) = 1_{\{0\}}(T) = P_{\mathcal{N}}$.)

(4) Let $T \in \mathcal{L}(\mathcal{H})$ be normal; prove that the following conditions on a complex number λ are equivalent:

(i) $\lambda \in \sigma_{ess}(T)$;

(ii) there exists an orthogonal direct-sum decomposition $\mathcal{H} = \mathcal{M} \oplus \mathcal{N}$, where dim $\mathcal{N} < \infty$, with respect to which T has the form $T = T_1 \oplus \lambda$, where $(T_1 - \lambda)$ is an invertible normal operator on $\mathcal{M}$;

(iii) there exists $\epsilon > 0$ such that $1_{\mathbb{D}_\epsilon + \lambda}(T) = 1_{\{\lambda\}}(T) = P_\lambda$,

where $\mathbb{D}_\epsilon + \lambda$ *denotes the* ϵ*-disc around the point* λ*, and* P_λ *is some finite-rank projection.*
(Hint: apply (3) above to $T - \lambda$*.)*

We now come to an important definition.

DEFINITION 4.4.7 *If* $T \in \mathcal{F}(\mathcal{H}_1, \mathcal{H}_2)$ *is a Fredholm operator, its* **index** *is the integer defined by*

$$ind\ T \ = \ dim(ker\ T) - dim(ker\ T^*).$$

Several elementary consequences of the definition are discussed in the following remark.

REMARK 4.4.8 (1) The index of a normal Fredholm operator is always 0. (Reason: If $T \in \mathcal{L}(\mathcal{H})$ is a normal operator, then $|T|^2 = |T^*|^2$, and the uniqueness of the square root implies that $|T| = |T^*|$; it follows that $ker\ T = ker\ |T| = ker\ T^*$.)

(2) It should be clear from the definitions that if $T = U|T|$ is the polar decomposition of a Fredholm operator, then $ind\ T = ind\ U$.

(3) If $\mathcal{H}_1$ and $\mathcal{H}_2$ are finite-dimensional, then $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2) = \mathcal{F}(\mathcal{H}_1, \mathcal{H}_2)$ and $ind\ T \ = \ dim\ \mathcal{H}_1 - dim\ \mathcal{H}_2\ \forall\ T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$; in particular, the index is independent of the operator in this case. (Reason: let us write $\rho = dim(ran\ T)$ (resp., $\rho^* = dim(ran\ T^*)$) and $\nu = dim(ker\ T)$ (resp., $\nu^* = dim(ker\ T^*)$) for the rank and nullity of T (resp., T^*); on the one hand, deduce from Exercise 4.2.6(3) that if $dim\ \mathcal{H}_i = n_i$, then $\rho = n_1 - \nu$ and $\rho^* = n_2 - \nu^*$; on the other hand, by Exercise 4.2.6(2), we find that $\rho = \rho^*$; hence,

$$ind\ T = \nu - \nu^* = (n_1 - \rho) - (n_2 - \rho) = n_1 - n_2\ .)$$

(4) If $S = UTV$, where $S \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_4), U \in \mathcal{L}(\mathcal{H}_3, \mathcal{H}_4), T \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_3), V \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, and if U and V are invertible (i.e., are 1-1 and onto), then S is a Fredholm operator if and only if T is, in which case, $ind\ S = ind\ T$. (This should be clear from Atkinson's theorem and the definition of the index.)

(5) Suppose $\mathcal{H}_i = \mathcal{N}_i \oplus \mathcal{M}_i$ and $dim\ \mathcal{N}_i < \infty$, for $i = 1, 2,$; suppose $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ is such that T maps $\mathcal{N}_1$ into $\mathcal{N}_2$, and such that T maps $\mathcal{M}_1$ 1-1 onto $\mathcal{M}_2$. Thus, with respect to these decompositions, T has the matrix decomposition

$$T = \begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix} ,$$

where D is invertible; then it follows from Atkinson's theorem that T is a Fredholm operator, and the assumed invertibility of D implies that $ind\ T = ind\ A = dim\ \mathcal{N}_1 - dim\ \mathcal{N}_2$ - see (3) above. □

Lemma 4.4.9 *Suppose $\mathcal{H}_i = \mathcal{N}_i \oplus \mathcal{M}_i$, for $i = 1, 2$; suppose $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ has the associated matrix decomposition*

$$T = \begin{bmatrix} A & B \\ C & D \end{bmatrix} ,$$

where $A \in \mathcal{L}(\mathcal{N}_1, \mathcal{N}_2), B \in \mathcal{L}(\mathcal{M}_1, \mathcal{N}_2), C \in \mathcal{L}(\mathcal{N}_1, \mathcal{M}_2)$, and $D \in \mathcal{L}(\mathcal{M}_1, \mathcal{M}_2)$; assume that D is invertible - i.e, D maps $\mathcal{M}_1$ 1-1 onto $\mathcal{M}_2$. Then

$$T \in \mathcal{F}(\mathcal{H}_1, \mathcal{H}_2) \Leftrightarrow (A - BD^{-1}C) \in \mathcal{F}(\mathcal{N}_1, \mathcal{N}_2) ,$$

and $ind\ T = ind\ (A - BD^{-1}C)$; further, if it is the case that $dim\ \mathcal{N}_i < \infty,\ i = 1, 2$, then T is necessarily a Fredholm operator and $ind\ T = dim \mathcal{N}_1 - dim \mathcal{N}_2$.

Proof : Let $U \in \mathcal{L}(\mathcal{H}_2)$ (resp., $V \in \mathcal{L}(\mathcal{H}_1)$) be the operator which has the matrix decomposition

$$U = \begin{bmatrix} 1_{\mathcal{N}_2} & -BD^{-1} \\ 0 & 1_{\mathcal{M}_2} \end{bmatrix} , \text{ (resp., } V = \begin{bmatrix} 1_{\mathcal{N}_1} & 0 \\ -D^{-1}C & 1_{\mathcal{M}_2} \end{bmatrix})$$

with respect to $\mathcal{H}_2 = \mathcal{N}_2 \oplus \mathcal{M}_2$ (resp., $\mathcal{H}_1 = \mathcal{N}_1 \oplus \mathcal{M}_1$).

Note that U and V are invertible operators, and that

$$UTV = \begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix} ;$$

since D is invertible, we see that $ker(UTV) = ker(A - BD^{-1}C)$ and that $ker(UTV)^* = ker(A - BD^{-1}C)^*$; also, it should be clear that UTV has closed range if and only if $(A - BD^{-1}C)$ has closed range; we thus see that T is a Fredholm operator precisely when $(A - BD^{-1}C)$ is Fredholm, and that $ind\ T = ind(A - BD^{-1}C)$ in that case. For the final assertion of the lemma (concerning finite-dimensional $\mathcal{N}_i$'s), appeal now to Remark 4.4.8(5). $\square$

We now state some simple facts in an exercise, before proceeding to establish the main facts concerning the index of Fredholm operators.

EXERCISE 4.4.10 *(1) Suppose* $D_0 \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ *is an invertible operator; show that there exists* $\epsilon > 0$ *such that if* $D \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ *satisfies* $||D - D_0|| < \epsilon$, *then* D *is invertible. (Hint: let* $D_0 = U_0|D_0|$ *be the polar decomposition; write* $D = U_0(U_0^*D)$, *note that* $||D - D_0|| = ||(U_0^*D - |D_0|)||$, *and that* D *is invertible if and only if* U_0^*D *is invertible, and use the fact that the set of invertible elements in the Banach algebra* $\mathcal{L}(\mathcal{H}_1)$ *form an open set.)*

(2) Show that a function $\phi : [0,1] \to \mathbb{Z}$ *which is locally constant, is necessarily constant.*

(3) Suppose $\mathcal{H}_i = \mathcal{N}_i \oplus \mathcal{M}_i, i = 1, 2$, *are orthogonal direct sum decompositions of Hilbert spaces.*

(a) Suppose $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ *is represented by the operator matrix*

$$T = \begin{bmatrix} A & 0 \\ C & D \end{bmatrix},$$

where A *and* D *are invertible operators; show, then, that* T *is also invertible and that* T^{-1} *is represented by the operator matrix*

$$T^{-1} = \begin{bmatrix} A^{-1} & 0 \\ -D^{-1}CA^{-1} & D^{-1} \end{bmatrix}.$$

(b) Suppose $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ *is represented by the operator matrix*

$$T = \begin{bmatrix} 0 & B \\ C & 0 \end{bmatrix},$$

where B is an invertible operator; show that $T \in \mathcal{F}(\mathcal{H}_1, \mathcal{H}_2)$ if and only if $C \in \mathcal{F}(\mathcal{N}_1, \mathcal{M}_2)$, and that if this happens, then ind $T =$ ind C.

THEOREM 4.4.11 *(a) $\mathcal{F}(\mathcal{H}_1, \mathcal{H}_2)$ is an open set in $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ and the function ind : $\mathcal{F}(\mathcal{H}_1, \mathcal{H}_2) \to \mathbb{C}$ is 'locally constant'; i.e., if $T_0 \in \mathcal{F}(\mathcal{H}_1, \mathcal{H}_2)$, then there exists $\delta > 0$ such that whenever $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ satisfies $||T - T_0|| < \delta$, it is then the case that $T \in \mathcal{F}(\mathcal{H}_1, \mathcal{H}_2)$ and ind $T =$ ind T_0.*

(b) $T \in \mathcal{F}(\mathcal{H}_1, \mathcal{H}_2), K \in \mathcal{K}(\mathcal{H}_1, \mathcal{H}_2) \Rightarrow (T + K) \in \mathcal{F}(\mathcal{H}_1, \mathcal{H}_2)$ and ind$(T + K)$ = ind T.

(c) $S \in \mathcal{F}(\mathcal{H}_2, \mathcal{H}_3), T \in \mathcal{F}(\mathcal{H}_1, \mathcal{H}_2) \Rightarrow ST \in \mathcal{F}(\mathcal{H}_1, \mathcal{H}_3)$ and ind$(ST) =$ ind $S +$ ind T.

Proof : (a) Suppose $T_0 \in \mathcal{F}(\mathcal{H}_1, \mathcal{H}_2)$. Set $\mathcal{N}_1 = ker\ T_0$ and $\mathcal{N}_2 = ker\ T_0^*$, so that $\mathcal{N}_i, i = 1, 2$, are finite-dimensional spaces and we have the orthogonal decompositions $\mathcal{H}_i = \mathcal{N}_i \oplus \mathcal{M}_i, i = 1, 2$, where $\mathcal{M}_1 = ran\ T_0^*$ and $\mathcal{M}_2 = ran\ T_0$. With respect to these decompositions of $\mathcal{H}_1$ and $\mathcal{H}_2$, it is clear that the matrix of T_0 has the form

$$T_0 = \left[\begin{array}{cc} 0 & 0 \\ 0 & D_0 \end{array} \right] ,$$

where the operator $D_0 : \mathcal{M}_1 \to \mathcal{M}_2$ is (a bounded bijection, and hence) invertible.

Since D_0 is invertible, it follows - see Exercise 4.4.10(1) - that there exists a $\delta > 0$ such that $D \in \mathcal{L}(\mathcal{M}_1, \mathcal{M}_2)$, $||D - D_0|| < \delta \ \Rightarrow\ D$ is invertible. Suppose now that $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ and $||T - T_0|| < \delta$; let

$$T = \left[\begin{array}{cc} A & B \\ C & D \end{array} \right]$$

be the matrix decomposition associated to T; then note that $||D - D_0|| < \delta$ and consequently D is an invertible operator. Conclude from Lemma 4.4.9 that T is a Fredholm operator and that

$$ind\ T \ = \ ind(A - BD^{-1}C) \ = \ dim\mathcal{N}_1 - dim\mathcal{N}_2 \ = \ ind\ T_0 \ .$$

(b) If T is a Fredholm operator and K is compact, as in (b), define $T_t = T + tK$, for $0 \leq t \leq 1$. It follows from Proposition 4.4.4 that each T_t is a Fredholm operator; further, it is a consequence of (a) above that the function $[0,1] \ni t \mapsto ind\ T_t$ is a locally constant function on the interval $[0,1]$; the desired conlcusion follows easily - see Exercise 4.4.10(2).

(c) Let us write $\mathcal{K}_1 = \mathcal{H}_1 \oplus \mathcal{H}_2$ and $\mathcal{K}_2 = \mathcal{H}_2 \oplus \mathcal{H}_3$, and consider the operators $U \in \mathcal{L}(\mathcal{K}_2), R \in \mathcal{L}(\mathcal{K}_1, \mathcal{K}_2)$ and $V \in \mathcal{L}(\mathcal{K}_1)$ defined, by their matrices with respect to the afore-mentioned direct-sum decompositions of these spaces, as follows:

$$U = \begin{bmatrix} 1_{\mathcal{H}_2} & 0 \\ -\epsilon^{-1} S & 1_{\mathcal{H}_3} \end{bmatrix}, \quad R = \begin{bmatrix} T & \epsilon 1_{\mathcal{H}_2} \\ 0 & S \end{bmatrix},$$

$$V = \begin{bmatrix} -\epsilon 1_{\mathcal{H}_1} & 0 \\ T & \epsilon^{-1} 1_{\mathcal{H}_2} \end{bmatrix},$$

where we first choose $\epsilon > 0$ to be so small as to ensure that R is a Fredholm operator with index equal to $ind\ T + ind\ S$; this is possible by (a) above, since the operator R_0, which is defined by modifying the definition of R so that the 'off-diagonal' terms are zero and the diagonal terms are unaffected, is clearly a Fredholm operator with index equal to the sum of the indices of S and T.

It is easy to see that U and V are invertible operators - see Exercise 4.4.10(3)(a) - and that the matrix decomposition of the product $URV \in \mathcal{F}(\mathcal{K}_1, \mathcal{K}_2)$ is given by:

$$URV = \begin{bmatrix} 0 & 1_{\mathcal{H}_2} \\ ST & 0 \end{bmatrix},$$

which is seen - see Exercise 4.4.10(3)(b) - to imply that $ST \in \mathcal{F}(\mathcal{H}_1, \mathcal{H}_3)$ and that $ind(ST) = ind\ R = ind\ S + ind\ T$, as desired. $\square$

EXAMPLE 4.4.12 Fix a separable infinite-dimensional Hilbert space $\mathcal{H}$; for definiteness' sake, we assume that $\mathcal{H} = \ell^2$. Let $S \in \mathcal{L}(\mathcal{H})$ denote the unilateral shift - see Example 2.4.15(1). Then, S is a Fredholm operator with $ind\ S = -1$, and $ind\ S^* = 1$; hence Theorem 4.4.11((c) implies that if $n \in \mathbb{N}$, then $S^n \in \mathcal{F}(\mathcal{H})$ and $ind(S^n) = -n$ and $ind(S^*)^n = n$; in particular, there exist operators with all possible indices.

Let us write $\mathcal{F}_n = \{T \in \mathcal{F}(\mathcal{H}) : ind\ T = n\}$, for each $n \in \mathbb{Z}$.

First consider the case $n = 0$. Suppose $T \in \mathcal{F}_0$; then it is possible to find a partial isometry U_0 with initial space equal to *ker* T and final space equal to *ker* T^*); then define $T_t = T + tU_0$. Observe that $t \neq 0 \Rightarrow T_t$ is invertible; and hence, the map $[0,1] \ni t \mapsto T_t \in \mathcal{L}(\mathcal{H})$ (which is clearly norm-continuous) is seen to define a path - see Exercise 4.4.13(1) - which is contained in $\mathcal{F}_0$ and connects T_0 to an invertible operator; on the other hand, the set of invertible operators is a path-connected subset of $\mathcal{F}_0$; it follows that $\mathcal{F}_0$ is path-connected.

Next consider the case $n > 0$. Suppose $T \in \mathcal{F}_n, n < 0$. Then note that $T(S^*)^n \in \mathcal{F}_0$ (by Theorem 4.4.11(c)) and since $(S^*)^n S^n = 1$, we find that $T = T(S^*)^n S^n \in \mathcal{F}_0 S^n$; conversely since Theorem 4.4.11(c) implies that $\mathcal{F}_0 S^n \subset \mathcal{F}_n$, we thus find that $\mathcal{F}_n = \mathcal{F}_0 S^n$.

For $n > 0$, we find, by taking adjoints, that $\mathcal{F}_n = \mathcal{F}_{-n}^* = (S^*)^n \mathcal{F}_0$.

We conclude that for all $n \in \mathbb{Z}$, the set $\mathcal{F}_n$ is path-connected; on the other hand, since the index is 'locally constant', we can conclude that $\{\mathcal{F}_n : n \in \mathbb{Z}\}$ is precisely the collection of 'path-components' (= maximal path-connected subsets) of $\mathcal{F}(\mathcal{H})$. $\square$

EXERCISE 4.4.13 *(1) A* **path** *in a topological space X is a continuous function $f : [0,1] \to X$; if $f(0) = x, f(1) = y$, then f is called a path joining (or connecting) x to y. Define a relation $\sim$ on X by stipulating that $x \sim y$ if and only if there exists a path joining x to y.*

Show that $\sim$ is an equivalence relation on X.

The equivalence classes associated to the relation $\sim$ are called the **path-components** *of X; the space X is said to be* **path-connected** *if X is itself a path component.*

(2) Let $\mathcal{H}$ be a separable Hilbert space. In this exercise, we regard $\mathcal{L}(\mathcal{H})$ as being topologised by the operator norm.

(a) Show that the set $\mathcal{L}_{sa}(\mathcal{H})$ of self-adjoint operators on $\mathcal{H}$ is path-connected. (Hint: Consider $t \mapsto tT$.)

(b) Show that the set $\mathcal{L}_+(\mathcal{H})$ of positive operators on $\mathcal{H}$ is path-connected. (Hint: Note that if $T \geq 0, t \in [0,1]$, then $tT \geq 0$.)

(c) Show that the set $GL_+(\mathcal{H})$ of invertible positive operators on $\mathcal{H}$ form a connected set. (Hint: If $T \in GL_+(\mathcal{H})$, use straight

line segments to first connect T to $||T|| \cdot 1$, and then $||T|| \cdot 1$ to 1.)

(d) Show that the set $\mathcal{U}(\mathcal{H})$ of unitary operators on $\mathcal{H}$ is path-connected. (Hint: If $U \in \mathcal{U}(\mathcal{H})$, find a self-adjoint A such that $U = e^{iA}$ - see Corollary 4.1.6(a) - and look at $U_t = e^{itA}$.)

We would like to conclude this section with the so-called 'spectral theorem for a general compact operator'. As a preamble, we start with an exercise which is devoted to 'algebraic (possibly non-orthogonal) direct sums' and associated non-self-adjoint projections.

EXERCISE 4.4.14 *(1) Let $\mathcal{H}$ be a Hilbert space, and let $\mathcal{M}$ and $\mathcal{N}$ denote closed subspaces of $\mathcal{H}$. Show that the following conditions are equivalent:*

(a) $\mathcal{H} = \mathcal{M} + \mathcal{N}$ and $\mathcal{M} \cap \mathcal{N} = \{0\}$;

(b) every vector $z \in \mathcal{H}$ is uniquely expressible in the form $z = x + y$ with $x \in \mathcal{M}, y \in \mathcal{N}$.

(2) If the equivalent conditions of (1) above are satisfied, show that there exists a unique $E \in \mathcal{L}(\mathcal{H})$ such that $Ez = x$, whenever z and x are as in (b) above. (Hint: note that $z = Ez + (z - Ez)$ and use the closed graph theorem to establish the boundedness of E.)

(3) If E is as in (2) above, then show that

(a) $E = E^2$;

(b) the following conditions on a vector $x \in \mathcal{H}$ are equivalent:

(i) $x \in$ ran E;

(ii) $Ex = x$.

(c) ker $E = \mathcal{N}$.

The operator E is said to be the 'projection on $\mathcal{M}$ along $\mathcal{N}$'.

(4) Show that the following conditions on an operator $E \in \mathcal{L}(\mathcal{H})$ are equivalent:

(i) $E = E^2$;

(ii) there exists a closed subspace $\mathcal{M} \subset \mathcal{H}$ such that E has the following operator-matrix with respect to the decomposition $\mathcal{H} = \mathcal{M} \oplus \mathcal{M}^\perp$:

$$E = \left[\begin{array}{cc} 1_{\mathcal{M}} & B \\ 0 & 0 \end{array} \right] ;$$

(iii) there exists a closed subspace $\mathcal{N} \subset \mathcal{H}$ such that E has the following operator-matrix with respect to the decomposition $\mathcal{H} = \mathcal{N}^\perp \oplus \mathcal{N}$:

$$E = \begin{bmatrix} 1_{\mathcal{N}^\perp} & 0 \\ C & 0 \end{bmatrix} ;$$

(iv) there exists closed subspaces $\mathcal{M}, \mathcal{N}$ satisfying the equivalent conditions of (1) such that E is the projection of $\mathcal{M}$ along $\mathcal{N}$.
(Hint: $(i) \Rightarrow (ii) :\ \mathcal{M} = ran\ E\ (=\ ker(1-E))$ is a closed subspace and $Ex = x \ \forall\ x \in \mathcal{M}$; since $\mathcal{M} = ran\ E$, (ii) follows. The implication $(ii) \Rightarrow (i)$ is verified by easy matrix-multiplication. Finally, if we let $(i)^*$ (resp., $(ii)^*$) denote the condition obtained by replacing E by E^* in condition (i) (resp., (ii)), then $(i) \Leftrightarrow (i)^* \Leftrightarrow (ii)^*$; take adjoints to find that $(ii)^* \Leftrightarrow (iii)$. The implication $(i) \Leftrightarrow (iv)$ is clear.)

(5) Show that the following conditions on an idempotent operator $E \in \mathcal{L}(\mathcal{H})$ - i.e., $E^2 = E$ - are equivalent:
(i) $E = E^*$;
(ii) $||E|| = 1$.
(Hint: Assume E is represented in matrix form, as in (4)(iii) above; notice that $x \in \mathcal{N}^\perp \Rightarrow ||Ex||^2 = ||x||^2 + ||Cx||^2$; conclude that $||E|| = 1 \ \Leftrightarrow\ C = 0$.)

(6) If E is the projection onto $\mathcal{M}$ along $\mathcal{N}$ - as above - show that there exists an invertible operator $S \in \mathcal{L}(\mathcal{H})$ such that $SES^{-1} = P_{\mathcal{M}}$. (Hint: Assume E and B are related as in (4)(ii) above; define

$$S = \begin{bmatrix} 1_{\mathcal{M}} & B \\ 0 & 1_{\mathcal{M}^\perp} \end{bmatrix} ;$$

deduce from (a transposed version of) Exercise 4.4.10 that S is invertible, and that

$$\begin{aligned} SES^{-1} &= \begin{bmatrix} 1_{\mathcal{M}} & B \\ 0 & 1_{\mathcal{M}^\perp} \end{bmatrix} \begin{bmatrix} 1_{\mathcal{M}} & B \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1_{\mathcal{M}} & -B \\ 0 & 1_{\mathcal{M}^\perp} \end{bmatrix} \\ &= \begin{bmatrix} 1_{\mathcal{M}} & 0 \\ 0 & 0 \end{bmatrix} . \end{aligned}$$

(7) Show that the following conditions on an operator $T \in \mathcal{L}(\mathcal{H})$ are equivalent:

(a) there exists closed subspaces $\mathcal{M}, \mathcal{N}$ as in (1) above such that

(i) $T(\mathcal{M}) \subset \mathcal{M}$ and $T|_{\mathcal{M}} = A$; and

(ii) $T(\mathcal{N}) \subset \mathcal{N}$ and $T|_{\mathcal{N}} = B$;

(b) there exists an invertible operator $S \in \mathcal{L}(\mathcal{H}, \mathcal{M} \oplus \mathcal{N})$ - where the direct sum considered is an 'external direct sum' - such that $STS^{-1} = A \oplus B$.

We will find the following bit of terminology convenient. Call operators $T_i \in \mathcal{L}(\mathcal{H}_i), i = 1, 2$, **similar** if there exists an invertible operator $S \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ such that $T_2 = ST_1S^{-1}$.

LEMMA 4.4.15 *The following conditions on an operator $T \in \mathcal{L}(\mathcal{H})$ are equivalent:*

(a) T is similar to an operator of the form $T_0 \oplus Q \in \mathcal{L}(\mathcal{M} \oplus \mathcal{N})$, where

(i) $\mathcal{N}$ is finite-dimensional;

(ii) T_0 is invertible, and Q is nilpotent.

(b) $T \in \mathcal{F}(\mathcal{H})$, $ind(T) = 0$ and there exists a positive integer n such that $ker\ T^n = ker\ T^m\ \forall\ m \geq n$.

Proof : $(a) \Rightarrow (b)$: If $STS^{-1} = T_0 \oplus Q$, then it is obvious that $ST^nS^{-1} = T_0^n \oplus Q^n$, which implies - because of the assumed invertibility of T_0 - that $ker\ T^n = S^{-1}(\{0\} \oplus ker\ Q^n)$, and hence, if $n = dim\mathcal{N}$, then for any $m \geq n$, we see that $kerT^m = S^{-1}(\{0\} \oplus \mathcal{N})$.

In particular, $kerT$ is finite-dimensional; similarly $ker\ T^*$ is also finite-dimensional, since $(S^*)^{-1}T^*S^* = T_0^* \oplus Q^*$; further,

$$ran\ T = S^{-1}(ran\ (T_0 \oplus Q)) = S^{-1}(\mathcal{M} \oplus (ran\ Q)) ,$$

which is closed since S^{-1} is a homeomorphism, and since the sum of the closed subspace $\mathcal{M} \oplus \{0\}$ and the finite-dimensional space $(\{0\} \oplus\ ran\ Q)$ is closed in $\mathcal{M} \oplus \mathcal{N}$ - see Exercise A.6.5(3). Hence T is a Fredholm operator.

Finally,

$$ind(T) = ind(STS^{-1}) = ind(T_0 \oplus Q) = ind(Q) = 0.$$

$(b) \Rightarrow (a)$: Let us write $\mathcal{M}_n = ran\ T^n$ and $\mathcal{N}_n = ker\ T^n$ for all $n \in \mathbb{N}$; then, clearly,

$$\mathcal{N}_1 \subset \mathcal{N}_2 \subset \cdots\ ;\ \mathcal{M}_1 \supset \mathcal{M}_2 \supset \cdots\ .$$

We are told that $\mathcal{N}_n = \mathcal{N}_m\ \forall\ m \geq n$. The assumption $ind\, T = 0$ implies that $ind\, T^m = 0\ \forall\ m$, and hence, we find that $dim(ker\, T^{*m}) = dim(ker\, T^m) = dim(ker\, T^n) = dim(ker\, T^{*n}) < \infty$ for all $m \geq n$. But since $ker\ T^{*m} = \mathcal{M}_m^\perp$, we find that $\mathcal{M}_m^\perp \subset \mathcal{M}_n^\perp$, from which we may conclude that $\mathcal{M}_m = \mathcal{M}_n\ \forall\ m \geq n$.

Let $\mathcal{N} = \mathcal{N}_n,\ \mathcal{M} = \mathcal{M}_n$, so that we have

$$\mathcal{N} = ker\ T^m \ and\ \mathcal{M} = ran\ T^m \quad \forall\ m \geq n\ . \tag{4.4.13}$$

The definitions clearly imply that $T(\mathcal{M}) \subset \mathcal{M}$ and $T(\mathcal{N}) \subset \mathcal{N}$ (since $\mathcal{M}$ and $\mathcal{N}$ are actually invariant under any operator which commutes with T^n).

We assert that $\mathcal{M}$ and $\mathcal{N}$ yield an algebraic direct sum decomposition of $\mathcal{H}$ (in the sense of Exercise 4.4.14(1)). Firstly, if $z \in \mathcal{H}$, then $T^n z \in \mathcal{M}_n = \mathcal{M}_{2n}$, and hence we can find $v \in \mathcal{H}$ such that $T^n z = T^{2n} v$; thus $z - T^n v \in ker\ T^n$; i.e., if $x = T^n v$ and $y = z - x$, then $x \in \mathcal{M}, y \in \mathcal{N}$ and $z = x + y$; thus, indeed $\mathcal{H} = \mathcal{M} + \mathcal{N}$. Notice that T (and hence also T^n) maps $\mathcal{M}$ onto itself; in particular, if $z \in \mathcal{M} \cap \mathcal{N}$, we can find an $x \in \mathcal{M}$ such that $z = T^n x$; the assumption $z \in \mathcal{N}$ implies that $0 = T^n z = T^{2n} x$; this means that $x \in \mathcal{N}_{2n} = \mathcal{N}_n$, whence $z = T^n x = 0$; since z was arbitrary, we have shown that $\mathcal{N} \cap \mathcal{M} = \{0\}$, and our assertion has been substantiated.

If $T_0 = T|_\mathcal{M}$ and $Q = T|_\mathcal{N}$, the (already proved) fact that $\mathcal{M} \cap \mathcal{N} = \{0\}$ implies that T^n is 1-1 on $\mathcal{M}$; thus T_0^n is 1-1; hence A is 1-1; it has already been noted that T_0 maps $\mathcal{M}$ onto $\mathcal{M}$; hence T_0 is indeed invertible; on the other hand, it is obvious that Q^n is the zero operator on $\mathcal{N}$. □

COROLLARY 4.4.16 *Let* $K \in \mathcal{K}(\mathcal{H})$*; assume* $0 \neq \lambda \in \sigma(K)$*; then* K *is similar to an operator of the form* $K_1 \oplus A \in \mathcal{L}(\mathcal{M} \oplus \mathcal{N})$*, where*

(a) $K_1 \in \mathcal{K}(\mathcal{M})$ *and* $\lambda \notin \sigma(K_1)$*; and*

(b) $\mathcal{N}$ *is a finite-dimensional space, and* $\sigma(A) = \{\lambda\}$.

Proof : Put $T = K - \lambda$; then, the hypothesis and Theorem 4.4.11 ensure that T is a Fredholm operator with $ind(T) = 0$. Consider the non-decreasing sequence

$$kerT \subset kerT^2 \subset \cdots \subset kerT^n \subset \cdots \ . \tag{4.4.14}$$

Suppose $ker\ T^n \ \neq kerT^{n+1}\ \forall\ n$; then we can pick a unit vector $x_n \in (ker\ T^{n+1}) \cap (ker\ T^n)^\perp$ for each n. Clearly the sequence $\{x_n\}_{n=1}^\infty$ is an orthonormal set. Hence, $lim_n||Kx_n|| = 0$ (by Exercise 4.4.17(3)).

On the other hand,

$$\begin{aligned} x_n \in ker\ T^{n+1} &\Rightarrow Tx_n \in ker\ T^n \\ &\Rightarrow \langle Tx_n, x_n\rangle = 0 \\ &\Rightarrow \langle Kx_n, x_n\rangle \ = \ \lambda \end{aligned}$$

contradicting the hypothesis that $\lambda \neq 0$ and the already drawn conclusion that $Kx_n \to 0$.

Hence, it must be the case that $ker\ T^n = ker\ T^{n+1}$ for some $n \in \mathbb{N}$; it follows easily from this that $ker\ T^n = ker\ T^m\ \forall\, m \geq n$.

Thus, we may conclude from Lemma 4.4.15 that there exists an invertible operator $S \in \mathcal{L}(\mathcal{H}, \mathcal{M} \oplus \mathcal{N})$ - where $\mathcal{N}$ is finite-dimensional - such that $STS^{-1} = T_0 \oplus Q$, where T_0 is invertible and $\sigma(Q) = \{0\}$; since $K = T + \lambda$, conclude that $SKS^{-1} \ = \ (T_0 + \lambda) \oplus (Q + \lambda)$; set $K_1 = T_0 + \lambda, A = Q + \lambda$, and conclude that indeed K_1 is compact, $\lambda \notin \sigma(K_1)$ and $\sigma(A) = \{\lambda\}$. □

EXERCISE 4.4.17 *(1) Let X be a metric space; if $x, x_1, x_2, \cdots \in X$, show that the following conditions are equivalent:*

(i) the sequence $\{x_n\}_n$ converges to x;

(ii) every subsequence of $\{x_n\}_n$ has a further subsequence which converges to x.

(Hint: for the non-trivial implication, note that if the sequence $\{x_n\}_n$ does not converge to x, then there must exist a subsequence whose members are 'bounded away from x'.)

(2) Show that the following conditions on an operator $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ are equivalent:

(i) T is compact;

(ii) if $\{x_n\}$ *is a sequence in* $\mathcal{H}_1$ *which converges weakly to 0 - i.e,* $\langle x, x_n \rangle \to 0 \ \forall \ x \in \mathcal{H}_1$ *- then* $||Tx_n|| \to 0$.

(iii) if $\{e_n\}$ *is any infinite orthonormal sequance in* $\mathcal{H}_1$*, then* $||Te_n|| \to 0$.

(Hint: for $(i) \Rightarrow (ii)$*, suppose* $\{y_n\}_n$ *is a subsequence of* $\{x_n\}_n$*; by compactness, there is a further subsequence* $\{z_n\}_n$ *of* $\{y_n\}_n$ *such that* $\{Tz_n\}_n$ *converges, to* z*, say; since* $z_n \to 0$ *weakly, deduce that* $Tz_n \to 0$ *weakly; this means* $z = 0$*, since strong convergence implies weak convergence; by (1) above, this proves (ii). The implication* $(ii) \Rightarrow (iii)$ *follows form the fact that any orthonormal sequence converges weakly to 0. For* $(iii) \Rightarrow (i)$*, deduce from Proposition 4.3.6(c) that if* T *is not compact, there exists an* $\epsilon > 0$ *such that* $\mathcal{M}_\epsilon = ran \ 1_{[\epsilon,\infty)}(|T|)$ *is infinite-dimensional; then any infinite orthonormal set*$\{e_n : n \in \mathbb{N}\}$ *in* $\mathcal{M}_\epsilon$ *would violate condition (iii).)*

We are finally ready to state the spectral theorm for a compact operator.

THEOREM 4.4.18 *Let* $K \in \mathcal{K}(\mathcal{H})$ *be a compact operator on a Hilbert space* $\mathcal{H}$*. Then,*

(a) $\lambda \in \sigma(K) - \{0\} \Rightarrow \lambda$ *is an eigenvalue of* K *and* λ *is 'isolated' in the sense that there exists* $\epsilon > 0$ *such that* $0 < |z - \lambda| < \epsilon \Rightarrow z \notin \sigma(K)$*;*

(b) if $\lambda \in \sigma(K) - \{0\}$*, then* λ *is an eigenvalue with 'finite algebraic multiplicity' in the strong sense described by Corollary 4.4.16;*

(c) $\sigma(K)$ *is countable, and the only possible accumulation point of* $\sigma(K)$ *is 0.*

Proof : Assertions (a) and (b) are immediate consequences of Corollary 4.4.16, while (c) follows immediately from (a). □

Chapter 5

Unbounded operators

5.1 Closed operators

This chapter will be concerned with some of the basic facts concerning unbounded operators on Hilbert spaces. We begin with some definitions. It will be necessary for us to consider operators which are not defined everywhere in a Hilbert space, but only on some linear subspace which is typically not closed; in fact, our primary interest will be in operators which are defined on a dense proper subspace. (Since we shall only consider Hilbert spaces, we adopt the following conventions throughout this chapter.)

Symbols $\mathcal{H}$ and $\mathcal{K}$ (and primed and subscripted variations thereof) will be reserved for separable complex Hilbert spaces, and the symbol $\mathcal{D}$ (and its variants) will be reserved for linear (typically non-closed) subspaces of Hilbert spaces.

DEFINITION 5.1.1 *A* **linear operator** *'from $\mathcal{H}$ to $\mathcal{K}$' is a linear map $T : \mathcal{D} \to \mathcal{K}$, where $\mathcal{D}$ is a linear subspace of $\mathcal{H}$; the subspace $\mathcal{D}$ is referred to as the* **domain** *(of definition) of the operator T, and we shall write $\mathcal{D} = dom(T)$ to indicate this relationship between $\mathcal{D}$ and T.*

The linear operator T is said to be **densely defined** *if $\mathcal{D} = domT$ is dense in $\mathcal{H}$.*

The reason for our interest in densely defined operators lies in the following fact.

Proposition 5.1.2 *Let $T : \mathcal{D}(\subset \mathcal{H}) \to \mathcal{K}$ be a linear operator; assume that T is densely defined.*

(a) The following conditions on a vector $y \in \mathcal{K}$ are equivalent:

(i) there exists a constant $C > 0$ such that $|\langle Tx, y\rangle| \leq C||x|| \;\forall\; x \in \mathcal{D}$;

(ii) there exists a vector $z \in \mathcal{H}$ such that $\langle Tx, y\rangle = \langle x, z\rangle$ for all $x \in \mathcal{D}$.

(b) Let $\mathcal{D}^*$ denote the set of all $y \in \mathcal{K}$ which satisfy the equivalent conditions of (a); then

(i) if $y \in \mathcal{D}^*$, there exists a unique $z \in \mathcal{H}$ satisfying the requirements listed in (a)(ii); and

(ii) there exists a unique linear operator $T^* : \mathcal{D}^* \to \mathcal{H}$ such that

$$\langle Tx, y\rangle = \langle x, T^*y\rangle \;\forall\; x \in \mathcal{D}, y \in \mathcal{D}^* .$$

Proof : (a) $(i) \Rightarrow (ii)$: If (i) is satisfied, then the mapping $\mathcal{D} \ni x \mapsto \langle Tx, y\rangle \in \mathbb{C}$ defines a bounded linear functional on $\mathcal{D}$ which consequently - see Exercise 1.5.5(1)(b) - has a unique extension to a bounded linear functional on $\overline{\mathcal{D}} = \mathcal{H}$; the assertion (ii) follows now from the Riesz representation.

$(ii) \Rightarrow (i)$: Obvious, by Cauchy-Schwarz.

(b) (i) The uniqueness of z is a direct consequence of the assumed density of $\mathcal{D}$.

(ii) If $y \in \mathcal{D}^*$, define T^*y to be the unique element z as in (a)(ii); the uniqueness assertion in (b)(i) above clearly implies the linearity of T^* (in exactly the same manner in which the linearity of the adjoint of a bounded operator was established). The operator T^* has the stipulated property by definition, and it is uniquely determined by this property in view of the uniqueness assertion (b)(i). □

Definition 5.1.3 *The* **adjoint** *of a densely defined operator T (as in Proposition 5.1.2) is the unique operator T^* whose domain and action are as prescribed in part (b) of Proposition 5.1.2. (It should be noted that we will talk about the adjoint of an operator* **only when** *it is densely defined; hence if we do talk about the adjoint of an operator, it will* **always be understood** *- even if it has not been explicitly stated - that the original operator is densely defined.)*

We now list a few useful examples of unbounded operators.

EXAMPLE 5.1.4 (1) Let $\mathcal{H} = L^2(X, \mathcal{B}, \mu)$, where $(X, \mathcal{B}, \mu)$ is a σ-finite measure space; given any measurable function $\phi : X \to \mathbb{C}$, let

$$\mathcal{D}_\phi \ = \ \{f \in L^2(\mu) : \ \phi f \in L^2(\mu)\} \ , \tag{5.1.1}$$

and define M_ϕ to be the operator given by

$$M_\phi f \ = \ \phi f \ , \ \forall \ f \in \mathcal{D}_\phi \ = \ dom \ M_\phi \ .$$

(2) If $k : X \times Y \to \mathbb{C}$ is a measurable function, where $(X, \mathcal{B}_X, \mu)$ and $(Y, \mathcal{B}_Y, \nu)$ are σ-finite measure spaces, we can associate the integral operator $Int \ k$, whose natural domain $\mathcal{D}_k$ is the set of those $g \in \mathcal{L}^2(Y, \nu)$ which satisfy the following two conditions: (a) $k(x, \cdot)g \in L^1(Y, \nu)$ for μ-almost every x, and (b) the function given by $((Int \ k)g)(x) \ = \ \int_X \ k(x, y)g(y) \ d\nu(y)$, (which, by (a), is μ-a.e. defined), satisfies $(Int \ k)g \in L^2(X, \mu)$.

(3) A very important source of examples of unbounded operators is the study of differential equations. To be specific, suppose we wish to study the differential expression

$$\tau \ = \ - \frac{d^2}{dx^2} + q(x) \tag{5.1.2}$$

on the interval $[a, b]$. By this we mean we want to study the passage $f \ \mapsto \ \tau f$, where $(\tau f)(x) = -\frac{d^2 f}{dx^2} + q(x)f(x)$, it being assumed, of course, that the function f is appropriately defined and is sufficiently 'smooth' as to make sense of τf.

The typical Hilbert space approach to such a problem is to start with the Hilbert space $\mathcal{H} = L^2([a, b]) = L^2([a, b], \mathcal{B}_{[a,b]}, m)$, where m denotes Lebesgue measure restricted to [a,b], and to study the operator $Tf \ = \ \tau f$ defined on the 'natural domain' $dom \ T$ consisting of those functions $f \in \mathcal{H}$ for which the second derivative f'' 'makes sense' and $\tau f \in L^2([a, b])$.

(4) Let $((\alpha^i_j))_{0 \leq i,j < \infty}$ be an arbitrary matrix of complex numbers, with rows and columns indexed as indicated. We wish to define the associated operator on ℓ^2 given by matrix multiplication; to be precise, we think of the typical element of ℓ^2 as a column vector $((x_j))$ of complex numbers such that $\sum_j |x_j|^2 < \infty$;

define $\mathcal{D}$ to be the set of those $x = (x_j) \in \ell^2$ which satisfy the following two conditions:

(i) $\sum_{j=0}^{\infty} |\alpha_j^i \, x_j| < \infty$ for all $i \geq 0$; and

(ii) if we define $y_i = \sum_{j=0}^{\infty} \alpha_j^i \, x_j$, then the vector $y = ((y_i))$ belongs to ℓ^2; in short, we require that $\sum_{i=0}^{\infty} |y_j|^2 < \infty$.

It is easily verified that $\mathcal{D}$ is a vector subspace of ℓ^2 and that the passage $x \mapsto y = Ax$ defines a linear operator A with domain given by $\mathcal{D}$. (The astute reader would have noticed that this example is a special case of example (2) above.)

It is an interesting exercise to determine when such a matrix yields a densely defined operator and what the adjoint looks like. The reader who pursues this line of thinking should, for instance, be able to come up with densely defined operators A such that *dom* $A^* = \{0\}$. □

As must have been clear with the preceding examples, it is possible to consider various different domains for the same 'formal expression' which defines the operator. This leads us to the notion of extensions and restrictions of operators, which is made precise in the following exercises.

Exercise 5.1.5 *(1) For $i = 1, 2$, let $\mathcal{D}_i \subset \mathcal{H}$ and let $T_i : \mathcal{D}_i \to \mathcal{K}$ be a linear operator. Show that the following conditions are equivalent:*

(i) $G(T_1) \subset G(T_2)$, where the graph $G(T)$ of a linear operator T 'from $\mathcal{H}$ to $\mathcal{K}$' is defined by $G(T) = \{(x, y) \in \mathcal{H} \oplus \mathcal{K} : x \in$ dom $T, \ y = Tx\}$;

(ii) $\mathcal{D}_1 \subset \mathcal{D}_2$ and $T_1 x = T_2 x \ \forall \ x \in \mathcal{D}_1$.

When these equivalent conditions are met, we say that T_2 is an **extension** *of T_1 or that T_1 is a* **restriction** *of T_2, and we denote this relationship by $T_2 \supset T_1$ or by $T_1 \subset T_2$.*

(2) (a) If S and T are linear operators, show that $S \subset T^*$ if and only if $\langle Tx, y \rangle = \langle x, Sy \rangle$ for all $x \in$ dom T, $y \in$ dom S.

(b) If $S \subset T$, and if S is densely defined, then show that (also T is densely defined so that T^* makes sense, and that) $S^* \supset T^*$.

We now come to one of the fundamental observations.

PROPOSITION 5.1.6 *Suppose $T : \mathcal{D} \to \mathcal{K}$ is a densely defined linear operator, where $\mathcal{D} \subset \mathcal{H}$. Then the graph $G(T^*)$ of the adjoint operator is a closed subspace of the Hilbert space $\mathcal{K} \oplus \mathcal{H}$; in fact, we have:*

$$G(T^*) = \{(Tx, -x) : x \in \mathcal{D}\}^\perp . \tag{5.1.3}$$

Proof : Let us write $\mathcal{S} = \{(Tx, -x) : x \in \mathcal{D}\}$; thus $\mathcal{S}$ is a linear subspace of $\mathcal{K} \oplus \mathcal{H}$. If $y \in dom\ T^*$, then, by definition, we have $\langle Tx, y\rangle = \langle x, T^*y\rangle \ \forall\ x \in \mathcal{D}$, or equivalently, $(y, T^*y) \in \mathcal{S}^\perp$.

Conversely, suppose $(y, z) \in \mathcal{S}^\perp$; this means that $\langle Tx, y\rangle + \langle -x, z\rangle = 0 \ \forall x \in \mathcal{D}$, or equivalently, $\langle Tx, y\rangle = \langle x, z\rangle \ \forall\ x \in \mathcal{D}$. Deduce from Proposition 5.1.2(a)(ii) that $y \in domT^*$ and that $T^*y = z$; thus, we also have $\mathcal{S}^\perp \subset G(T^*)$. □

DEFINITION 5.1.7 *A linear operator $T : \mathcal{D}\ (\subset \mathcal{H}) \to \mathcal{K}$ is said to be* **closed** *if its graph $G(T)$ is a closed subspace of $\mathcal{H} \oplus \mathcal{K}$.*

A linear operator T is said to be **closable** *if it has a closed extension.*

Thus proposition 5.1.6 shows that the adjoint of a densely defined operator is always a closed operator. Some further consequences of the preceding definitions and proposition are contained in the following exercises.

EXERCISE 5.1.8 *In the following exercise, it will be assumed that $\mathcal{D} \subset \mathcal{H}$ is a linear subspace.*

(a) Show that the following conditions on a linear subspace $\mathcal{S} \subset \mathcal{H} \oplus \mathcal{K}$ are equivalent:

(i) there exists some linear operator T 'from $\mathcal{H}$ to $\mathcal{K}$' such that $\mathcal{S} = G(T)$;

(ii) $\mathcal{S} \cap (\{0\} \oplus \mathcal{K}) = \{(0, 0)\}$.

(Hint: Clearly (i) implies (ii) since linear maps map 0 to 0; to see that (ii) implies (i), define $\mathcal{D} = P_1(\mathcal{S})$, where P_1 denotes the projection of $\mathcal{H} \oplus \mathcal{K}$ onto the first co-ordinate, and note that (ii) may be re-phrased thus: if $x \in \mathcal{D}$, then there exists a unique $y \in \mathcal{K}$ such that $(x, y) \in \mathcal{S}$.)

(b) Show that the following conditions on the linear operator $T : \mathcal{D} \to \mathcal{H}_2$ are equivalent:

(i) T is closable;

(ii) if $\{x_n\}_n$ is a sequence in $\mathcal{D}$ such that $x_n \to 0 \in \mathcal{H}$ and $Tx_n \to y \in \mathcal{K}$, then $y = 0$.
(Hint: (ii) is just a restatement of the fact that if $(0, y) \in \overline{G(T)}$, then $y = 0$. It is clear now that $(i) \Rightarrow (ii)$. Conversely, if condition (ii) is satisfied, we find that $\mathcal{S} = \overline{G(T)}$ satisfies condition (a)(ii) above, and it follows from this and (a) that $\mathcal{S}$ is the graph of an operator which is necessarily a closed extension - in fact, the closure, in the sense of (c) below - of T.)

(c) If $T : \mathcal{D} \to \mathcal{K}$ is a closable operator, then show that there exists a unique closed operator - always denoted by $\overline{T}$, and referred to as the **closure** *of T - such that $G(\overline{T})$ is the closure $\overline{G(T)}$ of the graph of T; and deduce that $\overline{T}$ is the smallest closed extension of T in the sense that it is a restriction of any closed extension of T.*

We now reap some consequences of Proposition 5.1.6.

COROLLARY 5.1.9 *Let $T : \mathcal{D}\ (\subset \mathcal{H}) \to \mathcal{K}$ be a densely defined linear operator. Then, T is closable if and only if T^* is densely defined, in which case $\overline{T} = T^{**}$.*

In particular, if T is a closed densely defined operator, then $T = T^{**}$.

Proof : (a) If T^* were densely defined, then $T^{**} = (T^*)^*$ makes sense and it follows from Exercise 5.1.5(2)(a), for instance, that $T \subset T^{**}$, and hence T^{**} is a closed extension of T.

Conversely, suppose T is closable. Let $y \in (dom\ T^*)^\perp$; note that $(y, 0) \in G(T^*)^\perp$ and deduce from equation 5.1.3 that $(y, 0)$ belongs to the closure of $\{(Tx, -x) : x \in dom\ T\}$; this, in turn, implies that $(0, y)$ belongs to $\overline{G(T)} = G(\overline{T})$, and consequently that $y = 0$. So 'T closable' implies 'T^* densely defined'.

Let $W : \mathcal{H} \oplus \mathcal{K} \to \mathcal{K} \oplus \mathcal{H}$ denote the (clearly unitary) operator defined by $W(x, y) = (y, -x)$, then equation 5.1.3 can be restated as

$$G(T^*) = W(\ G(T)\ )^\perp . \tag{5.1.4}$$

On the other hand, observe that $W^* : \mathcal{K} \oplus \mathcal{H} \to \mathcal{H} \oplus \mathcal{K}$ is the unitary operator given by $W^*(y, x) = (-x, y)$, and equation

5.1.4 (applied to T^* in place of T) shows that

$$\begin{aligned} G(T^{**}) &= (-W^*)(\ G(T^*)\ )^\perp \\ &= (\ W^*(G(T^*)\ )^\perp \\ &= (\ G(T)^\perp)^\perp \quad \text{(by 5.1.4)} \\ &= \overline{G(T)} \\ &= G(\overline{T})\ . \end{aligned}$$

□

We thus find that the class of closed densely defined operators between Hilbert spaces is closed under the adjoint operation (which is an 'involutory operation' on this class). We shall, in the sequel, be primarily interested in this class; we shall use the symbol $\mathcal{L}_c(\mathcal{H}, \mathcal{K})$ to denote the class of densely defined closed linear operators $T : dom\ T\ (\subset \mathcal{H}) \ \to \mathcal{K}$. Some more elementary facts about (domains of combinations of) unbounded operators are listed in the following exercise.

Exercise 5.1.10 *(a) The following conditions on a closed operator T are equivalent:*

(i) dom T is closed;

(ii) T is bounded.

(Hint: (i) ⇒ (ii) is a consequence of the closed graph theorem. (ii) ⇒ (i): if $x \in \overline{dom\ T}$, and if $\{x_n\}_n \subset dom\ T$ is such that $x_n \to x$, then $\{Tx_n\}_n$ converges, to y say; thus $(x, y) \in \overline{G(T)}$ and in particular (since T is closed) $x \in dom\ T$.)

(b) If T (resp. T_1) : dom T (resp. T_1) $(\subset \mathcal{H}) \ \to \mathcal{K}$ and S : dom S $(\subset \mathcal{K}) \ \to \mathcal{K}_1$ are linear operators, the composite ST, the sum $T + T_1$ and the scalar multiple αT are the operators defined 'pointwise' by the obvious formulae, on the following domains:

$$\begin{aligned} dom(ST) &= \{x \in dom\ T : Tx \in dom\ S\} \\ dom(T + T_1) &= dom\ T\ \cap\ dom\ T_1 \\ dom(\alpha T) &= dom\ T\ . \end{aligned}$$

Show that:

(i) $T + T_1 = T_1 + T$;

(ii) $S(\alpha T + T_1) \supset (\alpha ST + ST_1)$.

(iii) if ST is densely defined, then $T^*S^* \subset (ST)^*$.

(iv) if $A \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ is an everywhere defined bounded operator, then $(T + A)^* = T^* + A^*$ and $(AT)^* = T^*A^*$.

(c) In this problem, we only consider (one Hilbert space $\mathcal{H}$ and) linear operators 'from $\mathcal{H}$ into itself'. Let T be a linear operator, and let $A \in \mathcal{L}(\mathcal{H})$ (so A is an everywhere defined bounded operator). Say that T commutes with A if it is the case that $AT \subset TA$.

(i) Why is the inclusion oriented the way it is? (also answer the same question for the inclusions in (b) above!)

(ii) For T and A as above, show that

$$AT \subset TA \;\Rightarrow\; A^*T^* \subset T^*A^* \;;$$

in other words, if A and T commute, then A^* and T^* commute. (Hint: Fix $y \in dom\ T^* = dom\ A^*T^*$; it is to be verified that then, also $A^*y \in dom\ T^*$ and $T^*(A^*y) = A^*T^*y$; for this, fix $x \in dom\ T$; then by hypothesis, $Ax \in dom\ T$ and $TAx = ATx$; then observe that

$$\langle x, A^*T^*y\rangle = \langle Ax, T^*y\rangle = \langle T(Ax), y\rangle = \langle ATx, y\rangle = \langle Tx, A^*y\rangle \;,$$

which indeed implies - since $x \in dom\ T$ was arbitrary - that $A^*y \in dom\ T^*$ and that $T^*(A^*y) = A^*T^*y$.)

(d) For an arbitrary set $\mathcal{S} \subset \mathcal{L}_c(\mathcal{H})$ (of densely defined closed operators 'from $\mathcal{H}$ to itself'), define

$$\mathcal{S}' \;=\; \{A \in \mathcal{L}(\mathcal{H}) :\; AT \subset TA \;\forall\; T \in \mathcal{S}\}$$

and show that:

(i) $\mathcal{S}'$ is a unital weakly closed subalgebra of $\mathcal{L}(\mathcal{H})$;

(ii) if $\mathcal{S}$ is self-adjoint - meaning that $\mathcal{S} = \mathcal{S}^* = \{A^* : A \in \mathcal{S}\}$ - then $\mathcal{S}'$ is also self-adjoint, and is consequently a von Neumann algebra.

5.2 Symmetric and Self-adjoint operators

DEFINITION 5.2.1 *A densely defined linear operator* $T : \mathcal{D} \ (\subset \mathcal{H}) \ \to \mathcal{H}$ *is said to be* **symmetric** *if it is the case that* $T \subset T^*$.

A linear operator $T : \mathcal{D} \ (\subset \mathcal{H}) \ \to \mathcal{H}$ *is said to be* **self-adjoint** *if* $T = T^*$.

Note that symmetric operators are assumed to be densely defined - since we have to first make sense of their adjoint! Also, observe (as a consequence of Exercise 5.1.52(b)) that any (densely-defined) restriction of any symmetric (in particular, a self-adjoint) operator is again symmetric, and so symmetric operators are much more common than self-adjoint ones.

Some elementary consequences of the definitions are listed in the following proposition.

PROPOSITION 5.2.2 *(a) The following conditions on a densely defined linear operator* $T_0 : dom\ T_0 \ (\subset \mathcal{H}) \ \to \mathcal{H}$ *are equivalent:*

(i) T_0 *is symmetric;*

(ii) $\langle T_0 x, y\rangle \ = \ \langle x, T_0 y\rangle \ \forall \ x, y \in \ dom\ T_0$;

(iii) $\langle T_0 x, x\rangle \ \in \ \mathbb{R} \ \forall \ x \in \ dom\ T_0$.

(b) If T_0 *is a symmetric operator, then*

$$||(T_0 \pm i)x||^2 \ = \ ||T_0 x||^2 + ||x||^2 \ , \ \forall \ x \in dom\ T_0 \ , \qquad (5.2.5)$$

and, in particular, the operators $(T_0 \pm i)$ *are (bounded below, and consequently) 1-1.*

(c) If T_0 *is symmetric, and if* T *is a self-adjoint extension of* T, *then necessarily* $T_0 \subset T \subset T_0^*$.

Proof : (a) Condition (ii) is essentially just a re-statement of condition (i) - see Exercise 5.1.5(2)(a); the implication $(ii) \Rightarrow (iii)$ is obvious, while $(iii) \Rightarrow (ii)$ is a consequence of the polarisation identity.

(b) follows from (a)(iii) - simply 'expand' the left side and note that the cross terms cancel out under the hypothesis.

(c) This is an immediate consequence of Exercise 5.1.5(2)(b).

□

Example 5.2.3 Consider the operator $D = \frac{d}{dt}$ of differentiation, which may be applied on the space of functions which are at least once differentiable. Suppose we are interested in the interval [0,1]. We then regard the Hilbert space $\mathcal{H} = L^2[0,1]$ - where the measure in question is usual Lebesgue measure. Consider the following possible choices of domains for this operator:

$$\begin{aligned}
\mathcal{D}_0 &= C_c^\infty(0,1) \ (\subset \mathcal{H}) \\
\mathcal{D}_1 &= \{f \in C^1([0,1]) : f(0) = f(1) = 0\} \\
\mathcal{D}_2 &= C^1([0,1])
\end{aligned}$$

Let us write $T_j f = iDf, \ f \in \mathcal{D}_j \ , \ j = 0,1,2$. Notice that $T_0 \subset T_1 \subset T_2$, and that if $f, g \in \mathcal{D}_2$, then since $(fg)' = f'g + fg'$ and since $\int_0^1 h'(t)dt = h(1) - h(0)$, we see that

$$\begin{aligned}
f(1)\overline{g(1)} - f(0)\overline{g(0)} &= \int_0^1 (f\overline{g})'(t)dt \\
&= \int_0^1 (\ f(t)\overline{g'(t)} + f'(t)\overline{g(t)}\ )dt \ ,
\end{aligned}$$

and consequently we find (what is customarily referred to as the formula obtained by 'integration by parts'):

$$i(f(1)\overline{g(1)} - f(0)\overline{g(0)}) \ = \ -\langle f, iDg\rangle + \langle iDf, g\rangle \ .$$

In particular, we see that T_2 is not symmetric (since there exist $f, g \in \mathcal{D}_2$ such that $(f(1)\overline{g(1)} - f(0)\overline{g(0)}) \neq 0$), but that T_1 is indeed symmetric; in fact, we also see that $T_2 \subset T_1^* \subset T_0^*$. □

Lemma 5.2.4 *Suppose T : dom T $(\subset \mathcal{H}) \ \to \mathcal{H}$ is a (densely defined) closed symmetric operator. Suppose $\lambda \in \mathbb{C}$ is a complex number with non-zero imaginary part. Then $(T-\lambda)$ maps dom T 1-1 onto a closed subspace of $\mathcal{H}$.*

Proof : Let $\lambda = a + ib$, with $a, b \in \mathbb{R}, b \neq 0$. It must be observed, exactly as in equation 5.2.5 (which corresponds to the case $a = 0, b = \mp 1$ of equation 5.2.6), that

$$||(T-\lambda)x||^2 \ = \ ||(T-a)x||^2 \ + \ |b|^2||x||^2 \ \geq \ |b|^2||x||^2 \ , \quad (5.2.6)$$

with the consequence that $(T - \lambda)$ is indeed bounded below. In particular, $(T-\lambda)$ is injective. If $x_n \in dom\ T$ and if $(T-\lambda)x_n \to y$ (say), then it follows that $\{x_n\}$ is also a Cauchy sequence, and that if $x = lim_n x_n$, then $(x, y + \lambda x)$ belongs to the closure of the graph of the closed operator T; hence $x \in dom\ T$ and $(T - \lambda)x = y$, thereby establishing that $(T - \lambda)$ indeed maps $dom\ T$ onto a closed subspace . □

In the sequel, we shall write $ker\ S = \{x \in dom\ S : Sx = 0\}$ and $ran\ S = \{Sx : x \in dom\ S\}$, for any linear operator S.

LEMMA 5.2.5 *(a) If $T:\ dom\ T\ (\subset \mathcal{H})\ \to \mathcal{K}$ is a closed operator, then $ker\ T$ is a closed subspace of $\mathcal{H}$.*

(b) If $T:\ dom\ T\ (\subset \mathcal{H})\ \to \mathcal{K}$ is a densely defined operator, then

$$(ran\ T)^{\perp}\ =\ ker\ T^*\ .$$

Proof : (a) Suppose $x_n \to x$, where $x_n \in ker\ T\ \forall n$; then

$$(x_n, Tx_n)\ =\ (x_n, 0) \to (x, 0)\ \in \mathcal{H} \oplus \mathcal{K}\ ,$$

and the desired conclusion is a consequence of the assumption that $G(T)$ is closed.

(b) If $y \in \mathcal{K}$, note that

$$\begin{aligned} y \in (ran\ T)^{\perp} &\Leftrightarrow \langle Tx, y\rangle\ =\ 0\ \forall\ x \in dom\ T \\ &\Leftrightarrow y \in ker\ T^*\ , \end{aligned}$$

as desired. □

PROPOSITION 5.2.6 *The following conditions on a closed symmetric operator $T:\ dom\ T\ (\subset \mathcal{H})\ \to \mathcal{H}$ are equivalent:*

(i) $ran\ (T - i)\ =\ ran\ (T + i)\ =\ \mathcal{H}$;

(ii) $ker\ (T^* + i)\ =\ ker\ (T^* - i)\ =\ \{0\}$;

(iii) T is self-adjoint.

Proof : $(i) \Leftrightarrow (ii)$: This is an immediate consequence of Lemma 5.2.5(b), Exercise 5.1.10(b)(iv), and Lemma 5.2.4.

$(i) \Rightarrow (iii)$: Let $x \in dom\ T^*$ be arbitrary; since $(T - i)$ is assumed to be onto, we can find a $y \in dom\ T$ such that $(T-i)y\ =\ (T^*-i)x$; on the other hand, since $T \subset T^*$, we find that $(x-y) \in ker(T^* - i) = ran(T+i)^\perp = \{0\}$; in particular $x \in dom\ T$ and $T^*x = Tx$; i.e., also $T^* \subset T$, as desired.

$(iii) \Rightarrow (ii)$ If $T = T^*$, then $ker(T^* \pm i) = ker(T \pm i)$ which is trivial, by Proposition 5.2.2(b). □

We now turn to the important construction of the Cayley transform of a closed symmetric operator.

PROPOSITION 5.2.7 *Let $T_0 : dom\ T_0\ (\subset \mathcal{H}) \ \to \mathcal{H}$ be a closed symmetric operator. Let $\mathcal{R}_0(\pm) = ran(T_0 \pm i) = ker^\perp(T_0^* \mp i)$. Then,*

(a) there exists a unique partial isometry $U_{T_0} \in \mathcal{L}(\mathcal{H})$ such that

$$U_{T_0} z \ = \ \begin{cases} (T_0 - i)x & if\ z = (T_0 + i)x \\ 0 & if\ z \in \mathcal{R}_0(+)^\perp\ ; \end{cases} \tag{5.2.7}$$

furthermore,

(i) U_{T_0} has initial (resp., final) space equal to $\mathcal{R}_0(+)$ (resp., $\mathcal{R}_0(-)$);

(ii) 1 is not an eigenvalue of U_{T_0}.

We shall refer to U_{T_0} as the **Cayley transform** *of the closed symmetric operator T.*

(b) If $T \supset T_0$ is a closed symmetric extension of T_0, and if the initial and final spaces of the partial isometry U_T are denoted by $\mathcal{R}(+)$ and $\mathcal{R}(-)$ respectively, then

(i) U_T 'dominates' the partial isometry U_{T_0} in the sense that $\mathcal{R}(+) \supset \mathcal{R}_0(+)$ and $U_T z = U_{T_0} z,\ \forall\ z \in \mathcal{R}_0(+)$, so that also $\mathcal{R}(-) \supset \mathcal{R}_0(-)$; and

(ii) 1 is not an eigenvalue of U_T.

(c) Conversely suppose U is a partial isometry with initial space $\mathcal{R}$, say, and suppose (i) U 'dominates' U_{T_0} (meaning that $\mathcal{R} \supset \mathcal{R}_0(+)$ and $Uz = U_{T_0} z\ \forall\ z \in \mathcal{R}_0(+)$), and (ii) 1 is not an eigenvalue of U.

Then there exists a unique closed symmetric extension $T \supset T_0$ such that $U = U_T$.

Proof : (a) Equation 5.2.5 implies that the passage

$$\mathcal{R}_0(+) \ni z = (T+i)x \stackrel{U_T}{\mapsto} (T-i)x = y \in \mathcal{R}_0(-) \qquad (5.2.8)$$

defines an isometric (linear) map of $\mathcal{R}_0(+)$ onto $\mathcal{R}_0(-)$. Hence there does indeed exist a partial isometry with initial (resp. final) space $\mathcal{R}_0(+)$ (resp., $\mathcal{R}_0(-)$), which satisfies the rule 5.2.8; the uniqueness assertion is clear since a partial isometry is uniquely determined by its action on its initial space.

As for (ii), suppose $Uz = z$ for some partial isometry U. Then, $||z|| = ||Uz|| = ||UU^*Uz|| \leq ||U^*Uz||$; deduce from Exercise 2.3.15(2)(b) that $z = U^*Uz = U^*z$. Hence 1 is an eigenvalue of a partial isometry U if and only if it is an eigenvalue of U^*; and $ker(U-1) = ker(U^*-1) \subset ker(U^*U-1)$.

In particular, suppose $U_{T_0}z = z$ for some $z \in \mathcal{H}$. It then follows from the last paragraph that $z \in \mathcal{R}_0(+)$. Finally, if z, x, y are as in 5.2.8, then, observe that

$$\begin{aligned} z &= T_0x + ix, & y &= T_0x - ix, \\ x = \frac{1}{2i}(z-y) &= \frac{1}{2i}(1-U_{T_0})z, & T_0x &= \tfrac{1}{2}(z+y) = \tfrac{1}{2}(1+U)z \ ; \end{aligned} \qquad (5.2.9)$$

and so, we find that

$$U_{T_0}z = z \ \Rightarrow \ z = T_0x + ix, \text{ where } x = 0 \ \Rightarrow \ z = 0 \ .$$

(b) If $z \in \mathcal{R}_0(+)$, then there exists a (necessarily unique, by Lemma 5.2.4) $x \in dom\ T_0$, $y \in \mathcal{R}_0(-)$ as in 5.2.8; then, $z = T_0x + ix = Tx + ix \in \mathcal{R}(+)$ and $U_Tz = Tx - ix = T_0x - ix = y$, and hence U_T does 'dominate' U_{T_0}, as asserted in (i); assertion (ii) follows from (a)(ii) (applied with T in place of T_0).

(c) Suppose $U, \mathcal{R}$ are as in (c). Let $\mathcal{D} = \{x \in \mathcal{H} : \exists\ z \in \mathcal{R}$ such that $x = z - Uz\}$. Then, the hypothesis that U 'dominates' U_{T_0} shows that

$$\mathcal{D} \ = \ (1-U)(\mathcal{R}) \supset (1-U_{T_0})(\mathcal{R}_0(+)) = dom\ T_0$$

and hence $\mathcal{D}$ is dense in $\mathcal{H}$. This $\mathcal{D}$ will be the domain of the T that we seek.

If $x \in \mathcal{D}$, then by definition of $\mathcal{D}$ and the assumed injectivity of $(1-U)$, there exists a unique $z \in \mathcal{R}$ such that $2ix = z - Uz$. Define $Tx = \frac{1}{2}(z + Uz)$.

We first verify that T is closed; so suppose $G(T) \ni (x_n, y_n) \to (x, y) \in \mathcal{H} \oplus \mathcal{H}$. Thus there exist $\{z_n\}_n \subset \mathcal{R}$ such that $2ix_n = z_n - Uz_n$, and $2y_n = 2Tx_n = z_n + Uz_n$ for all n. Then

$$z_n = (y_n + ix_n) \text{ and } Uz_n = y_n - ix_n \; ; \tag{5.2.10}$$

hence $z_n \to y + ix = z$ (say) and $Uz = lim\; Uz_n = y - ix$; since $\mathcal{R}$ is closed, it follows that $z \in \mathcal{R}$ and that $z - Uz = 2ix,\; z + Uz = 2y$; hence $x \in \mathcal{D}$ and $Tx = y$, thereby verifying that T is indeed closed.

With x, y, z as above, we find that (for arbitrary $x \in dom\; T$)

$$\begin{aligned} \langle Tx, x\rangle &= \langle \frac{1}{2}(z + Uz), \frac{1}{2i}(z - Uz)\rangle \\ &= \frac{-1}{4i}\left[||z||^2 - ||Uz||^2 + \langle Uz, z\rangle - \langle z, Uz\rangle\right] \\ &= \frac{-1}{4i} 2i\; Im\langle Uz, z\rangle \\ &\in \mathbb{R} \end{aligned}$$

and conclude - see Proposition 5.2.2(a)(iii) - that T is indeed symmetric. It is clear from the definitions that indeed $U = U_T$. □

We spell out some facts concerning what we have called the Cayley transform in the following remark.

Remark 5.2.8 (1) The proof of Proposition 5.2.7 actually establishes the following fact: let $\mathcal{T}$ denote the set of all (densely defined) closed symmetric operators 'from $\mathcal{H}$ to itself'; then the assignment

$$\mathcal{T} \ni T \mapsto U_T \in \mathcal{U} \tag{5.2.11}$$

defines a bijective correspondence from $\mathcal{T}$ to $\mathcal{U}$, where $\mathcal{U}$ is the collection of all partial isometries U on $\mathcal{H}$ with the following two properties: (i) $(U-1)$ is 1-1; and (ii) $ran\; (U-1)U^*U$ is dense in $\mathcal{H}$; further, the 'inverse transform' is the map

$$\mathcal{U} \ni U \mapsto T = -i(1+U)(1-U)^{(-1)} \; , \tag{5.2.12}$$

where we have written $(1-U)^{(-1)}$ to denote the inverse of the 1-1 map $(1-U)|_{ran(U^*U)}$ - thus, the domain of T is just $ran((U-1)U^*U)$.

Note that the condition (ii) (defining the class $\mathcal{U}$ in the last paragraph) is equivalent to the requirement that $ker(U^*U(U^*-1)) = \{0\}$; note that $U^*U(U^*-1) = U^* - U^*U = U^*(1-U)$ (since U^* is a partial isometry); hence, condition (ii) is equivalent to the condition that $ker(U^*(1-U)) = \{0\}$; and this condition clearly implies that $(1-U)$ must be 1-1. Thus, $(ii) \Rightarrow (i)$, and we may hence equivalently define

$$\mathcal{U} = \{U \in \mathcal{L}(\mathcal{H}) : U = UU^*U,\ ker(U^* - U^*U) = \{0\}\} \ .$$

(2) Suppose T is a self-adjoint operator; then it follows from Proposition 5.2.6 that the Cayley transform U_T is a unitary operator, and this is the way in which we shall derive the spectral theorem for (possibly unbounded) self-adjoint operators from the spectral theorem for unitary operators. $\square$

It is time we introduced the terminology that is customarily used to describe many of the objects that we have already encountered.

Definition 5.2.9 *Let T be a closed (densely defined) symmetric operator. The closed subspaces*

$$\begin{aligned} \mathcal{D}_+(T) &= ker(T^* - i) &= \{x \in dom\ T^* : T^*x = ix\} \\ \mathcal{D}_-(T) &= ker(T^* + i) &= \{x \in dom\ T^* : T^*x = -ix\} \end{aligned}$$

are called the (positive and negative) **deficiency spaces** *of T, and the (cardinal numbers)*

$$\delta_\pm(T) \ = \ dim\ \mathcal{D}_\pm(T)$$

are called the **deficiency indices** *of T.*

Thus, if U_T is the Cayley transform of the closed symmetric operator T, then $\mathcal{D}_+(T) = kerU_T$ and $\mathcal{D}_-(T) = kerU_T^*$.

Corollary 5.2.10 *Let $T \in \mathcal{T}$ (in the notation of Remark 5.2.8(1)).*

(a) If $T \subset T_1$ and if also $T_1 \in \mathcal{T}$, then $\mathcal{D}_\pm(T) \subset \mathcal{D}_\pm(T_1)$.

(b) A necessary and sufficient condition for T to be self-adjoint is that $\delta_+(T) = \delta_-(T) = 0$.

(c) A necessary and sufficient condition for T to admit a self-adjoint extension is that $\delta_+(T) = \delta_-(T)$.

Proof : Assertions (a) and (b) are immediate consequences of Proposition 5.2.7(b) and Proposition 5.2.6.

As for (c), if T_1 is a symmetric extension of T, then, by Proposition 5.2.7(b), it follows that $\mathcal{D}_\pm(T_1) \supset \mathcal{D}_\pm(T)$. In particular, if T_1 is self-adjoint, then the Cayley transform U of T_1 is a unitary operator which dominates the Cayley transform of T and consequently maps $\mathcal{D}_+^\perp$ onto $\mathcal{D}_-(T)^\perp$; this implies that $\delta_+(T) = \delta_-(T)$.

Conversely, suppose $\delta_+(T) = \delta_-(T)$. Then there exists a unitary operator U_1 such that $U_1|_{\mathcal{D}_+(T)^\perp} = U_T$. (Why?) If it so happens that 1 is an eigenvalue of U_1, define

$$Ux = \begin{cases} -x & if\ Ux = x \\ U_1 x & if\ x \in ker^\perp(U_1 - 1) \end{cases}$$

and note that U is also a unitary operator which 'dominates' U_T and which does not have 1 as an eigenvalue; so that U must be the Cayley transform of a self-adjoint extension of T. □

5.3 Spectral theorem and polar decomposition

In the following exercises, we will begin the process of applying the spectral theorem for bounded normal operators to construct some examples of unbounded closed densely defined normal operators.

Exercise 5.3.1 *(1) Let $\mathcal{H} = L^2(X, \mathcal{B}, \mu)$ and let $\mathcal{C}$ denote the collection of all measurable functions $\phi : X \to \mathbb{C}$; for $\phi \in \mathcal{C}$, let M_ϕ be as in Example 5.1.4(1). Show that if $\phi, \psi \in \mathcal{C}$ are arbitrary, then,*

(a) M_ϕ *is densely defined.*
(b) M_ϕ *is a closed operator.*
(c) $M_\phi^* = M_{\overline{\phi}}$.
(d) $M_{\alpha\phi+\psi} \supset \alpha M_\phi + M_\psi$, *for all* $\alpha \in \mathbb{C}$.
(e) $M_\phi M_\psi \supset M_{\phi\psi} \subset M_\psi M_\phi$.
(f) $M_\phi M_{\overline{\phi}} = M_{\overline{\phi}} M_\phi$.
(Hint: (a) let $E_n = \{x \in X : |\phi(x)| \leq n\}$ *and let* $P_n = M_{1_{E_n}}$*; show that* $X = \cup_n E_n$*, hence the sequence* $\{P_n\}_n$ *converges strongly to 1, and note that ran* $P_n = \{f \in L^2(\mu) : f = 0$ *a.e. outside* $E_n\} \subset$ *dom* M_ϕ*; (b) if* $f_n \to f$ *in* $L^2(\mu)$*, then there exists a subsequence* $\{f_{n_k}\}_k$ *which converges to* f *a.e. (c) First note that dom* $M_{\overline{\phi}} =$ *dom* M_ϕ *and deduce from Exercise 5.1.5(2)(a) that* $M_{\overline{\phi}} \subset M_\phi^*$*; then observe that if* E_n, P_n *are as above, then ran* $P_n \subset$ *dom*M_ϕ^* *and in fact* $P_n M_\phi^* \subset M_\phi^* P_n = M_{\overline{\phi}1_{E_n}} \in \mathcal{L}(\mathcal{H})$*; i.e., if* $g \in$ *dom* M_ϕ^**, and* n *is arbitrary, then,* $P_n M_\phi^* g = \overline{\phi} 1_{E_n} g$ *a.e.; deduce (from the fact that* M_ϕ^* *is closed) that* $M_\phi^* g = \overline{\phi} g$*; i.e.,* $M_\phi^* \subset M_{\overline{\phi}}$*. (d) - (f) are consequences of the definition.)*

(2) Let $S_n : \mathcal{D}_n\ (\subset \mathcal{H}_n) \to \mathcal{K}_n$ *be a linear operator, for* $n = 1, 2, \cdots$*. Define* $S : \mathcal{D}\ (\subset \oplus_n \mathcal{H}_n) \to \oplus_n \mathcal{K}_n$ *by* $(Sx)_n = S_n x_n$*, where* $\mathcal{D} = \{x = ((x_n))_n : x_n \in \mathcal{D}_n\ \forall n,\ \sum_n ||S_n x_n||^2 < \infty\}$*. Then show that:*
(a) S *is a linear operator;*
(b) S *is densely defined if and only if each* S_n *is;*
(c) S *is closed if and only if each* S_n *is.*
(Hint: (a) is clear, as are the 'only if' statements in (b) and (c); if $\mathcal{D}_n$ *is dense in* $\mathcal{H}_n$ *for each* n*, then the set* $\{((x_n)) \in \oplus_n \mathcal{H}_n : x_n \in \mathcal{D}_n\ \forall n$ *and* $x_n = 0$ *for all but finitely many* $n\}$ *is (a subset of* $\mathcal{D}$ *which is) dense in* $\oplus_n \mathcal{H}_n$*; as for (c), if* $\mathcal{D} \ni x(k) \to x \in \oplus_n \mathcal{H}_n$ *and* $Sx(k) \to y \in \oplus_n \mathcal{K}_n$*, where* $x(k) = ((x(k)_n))$*,* $x = ((x_n))$ *and* $y = ((y_n))$*, then we see that* $x(k)_n \in$ *dom* $S_n\ \forall n$*,* $x(k)_n \to x_n$ *and* $S_n x(k)_n \to y_n$ *for all* n*; so the hypothesis that each* S_n *is closed would then imply that* $x_n \in$ *dom* S_n *and* $S_n x_n = y_n\ \forall\ n$*; i.e.,* $x \in$ *dom* S *and* $Sx = y$*.)*

(3) If S_n, S *are the operators in (2) above, with domains specified therein, let us use the notation* $S = \oplus_n S_n$*. If* S *is densely defined, show that* $S^* = \oplus_n\ S_n^*$*.*

(4) Let X be a locally compact Hausdorff space and let $\mathcal{B}_X \ni E \mapsto P(E)$ denote a separable spectral measure on X. To each measurable function $\phi : X \to \mathbb{C}$, we wish to construct a closed operator

$$\int_X \phi dP \ = \ \int_X \phi(\lambda) dP(\lambda) \tag{5.3.13}$$

satisfying the following constraints:

(i) if $P = P_\mu$, then $\int_X \phi dP$ agrees with the operator we denoted by M_ϕ in (1) above; and

(ii) $\int_X \phi d(\oplus_i P_i) \ = \ \oplus_i \int_X \phi dP_i$, for any countable direct sum of spectral measures, where the direct sum is interpreted as in (2) above.

(a) Show that there exists a unique way to define an assignment $\mathcal{C} \times \mathcal{P} \ni (\phi, P) \mapsto \int_X \phi dP \in \mathcal{L}_c(\mathcal{H})$ - where $\mathcal{P}$ denotes the class of all separable spectral measures on $(X, \mathcal{B})$ - in such a way that the requirements (i) and (ii) above, are satisfied. (Hint: by the Hahn-Hellinger theorem - see Remark 3.5.13 - there exist a probability measure $\mu : \mathcal{B}_X \to [0,1]$ and a partition $\{E_n : 0 \le n \le \aleph_0\} \subset \mathcal{B}_X$ such that $P \cong \oplus_{0 \le n \le \aleph_0} P^n_{\mu_{E_n}}$; appeal to (1)(a),(b) and (2)(a),(b) above.)

(b) (An alternative description of the operator $\int_X \phi dP$ - constructed in (a) above - is furnished by this exercise.)

Let $\phi : X \to \mathbb{C}$ be a measurable function and let $P : \mathcal{B}_X \to \mathcal{L}(\mathcal{H})$ be a separable spectral measure; set $E_n \ = \ \{x \in X : |\phi(x)| < n\}$ and $P_n = P(E_n)$. Then show that

(i) $\phi 1_{E_n} \in L^\infty(X, \mathcal{B}, \mu)$ and $T_n = \int_X \phi 1_{E_n} dP$ is a bounded normal operator, for each n;

(ii) in fact, if we set $T = \int_X \phi dP$, then show that $T_n = TP_n$;

(iii) the following conditions on a vector $x \in \mathcal{H}$ are equivalent:

(α) $\sup_n \ ||T_n x|| < \infty$;

(β) $\{T_n x\}_n$ is a Cauchy sequence;

(iv) if we define $\mathcal{D}^{(P)}_\phi$ to be the set of those vectors in $\mathcal{H}$ which satisfy the equivalent conditions of (iii), then $\mathcal{D}^{(P)}_\phi$ is precisely equal to dom $(\int_X \phi dP)$ and further,

$$(\int_X \phi dP)x \ = \ \lim_n T_n x \ \forall \ x \in \mathcal{D}^{(P)}_\phi .$$

(v) Show that

$$(\int_X \phi dP)^* = \int_X \overline{\phi} dP$$

and that

$$(\int_X \overline{\phi} dP) (\int_X \phi dP) = (\int_X \phi dP) (\int_X \overline{\phi} dP) .$$

(Hint: Appeal to (1)(c) and (3) of these exercises for the assertion about the adjoint; for the other, appeal to (1)(f) and (2).)

REMARK 5.3.2 (1) The operator $\int_X \phi dP$ constructed in the preceding exercises is an example of an unbounded normal operator, meaning a densely defined operator T which satisfies the condition $TT^* = T^*T$ - where, of course, this equality of two unbounded operators is interpreted as the equality of their graphs.

(2) Thus, for example, if T is a bounded normal operator on a separable Hilbert space $\mathcal{H}$, and if $P_T(E) = 1_E(T)$ for $E \in \mathcal{B}_{\sigma(T)}$, and if $\phi : \sigma(T) \to \mathbb{C}$ is any measurable function, then we shall define (the 'unbounded functional calculus' for T by considering) the (in general unbounded) operator $\phi(T)$ by the rule

$$\phi(T) = \int_{\sigma(T)} \phi dP_T . \tag{5.3.14}$$

It should be noted that it is not strictly necessary that the function ϕ be defined everywhere; it suffices that it be defined P a.e.; (i.e., in the complement of some set N such that $P(N) = 0$). Thus, for instance, so long as λ_0 is not an eigenvalue of T, we may talk of the operator $\phi(T)$ where $\phi(\lambda) = (\lambda - \lambda_0)^{-1}$ - this follows easily from Exercise 4.3.8.) □

We are now ready for the spectral theorem.

THEOREM 5.3.3 *(***Spectral Theorem for unbounded self-adjoint operators***)*

(a) If T is a self-adjoint operator on $\mathcal{H}$, then there exists a spectral measure (or equivalently, a projection-valued strongly continuous mapping) $\mathcal{B}_{\mathbb{R}} \ni E \mapsto P(E) \in \mathcal{L}(\mathcal{H})$ such that

$$T = \int_{\mathbb{R}} \lambda \, dP(\lambda) \tag{5.3.15}$$

in the sense of equation 5.3.14.

(b) Furthermore, the spectral measure associated with a self-adjoint operator is unique in the following sense: if T_i is a self-adjoint operator on $\mathcal{H}_i$ with associated spectral measures $P_i : \mathcal{B}_{\mathbb{R}} \to \mathcal{L}(\mathcal{H})$, for $i = 1, 2$, then the following conditions on a unitary operator $W : \mathcal{H}_1 \to \mathcal{H}_2$ are equivalent:

(i) $T_2 = WT_1W^*$;

(ii) $P_2(E) = WP_1(E)W^* \ \forall\ E \in \mathcal{B}_{\mathbb{R}}$.

Proof : (a) If $U = U_T$ is the Cayley-transform of A, then U is a unitary operator on $\mathcal{H}$; let $P_U : \mathcal{B}_{\sigma(U)} \to \mathcal{L}(\mathcal{H})$ be the associated spectral measure (defined by $P_U(E) = 1_E(U)$).

The functions

$$\mathrm{R} \ni t \stackrel{f}{\mapsto} \frac{t-i}{t+i} \quad , \quad (\mathbb{T} - \{1\}) \ni z \stackrel{g}{\mapsto} i\left(\frac{1+z}{1-z}\right) \ ,$$

are easily verified to be homeomorphisms which are inverses of one another.

Our earlier discussion on Cayley transforms shows that $U-1$ is 1-1 and hence $P_U(\{1\}) = 0$; hence $g \circ f = id_{\sigma(U)}$ (P_U *a.e.*) Our construction of the Cayley transform also shows that $T = g(U)$ in the sense of the unbounded functional calculus - see 5.3.14; it is a routine matter to now verify that equation 5.3.15 is indeed satisfied if we define the spectral measure $P : \mathcal{B}_{\mathbb{R}} \to \mathcal{L}(\mathcal{H})$ by the prescription $P(E) = P_U(f(E))$.

(b) This is an easy consequence of the fact that if we let $P_n^{(i)} = P_i([n-1, n))$ and $T_n^{(i)} = T_i|_{ranP_n}$, then $T_i = \oplus_{n \in \mathbb{Z}} T_n^{(i)}$. $\square$

We now introduce the notion, due to von Neumann, of an unbounded operator being **affiliated** to a von Neumann algebra. This relies on the notion - already discussed in Exercise 5.1.10(c) - of a bounded operator commuting with an unbounded one .

Proposition 5.3.4 *Let $M \subset \mathcal{L}(\mathcal{H})$ be a von Neumann algebra; then the following conditions on a linear operator T 'from $\mathcal{H}$ into itself' are equivalent:*

(i) $A'T \subset TA'$, $\forall\ A' \in M'$;

(ii) $U'T = TU'$, $\forall$ unitary $U' \in M'$;

(iii) $U'TU'^* = T$, $\forall$ unitary $U' \in M'$.

If it is further true that if $T = T^*$, then the preceding three conditions are all equivalent to the following fourth condition:

(iv) $P_T(E) \in M \; \forall \; E \in \mathcal{B}_{\mathbb{R}}$, where P_T denotes the unique spectral measure associated with T, as in Theorem 5.3.3.

If T and M are as above, then we shall say that T is affiliated with M and we shall write $T\eta M$.

Proof : $(i) \Rightarrow (ii)$: The hypothesis (i) amounts to the requirement that if $x \in dom\ T$, and if $A' \in M'$ is arbitrary, then $A'x \in \ dom\ T$ and $TA'x = A'Tx$. In particular, if U' is any unitary element of M', then $U'x, U'^*x \in \ dom\ T$ and $TU'x = U'Tx$ and $TU'^*x = U'^*Tx$. Since U' is a bijective map, it follows easily from this that we actually have $TU' = U'T$.

The implication $(ii) \Leftrightarrow (iii)$ is an easy exercise, while the implication $(ii) \Rightarrow (i)$ is a consequence of the simple fact that every element $A' \in M'$ is expressible as a linear combination of (no more that four) unitary elements of M'. (*Proof of the above fact:* by looking at the Cartesian decomposition, and using the fact that M is closed under taking 'real' and 'imaginary' parts, it is seen that it suffices to prove that any self-adjoint element of a C^*-algebra is a linear combination of two unitary elements; but this is true because, if A is self-adjoint and if $||A|| \leq 1$, then we may write $A = \frac{1}{2}(U_+ + U_-)$, where $U_\pm = g_\pm(A)$, and $g_\pm(t) \; = \; t \pm i\sqrt{1-t^2}$.)

Finally the implication $(iii) \Leftrightarrow (iv)$ - when T is self-adjoint - is a special case of Theorem 5.3.3(b) (applied to the case $T_i = T, i = 1, 2$.) $\square$

It should be noted that if T is a (possibly unbounded) self-adjoint operator on $\mathcal{H}$, if $\phi \mapsto \phi(T)$ denotes the associated 'unbounded functional calculus', and if $M \subset \mathcal{L}(\mathcal{H})$ is a von Neumann algebra of bounded operators on $\mathcal{H}$, then $T\eta M \;\Rightarrow \phi(T)\eta M$ for every measurable function $\phi : \mathbb{R} \to \mathbb{C}$; also, it should be clear - from the double commutant theorem, for instance - that if $T \in \mathcal{L}(\mathcal{H})$ is an everywhere defined bounded operator, then $T\eta M \;\Leftrightarrow T \in M$.

We collect some elementary properties of this notion of affiliation in the following exercises.

EXERCISE 5.3.5 *(1) Let* S, T *denote linear operators 'from* $\mathcal{H}$ *into itself', and let* $M \subset \mathcal{L}(\mathcal{H})$ *be a von Neumann algebra such that* $S, T \eta M$*. Show that*

(i) if T *is closable, then* $\overline{T} \eta M$*;*

(ii) if T *is densely defined, then also* $T^* \eta M$*;*

(iii) $ST \eta M$*.*

(2) Given any family $\mathcal{S}$ *of linear operators 'from* $\mathcal{H}$ *into itself', let* $\mathcal{W}_{\mathcal{S}}$ *denote the collection of all von Neumann algebras* $M \subset \mathcal{L}(\mathcal{H})$ *with the property that* $T \eta M \ \forall \ T \in \mathcal{S}$*; and define*

$$W^*(\mathcal{S}) \ = \ \cap \ \{M : M \in \mathcal{W}_{\mathcal{S}}\} \ .$$

Then show that:

(i) $W^*(\mathcal{S})$ *is the smallest von-Neumann subalgebra* M *of* $\mathcal{L}(\mathcal{H})$ *with the property that* $T \eta M \ \forall \ T \in \mathcal{S}$*;*

(ii) if $\mathcal{S} \subset \mathcal{T} \subset \mathcal{L}_c(\mathcal{H})$*, then*

$$W^*(\mathcal{S}) \subset W^*(\mathcal{T}) \ = W^*(\mathcal{T}^*) \ = \ (\mathcal{T} \cup \mathcal{T}^*)'' \ ,$$

where, of course, $\mathcal{T}^* \ = \ \{T^* : T \eta \mathcal{T}\}$*, and* $\mathcal{S}' \ = \ \{A \in \mathcal{L}(\mathcal{H}) : AT \subset TA \ \forall \ T \in \mathcal{S}\}$ *(as in Exercise 5.1.10(d)).*

(3) Let $(X, \mathcal{B})$ *be any measurable space and suppose* $\mathcal{B} \ni E \ \mapsto \ P(E) \in \mathcal{L}(\mathcal{H})$ *is a countably additive projection-valued spectral measure (taking values in projection operators in* $\mathcal{L}(\mathcal{H})$*), where* $\mathcal{H}$ *is assumed to be a separable Hilbert space with orthonormal basis* $\{e_n\}_{n \in \mathbb{N}}$*.*

For $x, y \in \mathcal{H}$*, let* $p_{x,y} : \mathcal{B} \to \mathbb{C}$ *be the complex measure defined by* $p_{x,y}(E) \ = \ \langle P(E)x, y \rangle$*. Let us write* $p_\alpha \ = \ \sum_n \alpha_n p_{e_n, e_n}$*, when either* $\alpha = ((\alpha_n))_n \in \ell^1$ *or* $\alpha \in \mathbb{R}_+^{\mathbb{N}}$ *(so that* p_α *makes sense as a finite complex measure or a (possibly infinite-valued, but)* σ*-finite positive measure defined on* $(X, \mathcal{B})$*.*

(a) Show that the following conditions on a set $E \in \mathcal{B}$ *are equivalent:*

(i) $P(E) = 0$*;*

(ii) $p_\alpha(E) = 0$ *for some* $\alpha = ((\alpha_n))_n$ *where* $\alpha_n > 0 \ \forall \ n \in \mathbb{N}$*;*

(iii) $p_\alpha(E) = 0$ *for every* $\alpha = ((\alpha_n))_n$ *where* $\alpha_n > 0 \ \forall \ n \in \mathbb{N}$*.*

(b) Show that the collection $\mathcal{N}_P = \{E \in \mathcal{B} : P(B) = 0\}$ *is a* σ*-ideal in* $\mathcal{B}$ *- meaning that if* $\{N_n : n \in \mathbb{N}\} \subset \mathcal{N}_P$ *and* $E \in \mathcal{B}$ *are arbitrary, then* $\cup_{n \in \mathbb{N}} N_n \in \mathcal{N}_P$ *and* $N_1 \cap E \in \mathcal{N}_P$*.*

(c) Let $\mathcal{L}^0(X, \mathcal{B})$ denote the algebra of all $(\mathcal{B}, \mathcal{B}_{\mathbb{C}})$-measurable complex-valued functions on X. Let Z_P denote the set of functions $h \in \mathcal{L}^0(X, \mathcal{B})$ such that $h = 0 \; P - a.e.$ (meaning, of course, that $P(\{x : h(x) \neq 0\}) = 0$). Verify that Z_P is an ideal in $\mathcal{L}^0(X, \mathcal{B})$ which is closed under pointwise convergence of sequences; in fact, show more generally, that if $\{f_n\}$ is a sequence in Z_P and if $f \in \mathcal{L}^0(X, \mathcal{B})$ is such that $f_n \to f \; P - a.e.$, then $f \in Z_P$.

(d) With the notation of (c) above, define the quotient spaces

$$\begin{aligned} L^0(X, \mathcal{B}, P) &= \mathcal{L}^0(X, \mathcal{B})/Z_P && (5.3.16) \\ L^\infty(X, \mathcal{B}, P) &= \mathcal{L}^\infty(X, \mathcal{B})/(\mathcal{L}^\infty(X, \mathcal{B}) \cap Z_P) && (5.3.17) \end{aligned}$$

where we write $\mathcal{L}^\infty(X, \mathcal{B})$ to denote the class of all bounded measuable complex-valued functions on X. Verify that $L^\infty(X, \mathcal{B}, P)$ is a Banach space with respect to the norm defined by

$$||f||_{L^\infty(X,\mathcal{B},P)} = \inf\{sup\{|f(x)| : x \in E\} : (X - E) \in \mathcal{N}_P\} .$$

(4) Let T be a self-adjoint operator on $\mathcal{H}$, and let $P_T : \mathcal{B}_{\mathbb{R}} \to \mathcal{L}(\mathcal{H})$ be the associated spectral measure as in Theorem 5.3.3. Show that:

(i) we have a **bounded functional calculus** *for T; i.e., there exists a unique *-algebra isomorphism $L^\infty(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, P_T) \ni \phi \mapsto \phi(T) \in W^*(\{T\})$ with the property that $1_E(T) \; (= P_T(E))$ for all $E \in \mathcal{B}_{\mathbb{R}}$; and*

(ii) if $S \in \mathcal{L}_c(\mathcal{H})$ is such that $S \eta W^*(\{T\})$, then there exists a $\phi \in L^0(X, \mathcal{B}_X, P)$ such that $S = \phi(T)$ in the sense that $S = \int \phi(\lambda) dP(\lambda)$.

Before we can get to the polar decomposition for unbounded operators, we should first identify just what we should mean by an unbounded self-adjoint operator being positive.

We begin with a lemma.

LEMMA 5.3.6 *Let $T \in \mathcal{L}_c(\mathcal{H}, \mathcal{K})$. Then,*

(a) for arbitrary $z \in \mathcal{H}$, there exists a unique $x \in dom(T^*T)$ such that $z = x + T^*Tx$;

(b) in fact, there exists an everywhere defined bounded operator $A \in \mathcal{L}(\mathcal{H})$ such that

(i) A is a positive 1-1 operator satisfying $||A|| \leq 1$; and

(ii) $(1 + T^*T)^{-1} = A$, meaning that ran A = $dom(1 + T^*T) = \{x \in dom\ T : Tx \in dom\ T^*\}$ and $Az + T^*T(Az) = z \ \forall z \in \mathcal{H}$.

Proof : Temporarily fix $z \in \mathcal{H}$; then, by Proposition 5.1.6, there exists unique elements $x \in dom(T)$ and $y \in dom(T^*)$ such that

$$(0, z) \ = \ (y, T^*y) \ - \ (Tx, -x) \ .$$

Thus, $z = T^*y + x$, where $y = Tx$; i.e., $x \in dom(T^*T)$ and $z = (1 + T^*T)x$, thereby establishing (a).

With the preceding notation, define $Az = x$, so that $A : \mathcal{H} \to \mathcal{H}$ is an everywhere defined (clearly) linear operator satisfying $(1 + T^*T)A = 1$, so that A is necessarily 1-1. Notice that

$$\langle Az, z \rangle \ = \ \langle x, z \rangle \ = \ ||x||^2 + ||Tx||^2 \ , \tag{5.3.18}$$

and hence $||Az||^2 = ||x||^2 \leq \langle Az, z \rangle \leq ||Az|| \ ||z||$. In other words, A is indeed a bounded positive contraction; i.e., $0 \leq A \leq 1$. Hence (b)(i) is true, and (b)(ii) is a consequence of the construction. $\square$

PROPOSITION 5.3.7 *(1) The following conditions on a closed densely defined operator $S \in \mathcal{L}_c(\mathcal{H})$ are equivalent:*

(i) S is self-adjoint and $\langle Sx, x \rangle \geq 0 \ \forall \ x \in dom\ S$;

(i)$'$ S is self-adjoint and $\sigma(S) \subset [0, \infty)$, where of course $\sigma(S)$ is called the spectrum of S, and $\rho(S) = \mathbb{C} - \sigma(S)$ is the resolvent set of S which consists of those complex scalars λ for which $(S - \lambda)^{-1} \in \mathcal{L}(\mathcal{H})$; (thus, $\lambda \notin \sigma(S)$ precisely when $(S - \lambda)$ maps dom S 1-1 onto all of $\mathcal{H}$;)

(ii) there exists a self-adjoint operator T acting in $\mathcal{H}$ such that $S = T^2$;

(iii) there exists $T \in \mathcal{L}_c(\mathcal{H}, \mathcal{K})$ (for some Hilbert space $\mathcal{K}$) such that $S = T^*T$.

A linear operator $S \in \mathcal{L}_c(\mathcal{H})$ is said to be a **positive self-adjoint** *operator if the three preceding equivalent conditions are met.*

(2) Furthermore, a positive operator has a unique positive square root - meaning that if S is a positive self-adjoint operator

acting in $\mathcal{H}$, then there exists a unique positive self-adjoint operator acting in $\mathcal{H}$ - which is usually denoted by the symbol $S^{\frac{1}{2}}$ - such that $S = (S^{\frac{1}{2}})^2$.

Proof: $(i) \Leftrightarrow (i)'$: If S is self-adjoint, let $P : \mathcal{B}_{\mathbb{R}}\mathcal{L}(\mathcal{H})$ be the (projection-valued) spectral measure asdsociated with S as in Theorem 5.3.3; as usual, if $x \in \mathcal{H}$, let $p_{x,x}$ denote the (genuine scalar-valued finite) positive measure defined by $p_{x,x}(E) = \langle P(E)x, x\rangle = ||P(E)x||^2$. Given that $S = \int_{\mathbb{R}} \lambda dP(\lambda)$, it is easy to see that the (second part of) condition (i) (resp., $(i)'$) amounts to saying that (a) $\int_{\mathbb{R}} \lambda dp_{x,x}(\lambda) > 0$ for all $x \in \mathcal{H}$ (resp., $(a)'$ P is supported in $[0, \infty)$ - meaning that $P((-\infty, 0)) = 0$); the proof of the equivalence $(i) \Leftrightarrow (i)'$ reduces now to the verification that conditions (a) and $(a)'$ are equivalent; but this is an easy verification.

$(i) \Rightarrow (ii)$: According to the hypothesis, we are given that $S = \int_{\mathbb{R}} \lambda \, dP(\lambda)$, where $P : \mathcal{B}_{\mathbb{R}} \to \mathcal{L}(\mathcal{H})$ is some spectral measure which is actually supported on the positive line $[0, \infty)$. Now, let $T = \int_{\mathbb{R}} \lambda^{\frac{1}{2}} \, dP(\lambda)$, and deduce (from Exercise 5.3.1(1)(c)) that T is indeed a self-adjoint operator and that $S = T^2$.

$(ii) \Rightarrow (iii)$: Obvious.

$(iii) \Rightarrow (i)$: It is seen from Lemma 5.3.6 that $(1+T^*T)^{-1} = A$ is a positive bounded operator A (which is 1-1). It follows that $\sigma(A) \subset [0, 1]$ and that $A = \int_{[0,1]} \lambda dP_A(\lambda)$, where $P_A : \mathcal{B}_{[0,1]} \to \mathcal{L}(\mathcal{H})$ is the spectral measure given by $P_A(E) = 1_E(A)$. Now it follows readily that T^*T is the self-adjoint operator given by $S = T^*T = \int_{[0,1]} (\frac{1}{\lambda} - 1) \, dP_A(\lambda)$; finally, since $(\frac{1}{\lambda} - 1) > 0$ $P - a.e.$, we find also that $\langle Sx, x\rangle \geq 0$.

(2) Suppose that S is a positive operator; and suppose T is some positive self-adjoint operator such that $S = T^2$. Let $P = P_T$ be the spectral measure associated with T, so that $T = \int_{\mathbb{R}} \lambda \, dP(\lambda)$. Define $M = W^*(\{T\})$ ($\cong L^\infty(\mathrm{R}, \mathcal{B}_{\mathbb{R}}, P)$); then $T\eta M$ by definition, and so - see Exercise 5.3.5(1)(iii) - also $S\eta M$. It follows - from Proposition 5.3.4 - that

$$S\eta M \;\Rightarrow\; 1_S(E) \in M \;\forall\; E \in \mathcal{B}_{\mathbb{R}^+} \;.$$

This implies that $W^*(\{S\}) = \{1_S(E) : E \in \mathcal{B}_{\mathbb{R}}\}'' \subset M$. If we use the symbol $S^{\frac{1}{2}}$ to denote the positive square root of S that was constructed in the proof of $(i) \Rightarrow (ii)$ above, note that

$S^{\frac{1}{2}}\ \eta\ W^*(\{S\}) \subset M$; let $\phi \in L^\infty(\mathbb{R}, \mathcal{B}, P)$ be such that $S = \phi(T)$; the properties of the operators concerned now imply that ϕ is a measurable function which is (i) non-negative P-a.e., and (ii) $\phi(\lambda)^2 = \lambda^2$ P-a.e.; it follows that $\phi(\lambda) = \lambda\ P - a.e.$, thus establishing the uniqueness of the positive square root. □

THEOREM 5.3.8 *Let $T \in \mathcal{L}_c(\mathcal{H}, \mathcal{K})$. Then there exists a unique pair $U, |T|$ of operators satisfying the following conditions:*
(i) $T = U|T|$;
(ii) $U \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ is a partial isometry;
(iii) $|T|$ is a positive self-adjoint operator acting in $\mathcal{H}$; and
(iv) ker T = ker U = ker $|T|$.

Furthermore, $|T|$ is the unique positive square root of the positive self-adjoint operator T^*T.

Proof: : As for the uniqueness assertion, first deduce from condition (iv) that $U^*U|T| = |T|$, and then deduce that $|T|$ has to be the unique positive square root of T^*T; the uniqueness of the U in the polar decomposition is proved, exactly as in the bounded case.

Let $P_n = 1_{T^*T}([0, n])$ and let $\mathcal{H}_n = ran\ P_n$. Then $\{P_n\}$ is an increasing sequence of projections such that $P_n \to 1$ strongly.

Observe also that $\mathcal{H}_n \subset dom\ \phi(T^*T)$ for every continuous (and more generally, any 'locally bounded measurable) function $\phi : \mathbb{R} \to \mathbb{C}$, and that, in particular,

$$x \in \mathcal{H}_n \Rightarrow x \in dom\ T^*T \text{ and } ||Tx|| = ||\ |T|\ x||\ . \quad (5.3.19)$$

(Reason: $||Tx||^2 = \langle Tx, Tx\rangle = \langle |T|^2 x, x\rangle = ||\ |T|x||^2$.)

An appeal to Exercise 3.4.12 now ensures the existence of an isometric operator $U_0 : \overline{ran\ |T|} \to \mathcal{K}$ such that $U_0(|T|x) = Tx$ for all $x \in \cup_n \mathcal{H}_n$. Let $U = U_0 \circ 1_{(0,\infty)}(|T|)$. Then it is clear that U is a partial isometry with initial space $ker^\perp|T|$, and that $U|T|x = Tx$ for all $x \in \cup_n \mathcal{H}_n$.

We now claim that the following conditions on a vector $x \in \mathcal{H}$ are equivalent:
(i) $x \in dom\ T$;
(ii) $\sup_n ||TP_n x|| < \infty$;
(ii)$'$ $\sup_n ||\ |T|P_n x|| < \infty$;

(i)′ $x \in \; dom(|T|)$.

The equivalence $(ii) \Leftrightarrow (ii)'$ is an immediate consequence of equation (5.3.19); then for any $x \in \mathcal{H}$ and $n \in \mathbb{N}$, we see that $|T|P_n x$ is the sum of the family $\{|T|(P_j - P_{j-1}) : 1 \leq j \leq n\}$ of mutually orthogonal vectors - where we write $P_0 = 0$ - and hence condition $(ii)'$ is seen to be equivalent to the convergence of the (orthogonal, as well as collapsing) sum

$$\sum_{j=1}^{\infty} |T|(P_j - P_{j-1})x \;=\; lim_n |T|P_n x \;.$$

Since $P_n x \to x$, the fact that $|T|$ is closed shows that the convergence of the displayed limit above would imply that $x \in \; dom\ |T|$ and that $|T|x = \; lim_n\ |T|P_n x$, hence $(ii)' \;\Rightarrow (i)'$.

On the other hand, if $x \in dom\ (|T|)$, then since $|T|$ and P_n commute, we have:

$$|T|x \;=\; \lim_n \; P_n|T|x \;=\; \lim_n |T|P_n x \;,$$

and the sequence $\{||\ (|T|P_n x)||\}_n$ is convergent and necessarily bounded, thus $(i)' \;\Rightarrow (ii)'$.

We just saw that if $x \in \; dom\ |T|$, then $\{|T|P_n x\}_n$ is a convergent sequence; the fact that the partial isometry U has initial space given by $\overline{\cup_n \mathcal{H}_n}$ implies that also $\{TP_n x \;=\; U(|T|P_n x)\}_n$ is convergent; since T is closed, this implies $x \in \; dom\ T$, and so $(i)' \Rightarrow (i)$.

We claim that the implication $(i) \Rightarrow (i)'$ would be complete once we prove the following assertion: if $x \in dom\ T$, then there exists a sequence $\{x_k\}_k \;\subset\; \cup_n\ \mathcal{H}_n$ such that $x_k \to x$ and $Tx_k \to Tx$. (Reason: if this assertion is true, if $x \in dom\ T$, and if $x_k \in \cup_n \mathcal{H}_n$ is as in the assertion, then $|||T|x_k - |T|x_l|| = ||Tx_k - Tx_l||$ (by equation (5.3.19)) and hence also $\{|T|x_k\}_k$ is a (Cauchy, hence convergent) sequence; and finally, since $|T|$ is closed, this would ensure that $x \in \; dom|T|$.)

In order for the assertion of the previous paragraph to be false, it must be the case that the graph $G(T)$ properly contains the closure of the graph of the restriction T_0 of T to $\cup_n \mathcal{H}_n$; equivalently, there must be a non-zero vector $y \in \; dom\ T$ such that (y, Ty) is orthogonal to (x, Tx) for every $x \in \cup_n \mathcal{H}_n$; i.e.,

$$\begin{aligned} x \in \cup_n \mathcal{H}_n \;\; & \Rightarrow \;\; \langle x, y \rangle \; + \; \langle Tx, Ty \rangle \; = \; 0 \\ & \Rightarrow \;\; \langle (1 + T^*T)x, y \rangle \; = \; 0 \; ; \end{aligned}$$

hence $y \in ((1 + T^*T)(\cup_n \mathcal{H}_n))^\perp$.

On the other hand, if $z \in$ *dom* T^*T $=$ *dom* $|T|^2$, notice that

$$(1 + T^*T)z \; = \; \lim_n P_n(1 + |T|^2)z \; = \; \lim_n (1 + |T|^2)P_n z$$

and hence $(ran(1 + T^*T))^\perp \; = \; ((1 + T^*T)(\cup_n \mathcal{H}_n))^\perp$.

Thus, we find that in order for the implication $(i) \Rightarrow (i)'$ to be false, we should be able to find a non-zero vector y such that $y \in (ran(1 + T^*T))^\perp$; but by Lemma 5.3.6(a), there is no such non-zero vector y. Thus, we have indeed completed the proof of the equation *dom* T $=$ *dom* $|T|$. Finally since the equation $T = U|T|$ has already been verified on *dom*$|T|$, the proof of the existence half of the polar decomposition is also complete.

□

Appendix A

Appendix

A.1 Some linear algebra

In this section, we quickly go through some basic linear algebra - i.e., the study of finite-dimensional vector spaces. Although we have restricted ourselves to complex vector spaces in the text, it might be fruitful to discuss vector spaces over general fields in this appendix. We begin by recalling the definition of a field.

DEFINITION A.1.1 *A* **field** *is a set - which we will usually denote by the symbol* $\mathbb{K}$ *- which is equipped with two binary operations called addition and multiplication, respectively, such that the following conditions are satisfied:*

(1) (Addition axioms) There exsits a map $\mathbb{K} \times \mathbb{K} \ni (\alpha, \beta) \mapsto (\alpha + \beta) \in \mathbb{K}$*, such that the following conditions hold, for all* $\alpha, \beta, \gamma \in \mathbb{K}$*:*

(i) (commutativity) $\alpha + \beta = \beta + \alpha$*;*

(ii) (associativity) $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$*;*

(iii) (zero) there exists an element in $\mathbb{K}$*, always denoted simply by 0, such that* $\alpha + 0 = \alpha$*;*

(iv) (negatives) there exists an element in $\mathbb{K}$*, always denoted by* $-\alpha$*, with the property that* $\alpha + (-\alpha) = 0$.

(2) (Multiplication axioms) There exsits a map $\mathbb{K} \times \mathbb{K} \ni (\alpha, \beta) \mapsto (\alpha\beta) \in \mathbb{K}$*, such that the following conditions hold, for all* $\alpha, \beta, \gamma \in \mathbb{K}$*:*

(i) (commutativity) $\alpha\beta = \beta\alpha$*;*

(ii) (associativity) $(\alpha\beta)\gamma = \alpha(\beta\gamma)$;

(iii) (one) there exists an element in $\mathbb{K}$*, always denoted simply by 1, such that* $1 \neq 0$ *and* $\alpha 1 = \alpha$;

(iv) (inverses) if $\alpha \neq 0$*, there exists an element in* $\mathbb{K}$*, always denoted by* α^{-1}*, with the property that* $\alpha\alpha^{-1} = 1$.

(3) (Distributive law) Addition and multiplication are related by the following axiom, valid for all $\alpha, \beta, \gamma \in \mathbb{K}$:
$\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$.

Some simple properties of a field are listed in the following exercises.

Exercise A.1.2 *Let* $\mathbb{K}$ *be a field.*

(1) Show that if $\alpha, \beta, \gamma \in \mathbb{K}$ *are such that* $\alpha + \beta = \alpha + \gamma$*, then* $\beta = \gamma$*; deduce, in particular, that the additive identity 0 is unique, and that the additive inverse* $-\alpha$ *is uniquely determined by* α.

(2) Show that if $\alpha, \beta, \gamma \in \mathbb{K}$ *are such that* $\alpha\beta = \alpha\gamma$*, and if* $\alpha \neq 0$*, then* $\beta = \gamma$*; deduce, in particular, that the multiplicative identity 1 is unique, and that the multiplicative inverse* α^{-1} *of a non-zero element* α *is uniquely determined by* α.

(3) Show that $\alpha \cdot 0 = 0$ *for all* $\alpha \in \mathbb{K}$*; and conversely, show that if* $\alpha\beta = 0$ *for some* $\beta \neq 0$*, show that* $\alpha = 0$*. (Thus, a field has no 'zero divisors'.)*

(4) Prove that if $\alpha_1, \cdots, \alpha_n \in \mathbb{K}$*, and if* $\pi \in S_n$ *is any permutation of* $\{1, \cdots, n\}$*, then* $\alpha_1 + (\alpha_2 + (\cdots + (\alpha_{n-1} + \alpha_n) \cdots))) = \alpha_{\pi(1)} + (\alpha_{\pi(2)} + (\cdots + (\alpha_{\pi(n-1)} + \alpha_{\pi(n)}) \cdots)))$ *, so that the expression* $\sum_{i=1}^n \alpha_i$ *may be given an unambiguous meaning.*

(5) If $\alpha_1 = \cdots = \alpha_n = \alpha$ *in (3), define* $n\alpha = \sum_{i=1}^n \alpha_i$*. Show that if* m, n *are arbitrary positive integers, and if* $\alpha \in \mathbb{K}$*, then*

(i) $(m + n)\alpha = m\alpha + n\alpha$;

(ii) $(mn)\alpha = m(n\alpha)$;

(iii) $-(n\alpha) = n(-\alpha)$;

(iv) if we define $0\alpha = 0$ *and* $(-n)\alpha = -(n\alpha)$ *(for* $n \in \mathbb{N}$*), then (i) and (ii) are valid for all* $m, n \in \mathbb{Z}$;

(v) $(n \cdot 1)\alpha = n\alpha$, *where the 1 on the left denotes the 1 in* $\mathbb{K}$ *and we write* $n \cdot 1$ *for* $\sum_{i=1}^{n} 1$.

REMARK A.1.3 If $\mathbb{K}$ is a field, there are two possibilities: either (i) $m \cdot 1 \neq 0 \ \forall m \in \mathbb{N}$ or (ii) there exists a positive integer m such that $m \cdot 1 = 0$; if we let p denote the smallest positive integer with this property, it follows from (5)(ii), (5)(v) and (3) of Exercise A.1.2 that p should necessarily be a prime number.

We say that the field $\mathbb{K}$ has **characteristic** equal to 0 or p according as possibility (i) or (ii) occurs for $\mathbb{K}$.

EXAMPLE A.1.4 (1) The sets of real ($\mathbb{R}$), complex ($\mathbb{C}$) and rational ($\mathbb{Q}$) numbers are examples of fields of characteristic 0.

(2) Consider $\mathbb{Z}_p$, the set of congruence classes of integers modulo p, where p is a prime number. If we define addition and multiplication 'modulo p', then $\mathbb{Z}_p$ is a field of characteristic p, which has the property that it has exactly p elements.

For instance, addition and multiplication in the field $\mathbb{Z}_3 = \{0, 1, 2\}$ are given thus:

$$0 + x = x \ \forall x \in \{0, 1, 2\}; 1 + 1 = 2, 1 + 2 = 0, 2 + 2 = 1;$$

$$0 \cdot x = 0, 1 \cdot x = x \ \forall x \in \{1, 2\}, 2 \cdot 2 = 1.$$

(3) Given any field $\mathbb{K}$, let $\mathbb{K}[t]$ denote the set of all polynomials in the indeterminate variable t, with coefficients coming from $\mathbb{K}$; thus, the typical element of $\mathbb{K}[t]$ has the form $p = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_n t^n$, where $\alpha_i \in \mathbb{K} \ \forall i$. If we add and multiply polynomials in the usual fashion, then $\mathbb{K}[t]$ is a commutative **ring** with identity - meaning that it has addition and multiplication defined on it, which satisfy all the axioms of a field except for the axiom demanding the existence of inverses of non-zero elements.

Consider the set of all expressions of the form $\frac{p}{q}$, where $p, q \in \mathbb{K}[t]$; regard two such expressions, say $\frac{p_i}{q_i}, i = 1, 2$ as being equivalent if $p_1 q_2 = p_2 q_1$; it can then be verified that the collection of all (equivalence classes of) such expressions forms a field $\mathbb{K}(t)$, with respect to the natural definitions of addition and multiplication.

(4)If $\mathbb{K}$ is any field, consider two possibilities: (a) char $\mathbb{K} = 0$; in this case, it is not hard to see that the mapping $\mathbb{Z} \ni m \rightarrow$

$m \cdot 1 \in \mathbb{K}$ is a 1-1 map, which extends to an embedding of the field $\mathbb{Q}$ into $\mathbb{K}$ via $\frac{m}{n} \mapsto (m \cdot 1)(n \cdot 1)^{-1}$; (b) char $\mathbb{K} = p \neq 0$; in this case, we see that there is a natural embedding of the field $\mathbb{Z}_p$ into $\mathbb{K}$. In either case, the 'subfield generated by 1' is called the **prime field** of $\mathbb{K}$; (thus, we have seen that this prime field is isomorphic to $\mathbb{Q}$ or $\mathbb{Z}_p$, according as the characteristic of $\mathbb{K}$ is 0 or p). $\square$

Before proceeding further, we first note that the definition of a vector space that we gave in Definition 1.1.1 makes perfectly good sense if we replace every occurrence of $\mathbb{C}$ in that definition with $\mathbb{K}$ - see Remark 1.1.2 - and the result is called a **vector space over the field** $\mathbb{K}$.

Throughout this section, the symbol V will denote a vector space over $\mathbb{K}$. As in the case of $\mathbb{C}$, there are some easy examples of vector spces over any field $\mathbb{K}$, given thus: for each positive integer n, the set $\mathbb{K}^n = \{(\alpha_1, \cdots, \alpha_n) : \alpha_i \in \mathbb{K} \ \forall i\}$ acquires the structure of a vector space over $\mathbb{K}$ with respect to coordinate-wise definitions of vector addition and scalar mulitplication. The reader should note that if $\mathbb{K} = \mathbb{Z}_p$, then $\mathbb{Z}_p^n$ is a vector space with p^n elements and is, in particular, a finite set! (The vector spaces $\mathbb{Z}_2^n$ play a central role in practical day-to-day affairs related to computer programming, coding theory, etc.)

Recall (from §1.1) that a subset $W \subset V$ is called a subspace of V if $x, y \in W, \alpha \in \mathbb{K} \Rightarrow (\alpha x + y) \in W$. The following exercise is devoted to a notion which is the counterpart, for vector spaces and subspaces, of something seen several times in the course of this text - for instance, Hilbert spaces and closed subspaces, Banach algebras and closed subalgebras, C^*-algebras and C^*-subalgebras, etc.; its content is that there exist two descriptions - one existential, and one 'constructive' - of the 'sub-object' and that both descriptions describe the same object.

EXERCISE A.1.5 *Let V be a vector space over $\mathbb{K}$.*

(a) If $\{W_i : i \in I\}$ is any collection of subspaces of V, show that $\cap_{i \in I} W_i$ is a subspace of V.

(b) If $S \subset V$ is any subset, define $\bigvee S = \cap \{W : W$ is a subspace of V, and $S \subset W\}$, and show that $\bigvee S$ is a subspace which contains S and is the smallest such subspace; we refer to $\bigvee S$ as the **subspace generated by S**; *conversely, if W is*

a subspace of V and if S is any set such that $W = \bigvee S$, we shall say that S **spans** *or generates W, and we shall call S a* **spanning set for** *W.*

(c) If $S_1 \subset S_2 \subset V$, then show that $\bigvee S_1 \subset \bigvee S_2$.

(d) Show that $\bigvee S = \{\sum_{i=1}^n \alpha_i x_i : \alpha_i \in \mathbb{K}, x_i \in S, n \in \mathbb{N}\}$, for any subset S of V. (The expression $\sum_{i=1}^n \alpha_i x_i$ is called a **linear combination** *of the vectors $x_1, \cdots, x_n$; thus the subspace spanned by a set is the collection of all linear combinations of vectors from that set.)*

LEMMA A.1.6 *The following conditions on a set $S \subset (V - \{0\})$ are equivalent:*

(i) if S_0 is a proper subset of S, then $\bigvee S_0 \neq \bigvee S$;

(ii) if $n \in \mathbb{N}$, if $x_1, x_2, \cdots, x_n$ are distinct elements of S, and if $\alpha_1, \alpha_2, \cdots, \alpha_n \in \mathbb{K}$ are such that $\sum_{i=1}^n \alpha_i x_i = 0$, then $\alpha_i = 0 \forall i$.

A set S which satisfies the above conditions is said to be **linearly independent***; and a set which is not linearly independent is said to be* **linearly dependent**.

Proof : $(i) \Rightarrow (ii)$: If $\sum_{i=1}^n \alpha_i x_i = 0$, with $x_1, \cdots, x_n$ being distinct elements of S and if some coefficient, say α_j, is not 0, then we find that $x_j = \sum_{i \neq j} \beta_i x_i$, where $\beta_i = -\frac{\alpha_i}{\alpha_j}$; this implies that $\bigvee(S - \{x_j\}) = \bigvee S$, contradicting the hypothesis (i).

$(ii) \Rightarrow (i)$: Suppose (ii) is satisfied and suppose $S_0 \subset (S - \{x\}$; (thus S_0 is a proper subset of S;) then we assert that $x \notin \bigvee S_0$; if this assertion were false, we should be able to express x as a linear combination, say $x = \sum_{i=1}^n \beta_i x_i$, where $x_1, \cdots, x_n$ are elements of S_0, where we may assume without loss of generality that the x_i's are all distinct; then, setting $x = x_0, \alpha_0 = 1, \alpha_i = -\beta_i$ for $1 \leq i \leq n$, we find that $x_0, x_1, \cdots, x_n$ are distinct elements of S and that the α_i's are scalars not all of which are 0 (since $\alpha_0 \neq 0$), such that $\sum_{i=0}^n \alpha_i x_i = 0$, contradicting the hypothesis (ii). □

PROPOSITION A.1.7 *Let S be a subset of a vector space V; the following conditions are equivalent:*

(i) S is a minimal spanning set for V;

(ii) S is a maximal linearly independent set in V.

A set S satisfying these conditions is called a (linear) **basis** *for V.*

Proof : $(i) \Rightarrow (ii)$: The content of Lemma A.1.6 is that if S is a minimal spanning set for $\bigvee S$, then (and only then) S is linearly independent; if $S_1 \supset S$ is a strictly larger set than S, then we have $V = \bigvee S \subset \bigvee S_1 \subset V$, whence $V = \bigvee S_1$; since S is a proper subset of S_1 such that $\bigvee S = \bigvee S_1$, it follows from Lemma A.1.6 that S_1 is not linearly independent; hence S is indeed a maximal linearly independent subset of V.

$(ii) \Rightarrow (i)$: Let $W = \bigvee S$; suppose $W \neq V$; pick $x \in V - W$; we claim that $S_1 = S \cup \{x\}$ is linearly independent; indeed, suppose $\sum_{i=1}^n \alpha_i x_i = 0$, where $x_1, \cdots, x_n$ are distinct elements of S_1, and not all α_i's are 0; since S is assumed to be linearly independent, it must be the case that there exists an index $i \in \{1, \cdots, n\}$ such that $x = x_i$ and $\alpha_i \neq 0$; but this would mean that we can express x as a linear combination of the remaining x_j's, and consequently that $x \in W$, conradicting the choice of x. Thus, it must be the case that S is a spanning set of V, and since S is linearly independent, it follows from Lemma A.1.6 that in fact S must be a minimal spanning set for V. □

Lemma A.1.8 *Let $L = \{x_1, \cdots, x_m\}$ be any linearly independent set in a vector space V, and let S be any spanning set for V; then there exists a subset $S_0 \subset S$ such that (i) $S - S_0$ has exactly m elements and (ii) $L \cup S_0$ is also a spanning set for V.*

In particular, any spanning set for V has at least as many elements as any linearly independent set.

Proof : The proof is by induction on m. If $m = 1$, since $x_1 \in V = \bigvee S$, there exist a decomposition $x_1 = \sum_{j=1}^n \alpha_j y_j$, with $y_j \in S$. Since $x_1 \neq 0$ (why?), there must be some j such that $\alpha_j y_j \neq 0$; it is then easy to see (upon dividing the previous equation by α_j - which is non-zero - and rearranging the terms of the resulting equation appropriately) that $y_j \in \bigvee\{y_1, \cdots, y_{j-1}, x_1, y_{j+1}, \cdots, y_n\}$, so that $S_0 = S - \{y_j\}$ does the job.

Suppose the lemma is valid when L has $(m-1)$ elements; then we can, by the induction hypothesis - applied to $L - \{x_m\}$ and S - find a subset $S_1 \subset S$ such that $S - S_1$ has $(m-1)$ elements, and such that $\bigvee(S_1 \cup \{x_1, \cdots, x_{m-1}\}) = V$. In particular, since $x_m \in V$, this means we can find vectors $y_j \in S_1$ and scalars $\alpha_i, \beta_j \in \mathbb{K}$, for $i < m, j \leq n$, such that $x_m = \sum_{i=1}^{m-1} \alpha_i x_i + \sum_{j=1}^{n} \beta_j y_j$. The assumed linear independence of L shows that ($x_m \neq 0$ and hence that) $\beta_{j_0} y_{j_0} \neq 0$, for some j_0. As in the last paragraph, we may deduce from this that $y_{j_0} \in \bigvee(\{y_j : 1 \leq j \leq n,\ j \neq j_0\} \cup \{x_1, \cdots, x_m\}$; this implies that $V = \bigvee(S_1 \cup \{x_1, \cdots, x_{m-1}\}) = \bigvee(S_0 \cup L)$, where $S_0 = S_1 - \{y_{j_0}\}$, and this S_0 does the job. $\Box$

EXERCISE A.1.9 *Suppose a vector space V over $\mathbb{K}$ has a basis with n elements, for some $n \in \mathbb{N}$. (Such a vector space is said to be finite-dimensional.) Show that:*

(i) any linearly independent set in V has at most n elements;

(ii) any spanning set for V has at least n elements;

(iii) any linearly independent set in V can be 'extended' to a basis for V.

COROLLARY A.1.10 *Suppose a vector space V over $\mathbb{K}$ admits a finite spanning set. Then it has at least one basis; further, any two bases of V have the same number of elements.*

Proof : If S is a finite spanning set for V, it should be clear that (in view of the assumed finiteness of S that) S must contain a minimal spanning set; i.e., S contains a finite basis - say $\mathcal{B} = \{v_1, \cdots, v_n\}$ - for V (by Proposition A.1.7).

If $\mathcal{B}_1$ is any other basis for V, then by Lemma A.1.8, (applied with $S = \mathcal{B}$ and L as any finite subset of $\mathcal{B}_1$) we find that $\mathcal{B}_1$ is finite and has at most n elements; by reversing the roles of $\mathcal{B}$ and $\mathcal{B}_1$, we find that any basis of V has exactly n elements. $\Box$

It is a fact, which we shall prove in §A.2, that every vector space admits a basis, and that any two bases have the same 'cardinality', meaning that a bijective correspondence can be established between any two bases; this common cardinality is called the (linear) **dimension** of the vector space, and denoted by $dim_{\mathbb{K}} V$.

In the rest of this section, we restrict ourselves to finite-dimensional vector spaces. If $\{x_1, \cdots, x_n\}$ is a basis for (the n-dimensional vector space) V, we know that any element x of

V may be expressed as a linear combination $\sum_{i=1}^{n} \alpha_i x_i$; the fact that the x_i's form a linearly independent set, ensures that the coefficients are uniquely determined by x; thus we see that the mapping $\mathbb{K}^n \ni (\alpha_1, \cdots, \alpha_n) \overset{T}{\mapsto} \sum_{i=1}^{n} \alpha_i x_i \in V$ establishes a bijective correspondence between $\mathbb{K}^n$ and V; this correspondence is easily see to be a linear isomorphism - i.e., the map T is 1-1 and onto, and it satisfies $T(\alpha x + y) = \alpha T x + T y$, for all $\alpha \in \mathbb{K}$ and $x, y \in \mathbb{K}^n$.

As in the case of $\mathbb{C}$, we may - and will - consider linear transformations between vector spaces over $\mathbb{K}$; once we have fixed bases $\mathcal{B}_1 = \{v_1, \cdots, v_n\}$ and $\mathcal{B}_2 = \{w_1, \cdots, w_m\}$ for the vector spaces V and W respectively, then there is a 1-1 correspondence between the set $L(V, W)$ of ($\mathbb{K}$-) linear transformations from V to W on the one hand, and the set $M_{m\times n}(\mathbb{K})$ of $m \times n$ matrices with entries coming from $\mathbb{K}$, on the other; the linear transformation $T \in L(V, W)$ corresponds to the matrix $((t^i_j)) \in M_{m\times n}(\mathbb{K})$ precisely when $T v_j = \sum_{i=1}^{m} t^i_j w_i$ for all $j = 1, 2, \cdots, n$. Again, as in the case of $\mathbb{C}$, when $V = W$, we take $\mathcal{B}_2 = \mathcal{B}_1$, and in this case, the assignment $T \mapsto ((t^i_j))$ sets up a linear isomorphism of the ($\mathbb{K}$-) vector space $L(V)$ onto the vector space $M_n(\mathbb{K})$, which is an isomorphism of $\mathbb{K}$-algebras - meaning that if products in $L(V)$ and $M_n(\mathbb{K})$ are defined by composition of transformations and matrix-multiplication (defined exactly as in Exercise 1.3.1(7)(ii)) respectively, then the passage (defined above) from linear transformations to matrices respects these products. (In fact, this is precisely the reason that matrix multiplication is defined in this manner.)

Thus the study of linear transformations on an n-dimensional vector space over $\mathbb{K}$ is exactly equivalent - once we have chosen some basis for V - to the study of $n \times n$ matrices over $\mathbb{K}$.

We conclude this section with some remarks concerning determinants. Given a matrix $A = ((a^i_j)) \in M_n(\mathbb{K})$, the determinant of A is a scalar (i.e., an element of $\mathbb{K}$), usually denoted by the symbol *det A*, which captures some important features of the matrix A. In case $\mathbb{K} = \mathbb{R}$, these features can be quantified as consisting of two parts: (a) the sign (i.e., positive or negative) of *det A* determines whether the linear transformation of $\mathbb{R}^n$ corresponding to A 'preserves' or 'reverses' the 'orientation' in $\mathbb{R}^n$; and (b) the absolute value of *det A* is the quantity by which volumes of regions in $\mathbb{R}^n$ get magnified under the application of the transformation corresponding to A.

The group of permutations of the set $\{1, 2, \cdots, n\}$ is denoted S_n; as is well known, every permutation is expressible as a product of transpositions (i.e., those which interchange two letters and leave everything else fixed); while the expression of a permutation as a product of transpositions is far from unique, it is nevertheless true - see [Art], for instance - that the number of transpositions needed in any decomposition of a fixed $\sigma \in S_n$ is either always even or always odd; a permutation is called *even* or *odd* depending upon which of these possibilities occurs for that permutation. The set A_n of even permutations is a subgroup of S_n which contains exactly half the number of elements of S_n; i.e., A_n has 'index' 2 in S_n.

The so-called **alternating character** of S_n is the function defined by

$$S_n \ni \sigma \mapsto \epsilon_\sigma = \begin{cases} 1 & if\ \sigma \in A_n \\ -1 & otherwise \end{cases} , \tag{A.1.1}$$

which is easily seen to be a homomorphism from the group S_n into the multiplicative group $\{+1, -1\}$.

We are now ready for the definition of the determinant.

DEFINITION A.1.11 *If $A = ((a^i_j)) \in M_n(\mathbb{K})$, the* **determinant** *of A is the scalar defined by*

$$det\ A = \sum_{\sigma \in S_n} \epsilon_\sigma \prod_{i=1}^{n} a^i_{\sigma(i)} .$$

The best way to think of a determinant stems from viewing a matrix $A = ((a^i_j)) \in M_n(\mathbb{K})$ as the ordered tuple of its rows; thus we think of A as $(a^1, a^2, \cdots, a^n)$ where $a^i = (a^i_1, \cdots, a^i_n)$ denotes the i-th row of the matrix A. Let $R : \mathbb{K}^n \times \overset{n\ terms}{\cdots} \times \mathbb{K}^n \to M_n(\mathbb{K})$ denote the map given by $R(a^1, \cdots, a^n) = ((a^i_j))$, and let $D : \mathbb{K}^n \times \overset{n\ terms}{\cdots} \times \mathbb{K}^n \to \mathbb{K}$ be defined by $D = det \circ R$. Then, R is clearly a linear isomorphism of vector spaces, while D is just the determinant mapping, but when viewed as a mapping from $\mathbb{K}^n \times \overset{n\ terms}{\cdots} \times \mathbb{K}^n$ into $\mathbb{K}$.

We list some simple consequences of the definitions in the following proposition.

PROPOSITION A.1.12 *(1) If A' denotes the transpose of the matrix A, then det $A =$ det A'.*

(2) The mapping D is an 'alternating multilinear form': i.e.,

(a) if $x^1, \cdots, x^n \in \mathbb{K}^n$ and $\sigma \in S_n$ are arbitrary, then,

$$D(x^{\sigma(1)}, \cdots, x^{\sigma(n)}) = \epsilon_\sigma D(x^1, x^2, \cdots, x^n) ;$$

and

(b) if all but one of the n arguments of D are fixed, then D is linear as a function of the remining argument; thus, for instance, if $x^1, y^1, x^2, x^3, \cdots, x^n \in \mathbb{K}^n$ and $\alpha \in \mathbb{K}$ are arbitrary, then

$$D(\alpha x^1 + y^1, x^2, \cdots, x^n) = \alpha D(x^1, x^2, \cdots, x^n) + D(y^1, x^2, \cdots, x^n).$$

Proof : (1) Let $B = A'$, so that $b^i_j = a^j_i \forall i, j$. Since $\sigma \mapsto \sigma^{-1}$ is a bijection of S_N such that $\epsilon_\sigma = \epsilon_{\sigma^{-1}}$ $\forall \sigma \in S_n$, we have

$$\begin{aligned} det\ A &= \sum_{\sigma \in S_n} \epsilon_\sigma \prod_{i=1}^n a^i_{\sigma(i)} \\ &= \sum_{\sigma \in S_n} \epsilon_\sigma \prod_{j=1}^n a_j^{\sigma^{-1}(j)} \\ &= \sum_{\sigma \in S_n} \epsilon_{\sigma^{-1}} \prod_{i=1}^n b^i_{\sigma^{-1}(i)} \\ &= \sum_{\pi \in S_n} \epsilon_\pi \prod_{i=1}^n b^i_{\pi(i)} \\ &= det\ B\ , \end{aligned}$$

as desired.

(2) (a) Let us write $y^i = x^{\sigma(i)}$. Then,

$$\begin{aligned} D(x^{\sigma(1)}, \cdots, x^{\sigma(x_n)}) &= D(y^1, \cdots, y^n) \\ &= \sum_{\pi \in S_n} \epsilon_\pi \prod_{i=1}^n y^i_{\pi(i)} \\ &= \sum_{\pi \in S_n} \epsilon_\pi \prod_{i=1}^n x^{\sigma(i)}_{\pi(i)} \\ &= \sum_{\pi \in S_n} \epsilon_\pi \prod_{j=1}^n x^j_{\pi(\sigma^{-1}(j))} \\ &= \sum_{\tau \in S_n} \epsilon_\tau \epsilon_\sigma \prod_{j=1}^n x^j_{\tau(j)} \\ &= \epsilon_\sigma\ D(x^1, \cdots, x^n)\ . \end{aligned}$$

(b) This is an immediate consequence of the definition of the determinant and the distributive law in $\mathbb{K}$. $\square$

EXERCISE A.1.13 *(a) Show that if $x^1, \cdots, x^n \in \mathbb{K}^n$ are vectors which are not all distinct, then $D(x^1, \cdots, x^n) = 0$. (Hint: if $x^i = x^j$, where $i \neq j$, let $\tau \in S_n$ be the transposition (ij), which switches i and j and leaves other letters unaffected; then, on the one hand $\epsilon_\tau = -1$, while on the other, $x^{\tau(k)} = x^k \ \forall k$; now appeal to the alternating character of the determinant, i.e., Proposition A.1.12(1).)*

(b) Show that if the rows of a matrix are linearly dependent, then its determinant must vanish. (Hint: if the i-th row is a linear combination of the other rows, use multilinearity of the determinant (i.e., Proposition A.1.12(2)) and (a) of this exercise.)

(c) Show that if a matrix A is not invertible, then det $A = 0$. (Hint: note that the transpose matrix A' is also not invertible, deduce from Exercise 1.3.1(4)(iii) that the (columns of A', or equivalently) the rows of A are linearly dependent; now use (b) above.

The way one computes determinants in practice is by 'expanding along rows' (or columns) as follows: first consider the sets $X^i_j = \{\sigma \in S_n : \sigma(i) = j\}$, and notice that, for each fixed i, S_n is partitioned by the sets $X^i_1, \cdots, X^i_n$. Suppose $A = ((a^i_j)) \in M_n(\mathbb{K})$; for arbitrary $i, j \in \{1, 2, \cdots, n\}$, let A^i_j denote the $(n-1) \times (n-1)$ matrix obtained by deleting the i-th row and the j-th column of A. Notice then that, for any fixed i, we have

$$\begin{aligned} det\ A &= \sum_{\sigma \in S_n} \epsilon_\sigma \prod_{k=1}^{n} a^k_{\sigma(k)} \\ &= \sum_{j=1}^{n} a^i_j \left(\sum_{\sigma \in X^i_j} \epsilon_\sigma \prod_{\substack{k=1 \\ k \neq i}}^{n} a^k_{\sigma(k)} \right) . \end{aligned} \tag{A.1.2}$$

After some careful book-keeping involving several changes of variables, it can be verified that the sum within the parentheses in equation A.1.2 is nothing but $(-1)^{i+j}$ *det* A^i_j, and we find the following useful prescription for computing the determinant, which is what we called 'expanding along rows':

$$det\ A \ = \sum_{j=1}^{n} \ (-1)^{i+j} a^i_j \ det\ A^i_j \ , \ \forall\ i\ . \tag{A.1.3}$$

An entirely similar reasoning (or an adroit application of Proposition A.1.12(1) together with equation A.1.3) shows that we can also 'expand along columns': i.e.,

$$det\ A \ = \sum_{i=1}^{n} \ (-1)^{i+j} a^i_j \ det\ A^i_j \ , \ \forall\ i\ . \tag{A.1.4}$$

We are now ready for one of the most important features of the determinant.

PROPOSITION A.1.14 *A matrix is invertible if and only if* $det\ A \neq 0$*; in case* $det\ A \neq 0$*, the inverse matrix* $B = A^{-1}$ *is given by the formula*

$$b^i_j \ = \ (-1)^{i+j} (det\ A)^{-1} \ det\ A^j_i \ .$$

Proof : That a non-invertible matrix must have vanishing determinant is the content of Exercise A.1.13. Suppose conversely that $d = det\ A \neq 0$. Define $b^i_j \ = \ (-)^{i+j} \frac{1}{d} \ det\ A^j_i$. We find that if $C = AB$, then $c^i_j \ = \ \sum_{k=1}^n a^i_k b^k_j, \ \forall i, j$. It follows directly from equation A.1.3 that $c^i_i = 1\ \forall i$. In order to show that C is the identity matrix, we need to verify that

$$\sum_{k=1}^{n} (-1)^{k+j} \ a^i_k \ det\ A^j_k \ = \ 0 \ , \ \forall i \neq j. \tag{A.1.5}$$

Fix $i \neq j$ and let P be the matrix with l-th row p^l, where

$$p^l \ = \ \begin{cases} a^l & if\ l \neq j \\ a^i & if\ l = j \end{cases}$$

and note that (a) the i-th and j-th rows of P are identical (and equal to a^i) - which implies, by Exercise A.1.13 that $det\ P = 0$; (b) $P^j_k = A^j_k\ \forall k$, since P and A differ only in the j-th row; and (c) $p^j_k = a^i_k\ \forall k$. Consequently, we may deduce that the expression on the left side of equation A.1.5 is nothing but the expansion, along the j-th row of the determinant of P, and thus equation A.1.5 is indeed valid.

Thus we have $AB = I$; this implies that A is onto, and since we are in finite dimensions, we may conclulde that A must also be 1-1, and hence invertible. $\square$

The second most important property of the determinant is contained in the next result.

PROPOSITION A.1.15 *(a) If $A, B \in M_n(\mathbb{K})$, then*

$$det(AB) \; = \; (det \; A)(det \; B) \; .$$

(b) If $S \in M_n(\mathbb{K})$ is invertible, and if $A \in M_n(\mathbb{K})$, then

$$det \; (SAS^{-1}) \; = \; det \; A \; . \tag{A.1.6}$$

Proof : (a) Let $AB = C$ and let a^i, b^i and c^i denote the i-th row of A, B and C respectively; then since $c^i_j \; = \; \sum_{k=1}^n a^i_k b^k_j$, we find that $c^i \; = \; \sum_{k=1}^n a^i_k b^k$. Hence we see that

$$\begin{aligned}
det \; (AB) &= D(c^1, \cdots, c^n) \\
&= D(\sum_{k_1=1}^n a^1_{k_1} b^{k_1}, \cdots, \sum_{k_n=1}^n a^n_{k_n} b^{k_n}) \\
&= \sum_{k_1,\cdots,k_n=1}^n a^1_{k_1} \cdots a^n_{k_n} D(b^{k_1}, \cdots, b^{k_n}) \\
&= \sum_{\sigma \in S_n} (\prod_{i=1}^n a^i_{\sigma(i)}) D(b^{\sigma(1)}, \cdots, b^{\sigma(n)}) \\
&= \sum_{\sigma \in S_n} (\prod_{i=1}^n a^i_{\sigma(i)}) \epsilon_\sigma D(b^1, \cdots, b^n) \\
&= (det \; A)(det \; B)
\end{aligned}$$

where we have used (i) the multilinearity of D (i.e., Proposition A.1.12(2)) in the third line, (ii) the fact that $D(x^i, \cdots, x^n) = 0$ if $x^i = x^j$ for some $i \neq j$, so that $D(b^{k_1}, \cdots, b^{k_n})$ can be non-zero only if there exists a permutation $\sigma \in S_n$ such that $\sigma(i) = k_i \; \forall i$, in the fourth line, and (iii) the fact that D is alternating (i.e., Proposition A.1.12(1)) in the fifth line.

(b) If I denotes the identity matrix, it is clear that $det\ I = 1$; it follows from (a) that $det\ S^{-1} = (det\ S)^{-1}$, and hence the desired conclusion (b) follows from another application of (a) (to a triple-product). □

The second part of the preceding proposition has the important consequence that we can unambiguously talk of the determinant of a linear transformation of a finite-dimensional vector space into itself.

COROLLARY A.1.16 *If V is an n-dimensional vector space over $\mathbb{K}$, if $T \in L(V)$, if $\mathcal{B}_1 = \{x_1, \cdots, x_n\}$ and $\mathcal{B}_2 = \{y_1, \cdots, y_n\}$ are two bases for V, and if $A = [T]_{\mathcal{B}_1}, B = [T]_{\mathcal{B}_2}$ are the matrices representing T with respect to the basis $\mathcal{B}_i$, for $i = 1, 2$, then there exists an invertible matrix $S \in M_n(\mathbb{K})$ such that $B = S^{-1}AS$, and consequently, $det\ B = det\ A$.*

Proof : Consider the unique (obviously invertible) linear transformation $U \in L(V)$ with the property that $Ux_j = y_j, 1 \leq j \leq n$. Let $A = ((a^i_j)), B = ((b^i_j))$, and let $S = ((s^i_j)) = [U]^{\mathcal{B}_2}_{\mathcal{B}_1}$ be the matrix representing U with respect to the two bases (in the notation of Exercise 1.3.1(7)(i)); thus, by definition, we have:

$$\begin{aligned} Tx_j &= \sum_{i=1}^n a^i_j x_i \\ Ty_j &= \sum_{i=1}^n b^i_j y_i \\ y_j &= \sum_{i=1}^n s^i_j x_i \ ; \end{aligned}$$

deduce that

$$\begin{aligned} Ty_j &= \sum_{i=1}^n b^i_j y_i \\ &= \sum_{i,k=1}^n b^i_j s^k_i x_k \\ &= \sum_{k=1}^n \left(\sum_{i=1}^n s^k_i b^i_j \right) \ x_k \ , \end{aligned}$$

while, also

$$\begin{aligned} Ty_j &= T\left(\sum_{i=1}^n s_j^i x_i\right) \\ &= \sum_{i=1}^n s_j^i T x_i \\ &= \sum_{i,k=1}^n s_j^i a_i^k x_k \\ &= \sum_{k=1}^n \left(\sum_{i=1}^n a_i^k s_j^i\right) x_k \ ; \end{aligned}$$

conclude that

$$[T]_{\mathcal{B}_1}^{\mathcal{B}_2} = SB = AS \ .$$

Since S is invertible, we thus have $B = S^{-1}AS$, and the proof of the corollary is complete (since the last statement follows from Proposition A.1.15(b)). □

Hence, given a linear transformation $T \in L(V)$, we may unambiguously define the determinant of T as the determinant of any matrix which represents T with respect to some basis for V; thus, $det\ T = det\ [T]_{\mathcal{B}}^{\mathcal{B}}$, where $\mathcal{B}$ is any (ordered) basis of V. In particular, we have the following pleasant consequence.

PROPOSITION A.1.17 *Let $T \in L(V)$; define the* **characteristic polynomial** *of T by the equation*

$$p_T(\lambda) = det\ (T - \lambda 1_V)\ ,\ \lambda \in \mathbb{K} \tag{A.1.7}$$

where 1_V denotes the identity operator on V. Then the following conditions on a scalar $\lambda \in \mathbb{K}$ are equivalent:
(i) $(T - \lambda 1_V)$ is not invertible;
(ii) there exists a non-zero vector $x \in V$ such that $Tx = \lambda x$;
(iii) $p_T(\lambda) = 0$.

Proof : It is easily shown, using induction and expansion of the determinant along the first row (for instance), that $p_T(\lambda)$ is a polynomial of degree n, with the coefficient of λ^n being $(-1)^n$.

The equivalence $(i) \Leftrightarrow (iii)$ is an immediate consequence of Corollary A.1.16, Proposition A.1.14. The equivalence $(i) \Leftrightarrow (ii)$

is a consequence of the fact that a 1-1 linear transformation of a finite-dimensional vector space into itself is necessarily onto and hence invertible. (Reason: if $\mathcal{B}$ is a basis for V, and if T is 1-1, then $T(\mathcal{B})$ is a linearly independent set with n elements in the n-dimensional space, and any such set is necessarily a basis.) $\Box$

DEFINITION A.1.18 *If T, λ, x are as in Proposition A.1.17(ii), then λ is said to be an* **eigenvalue** *of T, and x is said to be an* **eigenvector** *of T corresponding to the eigenvalue λ.*

We conclude this section with a couple of exercises.

EXERCISE A.1.19 *Let $T \in L(V)$, where V is an n-dimensional vector space over $\mathbb{K}$.*

(a) Show that there are at most n distinct eigenvalues of T. (Hint: a polynomial of degree n with coefficients from $\mathbb{K}$ cannot have $(n+1)$ distinct roots.)

(b) Let $\mathbb{K} = \mathbb{R}, V = \mathbb{R}^2$ and let T denote the transformation of rotation (of the plane about the origin) by 90^o in the counter-clockwise direction.

(i) Show that the matrix of T with respect to the standard basis $\mathcal{B} = \{(1,0),(0,1)\}$ is given by the matrix

$$[T]_{\mathcal{B}} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} ,$$

and deduce that $p_T(\lambda) = \lambda^2 + 1$. Hence, deduce that T has no (real) eigenvalue.

(ii) Give a geometric proof of the fact that T has no (real) eigenvalue.

(c) Let $p(\lambda) = (-1)^n\lambda^n + \sum_{k=0}^{n-1} \alpha_k \lambda^k$ be any polynomial of degree n with leading coefficient $(-1)^n$. Show that there exists a linear transformation on any n-dimensional space with the property that $p_T = p$. (Hint: consider the matrix

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & \beta_0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & \beta_1 \\ 0 & 1 & 0 & 0 & \cdots & 0 & \beta_2 \\ 0 & 0 & 1 & 0 & \cdots & 0 & \beta_3 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & \beta_{n-1} \end{bmatrix} ,$$

wher $\beta_j = (-1)^{n+1}\alpha_j, \ \forall 0 \leq j < n.$*)*

(d) Show that the following conditions on $\mathbb{K}$ *are equivalent:*

(i) every non-canstant polynomial p *with coefficients from* $\mathbb{K}$ *has a root - i.e., there exists an* $\alpha \in \mathbb{K}$ *such that* $p(\alpha) = 0$*;*

(ii) every linear transformation on a finite-dimensional vector space over $\mathbb{K}$ *has an eigenvalue.*

(Hint: Use Proposition A.1.17 and (c) of this exercise).

A.2 Transfinite considerations

One of the most useful arguments when dealing with 'infinite constructions' is the celebrated *Zorn's lemma*, which is equivalent to one of the basic axioms of set theory, the 'Axiom of Choice'. The latter axiom amounts to the statement that if $\{X_i : i \in I\}$ is a non-empty collection of non-empty sets, then their Cartesian product $\prod_{i \in I} X_i$ is also non-empty. It is a curious fact that this axiom is actually equivalent to Tychonoff's theorem on the compactness of the Cartesian product of compact spaces; in any case, we just state Zorn's Lemma below, since the question of 'proving' an axiom does not arise, and then discuss some of the typical ways in which it will be used.

LEMMA A.2.1 *(***Zorn's lemma***) Suppose a non-empty partially ordered set* $(\mathcal{P}, \leq)$ *satisfies the following condition: every totally ordered set in* $\mathcal{P}$ *has an upper bound (i.e., whenever* $\mathcal{C}$ *is a subset of* $\mathcal{P}$ *with the property that any two elements of* $\mathcal{C}$ *are comparable - meaning that if* $x, y \in \mathcal{C}$*, then either* $x \leq y$ *or* $y \leq x$ *- then there exists an element* $z \in \mathcal{P}$ *such that* $x \leq z \ \forall x \in \mathcal{C}$*).*

Then $\mathcal{P}$ *admits a* **maximal element** *- i.e., there exists an element* $z \in \mathcal{P}$ *such that if* $x \in \mathcal{P}$ *and* $z \leq x$*, then necessarily* $x = z$*.*

A prototypical instance of the manner in which Zorn's lemma is usually applied is contained in the proof of the following result.

PROPOSITION A.2.2 *Every vector space (over an arbitrary field* $\mathbb{K}$*), with the single exception of* $V = \{0\}$*, has a basis.*

Proof : In view of Proposition A.1.7, we need to establish the existence of a maximal linearly independent subset of the given vector space V.

Let $\mathcal{P}$ denote the collection of all linearly independent subsets of V. Clearly the set $\mathcal{P}$ is partially ordered with respect to inclusion. We show now that Zorn's lemma is applicable to this situation.

To start with, $\mathcal{P}$ is not empty. This is because V is assumed to be non-zero, and so, if x is any non-zero vector in V, then $\{x\} \in \mathcal{P}$.

Next, suppose $\mathcal{C} = \{S_i : i \in I\}$ is a totally ordered subset of $\mathcal{P}$; thus for each $i \in I$, we are given a linearly independent set S_i in $\mathcal{P}$; we are further told that whenever $i, j \in I$, either $S_i \subset S_j$ or $S_j \subset S_i$. Let $S = \cup_{i \in I} S_i$; we contend that S is a linearly independent set; for, suppose $\sum_{k=1}^{n} \alpha_k x_k = 0$ where $n \in \mathbb{N}, \alpha_k \in \mathbb{K}$ and $x_k \in S$ for $1 \leq k \leq n$; by definition, there exist indices $i_k \in I$ such that $x_k \in S_{i_k} \forall k$; the assumed total ordering on C clearly implies that we may assume, after re-labelling, if necessary, that $S_{i_1} \subset S_{i_2} \subset \cdots \subset S_{i_n}$. This means that $x_k \in S_{i_n} \ \forall \ 1 \leq k \leq n$; the assumed linear independence of S_{i_n} now shows that $\alpha_k = 0 \ \forall \ 1 \leq k \leq n$. This establishes our contention that S is a linearly independent set. Since $S_i \subset S \ \forall \ i \in I$, we have in fact verified that the arbitrary totally ordered set $\mathcal{C} \subset \mathcal{P}$ admits an upper bound, namely S.

Thus, we may conclude from Zorn's lemma that there exists a maximal element, say $\mathcal{B}$, in $\mathcal{P}$; i.e., we have produced a maximal linearly independent set, namely $\mathcal{B}$, in V, and this is the desired basis. $\square$

The following exercise lists a couple of instances in the body of this book where Zorn's lemma is used. In each case, the solution is an almost verbatim repetition of the above proof.

EXERCISE A.2.3 *(1) Show that every Hilbert space admits an orthonormal basis.*

(2) Show that every (proper) ideal in a unital algebra is contained in a maximal (proper) ideal. (Hint: If $\mathcal{I}$ is an ideal in an algebra $\mathcal{A}$, consider the collection $\mathcal{P}$ of all ideals of $\mathcal{A}$ which contain $\mathcal{I}$; show that this is a non-empty set which is partially ordered by inclusion; verify that every totally ordered set in $\mathcal{P}$

has an upper bound (namely the union of its members); show that a maximal element of $\mathcal{P}$ is necessarily a maximal ideal in $\mathcal{A}$.)

The rest of this section is devoted to an informal discussion of **cardinal numbers**. We will not be very precise here, so as to not get into various logical technicalities. In fact we shall not define what a cardinal number is, but we shall talk about the *cardinality* of a set.

Loosely speaking, we shall agree to say that two sets have the same **cardinality** if it is possible to set up a 1-1 correspondence between their members; thus, if X, Y are sets, and if there exists a 1-1 function f which maps X onto Y, we shall write $|X| = |Y|$ (and say that X and Y have the same cardinality). When this happens, let us agree to also write $X \sim Y$.

Also, we shall adopt the convention that the symbol $\coprod_{i \in I} X_i$ (resp., $A \coprod B$) denotes the union of a collection of (resp., two) sets which are pairwise disjoint.

We come now to an elementary proposition, whose obvious proof we omit.

Proposition A.2.4 *If $X = \coprod_{i \in I} X_i$ and $Y = \coprod_{i \in I} Y_i$ are partitions of two sets such that $|X_i| = |Y_i| \ \forall \ i \in I$, then also $|X| = |Y|$.*

Thus, 'cardinality is a completely additive function' (just as a measure is a countably additive function, while there is no 'countability restriction' now).

More generally than before, if X and Y are sets, we shall say that $|X| \leq |Y|$ if there exists some subset $Y_0 \subset Y$ such that $|X| = |Y_0|$. It should be clear that '$\leq$ ' is a reflexive (ie., $|X| \leq |X| \ \forall X$) and transitive (i.e., $|X| \leq |Y|$ and $|Y| \leq |Z|$ imply $|X| \leq |Z|$) relation.

The next basic fact is the so-called **Schroeder-Bernstein theorem**, which asserts that '$\leq$ ' is an anti-symmetric relation. (We are skirting close to serious logical problems here - such as: what is a cardinal number? does it make sense to talk of the set of all cardinal numbers? etc.; but we ignore such questions and hide behind the apology that we are only speaking loosely, as stated earlier. The reader who wants to know more about such logical subtleties might look at [Hal], for instance, for a primer, as well as an introduction to more serious (and less 'naive') literature, on the subject.)

THEOREM A.2.5 *Let X and Y be sets. Then,*

$$|X| \leq |Y|, \ |Y| \leq |X| \ \Rightarrow \ |X| = |Y| \ .$$

Proof : The hypothesis translates into the existence of 1-1 functions $f : X \to Y$ and $g : Y \to X$; to prove the theorem, we need to exhibit a 1-1 function which maps X **onto** Y.

Let $g(Y) = Y_0$. Set $h = g \circ f$. Then h is a 1-1 map of X into itself. Define $X_n = h^{(n)}(X)$ and $Y_n = h^{(n)}(Y_0)$, where $h^{(n)}$ denotes the composition of h with itself n times. The definitions imply that

$$X = X_0 \supset Y_0 \supset X_1 \supset Y_1 \supset \cdots \ ,$$

and consequently, we see that

$$X_\infty \ = \ \cap_{n=1}^\infty X_n \ = \ \cap_{n=1}^\infty Y_n \ .$$

It follows that we have the following partitions of X and Y_0 respectively:

$$\begin{aligned} X &= \left(\coprod_{n=0}^\infty (X_n - Y_n) \right) \ \coprod \ \left(\coprod_{n=0}^\infty (Y_n - X_{n+1}) \right) \ \coprod \ X_\infty \\ Y_0 &= \left(\coprod_{n=0}^\infty (Y_n - X_{n+1}) \right) \ \coprod \left(\coprod_{n=1}^\infty (X_n - Y_n) \right) \ \coprod \ X_\infty \ ; \end{aligned}$$

consider the function $\phi : X = X_0 \to Y_0$ defined by

$$\phi(x) \ = \ \begin{cases} h(x) & if \ x \notin \ X_\infty \\ x & if \ x \in X_\infty \end{cases}$$

It is clear that ϕ is a bijection and consequently, $g^{-1} \circ \phi$ is a bijection of X onto Y (where, of course, g^{-1} denotes the map from Y_0 to Y which is the inverse of the bijection $g : Y \to Y_0$).
□

The next step is to show that $\leq \ '$ is a 'total order'.

LEMMA A.2.6 *Let X and Y be arbitrary sets. Then, either $|X| \leq |Y|$ or $|Y| \leq |X|$.*

Proof : Assume, without loss of generality, that both X and Y are non-empty sets. In particular, there exists (singleton) subsets $X_0 \subset X, Y_0 \subset Y$ such that $|X_0| = |Y_0|$.

Let $\mathcal{P} = \{(A, f, B) : A \subset X, B \subset Y$, and $f : A \to B$ is a bijection$\}$. Then by the previous paragraph, $\mathcal{P} \neq \emptyset$. Define an ordering on $\mathcal{P}$ as follows: if $(A, f, B), (C, g, D) \in \mathcal{P}$, say that $(A, f, B) \leq (C, g, D)$ if it is the case that $A \subset C$ and that $g|_A = f$ (so that, in particular, we should also have $B \subset D$).

It should be easy to see that this defines a partial order on $\mathcal{P}$; further, if $\mathcal{C} = \{(A_i, f_i, B_i) : i \in I\}$ is a totally ordered set in $\mathcal{P}$, it should be clear that if we define $A = \cup_{i \in I} A_i$ and $B = \cup_{i \in I} B_i$, then there exists a unique map f of A onto B with the property that $f|_{A_i} = f_i$; in other words, (A, f, B) is an upper bound for $\mathcal{P}$.

Thus Zorn's lemma is applicable; if (X_0, f, Y_0) is a maximal element of $\mathcal{P}$ - whose existence is assured by Zorn's lemma - then it cannot be the case that both $X - X_0$ and $Y - Y_0$ are non-empty; for, in that case, we could arbitrarily pick $x \in X - X_0, y \in Y - Y_0$, and define the bijection F of $A = X_0 \cup \{x\}$ onto $B = Y_0 \cup \{y\}$ by $F(x) = y$ and $F|_{X_0} = f$; then the element (A, F, B) would contradict the maximality of (X_0, f, Y_0).

In conclusion, either $X_0 = X$ (in which case $|X| \leq |Y|$) or $Y_0 = Y$ (in which case $|Y| \leq |X|$). □

DEFINITION A.2.7 *A set X is said to be* **infinite** *if there exists a proper subset $X_0 \subset X$ (and $X_0 \neq X$) such that $|X| = |X_0|$. (Thus X is infinite if and only if there exists a 1-1 finction of X onto a proper subset of itself.)*

A set X is said to be **finite** *if it is not infinite.*

LEMMA A.2.8 *A set X is infinite if and only if $|\mathbb{N}| \leq |X|$.*

Proof : Suppose X is an infinite set; then there exists a 1-1 function $f : X \to X - \{x_0\}$, for some $x_0 \in X$. Inductively define $x_{n+1} = f(x_n)$ for all $n = 0, 1, 2, \cdots$. If we set $X_n = f^{(n)}(X)$, where $f^{(n)}$ denotes the composition of f with itself n times, it follows that $x_n \in X_n - X_{n+1}$ $\forall n$, whence $x_n \neq x_m$ $\forall n \neq m$; i.e., the map $n \mapsto x_n$ is a 1-1 map of $\mathbb{N}$ into X and hence $|\mathbb{N}| \leq |X|$.

Conversely, if $g : \mathbb{N} \to X$ is a 1-1 map, consider the map $f : X \to X$ defined by

$$f(x) = \begin{cases} g(n+1) & if\ x = g(n) \\ x & if\ x \notin g(\mathbb{N}) \end{cases}$$

and note that $f : X \to X - \{g(1)\}$ is a 1-1 map, whence X is infinite. □

The following bit of natural notation will come in handy; if X and Y are sets, let us write $|X| < |Y|$ if it is the case that $|X| \leq |Y|$ and $|X| \neq |Y|$.

The next result is a sort of 'Euclidean algorithm' for general cardinal numbers.

Proposition A.2.9 *Let X and Y be arbitrary sets. Suppose Y is not empty. Then, there exists a partition*

$$X = \left(\coprod_{i \in I} X_i \right) \coprod R , \qquad \text{(A.2.8)}$$

where I is some (possibly empty) index set, such that $|X_i| = |Y|\ \forall\ i \in I$, and $|R| < |Y|$; further, if I is infinite, then there exists a (possibly different) partition $X = \coprod_{i \in I} Y_i$ with the property that $|Y| = |Y_i|\ \forall i \in I$. (The R in equation A.2.8 is supposed to signify the 'remainder after having divided X by Y'.)

Proof : If it is not true that $|Y| \leq |X|$, then, by Lemma A.2.6, we must have $|X| < |Y|$ and we may set $I = \emptyset$ and $R = X$ to obtain a decomposition of the desired sort.

So suppose $|Y| \leq |X|$. Let $\mathcal{P}$ denote the collection of all families $\mathcal{S} = \{X_i : i \in I\}$ of pairwise disjoint subsets of X, where I is some non-empty index set, with the property that $|X_i| = |Y|$ $\forall i \in I$. The assumption $|Y| \leq |X|$ implies that there exists $X_0 \subset X$ such that $|Y| = |X_0|$; thus $\{X_0\} \in \mathcal{P}$, and consequently $\mathcal{P}$ is non-empty.

It should be clear that $\mathcal{P}$ is partially ordered by inclusion, and that if $\mathcal{C} = \{\mathcal{S}_\lambda : \lambda \in \Lambda\}$ is a totally ordered set in $\mathcal{P}$, then $\mathcal{S} = \cup_{\lambda \in \Lambda} \mathcal{S}_\lambda$ is an element of $\mathcal{P}$ which is 'greater than or equal to' (i.e., contains) each member $\mathcal{S}_\lambda$ of $\mathcal{C}$. Thus, every totally ordered set in $\mathcal{P}$ has an upper bound.

Hence, by Zorn's lemma, there exists a maximal element in $\mathcal{P}$, say $\mathcal{S} = \{X_i : i \in I\}$. Let $R = X - (\coprod_{i\in I} X_i)$. Clearly, then, equation A.2.8 is satisfied. Further, if it is not the case that $|R| < |Y|$, then, by Lemma A.2.6, we must have $|Y| = |A|$, with $A \subset R$; but then the family $\mathcal{S}' = \mathcal{S} \cup \{A\}$ would contradict the maximality of $\mathcal{S}$. This contradiction shows that it must have been the case that $|R| < |Y|$, as desired.

Finally, suppose we have a decomposition as in equation A.2.8, such that I is infinite. Then, by Lemma A.2.8, we can find a subset $I_0 = \{i_n : n \in \mathbb{N}\}$, where $i_n \neq i_m \ \forall n \neq m$. Since $|R| \leq |X_{i_1}|$, we can find a subset $R_0 \subset X_{i_1}$ such that $|R| = |R_0|$; notice now that

$$\begin{aligned} |X_{i_n}| &= |X_{i_{n+1}}| \ \forall \ n \in \mathbb{N} \\ |X_i| &= |X_i| \ \forall \ i \in I - I_0 \\ |R| &= |R_0| \end{aligned}$$

and deduce from Proposition A.2.4 that

$$\begin{aligned} |X| &= |\left(\coprod_{n=1}^{\infty} X_{i_n}\right) \ \coprod \ \left(\coprod_{i\in(I-I_0)} X_i\right) \ \coprod \ R| \\ &= |\left(\coprod_{n=2}^{\infty} X_{i_n}\right) \ \coprod \ \left(\coprod_{i\in(I-I_0)} X_i\right) \ \coprod \ R_0| \ ; \quad \text{(A.2.9)} \end{aligned}$$

since the set on the right of equation A.2.9 is clearly a subset of $Z = \coprod_{i\in I} X_i$, we may deduce - from Theorem A.2.5 - that $|X| = |Z|$; if $f : Z \to X$ is a bijection (whose existence is thus guaranteed), let $Y_i = f(X_i) \ \forall i \in I$; these Y_i's do what is wanted of them. □

EXERCISE A.2.10 *(0) Show that Proposition A.2.9 is false if Y is the empty set, and find out where the non-emptiness of Y was used in the proof presented above.*

(1) Why is Proposition A.2.9 a generalisation of the Euclidean algorithm (for natural numbers)?

PROPOSITION A.2.11 *A set X is infinite if and only if X is non-empty and there exists a partition $X = \ = \coprod_{n=1}^{\infty} A_n$ with the property that $|X| = |A_n| \ \forall n$.*

Proof : We only prove the non-trivial ('only if') implication.

Case (i) : $X = \mathbb{N}$.

Consider the following listing of elements of $\mathbb{N} \times \mathbb{N}$:

$$(1,1);\ (1,2),(2,1);\ (1,3),(2,2),(3,1);\ (1,4),\cdots,(4,1);\cdots\ ;$$

this yields a bijection $f : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$; let $A_n = f^{-1}(\{n\} \times \mathbb{N})$; then $\mathbb{N} = \coprod_{n \in \mathbb{N}} A_n$ and $|A_n| = |\mathbb{N}|\ \forall n \in \mathbb{N}$.

Case (ii) : X arbitrary.

Suppose X is infinite. Thus there exists a 1-1 function f of X onto a proper subset $X_1 \subset X$; let $Y = X - X_1$, which is non-empty, by hypothesis. Inductively define $Y_1 = Y$, and $Y_{n+1} = f(Y_n)$; note that $\{Y_n : n = 1, 2, \cdots\}$ is a sequence of pairwise disjoint subsets of X with $|Y| = |Y_n|\ \forall n$. By case (i), we may write $\mathbb{N} = \coprod_{n=1}^\infty A_n$, where $|A_n| = |\mathbb{N}|\ \forall n$. Set $W_n = \coprod_{k \in A_n} Y_k$, and note that $|W_n| = |Y \times \mathbb{N}|\ \forall n$. Set $R = X - \cup_{n=0}^\infty Y_n$, and observe that $X = (\coprod_{n=1}^\infty W_n) \coprod R$.

Deduce - from the second half of Proposition A.2.9(b) - that there exists an infinite set I such that $|X| = |Y \times \mathbb{N} \times I|$. Observe now that $Y \times \mathbb{N} \times I = \coprod_{n=1}^\infty (Y \times A_n \times I)$, and that $|Y \times \mathbb{N} \times I| = |Y \times A_n \times I|,\ \forall n$.

Thus, we find that we do indeed have a partition of the form $X = \coprod_{n=1}^\infty X_n$, where $|X| = |X_n|\ \forall n$, as desired. □

The following Corollary follows immediately from Proposition A.2.11, and we omit the easy proof.

COROLLARY A.2.12 *If X is an infinite set, then $|X \times \mathbb{N}| = |X|$.*

It is true, more generally, that if X and Y are infinite sets such that $|Y| \leq |X|$, then $|X \times Y| = |X|$; we do not need this fact and so we do not prove it here. The version with $Y = \mathbb{N}$ is sufficient, for instance, to prove the fact that makes sense of the dimension of a Hilbert space.

PROPOSITION A.2.13 *Any two orthonormal bases of a Hilbert space have the same cardinality, and consequently, it makes sense to define the* **dimension** *of the Hilbert space $\mathcal{H}$ as the 'cardinality' of any orthonormal basis.*

Proof : Suppose $\{e_i : i \in I\}$ and $\{f_j : j \in J\}$ are two orthonormal bases of a Hilbert space $\mathcal{H}$.

First consider the case when I or J is finite. Then, the conclusion that $|I| = |J|$ is a consequence of Corollary A.1.10.

We may assume, therefore, that both I and J are infinite.

For each $i \in I$, since $\{f_j : j \in J\}$ is an orthonormal set, we know, from Bessel's inequality - see Proposition 2.3.3 - that the set $J_i = \{j \in J : \langle e_i, f_j \rangle \neq 0\}$ is countable. On the other hand, since $\{e_i : i \in I\}$ is an orthonormal basis, each $j \in J$ must belong to at least one J_i (for some $i \in I$). Hence $J = \cup_{i \in I} J_i$. Set $K = \{(i, j) : i \in I, j \in J_i\} \subset I \times J$; thus, the projection onto the second factor maps K onto J, which clearly implies that $|J| \leq |K|$.

Pick a 1-1 function $f_i : J_i \to \mathbb{N}$, and consider the function $F : K \to I \times \mathbb{N}$ defined by $F(i, j) = (i, f_i(j))$. It must be obvious that F is a 1-1 map, and hence we find (from Corollary A.2.12) that

$$|J| \leq |K| = |F(K)| \leq |I \times \mathbb{N}| = |I|.$$

By reversing the roles of I and J, we find that the proof of the proposition is complete. □

EXERCISE A.2.14 *Show that every vector space (over any field) has a basis, and that any two bases have the same cardinality. (Hint: Imitate the proof of Proposition A.2.13.)*

A.3 Topological spaces

A topological space is a set X, where there is a notion of 'nearness' of points (although there is no metric in the background), and which (experience shows) is the natural context to discuss notions of continuity. Thus, to each point $x \in X$, there is singled out some family $\mathcal{N}(x)$ of subsets of X, whose members are called 'neighbourhoods of the point x'. (If X is a metric space, we say that N is a neighbourhood of a point x if it contains all points sufficiently close to x - i.e., if there exists some $\epsilon > 0$ such that $N \supset B(x, \epsilon) = \{y \in X : d(x, y) < \epsilon\}$.) A set is thought of as being 'open' if it is a neighbourhood of each of its points

- i.e., U is open if and only if $U \in \mathcal{N}(x)$ whenever $x \in U$. A topological space is an axiomatisation of this set-up; what we find is a simple set of requirements (or axioms) to be met by the family, call it τ, of all those sets which are open, in the sense above.

DEFINITION A.3.1 *A* **topological space** *is a set X, equipped with a distinguished collection τ of subsets of X, the members of τ being referred to as* **open** *sets, such that the following axioms are satisfied:*

(1) $X, \emptyset \in \tau$; thus, the universe of discourse and the empty set are assumed to be empty;

(2) $U, V \in \tau \;\Rightarrow\; U \cap V \in \tau$; (and consequently, any finite intersection of open sets is open); and

(3) $\{U_i : i \in I\} \subset \tau \;\Rightarrow\; \cup_{i \in I} U_i \in \tau$.

The family τ is called the **topology** *on X. A subset $F \subset X$ will be said to be closed precisely when its complement $X - F$ is open.*

Thus a topology on X is just a family of sets which contains the whole space and the empty set, and which is closed under the formation of finite intersections and arbitrary unions. Of course, every metric space is a topological space in a natural way; thus, we declare that a set is open precisely when it is expressible as a union of open balls. Some other easy (but perhaps somewhat pathological) examples of topological spaces are given below.

EXAMPLE A.3.2 (1) Let X be any set; define $\tau = \{X, \emptyset\}$. This is a topology, called the **indiscrete topology**.

(2) Let X be any set; define $\tau = 2^X = \{A : A \subset X\}$; thus, every set is open in this topology, and this is called the **discrete topology**.

(3) Let X be any set; the so-called 'co-finite topology' is defined by declaring a set to be open if it is either empty or if its complement is finite.

(4) Replacing every occurrence of the word 'finite' in (3) above, by the word 'countable', gives rise to a topology, called, naturally, the 'co-countable topology' on X. (Of course, this topology would be interesting only if the set X is uncountable

- just as the co-finite topology would be interesting only for infinite sets X - since in the contrary case, the resulting topology would degenerate into the discrete topology.) $\Box$

Just as convergent sequences are a useful notion when dealing with metric spaces, we will find that nets - see Definition 2.2.3, as well as Example 2.2.4(3) - will be very useful while dealing with general topological spaces. As an instance, we cite the following result:

PROPOSITION A.3.3 *The following conditions on a subset F of a topological space are equivalent:*

(i) F is closed; i.e., $U = X - F$ is open;

(ii) if $\{x_i : i \in I\}$ is any net in F which converges to a limit $x \in X$, then $x \in F$.

Proof : $(i) \Rightarrow (ii)$: Suppose $x_i \to x$ as in (ii); suppose $x \notin F$; then $x \in U$, and since U is open, the definition of a convergent net shows that there exists some $i_0 \in I$ such that $x_i \in U \; \forall i \geq i_0$; but this contradicts the assumption that $x_i \in F \; \forall i$.

$(ii) \Rightarrow (i)$: Suppose (ii) is satisfied; we assert that if $x \in U$, then there exists an open set $B_x \subset U$ such that $x \in B_x$. (This will exhibit U as the union $\cup_{x \in U} B_x$ of open sets and thereby establish (i).) Suppose our assertion were false; this would mean that there exists an $x \in U$ such that for every open neighbourhood V of x - i.e., an open set containing x - there exists a point $x_V \in V \cap F$. Then - see Example 2.2.4(3) - $\{x_V : V \in \mathcal{N}(x)\}$ would be a net in F which converges to the point $x \notin F$, and the desired contradiction has been reached. $\Box$

We gather a few more simple facts concerning closed sets in the next result.

PROPOSITION A.3.4 *Let X be a topological space. Let us temporarily write $\mathcal{F}$ for the class of all closed sets in X.*

(1) The family $\mathcal{F}$ has the following properties (and a topological space can clearly be equivalently defined as a set where there is a distinguished class $\mathcal{F}$ of subets of X which are closed sets and which satisfy the following properties):

(a) $X, \emptyset \in \mathcal{F}$;

(b) $F_1, F_2 \in \mathcal{F} \Rightarrow (F_1 \cup F_2) \in \mathcal{F}$*;*

(c) if I *is an arbitrary set, then* $\{F_i : i \in I\} \subset \mathcal{F} \Rightarrow \cap_{i \in I} F_i \in \mathcal{F}$*.*

(2) if $A \subset X$ *is any subset of* X*, the* **closure** *of* A *is the set - which will always be denoted by the symbol* $\overline{A}$ *- defined by*

$$\overline{A} = \cap \{F \in \mathcal{F} : A \subset F\} \; ; \tag{A.3.10}$$

then

(a) $\overline{A}$ *is a closed set, and it is the smallest closed set which contains* A*;*

(b) $A \subset B \Rightarrow \overline{A} \subset \overline{B}$*;*

(c) $x \in \overline{A} \Leftrightarrow U \cap A \neq \emptyset$ *for every open set containing* x*.*

Proof : The proof of (1) is a simple exercise in complementation (and uses nothing more than the so-called 'de Morgan's laws').

(2) (a) is a consequence of (1)(c), while (b) follows immediately from (a); and (c) follows from the definitions (and the fact that a set is closed precisely when its complement is open). □

The following exercise lists some simple consequences of the definition of the closure operation, and also contains a basic definition.

EXERCISE A.3.5 *(1) Let* A *be a subset of a topological space* X*, and let* $x \in A$*; then show that* $x \in \overline{A}$ *if and only if there is a net* $\{x_i : i \in I\}$ *in* A *such that* $x_i \to x$*.*

(2) If X *is a metric space, show that nets can be replaced by sequences in (1).*

(3) A subset D *of a topological space is said to be* **dense** *if* $\overline{D} = X$*. (More generally, a set* A *is said to be 'dense in a set* B*' if* $B \subset \overline{A}$*.) Show that the following conditions on the subset* $D \subset X$ *are equivalent:*

(i) D *is dense (in* X*);*

(ii) for each $x \in X$*, there exists a net* $\{x_i : i \in I\}$ *in* D *such that* $x_i \to x$*;*

(iii) if U *is any non-empty open set in* X*, then* $D \cap U \neq \emptyset$*.*

In dealing with metric spaces, we rarely have to deal explicitly with general open sets; open balls suffice in most contexts. These are generalised in the following manner.

PROPOSITION A.3.6 *(1) Let (X, τ) be a topological space. The following conditions on a a subcollection $\mathcal{B} \subset \tau$ are equivalent:*

(i) $U \in \tau \Rightarrow \exists \{B_i : i \in I\} \subset \mathcal{B}$ such that $U = \cup_{i \in I} B_i$;

(ii) $x \in U, U$ open $\Rightarrow \exists B \in \mathcal{B}$ such that $x \in B \subset U$.

A collection $\mathcal{B}$ satisfying these equivalent conditions is called a **base** *for the topology τ.*

(2) A family $\mathcal{B}$ is a base for some topology τ on X if and only if $\mathcal{B}$ satisfies the two following conditions:

(a) $\mathcal{B}$ covers X, meaning that $X = \cup_{B \in \mathcal{B}} B$; and

(b) $B_1, B_2 \in \mathcal{B}, x \in B_1 \cap B_2 \Rightarrow \exists B \in \mathcal{B}$ such that $x \in B \subset (B_1 \cap B_2)$.

The elementary proof of this proposition is left as an exercise for the reader. It should be fairly clear that if $\mathcal{B}$ is a base for a topology τ on X, and if τ' is any topology such that $\mathcal{B} \subset \tau'$, then, necessarily $\tau \subset \tau'$. Thus, if $\mathcal{B}$ is a base for a topology τ, then τ is the 'smallest topology' with respect to which all the members of $\mathcal{B}$ are open. However, as condition (ii) of Proposition A.3.6 shows, not any collection of sets can be a base for some topology. This state of affairs is partially remedied in the following proposition.

PROPOSITION A.3.7 *(a) Let X be a set and let $\mathcal{S}$ be an arbitrary family of subsets of X. Then there exists a smallest topology $\tau(\mathcal{S})$ on X such that $\mathcal{S} \subset \tau(\mathcal{S})$; we shall refer to $\tau(\mathcal{S})$ as the topology generated by $\mathcal{S}$.*

(b) Let $X, \mathcal{S}, \tau(\mathcal{S})$ be as in (a) above. Let $\mathcal{B} = \{X\} \cup \{B : \exists\, n \in \mathbb{N}$, and $S_1, S_2, \cdots, S_n \in \mathcal{S}$ such that $B = \cap_{i=1}^{n} S_i\}$. Then $\mathcal{B}$ is a base for the topology τ; in particular, a typical element of $\tau(\mathcal{S})$ is expressible as an arbitrary union of finite intersections of members of $\mathcal{S}$.

If (X, τ) is a topological space, a family $\mathcal{S}$ is said to be a **sub-base** *for the topology τ if it is the case that $\tau = \tau(\mathcal{S})$.*

Proof : For (a), we may simply define

$$\tau(\mathcal{S}) \;=\; \cap\{\tau' : \tau' \text{ is a topology and } \mathcal{S} \subset \tau'\} ,$$

and note that this does the job.

As for (b), it is clear that the family $\mathcal{B}$, as defined in (b), covers X and is closed under finite intersections; consequently, if we define $\tau = \{\cup_{B \in \mathcal{B}_0} B : \mathcal{B}_0 \subset \mathcal{B}\}$, it may be verified that τ is a topology on X for which $\mathcal{B}$ is a base; on the other hand, it is clear from the construction (and the definition of a topology) that if τ' is any topology which contains $\mathcal{S}$, then τ' must necessarily contain τ, and we find that, hence, $\tau = \tau(\mathcal{S})$. □

The usefulness of sub-bases may be seen as follows: very often, in wanting to define a topology, we will find that it is natural to require that sets belonging to a certain class $\mathcal{S}$ should be open; then the topology we seek is any topology which is at least large as $\tau(\mathcal{S})$, and we will find that this minimal topology is quite a good one to work with, since we know, by the last proposition, precisely what the open sets in this topology look like. In order to make all this precise, as well as for other reasons, we need to discuss the notion of continuity, in the context of topological spaces.

DEFINITION A.3.8 *A function $f : X \to Y$ between topological spaces is said to be:*

(a) continuous at the point $x \in X$, if $f^{-1}(U)$ is an open neighbourhood of x in the topological space X, whenever U is an open neighbourhood of the point $f(x)$ in the topological space Y;

(b) **continuous** *if it is continuous at each $x \in X$, or equivalently, if $f^{-1}(U)$ is an open set in X, whenever U is an open set in Y.*

The proof of the following elementary proposition is left as an exercise for the reader.

PROPOSITION A.3.9 *Let $f : X \to Y$ be a map between topological spaces.*

(1) if $x \in X$, then f is continuous at x if and only if $\{f(x_i) : i \in I\}$ is a net converging to $f(x)$ in Y, whenever $\{x_i : i \in I\}$ is a net converging to x in X;

(2) the following conditions on f are equivalent:

(i) f is continuous;

(ii) $f^{-1}(F)$ is a closed subset of X whenever F is a closed subset of Y;

(iii) $\{f(x_i) : i \in I\}$ is a net converging to $f(x)$ in Y, whenever $\{x_i : i \in I\}$ is a net converging to x in X;

(iv) $f^{-1}(B)$ is open in X, whenever B belongs to some base for the topology on Y;

(v) $f^{-1}(S)$ is open in X, whenever S belongs to some sub-base for the topology on Y.

(3) The composition of continuous maps is continuous; i.e., if $f : X \to Y$ and $g : Y \to Z$ are continuous maps of topological spaces, then $g \circ f : X \to Z$ is continuous.

We ae now ready to illustrate what we meant in our earlier discussion of the usefulness of sub-bases. Typically, we have the following situation in mind: suppose X is just some set, that $\{Y_i : i \in I\}$ is some family of topological spaces, and suppose we have maps $f_i : X \to Y_i, \forall i \in I$. We would want to topologise X in such a way that each of the maps f_i is continuous. This, by itself, is not difficult, since if X is equipped with the discrete topology, any map from X into any topological space would be continuous; but if we wnat to topologise X in an efficient, as well as natural, manner with respect to the requirement that each f_i is continuous, then the method of sub-bases tells us what to do. Let us make all this explicit.

PROPOSITION A.3.10 *Suppose $\{f_i : X \to X_i | i \in I\}$ is a family of maps, and suppose τ_i is a topology on X_i for each $i \in I$. Let $\mathcal{S}_i$ be an arbitrary sub-base for the topology τ_i. Define $\tau = \tau(\mathcal{S})$, where*

$$\mathcal{S} = \{f_i^{-1}(V_i) : V_i \in \mathcal{S}_i, i \in I\} .$$

Then,

(a) f_i is continuous as a mapping from the topological space (X, τ) into the topological space (X_i, τ_i), for each $i \in I$;

(b) the topology τ is the smallest topology on X for which (a) above is valid; and consequently, this topology is independent of the choice of the sub-bases $\mathcal{S}_i, i \in I$ and depends only upon the data $\{f_i, \tau_i, i \in I\}$.

This topology τ on X is called the **weak topology** *induced by the family $\{f_i : i \in I\}$ of maps, and we shall write $\tau = \tau(\{f_i : i \in I\})$.*

The proposition is an immediate consequence of Proposition A.3.9(2)(ii) and Proposition A.3.7.

EXERCISE A.3.11 *Let X, X_i, τ_i, f_i and $\tau = \tau(\{f_i : i \in I\})$ be as in Proposition A.3.10.*

(a) Suppose Z is a topological space and $g : Z \to X$ is a function. Then, show that g is continuous if and only if $f_i \circ g$ is continuous for each $i \in I$.

(b) Show that the family $\mathcal{B} = \{\cap_{j=1}^n f_{i_j}^{-1}(V_{i_j}) : i_1, \cdots, i_n \in I, n \in \mathbb{N}, V_{i_j} \in \tau_{i_j} \ \forall j\}$ is a base for the topology τ.

(c) Show that a net $\{x_\lambda : \lambda \in \Lambda\}$ converges to x in (X, τ) if and only if the net $\{f_i(x_\lambda) : \lambda \in \Lambda\}$ converges to $f_i(x)$ in (X_i, τ_i), for each $i \in I$.

As in a metric space, any subspace (= subset) of a topological space acquires a natural structure of a topological space in the manner indicated in the following exercise.

EXERCISE A.3.12 *Let (Y, τ) be a topological space, and let $X \subset Y$ be a subset. Let $i_{X \to Y} : X \to Y$ denote the inclusion map. Then the* **subspace topology** *on X (or the topology on X induced by the topology τ on Y) is, by definition, the weak topology $\tau(\{i_{X \to Y}\})$. This topology will be denoted by $\tau|_X$.*

(a) Show that $\tau|_X = \{U \cap X : U \in \tau\}$, or equivalently, that a subset $F \subset X$ is closed in $(X, \tau|_X)$ if and only if there exists a closed set F_1 in (Y, τ) such that $F = F_1 \cap X$.

(b) Show that if Z is some topological space and if $f : Z \to X$ is a function, then f is continuous when regarded as a map into the topological space $(X, \tau|_X)$ if and only if it is continuous when regarded as a map into the topological space (Y, τ).

One of the most important special cases of this construction is the **product** of topological spaces. Suppose $\{(X_i, \tau_i) : i \in I\}$ is an arbirary family of topological spaces. Let $X = \prod_{i \in I} X_i$ denote their Cartesian product. Then the **product topology** is, by definition, the weak topology on X induced by the family

$\{\pi_i : X \to X_i | i \in I\}$, where π_i denotes, for each $i \in I$, the natural 'projection' of X onto X_i. We shall denote this product topology by $\prod_{i \in I} \tau_i$. Note that if $\mathcal{B}_i$ is a base for the topology τ_i, then a base for the product topology is given by the family $\mathcal{B}$, where a typical element of $\mathcal{B}$ has the form

$$B = \{x \in X : \pi_i(x) \in B_i \ \forall i \in I_0\} ,$$

where I_0 is an arbitrary finite subset of I and $B_i \in \mathcal{B}_i$ for each $i \in I_0$. Thus, a typical basic open set is prescribed by constraining some *finitely* many co-ordinates to lie in specified basic open sets in the appropriate spaces.

Note that if Z is any set, then maps $F : Z \to X$ are in a 1-1 correspondence with families $\{f_i : Z \to X_i | i \in I\}$ of maps - where $f_i = \pi_i \circ f$; and it follows from Exercise A.3.11 that if Z is a topological space, then the mapping f is continuous precisely when each f_i is continuous. To make sure that you have really understood the definition of the product topology, you are urged to work out the following exercise.

EXERCISE A.3.13 *If (X, τ) is a topological space, and if I is any set, let X^I denote the space of functions $f : I \to X$.*

(a) Show that X^I may be identified with the product $\prod_{i \in I} X_i$, where $X_i = X \ \forall i \in I$.

(b) Let X^I be equipped with the product topology; fix $x_0 \in X$ and show that the set $D = \{f \in X^I : f(i) = x_0$ for all but a finite number of i's $\}$ is dense in X, meaning that if U is any open set in X, then $D \cap U \neq \emptyset$.

(c) If Λ is a directed set, show that a net $\{f_\lambda : \lambda \in \Lambda\}$ in X^I converges to a point $f \in X^I$ if and only if the net $\{f_\lambda(i) : \lambda \in \Lambda\}$ converges to $f(i)$, for each $i \in I$. (In other words, the product topology on X^I is nothing but the topology of 'pointwise convergence'.)

We conclude this section with a brief discussion of 'homeomorphisms'.

DEFINITION A.3.14 *Two topological spaces X and Y are said to be* **homeomorphic** *if there exists continuous functions $f : X \to Y, g : Y \to X$ such that $f \circ g = id_Y$ and $g \circ f = id_X$. A*

homeomorphism *is a map f as above - i.e., a continuous bijection between two topological spaces whose (set-theoretic) inverse is also continuous.*

The reader should observe that requiring that a function $f : X \to Y$ is a homeomorphism is more than just requiring that f is 1-1, onto and continuous; if only so much is required of the function, then the inverse f^{-1} may fail to be continuous; an example of this phenomenon is provided by the function $f : [0, 1) \to \mathbb{T}$ defined by $f(t) = exp(2\pi it)$.

The proof of the following proposition is elementary, and left as an exercise to the reader.

PROPOSITION A.3.15 *Suppose f is a continuous bijective map of a topological space X onto a space Y; then the following conditions are equivalent:*

(i) f is a homeomorphism;

(ii) f is an **open map** *- i.e., if U is an open set in X, then $f(U)$ is an open set in Y;*

(iii) f is a **closed map** *- i.e., if F is a closed set in X, then $f(F)$ is a closed set in Y.*

A.4 Compactness

This section is devoted to a quick review of the theory of compact spaces. For the uninitiated reader, the best way to get a feeling for compactness is by understanding this notion for subsets of the real line. The features of compact subsets of $\mathbb{R}$ (and, more generally, of any $\mathbb{R}^n$) are summarised in the following result.

THEOREM A.4.1 *The following conditions on a subset $K \subset \mathbb{R}^n$ are equivalent:*

(i) K is closed and bounded;

(ii) every sequence in K has a subsequence which converges to some point in K;

(iii) every **open cover** *of K has a finite sub-cover - i.e., if $\{U_i : i \in I\}$ is any collection of open sets such that $K \subset \cup_{i\in I} U_i$, then there exists a finite subcollection $\{U_{i_j} : 1 \leq j \leq n\}$ which still covers K (meaning that $K \subset \cup_{j=1}^n U_{i_j}$).*

(iv) Suppose $\{F_i : i \in I\}$ *is any family of closed subsets which has the* **finite intersection property** *with respect to* K *- i.e.,* $\cap_{i \in I_0} F_i \cap K \neq \emptyset$ *for any finite subset* $I_0 \subset I$*; then it is necessarily the case that* $\cap_{i \in I} F_i \cap K \neq \emptyset$.

A subset $K \subset \mathbb{R}$ *which satisfies the equivalent conditions above is said to be* **compact**. □

We will not prove this here, since we shall be proving more general statements.

To start with, note that the conditions (iii) and (iv) of Theorem A.4.1 make sense for any subset K of a topological space X, and are easily seen to be equivalent (by considering the equation $F_i = X - U_i$); also, while conditions (i) and (ii) makes sense in any topological space, they may not be very strong in general. What we shall see is that conditions (ii) - (iv) are equivalent for any subset of a metric space, and these are equivalent to a stronger version of (i). We begin with the appropriate definition.

DEFINITION A.4.2 *A subset* K *of a topological space is said to be* **compact** *if it satisfies the equivalent conditions (iii) and (iv) of Theorem A.4.1.*

'Compactness' is an *intrinsic property*, as asserted in the following exercise.

EXERCISE A.4.3 *Suppose* K *is a subset of a topological space* (X, τ)*. Then show that* K *is compact when regarded as a subset of the topological space* (X, τ) *if and only if* K *is compact when regarded as a subset of the topological space* $(K, \tau|_K)$ *- see Exercise A.3.12.*

PROPOSITION A.4.4 *(a) Let* $\mathcal{B}$ *be any base for the topology underlying a topological space* X*. Then, a subset* $K \subset X$ *is compact if and only if any cover of* K *by open sets, all of which belong to the base* $\mathcal{B}$*, admits a finite sub-cover.*

(b) A closed subset of a compact set is compact.

Proof : (a) Clearly every 'basic open cover' of a compact set admits a finite subcover. Conversely, suppose $\{U_i : i \in I\}$

is an open cover of a set K which has the property that every cover of K by members of $\mathcal{B}$ admits a finite subcover. For each $i \in I$, find a subfamily $\mathcal{B}_i \subset \mathcal{B}$ such that $U_i = \cup_{B \in \mathcal{B}_i} B$; then $\{B : B \in \mathcal{B}_i, i \in I\}$ is a cover of K by members of $\mathcal{B}$; so there exist $B_1, \cdots, B_n$ in this family such that $K \subset \cup_{i=1}^n B_i$; for each $j = 1, \cdots, n$, by construction, we can find $i_j \in I$ such that $B_j \subset U_{i_j}$; it follows that $K \subset \cup_{j=1}^n U_{i_j}$.

(b) Suppose $C \subset K \subset X$ where K is a compact subset of X and C is a subset of K which is closed in the subspace topology of K. Thus, there exists a closed set $F \subset X$ such that $C = F \cap K$. Now, suppose $\{U_i : i \in I\}$ is an open cover of C. Then $\{U_i : i \in I\} \cup \{X - F\}$ is an open cover of K; so there exists a finite subfamily $I_0 \subset I$ such that $K \subset \cup_{i \in I_0} U_i \cup (X - F)$, which clearly implies that $C \subset \cup_{i \in I_0} U_i$. □

COROLLARY A.4.5 *Let K be a compact subset of a metric space X. Then, for any $\epsilon > 0$, there exists a finite subset $N_\epsilon \subset K$ such that $K \subset \cup_{x \in N_\epsilon} B(x, \epsilon)$, (where, of course, the symbol $B(x, \epsilon) = \{y \in X : d(y, x) < \epsilon\}$ denotes the open ball with centre X and radius ϵ).*

Any set N_ϵ as above is called an ϵ-**net** *for K.*

Proof : The family of all open balls with (arbitrary centres and) radii bounded by $\frac{\epsilon}{2}$, clearly constitutes a base for the topology of the metric space X. (Verify this!) Hence, by the preceding proposition, we can find $y_1, \cdots, y_n \in X$ such that $K \subset \cup_{i=1}^n B(y_i, \frac{\epsilon}{2})$. We may assume, without loss of generality, that K is not contained in any proper sub-family of these n open balls; i.e., we may assume that there exists $x_i \in K \cap B(y_i, \frac{\epsilon}{2})$, $\forall i$. Then, clearly $K \subset \cup_{i=1}^n B(x_i, \epsilon)$, and so $N_\epsilon = \{x_1, \cdots, x_n\}$ is an ϵ-net for K. □

DEFINITION A.4.6 *A subset A of a metric space X is said to be* **totally bounded** *if, for each $\epsilon > 0$, there exists a finite ϵ-net for A.*

Thus, compact subsets of metric spaces are totally bounded. The next proposition provides an alternative criterion for total boundedness.

PROPOSITION A.4.7 *The following conditions on a subset A of a metric space X are equivalent:*

(i) A is totally bounded;

(ii) every sequence in A has a Cauchy subsequence.

Proof : $(i) \Rightarrow (ii)$: Suppose $S_0 = \{x_k\}_{k=1}^{\infty}$ is a sequence in A. We assert that there exist (a) a sequence $\{B_n = B(z_n, 2^{-n})\}$ of open balls (with shrinking radii) in X; and (b) sequences $S_n = \{x_k^{(n)}\}_{k=1}^{\infty}$, $n = 1, 2, \cdots$ with the property that $S_n \subset B_n$ and S_n is a subsequence of S_{n-1}, for all $n \geq 1$.

We construct the B_n's and the S_n's inductively. To start with, pick a finite $\frac{1}{2}$-net $N_{\frac{1}{2}}$ for A; clearly, there must be some $z_1 \in N_{\frac{1}{2}}$ with the property that $x_k \in B(z_1, \frac{1}{2})$ for infinitely many values of k; define $B_1 = B(z_1, \frac{1}{2})$ and choose S_1 to be any subsequence $\{x_k^{(1)}\}_{k=1}^{\infty}$ of S_0 with the property that $x_k^{(1)} \in B_1 \ \forall k$.

Suppose now that open balls $B_j = B(z_j, 2^{-j})$ and sequences $S_j, 1 \leq j \leq n$ have been chosen so that S_j is a subsequence of S_{j-1} which lies in $B_j \ \forall 1 \leq j \leq n$.

Now, let $N_{2^{-(n+1)}}$ be a finite $2^{-(n+1)}$-net for A; as before, we may argue that there must exist some $z_{n+1} \in N_{2^{-(n+1)}}$ with the property that if $B_{n+1} = B(z_{n+1}, 2^{-(n+1)})$, then $x_k^{(n)} \in B_{n+1}$ for infinitely many values of k; we choose S_{n+1} to be a subsequence of S_n which lies entirely in B_{n+1}.

Thus, by induction, we can conclude the existence of sequences of open balls B_n and sequences S_n with the asserted properties. Now, define $y_n = x_n^{(n)}$ and note that (i) $\{y_n\}$ is a subsequence of the initial sequence S_0, and (ii) if $n, m \geq k$, then $y_n, y_m \in B_k$ and hence $d(y_n, y_m) < 2^{1-k}$; this is the desired Cauchy subsequence.

$(ii) \Rightarrow (i)$: Suppose A is not totally bounded; then there exists some $\epsilon > 0$ such that A admits no finite ϵ-net; this means that given any finite subset $F \subset A$, there exists some $a \in A$ such that $d(a, x) \geq \epsilon \ \forall \ x \in F$. Pick $x_1 \in A$; (this is possible, since we may take $F = \emptyset$ in the previous sentence); then pick $x_2 \in A$ such that $d(x_1, x_2) \geq \epsilon$; (this is possible, by setting $F = \{x_1\}$ in the previous sentence); then (set $F = \{x_1, x_2\}$ in the previous sentence and) pick $x_3 \in A$ such that $d(x_3, x_i) \geq \epsilon$, $i = 1, 2$; and so on. This yields a sequence $\{x_n\}_{n=1}^{\infty}$ in A such that $d(x_n, x_m) \geq \epsilon \ \forall \ n \neq m$. This sequence clearly has no

Cauchy subsequence, thereby contradicting the assumption (ii), and the proof is complete. □

COROLLARY A.4.8 *If A is a totally bounded set in a metric space, so also are its closure $\overline{A}$ and any subset of A.*

REMARK A.4.9 The argument to prove $(i) \Rightarrow (ii)$ in the above theorem has a very useful component which we wish to single out; starting with a sequence $S = \{x_k\}$, we proceeded, by some method - which method was dictated by the specific purpose on hand, and is not relevant here - to construct sequences $S_n = \{x_k^{(n)}\}$ with the property that each S_{n+1} was a subsequence of S_n (and with some additional desirable properties); we then considered the sequence $\{x_n^{(n)}\}_{n=1}^{\infty}$. This process is sometimes referred to as the **diagonal argument**.

EXERCISE A.4.10 *(1) Show that any totally bounded set in a metric space is bounded, meaning that it is a subset of some open ball.*

(2) Show that a subset of $\mathbb{R}^n$ is bounded if and only if it is totally bounded. (Hint: by Corollary A.4.8 and (1) above, it is enough to establish the total-boundedness of the set given by $K = \{x = (x_1, \cdots, x_n) : |x_i| \leq N \ \forall i\}$; given ϵ, pick k such that the diameter of an n-cube of side $\frac{2}{k}$ is smaller than ϵ, and consider the points of the form $(\frac{i_1}{k}, \frac{i_2}{k}, \cdots, \frac{i_n}{k})$ in order to find an ϵ-net.)

We are almost ready to prove the announced equivalence of the various conditions for compactness; but first, we need a technical result which we take care of before proceeding to the desired assertion.

LEMMA A.4.11 *The following conditions on a metric space X are equivalent:*

(i) X is **separable** *- i.e., there exists a countable set D which is dense in X (meaning that $X = \overline{D}$);*

(ii) X satisfies the **second axiom of countability** *- meaning that there is a countable base for the topology of X.*

Proof : $(i) \Rightarrow (ii)$: Let $\mathcal{B} = \{B(x,r) : x \in D, \ 0 < r \in \mathbb{Q}\}$ where D is some countable dense set in X, and $\mathbb{Q}$ denotes the (countable) set of rational numbers. We assert that this is a base for the topology on X; for, if U is an open set and if $y \in U$, we may find $s > 0$ such that $B(y,s) \subset U$; pick $x \in D \cap B(y, \frac{s}{2})$ and pick a positive rational number r such that $d(x,y) < r < \frac{s}{2}$ and note that $y \in B(x,r) \subset U$; thus, for each $y \in U$, we have found a ball $B_y \in \mathcal{B}$ such that $y \in B_y \subset U$; hence $U = \cup_{y \in U} B_y$.

$(ii) \Rightarrow (i)$: If $\mathcal{B} = \{B_n\}_{n=1}^\infty$ is a countable base for the topology of X, (where we may assume that each B_n is a non-empty set, without loss of generality), pick a point $x_n \in B_n$ for each n and verify that $D = \{x_n\}_{n=1}^\infty$ is indeed a countable dense set in X. □

THEOREM A.4.12 *The following conditions on a subset K of a metric space X are equivalent:*

(i) K is compact;

(ii) K is complete and totally bounded;

(iii) every sequence in K has a subsequence which converges to some point of K.

Proof : $(i) \Rightarrow (ii)$: We have already seen that compactness implies total boundedness. Suppose now that $\{x_n\}$ is a Cauchy sequence in K. Let F_n be the closure of the set $\{x_m : m \geq n\}$, for all $n = 1, 2, \cdots$. Then $\{F_n\}_{n=1}^\infty$ is clearly a decreasing sequence of closed sets in X; further, the Cauchy criterion implies that $diam \ F_n \ \to 0$ as $n \to \infty$. By invoking the finite intersection property, we see that there must exist a point $x \in \cap_{n=1}^\infty F_n \cap K$. Since the diameters of the F_n's shrink to 0, we may conclude that (such an x is unique and that) $x_n \to x$ as $n \to \infty$.

$(ii) \Leftrightarrow (iii)$: This is an immediate consequence of Proposition A.4.7.

$(ii) \Rightarrow (i)$: For each $n = 1, 2, \cdots$, let $N_{\frac{1}{n}}$ be a $\frac{1}{n}$-net for K. It should be clear that $D = \cup_{n=1}^\infty N_{\frac{1}{n}}$ is a countable dense set in K. Consequently K is separable; hence, by Lemma A.4.11, K admits a countable base $\mathcal{B}$.

In order to check that K is compact in X, it is enough, by Exercise A.4.3, to check that K is compact in itself. Hence, by Proposition A.4.4, it is enough to check that any countable open cover of X has a finite sub-cover; by the reasoning that goes in to establish the equivalence of conditions (iii) and (iv) in Theorem A.4.1, we have, therefore, to show that if $\{F_n\}_{n=1}^{\infty}$ is a sequence of closed sets in K such that $\cap_{n=1}^{N} F_n \neq \emptyset$ for each $N = 1, 2, \cdots$, then necessarily $\cap_{n=1}^{\infty} F_n \neq \emptyset$. For this, pick a point $x_N \in \cap_{n=1}^{N} F_n$ for each N; appeal to the hypothesis to find a subsequence $\{y_n\}$ of $\{x_n\}$ such that $y_n \to x \in K$, for some point $x \in K$. Now notice that $\{y_m\}_{m=N}^{\infty}$ is a sequence in $\cap_{n=1}^{N} F_n$ which converges to x; conclude that $x \in F_n \forall n$; hence $\cap_{n=1}^{\infty} F_n \neq \emptyset$. □

In view of Exercise A.4.10(2), we see that Theorem A.4.12 does indeed generalise Theorem A.4.1.

We now wish to discuss compactness in general topological spaces. We begin with some elementary results.

Proposition A.4.13 *(a) Suppose $f : X \to Y$ is a continuous map of topological spaces; if K is a compact subset of X, then $f(K)$ is a compact subset of Y; in other words, a continuous image of a compact set is compact.*

(b) If $f : K \to \mathbb{R}$ is continuous, and if K is a compact set in X, then (i) $f(K)$ is bounded, and (ii) there exist points $x, y \in K$ such that $f(x) \leq f(z) \leq f(y) \ \forall \ z \in K$; in other words, a continuous real-valued function on a compact set is bounded and attains its bounds.

Proof : (a) If $\{U_i : i \in I\}$ is an open cover of $f(K)$, then $\{f^{-1}(U_i) : i \in I\}$ is an open cover of K; if $\{f^{-1}(U_i) : i \in I_0\}$ is a finite sub- cover of K, then $\{U_i : i \in I_0\}$ is a finite sub- cover of $f(K)$.

(b) This follows from (a), since compact subsets of $\mathbb{R}$ are closed and bounded (and hence contain their supremum and infimum). □

The following result is considerably stronger than Proposition A.4.4(a).

Theorem A.4.14 *(**Alexander sub-base theorem**)*

Let $\mathcal{S}$ be a sub-base for the topology τ underlying a topological space X. Then a subspace $K \subset X$ is compact if and only if any open cover of K by a sub-family of $\mathcal{S}$ admits a finite sub-cover.

Proof : Suppose $\mathcal{S}$ is a sub-base for the topology of X and suppose that any open cover of K by a sub-family of $\mathcal{S}$ admits a finite sub-cover. Let $\mathcal{B}$ be the base generated by $\mathcal{S}$ - i.e., a typical element of $\mathcal{B}$ is a finite intersection of members of $\mathcal{S}$. In view of Proposition A.4.4(a), the theorem will be proved once we establish that any cover of X by members of $\mathcal{B}$ admits a finite sub-cover.

Assume this is false; thus, suppose $\mathcal{U}_0 = \{B_i : i \in I_0\} \subset \mathcal{B}$ is an open cover of X, which does not admit a finite subcover. Let $\mathcal{P}$ denote the set of all subfamilies $\mathcal{U} \subset \mathcal{U}_0$ with the property that no finite sub-family of $\mathcal{U}$ covers X. (Notice that $\mathcal{U}$ is an open cover of X, since $\mathcal{U} \subset \mathcal{U}_0$.) It should be obvious that $\mathcal{P}$ is partially ordered by inclusion, and that $\mathcal{P} \neq \emptyset$ since $\mathcal{U} \in \mathcal{P}$. Suppose $\{\mathcal{U}_i : i \in I\}$ is a totally ordered subset of $\mathcal{P}$; we assert that then $\mathcal{U} = \cup_{i\in I}\mathcal{U}_i \in \mathcal{P}$. (Reason: Clearly $\mathcal{U}_0 \subset \cup_{i\in I}\mathcal{U}_i \subset \mathcal{B}$; further, suppose $\{B_1, \cdots, B_n\} \subset \cup_{i\in I}\mathcal{U}_i$; the assumption of the family $\{\mathcal{U}_i\}$ being totally ordered then shows that in fact there must exist some $i \in I$ such that $\{B_1, \cdots, B_n\} \subset \mathcal{U}_i$; by definition of $\mathcal{P}$, it cannot be the case that $\{B_1, \cdots, B_n\}$ covers X; thus, we have shown that no finite sub-family of $\mathcal{U}$ covers X. Hence, Zorn's lemma is applicable to the partially ordered set $\mathcal{P}$.

Thus, if the theorem were false, it would be possible to find a family $\mathcal{U} \subset \mathcal{B}$ which is an open cover of X, and further has the following properties: (i) no finite sub-family of $\mathcal{U}$ covers X; and (ii) $\mathcal{U}$ is a maximal element of $\mathcal{P}$ - which means that whenever $B \in \mathcal{B} - \mathcal{U}$, there exists a finite subfamily $\mathcal{U}_B$ of $\mathcal{U}$ such that $\mathcal{U}_B \cup \{B\}$ is a (finite) cover of X.

Now, by definition of $\mathcal{B}$, each element B of $\mathcal{U}$ has the form $B = S_1 \cap S_2 \cdots \cap S_n$, for some $S_1, \cdots, S_n \in \mathcal{S}$. Assume for the moment that none of the S_i's belongs to $\mathcal{U}$. Then property (ii) of $\mathcal{U}$ (in the previous paragraph) implies that for each $1 \leq i \leq n$, we can find a finite subfamily $\mathcal{U}_i \subset \mathcal{U}$ such that $\mathcal{U}_i \cup \{S_i\}$ is a finite open cover of X. Since this is true for all i, this implies that $\cup_{i=1}^n \mathcal{U}_i \cup \{B = \cap_{i=1}^n S_i\}$ is a finite open cover of X; but since $B \in \mathcal{U}$, we have produced a finite subcover of X from $\mathcal{U}$, which contradicts the defining property (i) (in the previous paragraph) of $\mathcal{U}$. Hence at least one of the S_i's must belong to $\mathcal{U}$.

Thus, we have shown that if $\mathcal{U}$ is as above, then each $B \in \mathcal{U}$ is of the form $B = \cap_{i=1}^n S_i$, where $S_i \in \mathcal{S}\ \forall i$ and further there exists at least one i_0 such that $S_{i_0} \in \mathcal{U}$. The passage $B \mapsto S_{i_0}$ yields a mapping $\mathcal{U} \ni B \mapsto S(B) \in \mathcal{U} \cap \mathcal{S}$ with the property that $B \subset S(B)$ for all $B \in \mathcal{U}$. Since $\mathcal{U}$ is an open cover of X by definition, we find that $\mathcal{U} \cap \mathcal{S}$ is an open cover of X by members of $\mathcal{S}$, which admits a finite sub-cover, by hypothesis.

The contradiction we have reached is a consequence of our assuming that there exists an open cover $\mathcal{U}_0 \subset \mathcal{B}$ which does not admit a finite sub-cover. The proof of the theorem is finally complete. □

We are now ready to prove the important result, due to Tychonoff, which asserts that if a family of topological spaces is compact, then so is their topological product.

THEOREM A.4.15 (**Tychonoff's theorem**)

Suppose $\{(X_i, \tau_i) : i \in I\}$ is a family of non-empty topological spaces. Let (X, τ) denote the product space $X = \prod_{i \in I} X_i$, equipped with the product topology. Then X is compact if and only if each X_i is compact.

Proof : Since $X_i = \pi_i(X)$, where π_i denotes the (continuous) projection onto the i-th co-ordinate, it follows from Proposition A.4.13(a) that if X is compact, so is each X_i.

Suppose conversely that each X_i is compact. For a subset $A \subset X_i$, let $A^i = \pi_i^{-1}(A)$. Thus, by the definition of the product topology, $\mathcal{S} = \{U^i : U \in \tau_i\}$ is a sub-base for the product topology τ. Thus, we need to prove the following: if J is any set, if $J \ni j \mapsto i(j) \in I$ is a map, if $A(i(j))$ is a closed set in $X_{i(j)}$ for each $j \in J$, and if $\cap_{j \in F} A(i(j))^{i(j)} \neq \emptyset$ for every finite subset $F \subset J$, then it is necessarily the case that $\cap_{j \in J} A(i(j))^{i(j)} \neq \emptyset$.

Let $I_1 = \{i(j) : j \in J\}$ denote the range of the mapping $j \mapsto i(j)$. For each fixed $i \in I_1$, let $J_i = \{j \in J : i(j) = i\}$; observe that $\{A(i(j)) : j \in J_i\}$ is a family of closed sets in X_i; we assert that $\cap_{j \in F} A(i(j)) \neq \emptyset$ for every finite subset $F \subset J_i$. (Reason: if this were empty, then, $\cap_{j \in F} A(i(j))^{i(j)} = (\cap_{j \in F} A(i(j)))^i$ would have to be empty, which would contradict the hypothesis.) We conclude from the compactness of X_i that $\cap_{j \in J_i} A(i(j)) \neq \emptyset$. Let x_i be any point from this intersection. Thus $x_i \in A(i(j))$ whenever $j \in J$ and $i(j) = i$. For $i \in I - I_1$, pick an arbitrary point $x_i \in X_i$. It is now readily verified that if $x \in X$ is the point such that $\pi_i(x) = x_i \ \forall i \in I$, then $x \in A(i(j))^{i(j)} \ \forall j \in J$, and the proof of the theorem is complete. $\Box$

Among topological spaces, there is an important subclass of spaces which exhibit some pleasing features (in that certain pathological situations cannot occur). We briefly discuss these spaces now.

DEFINITION A.4.16 *A topological space is called a* **Hausdorff space** *if, whenever x and y are two distinct points in X, there exist open sets U, V in X such that $x \in U, y \in V$ and $U \cap V = \emptyset$. Thus, any two distinct points can be 'separated' (by a pair of disjoint open sets). (For obvious reasons, the preceding requirement is sometimes also referred to as the Hausdorff separation axiom.)*

EXERCISE A.4.17 *(1) Show that any metric space is a Hausdorff space.*

(2) Show that if $\{f_i : X \to X_i : i \in I\}$ is a family of functions from a set X into topological spaces (X_i, τ_i), and if each (X_i, τ_i) is a Hausdorff space, then show that the weak topology $\tau(\{f_i : i \in I\})$ on X (which is induced by this family of maps) is a Hausdorff topology if and only if the f_i's separate points - meaning that whenever x and y are distinct points in X, then there exists an index $i \in I$ such that $f_i(x) \neq f_i(y)$.

(3) Show that if $\{(X_i, \tau_i) : i \in I\}$ is a family of topological spaces, then the topological product space $(\prod_{i \in I} X_i, \prod_{i \in I} \tau_i)$ is a Hausdorff space if and only if each (X_i, τ_i) is a Hausdorff space.

(4) Show that a subspace of a Hausdorff space is also a Hausdorff space (with respect to the subspace topology).

(5) Show that Exercise (4), as well as the 'only if' assertion of Exercise (3) above, are consequences of Exercise (2).

We list some elementary properties of Hausdorff spaces in the following Proposition.

PROPOSITION A.4.18 *(a) If (X, τ) is a Hausdorff space, then every finite subset of X is closed.*

(b) A topological space is a Hausdorff space if and only if 'limits of convergent nets are unique' - i.e., if and only if, whenever $\{x_i : i \in I\}$ is a net in X, and if the net converges to both $x \in X$ and $y \in X$, then $x = y$.

(c) Suppose K is a compact subset of a Hausdorff space X and suppose $y \notin K$; then there exist open sets U, V in X such that $K \subset U, y \in V$ and $U \cap V = \emptyset$; in particular, a compact subset of a Hausdorff space is closed.

(d) If X is a compact Hausdorff space, and if C and K are closed subsets of X which are disjoint - i.e., if $C \cap K = \emptyset$ - then there exist a pair of disjoint open sets U, V in X such that $K \subset U$ and $C \subset V$.

Proof : (a) Since finite unions of closed sets are closed, it is enough to prove that $\{x\}$ is closed, for each $x \in X$; the Hausdorff separation axiom clearly implies that $X - \{x\}$ is open.

(b) Suppose X is Hausdorff, and suppose a net $\{x_i : i \in I\}$ converges to $x \in X$ and suppose $x \neq y \in X$. Pick open sets U, V as in Definition A.4.16; by definition of convergence of a net, we can find $i_0 \in I$ such that $x_i \in U \; \forall i \geq i_0$; it follows then that $x_i \notin V \; \forall i \geq i_0$, and hence the net $\{x_i\}$ clearly does not converge to y.

Conversely, suppose X is not a Hausdorff space; then there exists a pair x, y of distinct points which cannot be separated. Let $\mathcal{N}(x)$ (resp., $\mathcal{N}(y)$) denote the directed set of open neighbourhoods of the point x (resp., y) - see Example 2.2.4(3). Let $I = \mathcal{N}(x) \times \mathcal{N}(y)$ be the directed set obtained from the Cartesian product as in Example 2.2.4(4). By the assumption that x and y cannot be separated, we can find a point $x_i \in U \cap V$ for each $i = (U, V) \in I$. It is fairly easily seen that the net $\{x_i : i \in I\}$ simultaneously converges to both x and y.

(c) Suppose K is a compact subset of a Hausdorff space X. Fix $y \notin K$; then, for each $x \in K$, find open sets U_x, V_x so that $x \in U_x, y \in V_x$ and $U_x \cap V_x = \emptyset$; now the family $\{U_x : x \in K\}$ is an open cover of the compact set K, and we can hence find $x_1, \cdots, x_n \in K$ such that $K \subset U = \cup_{i=1}^n U_{x_i}$; conclude that if $V = \cap_{i=1}^n V_{x_i}$, then V is an open neighbourhood of y such that $U \cap V = \emptyset$; and the proof of (c) is complete.

(d) Since closed subsets of compact spaces are compact, we see that C and K are compact. Hence, by (c) above, we may, for each $y \in C$, find a pair U_y, V_y of disjoint open subsets of X such that $K \subset U_y, y \in V_y$, and $U_y \cap V_y = \emptyset$. Now, the family $\{V_y : y \in C\}$ is an open cover of the compact set C, and hence we can find $y_1, \cdots, y_n \in C$ such that if $V = \cup_{i=1}^n V_{y_i}$ and $U = \cap_{i=1}^n U_{y_i}$, then U and V are open sets which satisfy the assertion in (d). □

COROLLARY A.4.19 *(1) A continuous mapping from a compact space to a Hausdorff space is a closed map (see Proposition A.3.15).*

(2) If $f : X \to Y$ is a continuous map, if Y is a Hausdorff space, and if $K \subset X$ is a compact subset such that $f|_K$ is 1-1, then $f|_K$ is a homeomorphism of K onto $f(K)$.

Proof : (1) Closed subsets of compact sets are compact; continuous images of compact sets are compact; and compact subsets of Hausdorff spaces are closed.

(2) This follows directly from (1) and Proposition A.3.15. □

Proposition A.4.18 shows that in compact Hausdorff spaces, disjoint closed sets can be separated (just like points). There are other spaces which have this property; metric spaces, for instance, have this property, as shown by the following exercise.

EXERCISE A.4.20 *Let (X, d) be a metric space. Given a subset $A \subset X$, define the distance from a point x to the set A by the formula*

$$d(x, A) \;=\; \inf\{d(x, a) : a \in A\} \; ; \qquad \text{(A.4.11)}$$

(a) Show that $d(x, A) = 0$ if and only if x belongs to the closure $\overline{A}$ of A.

(b) Show that

$$|d(x, A) \;-\; d(y, A)| \;\leq\; d(x, y) \;,\; \forall x, y \in X \;,$$

and hence conclude that $d : X \to \mathbb{R}$ is a continuous function.

(c) If A and B are disjoint closed sets in X, show that the function defined by

$$f(x) \;=\; \frac{d(x, A)}{d(x, A) + d(x, B)}$$

is a (meaningfully defined and uniformly) continuous function $f : X \to [0,1]$ *with the property that*

$$f(x) \;=\; \begin{cases} 0 & if\ x \in A \\ 1 & if\ x \in B \end{cases} \quad . \tag{A.4.12}$$

(d) Deduce from (c) that disjoint closed sets in a metric space can be separated (by disjoint open sets). (Hint: if the closed sets are A *and* B*, consider the set* U *(resp.,* V*) of points where the function* f *of (c) is 'close to 0' (resp., 'close to 1').*

The preceding exercise shows, in addition to the fact that disjoint closed sets in a metric space can be separated by disjoint open sets, that there exists lots of continuous real-valued functions on a metric space. It is a fact that the two notions are closely related. To get to this fact, we first introduce a convenient notion, and then establish the relevant theorem.

Definition A.4.21 *A topological space* X *is said to be* **normal** *if (a) it is a Hausdorff space, and (b) whenever* A, B *are closed sets in* X *such that* $A \cap B = \emptyset$*, it is possible to find open sets* U, V *in* X *such that* $A \subset U, B \subset V$ *and* $U \cap V = \emptyset$*.*

The reason that we had to separately assume the Hausdorff condition in the definition given above is that the Hausdorff axiom is not a consequence of condition (b) of the preceding definition. (For example, let $X = \{1, 2\}$ and let $\tau = \{\emptyset, \{1\}, X\}$; then τ is indeed a non-Hausdorff topology, and there do not exist a pair of non-empty closed sets which are disjoint from one another, so that condition (b) is vacuously satisfied.)

We first observe an easy consequence of normality, and state it as an exercise.

Exercise A.4.22 *Show that a Hausdorff space* X *is normal if and only if it satisfies the following condition: whenever* $A \subset W \subset X$*, where* A *is closed and* W *is open, then there exists an open set* U *such that* $A \subset U \subset \overline{U} \subset W$*. (Hint: consider* $B = X - W$*.)*

Thus, we find - from Proposition A.4.18(d) and Exercise A.4.20(d) - that compact Hausdorff spaces and metric spaces are examples of normal spaces. One of the results relating normal spaces and continuous functions is the useful *Urysohn's lemma* which we now establish.

THEOREM A.4.23 (**Urysohn's lemma**)
Suppose A and B are disjoint closed subsets of a normal space X; then there exists a continuous function $f : X \to [0,1]$ such that

$$f(x) \;=\; \begin{cases} 0 & if\ x \in A \\ 1 & if\ x \in B \end{cases} \;. \tag{A.4.13}$$

Proof : Write $\mathbb{Q}_2$ for the set of 'dyadic rational numbers' in [0,1]. Thus, $\mathbb{Q}_2 \;=\; \{\frac{k}{2^n} : n = 0, 1, 2, \cdots, 0 \leq k \leq 2^n\}$.

Assertion : There exist open sets $\{U_r : r \in \mathbb{Q}_2\}$ such that
(i) $A \subset U_0,\ U_1 = X - B$; and
(ii)

$$r, s \in \mathbb{Q}_2, r < s \;\Rightarrow\; \overline{U_r} \subset U_s \;. \tag{A.4.14}$$

Proof of assertion : , Define $\mathbb{Q}_2(n) \;=\; \{\frac{k}{2^n} : 0 \leq k \leq 2^n\}$, for $n = 0, 1, 2, \cdots$. Then, we clearly have $\mathbb{Q}_2 \;=\; \cup_{n=0}^{\infty} \mathbb{Q}_2(n)$. We shall use induction on n to construct $U_r, r \in \mathbb{Q}_2(n)$.

First, we have $\mathbb{Q}_2(0) \;=\; \{0, 1\}$. Define $U_1 = X - B$, and appeal to Exercise A.4.22 to find an open set U_0 such that $A \subset U_0 \subset \overline{U_0} \subset U_1$.

Suppose that we have constructed open sets $\{U_r : r \in \mathbb{Q}_2(n)\}$ such that the condition A.4.14 is satisfied whenever $r, s \in \mathbb{Q}_2(n)$. Notice now that if $t \in \mathbb{Q}_2(n+1) - \mathbb{Q}_2(n)$, then $t = \frac{2m+1}{2^{n+1}}$ for some unique integer $0 \leq m < 2^n$. Set $r = \frac{2m}{2^{n+1}}, s = \frac{2m+2}{2^{n+1}}$, and note that $r < t < s$ and that $r, s \in \mathbb{Q}_2(n)$. Now apply Exercise A.4.22 to the inclusion $\overline{U_r} \subset U_s$ to deduce the existence of an open set U_t such that $\overline{U_r} \subset U_t \subset \overline{U_t} \subset U_s$. It is a simple matter to verify that the condition A.4.14 is satisfied now for all $r, s \in \mathbb{Q}_2(n+1)$; and the proof of the assertion is also complete.

Now define the function $f : X \to [0,1]$ by the following prescription:

$$f(x) = \begin{cases} \inf\{t \in \mathbb{Q}_2 : x \in U_t\} & if\ x \in U_1 \\ 1 & if\ x \notin U_1 \end{cases}$$

This function is clearly defined on all of X, takes values in [0,1], is identically equal to 0 on A, and is identically equal to 1 on B. So we only need to establish the continuity of f, in order to complete the proof of the theorem. We check continuity of f at a point $x \in X$; suppose $\epsilon > 0$ is given. (In the following proof, we will use the (obvious) fact that $\mathbb{Q}_2$ is dense in [0,1].)

Case (i) : $f(x) = 0$: In this case, pick $r \in \mathbb{Q}_2$ such that $r < \epsilon$. The assumption that $f(x) = 0$ implies that $x \in U_r$; also $y \in U_r \Rightarrow f(y) \leq r < \epsilon$.

Case(ii) : $0 < f(x) < 1$. First pick $p, t \in \mathbb{Q}_2$ such that $f(x) - \epsilon < p < f(x) < t < f(x) + \epsilon$. By the definition of f, we can find $s \in \mathbb{Q}_2 \cap (f(x), t)$ such that $x \in U_s$. Then pick $r \in \mathbb{Q}_2 \cap (p, f(x))$ and observe that $f(x) > r \Rightarrow x \notin U_r \Rightarrow x \notin \overline{U_p}$. Hence we see that $V = U_s - \overline{U_p}$ is an open neighbourhood of x; it is also easy to see that $y \in V \ \Rightarrow p \leq f(y) \leq s$ and hence $y \in V \Rightarrow |f(y) - f(x)| < \epsilon$.

Case (iii) : $f(x) = 1$: The proof of this case is similar to part of the proof of Case (ii), and is left as an exercise for the reader. □

We conclude this section with another result concerning the existence of 'sufficiently many' continuous functions on a normal space.

Theorem A.4.24 (**Tietze's extension theorem**)

Suppose $f : A \to [-1,1]$ is a continuous function defined on a closed subspace A of a normal space X. Then there exists a continuous function $F : X \to [-1,1]$ such that $F|_A = f$.

Proof : Let us set $f = f_0$ (for reasons that will soon become clear).

Let $A_0 = \{x \in A : f_0(x) \leq -\frac{1}{3}\}$ and $B_0 = \{x \in A : f_0(x) \geq \frac{1}{3}\}$; then A_0 and B_0 are disjoint sets which are closed in A and hence also in X. By Urysohn's lemma, we can find a continuous function $g_0 : X \to [-\frac{1}{3}, \frac{1}{3}]$ such that $g_0(A_0) = \{-\frac{1}{3}\}$ and

$g_0(B_0) = \{\frac{1}{3}\}$. Set $f_1 = f_0 - g_0|_A$ and observe that $f_1 : A \to [-\frac{2}{3}, \frac{2}{3}]$.

Next, let $A_1 = \{x \in A : f_1(x) \leq -\frac{1}{3} \cdot \frac{2}{3}\}$ and $B_1 = \{x \in A : f_1(x) \geq \frac{1}{3} \cdot \frac{2}{3}\}$; and as before, construct a continuous function $g_1 : X \to [-\frac{1}{3} \cdot \frac{2}{3}, \frac{1}{3} \cdot \frac{2}{3}]$ such that $g_1(A_1) = \{-\frac{1}{3} \cdot \frac{2}{3}\}$ and $g_1(B_1) = \{\frac{1}{3} \cdot \frac{2}{3}\}$. Then define $f_2 = f_1 - g_1|_A = f_0 - (g_0 + g_1)|_A$, and observe that $f_2 : A \to [-(\frac{2}{3})^2, (\frac{2}{3})^2]$.

Repeating this argument indefinitely, we find that we can find (i) a sequence $\{f_n\}_{n=0}^{\infty}$ of continuous functions on A such that $f_n : A \to [-(\frac{2}{3})^n, (\frac{2}{3})^n]$, for each n; and (ii) a sequence $\{g_n\}_{n=0}^{\infty}$ of continuous functions on X such that $g_n(A) \subset [-\frac{1}{3} \cdot (\frac{2}{3})^n, \frac{1}{3} \cdot (\frac{2}{3})^n]$, for each n; and such that these sequences satisfy

$$f_n = f_0 - (g_0 + g_1 + \cdots + g_{n-1})|_A \ .$$

The series $\sum_{n=0}^{\infty} g_n$ is absolutely summable, and consequently, summable in the Banach space $C_b(X)$ of all bounded continuous functions on X. Let F be the limit of this sum. Finally, the estimate we have on f_n (see (i) above) and the equation displayed above show that $F|_A = f_0 = f$ and the proof is complete. □

EXERCISE A.4.25 *(1) Show that Urysohn's lemma is valid with the unit interval [0,1] replaced by any closed interval* $[a, b]$, *and thus justify the manner in which Urysohn's lemma was used in the proof of Tietze's extension theorem. (Hint: use appropriate 'affine maps' to map any closed interval homeomorphically onto any other closed interval.)*

(2) Show that Tietze's extension theorem remains valid if [0,1] is replaced by (a) $\mathbb{R}$, *(b)* $\mathbb{C}$, *(c)* $\mathbb{R}^n$.

A.5 Measure and integration

In this section, we review some of the basic aspects of measure and integration. We begin with the fundamental notion of a σ-algebra of sets.

DEFINITION A.5.1 *A* σ**-algebra** *of subsets of a set X is, by definition, a collection $\mathcal{B}$ of subsets of X which satisfies the following requirements:*

(a) $X \in \mathcal{B}$;

(b) $A \in \mathcal{B} \;\Rightarrow\; A^c \in \mathcal{B}$, *where* $A^c = X - A$ *denotes the complement of the set* A*; and*

(c) $\{A_n\}_{n=1}^{\infty} \subset \mathcal{B} \;\Rightarrow\; \cup_n A_n \in \mathcal{A}$.

A **measurable space** *is a pair* $(X, \mathcal{B})$ *consisting of a set* X *and a* σ*-algebra* $\mathcal{B}$ *of subsets of* X.

Thus, a σ-algebra is nothing but a collection of sets which contains the whole space and is closed under the formation of complements and countable unions. Since $\cap_n A_n = (\cup_n A_n^c)^c$, it is clear that a σ-algebra is closed under the formation of countable intersections; further, every σ-algebra always contains the empty set (since $\emptyset = X^c$).

EXAMPLE A.5.2 (1) The collection 2^X of all subsets of X is a σ-algebra, as is the two-element collection $\{\emptyset, X\}$. These are called the 'trivial' σ-algebras; if $\mathcal{B}$ is any σ-algebra of subsets of X, then $\{\emptyset, X\} \subset \mathcal{B} \subset 2^X$.

(2) The intersection of an arbitrary collection of σ-algebras of subsets of X is also a σ-algebra of subsets of X. It follows that if $\mathcal{S}$ is any collection of subsets of X, and if we set $\mathcal{B}(\mathcal{S}) = \cap\{\mathcal{B} : \mathcal{S} \subset \mathcal{B}, \mathcal{B}$ is a σ-algebra of subsets of $X\}$, then $\mathcal{B}(\mathcal{S})$ is a σ-algebra of subsets of X, and it is the smallest σ-algebra of subsets of X which contains the initial family $\mathcal{S}$ of sets; we shall refer to $\mathcal{B}(\mathcal{S})$ as the σ-algebra generated by $\mathcal{S}$.

(3) If X is a topological space, we shall write $\mathcal{B}_X = \mathcal{B}(\tau)$, where τ is the family of all open sets in X; members of this σ-algebra are called **Borel sets** and $\mathcal{B}_X$ is called the **Borel σ-algebra** of X.

(4) Suppose $\{(X_i, \mathcal{B}_i) : i \in I\}$ is a family of measurable spaces. Let $X = \prod_{i \in I} X_i$ denote the Cartesian product, and let $\pi_i : X \to X_i$ denote the i-th coordinate projection, for each $i \in I$. Then the **product σ-algebra** $\mathcal{B} = \prod_{i \in I} \mathcal{B}_i$ is the σ-algebra of subsets of X defined as $\mathcal{B} = \mathcal{B}(\mathcal{S})$, where $\mathcal{S} = \{\pi_i^{-1}(A_i) : A_i \in \mathcal{B}_i, i \in I\}$, and $(X, \mathcal{B})$ is called the 'product measurable space'.

(5) If $(X, \mathcal{B})$ is a measurable space, and if $X_0 \subset X$ is an arbitrary subset, define $\mathcal{B}|_{X_0} = \{A \cap X_0 : A \in \mathcal{B}\}$; equivalently, $\mathcal{B}|_{X_0} = \{i^{-1}(A) : A \in \mathcal{B}\}$, where $i : X_0 \to X$ denotes the inclusion map. The σ-algebra $\mathcal{B}|_{X_0}$ is called the σ-algebra induced by $\mathcal{B}$. $\square$

EXERCISE A.5.3 *(1) Let $\mathcal{H}$ be a separable Hilbert space. Let τ (resp., τ_w) denote the family of all open (resp., weakly open) sets in $\mathcal{H}$. Show that $\mathcal{B}_{\mathcal{H}} = \mathcal{B}(\tau) = \mathcal{B}(\tau_w)$. (Hint: It suffices to prove that if $x \in \mathcal{H}$ and if $\epsilon > 0$, then $\overline{B(x,\epsilon)} = \{y \in \mathcal{H} : ||y - x|| \leq \epsilon\} \in \mathcal{B}(\tau_w)$; pick a countable (norm-) dense set $\{x_n\}$ in $\mathcal{H}$, and note that $\overline{B(x,\epsilon)} = \cap_n \{y : |\langle (y - x), x_n \rangle| \leq \epsilon ||x_n||\}$.)*

(2) Let τ_n (resp., τ_s, τ_w) denote the set of all subsets of $\mathcal{L}(\mathcal{H})$ which are open with respect to the norm topology (resp., strong operator topology, weak operator topology) on $\mathcal{L}(\mathcal{H})$. If $\mathcal{H}$ is separable, show that $\mathcal{B}_{\mathcal{L}(\mathcal{H})} = \mathcal{B}(\tau_n) = \mathcal{B}(\tau_s) = \mathcal{B}(\tau_w)$. (Hint: use (1) above.)

We now come to the appropriate class of maps in the category of measurable spaces.

DEFINITION A.5.4 *If $(X_i, \mathcal{B}_i), i = 1, 2,$ are measurable spaces, then a function $f : X_1 \to X_2$ is said to be* **measurable** *if $f^{-1}(A) \in \mathcal{B}_1 \ \forall \ A \in \mathcal{B}_2$.*

PROPOSITION A.5.5 *(a) The composite of measurable maps is measurable.*

(b) If $(X_i, \mathcal{B}_i), i = 1, 2,$ are measurable spaces, and if $\mathcal{B}_2 = \mathcal{B}(\mathcal{S})$ for some family $\mathcal{S} \subset 2^{X_2}$, then a function $f : X_1 \to X_2$ is measurable if and only if $f^{-1}(A) \in \mathcal{B}_1$ for all $A \in \mathcal{S}$.

(c) If $X_i, i = 1, 2$ are topological spaces, and if $f : X_1 \to X_2$ is a continuous map, then f is measurable as a map between the measurable spaces $(X_i, \mathcal{B}_{X_i}), i = 1, 2$.

Proof: (a) If $(X_i, \mathcal{B}_i), i = 1, 2, 3,$ are measurable spaces, and if $f : X_1 \to X_2$ and $g : X_2 \to X_3$ are measurable maps, and if $A \in \mathcal{B}_3$, then $g^{-1}(A) \in \mathcal{B}_2$ (since g is measurable), and consequently, $(g \circ f)^{-1}(A) = f^{-1}(g^{-1}(A)) \in \mathcal{B}_1$ (since f is measurable).

(b) Suppose $f^{-1}(A) \in \mathcal{B}_1$ for all $A \in \mathcal{S}$. Since $(f^{-1}(A))^c = f^{-1}(A^c), f^{-1}(X_2) = X_1$ and $f^{-1}(\cup_i A_i) = \cup_i f^{-1}(A_i)$, it is seen that the family $\mathcal{B} = \{A \subset X_2 : f^{-1}(A) \in \mathcal{B}_1\}$ is a σ-algebra of subsets of $\mathcal{B}_2$ which contains $\mathcal{S}$, and consequently, $\mathcal{B} \supset \mathcal{B}(\mathcal{S}) = \mathcal{B}_2$, and hence f is measurable.

(c) Apply (b) with $\mathcal{S}$ as the set of all open sets in X_2. $\square$

Some consequences of the preceding proposition are listed in the following exercises.

EXERCISE A.5.6 *(1) Suppose* $\{(X_i, \mathcal{B}_i) : i \in I\}$ *is a family of measurable spaces. Let* $(X, \mathcal{B})$ *be the product measurable space, as in Example A.5.2(4). Let* $(Y, \mathcal{B}_0)$ *be any measurable space.*

(a) Show that there exists a bijective correspondence between maps $f : Y \to X$ *and families* $\{f_i : Y \to X_i\}_{i \in I}$ *of maps, this correspondence being given by* $f(y) = ((f_i(y)))$*; the map* f *is written as* $f = \prod_{i \in I} f_i$.

(b) Show that if f, f_i *are as in (a) above, then* f *is measurable if and only if each* f_i *is measurable. (Hint: Use the family* $\mathcal{S}$ *used to define the product* σ*-algebra, Proposition A.5.5(b), and the fact that* $\pi_i \circ f = f_i \ \forall i$*, where* π_i *denotes the projection of the product onto the i-th factor.)*

(2) If $(X, \mathcal{B})$ *is a measurable space, if* $f_i : X \to \mathbb{R}, i \in I$*, are measurable maps (with respect to the Borel* σ*-algebra on* $\mathbb{R}$*), and if* $F : \mathbb{R}^I \to \mathbb{R}$ *is a continuous map (where* $\mathbb{R}^I$ *is endowed with the product topology), then show that the map* $g : X \to \mathbb{R}$ *defined by* $g = F \circ (\prod_{i \in I} f_i)$ *is a* $(\mathcal{B}, \mathcal{B}_{\mathbb{R}})$*-measurable function.*

(3) If $(X, \mathcal{B})$ *is a measurable space, and if* $f, g : X \to \mathbb{R}$ *are measurable maps, show that each of the following maps is measurable:* $|f|, af + bg$ *(where* a, b *are arbitrary real numbers),* fg, $f^2 + 3f^3$, $x \mapsto sin(g(x))$.

(4) The **indicator function** *of a subset* $E \subset X$ *is the real-valued function on* X*, always denoted in these notes by the symbol* 1_E*, which is defined by*

$$1_E(x) \ = \ \left\{ \begin{array}{ll} 1 & if\ x \in E \\ 0 & if\ x \notin E \end{array} \right. .$$

Show that 1_E *is measurable if and only if* $E \in \mathcal{B}$*. (For this reason, the members of* $\mathcal{B}$ *are sometimes referred to as measurable* *sets. Also, the correspondence $E \to 1_E$ sets up a bijection between the family of subsets of X on the one hand, and functions from X into the 2-element set $\{0,1\}$, on the other; this is one of the reasons for using the notation 2^X to denote the class of all subsets of the set X.)*

In the sequel, we will have occasion to deal with possibly infinite-valued functions. We shall write $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty, -\infty\}$. Using any reasonable bijection between $\overline{\mathbb{R}}$ and $[-1,1]$ - say, the one given by the map

$$f(x) \ = \ \begin{cases} \pm 1 & if \ x = \pm\infty \\ \frac{x}{1+|x|} & if \ x \in \mathbb{R} \end{cases}$$

- we may regard $\overline{\mathbb{R}}$ as a (compact Hausdorff) topological space, and hence equip it with the corresponding Borel σ-algebra. If we have an extended real-valued function defined on a set X - i.e., a map $f : X \to \overline{\mathbb{R}}$ - and if $\mathcal{B}$ is a σ-algebra on X, we shall say that f is measurable if $f^{-1}(A) \in \mathcal{B} \ \forall \ A \in \mathcal{B}_{\overline{\mathbb{R}}}$. The reader should have little difficulty in verifying the following fact: given a map $f : X \to \overline{\mathbb{R}}$, let $X_0 = f^{-1}(\mathbb{R})$, $f_0 = f|_{X_0}$, and let $\mathcal{B}|_{X_0}$ be the induced σ-algebra on the subset X_0 - see Example A.5.2(5); then the extended real-valued function f is measurable if and only if the following conditions are satisfied: (i) $f^{-1}(\{a\}) \in \mathcal{B}$, for $a \in \{-\infty, \infty\}$, and (ii) $f_0 : X_0 \to \mathbb{R}$ is measurable.

We will, in particular, be interested in limits of sequences of functions (when they converge). Before systematically discussing such (pointwise) limits, we pause to recall some facts about the **limit-superior** and the **limit-inferior** of sequences.

Suppose, to start with, that $\{a_n\}$ is a sequence of real numbers. For each (temporarily fixed) $m \in \mathbb{N}$, define

$$b_m \ = \ inf_{n \geq m} a_n \ , \ c_m \ = \ sup_{n \geq m} a_n \ , \qquad \text{(A.5.15)}$$

with the understanding, of course, that $b_m = -\infty$ (resp., $c_m = +\infty$) if the sequence $\{a_n\}$ is not bounded below (resp., bounded above).

Hence,

$$n \geq m \ \Rightarrow \ b_m \leq a_n \leq c_m \ ;$$

and in particular,

$$b_1 \leq b_2 \leq \cdots \leq b_m \leq \cdots \leq c_m \leq \cdots \leq c_2 \leq c_1 \ ;$$

thus, $\{b_m : m \in \mathbb{N}\}$ (resp., $\{c_m : m \in \mathbb{N}\}$) is a non-decreasing (resp., non-increasing) sequence of extended real numbers. Consequently, we may find extended real numbers b, c such that $b = sup_m b_m = lim_m b_m$ and $c = inf_m c_m = lim_m c_m$.

We define $b = lim\ inf_n a_n$ and $c = lim\ sup_n a_n$. Thus, we have

$$lim\ inf_n\ a_n \ = \ \lim_{m \to \infty} \ \inf_{n \geq m} \ a_m \ = \sup_{m \in \mathbb{N}} \ \inf_{n \geq m} \ a_m \qquad \text{(A.5.16)}$$

and

$$lim\ sup_n\ a_n \ = \ \lim_{m \to \infty} \ \sup_{n \geq m} \ a_m \ = \inf_{m \in \mathbb{N}} \ \sup_{n \geq m} \ a_m \ . \qquad \text{(A.5.17)}$$

We state one useful alternative description of the 'inferior' and 'superior' limits of a sequence, in the following exercise.

EXERCISE A.5.7 *Let $\{a_n : n \in \mathbb{N}\}$ be any sequence of real numbers. Let L be the set of all 'limit-points' of this sequence; thus, $l \in L$ if and only if there exists some subsequence for which $l = \lim_{k \to \infty} a_{n_k}$. Thus $L \subset \overline{\mathbb{R}}$. Show that:*

(a) $lim\ inf\ a_n\ ,\ lim\ sup\ a_n\ \in\ L$; and

(b) $l \in L \ \Rightarrow\ lim\ inf\ a_n\ \leq l \leq\ lim\ sup\ a_n$.

(c) the sequence $\{a_n\}$ converges to an extended real number if and only if $lim\ inf\ a_n\ =\ lim\ sup\ a_n$.

If X is a set and if $\{f_n : n \in \mathbb{N}\}$ is a sequence of real-valued functions on X, we define the extended real-vallued functions $lim\ inf\ f_n$ and $lim\ sup\ f_n$ in the obvious pointwise manner, thus:

$$\begin{aligned} (lim\ inf\ f_n)(x) &= lim\ inf\ f_n(x), \\ (lim\ sup\ f_n)(x) &= lim\ sup\ f_n(x). \end{aligned}$$

PROPOSITION A.5.8 *Suppose $(X, \mathcal{B})$ is a measurable space, and suppose $\{f_n : n \in \mathbb{N}\}$ is a sequence of real-valued measurable functions on* X. *Then,* *(a) $sup_n f_n$ and $inf_n f_n$ are measurable extended real-valued functions on X;*

(b) $lim\ inf\ f_n$ and $lim\ sup\ f_n$ are measurable extended real-valued functions on X;

(c) the set $C = \{x \in X : \{f_n(x)\}$ converges to a finite real number$\}$ is measurable, i.e., $C \in \mathcal{B}$; further, the function $f : C \to \mathbb{R}$ defined by $f(x) = lim_n f_n(x)$ is measurable (with respect to the induced σ-algebra $\mathcal{B}|_C$. (In particular, the pointwise limit of a (pointwise) convergent sequence of measurable real-valued functions is also a measurable function.)

Proof : (a) Since $inf_n\ f_n = -\ sup_n(-f_n)$, it is enough to prove the assertion concerning suprema. So, suppose $f = sup_n\ f_n$, where $\{f_n\}$ is a sequence of measurable real-valued functions on X. In view of Proposition A.5.5(b), and since the class of sets of the form $\{t \in \overline{\mathbb{R}} : t > a\}$ - as a varies over $\mathbb{R}$ - generates the Borel σ-algebra $\mathcal{B}_{\overline{\mathbb{R}}}$, we only need to check that $\{x \in X : f(x) > a\} \in \mathcal{B}$, for each $a \in \mathbb{R}$; but this follows from the identity

$$\{x \in X : f(x) > a\} \ = \ \cup_{n=1}^{\infty} \{x \in X : f_n(x) > a\}$$

(b) This follows quite easily from repeated applications of (a) (or rather, the extension of (a) which covers suprema and infima of sequences of extended real-valued functions).

(c) Observe, to start with, that if Z is a Hausdorff space, then $\Delta = \{(z, z) : z \in Z\}$ is a closed set in $Z \times Z$; hence, if $f, g : X \to Z$ are $(\mathcal{B}, \mathcal{B}_Z)$-measurable functions, then $(f, g) : X \to Z \times Z$ is a $(\mathcal{B}, \mathcal{B}_{Z\times Z})$-measurable function, and hence $\{x \in X : f(x) = g(x)\} \in \mathcal{B}$.

In particular, if we apply the preceding fact to the case when $f = lim\ sup_n f_n$, $g = lim\ inf_n\ f_n$ and $Z = \overline{\mathbb{R}}$, we find that if D is the set of points $x \in X$ for which the sequence $\{f_n(x)\}$ converges to a point in $\overline{\mathbb{R}}$, then $D \in \mathcal{B}$. Since $\mathbb{R}$ is a measurable set in $\overline{\mathbb{R}}$, the set $F = \{x \in X : \ lim\ sup_n\ f_n(x) \ \in \mathbb{R}\}$ must also belong to $\mathcal{B}$, and consequently, $C = D \cap F \in \mathcal{B}$. Also since $f|_C$ is clearly a measurable function, all parts of (c) are proved. □

The following proposition describes, in a sense, how to construct all possible measurable real-valued functions on a measurable space $(X, \mathcal{B})$.

Proposition A.5.9 *Let* $(X, \mathcal{B})$ *be a measurable space.*

(1) The following conditions on a $(\mathcal{B}, \mathcal{B}_{\mathbb{R}})$*-measurable function* $f : X \to \mathbb{R}$ *are equivalent:*

(i) $f(X)$ *is a finite set;*

(ii) there exists $n \in \mathbb{N}$*, real numbers* $a_1, \cdots, a_n \in \mathbb{R}$ *and a partition* $X = \coprod_{i=1}^n E_i$ *such that* $E_i \in \mathcal{B}\ \forall i$*, and* $f = \sum_{i=1}^n a_i 1_{E_i}$*.*

A function satisfying the above conditions is called a **simple** *function (on* $(X, \mathcal{B})$*).*

(2) The following conditions on a $(\mathcal{B}, \mathcal{B}_{\mathbb{R}})$*-measurable function* $f : X \to \mathbb{R}$ *are equivalent:*

(i) $f(X)$ *is a countable set;*

(ii) there exists a countable set I*, real numbers* $a_n,\ n \in I$*, and a partiton* $X = \coprod_{n \in I} E_n$ *such that* $E_n \in \mathcal{B}\ \forall n \in I$*, and* $f = \sum_{n \in I} a_n 1_{E_n}$*.*

A function satisfying the above conditions is called an **elementary** *function (on* $(X, \mathcal{B})$*).*

(3) If $f : X \to \mathbb{R}$ *is any non-negative measurable function, then there exists a sequence* $\{f_n\}$ *of simple functions on* $(X, \mathcal{B})$ *such that:*

(i) $f_1(x) \leq f_2(x) \leq \cdots \leq f_n(x) \leq f_{n+1}(x) \leq \cdots$*, for all* $x \in X$*; and*

(ii) $f(x) = lim_n\ f_n(x) = sup_n\ f_n(x)\ \forall\ x \in X$*.*

Proof : (1) The implication $(ii) \Rightarrow (i)$ is obvious, and as for $(i) \Rightarrow (ii)$, if $f(X) = \{a_1, \cdots, a_n\}$, set $E_i = f^{-1}(\{a_i\})$, and note that these do what they are expected to do.

(2) This is proved in the same way as was (1).

(3) Fix a positive integer n, let $I_n^k = [\frac{k}{2^n}, \frac{k+1}{2^n})$ and define $E_n^k = f^{-1}(I_n^k)$, for $k = 0, 1, 2, \cdots$. Define $h_n = \sum_{k=1}^\infty \frac{k}{2^n} 1_{E_n^k}$. It should be clear that (the E_n^k's inherit measurability from f and consequently) h_n is an elementary function in the sense of (2) above.

The definitions further show that, for each $x \in X$, we have

$$f(x) - \frac{1}{2^n} < h_n(x) \leq h_{n+1}(x) \leq f(x)\ .$$

Thus, the h_n's form a non-decreasing sequence of non-negative elementary functions which converge uniformly to the function f. If we set $f_n = min\{h_n, n\}$, it is readily verified that these f_n's are simple functions which satisfy the requirements of (3).

□

Now we come to the all important notion of a measure. (Throughout this section, we shall use the word measure to denote what should be more accurately referred to as a positive measure; we will make the necessary distinctions when we need to.)

DEFINITION A.5.10 *Let $(X, \mathcal{B})$ be a measurable space. A* **measure** *on $(X, \mathcal{B})$ is a function $\mu : \mathcal{B} \to [0, \infty]$ with the following two properties:*

(0) $\mu(\emptyset) = 0$; and

(1) μ is **countably additive** *- i.e., if $E = \coprod_{n=1}^{\infty} E_n$ is a countable 'measurable' partition, meaning that $E, E_n \in \mathcal{B}\ \forall n$, then $\mu(E) = \sum_{n=1}^{\infty} \mu(E_n)$.*

A **measure space** *is a triple $(X, \mathcal{B}, \mu)$, consisting of a measurable space together with a measure defined on it.*

A measure μ is said to be **finite** *(resp., a* **probability measure***) if $\mu(X) < \infty$ (resp., $\mu(X) = 1$).*

We list a few elementary consequences of the definition in the next proposition.

PROPOSITION A.5.11 *Let $(X, \mathcal{B}, \mu)$ be a measure space; then,*

(1) μ is 'monotone': i.e., $A, B \in \mathcal{B},\ A \subset B \ \Rightarrow \mu(A) \leq \mu(B)$;

(2) μ is 'countably subadditive': i.e., if $E_n \in \mathcal{B}\ \forall n = 1, 2, \cdots$, then $\mu(\cup_{n=1}^{\infty} E_n), \leq \sum_{n=1}^{\infty} \mu(E_n)$;

(3) μ is 'continuous from below': i.e., if $E_1 \subset E_2 \subset \cdots \subset E_n \subset \cdots$ is an increasing sequence of sets in $\mathcal{B}$, and if $E = \cup_{n=1}^{\infty} E_n$, then $\mu(E) = lim_n \mu(E_n)$;

(4) μ is 'continuous from above if it is restricted to sets of finite measure': i.e., if if $E_1 \supset E_2 \supset \cdots \supset E_n \supset \cdots$ is a decreasing sequence of sets in $\mathcal{B}$, if $E = \cap_{n=1}^{\infty} E_n$, and if $\mu(E_1) < \infty$, then $\mu(E) = lim_n \mu(E_n)$.

Proof : (1) Since $B = A \coprod (B - A)$, deduce from the positivity and countable additivity of μ that

$$\begin{aligned}
\mu(A) &\leq \mu(A) + \mu(B - A) \\
&= \mu(A) + \mu(B - A) + \mu(\emptyset) + \mu(\emptyset) + \cdots \\
&= \mu(A \coprod (B - A) \coprod \emptyset \coprod \emptyset \coprod \cdots) \\
&= \mu(B) \; .
\end{aligned}$$

(2) Define $S_n = \cup_{i=1}^n E_i$, for each $n = 1, 2, \cdots$; then it is clear that $\{S_n\}$ is an increasing sequence of sets and that $S_n \in \mathcal{B}$ $\forall n$; now set $A_n = S_n - S_{n-1}$ for each n (with the understanding that $S_0 = \emptyset$); it is seen that $\{A_n\}$ is a sequence of pairwise disjoint sets and that $A_n \in \mathcal{B}$ for each n; finally, the construction also ensures that

$$S_n = \cup_{i=1}^n E_i = \coprod_{i=1}^n A_i, \; \forall n = 1, 2, \cdots \; . \tag{A.5.18}$$

Further, we see that

$$\cup_{n=1}^\infty E_n = \coprod_{n=1}^\infty A_n \; , \tag{A.5.19}$$

and hence by the assumed countable additivity and the already established monotonicity of μ, we see that

$$\begin{aligned}
\mu(\cup_{n=1}^\infty E_n) &= \mu(\coprod_{n=1}^\infty A_n) \\
&= \sum_{n=1}^\infty \mu(A_n) \\
&\leq \sum_{n=1}^\infty \mu(E_n) \; ,
\end{aligned}$$

since $A_n \subset E_n$ $\forall n$.

(3) If the sequence $\{E_n\}$ is already increasing, then, in the notation of the proof of (2) above, we find that $S_n = E_n$ and that $A_n = E_n - E_{n-1}$. Since countable additivity implies 'finite additivity' (by tagging on infinitely many copies of the empty set, as in the proof of (1) above), we may deduce from equation A.5.18 that

$$\mu(E_n) = \mu(S_n) = \sum_{i=1}^n \mu(A_i) \; ;$$

similarly, we see from equation A.5.19 that

$$\mu(E) \;=\; \sum_{n=1}^{\infty} \mu(A_n) \;;$$

the desired conclusion is a consequence of the last two equations.

(4) Define $F_n = E_1 - E_n$, note that (i) $F_n \in \mathcal{B}\ \forall n$, (ii) $\{F_n\}$ is an increasing sequence of sets, and (iii) $\cup_{n=1}^{\infty} F_n \;=\; E_1 - \cap_{n=1}^{\infty} E_n$; the desired conclusion is a consequence of one application of (the already proved) (3) above to the sequence $\{F_n\}$, and the following immediate consequence of (1): if $A, B \in \mathcal{B}$, $A \subset B$ and if $\mu(B) < \infty$, then $\mu(A) = \mu(B) - \mu(B - A)$. $\Box$

We now give some examples of measures; we do not prove the various assertions made in these examples; the reader interested in further details, may consult any of the standard texts on measure theory (such as [Hal1], for instance).

Example A.5.12 (1) Let X be any set, let $\mathcal{B} = 2^X$ and define $\mu(E)$ to be n if E is finite and has exactly n elements, and define $\mu(E)$ to be ∞ if E is an infinite set. This μ is easily verified to define a measure on $(X, 2^X)$, and is called the **counting measure** on X.

For instance, if $X = \mathbb{N}$, if $E_n = \{n, n+1, n+2, \cdots\}$, and if μ denotes counting measure on $\mathbb{N}$, we see that $\{E_n\}$ is a decreasing sequence of sets in $\mathbb{N}$, with $\cap_{n=1}^{\infty} E_n \;=\; \emptyset$; but $\mu(E_n) = \infty\ \forall n$, and so $\lim_n \mu(E_n) \;\neq\; \mu(\cap_n E_n)$; thus, if we do not restrict ourselves to sets of finite measure, a measure need not be continuous from above - see Proposition A.5.11(4).

(2) It is a fact, whose proof we will not go into here, that there exists a unique measure m defined on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ with the propertythat $m([a, b]) = b - a$ whenever $a, b \in \mathbb{R}, a < b$. This measure is called **Lebesgue measure** on $\mathbb{R}$. (Thus the Lebesgue measure of an interval is its length.)

More generally, for any $n \in \mathbb{N}$, there exists a unique measure m^n defined on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ such that if $B \;=\; \prod_{i=1}^n [a_i, b_i]$ is the n-dimensional 'box' with sides $[a_i, b_i]$, then $m^n(B) \;=\; \prod_{i=1}^n (b_i - a_i)$; this measure is called **n-dimensional Lebesgue measure**; thus, the n-dimensional Lebesgue measure of a box is just its (n-dimensional) volume.

(3) Suppose $\{(X_i, \mathcal{B}_i, \mu_i) : i \in I\}$ is an arbitrary family of 'probability spaces' : i.e., each $(X_i, \mathcal{B}_i)$ is a measurable space and μ_i is a probability measure defined on it. Let $(X, \mathcal{B})$ be the product measurable space, as in Example A.5.2(4). (Recall that $\mathcal{B}$ is the σ-algebra generated by all 'finite-dimensional cylinder sets'; i.e., $\mathcal{B} = \mathcal{B}(\mathcal{C})$, where a typical member of $\mathcal{C}$ is obtained by fixing finitely many co-ordinates (i.e., a finite set $I_0 \subset I$), fixing measurable sets $E_i \in \mathcal{B}_i \ \forall i \in I_0$, and looking at the 'cylinder C with base $\prod_{i \in I_0} E_i$' - thus $C = \{((x_i)) \in X : x_i \in E_i \ \forall i \in I_0\}$. It is a fact that there exists a unique probability measure μ defined on $(X, \mathcal{B})$ with the property that if C is as above, then $\mu(C) = \prod_{i \in I_0} \mu_i(E_i)$. This measure μ is called the product of the probability measures μ_i.

It should be mentioned that if the family I is finite, then the initial measures μ_i do not have to be probability measures for us to be able to construct the product measure. (In fact m^n is the product of n copies of m.) In fact, existence and uniqueness of the 'product' of finitely many measures can be established if we only impose the condition that each μ_i be a **σ-finite measure** - meaning that X_i admits a countable partition $X_i = \coprod_{n=1}^{\infty} E_n$ such that $E_n \in \mathcal{B}_i$ and $\mu_i(E_n) < \infty$ for all n.

It is only when we wish to construct the product of an infinite family of measures, that we have to impose the requirement that (at least all but finitely many of) the measures concerned are probability measures. $\Box$

Given a measure space $(X, \mathcal{B}, \mu)$ and a non-negative $(\mathcal{B}, \mathcal{B}_{\mathbb{R}})$-measurable function $f : X \to \mathbb{R}$, there is a unique (natural) way of assigning a value (in $[0, \infty]$) to the expression $\int f d\mu$. We will not go into the details of how this is proved; we will, instead, content ourselves with showing the reader how this 'integral' is computed, and leave the interested reader to go to such standard references as [Hal1], for instance, to become convinced that all these things work out as they are stated to. We shall state, in one long Proposition, the various features of the assignment $f \mapsto \int f d\mu$.

PROPOSITION A.5.13 *Let $(X, \mathcal{B}, \mu)$ be a measure space; let us write $\mathcal{M}_+$ to denote the class of all functions $f : X \to [0, \infty)$ which are $(\mathcal{B}, \mathcal{B}_{\mathbb{R}})$-measurable.*

(1) There exists a unique map $\mathcal{M}_+ \ni f \mapsto \int f d\mu \in [0,\infty]$ *which satisfies the following properties:*

(i) $\int 1_E d\mu = \mu(E) \; \forall \; E \in \mathcal{B}$;

(ii) $f, g \in \mathcal{M}_+, a, b \in [0,\infty) \Rightarrow \int (af + bg) d\mu = a \int f d\mu + b \int g d\mu$;

(iii) for any $f \in \mathcal{M}_+$, *we have*

$$\int f d\mu = \lim_n \int f_n d\mu ,$$

for any non-decreasing sequence $\{f_n\}$ *of simple functions which converges pointwise to* f. *(See Proposition A.5.9(3).)*

(2) Further, in case the measure μ *is* σ*-finite, and if* $f \in \mathcal{M}_+$, *then the quantity* ($\int f d\mu$) *which is called the integral of* f *with respect to* μ, *admits the following interpretation as 'area under the curve': let* $A(f) = \{(x,t) \in X \times \mathbb{R} : 0 \leq t \leq f(x)\}$ *denote the 'region under the graph of* f*'; let* $\mathcal{B} \otimes \mathcal{B}_{\mathbb{R}}$ *denote the natural product* σ*-algebra on* $X \times \mathbb{R}$ *(see Example A.5.2), and let* $\lambda = \mu \times m$ *denote the product-measure - see Example A.5.12(3) - of* μ *with Lebesgue measure on* $\mathbb{R}$; *then,*

(i) $A(f) \in \mathcal{B} \otimes \mathcal{B}_{\mathbb{R}}$; *and*

(ii) $\int f d\mu = \lambda(\; A(f) \;)$. □

EXERCISE A.5.14 *Let* $(X, \mathcal{B}, \mu)$ *be a measure space. Then show that*

(a) a function $f : X \to \mathbb{C}$ *is* $(\mathcal{B}, \mathcal{B}_{\mathbb{C}})$*-measurable if and only if* $Re\ f$ *and* $Im\ f$ *are* $(\mathcal{B}, \mathcal{B}_{\mathbb{R}})$*-measurable, where, of course* $Re\ f$ *andf* $Im\ f$ *denote the real- and imaginary- parts of* f, *respectively;*

(b) a function $f : X \to \mathbb{R}$ *is* $(\mathcal{B}, \mathcal{B}_{\mathbb{R}})$*-measurable if and only if* $f_\pm$ *are* $(\mathcal{B}, \mathcal{B}_{\mathbb{R}})$*-measurable, where* $f_\pm$ *are the 'positive and negative parts of* f *(as in the proof of Proposition 3.3.11(f), for instance);*

(c) if $f : X \to \mathbb{C}$ *then* $f = (f_1 - f_2) + i(f_3 - f_4)$, *where* $\{f_j : 1 \leq j \leq 4\}$ *are the non-negative functions defined by* $f_1 = (Re\ f)_+, f_2 = (Re\ f)_-, f_3 = (Im\ f)_+, f_4 = (Im\ f)_-$; *show further that* $0 \leq f_j(x) \leq |f(x)| \; \forall x \in X$;

(d) if $f, g : X \to [0,\infty)$ *are non-negative measurable functions on* X *such that* $f(x) \leq g(x) \; \forall x$, *then* $\int f d\mu \leq \int g d\mu$.

Using the preceding exercise, we can talk of the integral of appropriate complex-valued functions. Thus, let us agree to call a function $f : X \to \mathbb{C}$ **integrable** with respect to the measure μ if (i) f is $(\mathcal{B}, \mathcal{B}_{\mathbb{C}})$-measurable, and if (ii) $\int |f| d\mu < \infty$. (This makes sense since the measurability of f implies that of $|f|$.) Further, if $\{f_j : 1 \leq j \leq 4\}$ are as in Exercise A.5.14(d) above, then we may define

$$\int f d\mu \ = \ (\int f_1 d\mu - \int f_2 d\mu) + i(\int f_3 d\mu - \int f_4 d\mu) \ .$$

It is an easy exercise to verify that the set of μ-integrable functions form a vector space and that the map $f \to \int f d\mu$ defines a linear functional on this vector space.

Remark A.5.15 **The space $L^1(X, \mathcal{B}, \mu)$:**

If $(X, \mathcal{B}, \mu)$ is a measure space, let $\mathcal{L}^1 = \mathcal{L}^1(X, \mathcal{B}, \mu)$ denote the vector space of all μ-integrable functions $f : X \to \mathbb{C}$. Let $\mathcal{N} \ = \ \{f \in \mathcal{L}^1 : \int |f| d\mu = 0\}$; it is then not hard to show that in order for a measurable function f to belong to $\mathcal{N}$, it is necessary and sufficient that f vanish μ-**almost everywhere**, meaning that $\mu(\{f \neq 0\}) = 0$. It follows that $\mathcal{N}$ is a vector subspace of $\mathcal{L}^1$; define $L^1 = L^1(X, \mathcal{B}, \mu)$ to be the quotient space $\mathcal{L}^1/\mathcal{N}$; (thus two integrable functions define the same element of L^1 precisely when they agree μ-almost everywhere); it is true, furthermore, that the equation

$$||f||_1 \ = \ \int |f| d\mu$$

can be thought of as defining a norm on L^1. (Strictly speaking, an element of L^1 is not one function, but a whole equivalence class of functions, (any two of which agree a.e.), but the integral in the above equation depends only on the equivalence class of the function f; in other words, the above equation defines a semi-norm on $\mathcal{L}^1$, which 'descends' to a norm on the quotient space L^1.) It is further true that $L^1(X, \mathcal{B}, \mu)$ is a Banach space.

In an exactly similar manner, the spaces $L^p(X, \mathcal{B}, \mu)$ are defined; since the cases $p \in \{1, 2, \infty\}$ will be of importance for us, and since we shall not have to deal with other values of p in this book, we shall only discuss these cases.

The space $L^2(X, \mathcal{B}, \mu)$:

Let $\mathcal{L}^2 = \mathcal{L}^2(X, \mathcal{B}, \mu)$ denote the set of $(\mathcal{B}, \mathcal{B}_{\mathbb{C}})$-measurable functions $f : X \to \mathbb{C}$ such that $|f|^2 \in \mathcal{L}^1$; for $f, g \in \mathcal{L}^2$, define $\langle f, g \rangle = \int f\overline{g}d\mu$. It is then true that $\mathcal{L}^2$ is a vector space and that the equation $||f||_2 = \langle f, f \rangle^{\frac{1}{2}}$ defines a semi-norm on $\mathcal{L}^2$; as before, if we let $\mathcal{N} = \{f \in \mathcal{L}^2 : ||f||_2 = 0\}$ (= the set of measurable functions which are equal to 0 a.e.), then the semi-norm $|| \cdot ||_2$ descends to a genuine norm on the quotient space $L^2 = L^2(X, \mathcal{B}, \mu) = \mathcal{L}^2/\mathcal{N}$. As in the case of L^1, we shall regard elements of L^2 as functions (rather than equivalence classes of functions) with the proviso that we shall consider two functions as being the same if they agree a.e. (which is the standard abbreviation for the expression 'almost everywhere'); when it is necessary to draw attention to the measure in question, we shall use expressions such as $f = g$ $\mu{-}a.e.$ It is a fact - perhaps not too surprising - that L^2 is a Hilbert space with respect to the natural definition of inner product.

It should be mentioned that much of what was stated here for $p = 1, 2$ has extensions that are valid for $p \geq 1$; (of course, it is only for $p = 2$ that L^2 is a Hilbert space;) in general, we get a Banach space L^p with

$$||f||_p = \left(\int |f|^p d\mu\right)^{\frac{1}{p}} .$$

The space $L^\infty(X, \mathcal{B}, \mu)$:

The case L^∞ deserves a special discussion because of certain features of the norm in question; naturally, we wish, here, to regard bounded measurable functions with the norm being given by the 'supremum norm'; the only mild complication is caused by our being forced to consider two measurable functions as being identical if they agree a.e.; hence, what happens on a set of (μ-) measure zero should not be relevant to the discussion; thus, for instance, if we are dealing with the example of Lebesgue measure m on R, then since $m(A) = 0$ for every countable set, we should regard the function $f(x) = 0$ $\forall x$ as being the same as the function g which is defined to agree with f outside the set

$\mathbb{N}$ of natural numbers but satisfies $g(n) = n \; \forall \; n \in \mathbb{N}$; note that g is not bounded, while f is as trivial as can be.

Hence, define $\mathcal{L}^\infty = \mathcal{L}^\infty(X, \mathcal{B}, \mu)$ to be the class of all $(\mathcal{B}, \mathcal{B}_{\mathbb{C}})$-measurable functions which are **essentially bounded**; thus $f \in \mathcal{L}^\infty$ precisely when there exists $E \in \mathcal{B}$ (which could very well depend upon f) such that $\mu(X - E) = 0$ and f is bounded on E; thus the elements of $\mathcal{L}^\infty$ are those measurable functions which are 'bounded a.e.'

Define $||f||_\infty = \inf\{K > 0 : \; \exists \; N \in \mathcal{B}$ such that $\mu(N) = 0$ and $|f(x)| \leq K$ whenever $x \in X - N\}$.

Finally, we define $L^\infty = L^\infty(X, \mathcal{B}, \mu)$ to be the quotient space $L^\infty = \mathcal{L}^\infty / \mathcal{N}$, where, as before, $\mathcal{N}$ is the set of measurable functions which vanish a.e. (which is also the same as $\{f \in \mathcal{L}^\infty : ||f||_\infty = 0)$. It is then the case that L^∞ is a Banach space.

It is a fact that if μ is a finite measure (and even more generally, but not without some restrictions on μ), and if $1 \leq p < \infty$, then the Banach dual space of $L^p(X, \mathcal{B}, \mu)$ may be naturally identified with $L^q(X, \mathcal{B}, \mu)$, where the 'dual-index' q is related to p by the equation $\frac{1}{p} + \frac{1}{q} = 1$, or equivalently, q is defined thus:

$$q = \begin{cases} \frac{p}{p-1} & if \; 1 < p < \infty \\ \infty & if \; p = 1 \end{cases}$$

where the 'duality' is given by integration, thus: if $f \in L^p, g \in L^q$, then it is true that $fg \in L^1$ and if we define

$$\phi_g(f) = \int fg d\mu \; , \; f \in L^p$$

then the correspondence $g \mapsto \phi_g$ is the one which establishes the desired isometric isomorphism $L^q \cong (L^p)^*$. □

We list some of the basic results on integration theory in one proposition, for convenience of reference.

Proposition A.5.16 *Let $(X, \mathcal{B}, \mu)$ be a measure space.*

*(1) (**Fatou's lemma**) if $\{f_n\}$ is a sequence of non-negative measurable functions on X, then,*

$$\int \; \lim inf \; f_n \; d\mu \; \leq \; \lim inf \; (\int \; f_n d\mu) \; .$$

*(2) (**monotone convergence theorem**) Suppose $\{f_n\}$ is a sequence of non-negative measurable functions on X, suppose f is a non-negative measurable function, and suppose that for almost every $x \in X$, it is true that the sequence $\{f_n(x)\}$ is a non-decreasing sequence of real numbers which converges to $f(x)$; then,*

$$\int f d\mu \ = \ \lim \int f_n d\mu \ = \ \sup \int f_n d\mu \ .$$

*(3) (**dominated convergence theorem**) Suppose $\{f_n\}$ is a sequence in $L^p(X, \mathcal{B}, \mu)$ for some $p \in [1, \infty)$; suppose there exists $g \in L^p$ such that $|f_n| \leq g$ a.e., for each n; suppose further that the sequence $\{f_n(x)\}$ converges to $f(x)$ for μ-almost every x in X; then $f \in L^p$ and $||f_n - f||_p \to 0$ as $n \to \infty$.*

REMARK A.5.17 It is true, conversely, that if $\{f_n\}$ is a sequence which converges to f in L^p, then there exists a subsequence, call it $\{f_{n_k} : k \in \mathbb{N}\}$ and a $g \in L^p$ such that $|f_{n_k}| \leq g$ *a.e.*, for each k, and such that $\{f_{n_k}(x)\}$ converges to $f(x)$ for μ-almost every $x \in X$; all this is for $p \in [1, \infty)$. Thus, modulo the need for passing to a subsequence, the dominated convergence theorem essentially describes convergence in L^p, provided $1 \leq p < \infty$. (The reader should be able to construct examples to see that the statement of the dominated convergence theorem is not true for $p = \infty$.) $\square$

For convenience of reference, we now recall the basic facts concerning product measures (see Example A.5.12(3)) and integration with respect to them.

PROPOSITION A.5.18 *Let $(X, \mathcal{B}_X, \mu)$ and $(Y, \mathcal{B}_Y, \nu)$ be σ-finite measure spaces. Let $\mathcal{B} = \mathcal{B}_X \otimes \mathcal{B}_Y$ be the 'product σ-algebra', which is defined as the σ-algebra generated by all 'measurable rectangles' - i.e., sets of the form $A \times B, A \in \mathcal{B}_X, B \in \mathcal{B}_Y$.*

(a) There exists a unique σ-finite measure λ (which is usually denoted by $\mu \times \nu$ and called the product measure) defined on $\mathcal{B}$ with the property that

$$\lambda(A \times B) = \mu(A)\nu(B) \ , \ \forall \ A \in \mathcal{B}_X, \ B \in \mathcal{B}_Y \ ; \qquad \text{(A.5.20)}$$

(b) more generally than in (a), if $E \in \mathcal{B}$ is arbitrary, define the vertical (resp., horizontal) slice of E by $E^x = \{y \in Y : (x,y) \in E\}$ (resp., $E_y = \{x \in X : (x,y) \in E\}$); then

(i) $E^x \in \mathcal{B}_Y \ \forall \ x \in X$ (resp., $E_y \in \mathcal{B}_X \ \forall \ y \in Y$);

(ii) the function $x \mapsto \nu(E^x)$ (resp., $y \mapsto \mu(E_y)$) is a measurable extended real-valued function on X (resp., Y); and

$$\int_X \nu(E^x)\, d\mu(x) \ = \ \lambda(E) \ = \ \int_Y \mu(E_y)\, d\nu(y). \qquad \text{(A.5.21)}$$

*(c) (**Tonelli's theorem**) Let $f : X \times Y \ \rightarrow \mathbb{R}^+$ be a non-negative $(\mathcal{B}, \mathcal{B}_{\mathbb{R}})$-measurable function; define the vertical (resp., horizontal) slice of f to be the function $f^x : Y \rightarrow \mathbb{R}^+$ (resp., $f_y : X \rightarrow \mathbb{R}^+$) given by $f^x(y) = f(x,y) = f_y(x)$; then,*

(i) f^x (resp., f_y) is a $(\mathcal{B}_Y, \mathcal{B}_{\mathbb{R}})$-measurable (resp., $(\mathcal{B}_X, \mathcal{B}_{\mathbb{R}})$-measurable) function, for every $x \in X$ (resp., $y \in Y$);

(ii) the function $x \mapsto \ \int_Y f^x d\nu$ (resp., $y \mapsto \int_X f_y d\mu$) is a $(\mathcal{B}_X, \mathcal{B}_{\overline{\mathbb{R}}})$-measurable (resp., $(\mathcal{B}_Y, \mathcal{B}_{\overline{\mathbb{R}}})$-measurable) extended real-valued function on X (resp., Y), and

$$\int_X \left(\ \int_Y f^x d\nu \ \right) d\mu(x) \ = \ \int_{X \times Y} f d\lambda \ = \ \int_Y \left(\ \int_X f_y d\mu \ \right) d\nu(y) \ . \qquad \text{(A.5.22)}$$

*(d) (**Fubini's theorem**) Suppose $f \in L^1(X \times Y, \mathcal{B}, \lambda)$; if the vertical and horizontal slices of f are defined as in (c) above, then*

(i) f^x (resp., f_y) is a $(\mathcal{B}_Y, \mathcal{B}_{\mathbb{C}})$-measurable (resp., $(\mathcal{B}_X, \mathcal{B}_{\mathbb{C}})$-measurable) function, for μ-almost every $x \in X$ (resp., for ν-almost every $y \in Y$);

(ii) $f^x \in L^1(Y, \mathcal{B}_Y, \nu)$ for μ-almost all $x \in X$ (resp., $f_y \in L^1(X, \mathcal{B}_X, \mu)$ for ν-almost all $y \in Y$);

(iii) the function $x \mapsto \ \int_Y f^x d\nu$ (resp., $y \mapsto \int_X f_y d\mu$) is a μ-a.e. (resp., ν-a.e.) meaningfully defined $(\mathcal{B}_X, \mathcal{B}_{\mathbb{C}})$-measurable (resp., $(\mathcal{B}_Y, \mathcal{B}_{\mathbb{C}})$-measurable) complex-valued function, which defines an element of $L^1(X, \mathcal{B}_X, \mu)$ (resp., $L^1(Y, \mathcal{B}_Y, \nu)$); further,

$$\int_X \left(\ \int_Y f^x d\nu \ \right) d\mu(x) \ = \ \int_{X \times Y} f d\lambda \ = \ \int_Y \left(\ \int_X f_y d\mu \ \right) d\nu(y) \ . \qquad \text{(A.5.23)}$$

□

One consequence of Fubini's theorem, which will be used in the text, is stated as an exercise below.

EXERCISE A.5.19 *Let $(X, \mathcal{B}_X, \mu)$ and $(Y, \mathcal{B}_Y, \nu)$ be σ-finite measure spaces, and let $\{e_n : n \in N\}$ (resp., $\{f_m : m \in M\}$) be an orthonormal basis for $L^2(X, \mathcal{B}_X, \mu)$ (resp., $L^2(Y, \mathcal{B}_Y, \nu)$). Define $e_n \otimes f_m : X \times Y \to \mathbb{C}$ by $(e_n \otimes f_m)(x, y) = e_n(x) f_m(y)$. Assume that both measure spaces are 'separable' meaning that the index sets M and N above are countable. Show that $\{e_n \otimes f_m : n \in N, m \in M\}$ is an orthonormal basis for $L^2(X \times Y, \mathcal{B}_X \otimes \mathcal{B}_Y, \mu \times \nu)$. (Hint: Use Proposition A.5.5 and the definition of $\mathcal{B}_X \otimes \mathcal{B}_Y$ to show that each $e_n \otimes f_m$ is $(\mathcal{B}_X \otimes \mathcal{B}_Y, \mathcal{B}_{\mathbb{C}})$-measurable; use Fubini's theorem to check that $\{e_n \otimes f_m\}_{n,m}$ is an orthonormal set; now, if $k \in L^2(X \times Y, \mathcal{B}_X \otimes \mathcal{B}_Y, \mu \times \nu)$, apply Tonelli's theorem to $|k|^2$ to deduce the existence of a set $A \in \mathcal{B}_X$ such that (i) $\mu(A) = 0$, (ii) $x \notin A \Rightarrow k^x \in L^2(Y, \mathcal{B}_Y, \nu)$, and*

$$\begin{aligned} ||k||^2_{L^2(\mu\times\nu)} &= \int_X ||k^x||^2_{L^2(\nu)} \, d\mu(x) \\ &= \int_X \left(\sum_{m\in M} |\langle k^x, f_m\rangle|^2 \right) d\mu(x) \\ &= \sum_{m\in M} \int_X |\langle k^x, f_m\rangle|^2 \, d\mu(x) \ ; \end{aligned}$$

set $g_m(x) = \langle k^x, f_m\rangle$, $\forall \ m \in M$, $x \notin A$, and note that each g_m is defined μ-a.e., and that

$$\begin{aligned} ||k||^2_{L^2(\mu\times\nu)} &= \sum_m ||g_m||^2_{L^2(\mu)} \\ &= \sum_{m\in M} \sum_{n\in N} |\langle g_m, e_n\rangle|^2 \ ; \end{aligned}$$

note finally - by yet another application of Fubini's theorem - that $\langle g_m, e_n\rangle = \langle k, e_n \otimes f_m\rangle$, to conclude that $||k||^2 = \sum_{m,n} |\langle k, e_n \otimes f_m\rangle|^2$ and consequently that $\{e_n \otimes f_m : m \in M, n \in N\}$ is indeed an orthonormal basis for $L^2(\mu \times \nu)$.

We conclude this brief discussion on measure theory with a brief reference to absolute continuity and the fundamental Radon-Nikodym theorem.

Proposition A.5.20 *Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f : X \to [0, \infty)$ be a non-negative integrable function. Then the equation*

$$\nu(E) \;=\; \int_E f d\mu \;=\; \int 1_E f d\mu \qquad \text{(A.5.24)}$$

defines a finite measure on $(X, \mathcal{B})$, which is sometimes called the 'indefinite integral' of f; this measure has the property that

$$E \in \mathcal{B},\ \mu(E) = 0 \;\Rightarrow\; \nu(E) = 0\ . \qquad \text{(A.5.25)}$$

If measures μ and ν are related as in the condition A.5.25, the measure ν is said to be **absolutely continuous** with respect to μ.

The following result is basic, and we will have occasion to use it in Chapter III. We omit the proof. Proofs of various results in this section can be found in any standard reference on measure theory, such as [Hal1], for instance. (The result is true in greater generality than is stated here, but this will suffice for our needs.)

Theorem A.5.21 *(**Radon-Nikodym theorem**)*

Suppose μ and ν are finite measures defined on the measurable space $(X, \mathcal{B})$. Suppose ν is absolutely continuous with respect to μ. Then there exists a non-negative function $g \in L^1(X, \mathcal{B}, \mu)$ with the property that

$$\nu(E) \;=\; \int_E g d\mu\ ,\ \ \forall\ E \in \mathcal{B}\ ;$$

the function g is uniquely determined μ-a.e. by the above requirement, in the sense that if g_1 is any other function in $L^1(X, \mathcal{B}, \mu)$ which satisfies the above condition, then $g = g_1\ \mu - a.e.$

Further, it is true that if $f : X \to [0, \infty)$ is any measurable function, then

$$\int f d\nu \;=\; \int f g d\mu\ .$$

Thus, a finite measure is absolutely continuous with respect to μ if and only if it arises as the indefinite integral of some non-negative function which is integrable with respect to μ; the function g, whose μ-equivalence class is uniquely determined by ν, is called the **Radon-Nikodym derivative** of ν with respect to μ, and both of the following expressions are used to denote this relationship:

$$g \;=\; \frac{d\nu}{d\mu} \;,\; d\nu \;=\; g\; d\mu \;. \tag{A.5.26}$$

We will outline a proof of the uniqueness of the Radon-Nikodym derivative in the following sequence of exercises.

EXERCISE A.5.22 *Let $(X, \mathcal{B}, \mu)$ be a finite measure space. We use the notation $\mathcal{M}_+$ to denote the class of all non-negative measurable functions on $(X, \mathcal{B})$.*

(1) Suppose $f \in \mathcal{M}_+$ satisfies the following condition: if $E \in \mathcal{B}$ and if $a1_E \leq f \leq b1_E$ μ-a.e. Show that

$$a\mu(E) \;\leq\; \int_E f\; d\mu \;\leq\; b\mu(E) \;.$$

(2) If $f \in \mathcal{M}_+$ and if $\int_A f d\mu = 0$ for some $A \in \mathcal{B}$, show that $f = 0$ a.e. on A - i.e., show that $\mu(\{x \in A : f(x) > 0\}) = 0$. (Hint: Consider the sets $E_n = \{x \in A : \frac{1}{n} \leq f(x) \leq n\}$, and appeal to (1).)

(3) If $f \in L^1(X, \mathcal{B}, \mu)$ satisfies $\int_E f d\mu = 0$ for every $E \in \mathcal{B}$, show that $f = 0$ a.e.; hence deduce the uniqueness assertion in the Radon-Nikodym theorem. (Hint: Write $f = g + ih$, where g and h are real-valued functions; let $g = g_+ - g_-$ and $h = h_+ - h_-$ be the canonical decompositions of g and h as differences of non-negative functions; show that (2) is applicable to each of the functions $g_\pm, h_\pm$ and the set $A = X$.)

(4) Suppose $d\nu = g\; d\mu$ as in Theorem A.5.21. Let $A = \{x : g(x) > 0\}$, and let $\mu|_A$ be the measure defined by

$$\mu|_A(E) \;=\; \mu(E \cap A)\; \forall\; E \in \mathcal{B}\;; \tag{A.5.27}$$

show that

(i) $\mu|_A$ is absolutely continuous with respect to μ and moreover, $d(\mu|_A) = 1_A d\mu$; and

(ii) $\mu|_A$ and ν are **mutually absolutely continuous**, *meaning that each measure is absolutely continuous with respect to the other, or equivalently, that they have the same null sets. (Hint: use (2) above to show that $\nu(E) = 0$ if and only if g vanishes μ-a.e. on E.)*

Two measures are said to be **singular** with respect to one another if they are 'supported on disjoint sets'; explicitly, if μ and ν are measures defined on $(X, \mathcal{B})$, we write $\mu \perp \nu$ if it is possible to find $A, B \in \mathcal{B}$ such that $X = A \coprod B$ and $\mu = \mu|_A, \ \nu = \nu|_B$ (in the sense of equation A.5.27). The purpose of the next exercise is to show that for several purposes, deciding various issues concerning two measures - such as absolutely continuity of one with respect to the other, or mutual singularity, etc.) - is as easy as deciding corresponding issues concerning two sets - such as whether one is essentially contained in the other, or essentially disjoint from another, etc.

EXERCISE A.5.23 *Let $\{\mu_i : 1 \leq i \leq n\}$ be a finite collection of finite positive measures on $(X, \mathcal{B})$. Let $\lambda = \sum_{i=1}^n \mu_i$; then show that:*

(a) λ is a finite positive measure and μ_i is absolutely continuous with respect to λ, for each $i = 1, \cdots, n$;

(b) if $g_i = \frac{d\mu_i}{d\lambda}$, and if $A_i = \{g_i > 0\}$, then

(i) μ_i is absolutely continuous with respect to μ_j if and only if $A_i \subset A_j$ (mod λ) (meaning that $\lambda(A_i - A_j) = 0$); and

(ii) $\mu_i \perp \mu_j$ if and only if A_i and A_j are disjoint mod λ (meaning that $\lambda(A_i \cap A_j) = 0$).

We will find the following consequence of the Radon-Nikodym theorem (and Exercise A.5.23) very useful. (This theorem is also not stated here in its greatest generality.)

THEOREM A.5.24 *(**Lebesgue decomposition**)*

Let μ and ν be finite measures defined on $(X, \mathcal{B})$. Then there exists a unique decomposition $\nu = \nu_a + \nu_s$ with the property that (i) ν_a is absolutely continuous with respect to μ, and (ii) ν_s and μ are singular with respect to each other.

Proof : Define $\lambda = \mu + \nu$; it should be clear that λ is also a finite measure and that both μ and ν are absolutely continuous with respect to λ. Define

$$f = \frac{d\mu}{d\lambda}, \ g = \frac{d\nu}{d\lambda}, \ A = \{f > 0\}, \ B = \{g > 0\} \ .$$

Define the measures ν_a and ν_s by the equations

$$\nu_a(E) \;=\; \int_{E\cap A} g\; d\lambda \;,\; \nu_s(E) \;=\; \int_{E-A} g\; d\lambda \;;$$

the definitions clearly imply that $\nu \;=\; \nu_a + \nu_s$; further, the facts that ν_a is absolutely continuous with respect to μ, and that ν_s and μ are mutually singular, are both immediate consequences of Exercise A.5.23(b).

Suppose $\nu = \nu_1 + \nu_2$ is some other decomposition of ν as a sum of finite positive measures with the property that ν_1 is absolutely continuous with respect to μ and $\nu_2 \perp \mu$. Notice, to start with, that both the ν_i are necessarily absolutely continuous with respect to ν and hence also with respect to λ. Write $g_i \;=\; \frac{d\nu_i}{d\lambda}$ and $F_i = \{g_i > 0\}$, for $i = 1, 2$. We may now deduce from the hypotheses and Exercise A.5.23(b) that we have the following inclusions mod λ:

$$F_1, F_2 \subset B,\; F_1 \subset A,\; F_2 \subset (X - A)\;.$$

note that $\frac{d(\nu_a)}{d\lambda} = 1_A g$ and $\frac{d(\nu_s)}{d\lambda} = 1_{X-A} g$, and hence, we have the following inclusions (mod λ):

$$F_1 \;\subset\; A \cap B \;=\; \{\frac{d(\nu_a)}{d\lambda} > 0\}$$

(resp.,

$$F_2 \;\subset\; B - A \;=\; \{\frac{d(\nu_s)}{d\lambda} > 0\}\;);$$

whence we may deduce from Exercise A.5.23(b) that ν_1 (resp., ν_2) is absolutely continuous with respect to ν_a (resp., ν_s).

The equation $\nu = \nu_1 + \nu_2$ must necessarily imply (again because of Exercise A.5.23(b)) that $B \;=\; F_1 \cup F_2$ (mod λ), which can only happen if $F_1 = A \cap B$ and $F_2 = B - A$ (mod λ), which means (again by the same exercise) that ν_1 (resp., ν_2) and ν_a (resp., ν_s) are mutually absolutely continuous. $\Box$

A.6 The Stone-Weierstrass theorem

We begin this section by introducing a class of spaces which, although not necessarily compact, nevertheless exhibit several features of compact spaces and are closely related to compact spaces in a manner we shall describe. At the end of this section, we prove the very useful Stone-Weierstrass theorem concerning the algebra of continuous functions on such spaces.

The spaces we shall be concerned with are the so-called *locally compact* spaces, which we now proceed to describe.

DEFINITION A.6.1 *A topological space X is said to be* **locally compact** *if it has a base $\mathcal{B}$ of sets with the property that given any open set U in X and a point $x \in U$, there exists a set $B \in \mathcal{B}$ and a compact set K such that $x \in B \subset K \subset U$.*

Euclidean space $\mathbb{R}^n$ is easily seen to be locally compact, for every $n \in \mathbb{N}$. More examples of locally compact spaces are provided by the following proposition.

PROPOSITION A.6.2 *(1) If X is a Hausdorff space, the following conditions are equivalent:*

(i) X is locally compact;

(ii) every point in X has an open neighbourhood whose closure is compact.

(2) Every compact Hausdorff space is locally compact.

(3) If X is a locally compact space, and if A is a subset of X which is either open or closed, then A, with respect to the subspace topology, is locally compact.

Proof : (1) $(i) \Rightarrow (ii)$: Recall that compact subsets of Hausdorff spaces are closed - see Proposition A.4.18(c) - while closed subsets of compact sets are compact in any topological space - see Proposition A.4.4(b). It follows that if $\mathcal{B}$ is as in Definition A.6.1, then the closure of every set in $\mathcal{B}$ is compact.

$(ii) \Rightarrow (i)$: Let $\mathcal{B}$ be the class of all open sets in X whose closures are compact. For each $x \in X$, pick a neighbourhood U_x whose closure - call it K_x - is compact. Suppose now that U is any open neighbourhood of x. Let $U_1 = U \cap U_x$, which is also an open neighbourhood of x. Now, $x \in U_1 \subset U_x \subset K_x$.

Consider K_x as a compact Hausdorff space (with the subspace topology). In this compact space, we see that (a) $K_x - U_1$ is a closed, and hence compact, subset of K_x, and (b) $x \notin K_x - U_1$. Hence, by Proposition A.4.18(c), we can find open sets V_1, V_2 in K_x such that $x \in V_1, K_x - U_1 \subset V_2$ and $V_1 \cap V_2 = \emptyset$. In particular, this means that $V_1 \subset K_x - V_2 = F$ (say); the set F is a closed subset of the compact set K_x and is consequently compact (in K_x and hence also in X - see Exercise A.4.3). Hence F is a closed set in (the Hausdorff space) X, and consequently, the closure, in X, of V_1 is contained in F and is compact. Also, since V_1 is open in K_x, there exists an open set V in X such that $V_1 = V \cap K_x$; but since $V_1 \subset \overline{V_1} \subset F \subset U_1 \subset K_x$, we find that also $V_1 = V \cap K_x \cap U_1 = V \cap U_1$; i.e., V_1 is open in X.

Thus, we have shown that for any $x \in X$ and any open neighbourhood U of x, there exists an open set $V \in \mathcal{B}$ such that $x \in V_1 \subset \overline{V_1} \subset U$, and such that $\overline{V_1}$ is compact; thus, we have verified local compactness of X.

(2) This is an immediate consequence of (1).

(3) Suppose A is a closed subset of X. Let $x \in A$. Then, by (1), there exists an open set U in X and a compact subset K of X such that $x \in U \subset K$. Then $x \in A \cap U \subset A \cap K$; but $A \cap K$ is compact (in X, being a closed subset of a compact set, and hence also compact in A), and $A \cap U$ is an open set in the subspace topology of A. Thus A is locally compact.

Suppose A is an open set. Then by Definition A.6.1, if $x \in A$, then there exists sets $U, K \subset A$ such that $x \in U \subset K \subset A$, where U is open in X and K is compact; clearly this means that U is also open in A, and we may conclude from (1) that A is indeed locally compact, in this case as well. $\square$

In particular, we see from Proposition A.4.18(a) and Proposition A.6.2(3) that if X is a compact Hausdorff space, and if $x_0 \in X$, then the subspace $A = X - \{x_0\}$ is a locally compact Hausdorff space with respect to the subspace topology. The surprising and exceedingly useful fact is that every locally compact Hausdorff space arises in this fashion.

THEOREM A.6.3 *Let X be a locally compact Hausdorff space. Then there exists a compact Hausdorff space $\hat{X}$ and a point in $\hat{X}$ - which is usually referred to as the 'point at infinity', and denoted simply by ∞ - such that X is homeomorphic to the subspace*

$\hat{X} - \{\infty\}$ *of* $\hat{X}$. *The compact space* $\hat{X}$ *is customarily referred to as the* **one-point compactification** *of* X.

Proof : Define $\hat{X} = X \cup \{\infty\}$, where ∞ is an artificial point (not in X) which is adjoined to X. We make $\hat{X}$ a topological space as follows: say that a set $U \subset \hat{X}$ is open if either (i) $U \subset X$, and U is an open subset of X; or (ii) $\infty \in U$, and $\hat{X} - U$ is a compact subset of X.

Let us first verify that this prescription does indeed define a topology on $\hat{X}$. It is clear that $\emptyset$ and $\hat{X}$ are open according to our definition. Suppose U and V are open in $\hat{X}$; there are four cases; (i) $U, V \subset X$: in this case U, V and $U \cap V$ are all open sets in X; (ii) U is an open subset of X and $V = \hat{X} - K$ for some compact subset of X: in this case, since $X - K$ is open in X, we find that $U \cap V = U \cap (X - K)$ is an open subset of X; (iii) V is an open subset of X and $U = \hat{X} - K$ for some compact subset of X: in this case, since $X - K$ is open in X, we find that $U \cap V = V \cap (X - K)$ is an open subset of X; (iv) there exist compact subsets $C, K \subset X$ such that $U = \hat{X} - C, V == \hat{X} - K$: in this case, $C \cup K$ is a compact subset of X and $U \cap V = \hat{X} - (C \cup K)$. We find that in all four cases, $U \cap V$ is open in $\hat{X}$. A similar case-by-case reasoning shows that an arbitrary union of open sets in $\hat{X}$ is also open, and so we have indeed defined a topology on $\hat{X}$.

Finally it should be obvious that the subspace topology that X inherits from $\hat{X}$ is the same as the given topology.

Since open sets in X are open in $\hat{X}$ and since X is Hausdorff, it is clear that distinct points in X can be separated in $\hat{X}$. Suppose now that $x \in X$; then, by Proposition A.6.2(1), we can find an open neighbourhood of x in X such that the closure (in X) of U is compact; if we call this closure K, then $V = \hat{X} - K$ is an open neighbourhood of ∞ such that $U \cap V = \emptyset$. Thus $\hat{X}$ is indeed a Hausdorff space.

Suppose now that $\{U_i : i \in I\}$ is an open cover of $\hat{X}$. Then, pick a U_j such that $\infty \in U_j$; since U_j is open, the definition of the topology on $\hat{X}$ implies that $\hat{X} - U_j = K$ is a compact subset of X. Then we can find a finite subset $I_0 \subset I$ such that $K \subset \cup_{i \in I_0} U_i$; it follows that $\{U_i : i \in I_0 \cup \{j\}\}$ is a finite subcover, thereby establishing the compactness of $\hat{X}$. □

EXERCISE A.6.4 *(1) Show that X is closed in $\hat{X}$ if and only if X is compact. (Hence if X is not compact, then X is dense in $\hat{X}$; this is the reason for calling $\hat{X}$ a compactification of X (since it is 'minimal' in some sense).*

(2) The following table has two columns which are labelled X and $\hat{X}$ respectively; if the i-th row has spaces A and B appearing in the the first and second columns, respectively, show that B is homeomorphic to $\hat{A}$.

	X	$\hat{X}$
1.	$\{1, 2, 3, \cdots\}$	$\{0\} \cup \{\frac{1}{n} : n = 1, 2, \cdots\}$
2.	$[0, 1)$	$[0, 1]$
3.	$\mathbb{R}^n$	$S^n = \{x \in \mathbb{R}^{n+1} : \|\|x\|\|_2 = 1\}$

The purpose of the next sequence of exercises is to lead up to a proof of the fact that a normed vector space is locally compact if and only if it is finite-dimensional.

EXERCISE A.6.5 *(1) Let X be a normed space, with finite dimension n, say. Let ℓ^1_n be the vector space $\mathbb{C}^n$ equipped with the norm $||\cdot||_1$ - see Example 1.2.2(1). Fix a basis $\{x_1, \cdots, x_n\}$ for X and define $T : \ell^1_n \to X$ by $T(\alpha_1, \cdots, \alpha_n) = \sum_{i=1}^n \alpha_i x_i$.*

(a) Show that $T \in \mathcal{L}(\ell^1_n, X)$, and that T is 1-1 and onto.

(b) Prove that the unit sphere $S = \{x \in \ell^1_n : ||x|| = 1\}$ is compact, and conclude (using the injectivity of T) that $\inf\{||Tx|| : x \in S\} > 0$; hence deduce that T^{-1} is also continuous.

(c) Conclude that any finite dimensional normed space is locally compact and complete, and that any two norms on such a space are equivalent - meaning that if $||\cdot||_i : i = 1, 2$ are two norms on a finite-dimensional space X, then there exist constants $k, K > 0$ such that $k||x||_1 \leq ||x||_2 \leq K||x||_1$ for all $x \in X$.

(2) (a) Let X_0 be a closed subspace of a normed space X; suppose $X_0 \neq X$. Show that, for each $\epsilon > 0$, there exists a vector $x \in X$ such that $||x|| = 1$ and $||x - x_0|| \geq 1 - \epsilon$ for all $x_0 \in X_0$. (Hint: Consider a vector of norm $(1 - 2\epsilon)$ in the quotient space X/X_0 - see Exercise 1.5.3(3).)

(b) If X is an infinite dimensional normed space, and if $\epsilon > 0$ is arbitrary, show that there exists a sequence $\{x_n\}_{n=1}^\infty$ in X such that $||x_n|| = 1$ for all N, and $||x_n - x_m|| \geq (1 - \epsilon)$, whenever $n \neq$

m. (Hint : Suppose unit vectors $x_1, \cdots, x_n$ have been constructed so that any two of them are at a distance of at least $(1-\epsilon)$ from one another; let X_n be the subspace spanned by $\{x_1, \cdots, x_n\}$; deduce from (1)(c) above that X_n is a proper closed subspace of X and appeal to (2)(a) above to pick a unit vector x_{n+1} which is at a distance of at least $(1-\epsilon)$ from each of the first n x_i's.)

(c) Deduce that no infinite-dimensional normed space is locally compact.

(3) Suppose $\mathcal{M}$ and $\mathcal{N}$ are closed subspaces of a Banach space X and suppose $\mathcal{N}$ is finite-dimensional; show that $\mathcal{M}+\mathcal{N}$ is a closed subspace of X. (Hint: Let $\pi : X \to X/\mathcal{M}$ be the natural quotient mapping; deduce from (1) above that $\pi(\mathcal{N})$ is a closed subspace of $X/\mathcal{M}$, and that, consequently, $\mathcal{N}+\mathcal{M} = \pi^{-1}(\pi(\mathcal{N}))$ is a closed subspace of X.)

(4) Show that the conclusion of (3) above is not valid, in general, if we only assume that $\mathcal{N}$ and $\mathcal{M}$ are closed, even when X is a Hilbert space. (Hint: Let $T \in \mathcal{L}(\mathcal{H})$ be a 1-1 operator whose range is not closed - for instance, see Remark 1.5.15; and consider $X = \mathcal{H} \oplus \mathcal{H}$, $\mathcal{M} = \mathcal{H} \oplus \{0\}, \mathcal{N} = G(T) = \{(x, Tx) : x \in \mathcal{H}\}$.)

We now wish to discuss continuous (real or complex-valued) functions on a locally compact space. We start with an easy consequence of Urysohn's lemma, Tietze's extension theorem and one-point compactifications of locally compact Hausdorff spaces.

PROPOSITION A.6.6 *Suppose X is a locally compact Hausdorff space.*

(a) If A and K are disjoint subsets of X such that A is closed and K is compact, then there exists a continuous function $f : X \to [0,1]$ such that $f(A) = \{0\}, f(K) = \{1\}$.

(b) If K is a compact subspace of X, and if $f : K \to \mathbb{R}$ is any continuous function, then there exists a continuous function $F : X \to \mathbb{R}$ such that $F|_K = f$.

Proof : (a) In the one-point compactification $\hat{X}$, consider the subsets K and $B = A \cup \{\infty\}$. Then, $\hat{X} - B = X - A$ is open (in X, hence in $\hat{X}$); thus B and K are disjoint closed subsets in a compact Hausdorff space; hence, by Urysohn's lemma, we can find a continuous function $g : \hat{X} \to [0,1]$ such that $g(B) = \{0\}, g(K) = \{1\}$. Let $f = g|_X$.

(b) This follows from applying Tietze's extension theorem (or rather, from its extension stated in Exercise A.4.25(2)) to the closed subset K of the one-point compactification of X, and then restricting the continuous function (which extends $f : K \to \hat{X}$) to X. □

The next result introduces a very important function space. (By the theory discussed in Chapter 2, these are the most general commutative C^*-algebras.)

PROPOSITION A.6.7 *Let $\hat{X}$ be the one-point compactification of a locally compact Hausdorff space X. Let $C(\hat{X})$ denote the space of all complex-valued continuous functions on $\hat{X}$, equipped with the sup-norm $||\cdot||_\infty$.*

(a) The following conditions on a continuous function $f : X \to \mathbb{C}$ are equivalent:

(i) f is the uniform limit of a sequence $\{f_n\}_{n=1}^\infty \subset C_c(X)$; i.e., there exists a sequence $\{f_n\}_{n=1}^\infty$ of continuous functions $f_n : X \to \mathbb{C}$ such that each f_n vanishes outside some compact subset of X, with the property that the sequence $\{f_n\}$ of functions converges uniformly to f on X.

(ii) f is continuous and f vanishes at 'infinity' - meaning that for each $\epsilon > 0$, there exists a compact subset $K \subset X$ such that $|f(x)| < \epsilon$ whenever $x \notin K$;

(iii) f extends to a continuous function $F : \hat{X} \to \mathbb{C}$ with the property that $F(\infty) = 0$.

The set of functions which satisfy these equivalent conditions is denoted by $C_0(X)$.

(b) Let $\mathcal{I} = \{F \in C(\hat{X}) : F(\infty) = 0\}$; then $\mathcal{I}$ is a maximal ideal in $C(\hat{X})$, and the mapping $F \mapsto F|_X$ defines an isometric isomorphism of $\mathcal{I}$ onto $C_0(X)$.

Proof : (a) $(i) \Rightarrow (ii)$: If $\epsilon > 0$, pick n such that $|f_n(x) - f(x)| < \epsilon$; let K be a compact set such that $f_n(x) = 0 \ \forall \ x \notin K$; then, clearly $|f_n(x)| < \epsilon \ \forall \ x \notin K$.

$(ii) \Rightarrow (iii)$: Define $F : \hat{X} \rightarrow \mathbb{C}$ by

$$F(x) \ = \ \begin{cases} f(x) & if \ x \in X \\ 0 & if \ x = \infty \end{cases} \ ;$$

the function F is clearly continuous at all points of X (since X is an open set in $\hat{X}$; the continuity of F at ∞ follows from the hypothesis on f and the definition of open neighbourhoods of ∞ in $\hat{X}$ (as complements of compact sets in X).

$(iii) \Rightarrow (i)$: Fix n. Let $A_n = \{x \in \hat{X} : |F(x)| \geq \frac{1}{n}\}$ and $B_n = \{x \in \hat{X} : |F(x)| \leq \frac{1}{2n}\}$. Note that A_n and B_n are disjoint closed subsets of $\hat{X}$, and that in fact $A_n \subset X$ (so that, in particular, A_n is a compact subset of X). By Urysohn's theorem we can find a continuous function $\phi_n : \hat{X} \rightarrow [0,1]$ such that $\phi_n(A_n) = \{1\}$ and $\phi_n(B_n) = \{0\}$. Consider the function $f_n = (F\phi_n)|_X$; then $f_n : X \rightarrow \mathbb{C}$ is continuous; also, if we set $K_n = \{x \in \hat{X} : |F(x)| \geq \frac{1}{2n}\}$, then K_n is a compact subset of X and $X - K_n \subset B_n$, and so $f_n(x) = 0 \forall \ x \in X - K_n$. Finally, notice that f_n agrees with f on A_n, while if $x \notin A_n$, then

$$|f(x) - f_n(x)| \leq |f(x)| \ (1 + |\phi_n(x)|) \ \leq \frac{2}{n}$$

and consequently the sequence $\{f_n\}$ converges uniformly to f.

(b) This is obvious. □

Now we proceed to the Stone-Weierstrass theorem via its more classical special case, the Weierstrass approximation theorem.

THEOREM A.6.8 (**Weierstrass approximation theorem**) *The set of polynomials is dense in the Banach algebra $C[a,b]$, where $-\infty < a \leq b < \infty$.*

Proof : Without loss of generality - why? - we may restrict ourselves to the case where $a = 0, b = 1$.

We begin with the binomial theorem

$$\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} = (x+y)^n \ ; \qquad \text{(A.6.28)}$$

first set $y = 1 - x$ and observe that

$$\sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} = 1 \ . \tag{A.6.29}$$

Next, differentiate equation A.6.28 once (with respect to x, treating y as a constant) to get:

$$\sum_{k=0}^{n} k \binom{n}{k} x^{k-1} y^{n-k} = n(x+y)^{n-1} \ ; \tag{A.6.30}$$

multiply by x, and set $y = 1 - x$, to obtain

$$\sum_{k=0}^{n} k \binom{n}{k} x^k (1-x)^{n-k} = nx \ . \tag{A.6.31}$$

Similarly, another differentiation (of equation A.6.30 with respect to x), subsequent multiplication by x^2, and the specialisation $y = 1 - x$ yields the identity

$$\sum_{k=0}^{n} k(k-1) \binom{n}{k} x^k (1-x)^{n-k} = n(n-1)x^2 \ . \tag{A.6.32}$$

We may now deduce from the three preceding equations that

$$\begin{aligned} & \sum_{k=0}^{n} (k-nx)^2 \binom{n}{k} x^k (1-x)^{n-k} \\ = \ & n^2x^2 - 2nx \cdot nx + (nx + n(n-1)x^2) \\ = \ & nx(1-x) \ . \end{aligned} \tag{A.6.33}$$

In order to show that any complex-valued continuous function on $[0,1]$ is uniformly approximable by polynomials (with complex coefficients), it clearly suffices (by considering real and imaginary parts) to show that any real-valued continuous function on $[0,1]$ is uniformly approximable by polynomials with real coefficients.

So, suppose now that $f : [0,1] \to \mathbb{R}$ is a continuous real-valued function, and that $\epsilon > 0$ is arbitrary. Since $[0,1]$ is compact, the function f is bounded; let $M = ||f||_\infty$. Also, since f is uniformly continuous, we can find $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ whenever $|x - y| < \delta$. (If you do not know what uniform continuity means, the $\epsilon - \delta$ statement given here is the definition; try to use the compactness of $[0,1]$ and prove this assertion directly.)

We assert that if $p(x) = \sum_{k=0}^{n} f(\frac{k}{n}) \binom{n}{k} x^k (1-x)^{n-k}$, and if n is sufficiently large, than $||f - p|| < \epsilon$. For this, first observe - thanks to equation A.6.29 - that if $x \in [0,1]$ is temporarily fixed, then

$$\begin{aligned} |f(x) - p(x)| &= |\sum_{k=0}^{n} (f(x) - f(\frac{k}{n})) \binom{n}{k} x^k (1-x)^{n-k} | \\ &\leq S_1 + S_2 , \end{aligned}$$

where $S_i = |\sum_{k \in I_i} (f(x) - f(\frac{k}{n})) \binom{n}{k} x^k (1-x)^{n-k}|$, and the sets I_i are defined by $I_1 = \{k : 0 \leq k \leq n, |\frac{k}{n} - x| < \delta\}$ and $I_2 = \{k : 0 \leq k \leq n, |\frac{k}{n} - x| \geq \delta\}$.

Notice now that, by the defining property of δ, that

$$\begin{aligned} S_1 &\leq \sum_{k \in I_1} \epsilon \binom{n}{k} x^k (1-x)^{n-k} \\ &\leq \epsilon \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \\ &= \epsilon , \end{aligned}$$

while

$$\begin{aligned} S_2 &\leq 2M \sum_{k \in I_2} \binom{n}{k} x^k (1-x)^{n-k} \\ &\leq 2M \sum_{k \in I_2} (\frac{k - nx}{n\delta})^2 \binom{n}{k} x^k (1-x)^{n-k} \\ &\leq \frac{2M}{n^2 \delta^2} \sum_{k=0}^{n} (k - nx)^2 \binom{n}{k} x^k (1-x)^{n-k} \\ &= \frac{2Mx(1-x)}{n\delta^2} \\ &\leq \frac{M}{2n\delta^2} \\ &\rightarrow 0 \; as \; n \rightarrow \infty , \end{aligned}$$

where we used (a) the identity A.6.33 in the 4th line above, and (b) the fact that $x(1-x) \leq \frac{1}{4}$ in the 5th line above; and the proof is complete. $\square$

We now proceed to the **Stone-Weierstrass theorem**, which is a very useful generalisation of the Weierstrass theorem.

THEOREM A.6.9 *(a)* (Real version) *Let X be a compact Hausdorff space; let $\mathcal{A}$ be a (not necessarily closed) subalgebra of the real Banach algebra $C_{\mathbb{R}}(X)$; suppose that $\mathcal{A}$ satisfies the following conditions:*

(i) $\mathcal{A}$ contains the constant functions (or equivalently, $\mathcal{A}$ is a unital sub-algebra of $C_{\mathbb{R}}(X)$); and

(ii) $\mathcal{A}$ separates points in X - meaning, of course, that if x, y are any two distinct points in X, then there exists $f \in \mathcal{A}$ such that $f(x) \neq f(y)$.
Then, $\mathcal{A}$ is dense in $C_{\mathbb{R}}(X)$.

(b) (Complex version) *Let X be as above, and suppose $\mathcal{A}$ is a self-adjoint subalgebra of $C(X)$; suppose $\mathcal{A}$ satisfies conditions (i) and (ii) above. Then, $\mathcal{A}$ is dense in $C(X)$.*

Proof : To begin with, we may (replace $\mathcal{A}$ by its closure), consequently assume that $\mathcal{A}$ is closed, and seek to prove that $\mathcal{A} = C_{\mathbb{R}}(X)$ (resp., $C(X)$).

(a) Begin by noting that since the function $t \mapsto |t|$ can be uniformly approximated on any compact interval of $\mathbb{R}$ by polynomials (thanks to the Weierstrass' theorem), it follows that $f \in \mathcal{A} \Rightarrow |f| \in \mathcal{A}$; since $x \vee y = \max\{x, y\} = \frac{x+y+|x-y|}{2}$, and $x \wedge y = \min\{x, y\} = \frac{x+y-x\vee y}{2}$, it follows that $\mathcal{A}$ is a 'sub-lattice' of $C_{\mathbb{R}}(X)$ - meaning that $f, g \in \mathcal{A} \Rightarrow f \vee g, f \wedge g \in \mathcal{A}$.

Next, the hypothesis that $\mathcal{A}$ separates points of X (together with the fact that $\mathcal{A}$ is a vector space containing the constant functions) implies thatif x, y are distinct points in X and if $s, t \in \mathbb{R}$ are arbitrary, then there exists $f \in \mathcal{A}$ such that $f(x) = s, f(y) = t$. (Reason: first find $f_0 \in \mathcal{A}$ such that $f_0(x) = s_0 \neq t_0 = f_0(y)$; next, find constants $a, b \in \mathbb{R}$ such that $as_0 + b = s$ and $at_0 + b = t$, and set $f = af_0 + b1$, where, of course, 1 denotes the constant function 1.)

Suppose now that $f \in C_{\mathbb{R}}(X)$ and that $\epsilon > 0$. Temporarily fix $x \in X$. For each $y \in X$, we can - by the previous paragraph - pick $f_y \in \mathcal{A}$ such that $f_y(x) = f(x)$ and $f_y(y) = f(y)$. Next, choose an open neighbourhood U_y of y such that $f_y(z) > f(z) - \epsilon\, \forall\, z \in U_y$. Then, by compactness, we can find $\{y_1, \cdots, y_n\} \subset X$

such that $X = \cup_{i=1}^n U_{y_i}$. Set $g^x = f_{y_1} \vee f_{y_2} \vee \cdots \vee f_{y_n}$, and observe that $g^x(x) = f(x)$ and that $g^x(z) > f(z) - \epsilon \; \forall \; z \in X$.

Now, we can carry out the procedure outlined in the preceding paragraph, for each $x \in X$. If g^x is as in the last paragraph, then, for each $x \in X$, find an open neighbourhood V_x of x such that $g^x(z) < f(z) + \epsilon \; \forall \; z \in V_x$. Then, by compactness, we may find $\{x_1, \cdots, x_m\} \subset X$ such that $X = \cup_{j=1}^m V_{x_j}$; finally, set $g = g^{x_1} \wedge g^{x_2} \wedge \cdots \wedge g^{x_m}$, and observe that the construction implies that $f(z) - \epsilon < g(z) < f(z) + \epsilon \; \forall \; z \in X$, thereby completing the proof of (a).

(b) This follows easily from (a), upon considering real- and imaginary- parts. (This is where we require that $\mathcal{A}$ is a self-adjoint subalgebra in the complex case.) □

We state some useful special cases of the Stone-Weierstrass theorem in the exercise below.

EXERCISE A.6.10 *In each of the following cases, show that the algebra $\mathcal{A}$ is dense in $C(X)$:*

(i) $X = \mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$, $\mathcal{A} = \bigvee\{z^n : n \in \mathbb{Z}\}$*; thus, $\mathcal{A}$ is the class of 'trigonometric polynomials';*

(ii) X a compact subset of $\mathbb{R}^n$, and $\mathcal{A}$ is the set of functions of the form $f(x_1, \cdots, x_n) = \sum_{k_1,\cdots,k_n=0}^N \alpha_{k_1,\cdots,k_n} x_1^{k_1} \cdots x_n^{k_n}$, *where* $\alpha_{k_1,\cdots,k_n} \in \mathbb{C}$;

(iii) X a compact subset of $\mathbb{C}^n$, and $\mathcal{A}$ is the set of (polynomial) functions of the form

$$f(z_1, \cdots, z_n) = \sum_{k_1,l_1,\cdots,k_n,l_n=0}^{N} \alpha_{k_1,l_1,\cdots,k_n,l_n} z_1^{k_1} \overline{z_1}^{l_1} \cdots z_n^{k_n} \overline{z_n}^{l_n} \;;$$

(iv) $X = \{1, 2, \cdots, N\}^{\mathbb{N}}$, *and* $\mathcal{A} = \{\omega_{k_1,\cdots,k_n} : n \in \mathbb{N}, 1 \leq k_1, \cdots, k_n \leq N\}$, *where*

$$\omega_{k_1,\cdots,k_n}((x_1, x_2, \cdots)) = exp\left(\frac{2\pi i \sum_{j=1}^n k_j x_j}{N}\right) .$$

The 'locally compact' version of the Stone-Weierstrass theorem is the content of the next exercise.

EXERCISE A.6.11 *Let $\mathcal{A}$ be a self-adjoint subalgebra of $C_0(X)$, where X is a locally compact Hausdorff space. Suppose $\mathcal{A}$ satisfies the following conditions:*

(i) if $x \in X$, then there exists $f \in \mathcal{A}$ such that $f(x) \neq 0$; and

(ii) $\mathcal{A}$ separates points.

Then show that $\mathcal{A} = C_0(X)$. (Hint: Let $\mathcal{B} = \{F + \alpha 1 : f \in \mathcal{A}, \alpha \in \mathbb{C}\}$, where 1 denotes the constant function on the one-point compactification $\hat{X}$, and F denotes the unique continuous extension of f to $\hat{X}$; appeal to the already established compact case of the Stone-Weierstrass theorem.)

A.7 The Riesz Representation theorem

This brief section is devoted to the statement and some comments on the Riesz representation theorem - which is an identification of the Banach dual space of $C(X)$, when X is a compact Hausdorff space.

There are various formulations of the Riesz representation theorem; we start with one of them.

THEOREM A.7.1 *(**Riesz representation theorem**)*

Let X be a compact Hausdorff space, and let $\mathcal{B} = \mathcal{B}_X$ denote the Borel σ-algebra (which is generated by the class of all open sets in X).

(a) Let $\mu : \mathcal{B} \to [0, \infty)$ be a finite measure; then the equation

$$\phi_\mu(f) = \int_X f d\mu \tag{A.7.34}$$

defines a bounded linear functional $\phi_\mu \in C(X)^*$ which is positive - meaning, of course, that $f \geq 0 \Rightarrow \phi_\mu(f) \geq 0$.

(b) Conversely, if $\phi \in C(X)^*$ is a bounded linear functional which is positive, then there exists a unique finite measure $\mu : \mathcal{B} \to [0, \infty)$ such that $\phi = \phi_\mu$.

Before we get to other results which are also referred to as the Riesz representation theorem, a few words concerning general 'complex measures' will be appropriate. We gather some facts concerning such 'finite complex measures' in the next proposition. We do not go into the proof of the proposition; it may be found in any standard book on analysis - see [Yos], for instance.

PROPOSITION A.7.2 *Let* $(X, \mathcal{B})$ *be any measurable space. Suppose* $\mu : \mathcal{B} \to \mathbb{C}$ *is a 'finite complex measure' defined on* $\mathcal{B}$ *- i.e., assume that* $\mu(\emptyset) = 0$ *and that* μ *is countably additive in the sense that whenever* $\{E_n : 1 \leq n < \infty\}$ *is a sequence of pairwise disjoint members of* $\mathcal{B}$*, then it is the case that the series* $\sum_{n=1}^{\infty} \mu(E_n)$ *of complex numbers is convergent to the limit* $\mu(\cup_{n=1}^{\infty} E_n)$.

Then, the assignment

$$\mathcal{B} \ni E \ \mapsto \ |\mu|(E) \ = \ \sup\{\sum_{n=1}^{\infty} |\mu(E_n)| : \ E = \coprod_{n=1}^{\infty} E_n, E_n \in \mathcal{B} \ \}$$

defines a finite positive measure $|\mu|$ *on* $(X, \mathcal{B})$.

Given a finite complex measure μ on $(X, \mathcal{B})$ as above, the positive measure $|\mu|$ is usually referred to as the *total variation measure* of μ, and the number $||\mu|| = |\mu|(X)$ is referred to as the **total variation** of the measure μ. It is an easy exercise to verify that the assignment $\mu \mapsto ||\mu||$ satisfies all the requirements of a norm, and consequently, the set $M(X)$ of all finite complex measures has a natural structure of a normed space.

THEOREM A.7.3 *Let* X *be a compact Hausdorff space. Then the equation*

$$\phi_\mu(f) \ = \ \int_X f d\mu$$

yields an isometric isomorphism $M(X) \ni \mu \mapsto \phi_\mu \in C(X)^*$ *of Banach spaces.*

We emphasise that a part of the above statement is the asserton that the norm of the bounded linear functional ϕ_μ is the total variation norm $||\mu||$ of the measure μ.

Bibliography

[Ahl] Lars V. Ahlfors, *Complex Analysis*, McGraw-Hill, 1966.

[Art] M. Artin, *Algebra*, Prentice Hall of India, New Delhi, 1994.

[Arv] W. Arveson, *An invitation to C^*-algebra*, Springer-Verlag, New York, 1976.

[Ban] S. Banach, *Opérations linéaires*, Chelsea, New York, 1932.

[Hal] Paul R. Halmos, *Naive Set Theory*, Van Nostrand, Princeton, 1960.

[Hal1] Paul R. Halmos, *Measure Theory*, Van Nostrand, Princeton, 1950.

[Hal2] Paul R. Halmos, *Finite-dimensional Vector Spaces*, Van Nostrand, London, 1958.

[MvN] F.A. Murray and John von Neumann, *On Rings of Operators*, Ann. Math., 37, (1936), 116-229.

[Rud] Walter Rudin, *Functional Analysis*, Mc-Graw Hill, 1973.

[Ser] J.- P. Serre, *Linear Representations of Finite Groups*, Springer-Verlag, New York, 1977.

[Sim] George F. Simmons, *Topology and Modern Analysis*, McGraw-Hill Book Company, 1963.

[Sto] M.H. Stone, *Linear transformations in Hilbert Space and their applications to analysis*, A.M.S., New York, 1932.

[Yos] K. Yosida, *Functional Analysis*, Springer, Berlin, 1965.

Index