Introduction

These are notes for a King's College course to fourth year undergraduates and MSc students. They cover the theoretical development of operators on Hilbert space up to the spectral theorem for bounded selfadjoint operators. I have tried to make the treatment as elementary as possible and to include only what is essential to the main development. The proofs are the simplest and most intuitive that I know. The exercises are culled from various sources; some of them are more or less original. They are designed to be worked as the course progresses and, in some cases, they anticipate results and techniques that come in the later theorems of the course.

It should be emphasized that these notes are not complete. Although the theoretical development is covered rather fully, examples, illustrations and applications which are essential for the understanding of the subject, are left to be covered in the lectures. There are good reasons for doing this. Experience has shown that audiences lose concentration if they are provided with comprehensive notes which coincide with the lectures. Also, in many cases examples and such are best treated in a less formal way which is more suited to oral presentation. In this way it is possible to cater for different sections of an audience with a mixed background. A formal proof may be indicated to some while others may have to take certain statements on trust. This is especially the case when integration spaces are involved.

I would like to thank the many students and colleagues who have pointed out errors and obscurities in earlier versions of these notes. A few proofs contain some sentences in square brackets. These indicate explanations that I consider rather obvious and should be superfluous to a formal proof but were added in response to some query.

For the benefit of a wider audience, here is a brief indication of what might be covered to supplement the notes and also a few comments.

Section 1. Examples of inner product spaces : $\ell^2_n(= \mathbb{C}^n)$ and ℓ^2. Continuous functions on $[a, b]$ with $\langle f, g\rangle = \int_a^b f(t)\overline{g(t)}\,dt$ and problems with extending to larger classes of functions (equivalence classes as elements of the space). Completeness of $\ell^2_n(= \mathbb{C}^n)$ and ℓ^2 (and continuous functions on $[a, b]$ not complete).

$L^2[a, b]$. Some brief discussion of the Lebesgue integral. The following statement to be known or accepted: there is a definition of the integral such that the (equivalence classes) of all functions f such that $\int_a^b |f(t)|^2\,dt$ exists and is finite forms a Hilbert space with the inner product $\langle f, g\rangle = \int_a^b f(t)\overline{g(t)}\,dt$ (that is, it is complete). Some more general L^2 spaces might be mentioned (e.g. $L^2(S)$ where $S = [a, b] \times [a, b]$ or some other subset of $\mathbb{R}^n$).

Examples of normed spaces which cannot be Hilbert spaces because they do not satisfy the parallelogram law ($C[a, b], \ell^p_n (n = 2)$ for $p \neq 2$).

Examples of normed spaces with closed convex sets where the distance from a point is not attained uniquely (e.g. the unit ball in ℓ^1_2 with the point $(1, 1)$) or not attained at all (e.g. the space $X = \{f \in C[0, 1] : f(0) = 0\}$, the set $\{f \in X : \int_0^1 f(t)\,dt = 1\}$ and the zero function as the point).

Some indications of the applications of the minimum distance theorem, e.g. to approximation theory and optimal control theory.

Section 2. This section is supplemented by specific examples of operators on ℓ^2 and $L^2[a,b]$. These include diagonal operators, shifts (forward, backward and weighted) on ℓ^2, the bilateral shift on $\ell^2(\mathbb{Z})$ and the following operators on $L^2[a,b]$.

Multiplication operator : $(M_\phi f)(x) = \phi(x).f(x) \qquad \phi$ (essentially) bounded.

Fredholm integral operator : $(Kf)(x) = \int_a^b k(x,t)f(t)\,dt$ where $\int\int |k|^2 < \infty$,

with the Volterra operator, $(Vf)(x) = \int_a^x f(t)\,dt$ as a special case.

The boundedness of these operators should be established and the adjoints identified. Other examples of finding adjoints (similar to those in the exercises) might be done.

Section 3. The main additional topic for this section is the connection to classical Fourier series. The fact that the normalized trigonometric functions for an orthonormal basis of $L^2[-\pi,\pi]$ should be established or accepted. One route uses the Stone-Weierstrass Theorem and the density of $C[-\pi,\pi]$ in $L^2[-\pi,\pi]$. Inevitably, this requires background in metric spaces and Lebesgue theory. Note that this fact is also established, albeit in a roundabout way, by the work in Section 7.

The projection onto a subspace can be written down in terms of an orthonormal basis of the subspace : $P_N h = \sum \langle h, y_i\rangle y_i$ where $\{y_i\}$ is an orthonormal basis of N.

Applying the Gram-Schmidt process to the polynomials in the Hilbert space $L^2[-1,1]$ gives (apart from constant factors) the Legendre polynomials. Similarly the Hermite and the Laguerre polynomials arise from orthonormalizing $\{x^n e^{-x^2}\}$ and $\{x^n e^{-x}\}$ in the spaces $L^2[-\infty,\infty]$ and $L^2[0,\infty]$ respectively.

Section 4. Some of the operators introduced in Section 2 should be examined for compactness. In particular, the conditions for a diagonal operator on ℓ^2 to be compact should be established.

Theorem 4.4 and Lemma 4.3 on which its proof depends are the only results in these notes which are not strictly needed for what comes later. Note that this result is not valid in general Banach spaces.

Section 5. The spectra of some specific operators should be identified. In particular, the spectrum of M_ϕ where $\phi(x) = x$ on $L^2[0,1]$ should be identified as $[0,1]$ and the fact that M_ϕ has no eigenvalues should be noted.

Section 6. The fact that the Volterra operator has no eigenvalues should be established, hence showing that some compact operators may have spectrum equal to $\{0\}$.

It is useful to review the orthogonal diagonalization of real symmetric matrices and/or unitary diagonalization of Hermitian (i.e. selfadjoint) matrices. It is instructive to re-write both these results and Theorem 6.9 in terms of projections onto eigenspaces.

Section 7. This section is motivated by an informal discussion of of the Green's function as the response of a system to the input of a unit pulse. This is illustrated by the elementary example of finding the shape under gravity of a heavy string (of variable density) fixed at $(0,0)$ and $(\ell,0)$. This is found by calculating the (triangular) shape $k(x,t)$ of a weightless string with a unit weight at $x = t$ and then using an integration process. The differential equation is also found and shown to give the same answer. (Naturally, usual elementary applied mathematical assumptions - small displacements, constant tension - apply.)

Additionally, a brief, very informal discussion of delta functions and the Green's function as the solution of the system with the function f being a delta function is of interest.

It should be stressed, however, that the proof of Theorem 7.1 is purely elementary and quite independent of the discussions above.

Section 8. The most important part of this final section is Theorem 8.3, the continuous functional calculus. This is sufficient for the vast majority of applications of the spectral theorem for bounded self-adjoint operators. These include, for example, the polar decomposition and the properties of e^{At} and these are done in the course.

The approach here to the general spectral theorem is elementary and very pedestrian. It should be noted that, given the appropriate background, there are more elegant ways. These include using the identification of the dual of $C[m, M]$ (actually of $C(\sigma)$) as a space of measures. There is also a Banach algebra treatment.

In an elementary course such as this, the technicalities of the spectral theorem need not be strongly emphasized. However, a down to earth approach should clarify the meaning of theorem and remove the mystery often attached by students to these operator integrals.

1 Elementary properties of Hilbert space

Definition A (complex) *inner (scalar) product space* is a vector space $\mathcal{H}$ together with a map $\langle \cdot\,,\cdot\rangle : \mathcal{H}\times\mathcal{H}\to\mathbb{C}$ such that, for all $x,y,z\in\mathcal{H}$ and $\lambda,\mu\in\mathbb{C}$,

1. $\langle \lambda x+\mu y, z\rangle = \lambda\langle x,z\rangle + \mu\langle y,z\rangle$,
2. $\langle x,y\rangle = \overline{\langle y,x\rangle}$,
3. $\langle x,x\rangle \geq 0$, and $\langle x,x\rangle = 0 \Longleftrightarrow x=0\,.$

   Properties 1,2 and 3 imply
4. $\langle (x, \lambda y+\mu z\rangle = \bar\lambda\langle x,y\rangle + \bar\mu\langle x,z\rangle$,
5. $\langle x,0\rangle = \langle 0,x\rangle = 0$.

Theorem 1.1 (Cauchy-Schwartz inequality)

$$|\langle x,y\rangle| \leq \langle x,x\rangle^{1/2}\langle y,y\rangle^{1/2}, \quad \forall x,y\in\mathcal{H}.$$

Proof. For all λ we have $\langle \lambda x+y, \lambda x+y\rangle \geq 0$. That is, for real λ

$$\lambda^2\langle x,x\rangle + \lambda(\langle x,y\rangle + \langle y,x\rangle) + \langle y,y\rangle \geq 0\,.$$

In the case that $\langle x,y\rangle$ is real, we have that the discriminant ("b^2-4ac") of this quadratic function of λ is negative which gives the result.

In general, put $x_1 = e^{-i\theta}x$ where θ is the argument of the complex number $\langle x,y\rangle$. Then $\langle x_1,y\rangle = e^{i\theta}\langle x,y\rangle = |\langle x,y\rangle|$ is real and $\langle x_1,x_1\rangle = \langle x,x\rangle$. Applying the above to x_1,y gives the required result. ■

[Alternatively, put $\lambda = -\frac{\langle y,x\rangle}{\langle x,x\rangle}$ in $\langle \lambda x+y, \lambda x+y\rangle \geq 0$.]

Theorem 1.2

$$\|x\| = \langle x,x\rangle^{1/2} \textit{ is a norm on } \mathcal{H}.$$

Proof. The facts that $\|x\|\geq 0$, $\|x\| = 0 \Longleftrightarrow x=0$ and $\|\lambda x\| = \langle \lambda x,\lambda x\rangle^{\frac{1}{2}} = |\lambda|.\|x\|$ are all clear from the equivalent properties of the inner product. For the triangle inequality,

$$\begin{aligned}\|x+y\|^2 &= \|x\|^2+\|y\|^2+\langle x,y\rangle+\langle y,x\rangle\\ &\leq \|x\|^2+\|y\|^2+2|\langle x,y\rangle|\\ &\leq \|x\|^2+\|y\|^2+2\|x\|.\|y\| \qquad \text{using (1.1)}\\ &= (\|x\|+\|y\|)^2\,.\ \blacksquare\end{aligned}$$

Lemma 1.3 (Polarization identity)

$$\langle x,y\rangle = \frac{1}{4}\left[\|x+y\|^2 - \|x-y\|^2 + i\|x+iy\|^2 - i\|x-iy\|^2\right]\,.$$

Proof.

$$\begin{aligned} \|x+y\|^2 &= \|x\|^2 + \|y\|^2 + \langle x,y\rangle + \langle y,x\rangle \\ -\|x-y\|^2 &= -\|x\|^2 + -\|y\|^2 + \langle x,y\rangle + \langle y,x\rangle \\ i\|x+iy\|^2 &= i\|x\|^2 + i\|y\|^2 + \langle x,y\rangle - \langle y,x\rangle \\ -i\|x-iy\|^2 &= -i\|x\|^2 + -i\|y\|^2 + \langle x,y\rangle - \langle y,x\rangle\,. \end{aligned}$$

Adding the above gives the result. ■

Lemma 1.4 (Paralellogram law)

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2\,.$$

Proof.

$$\begin{aligned} \|x+y\|^2 &= \|x\|^2 + \|y\|^2 + \langle x,y\rangle + \langle y,x\rangle \\ \|x-y\|^2 &= \|x\|^2 + \|y\|^2 - \langle x,y\rangle - \langle y,x\rangle\,. \end{aligned}$$

Adding the above gives the result. ■

Definition x is said to be *orthogonal* to y if $\langle x,y\rangle = 0$; we write $x \perp y$.

Lemma 1.5 (Theorem of Pythagoras)

$$\langle x,y\rangle = 0 \Longrightarrow \|x\|^2 + \|y\|^2 = \|x+y\|^2\,.$$

Proof. Obvious. ■

Definition. If $\mathcal{H}$ is an inner product space and $(\mathcal{H}, \|\cdot\|\ )$ is complete then $\mathcal{H}$ is called a *Hilbert space.*

A set C (in a vector space) is *convex* if

$$x, y \in C \Longrightarrow \alpha x + (1-\alpha)y \in C \text{ whenever } 0 \le \alpha \le 1\,.$$

In a metric space, the distance from a point x to a set S is $d(x,S) = \inf\{\|x-s\| : s \in S\}$.

Theorem 1.6 *If K is a closed convex set in a Hilbert space $\mathcal{H}$ and $h \in \mathcal{H}$ then there exists a unique $k \in K$ such that*

$$d(h,K) = \|h-k\|.$$

Proof. Let $C = K - h = \{k - h : k \in K\}$. Note that C is also closed and convex, $d(h,K) = d(0,C)$ and if $c = k - h \in C$ is of minimal norm then k is the required element of K. Therefore it is sufficient to prove the theorem for the case $h = 0$.

Let $d = d(0, C) = \inf_{c\in C} \|c\|$. The $\|c\| \geq d$ for all $c \in C$. Choose a sequence (c_n) such that $(\|c_n\|) \to d$. Using the parallelogram law (Lemma 1.4),

$$\|c_n + c_m\|^2 + \|c_n - c_m\|^2 = 2\|c_n\|^2 + 2\|c_m\|^2 .$$

But, since C is convex, $\frac{c_n+c_m}{2} \in C$ and so $\|\frac{c_n+c_m}{2}\| \geq d$; that is $\|c_n + c_m\|^2 \geq 4d^2$. Therefore

$$\begin{aligned} 0 \leq \|c_n - c_m\|^2 &= 2\|c_n\|^2 + 2\|c_m\|^2 - \|c_n + c_m\|^2 \\ &\leq 2(\|c_n\|^2 + \|c_m\|^2) - 4d^2 \to 0 \qquad (*) \end{aligned}$$

as $n, m \to \infty$. It follows easily that (c_n) is a Cauchy sequence. [Since $(\|c_n\|) \to d$, given $\epsilon > 0$, there exists n_0 such that for $n > n_0$, $2(\|c_n\|^2 - d^2) < \frac{\epsilon^2}{2}$. Then (*) shows that for $n, m > n_0$, $\|c_n - c_m\| < \epsilon$.] Since $\mathcal{H}$ is complete and C is closed, (c_n) converges to an element $c \in C$ and $\|c\| = \lim_{n\to\infty} \|c_n\| = d$.

To prove uniqueness, suppose also that $c' \in C$ with $\|c'\| = d$. The same calculation as for (*) (with $c_n = c$ and $c_m = c'$) shows that

$$0 \leq \|c - c'\|^2 \leq 2\|c\|^2 + 2\|c'\|^2 - \|c + c'\|^2 \leq 2d^2 + 2d^2 - 4d^2 = 0$$

and so $c = c'$. ■

Lemma 1.7 *If N is a closed subspace of a Hilbert space $\mathcal{H}$ and $h \in \mathcal{H}$ then*

$$d(h, N) = \|h - n_0\| \text{ if and only if } \langle h - n_0, n\rangle = 0 \text{ for all } n \in N .$$

Proof. Suppose $d(h, N) = \|h - n_0\|$. Write $z = h - n_0$. Then for all non-zero $n \in N$,

$$\begin{aligned} \|z\|^2 &\leq \left\| z - \frac{\langle z, n\rangle n}{\|n\|^2} \right\|^2 \\ &= \|z\|^2 - \frac{2|\langle z, n\rangle|^2}{\|n\|^2} + \frac{|\langle z, n\rangle|^2}{\|n\|^2} \\ &= \|z\|^2 - \frac{|\langle z, n\rangle|^2}{\|n\|^2} \end{aligned}$$

so $\langle z, n\rangle = 0$.

Conversely if $h - n_0 \perp N$ then, by Pythagoras (Lemma 1.5) for all $n \in N$,

$$\|h - n\|^2 = \|h - n_0 + n_0 - n\|^2 = \|h - n_0\|^2 + \|n_0 - n\|^2 \geq \|h - n_0\|^2 .$$

Hence $\inf_{n\in N} \|h - n\|$ is attained at n_0. ■

Note that the above proof is putting a geometrical argument into symbolic form. The quantity $\frac{\langle z,n\rangle n}{\|n\|^2}$ is the "resolution of the vector z in the direction of n".

In these notes the term *subspace* (of a Hilbert space) will always mean a *closed* subspace. The justification for this is that the prefix "sub" refers to a substructure; so the subspace should be a Hilbert space in its own right, that is, it should be complete. But it is an easy fact that a subset of a complete space is complete if and only if it is closed.

Definition. Given a subset S of $\mathcal{H}$ the orthogonal complement $S^{\perp}$ is defined by

$$S^{\perp} = \{x : \langle x, s\rangle = 0 \text{ for all } s \in S\}.$$

Corollary 1.8 *If N is a (closed) subspace of a Hilbert space $\mathcal{H}$,*

$$N^{\perp} = (0) \Longleftrightarrow N = \mathcal{H}$$

Proof. Clearly, if $N = \mathcal{H}$ then $N^{\perp} = (0)$. For the converse, if $N \neq \mathcal{H}$ take $h \notin N$, Then there is $n_0 \in N$ such that $d(h, N) = \|h - n_0\|$ and the Lemma shows that $0 \neq h - n_0 \perp N$, so $N^{\perp} \neq (0)$. ■

Lemma 1.9 *For subsets of a Hilbert space $\mathcal{H}$*

(i) $S^{\perp}$ is a closed subspace,
(ii) $S_1 \supseteq S_2 \Longrightarrow S_1^{\perp} \subseteq S_2^{\perp}$,
(iii) $S \subseteq S^{\perp\perp}$,
(iv) $S^{\perp} = S^{\perp\perp\perp}$,
(v) $S \cap S^{\perp} = (0)$.

Proof. (i) Clearly $S^{\perp}$ is a vector subspace. To show it is closed, let $t_n \in S^{\perp}$ be a sequence converging to t. Then, by the continuity of the inner product, for all $s \in S$,

$$\langle t, s\rangle = \lim_{n\to\infty} \langle t_n, s\rangle = 0$$

so $t \in S^{\perp}$. [In grim detail, $|\langle t, s\rangle| = |\langle t - t_n, s\rangle| \leq \|t - t_n\|.\|s\| \to 0$, so, since $\langle t, s\rangle$ does not depend on n, $\langle t, s\rangle = 0$.]

(ii) and (iii) are clear. For (iv), apply (iii) to $S^{\perp}$ yields $S^{\perp} \subseteq S^{\perp\perp\perp}$, and applying (ii) to (iii) to gives the reverse inclusion. For (v), if $x \in S \cap S^{\perp}$ then $\langle x, x\rangle = 0$ so $x = 0$. ■

Lemma 1.10 *If M and N are orthogonal subspaces of a Hilbert space then $M \oplus N$ is closed.*

Proof. Note that since $N \perp M$, we have that $N \cap M = (0)$ and the sum $M + N$ is automatically direct. Let $z_n \in M \oplus N$ such that $(z_n) \to z$. We need to show that $z \in M \oplus N$. Now $z_n = x_n + y_n$ with $x_n \in N$ and $y_n \in M$. Therefore, using Pythagoras (Lemma 1.5) since $M \perp N$,

$$\|z_{n+p} - z_n\|^2 = \|x_{n+p} - x_n\|^2 + \|y_{n+p} - y_n\|^2.$$

As (z_n) is convergent, it is a Cauchy sequence. If follows easily from the above that both (x_n) and (y_n) are Cauchy sequences so, since $\mathcal{H}$ is complete, (x_n) and (y_n) both converge. Call the limits x and y. Then, since M and N are closed subspaces, $x \in M$ and $y \in N$. Thus $z = \lim(x_n + y_n) = x + y \in M \oplus N$. ■

Theorem 1.11 *If N is a subspace of a Hilbert space $\mathcal{H}$ then $N \oplus N^\perp = \mathcal{H}$.*

Proof. From above, $N \oplus N^\perp$ is a (closed) subspace. Also, if $x \in (N \oplus N^\perp)^\perp$ then $x \in N^\perp \cap N^{\perp\perp}$ so $x = 0$. Therefore, from Corollary 1.8, $N \oplus N^\perp = \mathcal{H}$.

Corollary 1.12

(i) *If N is a subspace then $N^{\perp\perp} = N$.*

(ii) *For any subset S of a Hilbert space $\mathcal{H}$, $S^{\perp\perp}$ is the smallest subspace containing S.*

Proof. (i) From Lemma 1.9 (iii) $N \subseteq N^{\perp\perp}$. Since $\mathcal{H} = N \oplus N^\perp$, if $x \in N^{\perp\perp}$ then $x = s + t$ with $s \in N$ and $t \in N^\perp$. But then also $t = x - s \in N^{\perp\perp}$, so $t = 0$ and $x = s \in N$.

(ii) Clearly $S^{\perp\perp}$ is a subspace containing S. If M is any subspace containing S then (Lemma 1.9 (ii)) $S^{\perp\perp} \subseteq M^{\perp\perp} = M$. ■

Exercises 1

1. For a Hilbert space H, show that the inner product, considered as a map from $H \times H$ to $\mathbb{C}$, is continuous.

2. Let H_1 and H_2 be Hilbert spaces. Let H be the set of ordered pairs $H_1 \times H_2$ with addition and multiplication defined (in the usual way) as follows:

$$\begin{aligned}(h_1, h_2) + (g_1, g_2) &= (h_1 + g_1, h_2 + g_2) \\ \alpha(h_1, h_2) &= (\alpha h_1, \alpha h_2).\end{aligned}$$

Show that the inner product defined by

$$\langle (h_1, h_2), (g_1, g_2) \rangle = \langle h_1, g_1 \rangle + \langle h_2, g_2 \rangle$$

satisfies the axioms for an inner product and H with this inner product is a Hilbert space. [H is called the (Hilbert space) direct sum of H_1 and H_2. One writes $H = H_1 \oplus H_2$.]

3. Prove that in the Cauchy–Schwarz inequality $|\langle x, y \rangle| \le \|x\|\|y\|$ the equality holds iff the vectors x and y are linearly dependent.

4. For which real α does the function $f(t) = t^\alpha$ belong to

(i) $L^2[0,1]$ (ii) $L^2[1,\infty]$ (iii) $L^2[0,\infty]$?

5. Let x and y be vectors in an inner product space. Given that $\| x+y \| = \| x \| + \| y \|$, show that x and y are linearly dependent.

6. Let $W[0,1]$ be the space of complex-valued functions which are continuously differentiable on $[0,1]$. Show that,

$$\langle f, g \rangle = \int_0^1 \{f(t)\overline{g(t)} + f'(t)\overline{g'(t)}\}\, dt$$

defines an inner product on $W[0,1]$.

7. Prove that in a complex inner product space the following equalities hold:

$$\begin{aligned}\langle x, y \rangle &= \frac{1}{N}\sum_{k=1}^{N} \|x + e^{2\pi i k/N} y\|^2 e^{2\pi i k/N} \quad \text{for } N \ge 3, \\ \langle x, y \rangle &= \frac{1}{2\pi}\int_0^{2\pi} \|x + e^{i\theta} y\|^2 e^{i\theta} d\theta\,.\end{aligned}$$

[This generalizes the polarization identity.]

8. Let M and N be closed subspaces of a Hilbert space. Show that

(i) $(M+N)^\perp = M^\perp \cap N^\perp$ (ii) $(M \cap N)^\perp = \overline{M^\perp + N^\perp}$.

9. Show that the vector subspace of ℓ^2 spanned by sequences of the form $(1, \alpha, \alpha^2, \alpha^3, \ldots)$, where $0 \le \alpha < 1$, is dense in ℓ^2.

A challenging but not very important exercise :

10. Show that, for any four points of a Hilbert space,

$$\| x - z \| \,.\, \| y - t \| \le \| x - y \| \,.\, \| z - t \| + \| y - z \| \,.\, \| x - t \| \,.$$

2 Linear Operators.

Some of the results in this section are stated for normed linear spaces but they will be used in the sequel only for Hilbert spaces.

Lemma 2.1 *Let X and Y be normed linear spaces and let $L : X \to Y$ be a linear map. Then the following are equivalent :*

1. *L is continuous;*
2. *L is continuous at 0;*
3. *there exists a constant K such that $\|Lx\| \leq K\|x\|$ for all $x \in X$.*

Proof. 1 implies 2 is obvious. If 2 holds, take any $\epsilon > 0$. Continuity at 0 shows that there is a corresponding $\delta > 0$ such that $\|LX\| < \epsilon$ whenever $\|x\| < \delta$. Take some c with $0 < c < \delta$. Then for any $x \neq 0$, $\left\|\frac{cx}{\|x\|}\right\| = c < \delta$ and so

$$\left\|L\left(\frac{cx}{\|x\|}\right)\right\| = c\frac{\|Lx\|}{\|x\|} < \epsilon.$$

This shows that $\|Lx\| < K\|x\|$ where $K = \frac{\epsilon}{c}$.

If 3 holds, to show continuity at any point x_0, note that

$$\|Lx - Lx_0\| = \|L(x - x_0)\| \leq K\|x - x_0\|.$$

Therefore, given any $\epsilon > 0$, let $\delta = \frac{\epsilon}{K}$. Then if $\|x - x_0\| < \delta$ we have $\|Lx - Lx_0\| < \epsilon$.
■

The set of all continuous (bounded) linear maps $X \to Y$ is denoted by $\mathcal{B}(X, Y)$. When $X = Y$ we write $\mathcal{B}(X)$.

For $L \in \mathcal{B}(X, Y)$, define $\|L\| = \sup_{x\neq 0} \frac{\|Lx\|}{\|x\|}$.

Exercise. $\|\cdot\|$ is a norm on $\mathcal{B}(X, Y)$ and

$$\|L\| = \sup_{x\neq 0} \frac{\|Lx\|}{\|x\|} = \sup_{\|x\|\leq 1} \|Lx\| = \sup_{\|x\|=1} \|Lx\|.$$

If Y is complete then so is $\mathcal{B}(X, Y)$

When $Y = \mathbb{C}$ then $\mathcal{B}(X, \mathbb{C})$ is called the *dual* of X and denoted by X' (sometimes by X^*). The elements of the dual are called (continuous) *linear functionals.*

We shall be concerned with Hilbert spaces; $\mathcal{H}$ will always denote a Hilbert space.

Theorem 2.2 (Riesz representation theorem) *Every linear functional f on $\mathcal{H}$ is of the form*

$$f(x) = \langle x, h\rangle$$

for some $h \in \mathcal{H}$, where $\|f\| = \|h\|$.

Proof. If $f = 0$, take $h = 0$. For $f \neq 0$ then $N = f^{-1}(0) = \{x : f(x) = 0\} \neq \mathcal{H}$. Also, since f is continuous, N is closed. Thus $N^{\perp} \neq (0)$ so take $y \perp N$. Then $f(y) \neq 0$. Write $z = \frac{y}{f(y)}$ so that $f(z) = 1$ [using $f(\alpha x) = \alpha f(x)$]. For any $x \in \mathcal{H}$

$$f\left(x - f(x)z\right) = f(x) - f(x).f(z) = 0 \text{ and so } x - f(x)z \in N\,.$$

Since $z \perp N$,

$$\langle x - f(x)z, z\rangle = \langle x, z\rangle - f(x)\|z\|^2 = 0.$$

Writing $h = \frac{z}{\|z\|^2}$ we obtain

$$f(x) = \langle x, h\rangle\,.$$

For the norm, note that $|f(x)| = |\langle x, h\rangle| \leq \|x\|.\|h\|$ so $\|f\| \leq \|h\|$. Also

$$\|f| = \sup_{x\neq 0}\frac{|f(x)|}{\|x\|} \geq \frac{|f(h)|}{\|h\|} = \|h\|\,.$$

■

Note that the result $\|f\| = \|h\|$ shows that the correspondence between $\mathcal{H}$ and its dual is one to one.

Lemma 2.3 (Polarization identity for operators)

$$\begin{aligned}\langle Ax, y\rangle &= \frac{1}{4}[\langle A(x+y),(x+y)\rangle - \langle A(x-y),(x-y)\rangle \\ &\quad +i\langle A(x+iy),(x+iy)\rangle - i\langle A(x-iy),(x-iy)\rangle]\,.\end{aligned}$$

Proof.

$$\begin{aligned}\langle A(x+y),(x+y)\rangle &= \langle Ax,x\rangle + \langle Ay,y\rangle + \langle Ax,y\rangle + \langle Ay,x\rangle \\ -\langle A(x-y),(x-y)\rangle &= -\langle Ax,x\rangle - \langle Ay,y\rangle + \langle Ax,y\rangle + \langle Ay,x\rangle \\ i\langle A(x+iy),(x+iy)\rangle &= i\langle Ax,x\rangle + i\langle Ay,y\rangle + \langle Ax,y\rangle - \langle Ay,x\rangle \\ -i\langle A(x-iy),(x-iy)\rangle &= -i\langle Ax,x\rangle - i\langle Ay,y\rangle + \langle Ax,y\rangle - \langle Ay,x\rangle\,.\end{aligned}$$

Adding the above gives the result. ■

Corollary 2.4 *If $\langle Ax, x\rangle = 0$ for all $x \in \mathcal{H}$ then $A = 0$.*

Proof. If $\langle Ax, x\rangle = 0$ for all $x \in \mathcal{H}$ the above shows that $\langle Ax, y\rangle = 0$ for all $x, y \in \mathcal{H}$ and so using $y = Ax$ it follows that $\|Ax\|^2 = 0$ for all $x \in \mathcal{H}$. Thus $A = 0$. ■

Definition Let $\mathcal{H}$ be a Hilbert space. A *bilinear form* (also called a *sesquilinear form*) ϕ on $\mathcal{H}$ is a map $\phi : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ such that

$$\begin{aligned}\phi(\alpha x + \beta x', y) &= \alpha\phi(x,y) + \beta\phi(x',y) \\ \phi(x, \alpha y + \beta y') &= \bar{\alpha}\phi(x,y) + \bar{\beta}\phi(x,y')\,.\end{aligned}$$

A bilinear form is said to be *bounded* if, for some constant K, $|\phi(x,y)| \leq K\|x\|.\|y\|$ for all $x, y \in \mathcal{H}$.

Theorem 2.5 (Riesz) *Every bounded bilinear form ϕ on $\mathcal{H}$ is of the form*

$$\phi(x,y) = \langle Ax, y\rangle$$

for some $A \in \mathcal{B}(\mathcal{H})$.

Proof. Consider x fixed for the moment. Then $\phi(x,y)$ is conjugate linear in y, so that $\overline{\phi(x,y)}$ is linear in y. Using Theorem 2.2 we have that there is a (unique) $h \in \mathcal{H}$,

$$\overline{\phi(x,y)} = \langle y, h\rangle\,, \text{ that is } \phi(x,y) = \langle h, y\rangle\,.$$

One can find such an h corresponding to each $x \in \mathcal{H}$. Define a function $\mathcal{H} \to \mathcal{H}$ by $Ax = h$. Then A is linear since , for all x, x', y,

$$\langle A(x+x'), y\rangle \; = \phi((x+x'), y) = \phi(x,y) + \phi(x', y) = \langle Ax, y\rangle \; + \langle Ax', y\rangle$$

so $A(x+x') = Ax + Ax'$ [since $A(x+x') - Ax - Ax' \in \mathcal{H}^{\perp} = (0)$]. Similarly $A(\alpha x) = \alpha Ax$. Also,

$$\|Ax\| = \sup_{y\neq 0} \frac{\langle Ax, y\rangle}{\|y\|} = \sup_{y\neq 0} \frac{|\phi(x,y)|}{\|y\|} \leq K\|x\|$$

so A is continuous. ■

Definition The adjoint. Let $A \in \mathcal{B}(\mathcal{H})$. Then $\psi(x,y) = \langle x, Ay\rangle$ is a bounded bilinear form on $\mathcal{H}$ so, by Theorem 2.5 there is an operator $A^* \in \mathcal{B}(\mathcal{H})$ such that

$$\langle A^* x, y\rangle = \psi(x,y) = \langle x, Ay\rangle\,.$$

A^* is called the *adjoint* of A.

Exercise.

(i) $(A^*)^* = A$,

(ii) $(\lambda A)^* = \bar{\lambda} A^*$,

(iii) $(A+B)^* = A^* + B^*$,

(iv) $(AB)^* = B^* A^*$,

(v) $\|A\| = \|A^*\|$.

Note. Bilinear forms could have been defined as maps ϕ from $\mathcal{H} \times \mathcal{K}$ to $\mathbb{C}$ where $\mathcal{H}$ and $\mathcal{K}$ are different Hilbert spaces. All the above can be done with essentially no change; (the adjoint of $A \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ is then an operator in $\mathcal{B}(\mathcal{K}, \mathcal{H})$).

Definition.

If	$A = A^*$	then A is said to be *selfadjoint*.
If	$AA^* = A^*A$	then A is said to be *normal*.
If	$UU^* = U^*U = I$	then U is said to be *unitary*.

Projections.

Let N be a closed subspace of $\mathcal{H}$. Then from Theorem 1.11,

$$\mathcal{H} = N \oplus N^{\perp}$$

that is, any $h \in \mathcal{H}$ has a unique decomposition as $h = x + y$ with $x \in N$ and $y \in N^{\perp}$.

The *orthogonal projection* P onto N is defined by $Ph = x$ (where $h = x + y$ is the decomposition above). Note that then $y = (I - P)h$ and $I - P$ is the projection onto $N^{\perp}$.

In this course we shall not consider projections that are not orthogonal and usually call these operators "projections".

Lemma 2.6 *Let N be a closed subspace of $\mathcal{H}$ and let P be the orthogonal projection onto N. Then*
(i) P *is linear,*
(ii) $\|P\| = 1$ *(unless $N = 0$),*
(iii) $P^2 = P$,
(iv) $P^* = P$.

Also, if $E \in \mathcal{B}(\mathcal{H})$ satisfies $E = E^2 = E^*$ then E is the (orthogonal) projection onto some (closed) subspace.

Proof. (i) Let $h, h' \in \mathcal{H}$ and suppose $h = x + y$ and $h' = x' + y'$ are the unique decompositions of h and h' with $x, x' \in N$ and $y, y' \in N^{\perp}$. Then $\alpha h + \beta h' = (\alpha x + \beta x') + (\alpha y + \beta y')$ is the decomposition of $\alpha h + \beta h'$ and

$$P(\alpha h + \beta h') = \alpha x + \beta x' = \alpha Ph + \beta Ph' .$$

(ii) If $h = x + y$ with $x \in N$ and $y \in N^{\perp}$,

$$\|Ph\|^2 = \|x\|^2 \leq \|x\|^2 + \|y\|^2 = \|h\|^2$$

and so $\|P\| \leq 1$. But if $0 \neq h \in N$ then $Ph = h$ and so $\|P\| = 1$.

(iii) If $h \in N$ [then $h = h + 0$ is the decomposition of h and] $Ph = h$. But for any $h \in \mathcal{H}$, $Ph \in N$ so $P(Ph) = Ph$, that is, $P^2 = P$.

(iv) If $h = x + y$ and $h' = x' + y'$ with $x, x' \in N$ and $y, y' \in N^{\perp}$.

$$\langle Ph, h' \rangle = \langle x, x' + y' \rangle = \langle x, x' \rangle$$

since $x \in N$ and $y' \in N^{\perp}$. Similarly $\langle h, Ph' \rangle = \langle x, x' \rangle$ and so $P = P^*$.

Finally, if $E \in \mathcal{B}(\mathcal{H})$ satisfies $E = E^2 = E^*$ let $N = \{x : Ex = x\}$. Then $N = \ker(I - E)$, so N is closed. For any $h \in \mathcal{H}$, write

$$h = Eh + (I - E)h .$$

Then $Eh \in N$ since $E(Eh) = E^2 h = Eh$ and $(I - E)h \perp N$ since if $x \in N$, $Ex = x$ and

$$\langle (I - E)h, x \rangle = \langle (I - E)h, Ex \rangle = \langle E^*(I - E)h, x \rangle = \langle (E^2 - E)h, x \rangle = 0.$$

This shows that E is the projection onto N. ∎

Lemma 2.7 *If P is the orthogonal projection onto a subspace N then for all $h \in \mathcal{H}$,*

$$d(h, N) = \|(I - P)h\|.$$

Proof. For any $h \in \mathcal{H}$ we have $Ph \in N$ and $\langle (I - P)h, n \rangle = 0$ for all $n \in N$. Therefore from Lemma 1.7

$$d(h, N) = \|h - Ph\| = \|(I - P)h\|.$$

■

Lemma 2.8 *Let $A \in \mathcal{B}(\mathcal{H})$ and P be the orthogonal projection onto a subspace N.*

(i) N *is invariant under* $A \Longleftrightarrow AP = PAP$.
(ii) $N^{\perp}$ *is invariant under* $A \Longleftrightarrow PA = PAP$.

If $A = A^*$ then N is invariant under $A \Longleftrightarrow N^{\perp}$ is invariant under $A \Longleftrightarrow PA = AP$.

Proof. (i) $\Longrightarrow$ Suppose $An \in N$ for all $n \in N$. Then since $Ph \in N$ for all $h \in \mathcal{H}$, we have $APh \in N$. Therefore then $PAPh = APh$ [since $Pn = n$ for all $n \in N$].

$\Longleftarrow$ If $n \in N$ then $Pn = n$ and so $An = APn = PAPn \in N$ [since N is the range of P].

(ii) The projection onto $N^{\perp}$ is $I - P$. Trivial algebra shows that

$$A(I - P) = (I - P)A(I - P) \Longleftrightarrow PA = PAP$$

and so (ii) follows from (i).

Finally $AP = PAP \Longleftrightarrow (AP)^* = (PAP)^* \Longleftrightarrow PA = PAP$ since $A = A^*$. If these equalities hold then $PA = AP$. ■

Exercises 2

1. Let $X \in \mathcal{B}(\mathcal{H})$. Show that :

(i) X is selfadjoint $\iff$ $\langle Xx, x\rangle$ is real for all $x \in \mathcal{H}$,

(ii) X is normal $\iff \| Xx \| = \| X^*x \|$ for all $x \in \mathcal{H}$,

(iii) X is unitary $\iff \| Xx \| = \| X^*x \| = \| x \|$ for all $x \in \mathcal{H}$.

2. Let S be the one-dimensional subspace of ℓ^2 spanned by the element $(1, -1, 0, 0, \ldots)$. Show explicitly that any element $x = (\xi_k) \in \ell^2$ can be written as $x = x_1 + x_2$ where $x_1 \in S$ and $x_2 \perp S$.

3. Let A be a selfadjoint operator such that for all $x \in H$, $\| Ax \| \geq c \| x \|$, where c is a positive constant. Show that A has a continuous inverse.

[Hints : Show (i) A is injective, (ii) the range of A is closed (iii) $(\mathrm{ran}(A))^\perp = (0)$.] Note that the selfadjointness condition **is** needed – consider the operator S on ℓ^2 defined by $S(\xi_1, \xi_2, \xi_3, \ldots) = (0, \xi_1, \xi_2, \xi_3, \ldots)$.

4. The operators D and W on ℓ^2 are defined by

$$D(\xi_1, \xi_2, \xi_3, \ldots) = (\alpha_1\xi_1, \alpha_2\xi_2, \alpha_3\xi_3, \ldots)\,,$$

$$W(\xi_1, \xi_2, \xi_3, \ldots) = (0, \alpha_1\xi_1, \alpha_2\xi_2, \alpha_3\xi_3, \ldots)\,,$$

where (α_n) is a bounded sequence of complex numbers. Show that W and D are bounded operators and find their adjoints.

5. Given that $X \in \mathcal{B}(\mathcal{H})$ is invertible (that is, there exists $X^{-1} \in \mathcal{B}(\mathcal{H})$ such that $XX^{-1} = X^{-1}X = I$) prove that X^* is invertible and $(X^*)^{-1} = (X^{-1})^*$.

6. Find the adjoint of the operator V defined on $L^2[0,1]$ by $(Vf)(x) = \int_0^x f(t)\,dt$.

7. Let $T : L^2[0,1] \to L^2[0,1]$ be defined by

$$(Tf)(x) = \sqrt{2x}\, f(x^2)\,.$$

Find the adjoint of T and deduce that T is unitary.

8. Let E, F be the orthogonal projections onto subspaces M and N respectively. Prove that,

(i) $EF = F \iff N \subseteq M \iff E - F$ is an orthogonal projection,

(ii) $EF = 0 \iff N \subseteq M^\perp \iff E + F$ is an orthogonal projection,

(iii) $EF = FE \iff E + F - FE$ is an orthogonal projection.

9. The operator $A \in \mathcal{B}(\mathcal{H})$ satisfies $Ax = x$ for some $x \in \mathcal{H}$. Prove that $A^*x = x + y$ where $y \perp x$. If, further, $\|A\| \leq 1$, show that $A^*x = x$.

Suppose that $E^2 = E$ and $\|E\| = 1$. Use the above to show that $\mathrm{ran}(E) = \mathrm{ran}(E^*)$ and $\ker(E) = \ker(E^*)$ and deduce that $E = E^*$ (so that E is the orthogonal projection onto some subspace of $\mathcal{H}$).

10. Let L_o and L_e be subspaces of $L^2[-1,1]$ defined by

$$\begin{aligned} L_o &= \{f : f(t) = -f(-t) \quad \text{(almost everywhere)}\} \\ L_e &= \{f : f(t) = f(-t) \qquad \text{(almost everywhere)}\}. \end{aligned}$$

Show that $L_o \oplus L_e = L^2[-1,1]$ and find the projections of $L^2[0,1]$ onto L_o and L_e. Find expressions for the distances of any element f to L_o and L_e. Calculate the values in the specific case where $f(t) = t^2 + t$.

11. Let M and N be vector subspaces of $\mathcal{H}$ such that $M \perp N$ and $M \dot{+} N = \mathcal{H}$. Prove that M and N are closed.

12. Show that the set of sequences $x = (\xi_n)$ such that $\sum_n n^2 |\xi_n|^2$ converges, forms a dense subset of l^2.

Define the operator D on l^2 by

$$D(\xi_1, \xi_2, \xi_3, \ldots) = (\xi_1, \tfrac{1}{2}\xi_2, \tfrac{1}{3}\xi_3, \ldots)\,,$$

and let M and N be linear subspaces of $l^2 \oplus l^2$ defined by

$$M = \{(x, 0) : x \in l^2\} \quad N = \{(x, Dx) : x \in l^2\}\,.$$

Observe that M is closed and use the continuity of D to show that N is also closed. Show that $M \cap N = (0)$ and that the algebraic direct sum of M and N is dense in $l^2 \oplus l^2$ but is not equal to $l^2 \oplus l^2$ (and so it is not closed).

3 Orthonormal Sets.

Definition. A set $\mathcal{S}$ of vectors of $\mathcal{H}$ is said to be *orthonormal* if
1. $\|x\| = 1$ for all $x \in \mathcal{S}$,
2. $\langle x, y\rangle = 0$ if $x \neq y$ and $x, y \in \mathcal{S}$.

Lemma 3.1 (Bessel's inequality) *If $\{x_i : 1 \le i \le n\}$ is a finite orthonormal set then, for any $h \in \mathcal{H}$, writing $\alpha_i = \langle h, x_i\rangle$,*

$$\sum_{i=1}^{n} |\alpha_i|^2 \le \|h\|^2.$$

(Note that the case $n = 1$ is the Cauchy-Schwartz inequality)

Proof. Let $h \in \mathcal{H}$. Then

$$\begin{aligned}
\|h - \sum_{i=1}^{n} \alpha_i x_i\|^2 &= \langle h - \sum_{i=1}^{n} \alpha_i x_i, h - \sum_{i=1}^{n} \alpha_i x_i\rangle \\
&= \|h\|^2 - \langle h, \sum_{i=1}^{n} \alpha_i x_i\rangle - \langle \sum_{i=1}^{n} \alpha_i x_i, h\rangle + \sum_{i,j=1}^{n} \alpha_i \overline{\alpha_j} \langle x_i, x_j\rangle \\
&= \|h\|^2 - \langle h, \sum_{i=1}^{n} \alpha_i x_i\rangle - \langle \sum_{i=1}^{n} \alpha_i x_i, h\rangle + \sum_{i=1}^{n} |\alpha_i|^2 \\
&= \|h\|^2 - \sum_{i=1}^{n} |\alpha_i|^2 \ge 0. \qquad \blacksquare
\end{aligned}$$

Lemma 3.2 *If $\{x_i : i = 1, 2, 3 \cdots\}$ is an orthonormal sequence then, for any $h \in \mathcal{H}$, writing $\alpha_i = \langle h, x_i\rangle$,*

$$\sum_{i=1}^{\infty} \alpha_i x_i$$

converges to a vector h' such that $\langle h - h', x_i\rangle = 0$ for all i.

Proof. Put $h_r = \sum_{i=1}^{r} \alpha_i x_i$. Then

$$\begin{aligned}
\|h_{r+p} - h_r\|^2 &= \left\| \sum_{i=r+1}^{r+p} \alpha_i x_i \right\|^2 \\
&= \sum_{i=r+1}^{r+p} |\alpha_i|^2 .
\end{aligned}$$

Now $\sum_{i=1}^{n} |\alpha_i|^2 \le \|h\|^2$ for all n and so $\sum_{i=1}^{\infty} |\alpha_i|^2$ is convergent and so is a Cauchy series. Hence, given any $\epsilon > 0$ there exists n_0 such that for $r > n_0, p > 0$ we have $\sum_{i=r+1}^{r+p} |\alpha_i|^2 < \epsilon^2$; that is, $\|h_{r+p} - h_r\| < \epsilon$. Therefore (h_r) is a Cauchy sequence and, since $\mathcal{H}$ is complete, it is convergent. Call its limit h'.

For any fixed i, $\langle h - h_r, x_i\rangle = 0$ for all $r > i$. Now let $r \to \infty$. Then it follows from the continuity of the inner product that $\langle h - h', x_i\rangle = 0$ for all i. $\blacksquare$

Theorem 3.3 *Let $\{x_i : i = 1, 2, 3 \cdots\}$ be an orthonormal sequence. The the following statements are equivalent.*

(i) $\{x_i : i = 1, 2, 3 \cdots\}$ *is maximal (that is, it is not a proper subset of any orthonormal set).*

(ii) *If $\alpha_i = \langle h, x_i \rangle = 0$, for all i then $h = 0$.*

(iii) (Fourier expansion) *For all $h \in \mathcal{H}$ we have $h = \sum_{i=1}^{\infty} \alpha_i x_i$.*

(iv) (Parseval's relation) *For all $h, g \in \mathcal{H}$ we have $\langle h, g \rangle = \sum_{i=1}^{\infty} \alpha_i \overline{\beta_i}$.*

(v) (Bessel's equality) *For all $h \in \mathcal{H}$ we have $\|h\|^2 = \sum_{i=1}^{\infty} |\alpha_i|^2$.*

(In the above, $\alpha_i = \langle h, x_i \rangle$ and $\beta_i = \langle g, x_i \rangle$.)

Proof. (i) $\Longrightarrow$ (ii). If (ii) is false then adding $\frac{h}{\|h\|}$ to the set $\{x_i : i = 1, 2, 3 \cdots\}$ gives a larger orthonormal set, contradicting (i).

(ii) $\Longrightarrow$ (iii). Let $h' = \sum_{i=1}^{\infty} \alpha_i x_i$ (this exists, by Lemma 3.2). Then $\langle h - h', x_i \rangle = 0$ for all i and so $h = h'$ by (ii)¿

(iii) $\Longrightarrow$ (iv). Let $h_r = \sum_{i=1}^{r} \alpha_i x_i$ and $g_s = \sum_{i=1}^{s} \beta_i x_i$. Then

$$\langle h_r, g_s \rangle = \sum_{i=1}^{\min[r,s]} \alpha_i \overline{\beta_i} \,.$$

Let $r \to \infty$ and $s \to \infty$. Using the continuity of the inner product, it follows that

$$\langle h, g \rangle = \sum_{i=1}^{\infty} \alpha_i \overline{\beta_i} \,.$$

(iv) $\Longrightarrow$ (v). Put $g = h$ in (iv).

(v) $\Longrightarrow$ (i). If $\{x_i : i = 1, 2, 3 \cdots\}$ is not maximal and can be enlarged by adding z then $\langle z, x_i \rangle = 0$ for all i but also

$$1 = \|z\|^2 = \sum_{i=1}^{\infty} |\langle z, x_i \rangle\|^2 = 0$$

which give a contradiction. ■

Definition. A maximal orthonormal sequence is called an *orthonormal basis.*

Clearly the concept of an orthonormal basis refers to a *set* of vectors so that any permutation of such a set is still an orthonormal basis. It follows that the series giving the fourier expansion of a vector can be re-arranged arbitrarily without altering its convergence or its sum. Such a series is said to be *unconditionally convergent.*

Theorem 3.4 (Gram-Schmidt process) *Let $\{y_i : i = 1, 2, 3\cdots\}$ be a sequence of vectors of $\mathcal{H}$. Then there exists an orthonormal sequence $\{x_i : i = 1, 2, 3\cdots\}$ such that, for each integer k*

$$\operatorname{span}\{x_1, x_2, x_3 \cdots x_k\} \supseteq \operatorname{span}\{y_1, y_2, y_3 \cdots y_k\}.$$

If $\{y_i : i = 1, 2, 3\cdots\}$ is a linearly independent set, then the above inclusion is an equality for each k.

Proof. First consider the case when $\{y_i : i = 1, 2, 3\cdots\}$ is a linearly independent set. Define

$$\begin{aligned}
u_1 &= y_1 & x_1 &= \frac{u_1}{\|u_1\|}, \\
u_2 &= y_2 - \langle y_2, x_1\rangle x_1 & x_2 &= \frac{u_2}{\|u_2\|}, \\
&\vdots & &\vdots \\
u_n &= y_n - \sum_{i=1}^{n-1}\langle y_n, x_i\rangle x_i & x_n &= \frac{u_n}{\|u_n\|}.
\end{aligned}$$

Easy inductive arguments show that for each k, $x_{k-1} \in \operatorname{span}\{y_1, y_2 \cdots, y_{k-1}\}$ and, since $\{y_i : i = 1, 2, 3\cdots\}$ is linearly independent, that $u_k \neq 0$. A further easy induction shows that for $r < n$ we have $\langle u_n, x_r\rangle = 0$ and so it follows easily that $\{x_i : i = 1, 2, 3\cdots\}$ is an orthonormal sequence. In the general case, the same construction applies except that whenever y_r is a linear combination of $y_1, y_2, \cdots, y_{r-1}$ this element is ignored.

It is clear in general that x_r is a linear combination of on $y_1, y_2, \cdots, y_{r-1}$ and so the inclusion

$$\operatorname{span}\{x_1, x_2, x_3 \cdots x_k\} \supseteq \operatorname{span}\{y_1, y_2, y_3 \cdots y_k\}.$$

is obvious. When $\{y_i : i = 1, 2, 3\cdots\}$ is a linearly independent set we have equality, since both sides have dimension k. ■

For the rest of this course we shall often need to assume that the Hilbert space we consider has a countable orthonormal basis. This is true for all spaces considered in the applications. The restriction is a rather technical matter and could be avoided but this would entail a discussion of uncountable sums. A few statements would also need modification.

The appropriate way to state this restriction is to say that the Hilbert space we consider is *separable*. For our purposes one could say that a Hilbert space is separable if it has a countable orthonormal basis, and take this as the definition of separability. However, this is a more general notion: recall that a metric space is defined to be *separable* if it has a countable, dense subset. The proposition which follows connects these ideas.

Proposition

1. In a separable Hilbert space, every orthonormal set is countable.

2. A Hilbert space is separable if and only if it has a (countable) orthonormal basis.

We shall not prove this in detail, but here is a sketch. For each element x_α of an orthonormal set $\{x_\alpha\}_{\alpha\in A}$, let B_α be the open ball centre x_α, radius $\frac{\sqrt{2}}{2}$. Since these balls are disjoint and since every open set must contain at least one element of a dense subset, it is clear that if the space is separable the orthonormal set must be countable. For 2, applying the Gram-Schmidt process to a countable dense subset results in an orthonormal sequence that is easily proved to be a basis. Conversely, if $\{x_n : n = 1, 2, 3, \cdots\}$ is a countable orthonormal basis, tedious but routine arguments show that the set

$$S = \{\sum_{n=1}^{N} r_n x_n : N \text{ finite }, r_n \text{ rational } \}$$

is countable. It is clearly dense because the closure of S is a subspace containing an orthonormal basis.

Exercises 3

1. Let $\{N_i\}$ be a sequence of mutually orthogonal subspaces of a Hilbert space H and let $\{E_i\}$ be the sequence of projections onto $\{N_i\}$. Show that for each $x \in H$,

(i) for a finite subset $\{N_i\}_{i=1}^n$ of $\{N_i\}$, $\sum_{i=1}^n \| E_i x \|^2 \leq \| x \|^2$.

(ii) $\sum_{i=1}^\infty E_i x$ converges to some $h \in H$ which satisfies $(x-h) \perp N_i$ for each i.

Show further that the following are equivalent :

(a) $\{N_i\}$ is not a proper subset of any orthogonal set of subspaces of H,

(b) $h \perp N_i$ for all $i \Rightarrow h = 0$,

(c) for each $x \in H$, $x = \sum_{i=1}^\infty E_i x$,

(d) for each $x, y \in H$, $\langle x, y\rangle = \sum_{i=1}^\infty \langle E_i x, E_i y\rangle$,

(e) for each $x \in H$, $\| x \|^2 = \sum_{i=1}^\infty \| E_i x \|^2$.

[Hints : You will need Q. 8 (ii) of Sheet 2 or its equivalent. Note that under the given conditions, since $E_i = E_i^* = E_i^2$, $\langle E_i x, E_j y\rangle = \langle E_j E_i x, y\rangle = 0$ if $i \neq j$.]

2. Find the first three functions obtained by applying the Gram- Schmidt process to the elements $\{t^n : n = 0, 1, \ldots\}$ of $L^2[-1,1]$. [Note: apart from constant factors, this process yields the Legendre polynomials.] Use your results and the theory developed in lectures to find a, b and c which minimises the quantity

$$\int_{-1}^{1} |t^4 - a - bt - ct^2|^2 \, dt \, .$$

3. Let N be a subspace of $L^2[0,1]$ with the property that for some fixed constant C and each $f \in N$,

$$|f(x)| \leq C\|f\| \qquad \text{almost everywhere .}$$

Prove that N is finite dimensional.

[Hint: for any orthonormal subset $f_1, f_2, \ldots f_n$, of N, evaluate, for any fixed y, the norm of g where

$$g(x) = \sum_{i=1}^{n} \overline{f_i(y)} f_i(x) \, .$$

Deduce that $\sum_{i=1}^n |f_i(y)|^2 \leq C^2$ and integrate this relation with respect to y.]

4 Compact Operators.

Definition. An operator $K \in \mathcal{B}(\mathcal{H})$ is said to be *compact* if for every bounded set $\mathcal{S}$ of vectors of $\mathcal{H}$ the set $\overline{\{Ks : s \in \mathcal{S}\}}$ is compact.

Equivalently :

Definition. An operator $K \in \mathcal{B}(\mathcal{H})$ is said to be *compact* if for every bounded sequence (x_n) of vectors of $\mathcal{H}$ the sequence (Kx_n) has a convergent subsequence.

We shall denote the set of all compact operators on $\mathcal{H}$ by $\mathcal{K}(\mathcal{H})$.

Definition. The *rank* of an operator is the dimension of its range.

Note that every operator of finite rank is compact. This is an immediate consequence of the Bolzano-Weierstrass theorem which states that every bounded sequence in $\mathbb{C}^n$ has a convergent subsequence. Note also that the identity operator on a Hilbert space $\mathcal{H}$ is compact if and only if $\mathcal{H}$ is finite-dimensional.

Theorem 4.1 *$\mathcal{K}(\mathcal{H})$ is an ideal of $\mathcal{B}(\mathcal{H})$.*

Proof. We need to show that, if $A, B \in \mathcal{K}(\mathcal{H})$ and $T \in \mathcal{B}(\mathcal{H})$ then $\alpha A, A + B, TA$ and AT are all in $\mathcal{K}(\mathcal{H})$. That is, for any a bounded sequence (x_n), we must show that $(\alpha Ax_n), ([A + B]x_n), (TAx_n)$ and (ATx_n) all have convergent subsequences.

Since A is compact, (Ax_n) has a convergent subsequence (Ax_{n_i}). Then clearly (αAx_{n_i}) is a convergent subsequence of (αAx_n) showing that αA is compact. Also, (x_{n_i}) is a bounded sequence and so, since B is compact, (Bx_{n_i}) has a convergent subsequence $(Bx_{n_{i_j}})$. Then $([A+B]x_{n_{i_j}})$ is a convergent subsequence of $([A+B]x_n)$, showing that $A + B$ is compact.

Again, since $T \in \mathcal{B}(\mathcal{H})$, T is continuous and so (TAx_{n_i}) is a convergent subsequence of (TAx_n) showing that TA is compact. The proof for AT is slightly different. Here, since (x_n) is bounded and $\|Tx_n\| \leq \|T\|.\|x_n\|$ we have that (Tx_n) is bounded and so, since A is compact, (ATx_n) has a convergent subsequence, showing that TA is compact. ■

A consequence of the above theorem is that, it $\mathcal{H}$ is infinite-dimensional then and $T \in \mathcal{B}(\mathcal{H})$ has an inverse $T^{-1} \in \mathcal{B}(\mathcal{H})$ then T is not compact.

Theorem 4.2 *$\mathcal{K}(\mathcal{H})$ is closed.*

Proof. Let (K_n) be a sequence of compact operators converging to K. To show that K is compact, we need to show that if (x_i) is a bounded sequence the (Kx_i) has a convergent subsequence.

Let (x_i^1) be a subsequence of (x_i) such that $(K_1x_i^1)$ is convergent,
let (x_i^2) be a subsequence of (x_i^1) such that $(K_2x_i^2)$ is convergent,
let (x_i^3) be a subsequence of (x_i^2) such that $(K_3x_i^3)$ is convergent,
and continue in this way.

[The notation above is slightly unusual and is adopted to avoid having to use subscripts on subscripts on $\cdots$.]

Let $z_i = x_i^i$. Then (z_i) is a subsequence of (x_i). Also, for each n, apart from the first n terms, (z_i) is a subsequence of (x_i^n) and so (K_nz_i) is convergent.

We now show that (Kz_i) is convergent by showing that it is a Cauchy sequence. For all i, j, n we have

$$\begin{aligned}\|Kz_i - Kz_j\| &= \|(K-K_n)z_i + K_nz_i - K_nz_j - (K-K_n)z_j\| \\ &\le \|K-K_n\|(\|z_i\| + \|z_j\|) + \|K_n(z_i - z_j)\|\,.\end{aligned}$$

Let $\epsilon > 0$ be given. Since $(K_n) \to K$ we can find n_0 such that $\|K - K_n\| < \frac{\epsilon}{4c}$ for $n > n_0$ where c satisfies $\|x_i\| \le c$ for the bounded sequence (x_i). Choose one fixed such n. Now, since (K_nz_i) converges, it is a Cauchy sequence and so there is an i_0 such that for $i > i_0, j > i_0$ we have $\|K_nz_i - K_nz_j\| < \frac{\epsilon}{2}$. Combining these with the displayed inequality shows that for $i > i_0, j > i_0$, $\|Kz_i - Kz_j\| < \epsilon$ so (Kz_i) is convergent as required. ■

Example. The operator K on $L^2[a,b]$ defined by

$$(Kf)(x) = \int_a^b k(x,t)f(t)\,dt\,,$$

where $\int_a^b \int_a^b |k(x,t)|^2\,dx\,dt = M^2 < \infty$, is compact.

We have already seen that operators of the above type are continuous with $\|K\| \le M$ (Recall that $k(x,t)$ is called the *kernel* of the integral operator K). We shall show that K is the norm limit of a sequence of finite rank operators. Note that if $k(x,t)$ is of the form $u(x)v(t)$ then

$$(Rf)(x) = \int_a^b u(x)v(t)f(t)\,dt \; = \langle f, \overline{v}\rangle u = (\overline{v} \otimes u)f$$

is a rank one operator.

Let S be the square $[a,b] \times [a,b]$. We shall apply Hilbert space theory to $L^2(S)$ which is a Hilbert space of functions of 2 variables with the inner product

$$\langle \phi, \psi \rangle = \int_a^b \int_a^b \phi(x,t)\overline{\psi(x,t)}\,dx\,dt\,.$$

Let (u_i) be an orthonormal basis of $L^2[a,b]$. Then $(u_i(x)u_j(t))_{i,j=1}^{\infty}$ is an orthonormal basis of $L^2(S)$. Indeed,

$$\begin{aligned}\langle u_i(x)u_j(t), u_k(x)u_l(t)\rangle &= \int_a^b \int_a^b u_i(x)u_j(t)\overline{u_k(x)u_l(t)}\,dx\,dt \\ &= \int_a^b u_i(x)\overline{u_k(x)}\,dx \int_a^b u_j(t)\overline{u_l(t)}\,dt = 0\end{aligned}$$

unless $i = k$ and $j = l$, in which case the integral is 1. Thus $(u_i(x)u_j(t))_{i,j=1}^{\infty}$ is an orthonormal sequence. To show that it is a basis, suppose $\phi(x,t) \perp u_i(x)u_j(t)$ for all i, j. Then

$$0 = \int_a^b \int_a^b \phi(x,t)\overline{u_i(x)u_j(t)}\,dx\,dt = \int_a^b \left(\int_a^b \phi(x,t)\overline{u_i(x)}\,dx\right)\overline{u_j(t)}\,dt\,.$$

This shows that, for each i, the function $\int_a^b \phi(x,t)\overline{u_i(x)}\,dx$ of t is orthogonal to $u_j(t)$ for each j. Therefore, since (u_j) is a basis of $L^2[a,b]$, it is (equivalent to) the zero function. Then, for fixed t the function $\phi(x,t)$ is orthogonal to $u_i(x)$ for each i and so it is zero.

Returning to the operator K, note that $k \in L^2(S)$. Therefore, by Theorem 3.3 (iii) it has a fourier expansion using the basis $(u_i u_j)$ of the type

$$k(x,t) = \sum_{i,j=1}^{\infty} \alpha_{ij} u_i(x) u_j(t)\,.$$

Thus, writing $k_n(x,t) = \sum_{i,j=1}^{n} \alpha_{ij} u_i(x) u_j(t)$ and

$$(K_n f)(x) = \int_a^b k_n(x,t) f(t)\, dt\,,$$

we have that K_n is a finite rank operator (of rank at most n^2). Note that $K - K_n$ is an integral operator (of the same type as K) with kernel $k(x,t) - k_n(x,t)$. Thus

$$\|K - K_n\|^2 \le \int_a^b \int_a^b |k(x,t) - k_n(x,t)|^2\, dx\, dt = \|k - k_n\|^2_{L^2(S)}$$

and the right hand side $\to 0$. Therefore Theorem 4.2 shows that K is compact.

Lemma 4.3 *Let K be a compact operator on $\mathcal{H}$ and suppose (T_n) is a bounded sequence in $\mathcal{B}(\mathcal{H})$ such that, for each $x \in \mathcal{H}$ the sequence $(T_n x)$ converges to Tx, where $T \in \mathcal{B}(\mathcal{H})$. Then $(T_n K)$ converges to TK in norm.*

Briefly, the above can be rephrased as :

If $K \in \mathcal{K}(\mathcal{H})$ and $\|T_n x - Tx\| \to 0$ for all $x \in \mathcal{H}$ then $\|T_n K - TK\| \to 0$.

In words : multiplying by a compact operator on the right converts a pointwise convergent sequence of operators into a norm convergent one.

Proof. Since (T_n) is a bounded sequence, $\|T_n\| \le C$ for some constant C. Then for all $x \in \mathcal{H}$, $\|Tx\| = \lim_n \|T_n x\| \le C\|x\|$ and so $\|T\| \le C$.

Let K be compact and suppose that $\|TK - T_n K\| \not\to 0$. Then there exists some $\delta > 0$ and a subsequence $(T_{n_i} K)$ such that $\|TK - T_{n_i} K\| > \delta$. Choose unit vectors (x_{n_i}) of $\mathcal{H}$ such that $\|(TK - T_{n_i} K)x_{n_i}\| > \delta$. [That this can be done follows directly from the definition of the norm of an operator.] Using the fact that K is compact, we can find a subsequence (x_{n_j}) of (x_{n_i}) such that (Kx_{n_j}) is convergent. Let the limit of this sequence be y. Then for all j

$$\delta < \|(TK - T_{n_j} K)x_{n_j}\| \le \|(T - T_{n_j})(Kx_{n_j} - y)\| + \|(T - T_{n_j})y\|\,.$$

Now, using the convergence of (Kx_{n_j}) to y, there exists n so that, for $n_j > n$, $\|Kx_{n_j} - y\| < \frac{\delta}{8C}$. Also, using the convergence of (T_{n_j}) to T, there exists m so that, for $n_j > m$, $\|(T - T_{n_j})y\| < \frac{\delta}{4}$. Then, for $j > \max[n,m]$ the right hand side of the displayed inequality is less than $\frac{\delta}{2}$, and this contradiction shows that the supposition that $\|TK - T_n K\| \not\to 0$ is false. ■

The theorem below is true for all Hilbert spaces, but we shall only prove it for the case when the space is separable.

Theorem 4.4 *Every compact operator on $\mathcal{H}$ is a norm limit of a sequence of finite rank operators.*

Proof. Let x_i be an orthonormal basis of $\mathcal{H}$. Define P_n by

$$P_n h = \sum_{i=1}^{n} \langle h, x_i \rangle x_i \,.$$

[Note that P_n is the projection onto $\operatorname{span} x_1, x_2, \cdots, x_n$. Also, P_n could be written as $P_n h = \sum_{i=1}^{n} x_i \otimes x_i$.] From Theorem 3.3(iii), for all $x \in \mathcal{H}$, $P_n x$ converges to x (that is, P_n converges pointwise to the identity operator I). Now, if K is any compact operator, $P_n K$ is of finite rank and, from Theorem 4.3 $(P_n K)$ converges to K in norm. ∎

Exercises 4

1. Let T be the operator on $l^2 \oplus l^2$ defined by $T(x, y) = (0, x)$. Show that $T^2 = 0$ and that T is not compact.

2. Let (x_n) be an orthonormal sequence in a Hilbert space H and let (α_n) be a bounded sequence of complex numbers. Prove that the operator A defined by

$$Ax = \sum_{n=1}^{\infty} \alpha_n \langle x, x_n \rangle x_n$$

is bounded with

$$\| A \| \leq \sup_n |\alpha_n| \,.$$

Hence prove that if $\lim_{n\to\infty} (\alpha_n) = 0$ then A is compact.

Show that, when $m \neq n$,

$$\| Ax_m - Ax_n \|^2 = |\alpha_m|^2 + |\alpha_n|^2 \,.$$

Hence prove that, conversely if $\lim_{n\to\infty} (\alpha_n) \neq 0$ then A is not compact.

3. Given that K^*K is compact, prove that K is compact.

[Hint: if (K^*Kx_n) is convergent, prove that (Kx_n) is a Cauchy sequence.]

4. Let K be a compact operator. Using the hints below, prove that for any orthonormal sequence $\{x_n\}$, $(Kx_n) \to 0$ as $n \to \infty$

Hints: Observe that, for any vector z, $\langle x_n, z \rangle \to 0$. [A result of the course states that $\sum |\langle x_n, z \rangle|^2$ is convergent.] Apply this, with $z = K^*y$ for any y, and show that no subsequence of (Kx_n) can converge to a non-zero vector.

5. Let A_n be a bounded sequence in $\mathcal{B}(\mathcal{H})$ such that, for all $x, y \in \mathcal{H}, \lim_{n\to\infty} \langle A_n x, y \rangle = 0$. Prove that, for any compact operator K,

$$\lim_{n\to\infty} \|KA_nK\| = 0 \,.$$

[Use the ideas in the proof of Lemma 4.3.]

5 The Spectrum.

Definition. The *spectrum* of an operator T is the set of all complex numbers λ such that $\lambda I - T$ has no inverse in $\mathcal{B}(\mathcal{H})$.

The *spectrum* of T is denoted by $\sigma(T)$.

The complement (in $\mathbb{C}$) of $\sigma(T)$, that is, the set of all complex numbers λ such that $\lambda I - T$ has an inverse in $\mathcal{B}(\mathcal{H})$, is called the *resolvent set* of T and is denoted by $\rho(T)$.

For any element T of $\mathcal{B}(\mathcal{H})$, it is a fact that $\sigma(T)$ is a non-empty compact subset of $\mathbb{C}$. We shall not need this general fact in this course. For the two classes of operators that we shall be concerned with (compact operators and selfadjoint operators) the required facts about the spectrum will be established by simple methods.

Note that every eigenvalue of an operator T is in the spectrum of T.

Also, if the K is a compact operator on an infinite-dimensional Hilbert space then $0 \in \sigma(K)$ (this merely repeats the fact that K does not have an inverse).

Lemma 5.1 *Let T be an operator such that for all $x \in \mathcal{H}$, $\| Tx \| \geq c \| x \|$, where c is a positive constant. Then the range of T is closed.*

Proof. Let (y_n) be a convergent sequence of elements of $\operatorname{ran}(T)$ converging to y. Then $y_n = Tx_n$ for some sequence (x_n) and we need to show that $y = Tx$ for some x.

Since (y_n) is convergent it is a Cauchy sequence. Now,

$$\|x_n - x_m\| \leq \tfrac{1}{c}\|T(x_n - x_m)\| = \tfrac{1}{c}\|y_n - y_m)\|$$

so it follows easily that (x_n) is a Cauchy sequence and so convergent to some element x. Then, since T is continuous, $y = \lim y_n = \lim Tx_n = \lim Tx$, as required. ■

Corollary 5.2 *If T is an in the lemma, the range of T^n is closed for each positive integer n.*

Proof. $\|T^n x\| \geq c^n \|x\|$ for all $x \in \mathcal{H}$. ■

We now derive some simple properties of the spectrum of a selfadjoint operator. For the rest of this section, A will denote a selfadjoint operator. Recall that $\langle Ax, x\rangle$ is real for all x since $\overline{\langle Ax, x\rangle} = \langle x, Ax\rangle = \langle A^*x, x\rangle = \langle Ax, x\rangle$.

Lemma 5.3

$$\|A\| = \sup_{\|x\| \leq 1} |\langle Ax, x\rangle| \, .$$

Proof. Let $k = \sup_{\|z\| \leq 1} |\langle Az, z\rangle|$. Then $|\langle Ax, x\rangle| \leq k\|x\|^2$ for all x and, from the Cauchy-Schwatrz in equality, $k \leq \|A\|$. Since

$$\|A\| = \sup_{\|x\| \leq 1} \|Ax\| = \sup_{\|x\| \leq 1} \sup_{\|y\| \leq 1} |\langle Ax, y\rangle| \, ,$$

to show that $\|A\| \leq k$, we need to show that $|\langle Ax, y\rangle| \leq k$ whenever $\|x\| \leq 1$ and $\|y\| \leq 1$. It is sufficient to prove this when $\langle Ax, y\rangle$ is real, since if $|\langle Ax, y\rangle| = e^{i\theta}\langle Ax, y\rangle$

then applying the result for the real case for $\langle Ax', y\rangle$ where $x' = e^{i\theta}x$, proves the general result.

Now, using the polarization identity (Lemma 2.3) and the paralellogram law (Lemma 1.4),

$$\begin{aligned} 4\langle Ax, y\rangle &= \langle A(x+y),(x+y)\rangle - \langle A(x-y),(x-y)\rangle \\ &\quad +i[\langle A(x+iy),(x+iy)\rangle - \langle A(x-iy),(x-iy)\rangle] \\ &\leq k\{\|x+y\|^2 + \|x-y\|^2\} \\ &= k(2\|x\|^2 + 2\|y\|^2) \leq 4k\,, \end{aligned}$$

(the expression in square brackets being zero since $\langle Ax, y\rangle$ is real). ■

Note that $\sup_{\|x\|\leq 1} |\langle Ax, x\rangle| = \sup_{\|x\|=1} |\langle Ax, x\rangle|$. We write

$$m = \inf_{\|x\|=1} \langle Ax, x\rangle \quad \text{and} \quad M = \sup_{\|x\|=1} \langle Ax, x\rangle\,.$$

Corollary 5.4 *For all $T \in \mathcal{B}(\mathcal{H})$*

$$\|T^*T\| = \|T\|^2\,.$$

Proof. Since T^*T is selfadjoint,

$$\|T^*T\| = \sup_{\|x\|\leq 1} |\langle T^*Tx, x\rangle| = \sup_{\|x\|\leq 1} \|Tx\|^2 = \|T\|^2\,.$$

■

The key to the next result is proving that $\| (\lambda I - A)x \| \geq c \| x \|$ whenever $\lambda \notin [m, M]$. This is done by a single calculation in the body of the proof. However, it can also be established by a sequence of simpler proofs as follows. Note that, if X is selfadjoint then

$$\begin{aligned} \|(iI - X)x\|^2 &= \langle (iI - X)x, (iI - X)x\rangle \\ &= \|x\|^2 + \|Xx\|^2 - i\langle x, Xx\rangle + i\langle Xx, x\rangle \\ &= \|Xx\|^2 + \|x\|^2 \geq \|x\|^2\,. \end{aligned}$$

Thus, if $\lambda = \xi + i\eta$ is not real (i.e. $\eta \neq 0$), then, using the above result for $X = \frac{1}{\eta}(A - \xi I)$, we have

$$\| (\lambda I - A)x \| = \| \eta(iI - X)x \| \geq |\eta|\|x\|\,.$$

If λ is real with $\lambda > M$, we have that for $\|x\| = 1$,

$$\|(\lambda I - A)x\| = \sup_{\|y\|\leq 1} |\langle (\lambda I - A)x, y\rangle| \geq \langle (\lambda I - A)x, x\rangle \geq \lambda - M$$

so that (dividing by $\|x\|$) we have $\|(\lambda I - A)x\| \geq (\lambda - M)\|x\|$ for all x. A similar proof holds when $\lambda < m$.

Theorem 5.5
(i) $\sigma(A) \subseteq [m, M]$,
(ii) *$m \in \sigma(A)$ and $M \in \sigma(A)$.*

Proof. (i) Suppose $\lambda \notin [m, M]$ and let $d = \text{dist}(\lambda, [m, M])$. Let $x \in \mathcal{H}$ be any unit vector and write $\alpha = \langle Ax, x\rangle$. Then $\langle(\alpha I - A)x, x\rangle = \langle x, (\alpha I - A)x\rangle = 0$ and

$$\begin{aligned}
\|(\lambda I - A)x\|^2 &= \|[\lambda I - \alpha I + (\alpha I - A)]x\|^2 \\
&= \langle[\lambda I - \alpha I + (\alpha I - A)]x, [\lambda I - \alpha I + (\alpha I - A)]x\rangle \\
&= |\lambda - \alpha|^2\|x\|^2 + (\bar{\alpha} - \bar{\lambda})\langle(\alpha I - A)x, x\rangle \\
&\qquad +(\alpha - \lambda)\langle x, (\alpha I - A)x\rangle + \|(\alpha I - A)x\|^2 \\
&\geq |\lambda - \alpha|^2 \geq d^2 .
\end{aligned}$$

It follows that $\|(\lambda I - A)x\| \geq d\|x\|$ [apply the above for $\frac{x}{\|x\|}$]. Hence $\lambda I - A$ is injective and, by Lemma 5.1, it has closed range. Further, if $0 \neq z \perp \text{ran}(\lambda I - A)$ then $0 = \langle(\lambda I - A)x, z\rangle = \langle x, (\bar{\lambda} I - A)z\rangle$ for all x and so $(\bar{\lambda} I - A)z = 0$. But this is impossible, since, from above, noting that $d = \text{dist}(\lambda, [m, M]) = \text{dist}(\bar{\lambda}, [m, M])$, we have $\|\bar{\lambda} I - A)z\| \geq d\|z\|$. Therefore, $\text{ran}(\lambda I - A) = \mathcal{H}$, (being both dense and closed).

Therefore, for any $y \in \mathcal{H}$, there is a unique $x \in \mathcal{H}$ such that $y = (\lambda I - A)x$. Define $(\lambda I - A)^{-1}y = x$. Then $\|y\| \geq d\|x\|$ so

$$\|(\lambda I - A)^{-1}y\| = \|x\| \leq \tfrac{1}{d}\|y\|$$

showing that $(\lambda I - A)^{-1} \in \mathcal{B}(\mathcal{H})$ (i.e. it is continuous). Thus $\lambda \notin \sigma(A)$, proving (i).

(ii) From Lemma 5.3, $\|A\|$ is either M or $-m$. If $\|A\| = M = \sup_{\|x\|=1}\langle Ax, x\rangle$; (if $\|A\| = -m$ the same proofs, applied to $-A$, hold) there exists a sequence (x_n) of unit vectors such that $(\langle Ax_n, x_n\rangle) \to M$. Then

$$\|(A - MI)x_n\|^2 = \|Ax_n\|^2 + M^2 - 2M\langle Ax_n, x_n\rangle \leq 2M^2 - 2M\langle Ax_n, x_n\rangle \to 0 .$$

Hence $A - MI$ has no inverse in $\mathcal{B}(\mathcal{H})$ [since if X were such an operator, $1 = \|x_n\| = \|X(A - MI)x_n\| \leq \|X\|.\|X(A - MI)x_n\| \to 0$] and so $M \in \sigma(A)$. For m, note that

$$\sup_{\|x\|=1} \langle(MI - A)x, x\rangle = M - m = \|MI - A\|$$

since $\inf_{\|x\|=1}\langle(MI - A)x, x\rangle = 0$. Applying the result just proved to the operator $MI - A$ shows that $M - m \in \sigma(MI - A)$, that is, $(M - m)I - (MI - A) = A - mI$ has no inverse. Hence $m \in \sigma(A)$. ■

The *spectral radius*, $\nu(T)$, of an operator T is defined to be

$$\nu(T) = \sup\{|\lambda| : \lambda \in \sigma(T)\} .$$

Thus we have shown that the spectrum of a selfadjoint operator is non-empty and real and its norm is equal to its spectral radius.

Exercises 5

1. Let $X, T \in \mathcal{B}(\mathcal{H})$ and suppose that X is invertible. Prove that $\sigma(T) = \sigma(X^{-1}TX)$.

2. Let $A \in \mathcal{B}(\mathcal{H})$ be a selfadjoint operator. Show that $U = (A - iI)(A + iI)^{-1}$ is a unitary operator.

6 The Spectral analysis of compact operators.

In this section K will always denote a compact operator.

Theorem 6.1 *If $\lambda \neq 0$ then either λ is an eigenvalue of K or $\lambda \in \rho(K)$.*

Proof. Suppose that $\lambda \neq 0$ is not an eigenvalue of K. We show that $\lambda \in \rho(K)$. The proof of this is in several stages.

(a) *For some $c > 0$, we have that that $\|(\lambda I - K)x\| \geq c\|x\|$ for all $x \in \mathcal{H}$.*
Suppose this is false. Then the inequality fails for $c = \frac{1}{k}$ for $k = 1, 2, \cdots$. Therefore there is a sequence of unit vectors such that

$$\|(\lambda I - K)x_k\| \leq \frac{1}{k},$$

that is, $((\lambda I - K)x_k) \to 0$. Applying the condition that K is compact, there is a subsequence (x_{k_i}) such that (Kx_{k_i}) is convergent. Call its limit y. Then

$$x_{k_i} = \frac{1}{\lambda}((\lambda I - K)x_{k_i} + Kx_{k_i})$$

and so $(x_{k_i}) \to \frac{y}{\lambda}$. Since (x_{k_i}) is a sequence of unit vectors, $y \neq 0$. But then,

$$(\lambda I - K)y = \lim_{i\to\infty}(\lambda I - K)x_{k_i} = \lambda\frac{y}{\lambda} - y = 0\,.$$

This contradicts the fact that λ is not an eigenvalue, so (a) is established.

(b) $\operatorname{ran}(\lambda I - K) = \mathcal{H}$.
Let $H_n = \operatorname{ran}(\lambda I - K)^n$ and write $H_0 = \mathcal{H}$. It follows from (a) using Lemma 5.1 that (H_n) is a sequence of closed subspaces. Also

$$(\lambda I - K)H_n = H_{n+1}$$

$$H_0 \supseteq H_1 \supseteq H_2 \supseteq H_3 \supseteq \cdots .$$

Note that, if $y \in H_n$ then $Ky = ((K - \lambda I)y + \lambda y) \in H_n$ so that $K(H_n) \subseteq H_n$.

We now use the compactness of K to show that the inclusion $H_n \subseteq H_{n+1}$ is not always proper. Suppose, on the contrary that

$$H_0 \supset H_1 \supset H_2 \supset H_3 \supset \cdots .$$

Using Lemma 1.7, for each n we can find a unit vector x_n such that $x_n \in H_n$ and $x_n \perp H_{n+1}$. We show that (Kx_n) cannot have a Cauchy subsequence. Indeed, if $m > n$

$$\begin{aligned} Kx_n - Kx_m &= (K - \lambda I)x_n + \lambda x_n - Kx_m \\ &= \lambda x_n + [(K - \lambda I)x_n - Kx_m] \\ &= \lambda x_n + z \end{aligned}$$

where $z \in H_{n+1}$ [$Kx_m \in H_m \subseteq H_{n+1}$ follows from $m > n$]. Thus

$$\|Kx_n - Kx_m\|^2 = |\lambda|^2 + \|z\|^2 \geq |\lambda|^2$$

and so (Kx_n) has no convergent Cauchy subsequence. Therefore, the inclusion is not always proper. Let k be the smallest integer such that $H_k = H_{k+1}$. If $k \neq 0$ then choose $x \in H_{k-1} \setminus H_k$. Then $(\lambda I - K)x \in H_k = H_{k+1}$ and so, for some y,

$$(\lambda I - K)x = (\lambda I - K)^{k+1}y = (\lambda I - K)z$$

where $z = (\lambda I - K)^k y \in H_k$. Now $x \notin H_k$ so $x - z \neq 0$ and

$$(\lambda I - K)(x - z) = 0$$

contradicting the fact that λ is not an eigenvalue. Therefore $k = 0$, that is $\mathrm{ran}(\lambda I - K) = H_1 = H_0 = \mathcal{H}$.

(c) *Completing the proof.* This is done exactly as in Theorem 5.5 (i). For any $y \in \mathcal{H}$, there is a unique $x \in \mathcal{H}$ such that $y = (\lambda I - K)x$. Define $(\lambda I - K)^{-1}y = x$. Then $\|y\| \geq c\|x\|$ so

$$\|(\lambda I - K)^{-1}y\| = \|x\| \leq \tfrac{1}{c}\|y\|$$

showing that $(\lambda I - K)^{-1} \in \mathcal{B}(\mathcal{H})$ (i.e. it is continuous). Thus $\lambda \notin \sigma(K)$. ■

Lemma 6.2 *If $\{x_n\}$ are eigenvectors of K corresponding to different eigenvalues $\{x_n\}$, then $\{x_n\}$ is a linearly independent set.*

Proof. This is exactly as in an elementary linear algebra course. Suppose the statement is false and k is the first integer such that $x_1, x_2, \cdots, x_k$ is linearly dependent. Then $\sum_{i=1}^{k} \alpha_i x_i = 0$ and $\alpha_k \neq 0$. Also, by hypothesis $Kx_i = \lambda_i x_i$ with the λ_i's all different. Now $x_k = \sum_{i=1}^{k-1} \beta_i x_i$ (where $\beta_i = -\alpha_i/\alpha_k$) and so

$$0 = (\lambda_k I - K)x_k = \sum_{i=1}^{k-1} (\lambda_k - \lambda_i)\beta_i x_i$$

showing that $x_1, x_2, \cdots, x_{k-1}$ is linearly dependent, contradicting the definition of k. ■

Theorem 6.3 *$\sigma(K)\setminus\{0\}$ consists of eigenvalues with finite-dimensional eigenspaces. The only possible point of accumulation of $\sigma(K)$ is 0.*

Proof. Let λ be any non-zero eigenvalue and let $N = \{x : Kx = \lambda x\}$ be the eigenspace of λ. If N is not finite-dimensional, we can find an orthonormal sequence (x_n) of elements of N [apply the Gram-Schmidt process (Theorem 3.4) to any linearly independent sequence]. Then

$$\|Kx_n - Kx_m\|^2 = \|\lambda x_n - \lambda x_m\|^2 = 2|\lambda|^2$$

which is impossible, since K is compact.

To show that $\sigma(K)$ has no points of accumulation other than (possibly) 0, we show that $\{\lambda \in \mathbb{C} : |\lambda| > \delta\} \cap \sigma(K)$ is finite for any $\delta > 0$. Suppose this is false and there is a sequence of distinct eigenvalues (λ_i) with $|\lambda_i| > \delta$ for all i. Then we have vectors x_i with $Kx_i = \lambda_i x_i$.

Let $H_n = \mathrm{span}\{x_1, x_2, \cdots, x_n\}$. Then, since $\{x_n\}$ is a linearly independent set, we have the proper inclusions

$$H_1 \subset H_2 \subset H_3 \subset H_4 \subset \cdots .$$

It is easy to see that $K(H_n) \subseteq H_n$ and $(\lambda_n I - K)H_n \subseteq H_{n-1}$. Choose, as in Theorem 6.1 a sequence of unit vectors (y_n) with $y_n \in H_n$ and $y_n \perp H_{n-1}$. Then, for $n > m$,

$$Ky_n - Ky_m = \lambda_n y_n - [(\lambda_n I - K)y_n - Ky_m] .$$

Since $(\lambda_n I - K)y_n \in H_{n-1}$ and $Ky_m \in H_m \subseteq H_{n-1}$, the vector in square brackets is in H_{n-1}. Therefore, since $y_n \perp H_{n-1}$,

$$\|Ky_n - Ky_m\| > |\lambda_n| > \delta$$

showing that (Ky_n) has no convergent subsequence. ■

Corollary 6.4 *The eigenvalues of K are countable and whenever they are put into a sequence (λ_i) we have that $\lim_{\to\infty} \lambda_i = 0$.*

Proof. [The set of all eigenvalues is (possibly) 0 together with the countable union of the finite sets of eigenvalues $> \frac{1}{n}, (n = 1, 2, \cdots)$.

If $\epsilon > 0$ is given then, since $\lambda : \lambda$ an eigenvalue of $K, |\lambda| \geq \epsilon$ is finite, we have that $|\lambda_i| < \epsilon$ for all but a finite number of values of i. Hence $(\lambda_i) \to 0$.] ■

Corollary 6.5 *If A is a compact selfadjoint operator then $\|A\|$ equals its eigenvalue of largest modulus.*

Proof. This is immediate from Theorem 5.5 (ii). ■

The Fredholm alternative. For any scalar μ, either

$$(I - \mu K)^{-1} \text{ exists}$$

or the equation

$$(I - \mu K)x = 0$$

has a finite number of linearly independent solutions.

(Fredholm formulated this result for the specific operator $(Kf)(x) = \int_a^b k(x,t)f(t)\,dt$. In fact, he said : EITHER the integral equation

$$f(x) - \mu \int_a^b k(x,t)f(t)\,dt \ = g(x)$$

has a unique solution, OR the associated homogeneous equation

$$f(x) - \mu \int_a^b k(x,t)f(t)\,dt \ = 0$$

has a finite number of linearly independent solutions.)

We now turn to compact selfadjoint operators. For the rest of this section A will denote a compact selfadjoint operator.

Note that every eigenvalue of of A is real. This is immediate from Theorem 5.5, but can be proved much more simply since if $Ax = \lambda x$, where x is a unit eigenvector,

$$\bar{\lambda} = \overline{\langle Ax, x\rangle} = \langle x, Ax\rangle = \langle A^*x, x\rangle = \langle Ax, x\rangle = \lambda .$$

Lemma 6.6 *Distinct eigenspaces of A are mutually orthogonal.*

Proof. Let x and y be eigenvectors corresponding to distinct eigenvalues λ and μ. Then,

$$\lambda\langle x, y\rangle = \langle Ax, y\rangle = \langle x, A^*y\rangle = \langle x, Ay\rangle = \bar{\mu}\langle x, y\rangle = \mu\langle x, y\rangle$$

(since μ is real) and so $\langle x, y\rangle = 0$. ■

Theorem 6.7 *If A is a compact selfadjoint operator on a Hilbert space $\mathcal{H}$ then $\mathcal{H}$ has an orthonormal basis consisting of eigenvectors of A.*

Proof. Let $(\lambda_i)_{i=1,2,\ldots}$ be the sequence of all the non-zero eigenvalues of A and let N_i be the eigenspace of λ_i. Take an orthonormal basis of each N_i and an orthonormal basis of $N_0 = \ker A$. Let (x_n) be the union of all these, put into a sequence. It follows from Lemma 6.6 that this sequence is orthonormal.

Let $M = \{z : z \perp x_n \text{ for all } n\}$. Then, if $y \in M$ we have that $\langle x_n, Ay\rangle = \langle Ax_n, y\rangle = \lambda_n\langle x_n, y\rangle = 0$ and so $A(M) \subseteq M$. Therefore A with its domain restricted to M is a compact selfadjoint operator on the Hilbert space M. Clearly this operator is selfadjoint [$\langle Ax, y\rangle = \langle x, Ay\rangle$ for all $x, y \in \mathcal{H}$ so certainly for all $x, y \in M$]. Also it cannot be have a no-zero eigenvector [for then $M \cap N_k \neq (0)$ for some $k > 0$]. Therefore, by Corollary 6.5, it is zero. But then $M \subseteq N_0$. But also $M \perp N_0$ and so $M = (0)$. Therefore (x_n) is a basis. ■

Corollary 6.8 *Then there is an orthonormal basis $\{x_n\}$ of $\mathcal{H}$ such that, for all h,*

$$Ah = \sum_{n=1}^{\infty} \lambda_n\langle h, x_n\rangle x_n\,.$$

Proof. Let (x_n) be the basis found in the Theorem and let $\lambda_n = \langle Ax_n, x_n\rangle$ (this is merely re-labeling the eigenvalues. The from Theorem 3.3 (iii), for any $h \in \mathcal{H}$,

$$h = \sum_{n=1}^{\infty} \langle h, x_n\rangle x_n\,.$$

Acting on this by A, since A is continuous and $Ax_n = \lambda_n x_n$ we have that

$$Ah = \sum_{n=1}^{\infty} \lambda_n\langle h, x_n\rangle x_n\,.$$

■

Theorem 6.9 *If A is a compact selfadjoint operator on a Hilbert space $\mathcal{H}$ then there is an orthonormal basis $\{x_n\}$ of $\mathcal{H}$ such that*

$$A = \sum_{n=1}^{\infty} \lambda_n(x_n \otimes x_n)$$

where the series is convergent in norm.

Proof. Let $\{x_n\}$ be the basis found as above so that $Ax_n = \lambda_n x_n$ and

$$Ah = \sum_{n=1}^{\infty} \lambda_n\langle h, x_n\rangle x_n\,.$$

Note that $(\lambda_n) \to 0$. Let

$$A_k = \sum_{n=1}^{k} \lambda_n (x_n \otimes x_n)\,.$$

We need to show that $\|A - A_k\| \to 0$ as $k \to \infty$.

Now

$$(A - A_k)h = \sum_{n=k+1}^{\infty} \lambda_n \langle h, x_n \rangle x_n\,.$$

and, using Theorem 3.3 (v)

$$\begin{aligned}
\|(A - A_k)h\|^2 &= \sum_{n=k+1}^{\infty} |\lambda_n \langle h, x_n \rangle|^2 \\
&\leq \sup_{n \geq k+1} |\lambda_n|^2 \sum_{n=k+1}^{\infty} |\langle h, x_n \rangle|^2 \\
&\leq \sup_{n \geq k+1} |\lambda_n|^2 \sum_{n=1}^{\infty} |\langle h, x_n \rangle|^2 \\
&= \sup_{n \geq k+1} |\lambda_n|^2 \, \|h\|^2\,.
\end{aligned}$$

Thus $\|(A - A_k)\| \leq \sup_{n \geq k+1} |\lambda_n|$, and so since $(\lambda_n) \to 0$, we have that $\|A - A_k\| \to 0$ as $k \to \infty$. ■

Alternatively, Theorem 4.4 may be used to prove the above result. Let $\{x_n\}$ and A_k and A be as above and let

$$P_k = \sum_{n=1}^{k} (x_n \otimes x_n)\,.$$

Then, since $\{x_n\}$ is a basis, Theorem 3.3 (iii) shows that (P_k) converges pointwise to the identity operator I. Since $A_k = AP_k$, Theorem 4.4 shows that (A_k) converges to A in norm.

Exercises 6

1. Let K be a compact operator on a Hilbert space $\mathcal{H}$ and let $\lambda \neq 0$ be an eigenvalue of K. Show that $\lambda I - K$ has closed range. [Hint : let $N = \ker(\lambda I - K)$ and let $M = N^{\perp}$. If $y \in \operatorname{ran}(\lambda I - K)$, show that $y = \lim_{n\to\infty}(\lambda I - K)z_n$ with $z_n \in M$. Now imitate the proof for the case when λ is not an eigenvalue.]

2. Find the norm of the compact operator V defined on $L^2[0,1]$ by

$$(Vf)(x) = \int_0^x f(t)\,dt$$

.

Hints: Use Corollary 5.4 and the fact that the norm of the compact selfadjoint operator V^*V is given by its largest eigenvalue. Now use the result of Exercises 2 Question 6 to show that if f satisfies $V^*Vf = \lambda f$ then it satisfies

$$\begin{cases} \lambda f'' + f = 0 \\ f(1) = 0, \qquad f'(0) = 0. \end{cases}$$

[You may assume that any vector in the range of V^*V (being in the range of two integrations) is twice differentiable (almost everywhere).]

Note that a direct approach to evaluating $\| V \|$ seems to be very difficult (try it !).

3. Let $\{x_n\}$ be an orthonormal basis of $\mathcal{H}$ and suppose that $T \in \mathcal{B}(\mathcal{H})$ is such that the series $\sum_{n=1}^{\infty} \|Tx_n\|^2$ converges. Prove that

(i) T is compact,

(ii) $\sum_{n=1}^{\infty} \|Ty_n\|^2$ converges for every orthonormal basis $\{y_n\}$ of $\mathcal{H}$ and for the sum is the same for every orthonormal basis.

Note : an operator satisfying the above is called a *Hilbert-Schmidt* operator.

Hints: (i) write $h \in \mathcal{H}$ as a Fourier series, $h = \sum_{i=1}^{\infty} \alpha_i x_i$ where $\alpha_i = \langle h, x_i\rangle$. Define $T_n h = \sum_{i=1}^{n} \alpha_i T x_i$ and show that

$$\|(T - T_n)h\|^2 \leq \left(\sum_{n+1}^{\infty} |\alpha_i|.\|Tx_i\|\right) \leq \|h\|^2.\left(\sum_{n+1}^{\infty} \|Tx_i\|^2\right).$$

(ii) Take an orthonormal basis ϕ_k of $\mathcal{H}$ consisting of eigenvectors of the compact operator T^*T. Observe that if $T^*T\phi_k = \mu_k\phi_k$ then $\mu_k = \langle T^*T\phi_k, \phi_k\rangle = \|T\phi_k\|^2 \geq 0$. Now use the spectral theorem for T^*T to prove that if for any orthonormal basis $\{x_n\}$, $\sum_{n=1}^{\infty} \|Tx_n\|^2$ converges then

$$\sum_{n=1}^{\infty} \|Tx_n\|^2 = \sum_{n=1}^{\infty} \langle T^*Tx_n, x_n\rangle = \sum_{k=1}^{\infty} \mu_k\,.$$

Note that for a double infinite series with all terms positive, the order of summation may be interchanged.

7 The Sturm-Liouville problem.

In this section we shall discuss the differential operator

$$Ly = \frac{d}{dx}\left(p(x)\frac{dy}{dx}\right) + q(x)y$$

acting on functions y defined on a closed bounded interval $[a,b]$. We shall assume that $p(x) > 0$ and $q(x)$ real for $a \le x \le b$.

We make further assumptions that may be summarized, broadly speaking, by saying that "everything makes sense". Specifically we need L to act on functions that are twice differentiable and whose second derivatives are in $L^2[a,b]$. We also need to have that p is differentiable with p' continuous on $[a,b]$.

We shall be concerned with solving the problem

$$Ly = \frac{d}{dx}\left(p(x)\frac{dy}{dx}\right) + q(x)y = f(x) \qquad (*)$$

subject to boundary conditions

$$\left.\begin{aligned} \alpha_1 y(a) + \alpha_2 y'(a) &= 0 \\ \beta_1 y(b) + \beta_2 y'(b) &= 0 \end{aligned}\right\} \qquad (\dagger)$$

Where $\alpha_1, \alpha_2, \beta_1$ and β_2 are real and $\alpha_1\alpha_2 \neq 0,\ \beta_1\beta_2 \neq 0$.

Note. The following calculation is of interest because it shows that L satisfies a symmetry condition that would, for a bounded operator, make it self-adjoint. However, it will not be used in the sequel. If L is restricted to act on the set of functions that satisfy the boundary conditions, then $\langle Ly, z\rangle = \langle y, Lz\rangle$. Indeed,

$$\begin{aligned} \langle -Ly, z\rangle + \langle y, Lz\rangle &= \int_a^b \left[\frac{d}{dt}\left(p(t)\frac{d\bar{z}(t)}{dt}\right)y - \frac{d}{dt}\left(p(t)\frac{dy}{dt}\right)\bar{z}\right]dt + \int_a^b (-qy\bar{z} + yq\bar{z})\,dt \\ &= \left[p(t)\frac{d\bar{z}}{dt}y(t) - p(t)\frac{dy}{dt}\bar{z}(t)\right]_a^b + \int_a^b p(t)\left(\frac{dy}{dt}\frac{d\bar{z}}{dt} - \frac{d\bar{z}}{dt}.\frac{dy}{dt}\right)dt \end{aligned}$$

and, as the integrals on the right of each line are 0, this will vanish if

$$p(b)\left[\bar{z}'(b)y(b) - \bar{z}(b)y'(b)\right] = p(a)\left[\bar{z}'(a)y(a) - \bar{z}(a)y'(a)\right].$$

But $\bar{z}'(a)y(a) - \bar{z}(a)y'(a)$ is the determinant of the 2×2 system

$$\begin{aligned} \xi\bar{z}(a) + \eta\bar{z}'(a) &= 0\,, \\ \xi y(a) + \eta y'(a) &= 0\,, \end{aligned}$$

which has the non-trivial solution $(\xi, \eta) = (\alpha_1, \alpha_2)$ when y and z satisfy the boundary conditions ($\dagger$). So $\bar{z}'(a)y(a) - \bar{z}(a)y'(a) = 0$ and similarly $\bar{z}'(b)y(b) - \bar{z}(b)y'(b) = 0$. Therefore $\langle Ly, z\rangle = \langle y, Lz\rangle$.

We shall be looking for eigenvalues and eigenfunctions of L that satisfy the conditions ($\dagger$); that is, for scalars λ and corresponding functions f that satisfy ($\dagger$) and the equation $Lf = \lambda f$. We make the additional assumption that $\lambda = 0$ is not an eigenvalue of the system. This is quite a reasonable assumption, since if it fails then the problem (*), subject to ($\dagger$) does not have a unique solution [an arbitrary multiple of the eigenfunction corresponding to $\lambda = 0$ could be added to any solution to obtain another solution].

Theorem 7.1 (Existence of the Green's function.) *Under the assumptions stated above, the problem (*), subject to (†) has the solution*

$$y(x) = \int_a^b k(x,t)\, f(t)\, dt$$

where $k(x,t)$ is real-valued and continuous on the square $[a,b] \times [a,b]$.

Proof. From the elementary theory of the initial value problem for linear differential equations, (also from Questions 1,2 and 3 of Exercises 6) we have that there is a unique function u such that $Lu = 0,\ u(a) = -\alpha_2, u'(a) = \alpha_1$. It follows easily that every solution of $Ly = 0,\ \alpha_1 y(a) + \alpha_2 y'(a) = 0$ is a scalar multiple of u. Similarly we have a unique v such that $Lv = 0, v(b) = -\beta_2, v'(b) = \beta_1$. The assumption that 0 is not an eigenvalue implies that u and v are linearly independent [if u were a multiple of v then it would be an eigenfunction].

Let

$$k(x,t) = \begin{cases} l\, u(x) & a \le x \le t \\ m\, v(x) & t \le x \le b \end{cases}$$

where (for fixed t) l, m are constants to be chosen. [Our motivational work suggests that we require $k(x,t)$ to be continuous and $p(x).\frac{\partial}{\partial x}k(x,t)$ to have a unit discontinuity at $x = t$.] Choose l, m such that

$$\begin{aligned} m.v(t) - l.u(t) &= 0, \\ p(t)[m.v'(t) - l.u'(t)] &= 1. \end{aligned}$$

Solving for l, m gives

$$l = \frac{v(t)}{\Delta} \qquad m = \frac{u(t)}{\Delta}$$

where $\Delta = p(t)(v'u - u'v) = pJ(u,v)$, where J is the Jacobean and hence non-zero [since u and v are independent]. Also,

$$\frac{d\Delta}{dt} = u(pv')' + u'(pv') - v(pu')' - v'(pu') = -quv + vqu = 0\,,$$

so Δ is a constant (i.e. also independent of t). [One can see, independently of the theory of Jacobeans, that $\Delta \neq 0$ since otherwise, at some point t_0

$$\begin{aligned} \xi.u(t_0) + \eta.v(t_0) &= 0, \\ \xi.u'(t_0) + \eta.v'(t_0) &= 0 \end{aligned}$$

has a non-trivial solution (ξ, η). Then $y = \xi.u + \eta.v$ is a solution of $Ly = 0, y(t_0) = y'(t_0) = 0$ and so $\xi.u + \eta.v$ is identically 0, contradicting the linear independence of u and v.] Hence we have that

$$k(x,t) = \begin{cases} \frac{u(x).v(t)}{\Delta} & a \le x \le t\,, \\ \frac{u(t).v(x)}{\Delta} & t \le x \le b\,. \end{cases}$$

To complete the proof, we just verify directly that

$$y(x) = \int_a^b k(x,t)\, f(t)\, dt$$

is the required solution. First note that when $x = a$ we have $x \le t$ throughout the range of integration and so

$$y(a) = \frac{1}{\Delta} \int_a^b u(a).v(t)\, f(t)\, dt$$

and, since $y'(x) = \displaystyle\int_a^b \frac{\partial k(x,t)}{\partial x}\, f(t)\, dt,$

$$y'(a) = \frac{1}{\Delta} \int_a^b u'(a).v(t)\, f(t)\, dt\,.$$

Therefore $\alpha_1 y(a) + \alpha_2 y'(a) = 0$ since u satisfies the boundary condition at $x = a$. Similarly $\beta_1 y(b) + \beta_2 y'(b) = 0$.

We now substitute into the equation. For notational convenience we substitute $\Delta\, y(x)$ (remembering that Δ is constant). Since $u(x), v(x)$ can be taken outside the integration, we obtain

$$\begin{aligned}
\Delta\, y(x) &= \Delta \int_a^b k(x,t)\, f(t)\, dt = v(x) \int_a^x u(t)\, f(t)\, dt + u(x) \int_x^b v(t)\, f(t)\, dt \\
(\Delta\, y(x))' &= v(x)u(x)f(x) + v'(x) \int_a^x u(t)\, f(t)\, dt \\
&\qquad - u(x)v(x)f(x) + u'(x) \int_x^b v(t)\, f(t)\, dt \\
(p(x)(\Delta\, y(x))')' &= (pv')' \int_a^x u(t)\, f(t)\, dt + pv'uf + (pu')' \int_x^b v(t)\, f(t)\, dt - pu'vf\,.
\end{aligned}$$

Therefore

$$\begin{aligned}
(p(x)(\Delta\, y(x))')' + q\Delta y &= [(pv')' + qv] \int_a^x u(t)\, f(t)\, dt \\
&\qquad + [(pu')' + qu] \int_x^b v(t)\, f(t)\, dt + pf(v'u - u'v) \\
&= f.\Delta
\end{aligned}$$

since u and v are solutions of $Ly = 0$. ■

We can now apply the results of Section 6 to draw conclusions about the eigenfunctions and eigenvalues of the Sturm-Liouville system (*), (†).

Define the operator K by

$$(Kf)(x) = \int_a^b k(x,t)\, f(t)\, dt\,.$$

Since k is continuous on $[a,b]\times[a,b]$ it is clear that $\int\int |k|^2 < \infty$ and so, as shown in Section 4, K is compact.

Let $\mathcal{D}$ be the set of functions y such that Ly exists as a function in $L^2[a,b]$ and satisfies the boundary conditions (†). (There is a little technical hand waving here. A more precise statement is: $y \in \mathcal{D}$ (which is an equivalence class of functions) if there is a representative y such that y is differentiable and $p.(y)'$ has a derivative almost everywhere such that $(p.(y)')' \in L^2[a,b]$. Informally $\mathcal{D}$ is the all $y \in L^2[a,b]$ that qualify as solutions of (*),(†) for some right hand side.) If $y \in \mathcal{D}$ and $f = Ly$ then Theorem 7.1 shows that $Kf = K(Ly) = y$, that is KL acts like the identity on $\mathcal{D}$.

In the other order, it follows from the proof of Theorem 7.1, that for every $f \in L^2[a,b]$ $Kf \in \mathcal{D}$. (The verification that Kf is a solution of $Ly = f$ explicitly shows this. Naturally, for the most general f, the differentiation of expressions like $\int_a^x u(t)\, f(t)\, dt$ one requires the relevant background from Lebesgue integration.) Also, from Theorem 7.1, $LKf = f$. Thus $LK = I$.

Note that L fails to be an inverse of K since it is not defined on the whole of $L^2[a,b]$, the Hilbert space in question. Indeed, since K is compact, it cannot be invertible. However, L *is* defined on a dense subset.

We use these notations and observations in the statements a proofs below.

Theorem 7.2 *(i)* *The operator K does not have $\lambda = 0$ as an eigenvalue.*
(ii) *λ is an eigenvalue of K if an only if $\mu = \frac{1}{\lambda}$ is an eigenvalue of the system* (*),(†).

Consequently, the system (*),(†)

1. *has a countable sequence (μ_i) of real eigenvalues such that $(|\mu_i|) \to \infty$;*
2. *has eigenfunctions which form an orthonormal basis of $L^2[a,b]$;*
3. *has finite-dimensional eigenspaces.*

Proof. (i) If $f \neq 0$ the solution of $Ly = f$ is Kf and cannot be $y = 0$. Therefore 0 is not an eigenvalue of K.

(ii) If λ is an eigenvalue of K the $K\phi = \lambda\phi$ and since ϕ is in the range of K, from the discussion above, $\phi \in \mathcal{D}$. Then $\lambda L\phi = LK\phi = \phi$ so that

$$L\phi = \frac{1}{\lambda}\phi = \mu\phi,$$

and μ is an eigenvalue of (*),(†).

Conversely, if μ is an eigenvalue of (*),(†), by assumption $\mu \neq 0$. We then have $L\phi = \mu\phi$ and $\phi \in \mathcal{D}$. Then

$$KL\phi = \phi = \mu K\phi$$

and so $K\phi = \lambda\phi$ where $\lambda = \frac{1}{\mu}$ is an eigenvalue of K.

The consequences are immediate deductions from the results of Section 6 (principally 6.3, 6.4 and 6.7). Note that in this case the set of eigenvalues of K cannot be finite because this would imply (by Corollary 6.8) that K vanishes on a non-zero (in fact, infinite-dimensional) subspace. ■

The most important result arising from this is consequence 2, since, for example, this is what justifies the expansions that are required in solving partial differential equations by the method of separation of variables.

Note. The assumption that 0 is not an eigenvalue of the system is not an essential restriction. For any constant c, the eigenfunctions of L and $L+c$ are the same and the eigenvalues of $L+c$ are $\lambda + c$ whenever λ is an eigenvalue L. It is a fact that, by adding a suitable constant to q we can always ensure that $\lambda = 0$ is not an eigenvalue of the system. For example, if the boundary conditions are $y(a) = y(b) = 0$, choose c so that $c + q$ does not change sign in $[a, b]$; for definiteness, assume $c + q(t) < 0$ for $a \leq t \leq b$. Let u be the (unique) solution of

$$L_c y = \frac{d}{dt}\left(p(t)\frac{dy}{dt}\right) + [c + q(t)]y = 0, \qquad y(a) = 0, y'(a) = 1\,.$$

Any solution of $L_c y = 0,\ y(a) = 0$ is a multiple of u so to show that 0 is not an eigenvalue we must show that $u(b) \neq 0$.

Since $u'(a) > 0$ and $u(a) = 0$, it follows that u is strictly positive in some interval $(a, a+\delta)$. Suppose z is the smallest zero of u that is $> a$ (if any). Then u' must vanish between a and z. If $z \leq b$ then $(pu')' = -[c+q]u$ is positive in (a, z) and so pu' is increasing in (a, z). But pu' is strictly positive at a so it is strictly positive in (a, z) contradicting that u' vanishes between a and z. Hence 0 is not an eigenvalue of $L_c y = 0, y \in \mathcal{D}$.

Similar, but more complicated arguments can be used for other boundary conditions (see Dieudonne, "Foundations of modern analysis", Chapter XI, Section 7).

The trigonometric functions form an orthonormal basis of $L^2[-\pi, \pi]$. This fact can be deduced from the work of the present section. [The trigonometric functions are actually the eigenfunctions of $y'' = 0$ subject to periodic boundary conditions $y(-\pi) = y(\pi), y'(-\pi) = y'(\pi)$, and such systems are not covered by our work; however the device below gives us the result.]

The eigenfunctions of the system $y'' = 0, y(0) = y(\pi) = 0$ are the functions $\{\sin nt : n = 1, 2, 3, \cdots\}$. Therefore, from Theorem 7.2 these form an orthonormal basis of $L^2[0, \pi]$. Similarly the functions $\{\cos nt : n = 0, 1, 2, 3, \cdots\}$ are the eigenfunctions of the system $y'{}' = 0, y'(0) = y'(\pi) = 0$ and so also form an orthonormal basis of $L^2[0, \pi]$. (It is true that

0 is an eigenvalue of the latter system, but this is covered by the note above. Alternatively, on can consider the system $y'' + ky = 0, y'(0) = y'(\pi) = 0$ for a suitable constant k – any non-integral k will do).

Now suppose that $f \in L^2[-\pi, \pi]$ is orthogonal to all the trigonometric functions. Then a simple change of variable shows that for each integer n,

$$0 = \int_{-\pi}^{\pi} f(t) \sin nt \, dt = \int_{-\pi}^{0} f(t) \sin nt \, dt + \int_{0}^{\pi} f(t) \sin nt \, dt = \int_{0}^{\pi} [f(t) - f(-t)] \sin nt \, dt \, .$$

Therefore $f(t) = f(-t)$ (almost everywhere) for $0 \leq t \leq \pi$, showing that f is an even function on $[-\pi, \pi]$. A similar calculation with cosines shows that f is also an odd function on $[-\pi, \pi]$. Thus $f = 0$ (almost everywhere) and the fact is proved.

Exercises 7

1. Let X be a Banach space (that is, a normed linear space that is complete). A series $\sum_n x_n$ in X is said to be **absolutely convergent** if the series $\sum_n \| x_n \|$ of real numbers is converegent. Prove that an absolutely convergent series in a Banach space is convergent. [Hint: prove that the sequence of partial sums is Cauchy.]

Existence theory for linear initial value problems using operator theory.

2. Let $k(x,t)$ be bounded and square integrable over the square $[a,b] \times [a,b]$, (in particular this will hold if k is continuous on $[a,b] \times [a,b]$). Define $K : L^2[a,b] \to L^2[a,b]$ by

$$(Kf)(x) = \int_a^x k(x,t)f(t)\,dt.$$

Prove that $\| K \| \leq (b-a)M$ where M is a bound for k in $[a,b] \times [a,b]$.

Let k_n be defined inductively by $k_1 = k$ and $k_n(x,t) = \int_t^x k(x,s)k_{n-1}(s,t)\,ds$. Prove that

$$(K^n f)(x) = \int_a^x k_n(x,t)f(t)\,dt.$$

Show by induction that

$$|k_n(x,t)| \leq M^n \frac{|x-t|^{n-1}}{(n-1)!}.$$

Using this result and the formal binomial expansion of $(I-K)^{-1}$, deduce from Question 1 that $(I-K)$ has an inverse in $\mathcal{B}(\mathcal{H})$. [Hint : after proving absolute convergence, verify by multiplication that the sum of the formal expansion is the inverse of $(I-K)$.] For each $\lambda \neq 0$, observe that $\frac{K}{\lambda}$ is an operator of the same type as K and deduce that $(\lambda - K)$ has an inverse in $\mathcal{B}(\mathcal{H})$.

3. Consider the initial value problem

$$(*) \qquad \begin{cases} y^{(n)}(x) + p_1(x)y^{(n-1)}(x) + \ldots + p_n(x)y(x) = f(x) \\ y(0) = y'(0) = \ldots = y^{(n-1)}(0) = 0 \end{cases}$$

where p_i and f are continuous. By putting $u(x) = y^{(n)}(x)$, show that, for any $b > 0$, this problem reduces to

$$(\dagger) \qquad (I-K)u = f$$

where K is an operator on $L^2[0,b]$ of the type considered in Question 6.

[Hint : show that $y^{(n-r)}(x) = \int_0^x \frac{(x-t)^{r-1}}{(r-1)!} u(t)\,dt$.]

Prove that $(\dagger)$ has a unique solution in $L^2[0,b]$ for each $b > 0$. By quoting appropriate theorems show that this solution is continuous (strictly, that the equivalence class of this solution contains a continuous function). Deduce that there is a unique continuous function with n continuous derivatives that satisfies $(*)$ in $[0,\infty)$.

Note that the general initial value problem

$$(**) \qquad \begin{cases} y^{(n)}(x) + p_1(x)y^{(n-1)}(x) + \ldots + p_n(x)y(x) = f(x) \\ y(0) = a_0, y'(0) = a_1, \ldots = y^{(n-1)}(0) = a_{n-1} \end{cases}$$

can be transformed into the form (*) by changing the dependent variable from y to z where

$$z(x) = y(x) - \sum_{k=0}^{n-1} a_k x^k$$

and thus $(**)$ also has a unique solution.

4. Find a Green's function for the system

$$y'' = f, \qquad y(0) = y(1) = 0\,.$$

Check your answer by verifying that it gives $x(x-1)$ as the solution when $f = 2$. Evaluate the eigenvalues and eigenfunctions of

$$y'' = \lambda y, \qquad y(0) = y(1) = 0\,,$$

and consequently find an orthonormal basis of $L^2[0,1]$.

5. Repeat the above question with the system

$$y'' = f, \qquad y(0) + y'(0) = y(1) - y'(1) = 0\,.$$

8 The Spectral analysis of selfadjoint operators.

In this section A will always denote a selfadjoint operator with spectrum $\sigma = \sigma(A)$ where $\sigma(A) \subseteq [m, M]$ as defined in Section 6.

Theorem 8.1 (Spectral Mapping Theorem) *Let $T \in \mathcal{B}(\mathcal{H})$. For any polynomial p,*

$$\sigma(p(T)) = \{p(\lambda) : \lambda \in \sigma(T)\}.$$

Proof. Let p be any polynomial. By the elementary scalar remainder theorem, $x - \lambda$ is a factor of $p(x) - p(\lambda)$, that is, $p(x) - p(\lambda) = (x - \lambda)q(x)$ for some polynomial q. Therefore,

$$p(T) - p(\lambda)I = (T - \lambda I)q(T).$$

Now if $p(\lambda) \notin \sigma(p(T))$ then $p(T) - p(\lambda)I$ has an inverse X and so

$$I = X.[p(T) - p(\lambda)I] = X.q(T).(T - \lambda I) = [p(T) - p(\lambda)I].X = (T - \lambda I).q(T).X.$$

Therefore $(T - \lambda I)$ has both a left inverse (namely $X.q(T)$) and a right inverse $(q(T).X)$. An easy algebraic argument shows that these are equal and $(T - \lambda I)$ has an inverse, that is, $\lambda \notin \sigma(T)$. That is, if $\lambda \in \sigma(T)$ then $p(\lambda) \in \sigma(p(T))$.

Conversely, if $k \in \sigma(p(T))$ then the polynomial $p(x) - k$ factors over the complex field into linear factors: $p(x) - k = (x - \lambda_1)(x - \lambda_2)(x - \lambda_3)\cdots(x - \lambda_n)$. Where $\lambda_1, \lambda_2, \lambda_3 \cdots \lambda_n$ are the roots of $p(x) = k$. Then

$$p(T) - kI = (T - \lambda_1 I)(T - \lambda_2 I)(T - \lambda_3 I)\cdots(T - \lambda_n I),$$

and these factors clearly commute. If each $T - \lambda_i I$ has an inverse then the product of all these would be an inverse of $p(T) - kI$. Therefore $T - \lambda_i I$ fails to have an inverse for at least one root λ_i of $p(x) = k$. That is $k = p(\lambda_i)$ for some $\lambda_i \in \sigma(T)$. ■

Corollary 8.2 *If A is a selfadjoint operator then, for any polynomial p,*

$$\|p(A)\| = \sup\{|p(\lambda)| : \lambda \in \sigma(A)\}$$

Proof. If p is real, $p(A)$ is selfadjoint. We know from Theorem 5.5 that for selfadjoint operators, the norm equals the spectral radius. Therefore, using the result of the theorem,

$$\|p(A)\| = \sup\{|k| : k \in \sigma(p(A))\} = \sup\{|p(\lambda)| : \lambda \in \sigma(A)\}.$$

For the general case, since $(\bar{p}.p)(A) = \bar{p}(A).p(A) = p(A)^*.p(A)$ we have, using Corollary 5.4

$$\begin{aligned}\|p(A)\|^2 = \|p(A)^*.p(A)\| &= \sup\{|\bar{p}p(\lambda)| : \lambda \in \sigma(A)\} \\ &= \sup\{|p(\lambda)|^2 : \lambda \in \sigma(A)\} \\ &= (\sup\{|p(\lambda)| : \lambda \in \sigma(A)\})^2. \ ■\end{aligned}$$

Definition. The functional calculus. Let $f \in C[m, M]$. From Weierstrass' approximation theorem there is a sequence (p_n) of polynomials converging to f uniformly on $[m, M]$. [If $f = g + ih$ is complex, (p_n) is found by combining sequences approximating g and h.] The operator $f(A)$ is defined by

$$f(A) = \lim_{n\to\infty} p_n(A).$$

Theorem 8.3 *The operator $f(A) \in \mathcal{B}(\mathcal{H})$ and is well defined. The map $f \mapsto f(A)$ is a *-algebra homomorphism of $C[m, M]$ into $\mathcal{B}(\mathcal{H})$ and*

$$\|f(A)\| = \sup\{|f(\lambda)| : \lambda \in \sigma(A)\}\,.$$

Proof. We first show that the above definition determines an operator $f(A)$. Let p_n be a sequence of polynomials converging to f uniformly on $[m, M]$; that is, in the norm of $C[m, M]$. Then Corollary 8.2 shows that

$$\begin{aligned}\|p_n(A) - p_m(A)\| &= \sup\{|p_n(\lambda) - p_m(\lambda)| : \lambda \in \sigma(A)\} \\ &\leq \sup\{|p_n(\lambda) - p_m(\lambda)| : \lambda \in [m, M]\} = \|p_n - p_m\|\,.\end{aligned}$$

Since (p_n) is a Cauchy sequence in $C[m, M]$, it is clear that $(p_n(A))$ is a Cauchy sequence in $\mathcal{B}(\mathcal{H})$ and so is convergent. To show that this defines a unique operator we must show that it is independent of the choice of the sequence. Suppose that (p_n) and (q_n) are two sequence of polynomials converging to f uniformly on $[m, M]$. Write $\lim p_n(A) = X$ and $\lim q_n(A) = Y$. Then

$$\begin{aligned}\|X - Y\| &= \|X - p_n(A) + p_n(A) - q_n(A) + q_n(A) - Y\| \\ &\leq \|X - p_n(A)\| + \|p_n(A) - q_n(A)\| + \|q_n(A) - Y\| \\ &\leq \|X - p_n(A)\| + \|q_n(A) - Y\| + \sup_{m \leq t \leq M} |p_n(t) - q_n(t)|\end{aligned}$$

and, since $\|X - Y\|$ is independent of n and the right hand side $\to 0$ as $n \to \infty$, this implies that $X = Y = f(A)$ and that the limit depends only on the function f.

To demonstrate the *-algebra homomorphism statement, we need to show that,

$$\begin{aligned}(\alpha f + \beta g)(A) &= \alpha f(A) + \beta g(A) \\ (f.g)(A) &= f(A).g(A) \\ \bar{f}(A) &= f(A)^*\,.\end{aligned}$$

These follow easily from the definitions since, if (p_n) and (q_n) are sequences of polynomials converging uniformly on $[m, M]$ to f and g respectively, then

$$(\alpha f + \beta g)(A) = \lim_{n\to\infty} (\alpha p_n + \beta q_n)(A) = \alpha \lim_{n\to\infty} p_n(A) + \beta \lim_{n\to\infty} q_n(A) = \alpha f(A) + \beta g(A)$$

and the other statements are easily proved in the same way.

Finally,

$$\|f(A)\| = \|\lim_{n\to\infty} p_n(A)\| = \lim_{n\to\infty} \|p_n(A)\| = \lim_{n\to\infty} \sup_{\lambda\in\sigma} |p_n(\lambda)| = \sup_{\lambda\in\sigma} |f(\lambda)|\,.$$

■

The extension of the functional calculus. We wish to extend the functional calculus to limits of pointwise monotonically convergent sequences of continuous functions. (The immediate goal is to attach a meaning to $\chi_{[-\infty,t]}(A)$. Note that $\chi_{[-\infty,t]}$ is a real function which is equal to its square. Our hope is that $E_t = \chi_{[-\infty,t]}(A)$ will satisfy $E_t = E_t^2 = E_t^*$; that is, that E_t is a projection.)

We first need a definition some technical results.

Definition. An operator A is said to be positive (in symbols $A \geq 0$) if $\langle Ax, x\rangle \geq 0$ for all $x \in \mathcal{H}$. We write $A \geq B$ to mean $A - B \geq 0$.

Note that, from the polarization identity, Theorem 2.3 and one of the exercises, any positive operator is selfadjoint. Also, it follows from Theorem 5.5 that $A \geq 0$ if and only if $\sigma(A) \subseteq \mathbb{R}^+$.

Lemma 8.4 *Let A be a positive operator. Then*
(i) $|\langle Ax, y\rangle| \leq \langle Ax, x\rangle.\langle Ay, y\rangle$
(ii) $\|Ax\| \leq \|A\|\langle Ax, x\rangle$.

Proof. (i) This is proved in exactly the same way as Theorem 1.1 from the fact that $\langle A(\lambda x + y), (\lambda x + y)\rangle \geq 0$ for all λ.

(ii) The result is obvious if $Ax = 0$. If $Ax \neq 0$, using (i) with $y = Ax$ we have

$$\begin{aligned} \|Ax\|^4 = |\langle Ax, Ax\rangle|^2 &\leq \langle Ax, x\rangle.\langle A^2x, Ax\rangle \\ &\leq \langle Ax, x\rangle.\|A^2x\|.\|Ax\| \\ &\leq \langle Ax, x\rangle.\|A\|.\|Ax\|^2 \end{aligned}$$

and the result follows on dividing by $\|Ax\|^2$. ■

Theorem 8.5 *Let (A_n) be a decreasing sequence of positive operators. Then (A_n) converges pointwise (strongly) to an operator A such that $0 \leq A \leq A_n$ for all n.*

Proof. Note that for each n, using Lemma normsa, $\|A_n\| = sup_{\|x\|=1}\langle A_nx, x\rangle \leq sup_{\|x\|=1}\langle A_1x, x\rangle = \|A_1\|$. The hypothesis shows that, for each $x \in \mathcal{H}$, the sequence $(\langle A_nx, x\rangle)$ is a decreasing sequence of positive real numbers and hence is convergent. If $m > n$ then $A_m - A_n \geq 0$ and from Lemma 8.4 (ii),

$$\begin{aligned} \|(A_m - A_n)x\|^2 &\leq \|A_m - A_n\|.\langle (A_m - A_n)x, x\rangle \\ &\leq 2.\|A_1\|.[\langle A_mx, x\rangle - \langle A_nx, x\rangle] \end{aligned}$$

and this shows that (A_n) is a Cauchy sequence, and so convergent. Call the limit Ax. It is routine to show that A is linear, bounded, selfadjoint and $0 \leq A \leq A_n$ for all n. [E.g. to show that A is selfadjoint, we write

$$\langle Ax, y\rangle = \lim_{n\to\infty} \langle A_nx, y\rangle = \lim_{n\to\infty} \langle x, A_ny\rangle = \langle x, Ay\rangle\,.$$

■

Lemma 8.6 *If the sequences $(A_n), (B_n)$ converge pointwise (strongly) to an operators A, B respectively (and if $\|A_n\|$ is bounded) then $(A_n.B_n) \to AB$ pointwise.*

[Note that pointwise convergence is convergence in a topology on $\mathcal{B}(\mathcal{H})$ called the "strong operator topology"; hence the alternative terminology. Note also that by a theorem of Banach space theory (the Uniform Boundedness Theorem) the condition in brackets is implied by the convergence of (A_n).]

Proof. For any $x \in \mathcal{H}$ we have

$$\|(A_nB_n - AB)x\| = \|(A_nB_n - A_nB + A_nB - AB)x\| \leq \|A_n\|.\|(B_n - B)x\| + \|(A_n - A)Bx\|$$

and the right hand side $\to 0$ from the pointwise convergence of (A_n) and (B_n) [at the points Bx and x respectively]. ■

Definition. The extended functional calculus. Let ϕ be a positive function on $[m, M]$ that is the pointwise limit of a decreasing sequence (f_n) of positive functions $\in C[m, M]$. The operator $\phi(A)$ is defined as the pointwise limit of the sequence $(f_n(A))$.

It is a fact that $\phi(A)$ is well defined (that is, it depends only on ϕ and not on the approximating sequence).

Lemma 8.7 *Let ϕ, ψ be a positive functions on $[m, M]$ that are the pointwise limits of a decreasing sequences of positive functions $\in C[m, M]$. Then*
(i) $\phi(A) + \psi(A) = (\phi + \psi)(A)$.
(ii) $\phi(A).\psi(A) = (\phi.\psi)(A)$.
(iii) *If X commutes with A then X commutes with $\phi(A)$.*

Proof. (i) Let (f_n) and (g_n) be decreasing sequences of functions that are continuous on $[m, M]$ and converge pointwise $[m, M]$ to ϕ and ψ respectively. Then (f_n+g_n) is a decreasing sequence of continuous functions converging pointwise to $\phi + \psi$ and

$$(\phi + \psi)(A) = \lim_{n\to\infty}(f_n + g_n)(A) = \lim_{n\to\infty} f_n(A) + \lim_{n\to\infty} g_n(A) = \phi(A) + \psi(A)$$

where the limits indicate pointwise convergence in $\mathcal{H}$.

(ii) As in (i) $(f_n.g_n)$ is a decreasing sequence of continuous functions converging pointwise to $\phi.\psi$ and

$$(\phi.\psi)(A) = \lim_{n\to\infty}(f_n.g_n)(A) = \lim_{n\to\infty} f_n(A). \lim_{n\to\infty} g_n(A) = \phi(A).\psi(A),$$

using Lemma 8.6, where the limits indicate pointwise convergence in $\mathcal{H}$.

(iii) If X commutes with A then X commutes with $f(A)$ for every $f \in C[m, M]$ since if (p_n) is a sequence of polynomials converging uniformly to f on $[m, M]$,

$$X.f(A) = \lim_{n\to\infty} X.p_n(A) = \lim_{n\to\infty} p_n(A).X = f(A).X .$$

Now if (f_n) is a decreasing sequences of functions in $C[m, M]$ that converge pointwise $[m, M]$ to ϕ then, using Lemma 8.6,

$$X.\phi(A) = \lim_{n\to\infty} X.f_n(A) = \lim_{n\to\infty} f_n(A).X = \phi(A).X . \quad \blacksquare$$

For every real λ we define the operator E_λ as follows: let

$$f_{n,\lambda} = \begin{cases} 1 & t < \lambda \\ 1 - n(t - \lambda) & \lambda \le t \le \lambda + \frac{1}{n} \\ 0 & t > \lambda + \frac{1}{n} \end{cases}$$

Then $(f_{n,\lambda})$ is is a decreasing sequence of continuous functions converging pointwise to $\chi_{[-\infty,\lambda]}$. Now let $E_\lambda = \chi_{[-\infty,\lambda]}(A)$. If $\lambda < m$ then for all $t \in [m, M]$ we have $f_{n,\lambda}(t) = 0$ for all sufficiently large n and so $E_\lambda = 0$. Similarly, $E_\lambda = I$ for $\lambda \ge M$.

Note that

$$E_\lambda.E_\mu = E_\nu \text{ where } \nu = \min[\lambda, \mu] .$$

It follows that $E_\lambda = E_\lambda^2$. Also $E_\lambda \ge 0$ and so $E_\lambda = E_\lambda^*$.

A family of projections with these properties is called a **bounded resolution of the identity**.

We call the family $\{E_\lambda : -\infty < \lambda < \infty\}$ as obtained above the bounded resolution of the identity for the operator A. It is fact that it is (essentially) uniquely determined by A.

We say that the integral

$$\int_m^M f(\lambda)\, dE_\lambda$$

of a function with respect to a bounded resolution of the identity exists and is equal to T if, given any $\epsilon > 0$ there exists a partition

$$\lambda_0 < \lambda_1 < \lambda_2 < \cdots \lambda_{n-1} < \lambda_n$$

with $\lambda_0 < m$ and $\lambda_n > M$ such that, for any $\xi \in (\lambda_{i-1}, \lambda_i]$,

$$\left\| T - \sum_{i=1}^{n} f(\xi_i)(E_{\lambda_i} - E_{\lambda_{i-1}}) \right\| < \epsilon .$$

For the proof of the Spectral Theorem, we need the following lemmas.

Lemma 8.8 *Let $\{\Delta_i : 1 \le i \le n\}$ be orthogonal projections such that $I = \sum_{i=0}^{n} \Delta_i$ and $\Delta_i.\Delta_j = 0$ when $i \neq j$. If $X \in \mathcal{B}(\mathcal{H})$ commutes with each Δ_i then*

$$\|X\| = \max_{1 \le i \le n} \|\Delta_i X \Delta_i\| .$$

Proof. Let $h \in \mathcal{H}$. Then

$$\|h\|^2 = \|\sum_{i=0}^{n} \Delta_i h\|^2 = \left\langle \sum_{i=0}^{n} \Delta_i h, \sum_{j=0}^{n} \Delta_j h \right\rangle = \sum_{i=0}^{n} \|\Delta_i h\|^2$$

since the cross terms are zero. Therefore, since $X\Delta_i = X\Delta_i^2 = \Delta_i X \Delta_i = \Delta_i X$

$$\begin{aligned} \|Xh\|^2 &= \sum_{i=0}^{n} \|\Delta_i X h\|^2 = \sum_{i=0}^{n} \|\Delta_i X \Delta_i . \Delta_i h\|^2 \\ &\le \sum_{i=0}^{n} \|\Delta_i X \Delta_i\|^2 . \|\Delta_i h\|^2 \\ &\le \max_{1 \le i \le n} \|\Delta_i X \Delta_i\|^2 . \sum_{i=0}^{n} \|\Delta_i h\|^2 \\ &= \max_{1 \le i \le n} \|\Delta_i X \Delta_i\|^2 . \|h\|^2 . \end{aligned}$$

Thus $\|X\| \le \max_{1 \le i \le n} \|\Delta_i X \Delta_i\|$. But the opposite inequality is clear, since for each i we have $\|\Delta_i X \Delta_i\| \le \|\Delta_i\|.\|X\|.\|\Delta_i\| = \|X\|$. ■

Corollary 8.9

$$\left\| \int_m^M f(\lambda)\, dE_\lambda \right\| \le \sup_{m \le t \le M} |f(t)| .$$

Proof. From the lemma,

$$\left\| \sum_{i=1}^{n} f(\xi_i)(E_{\lambda_i} - E_{\lambda_{i-1}}) \right\| = \max_{1 \le i \le n} \|f(\xi_i)(E_{\lambda_i} - E_{\lambda_{i-1}})\| = \max_{1 \le i \le n} |f(\xi_i)| \le \sup_{m \le t \le M} |f(t)| .$$

Since the integral is approximated arbitrarily closely in norm by these sums, the result follows. ■

Lemma 8.10 *For $\lambda > \mu$,*

$$\mu(E_\lambda - E_\mu) \le (E_\lambda - E_\mu) A (E_\lambda - E_\mu) \le \lambda(E_\lambda - E_\mu) .$$

Proof. First note the general fact that if $S \leq T$ then for any operator X we have $X^*SX \leq X^*TX$. This is clear since for any $h \in \mathcal{H}$,

$$\langle X^*SXh, h\rangle = \langle SXh, Xh\rangle \leq \langle TXh, Xh\rangle = \langle X^*TXh, h\rangle\,.$$

Next we claim that $AE_\lambda \leq \lambda E_\lambda$. Let $f_{\lambda,n}$ be as in the definition of E_λ and define $g_n(t) = (\lambda - t)f_{\lambda,n} + \frac{1}{n}$. It is easy to see that $g_n \geq 0$. [This is obvious when $t \leq \lambda$, and also when $t \geq \lambda + \frac{1}{n}$, for then $f_{\lambda,n}(t) = 0$; for $t \in (\lambda, \lambda + \frac{1}{n})$ we have $\lambda - t > -\frac{1}{n}$ so, since $0 \leq f_{\lambda,n}(t) \leq 1$, it follows that $g_n(t) \geq 0$.] Also (g_n) is decreasing (this is an elementary but tedious exercise) and so the pointwise (strong) limit of $(g_n(A))$ exists and is positive. But $(g_n(t))$ is pointwise convergent to $\lambda\chi(-\infty, \lambda](t) - t\chi(-\infty, \lambda](t)$. Therefore

$$\lambda E_\lambda - AE_\lambda \geq 0$$

proving the claim.

Note that E_λ commutes with A and $AE_\lambda^2 = AE_\lambda$. Therefore we have that

$$E_\lambda AE_\lambda = AE_\lambda^2 \leq \lambda E_\lambda\,.$$

We now use that $(I - E_\mu) = (I - E_\mu)^*$ and the general fact from the start of the proof to conclude that

$$(E_\lambda - E_\mu)A(E_\lambda - E_\mu) = (I - E_\mu)E_\lambda AE_\lambda(I - E_\mu) \leq (I - E_\mu)\lambda E_\lambda(I - E_\mu) = \lambda(E_\lambda - E_\mu)\,.$$

The fact that $\mu(E_\lambda - E_\mu) \leq (E_\lambda - E_\mu)A(E_\lambda - E_\mu)$ is proved in an exactly similar way. ■

Theorem 8.11 (The Spectral Theorem for bounded selfadjoint operators.) *For any bounded selfadjoint operator A there exists a bounded resolution of the identity E_λ such that*

$$A = \int_m^M \lambda\, dE_\lambda\,.$$

Proof. Let E_λ be the resolution of the identity as found above. Let $\epsilon > 0$ be given. Choose $\lambda_0 < \lambda_1 < \lambda_2 < \cdots \lambda_n$ with $\lambda_0 < m,\ \lambda_n > M$ such that $0 \leq \lambda_i - \lambda_{i-1} < \epsilon$. Write $\Delta_i = E_{\lambda_i} - E_{\lambda_{i-1}}$. From Lemma 8.10 we have that

$$\lambda_{i-1}\Delta_i \leq \Delta_i A\Delta_i \leq \lambda_i\Delta_i\,.$$

Hence, for any $\xi_i \in [\lambda_{i-1}, \lambda_i]$,

$$(\lambda_{i-1} - \xi_i)\Delta_i \leq \Delta_i A\Delta_i - \xi_i\Delta_i \leq (\lambda_i - \xi_i)\Delta_i\,.$$

Note that when $\|h\| = 1$, since Δ_i is an orthogonal projection, $0 \leq \langle\Delta_i h, h\rangle = \|\Delta_i h\|^2 \leq 1$. Therefore, using Theorem 5.3

$$\|\Delta_i A\Delta_i - \xi_i\Delta_i\| \leq \max[|\lambda_i - \xi_i|, |\lambda_{i-1} - \xi_i|] < \epsilon\,.$$

Observe that $\{\Delta_i : 1 \leq i \leq n\}$ satisfies the hypotheses of Lemma 8.8 and that $X = (A - \sum_{i=1}^n \xi_i\Delta_i)$ commutes with each Δ_i. Therefore, applying Lemma 8.8 we have that

$$\|(A - \sum_{i=1}^n \xi_i\Delta_i)\| = \max_{1\leq i\leq n} \|\Delta_i A\Delta_i - \xi_i\Delta_i\| < \epsilon$$

and this is exactly what is required. ■

$\Delta_i.\Delta_j = 0$ when $i \neq j$.

Corollary 8.12 *If f is continuous on $[m, M]$ then*

$$f(A) = \int_m^M f(\lambda)\, dE_\lambda \,.$$

Proof. For any integer k, choose $n, \xi_1, \xi_2 \cdots \xi_n, \lambda_0, \lambda_1, \ldots, \lambda_n$ as in the Theorem so that

$$\left\| (A - \sum_{i=1}^{n} \xi_i \Delta_i) \right\| < \frac{1}{k}$$

and write $\mathcal{I}_k = \sum_{i=1}^{n} \xi_i \Delta_i$. Then $\lim_{k\to\infty} \mathcal{I}_k = A$. Therefore, for any integer r,

$$A^r = \lim_{k\to\infty} \mathcal{I}_k^r \,.$$

But

$$\mathcal{I}_k^r = \left(\sum_{i=1}^{n} \xi_i \Delta_i \right)^r = \sum_{i=1}^{n} \xi_i^r \Delta_i \,,$$

and the right hand side is the approximating sum to $\int_m^M \lambda^r \, dE_\lambda$. Therefore

$$A^r = \int_m^M \lambda^r \, dE_\lambda \,.$$

and, by taking linear combinations,

$$p(A) = \int_m^M p(\lambda)\, dE_\lambda$$

for all polynomials p.

For $f \in C[m, M]$, given any $\epsilon > 0$, choose a polynomial p_ϵ such that

$$\sup_{m \le \lambda \le M} |f(\lambda) - p(\lambda)| < \epsilon \,.$$

Then, from Theorem 8.3 $\|f(A) - p(A)\| < \epsilon$ and

$$\begin{aligned}
\left\| f(A) - \int_m^M f(\lambda)\, dE_\lambda \right\| &\le \left\| f(A) - \int_m^M p(\lambda)\, dE_\lambda \right\| + \left\| \int_m^M p(\lambda)\, dE_\lambda - \int_m^M f(\lambda)\, dE_\lambda \right\| \\
&= \|f(A) - p(A)\| + \left\| \int_m^M p(\lambda) - f(\lambda)\, dE_\lambda \right\| \\
&< 2\epsilon \,,
\end{aligned}$$

the last step using Corollary 8.9. Since ϵ is arbitrary, it follows that

$$f(A) = \int_m^M f(\lambda)\, dE_\lambda \,.$$

■

Exercises 8

1. (More spectral mapping results.) If $X \in \mathcal{B}(\mathcal{H})$ show that

(i) $\sigma(X^*) = \{\lambda : \bar{\lambda} \in \sigma(X)\}$,

(ii) if X is invertible then $\sigma(X^{-1}) = \{\lambda^{-1} : \lambda \in \sigma(X)\}$.

Deduce that every member of the spectrum of a unitary operator has modulus 1.

2. For any selfadjoint operator A, prove that $\ker A = (\operatorname{ran} A)^{\perp}$.

3. Let A and B be selfadjoint operators. Show that if there exists an invertible operator T such that $T^{-1}AT = B$ then there exists a unitary operator U such that $U^*AU = B$ (that is, if A and B are similar then they are unitarily equivalent).

[Hint: use the polar decomposition of T.]

4. Suppose $X \in \mathcal{B}(\mathcal{H})$ is selfadjoint and $\| X \| \leq 1$. Observe that $X + i\sqrt{I - X^2}$ can be defined and prove that it is unitary. Deduce that any operator T can be written as a linear combination of at most 4 unitary operators. [First write $T = X + iY$ with X, Y selfadjoint.]

5. (i) Use results on the spectrum show that $A \geq kI$ (where $k \in \mathbb{R}$) if and only if for all $\lambda \in \sigma(A)$, $\lambda \geq k$. [Note that A is selfadjoint, since $\langle Ax, x\rangle$ is real – make sure you know how this follows!] Deduce that if $A \geq I$ then $A^n \geq I$ for every positive integer n. [Alternatively factorise $A^n - I$.]

(ii) If B and C commute and $B \geq C \geq 0$ then prove that $B^n \geq C^n$ for every positive integer n. [Factorise $B^n - C^n$.]

6. Let U be both selfadjoint and unitary. Prove that $\sigma(U) = \{-1, 1\}$ (unless $U = \pm I$). [Use Question 1.] Use the spectral theorem to find an orthogonal projection E such that $U = 2E - I$; (alternatively, if you are given the result it is trivial to verify that $E = \frac{1}{2}(U + I)$ is a suitable E). Note that a self adjoint isometry must be unitary [use Question 2]. Deduce that the only positive isometry is I.

[Definition: V is an isometry if $\|Vh\| = \|h\|$ for all $h \in \mathcal{H}$.]