

trim

**TEXTS AND READINGS
IN MATHEMATICS**

50

**Notes on
Functional Analysis**

Rajendra Bhatia

HAR HINDUSTAN
BOOK AGENCY

Notes on Functional Analysis

Rajendra Bhatia

Indian Statistical Institute

Delhi

**HINDUSTAN
BOOK AGENCY**

Published in India by

Hindustan Book Agency (India)
P 19 Green Park Extension
New Delhi 110 016
India

email: info@hindbook.com
<http://www.hindbook.com>

Copyright © 2009, Hindustan Book Agency (India)

No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission from the copyright owner, who has also the sole right to grant licences for translation into other languages and publication thereof.

All export rights for this edition vest exclusively with Hindustan Book Agency (India). Unauthorized export is a violation of Copyright Law and is subject to legal action.

ISBN 978-81-85931-89-0

Preface

These notes are a record of a one semester course on Functional Analysis that I have given a few times to the second year students in the Master of Statistics program at the Indian Statistical Institute, Delhi.

I first taught this course in 1987 to an exceptionally well prepared batch of five students, three of whom have gone on to become very successful mathematicians. Ten years after the course one of them suggested that my lecture notes could be useful for others. I had just finished writing a book in 1996 and was loathe to begin another soon afterwards. I decided instead to prepare an almost verbatim record of what I said in the class the next time I taught the course. This was easier thought than done. The notes written in parts over three different years of teaching were finally ready in 2004.

This background should explain the somewhat unusual format of the book. Unlike the typical text it is not divided into chapters and sections, and it is neither self-contained nor comprehensive. The division is into lectures each corresponding to a 90 minutes class room session. Each is broken into small units that are numbered.

Prerequisites for this course are a good knowledge of Linear Algebra, Real Analysis, Lebesgue Integrals, Metric Spaces, and the rudiments of Set Topology. Traditionally, all these topics are taught before Functional Analysis, and they are used here without much ado. While all major ideas are explained in full, several smaller details are left as exercises. In addition there are other exercises of varying difficulty, and all students are encouraged to do as many of them as they can.

The book can be used by hard working students to learn the basics of Functional Analysis, and by teachers who may find the division into lectures helpful in planning

their courses. It could also be used for training and refresher courses for Ph.D. students and college teachers.

The contents of the course are fairly standard; the novelties, if any, lurk in the details. The course begins with the definition and examples of a Banach space and ends with the spectral theorem for bounded self-adjoint operators in a Hilbert space. Concrete examples and connections with classical analysis are emphasized where possible. Of necessity many interesting topics are left out.

There are two persons to whom I owe special thanks. The course follows, in spirit but not in detail, the one I took as a student from K. R. Parthasarathy. In addition I have tried to follow his injunction that each lecture should contain (at least) one major idea. Ajit Iqbal Singh read the notes with her usual diligence and pointed out many errors, inconsistencies, gaps and loose statements in the draft version. I am much obliged for her help. Takashi Sano read parts of the notes and made useful suggestions. I will be most obliged to alert readers for bringing the remaining errors to my notice so that a revised edition could be better.

The notes have been set into type by Anil Shukla with competence and care and I thank him for the effort.

A word about notation

To begin with I talk of real or complex vector spaces. Very soon, no mention is made of the field. When this happens, assume that the space is complex. Likewise I start with normed linear spaces and then come to Banach spaces. If no mention is made of this, assume that X stands for a complete normed linear space.

I do not explicitly mention that a set has to be nonempty or a vector space nonzero for certain statements to be meaningful. Bounded linear functionals, after some time are called linear functionals, and then just functionals. The same happens to bounded linear operators.

A sequence is written as $\{x_n\}$ or simply as “the sequence x_n ”.

Whenever a general measure space is mentioned, it is assumed to be σ -finite.

The symbol $\overline{E}$ is used for two different purposes. It could mean the closure of the subset E of a topological space, or the complex conjugate of a subset E of the complex plane. This is always clear from the context, and there does not seem any need to discard either of the two common usages.

There are twenty six Lectures in this book. Each of these has small parts with numbers. These are called Sections. A reference such as “Section m ” means the section numbered m in the same Lecture. Sections in other lectures are referred to as “Section m in Lecture n ”. An equation number $(m.n)$ means the equation numbered n in Lecture m .

Do I contradict myself?

Very well then I contradict myself

(I am large, I contain multitudes)

—Walt Whitman

Contents

Lecture	1	Banach Spaces	1
Lecture	2	Dimensionality	11
Lecture	3	New Banach Spaces from Old	19
Lecture	4	The Hahn-Banach Theorem	28
Lecture	5	The Uniform Boundedness Principle	36
Lecture	6	The Open Mapping Theorem	42
Lecture	7	Dual Spaces	49
Lecture	8	Some Applications	58
Lecture	9	The Weak Topology	66
Lecture	10	The Second Dual and the Weak* Topology	73
Lecture	11	Hilbert Spaces	81
Lecture	12	Orthonormal Bases	93
Lecture	13	Linear Operators	103
Lecture	14	Adjoint Operators	111
Lecture	15	Some Special Operators in Hilbert Space	119
Lecture	16	The Resolvent and The Spectrum	129
Lecture	17	Subdivision of the Spectrum	139
Lecture	18	Spectra of Normal Operators	146
Lecture	19	Square Roots and the Polar Decomposition	155
Lecture	20	Compact Operators	163
Lecture	21	The Spectrum of a Compact Operator	170
Lecture	22	Compact Operators and Invariant Subspaces	178
Lecture	23	Trace Ideals	187
Lecture	24	The Spectral Theorem -I	198
Lecture	25	The Spectral Theorem -II	209
Lecture	26	The Spectral Theorem -III	219
Index			230

Banach Spaces

The subject Functional Analysis was created at the beginning of the twentieth century to provide a unified framework for the study of problems that involve *continuity* and *linearity*. The basic objects of study in this subject are Banach spaces and linear operators on these spaces.

1. Let X be a vector space over the field $\mathbb{F}$, where $\mathbb{F}$ is either the field $\mathbb{R}$ of real numbers or the field $\mathbb{C}$ of complex numbers. A *norm* $\| \cdot \|$ on X is a function that assigns to each element of X a nonnegative real value, and has the following properties:

- (i) $\|x\| = 0$ if, and only if, $x = 0$.
- (ii) $\|\alpha x\| = |\alpha| \|x\|$, for all $\alpha \in \mathbb{F}, x \in X$.
- (iii) $\|x + y\| \leq \|x\| + \|y\|$, for all $x, y \in X$.

Property (iii) is called the *triangle inequality*.

A vector space equipped with a norm is called a *normed vector space* (or a *normed linear space*).

From the norm arises a *metric* on X given by $d(x, y) = \|x - y\|$. If the metric space (X, d) is complete, we say that X is a *Banach space*. (Stefan Banach was a Polish mathematician, who in 1932 wrote the book *Théorie des Opérations Linéaires*, the first book on Functional Analysis.)

It follows from the triangle inequality that

$$| \|x\| - \|y\| | \leq \|x - y\|.$$

This shows that the norm is a continuous function on X .

Examples Aplenty

2. The absolute value $|\cdot|$ is a norm on the space $\mathbb{F}$, and with this $\mathbb{F}$ is a Banach space.

3. The *Euclidean space* $\mathbb{F}^n$ is the space of n -vectors $x = (x_1, \dots, x_n)$ with the norm

$$\|x\|_2 := \left(\sum_{j=1}^n |x_j|^2 \right)^{1/2}.$$

4. For each real number $p, 1 \leq p < \infty$ the space ℓ_p^n is the space $\mathbb{F}^n$ with the p -norm of a vector $x = (x_1, \dots, x_n)$ defined as

$$\|x\|_p = \left(\sum_{j=1}^n |x_j|^p \right)^{\frac{1}{p}}.$$

The ∞ -norm of x is defined as

$$\|x\|_\infty = \max_{1 \leq j \leq n} |x_j|.$$

It is easy to see that $\|x\|_p$ is a norm in the special cases $p = 1, \infty$. For other values of p , the proof goes as follows.

(i) For each $1 \leq p \leq \infty$, its *conjugate index* (the *Hölder conjugate*) is the index q that satisfies the equation

$$\frac{1}{p} + \frac{1}{q} = 1.$$

If $1 < p < \infty$, and $a, b \geq 0$, then

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}. \quad (1.1)$$

This is called the *generalised arithmetic-geometric mean inequality* or *Young's inequality*. (When $p = 2$, this is the arithmetic-geometric mean inequality.)

(ii) Given two vectors x and y , let xy be the vector with coordinates $(x_1y_1, \dots, x_ny_n)$.

Use (1.1) to prove the *Hölder inequality*

$$\|xy\|_1 \leq \|x\|_p \|y\|_q, \quad (1.2)$$

for all $1 \leq p \leq \infty$. When $p = 2$, this is the more familiar *Cauchy-Schwarz inequality*.

(iii) Use (1.2) to prove the *Minkowski inequality*

$$\|x + y\|_p \leq \|x\|_p + \|y\|_p. \quad (1.3)$$

5. The justification for the symbol $\|\cdot\|_\infty$ is the fact

$$\lim_{p \rightarrow \infty} \|x\|_p = \|x\|_\infty.$$

6. Why did we restrict ourselves to $p \geq 1$? Let $0 < p < 1$ and take the same definition of $\|\cdot\|_p$ as above. Find two vectors x and y in $\mathbb{F}^2$ for which the triangle inequality is violated.

7. A slight modification of Example 4 is the following. Let $\alpha_j, 1 \leq j \leq n$ be given positive numbers. Then, for each $1 \leq p < \infty$,

$$\|x\| := \left(\sum \alpha_j |x_j|^p \right)^{1/p},$$

is a norm.

All the spaces in the examples above are finite-dimensional and are Banach spaces when equipped with the norms we have defined.

8. Let $C[0, 1]$ be the space of (real or complex valued) continuous functions on the interval $[0, 1]$. Let

$$\|f\| = \sup_{0 \leq t \leq 1} |f(t)|.$$

Then $C[0, 1]$ is a Banach space.

The space consisting of all polynomial functions (of all degrees) is a subspace of $C[0, 1]$. This subspace is not complete. Its completion is the space $C[0, 1]$.

9. More generally, let X be any compact metric space, and let $C(X)$ be the space of (real or complex valued) continuous functions on X . Let

$$\|f\| := \sup_{x \in X} |f(x)|.$$

It is clear that this defines a norm. The completeness of $C(X)$ is proved by a typical use of epsilonics. This argument is called the $\varepsilon/3$ argument.

Let f_n be a Cauchy sequence in $C(X)$. Then for every $\varepsilon > 0$ there exists an integer N such that for $m, n \geq N$ and for all x

$$|f_n(x) - f_m(x)| \leq \varepsilon.$$

So, for every x , the sequence $f_n(x)$ converges to a limit (in $\mathbb{F}$) which we may call $f(x)$. In the inequality above let $m \rightarrow \infty$. This gives

$$|f_n(x) - f(x)| \leq \varepsilon$$

for $n \geq N$ and for all x . In other words, the sequence f_n converges uniformly to f . We now show that f is continuous. Let x be any point in X and let ε be any positive number. Choose N such that $|f_N(z) - f(z)| \leq \varepsilon/3$ for all $z \in X$. Since f_N is continuous at x , there exists δ such that $|f_N(x) - f_N(y)| \leq \varepsilon/3$ whenever $d(x, y) \leq \delta$. Hence, if $d(x, y) \leq \delta$, then

$$|f(x) - f(y)| \leq |f(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f(y)|.$$

Each of the three terms on the right hand side of this inequality is bounded by $\varepsilon/3$. Thus $|f(x) - f(y)| \leq \varepsilon$, and f is continuous at x .

10. For each natural number r , let $C^r[0, 1]$ be the space of all functions that have continuous derivatives $f^{(1)}, f^{(2)}, \dots, f^{(r)}$ of order upto r . (As usual, the derivatives are one-sided limits at the endpoints 0 and 1.) Let

$$\|f\| := \sum_{j=0}^r \sup_{0 \leq t \leq 1} |f^{(j)}(t)|.$$

The space $C^r[0, 1]$ is a Banach space with this norm. (Recall that if the sequences f_n and f'_n converge uniformly on $[0, 1]$ to f, g respectively, then f is differentiable and $f' = g$.)

11. Now let X be any metric space, not necessarily compact, and let $C(X)$ be the

space of *bounded* continuous functions on X . Let

$$\|f\| := \sup_{x \in X} |f(x)|.$$

Then $C(X)$ is a Banach space.

Sequence Spaces

12. An interesting special case of Example 11 is obtained by choosing $X = \mathbb{N}$, the set of natural numbers. The resulting space is then the space of *bounded sequences*. This is the space ℓ_∞ ; if $x = (x_1, x_2, \dots)$ is an element of this space then its norm is

$$\|x\|_\infty := \sup_{1 \leq j < \infty} |x_j|.$$

13. Let c be the subspace of ℓ_∞ that consists of all convergent sequences. Use an $\varepsilon/3$ argument to show that it is a closed subspace of ℓ_∞ .

Let c_0 be the collection of all sequences converging to 0. This is also a closed linear subspace of ℓ_∞ .

We use the symbol c_{00} to denote the collection of all sequences whose terms are zero after some stage. This is a linear subspace of ℓ_∞ , but is not closed. The space c_0 is the *completion* of c_{00} (the smallest closed space in ℓ_∞ that contains c_{00}).

14. For each real number $1 \leq p < \infty$, let ℓ_p be the collection of all sequences $x = (x_1, x_2, \dots)$ such that $\sum_{j=1}^{\infty} |x_j|^p < \infty$.

(i) Use the convexity of the function $f(t) = t^p$ on $[0, \infty)$ to show that ℓ_p is a vector space.

(ii) Note that $\ell_p \subset c_0 \subset \ell_\infty$.

(iii) The inclusions in (ii) are proper. (Consider the sequence with terms $x_n = \frac{1}{\log n}$.)

(iv) The space ℓ_p for any $1 \leq p < \infty$ is not closed in ℓ_∞ .

(v) For $x \in \ell_p$, define

$$\|x\|_p := \left(\sum_{j=1}^{\infty} |x_j|^p \right)^{1/p}.$$

Show that this is a norm. Imitate the steps in Example 4. Some modifications are necessary. The Hölder inequality (1.2) is now the statement: *if $x \in \ell_p$ and $y \in \ell_q$, then their termwise product xy is in ℓ_1 and the inequality (1.2) holds.* With this norm ℓ_p is a Banach space.

(vi) Let $1 \leq p < p' < \infty$. If the series $\Sigma |x_j|^p$ converges, then so does $\Sigma |x_j|^{p'}$. Thus the vector space ℓ_p is contained in $\ell_{p'}$. Further, for every $x \in \ell_p$ we have

$$\|x\|_{p'} \leq \|x\|_p. \quad (1.4)$$

This inequality can be proved as follows. Assume first that $\|x\|_p = 1$. Then $|x_j| \leq 1$ for all j , and hence, $|x_j|^{p'} \leq |x_j|^p$. This shows that

$$\Sigma |x_j|^{p'} \leq \Sigma |x_j|^p = 1,$$

and the inequality (1.4) follows. If x is an arbitrary element of ℓ_p , then let $y = x/\|x\|_p$. Then $\|y\|_p = 1$, and hence, $\|y\|_{p'} \leq \|y\|_p$. This shows (1.4) is true for all $x \in \ell_p$.

Lebesgue Spaces

15. Let I be the interval $[0, 1]$ with the Lebesgue measure μ . Let X be the collection of all bounded measurable functions on I , and for $f \in X$ let

$$\|f\| := \sup_{t \in I} |f(t)|.$$

Then X is a Banach space. (To prove completeness, recall that uniform convergence of a sequence f_n is enough to ensure that the limit f is measurable.)

16. Since sets of measure zero are of no consequence, it is more natural to consider *essentially bounded* functions rather than bounded ones. Let f be a measurable function on I . If there exists an $M > 0$ such that

$$\mu(\{t \in I : |f(t)| > M\}) = 0,$$

we say f is essentially bounded. The infimum of all such M is called the *essential*

supremum of $|f|$, and is written as

$$\|f\|_\infty = \text{ess sup } |f|.$$

The collection of all (equivalence classes of) such functions is the space $L_\infty[0, 1]$. It is a Banach space with this norm.

17. For $1 \leq p < \infty$, let $L_p[0, 1]$ be the collection of all measurable functions on $[0, 1]$ for which $\int_0^1 |f(t)|^p dt$ is finite. Then $L_p[0, 1]$ is a vector space and

$$\|f\|_p := \left(\int_0^1 |f(t)|^p dt \right)^{1/p}$$

is a norm on it. To prove this, one uses versions of Hölder and Minkowski inequalities (1.2) and (1.3) in which sums are replaced by integrals.

The completeness of $L_p[0, 1]$ is standard measure theory. The assertion that $L_p[0, 1]$ is complete is called the *Riesz–Fischer Theorem*. (Warning: There are other theorems going by the same name.)

18. The interval I can be replaced by a general measure space $(X, \mathcal{S}, \mu)$ in which X is a set, $\mathcal{S}$ a σ -algebra of subsets of X , and μ any measure. The spaces $L_p(X, \mathcal{S}, \mu)$, $1 \leq p \leq \infty$, can then be defined in the same way as above. (It is often necessary to put some restrictions like σ -finiteness to prevent unruly behaviour of different sorts.) When $X = \mathbb{N}$, and μ is the counting measure, we get sequence spaces.

If $\mu(X)$ is finite, and $1 \leq p < p' \leq \infty$, then the space $L_{p'}$ is a linear subspace of L_p . In this case we have

$$\|f\|_p \leq \mu(X)^{1/p-1/p'} \|f\|_{p'} \quad (1.5)$$

for all $f \in L_{p'}$. (This can be seen using the Hölder inequality, choosing one of the functions to be identically 1.) This is just the opposite of the behaviour of sequence spaces in Example 14.

If $\mu(X) = \infty$, no inclusion relations of this kind can be asserted in general.

Separable Spaces

A metric space is called *separable* if it has a subset that is countable and dense. Separable Banach spaces are easier to handle than nonseparable ones. So, it is of interest to know which spaces are separable.

19. The space $C[0, 1]$ is separable. Polynomials with rational coefficients are dense in this space.

20. For $1 \leq p < \infty$, the space c_{00} is dense in ℓ_p . Within this space those that have rational entries are dense. So the spaces ℓ_p , $1 \leq p < \infty$ are separable.

21. The space ℓ_∞ is not separable. Consider the set S of sequences whose terms are 0 or 1. Then S is an uncountable subset of ℓ_∞ . (It is uncountable because every point in the unit interval has a binary decimal expansion and thus corresponds to a unique element of S .) If x, y are any two distinct elements of S , then $\|x - y\|_\infty = 1$. So the open balls $B(x, 1/2)$, with radii $1/2$ and centred at points $x \in S$, form an uncountable disjoint collection. Any dense set in ℓ_∞ must have at least one point in each of these balls, and hence can not be countable.

The subspace c_0 of ℓ_∞ is separable (c_{00} is dense in it) as is the subspace c (consider sequences whose terms are constant after some stage).

22. For $1 \leq p < \infty$, the spaces $L_p[0, 1]$ are separable. Continuous functions are dense in each of them. The space $L_\infty[0, 1]$ is not. (Consider the characteristic functions of the intervals $[0, t]$, $0 \leq t \leq 1$).

23. What about the spaces $L_p(X, \mathcal{S}, \mu)$? These can not be “smaller” than the spaces $(X, \mathcal{S}, \mu)$. If we put $d(E, F) = \mu(E \Delta F)$, where $E \Delta F$ is the symmetric difference of the sets E and F , then $d(E, F)$ is a metric on $\mathcal{S}$. It can be proved (with standard but elaborate measure theory) that for $1 \leq p < \infty$, the space $L_p(X, \mathcal{S}, \mu)$ is separable if and only if the metric space $(\mathcal{S}, d)$ is separable. Further, this condition is satisfied if and only if the σ -algebra $\mathcal{S}$ is countably generated. (The statements about ℓ_p and

$L_p[0, 1]$, $1 \leq p < \infty$ are included in this more general set up.)

More examples

24. A function f on $[0, 1]$ is said to be *absolutely continuous* if, given $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\sum_{i=1}^n |f(t'_i) - f(t_i)| < \varepsilon$$

for every finite disjoint collection of intervals $\{(t_i, t'_i)\}$ in $[0, 1]$ with $\sum_{i=1}^n |t'_i - t_i| < \delta$.

The *Fundamental Theorem of Calculus* says that if f is absolutely continuous, then it is differentiable almost everywhere, its derivative f' is in $L_1[0, 1]$, and $f(t) = \int_0^t f'(s)ds + f(0)$ for all $0 \leq t \leq 1$. Conversely, if g is any element of $L_1[0, 1]$, then the function G defined as $G(t) = \int_0^t g(s)ds$ is absolutely continuous, and then G' is equal to g almost everywhere.

For each natural number r , let $L_p^{(r)}[0, 1]$ be the collection of all $(r - 1)$ times continuously differentiable functions f on $[0, 1]$ with the properties that $f^{(r-1)}$ is absolutely continuous and $f^{(r)}$ belongs to $L_p[0, 1]$. For f in this space define

$$\|f\| := \|f\|_p + \|f^{(1)}\|_p + \cdots + \|f^{(r)}\|_p.$$

Then $L_p^{(r)}[0, 1]$, $1 \leq p < \infty$ is a Banach space. (The proof is standard measure theory.) These are called *Sobolev spaces* and are used often in the study of differential equations.

25. Let D be the unit disk in the complex plane and let X be the collection of all functions analytic on D and continuous on its closure $\bar{D}$. For f in X , let

$$\|f\| := \sup_{z \in \bar{D}} |f(z)|.$$

Then X is a Banach space with this norm. (The uniform limit of analytic functions is analytic. Use the theorems of Cauchy and Morera.)

Caveat

We have now many examples of Banach spaces. We will see some more in the course. Two remarks must be made here.

There are important and useful spaces in analysis that are vector spaces and have a natural topology on them that does not arise from any norm. These are *topological vector spaces* that are not normed spaces. The spaces of *distributions* used in the study of differential equations are examples of such spaces.

All the examples that we gave are not hard to describe and come from familiar contexts. There are Banach spaces with norms that are defined inductively and are not easy to describe. These Banach spaces are sources of counterexamples to many assertions that seem plausible and reasonable. There has been a lot of research on these exotic Banach spaces in recent decades.

Dimensionality

Algebraic (Hamel)Basis

1. Let X be a vector space and let S be a subset of it. We say S is *linearly independent* if for every finite subset $\{x_1, \dots, x_n\}$ of S , the equation

$$a_1x_1 + \dots + a_nx_n = 0 \tag{2.1}$$

holds if and only if $a_1 = a_2 = \dots = a_n = 0$. A (finite) sum like the one in (2.1) is called a *linear combination* of $x_1, \dots, x_n$.

Infinite sums have a meaning only if we have a notion of convergence in X .

2. A linearly independent subset B of a vector space X is called a *basis* for X if every element of X is a linear combination of (a finite number of) elements of B . To distinguish it from another concept introduced later we call this a *Hamel basis* or an *algebraic basis*.

Every (nonzero) vector space has an algebraic basis. This is proved using Zorn's Lemma. We will use this Lemma often.

Zorn's Lemma

3. Let X be any set. A binary relation $\leq$ on X is called a *partial order* if it satisfies three conditions

- (i) $x \leq x$ for all $x \in X$, (*reflexivity*)

- (ii) if $x \leq y$ and $y \leq x$, then $x = y$, (*antisymmetry*)
- (iii) if $x \leq y$ and $y \leq z$, then $x \leq z$. (*transitivity*)

A set X with a partial order is called a *partially ordered set*.

The sets $\mathbb{N}, \mathbb{Q}, \mathbb{R}$ (natural numbers, rational numbers, and real numbers) are partially ordered if $x \leq y$ means “ x is less than or equal to y ”. Another partial order on $\mathbb{N}$ can be defined by ordaining that $x \leq y$ means “ x divides y ”. The class of all subsets of a given set is partially ordered if we say $E \leq F$ whenever $E \subseteq F$.

An element x_0 of a partially ordered set X is called a *maximal element* if there is no element bigger than it; i.e., $x_0 \leq x$ if and only if $x = x_0$. Such an element need not exist; and if it does it need not be unique.

Let E be a subset of a partially ordered set X . An element x_0 of X is an *upper bound* for E if $x \leq x_0$ for all $x \in E$. We say E is *bounded above* if an upper bound for E exists.

- A partially ordered set X is *totally ordered* if in addition to the conditions (i) - (iii), the binary relation $\leq$ satisfies a fourth condition:
- (iv) if $x, y \in X$, then either $x \leq y$ or $y \leq x$.

Zorn’s Lemma says:

If X is a partially ordered set in which every totally ordered subset is bounded above, then X contains a maximal element.

This Lemma is logically *equivalent* to the *Axiom of Choice* (in the sense that one can be derived from the other). This axiom says that if $\{X_\alpha\}$ is any family of sets, then there exists a set Y that contains exactly one element from each X_α .

See J.L. Kelley, *General Topology* for a discussion.

- 4. Exercises.** (i) Use Zorn’s Lemma to show that every vector space X has an algebraic basis. (This is a maximal linearly independent subset of X .)

- (ii) Show that any two algebraic bases of X have the same cardinality. This is called the *dimension* of X , written as $\dim X$.
- (iii) If B is an algebraic basis for X then every element of X can be written uniquely as a linear combination of elements of B .
- (iv) Two vector spaces X and Y are isomorphic if and only if $\dim X = \dim Y$.

5. The notion of an algebraic basis is not of much use in studying Banach spaces since it is not related to any topological property. We will see if X is a Banach space, then either $\dim X < \infty$ or $\dim X \geq c$, the cardinality of the continuum. Thus there is no Banach space whose algebraic dimension is countably infinite.

Topological (Schauder) Basis

6. Let $\{x_n\}$ be a sequence of elements of a Banach space X . We say that the series $\sum_{n=1}^{\infty} x_n$ *converges* if the sequence $s_N = \sum_{n=1}^N x_n$ of its partial sums has a limit in X .

7. A sequence $\{x_n\}$ in a Banach space X is a *topological basis* (*Schauder basis*) for X if every element x of X has a unique representation $x = \sum_{n=1}^{\infty} a_n x_n$. Note that the order in which the elements x_n are enumerated is important in this definition.

A Schauder basis is necessarily a linearly independent set.

8. If $\{x_n\}$ is a Schauder basis for a Banach space X , then the collection of all finite sums $\sum_{n=1}^N a_n x_n$, in which a_n are scalars with rational real and imaginary parts, is dense in X . So, X is separable. Thus a nonseparable Banach space can not have a Schauder basis.

For $n = 1, 2, \dots$, let e_n be the vector with all entries zero except an entry 1 in the n th place. Then $\{e_n\}$ is a Schauder basis for each of the spaces ℓ_p , $1 \leq p < \infty$, and for the space c_0 .

9. Is there any obvious Schauder basis for the space $C[0, 1]$ of real functions? The one constructed by Schauder is described below.

Exercise. Let $\{r_i : i \geq 1\}$ be an enumeration of dyadic rationals in $[0, 1]$: $0, 1, \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}, \frac{1}{16}, \frac{3}{16}, \dots$ Let $f_1(t) \equiv 1, f_2(t) = t$; and for $n > 2$ define f_n as follows. Let $f_n(r_j) = 0$ if $j < n, f_n(r_n) = 1$, and let f_n be linear between any two neighbours among the first n dyadic rationals. Draw the graphs of f_3, f_4 and f_5 . Show that every element g of $C[0, 1]$ has a unique representation $g = \sum a_i f_i$:

(i) Note a_1 must be $g(0)$;

(ii) a_2 must be $g(1) - a_1$;

(iii) proceed inductively to see that

$$a_n = g(r_n) - \sum_{i=1}^{n-1} a_i f_i(r_n);$$

(iv) draw the graph of $\sum_{i=1}^n a_i f_i$;

(v) since the sequence r_i is dense in $[0, 1]$, these sums converge uniformly to g , as $n \rightarrow \infty$.

Note that $\|f_n\| = 1$ for all n . Thus we have a *normalised* basis for $C[0, 1]$.

10. Does every separable Banach space have a Schauder basis?

This question turns out to be a difficult one. In 1973, P. Enflo published an example to show that the answer is in the negative. (This kind of problem has turned out to be slippery ground. For example, it is now known that every l_p space with $p \neq 2$ has a *subspace* without a Schauder basis.)

Equivalence of Norms

11. Let $\|\cdot\|$ and $\|\cdot\|'$ be two norms on a vector space X . We say these norms are *equivalent* if there exist positive real numbers C and C' such that

$$\|x\| \leq C\|x\|', \quad \|x\|' \leq C'\|x\|$$

for all x . Clearly this is an equivalence relation between norms.

The metrics arising from equivalent norms are equivalent. Any sequence that converges in the metric induced by a norm also converges in the one induced by an equivalent norm.

We will see that if X is finite dimensional, then all norms on X are equivalent to one another.

12. Let $x_1, \dots, x_n$ be orthonormal vectors in the Euclidean space $\mathbb{C}^n$. Then for all scalars $a_1, \dots, a_n$,

$$\|a_1x_1 + \dots + a_nx_n\|_2^2 = |a_1|^2 + \dots + |a_n|^2. \quad (2.2)$$

The next lemma provides a good working substitute for this. It says that if $x_1, \dots, x_n$ are linearly independent vectors in any Banach space, then the norm of any linear combination $a_1x_1 + \dots + a_nx_n$ can not be too small.

Lemma. Let $\{x_1, \dots, x_n\}$ be linearly independent vectors in any normed linear space X . Then there exists a constant $C > 0$, such that for all scalars $a_1, \dots, a_n$

$$\|a_1x_1 + \dots + a_nx_n\| \geq C(|a_1| + \dots + |a_n|). \quad (2.3)$$

Proof. Divide both sides of the inequality (2.3) by $|a_1| + \dots + |a_n|$. The problem reduces to showing that there exists C , such that if $\sum |a_j| = 1$, then

$$\|a_1x_1 + \dots + a_nx_n\| \geq C.$$

If this were not the case, for each positive integer m there would exist $a_1^{(m)}, \dots, a_n^{(m)}$ with $\sum |a_j^{(m)}| = 1$ such that

$$\|a_1^{(m)}x_1 + \dots + a_n^{(m)}x_n\| < \frac{1}{m}. \quad (2.4)$$

The sequence $(a_1^{(m)}, \dots, a_n^{(m)})$ indexed by m is a bounded sequence in $\mathbb{C}^n$. So, by the Bolzano-Weierstrass Theorem it has a convergent subsequence. The limit of this subsequence is an n -tuple $(a_1, \dots, a_n)$ with $\sum |a_j| = 1$. Since x_j are linearly independent, this means

$$a_1x_1 + \dots + a_nx_n \neq 0.$$

This contradicts (2.4) which says that $a_1^{(m)}x_1 + \dots + a_n^{(m)}x_n$ converges to zero as $m \rightarrow \infty$. ■

13. Theorem. Any two norms on a finite dimensional vector space are equivalent.

Proof. Let $\{x_1, \dots, x_n\}$ be a basis for X . If $x = a_1x_1 + \dots + a_nx_n$, set

$$\|x\|_1 = |a_1| + \dots + |a_n|.$$

This is a norm on X . Let $\|\cdot\|$ be any other norm. By the Lemma in 12, there exists a constant C such that

$$\|x\| \geq C\|x\|_1.$$

On the other hand if $C' = \max \|x_j\|$, then

$$\|x\| \leq \sum_j |a_j| \|x_j\| \leq C' \sum |a_j| = C'\|x\|_1.$$

Thus $\|\cdot\|$ and $\|\cdot\|_1$ are equivalent. ■

14. Exercises. (i) Consider the space $\mathbb{C}^n$ with the p -norms $1 \leq p \leq \infty$. Given two indices p and p' , find the smallest numbers $C_{p,p'}$ such that

$$\|x\|_p \leq C_{p,p'} \|x\|_{p'} \text{ for all } x.$$

- (ii) Find an example of an infinite-dimensional vector space with two inequivalent norms.
- (iii) Show that every finite-dimensional normed linear space is a Banach space.

Local Compactness

15. By the Heine–Borel Theorem, the closed unit ball $\{x : \|x\|_2 \leq 1\}$ is a compact subset of $\mathbb{C}^n$. It follows that any closed ball (with any centre and radius) is compact.

A topological space in which every point has a neighbourhood with compact closure is called *locally compact*. This property is the next best thing to compactness.

We have seen that all norms on a finite-dimensional space are equivalent. So, the Heine–Borel Theorem says that all finite-dimensional normed spaces are locally compact. We will see that no infinite-dimensional space has this property.

16. **F. Riesz's Lemma.** Let M be a proper closed linear subspace of a normed linear space X . Then for each $0 < t < 1$, there exists a unit vector x_t in X such that

$$\text{dist}(x_t, M) \geq t.$$

(The *distance* of a vector x from a subspace M is the number $\text{dist}(x, M) = \inf \{\|x - m\| : m \in M\}$.)

Proof. Choose any vector u not in M , and let $d = \text{dist}(u, M)$. Since M is closed, $d > 0$. For each $0 < t < 1$, $d/t > d$. Hence, by the definition of d , there exists $x_0 \in M$ such that

$$d \leq \|u - x_0\| \leq d/t.$$

If $x_t := \frac{u - x_0}{\|u - x_0\|}$, then for all $x \in M$

$$\begin{aligned} \|x - x_t\| &= \frac{1}{\|u - x_0\|} \|(\|u - x_0\|)x - u + x_0\| \\ &= \frac{1}{\|u - x_0\|} \|u - x_1\|, \end{aligned}$$

where $x_1 = \|u - x_0\|x + x_0$. Note that for each x in M the vector x_1 is also in M .

So, by the definition of d

$$\|x - x_t\| \geq \frac{d}{\|u - x_0\|} \geq t.$$

17. Exercises. (i) If X is finite-dimensional, its unit sphere $S := \{x : \|x\| = 1\}$ is compact. Use this to show that there exists a unit vector x such that $\text{dist}(x, M) = 1$.

(ii) This need not be true if X is infinite-dimensional. Show that the choice

$$\begin{aligned} X &= \{f \in C[0, 1] : f(0) = 0\} \\ M &= \{f \in X : \int_0^1 f = 0\} \end{aligned}$$

provides a counter example.

18. Theorem. In any infinite-dimensional normed linear space the closed unit ball cannot be compact.

Proof. Choose any unit vector x_1 in X and let M_1 be its linear span. By Riesz's Lemma, there exists a unit vector x_2 such that $\text{dist}(x_2, M_1) > 1/2$, and hence, $\|x_2 - x_1\| > 1/2$. Let M_2 be the linear span of x_1 and x_2 . Repeat the argument. This leads to a sequence x_n of unit vectors at distance greater than $1/2$ from each other. So, the unit ball is not compact. ■

Thus a normed linear space is locally compact if and only if it is finite-dimensional. This famous theorem was first proved by F. Riesz.

New Banach Spaces from Old

Quotient Spaces

1. Let X be a vector space and M a subspace of it. Say that two elements x and y of X are *equivalent*, $x \sim y$, if $x - y \in M$. This is an equivalence relation on X . The coset of x under this relation is the set

$$\tilde{x} = x + M := \{x + m : m \in M\}.$$

Let $\tilde{X}$ be the collection of all these cosets. If we set

$$\begin{aligned}\tilde{x} + \tilde{y} &= \widetilde{x + y}, \\ \alpha \tilde{x} &= \widetilde{\alpha x},\end{aligned}$$

then $\tilde{X}$ is a vector space with these operations.

The zero element of $\tilde{X}$ is M . The space $\tilde{X}$ is called the *quotient* of X by M , written as X/M .

If $X = \mathbb{R}^2$, a non-trivial subspace of it is a line through the origin. The space $\tilde{X}$ is then the collection of all lines parallel to this.

2. Let X be a normed linear space and let M be a closed subspace. Let $\tilde{X} = X/M$ and define

$$\|\tilde{x}\| = \text{dist}(x, M) = \inf_{m \in M} \|x - m\|.$$

Then this is a norm on $\tilde{X}$. (To make sure that $\|\tilde{x}\|$ is a norm we need M to be closed.) Note that we can also write

$$\|\tilde{x}\| = \inf_{m \in M} \|x + m\|.$$

We will show that if X is complete, then so is $\tilde{X}$.

3. We say that a sequence x_n in a normed linear space X is *summable* if the series $\sum x_n$ is convergent, and *absolutely summable* if the series $\sum \|x_n\|$ is convergent.

Exercise. A normed linear space is complete if and only if every absolutely summable sequence in it is summable.

4. **Theorem.** Let X be a Banach space and M a closed subspace of it. Then the quotient X/M is also a Banach space.

Proof. Let $\tilde{x}_n$ be an absolutely summable sequence in $\tilde{X}$. We will show that $\tilde{x}_n$ is summable. For each n , choose $m_n \in M$ such that

$$\|x_n - m_n\| \leq \inf_{m \in M} \|x_n - m\| + \frac{1}{2^n} = \|\tilde{x}_n\| + \frac{1}{2^n}.$$

Since $\sum \|\tilde{x}_n\|$ is convergent, the sequence $x_n - m_n$ in X is absolutely summable, and hence summable. Let

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

The coset $\tilde{y}$ is a natural candidate for being the limit of the series $\sum \tilde{x}_n$. Indeed,

$$\begin{aligned} \left\| \sum_{n=1}^N \tilde{x}_n - \tilde{y} \right\| &= \left\| \left(\sum_{n=1}^N x_n - y \right) \sim \right\| \\ &= \inf_{m \in M} \left\| \sum_{n=1}^N x_n - y - m \right\| \\ &\leq \left\| \sum_{n=1}^N x_n - y - \sum_{n=1}^N m_n \right\| \\ &= \left\| \sum_{n=1}^N (x_n - m_n) - y \right\|. \end{aligned}$$

The right hand side goes to zero as $N \rightarrow \infty$. This shows $\tilde{x}_n$ is summable. ■

5. Exercises. (i) Let X be the Euclidean space $\mathbb{C}^n$ and let $M = \mathbb{C}^k$, $1 \leq k < n$. Show that X/M is isomorphic to the orthogonal complement of $\mathbb{C}^k$.

(ii) Let $X = C[0, 1]$ and let $M = \{f : f(0) = 0\}$. Show that X/M is isomorphic to the scalars $\mathbb{C}$. (Identify the scalar z with the constant function with value z everywhere.)

Bounded Linear Operators

6. Let X, Y be vector spaces. A linear map from X to Y is called a *linear operator*.

Suppose X and Y are normed linear spaces. Let

$$M = \sup_{\|x\|=1} \|Ax\|. \quad (3.1)$$

If M is finite, we say that A is a *bounded linear operator*. From the definition of M we see that

$$\|Ax\| \leq M \|x\| \quad \text{for all } x \in X, \quad (3.2)$$

and hence, every bounded linear operator is a continuous map from X into Y . (If $x_n \rightarrow x$ in X , then $Ax_n \rightarrow Ax$ in Y .)

Now suppose A is a continuous linear map from X to Y . By the $\varepsilon - \delta$ definition of continuity, there exists a number $\delta > 0$ such that $\|Ax\| \leq 1$ whenever $\|x\| \leq \delta$. If x is a vector in X , with $\|x\| = 1$, then $\|\delta x\| = \delta$. Hence $\delta \|Ax\| \leq 1$, and $\|Ax\| \leq 1/\delta$. Thus A is bounded.

Thus a linear operator is continuous if and only if it is bounded. If a linear operator is continuous at 0, then it is continuous everywhere.

The set of all bounded linear operators from X to Y is denoted as $\mathcal{B}(X, Y)$. This is a vector space.

7. For A in $\mathcal{B}(X, Y)$ let

$$\|A\| := \sup_{\|x\|=1} \|Ax\|.$$

It is easy to see the following

- (i) $\|Ax\| \leq \|A\| \|x\|$ for all x .
- (ii) $\|A\| = \inf\{M : \|Ax\| \leq M\|x\| \text{ for all } x\}$.
- (iii) $\|A\| = \sup_{\|x\| \leq 1} \|Ax\|$.
- (iv) $\mathcal{B}(X, Y)$ is a normed linear space with this norm.

8. Elements of $\mathcal{B}(X, \mathbb{F})$ are called *bounded linear functionals* on X .

9. Each $m \times n$ matrix gives rise to a bounded linear operator from $\mathbb{C}^n$ into $\mathbb{C}^m$.

Each element u of $\mathbb{C}^n$ gives rise to a linear functional via the map $x \mapsto u \cdot x$, where $u \cdot x$ is the dot product.

Let $X = L_\infty[0, 1]$. Then the map $f \mapsto \int_0^1 f(t)dt$ is a bounded linear functional.

10. Let $X = L_2[0, 1]$. Let $K(x, y)$ be a measurable function on $[0, 1] \times [0, 1]$ such that

$$\int_0^1 \int_0^1 |K(x, y)|^2 dx dy < \infty. \quad (3.3)$$

We say that K is a *square integrable kernel*. For f in X , define Kf as

$$(Kf)(x) = \int_0^1 K(x, y)f(y)dy. \quad (3.4)$$

By Fubini's Theorem, it follows from (3.3) that

$$\int_0^1 |K(x, y)|^2 dy < \infty \text{ a.e. } x,$$

and then by the Schwarz inequality

$$|\int_0^1 K(x, y)f(y)dy|^2 \leq \int_0^1 |K(x, y)|^2 dy \int_0^1 |f(y)|^2 dy < \infty \text{ a.e. } x.$$

Thus Kf is a well defined bounded measurable function and hence is in $L_2[0, 1]$. So, the map $f \mapsto Kf$ is a linear operator on $L_2[0, 1]$. It is easy to see that it is a bounded

linear operator.

Indeed

$$\|Kf\|_2^2 \leq \left[\int_0^1 \int_0^1 |K(x,y)|^2 dx dy \right] \|f\|_2^2.$$

There is nothing special about $[0, 1]$ here. It could be replaced by any bounded or unbounded interval of the real line. The square integrability condition (3.3) is sufficient, but not necessary, for the operator in (3.4) to be bounded.

11. Let $X = C[0, 1]$ and let $K(x, y)$ be a continuous function on $[0, 1] \times [0, 1]$. For $f \in X$, let Kf be a new function defined as

$$(Kf)(x) = \int_0^1 K(x, y)f(y)dy. \quad (3.5)$$

Show that $f \mapsto Kf$ is a bounded linear operator on X .

The condition that $K(x, y)$ is continuous in (x, y) is sufficient, but not necessary, to ensure that the operator K is bounded. For example the operator K in (3.5) is bounded if $\lim_{x_n \rightarrow x} \int_0^1 |K(x_n, y) - K(x, y)|dy = 0$.

The operators K defined in (3.4) and (3.5) are said to be *integral kernel operators* induced by the kernel $K(x, y)$. They are obvious generalisations of operators on finite-dimensional spaces induced by matrices. Many problems in mathematical physics are solved by formulating them as integral equations. Integral kernel operators are of great interest in this context.

12. Let X, Y, Z be normed linear spaces, if $A \in \mathcal{B}(Y, Z)$ and $B \in \mathcal{B}(X, Y)$, then $AB \in \mathcal{B}(X, Z)$ and

$$\|AB\| \leq \|A\| \|B\|. \quad (3.6)$$

The space $\mathcal{B}(X, X)$ is written as $\mathcal{B}(X)$ to save space (and breath). It is a vector space, and two of its elements can be multiplied. The multiplication behaves nicely with respect to addition: $A(B + C) = AB + AC$, and $(A + B)C = AC + BC$.

13. If X is any normed linear space and Y a Banach space, then $\mathcal{B}(X, Y)$ is a Banach

space. To see this note that if A_n is a Cauchy sequence in $\mathcal{B}(X, Y)$ then $A_n x$ is a Cauchy sequence in Y for each $x \in X$. Let $Ax = \lim A_n x$. The operator A is linear. Further

$$\|Ax\| = \lim \|A_n x\| \leq (\sup \|A_n\|)\|x\|.$$

Since A_n is Cauchy, $\sup \|A_n\|$ is finite. Show that $\|A_n - A\| \rightarrow 0$.

14. Let $\mathcal{A}$ be a vector space. We say $\mathcal{A}$ is an *algebra* if there is a rule for multiplication on $\mathcal{A}$ that satisfies four conditions

$$(i) \quad a(bc) = (ab)c,$$

$$(ii) \quad a(b + c) = ab + ac,$$

$$(iii) \quad (a + b)c = ac + bc,$$

$$(iv) \quad \lambda(ab) = (\lambda a)b = a(\lambda b),$$

for all a, b, c in A and for all scalars λ .

If A has a norm that is *submultiplicative*; i.e., $\|ab\| \leq \|a\| \|b\|$, then we say that $\mathcal{A}$ is a *normed algebra*. If $\mathcal{A}$ is complete in the metric induced by this norm, then $\mathcal{A}$ is called a *Banach algebra*.

Our discussion above shows that if X is a Banach space, then $\mathcal{B}(X)$ is a Banach algebra. The study of Banach algebras has been a major theme in Functional Analysis.

15. **Exercise.** Let A be an $n \times n$ matrix. Let X be the space $\mathbb{C}^n$ with the ℓ_p norm, where $p = 1$, or ∞ . Regard A as a linear operator on X and find an expression for its norm in each of the two cases.

16. The space $\mathcal{B}(X, \mathbb{F})$ is called the *dual space* of X . The symbol X^* is conventionally used for it and its elements are called (*bounded*) *linear functionals* on X .

When $\mathbb{F} = \mathbb{C}$, some times it is more convenient to consider *conjugate linear functionals* instead of linear ones: instead of demanding $f(\alpha x) = \alpha f(x)$ we demand $f(\alpha x) = \bar{\alpha} f(x)$. Some times the collection of such functionals is called the dual space. This is just a matter of preference, it makes some formulas slightly simpler and some others a little awkward.

17. We should emphasize that we will use the word functional to mean a bounded linear functional.

If X is any vector space, then the *algebraic dual* of X is the vector space X' consisting of all linear functional on X . If X is finite-dimensional and $e_1, \dots, e_n$ is a basis for it, then the linear functionals $f_1, \dots, f_n$ defined by $f_i(e_j) = \delta_{ij}$ constitute a basis for the dual space X' . Thus X and X' are isomorphic vector spaces.

Let X be an infinite-dimensional vector space with a Hamel basis $\{e_1, e_2, \dots\}$. Now the linear functionals $\{f_1, f_2, \dots\}$ defined as above no longer constitute a Hamel basis for X' . (Consider the linear functional f defined as $f(e_{2n}) = 1$, $f(e_{2n+1}) = 0$, for all n .) In fact the dimension of X' is uncountable. Thus for infinite-dimensional spaces the concept of algebraic dual is not as useful as in finite dimensions. For Banach spaces, the dual X^* (sometimes called the *topological dual*) turns out to be a very useful object.

Two examples

18. The kernel

$$K(x, y) = \frac{1}{\sqrt{2\pi}} e^{-ixy} \quad (3.7)$$

is called the *Fourier kernel*. Let

$$(\mathcal{F}f)(x) := \int_{-\infty}^{\infty} K(x, y) f(y) dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ixy} f(y) dy. \quad (3.8)$$

Then $\mathcal{F}f$ is called the *Fourier transform* of f . $\mathcal{F}$ is a bounded linear map from $L_1(\mathbb{R})$ into $L_{\infty}(\mathbb{R})$. Its norm is $1/\sqrt{2\pi}$. It is a basic fact in Fourier analysis that $\mathcal{F}$ induces a bijective map of $L_2(\mathbb{R})$ onto itself and in this case its norm is 1. Note that the Fourier kernel is *not* square integrable on $\mathbb{R} \times \mathbb{R}$.

19. Let $\mathbb{R}_+ = [0, \infty)$ and consider the kernel

$$K(x, y) := e^{-xy} \quad (3.9)$$

on $\mathbb{R}_+ \times \mathbb{R}_+$. The associated integral operator on $L_2(\mathbb{R}_+)$ is called the *Laplace transform* $\mathcal{L}$; i.e.,

$$(\mathcal{L}f)(x) = \int_0^{\infty} e^{-xy} f(y) dy. \quad (3.10)$$

Like the Fourier kernel, the kernel (3.9) too is not square integrable. Like the Fourier transform, the Laplace transform is a bounded operator on $L_2(\mathbb{R}_+)$ and is bijective. Let us calculate the norm of this operator on $L_2(\mathbb{R}_+)$.

Let $g = \mathcal{L}f$. Then

$$\begin{aligned} |g(x)|^2 &= \left| \int_0^{\infty} e^{-xy} f(y) dy \right|^2 \\ &= \left| \int_0^{\infty} \left(e^{-xy/2} y^{-1/4} \right) \left(e^{-xy/2} y^{1/4} f(y) \right) dy \right|^2. \end{aligned}$$

So by the Schwarz inequality

$$|g(x)|^2 \leq \int_0^{\infty} e^{-xy} y^{-1/2} dy \int_0^{\infty} e^{-xy} y^{1/2} |f(y)|^2 dy. \quad (3.11)$$

The first integral in (3.11) can be evaluated by a change of variables :

$$\begin{aligned} \int_0^{\infty} e^{-xy} y^{-1/2} dy &= x^{-1/2} \int_0^{\infty} e^{-u} u^{-1/2} du \\ &= x^{-1/2} 2 \int_0^{\infty} e^{-t^2} dt = \sqrt{\pi} x^{-1/2} \end{aligned}$$

Putting this into (3.11) and integrating we get

$$\|g\|^2 = \int_0^\infty |g(x)|^2 dx \leq \sqrt{\pi} \int_0^\infty \int_0^\infty x^{-1/2} e^{-xy} y^{1/2} |f(y)|^2 dy dx. \quad (3.12)$$

Change the order of integration. The integral with respect to x can be evaluated once again by a change of variables.

$$\int_0^\infty e^{-xy} x^{-1/2} dx = \sqrt{\pi} y^{-1/2}.$$

Hence we have from (3.12)

$$\|g\|^2 \leq \pi \int_0^\infty |f(y)|^2 dy = \pi \|f\|^2.$$

This shows that

$$\|\mathcal{L}\| \leq \sqrt{\pi}. \quad (3.13)$$

Actually the two sides of (3.13) are equal. How does one prove this? This would clearly be the case if we could produce a function f such that

$$\frac{\|\mathcal{L}f\|}{\|f\|} = \sqrt{\pi}. \quad (3.14)$$

The calculation above suggests a candidate. If $f(y) = y^{-1/2}$, then $\mathcal{L}f = \sqrt{\pi} f$. That seems more than what we need for (3.14). However, the argument is flawed since this function is *not* in $L_2(\mathbb{R}_+)$. What saves it is the observation that to prove $\|\mathcal{L}\| = \sqrt{\pi}$ we do not need (3.14). It is adequate to produce a sequence of functions f_n for which

$$\frac{\|\mathcal{L} f_n\|}{\|f_n\|} \rightarrow \sqrt{\pi}.$$

Exercise. Let $0 < a < b < \infty$ and let $f(y) = y^{-1/2}$ when $y \in [a, b]$, and $f(y) = 0$ outside this interval. Show that for each $\varepsilon > 0$ we can choose a and b such that $\|\mathcal{L}f\|^2 \geq (1 - \varepsilon)\pi \|f\|^2$. This shows that $\|\mathcal{L}\| = \sqrt{\pi}$.

The Hahn-Banach Theorem

The Hahn-Banach Theorem (H.B.T.) is called one of the three basic principles of linear analysis—the two others are the Uniform Boundedness Principle and the Open Mapping Theorem. We will study them in the next three lectures. The H.B.T. has several versions and several corollaries.

In essence the theorem says that we can extend a linear functional from a subspace to all of a vector space without increasing its size.

A real valued function p on a vector space X is called a *sublinear functional* if it is subadditive and positively homogenous; i.e.,

$$p(x + y) \leq p(x) + p(y) \quad \text{for all } x, y \in X,$$

$$p(\alpha x) = \alpha p(x) \quad \text{for } \alpha \geq 0, x \in X.$$

A norm is an example of a sublinear functional.

The H.B.T. for real vector spaces

1. Theorem. Let X be a real vector space and p a sublinear functional on it. Let X_0 be a subspace of X and let f_0 be a linear functional on X_0 such that $f_0(x) \leq p(x)$ for all $x \in X_0$. Then there exists a linear functional f on X such that $f(x) = f_0(x)$ whenever $x \in X_0$, and $f(x) \leq p(x)$ for all $x \in X$.

Proof. The idea is simple. Let x_1 be a vector outside X_0 . We will first extend f_0 to

the space spanned by X_0 and x_1 , and then use Zorn's Lemma. Let

$$X_1 := \{x + \alpha x_1 : x \in X_0, \alpha \in \mathbb{R}\}.$$

The representation of each element of X_1 in the form $x + \alpha x_1$ is unique. For every pair x, y in X_0

$$f_0(x) + f_0(y) = f_0(x + y) \leq p(x + y) \leq p(x + x_1) + p(y - x_1).$$

So,

$$f_0(y) - p(y - x_1) \leq p(x + x_1) - f_0(x).$$

Let

$$\begin{aligned} a &= \sup_{y \in X_0} [f_0(y) - p(y - x_1)], \\ b &= \inf_{x \in X_0} [p(x + x_1) - f_0(x)]. \end{aligned}$$

Then $a \leq b$. Choose any number c such that $a \leq c \leq b$. Then for all $x \in X_0$

$$f_0(x) - c \leq p(x - x_1),$$

$$f_0(x) + c \leq p(x + x_1).$$

Let α be any nonnegative real number, and multiply both sides of these inequalities by α . Then replace αx by x . This gives

$$f_0(x) - \alpha c \leq p(x - \alpha x_1),$$

$$f_0(x) + \alpha c \leq p(x + \alpha x_1),$$

for all $x \in X_0$, and $\alpha \geq 0$. Hence

$$f_0(x) + \alpha c \leq p(x + \alpha x_1)$$

for all $x \in X_0$ and for all $\alpha \in \mathbb{R}$. If we define

$$f_1(x + \alpha x_1) = f_0(x) + \alpha c,$$

then we get a linear functional f_1 on X_1 and $f_1(y) \leq p(y)$ for all $y \in X_1$.

Thus we have obtained an extension of f_0 to X_1 . Note this extension is *not unique* since it is defined in terms of c , an arbitrary number between a and b . If $X_1 = X$, we are done. If not, we can repeat the argument above extending f_1 to a bigger subspace of X . Does this process of extending by one dimension at a time eventually exhaust all of X ? We do not know this, and to overcome the difficulty we employ Zorn's Lemma.

Let $\mathcal{F}$ be the collection of all ordered pairs (Y, f) where Y is a subspace of X that contains X_0 , and f is a linear functional on Y that reduces to f_0 on X_0 and is dominated by p on Y . Define a partial order $\leq$ on $\mathcal{F}$ by saying that $(Y_1, f_1) \leq (Y_2, f_2)$ if Y_2 is a linear space that contains Y_1 and $f_2 = f_1$ on Y_1 . Let $\mathcal{G} = \{(Y_\alpha, f_\alpha)\}_{\alpha \in \Lambda}$ be a totally ordered subset of $\mathcal{F}$. Then the pair (Y, g) , where $Y = \bigcup_{\alpha \in \Lambda} Y_\alpha$ and $g(x) = f_\alpha(x)$ for $x \in Y_\alpha$, is an element of $\mathcal{F}$ and is an upper bound for $\mathcal{G}$. Therefore, by Zorn's Lemma, $\mathcal{F}$ has a maximal element. Let (Y_∞, f_∞) be this maximal element. If $Y_\infty \neq X$, then we could extend (Y_∞, f_∞) by adding one dimension as before. But then (Y_∞, f_∞) would not have been maximal. Thus $Y_\infty = X$ and if we put $f = f_\infty$, then f is a linear functional on X with the required properties. ■

The H.B.T. for complex vector spaces

2. Theorem. Let X be a (complex) vector space and p a sublinear functional on it. Let X_0 be a subspace of X and f_0 a linear functional on X_0 such that $\operatorname{Re} f_0(x) \leq p(x)$ for all $x \in X_0$. Then there exists a linear functional f on X such that $f(x) = f_0(x)$ whenever $x \in X_0$, and $\operatorname{Re} f(x) \leq p(x)$ for all $x \in X$.

Proof. Regard X as a vector space over $\mathbb{R}$ by restricting the scalars to real numbers. Let $g_0(x) = \operatorname{Re} f_0(x)$ for all $x \in X_0$. Then g_0 is a real linear functional on X_0 dominated by the sublinear functional p . So, g_0 can be extended to a real linear functional g on X dominated by p . Note that

$$g_0(ix) = \operatorname{Re} f_0(ix) = \operatorname{Re} if_0(x) = -\operatorname{Im} f_0(x).$$

So,

$$f_0(x) = g_0(x) - ig_0(ix) \quad \text{for all } x \in X_0.$$

This suggests that we define

$$f(x) = g(x) - ig(ix) \quad \text{for all } x \in X. \quad (4.1)$$

Then note that

$$\operatorname{Re} f(x) = g(x) \leq p(x) \quad \text{for all } x \in X.$$

So far we can say only that f is *real* linear: i.e. $f(x + y) = f(x) + f(y)$ and $f(\alpha x) = \alpha f(x)$ for $\alpha \in \mathbb{R}$. Let $\alpha + i\beta$ be any complex number. Then using (4.1) we see that

$$\begin{aligned} f((\alpha + i\beta)x) &= f(\alpha x + \beta ix) = \alpha f(x) + \beta f(ix) \\ &= \alpha f(x) + \beta[g(ix) - ig(-x)] \\ &= \alpha f(x) + \beta[g(ix) + ig(x)] \\ &= \alpha f(x) + i\beta[g(x) - ig(ix)] \\ &= \alpha f(x) + i\beta f(x) = (\alpha + i\beta)f(x). \end{aligned}$$

So f is complex linear as well. ■

The H.B.T. for normed linear spaces

3. This is the original version proved by F. Hahn in 1926.

Theorem. Let X be a normed linear space. Let X_0 be a subspace of it and let f_0 be a linear functional on X_0 such that $|f_0(x)| \leq C\|x\|$ for all $x \in X_0$ and some $C > 0$. Then there exists a linear functional f on X such that $f(x) = f_0(x)$ for all $x \in X_0$ and $|f(x)| \leq C\|x\|$ for all $x \in X$.

Proof. We will use the versions of H.B.T. proved in 1 and 2. We give the proof for real spaces and leave the complex case as an exercise.

Let $p(x) = C\|x\|$. This is a sublinear functional. Since $f_0(x) \leq p(x)$ for all $x \in X_0$, we can find a linear functional f on X that reduces to f_0 on X_0 and such that $f(x) \leq p(x)$ for all $x \in X$.

Since $p(-x) = p(x)$, it follows that $f(-x) \leq p(x)$; i.e., $-f(x) \leq p(x)$. So $|f(x)| \leq p(x) = C\|x\|$ for all $x \in X$. So the theorem is proved for real spaces. ■

The theorem says that a linear functional on X_0 can be extended to X *without increasing its norm*.

Corollaries of the H.B.T.

4. Proposition. Let X_0 be a subspace of a normed linear space X , and let x_1 be a vector such that $\text{dist}(x_1, X_0) = \delta > 0$. Then there exists a linear functional f on X such that

$$\|f\| = 1, \quad f(x_1) = \delta, \quad \text{and} \quad f(x) = 0 \quad \text{for all } x \in X_0.$$

Proof. Let X_1 be the linear span of X_0 and x_1 . Every vector in X_1 can be written uniquely as $y = x + \alpha x_1$ with $x \in X_0$, $\alpha \in \mathbb{C}$. Let $f_1(y) = \alpha\delta$. Then f_1 is a linear functional on X_1 , $f_1(x_1) = \delta$ and $f_1(x) = 0$ for all $x \in X_0$. If we show $\|f_1\| = 1$, the proposition would follow from the H.B.T.

Let x be any element of X_0 and let $\alpha \neq 0$. Then

$$\begin{aligned} |f_1(x + \alpha x_1)| &= |\alpha|\delta \leq |\alpha| \left\| \frac{x}{\alpha} + x_1 \right\| \quad (\text{see the definition of } \delta) \\ &= \|x + \alpha x_1\|. \end{aligned}$$

So $\|f_1\| \leq 1$. Note that for each $x \in X_0$, $|f_1(x - x_1)| = \delta$. Choose a sequence $x_n \in X_0$ such that $\|x_n - x_1\| \rightarrow \delta$. For this sequence $|f_1(x_n - x_1)|/\|x_n - x_1\|$ converges to 1. Hence $\|f_1\| = 1$. ■

5. Exercise. For each nonzero vector x_0 in a normed linear space X there exists a

linear functional f on X such that $\|f\| = 1$ and $f(x_0) = \|x_0\|$.

This shows that the norm of x can be expressed as

$$\|x\| = \sup_{f \in X^*, \|f\|=1} |f(x)|. \quad (4.2)$$

For each pair of distinct vectors x_1, x_2 in X , there exists a linear functional f on X such that $\|f\| = 1$ and $f(x_1) \neq f(x_2)$.

This last assertion is expressed by saying the space X^* *separates points* of X .

6. Theorem. Let X be a Banach space. If the dual space X^* is separable, then so is X .

Proof. Choose a countable dense set $\{f_n\}$ in X^* . For each n , choose $x_n \in X$ such that $\|x_n\| = 1$ and $|f_n(x_n)| \geq \frac{1}{2}\|f_n\|$. Let $\mathcal{D}$ be the collection of all rational linear combinations of elements of the set $\{x_n\}$. Then $\mathcal{D}$ is countable. Its closure $\bar{\mathcal{D}}$ is a subspace of X . If $\bar{\mathcal{D}} \neq X$, we can choose a linear functional f on X such that $\|f\| = 1$ and $f(x) = 0$ for all $x \in \bar{\mathcal{D}}$.

Since $\{f_n\}$ is dense in X^* , there exists a subsequence f_m converging to f . Note that

$$\frac{1}{2}\|f_m\| \leq |f_m(x_m)| = |(f_m - f)(x_m)| \leq \|f_m - f\|.$$

Thus $\|f_m\| \rightarrow 0$. Since $\|f_m\| \rightarrow \|f\|$ and $\|f\| = 1$, this is a contradiction. ■

We will see that $\ell_1^* = \ell_\infty$. So, the converse of the Theorem is not true.

7. Exercise. Let X_0 be a proper closed subspace of X . Show X_0 is nowhere dense in X . (It can not contain any ball.)

Banach Limits

8. Let ℓ_∞ be the space of *real* bounded sequences. A linear functional on this space is called a *Banach limit* if

- (i) $f(x_1, x_2, \dots) \geq 0$ if all $x_n \geq 0$.
- (ii) $f(x_2, x_3, \dots) = f(x_1, x_2, x_3, \dots)$.
- (iii) $f(1, 1, 1, \dots) = 1$.

We will show that such a linear functional exists.

Consider the subspace c in ℓ_∞ consisting of all convergent sequences. For an element $x = (x_1, x_2, \dots)$ of c let $f_0(x) = \lim x_n$.

This is a linear functional on c . For any $x = (x_1, x_2, \dots)$ in ℓ_∞ , define

$$p(x) = \inf \left\{ \overline{\lim}_{n \rightarrow \infty} \frac{1}{r} \sum_{j=1}^r x_{n+k_j} \right\}, \quad (4.3)$$

where the inf is over all choices of positive integers $r; k_1, \dots, k_r$.

Exercises. (i) Show that p is a sublinear functional.

(ii) Show that $p(x) \leq \overline{\lim} x_n$.

(iii) Show that $f_0(x) = p(x)$ for all $x \in c$.

Hence, by the H.B.T., there exists a linear functional f on ℓ_∞ such that

$$f(x) \leq p(x) \quad \text{for all } x \in \ell_\infty.$$

(iv) Show that

$$\underline{\lim} x_n \leq f(x) \leq \overline{\lim} x_n \quad \text{for all } x \in \ell_\infty.$$

(v) Let S be the linear operator on ℓ_∞ defined as

$$S((x_1, x_2, \dots)) = (x_2, x_3, \dots).$$

Show that

$$p(x - Sx) \leq 0 \quad \text{for all } x.$$

(vi) Show that $f(x) = f(Sx)$ for all x .

This shows the existence of a Banach limit.

Exercises. (i) A sequence in ℓ_∞ is called *almost convergent* if all its Banach limits are equal.

Show that x is almost convergent if

$$p(x) = -p(-x),$$

where p is defined by (4.3).

(ii) The sequence x is almost convergent and its Banach limit is ℓ if and only if

$$\lim_{r \rightarrow \infty} \frac{x_n + x_{n+1} + \cdots + x_{n+r-1}}{r} = \ell,$$

and the convergence is uniform in n .

Exercise. Find the Banach limit of the sequence $x = (1, 0, 1, 0, \dots)$.

9. The Hahn-Banach theorem has other geometric versions concerning separation properties of convex sets.

Let f be a nonzero linear functional on X . The set $\{x : f(x) = c\}$ is called a *hyperplane*.

Let X be a real normed linear space and let K be an open convex set in X . One geometric version of H. B. T. says that any point y not in K can be separated from K by a hyperplane; i.e., there is a linear functional f on X with $f(y) = c$ and $f(x) < c$ for all $x \in K$.

The Uniform Boundedness Principle

The *Baire Category Theorem* says that a complete metric space cannot be the union of a countable number of nowhere dense sets. This has several very useful consequences. One of them is the Uniform Boundedness Principle (U.B.P.) also called the *Banach-Steinhaus Theorem*.

The U.B.P.

1. **Theorem.** Let X be a Banach space, and let $\{p_\lambda\}$ be a family of continuous nonnegative functions on X , each satisfying the conditions

$$\begin{aligned} p_\lambda(x+y) &\leq p_\lambda(x) + p_\lambda(y) \text{ for all } x, y, \\ p_\lambda(-x) &= p_\lambda(x) \text{ for all } x. \end{aligned}$$

Suppose for each x ,

$$\sup_{\lambda} p_\lambda(x) < \infty.$$

Then

$$\sup_{\lambda} \sup_{\|x\| \leq 1} p_\lambda(x) < \infty.$$

(The hypothesis is that the family $\{p_\lambda\}$ is bounded at each point x ; the conclusion is that it is uniformly bounded over the unit ball of X .)

Proof. For each n , let

$$C_n := \{x : \sup_{\lambda} p_\lambda(x) \leq n\}.$$

We can also write

$$C_n = \cap_{\lambda} \{x : p_{\lambda}(x) \leq n\}.$$

Since p_{λ} are continuous, C_n is closed. By the hypothesis $X = \cup_n C_n$. So, by the Baire Category Theorem, there exists an n_0 such that the set C_{n_0} contains a closed ball $\bar{B}(x_0, r)$.

Let x be any element of X such that $\frac{1}{2}\|x\| \leq r$. Then the vectors $x_0 \pm x/2$ are in the ball $\bar{B}(x_0, r)$. Since

$$x = x_0 + \frac{x}{2} - (x_0 - \frac{x}{2})$$

we have

$$p_{\lambda}(x) \leq p_{\lambda}(x_0 + \frac{x}{2}) + p_{\lambda}(x_0 - \frac{x}{2}) \leq 2n_0.$$

This is true for all x with $\|x\| \leq 2r$. Hence,

$$\sup_{\lambda} \sup_{\|x\| \leq 2r} p_{\lambda}(x) \leq 2n_0 < \infty.$$

If $1 \leq 2r$, the proof is over. If this is not the case, choose a positive integer $m > 1/2r$.

Now if $\|x\| \leq 1$, then $\|x/m\| < 2r$, and

$$p_{\lambda}(x) \leq mp_{\lambda}(\frac{x}{m}) \leq 2mn_0.$$

So, $\sup_{\lambda} \sup_{\|x\| \leq 1} p_{\lambda}(x) \leq 2mn_0 < \infty$. ■

2. Corollary. Let X be a Banach space and let $\{A_{\alpha}\}$ be a family of bounded linear operators from X into a normed linear space Y . Suppose for each $x \in X$

$$\sup_{\alpha} \|A_{\alpha}x\| < \infty.$$

Then

$$\sup_{\alpha} \|A_{\alpha}\| < \infty.$$

3. The completeness of X is an essential requirement in the U.B.P. Consider the space $X = c_{00}$ in ℓ_{∞} . On this space define for each n , a linear functional f_n as

$f_n(x) = nx_n$. Then for each x in c_{00}

$$\sup_n |f_n(x)| < \infty$$

(because the terms of the sequence x are zero after some stage). However, $\|f_n\| = n$, and hence

$$\sup_n \|f_n\| = \infty.$$

Typical Applications of the U.B.P.

4. Proposition. Let $\{f_n\}$ be a sequence of bounded linear functionals on a Banach space X . Suppose for each x , $f_n(x)$ converges to a limit $f(x)$. Then f is a bounded linear functional.

Proof. It is easy to see that f is linear. For each x , the sequence $\{f_n(x)\}$ is convergent, hence bounded; i.e., there exists a number $K(x)$ such that

$$\sup_n |f_n(x)| = K(x) < \infty.$$

Hence, by the U.B.P., there exists a number K such that

$$\sup_n \sup_{\|x\| \leq 1} |f_n(x)| \leq K.$$

Hence $\sup_{\|x\| \leq 1} |f(x)| \leq K$. ■

In general, the pointwise limit of continuous functions is not continuous. The proposition just proved says that this is the case when the functions involved are linear functionals.

5. Proposition. Let X, Y, Z be Banach spaces. Let A_n be a sequence in $\mathcal{B}(X, Y)$ such that $A_n x$ converges to Ax for each $x \in X$, and B_n a sequence in $\mathcal{B}(Y, Z)$ such that $B_n(y)$ converges to By for each $y \in Y$. Then $B_n A_n x$ converges to BAx for each $x \in X$.

Proof. For each x , the sequence $\|A_n x\|$ is convergent, hence bounded. So, by the U.B.P. the sequence $\|A_n\|$ is bounded. This is true for $\|B_n\|$ also. Note that

$$\begin{aligned}\|B_n A_n x - B A x\| &= \|B_n(A_n - A)x + (B_n - B)A x\| \\ &\leq \|B_n\| \|(A_n - A)x\| + \|(B_n - B)A x\|.\end{aligned}$$

As $n \rightarrow \infty$ both the terms on the right go to zero. ■

Divergence of Fourier Series

6. Let X be the Banach space of continuous functions on the interval $[-\pi, \pi]$. The *Fourier coefficients of a function f in X* are the numbers

$$a_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-int} dt. \quad (5.1)$$

The *Fourier series* of f is the series

$$\sum_{n=-\infty}^{\infty} a_n e^{int}. \quad (5.2)$$

One of the basic questions in the study of such series is whether this series converges at each point t in $[-\pi, \pi]$, and if so, is its sum equal to $f(t)$?

An example to show that this is not always the case was constructed by Du Bois-Raymond in 1876. The idea was to construct successively worse functions and take their limit. This is called *condensation of singularities* and eventually it led to the discovery of the U.B.P.

Using the U.B.P. it is possible to give a *soft proof* of the existence of a continuous function whose Fourier series diverges at some point. A soft proof means that the messy construction of an explicit example is avoided. Such a proof is given below.

7. For each f , let

$$\Lambda_N(f) = \sum_{n=-N}^N a_n$$

be the partial sum of the series (5.2) at $t = 0$. For each N , this is a linear functional on X . We have

$$\Lambda_N(f) = \int_{-\pi}^{\pi} f(t) D_N(t) dt,$$

where

$$D_N(t) = \frac{1}{2\pi} \sum_{n=-N}^N e^{int}$$

is called the *Dirichlet kernel*. One can see that

$$D_N(t) = \frac{1}{2\pi} \frac{\sin(N + \frac{1}{2})t}{\sin \frac{t}{2}},$$

and using this

$$\lim_{N \rightarrow \infty} \int_{-\pi}^{\pi} |D_N(t)| dt = \infty. \quad (5.3)$$

Note that $\|\Lambda_N\| \leq \int_{-\pi}^{\pi} |D_N(t)| dt$. For a fixed N , let $g_N(t) = \text{sgn } D_N(t)$. This is a step function and can be approximated by continuous functions of norm 1; i.e., there exist ϕ_m in X such that $\|\phi_m\| = 1$ and $\lim \phi_m(t) = g_N(t)$ for every t . Hence, by the Dominated Convergence Theorem

$$\lim_{m \rightarrow \infty} \Lambda_N(\phi_m) = \int_{-\pi}^{\pi} g_N(t) D_N(t) dt = \int_{-\pi}^{\pi} |D_N(t)| dt.$$

Thus, $\|\Lambda_N\| = \int_{-\pi}^{\pi} |D_N(t)| dt$ and by (5.3) $\|\Lambda_N\|$ is unbounded. Hence by the U.B.P, there exists an f in X for which $|\Lambda_N(f)|$ is unbounded; i.e., the Fourier series of f diverges at 0.

Exercises.

8. A subset of a metric space is said to be *meagre (of first category)* if it is the union of a countable family of nowhere dense sets.

Let X, Y be Banach spaces and let $\mathcal{S}$ be a subset of $\mathcal{B}(X, Y)$. Suppose there exists a point $x_0 \in X$ such that the set $\{Ax_0 : A \in \mathcal{S}\}$ is unbounded. Show that the

set

$$\{x \in X : \sup_{A \in \mathcal{S}} \|Ax\| < \infty\}$$

is meagre in X . (Examine the proof of the U.B.P.).

9. For each t in $[-\pi, \pi]$ consider the set of all f in $C[-\pi, \pi]$ for which the partial sums of its Fourier series $\sum_{n=-N}^N a_n e^{int}$ are bounded. Show that this set is meagre in $C[-\pi, \pi]$.

10. Show that there exists a continuous function on $[-\pi, \pi]$ whose Fourier series diverges at each point of a dense set in $[-\pi, \pi]$.

The Open Mapping Theorem

Theorems that tell us that a continuous map is also open under some simple conditions play a very important role in analysis. The open mapping theorem is one such result.

1. **Theorem.** Let X, Y be Banach spaces and let A be a bounded linear operator from X to Y . If A is surjective, then it is an open map (i.e., the image of every open set under A is open).

A few comments before the proof might be helpful.

In the presence of linearity, continuity arguments are often simpler. A *translation* on X is a map of the form $T(x) = x + x_0$, and a *dilation* one of the form $Tx = \alpha x, \alpha \neq 0$. If X is a normed linear space, then all translations and dilations are homeomorphisms of X . If we show that the image under A of some open ball around 0 in X contains an open ball around 0 in Y , then it would follow that the image of every open ball contains an open ball, and hence A is open.

If E and F are two subsets of a vector space X , then $E + F$ stands for the set $\{x + y : x \in E, y \in F\}$, and αE for the set $\{\alpha x : x \in E\}$. Clearly $2E \subset E + E$. If E is a convex set, then $2E = E + E$. In particular this is true when E is any ball in a normed linear space. The closure of a convex set is convex, and the image of a convex set under a linear map is convex.

We will use the notation $B_X(x_0, r)$ for the open ball of radius r around the point x_0 in X .

Proof of the theorem. Let $E = A(B_X(0, 1))$, and let F be its closure. The first step of the proof consists of showing that F contains an open ball $B_Y(0, 2\varepsilon)$, and the second step of showing that this implies that E contains the ball $B_Y(0, \varepsilon)$. We have observed that this would suffice for proving the theorem.

Since A is a surjective linear map, we have $Y = \cup_{n=1}^{\infty} A(B_X(0, n)) = \cup_{n=1}^{\infty} nE$. Since the space Y is complete, the Baire category theorem tells us that for some m the set $\overline{mE} = mF$ has a nonempty interior. Hence F contains some open ball, say $B_Y(y_0, 4\varepsilon)$. The point y_0 , being in F , can be expressed as $y_0 = \lim Ax_n$, where x_n is a sequence in $B_X(0, 1)$. The points $-x_n$ are also in $B_X(0, 1)$, and hence $-y_0$ is in F . Thus $B_Y(0, 4\varepsilon) = B_Y(y_0, 4\varepsilon) - y_0 \subset F + F = 2F$, and hence $B_Y(0, 2\varepsilon) \subset F$. The first step is over.

Let y be any point of $B_Y(0, 2\varepsilon)$. Since $y \in F$, there exists a point y_1 in E such that $\|y - y_1\| < \varepsilon$. In other words, $y - y_1$ is a point of $B_Y(0, \varepsilon)$ which in turn is a subset of $\frac{1}{2}F$. Repeating the argument, we can find a point y_2 in the set $\frac{1}{2}E$ such that $y - (y_1 + y_2)$ is in $B_Y(0, \varepsilon/2)$, a subset of $\frac{1}{4}F$. Thus we have a sequence $y_n = Ax_n$, where $\|x_n\| < 1/2^{n-1}$, and $\|y - (y_1 + \cdots + y_n)\| < \varepsilon/2^{n-1}$. Let $x = \sum_{n=1}^{\infty} x_n$. Then $\|x\| < \sum_{n=1}^{\infty} 1/2^{n-1} = 2$, and $y = \sum_{n=1}^{\infty} y_n = \sum_{n=1}^{\infty} Ax_n = Ax$. We have shown that every point of $B_Y(0, 2\varepsilon)$ is in the set $2E$. Hence every point of $B_Y(0, \varepsilon)$ is in E , and this completes the proof. ■

2. The Inverse Mapping Theorem. Let X, Y be Banach spaces. If a bounded linear operator A from X to Y is bijective, then the inverse A^{-1} is a bounded linear operator.

This is an immediate consequence of the Open Mapping Theorem. (Recall that the inverse of a linear operator, if it exists, is linear.)

3. Remark. The crucial part of the hypothesis is that A is surjective. If the range $\text{ran}(A)$ were always closed, the theorem would be trivial: we would just say if A is injective then its inverse A^{-1} from the Banach space $\text{ran}(A)$ to X is a bounded linear

operator. However, $\text{ran}(A)$ is not always closed. For example, let A be the map on ℓ_2 that sends the sequence $\{x_n\}$ to $\{x_n/n\}$. Then $\text{ran}(A)$ contains all sequences $\{y_n\}$ for which

$$\sum n^2 |y_n|^2 < \infty.$$

In particular, $\text{ran}(A)$ contains the space c_{00} , and hence it is dense in ℓ_2 . If it were closed it would be all of ℓ_2 . But that can not be, since the sequence $\{\frac{1}{n}\}$ is not in $\text{ran}(A)$.

This example shows that the inverse of a bounded linear operator from a Banach space onto an incomplete normed linear space need not be bounded. Rephrasing this example in terms of infinite matrices makes the picture clearer. The operator A acts on the standard basis $\{e_n\}$ as $Ae_n = e_n/n$, and hence it is represented by the infinite diagonal matrix $\text{diag}(1, 1/2, 1/3, \dots)$. Clearly A is injective and $\|A\| = 1$. However, the linear operator A^{-1} from $\text{ran}(A)$ into ℓ_2 corresponds to the diagonal matrix $\text{diag}(1, 2, 3, \dots)$, and is not bounded.

The Closed Graph Theorem

4. If X, Y are vector spaces, then their direct sum $X \oplus Y$ is the collection of ordered pairs (x, y) with $x \in X, y \in Y$, and with vector space operations defined as usual. If X, Y are normed linear spaces, we define

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

This is a norm on $X \oplus Y$. If X, Y are Banach spaces, then $X \oplus Y$ is a Banach space.

The maps $P_1(x, y) := x$ and $P_2(x, y) := y$ are called the *projections* onto X and Y . They are linear and continuous.

If A is a linear operator from X to Y , its *graph* is the set $\{(x, Ax) : x \in X\}$. This is a linear subspace of $X \oplus Y$.

5. **Theorem.** Let X, Y be Banach spaces and let A be a linear map from X to Y .

Then A is bounded if and only if its graph is a closed subspace of $X \oplus Y$.

Proof. Let $G(A)$ be the graph of A . It is easy to see that if A is continuous, then $G(A)$ is closed.

If $G(A)$ is closed, then it is a Banach space (in the space $X \oplus Y$). For an element (x, Ax) of $G(A)$, let $P_1(x, Ax) = x$, $P_2(x, Ax) = Ax$. Then P_1, P_2 are continuous linear maps from $G(A)$ into X, Y , respectively. The map P_1 is a bijection. So, its inverse P_1^{-1} is a continuous map from X onto $G(A)$, by the Open Mapping Theorem. Since $A = P_2 P_1^{-1}$, A is also continuous. ■

6. What does this theorem say? Let f be any map from X to Y . To say that f is continuous means that if $x_n \rightarrow x$ in X , then the sequence $f(x_n)$ converges to a limit y in Y and $y = f(x)$. The Closed Graph Theorem says that if f is a *linear* map between Banach spaces, then to prove its continuity we have to show that if $x_n \rightarrow x$ in X and $f(x_n) \rightarrow y$ in Y , then $y = f(x)$. This makes it easier to check whether a linear map is continuous.

The assertion of the theorem is not always true if X or Y is not complete. For example, let $Y = C[0, 1]$ and let X be the linear subspace of Y consisting of functions that have continuous derivatives. The derivative map $Af = f'$ is a linear operator from X into Y . It is not continuous but its graph is closed.

7. **Exercise.** Let X be a vector space with two norms $\|\cdot\|_1$ and $\|\cdot\|_2$, both of which make it a Banach space. Suppose there exists a constant C such that $\|x\|_1 \leq C\|x\|_2$ for all x . Then there exists a constant D such that $\|x\|_2 \leq D\|x\|_1$ for all x .

8. **Exercise.** Let X be a Banach space with a Schauder basis $\{x_n\}$. Let $\{a_n(x)\}$ be the coefficients of x in this basis; i.e., let $x = \sum a_n(x)x_n$. Show that each a_n is a bounded linear functional on X .

[Hint: Consider the space Y consisting of all sequences $a = (a_1, a_2, \dots)$ for which the series $\sum a_n x_n$ converges in X . Define the norm of such a sequence as

$$\|a\| = \sup_n \left\| \sum_{j=1}^n a_j x_j \right\|.$$

Show that Y is a Banach space with this norm. The map $T(a) = \sum a_n x_n$ is a bounded linear operator from Y onto X . Use the Inverse Mapping Theorem now.]

Some Applications of the Basic Principles

9. **Exercise.** The algebraic dimension of any infinite-dimensional Banach space can not be countable. (If X has a countable Hamel basis then X can be expressed as a countable union of nowhere dense sets.)

10. **Exercise.** The algebraic dimension of ℓ_∞ is c , the cardinality of the continuum. Hints : For each t in $(0, 1)$ let $x_t = (1, t, t^2, \dots)$. Then $x_t \in \ell_\infty$ and the family $\{x_t : 0 < t < 1\}$ is linearly independent. One way of seeing this is by observing that the Vandermonde determinant

$$\begin{vmatrix} 1 & \dots & 1 \\ t_1 & \dots & t_n \\ \cdot & \dots & \cdot \\ t_1^{n-1} & \dots & t_n^{n-1} \end{vmatrix} = \prod_{i>j} (t_i - t_j)$$

is non zero if $t_i \neq t_j$. Thus $\dim \ell_\infty \geq c$. Since the cardinality of ℓ_∞ as a set is also c (why?) it follows that $\dim \ell_\infty = c$.

11. **Proposition.** Every infinite-dimensional Banach space X contains a vector space that is algebraically isomorphic to ℓ_∞ .

Proof. Let f_1 be a nonzero continuous linear functional on X . Let Z_1 be its kernel. Then Z_1 is a closed linear subspace of X and its codimension is one. Choose a vector

$x_1 \in X \setminus Z_1$ (the complement of Z_1 in X) with $\|x_1\| = 1$.

Now let f_2 be a nonzero continuous linear functional on Z_1 and let Z_2 be its kernel. Choose a vector $x_2 \in Z_1 \setminus Z_2$ with $\|x_2\| = 1/2$. Continuing this process we get a decreasing sequence of subspaces $X \supset Z_1 \supset Z_2 \supset \dots$, and a sequence of vectors x_n such that $\|x_n\| = 1/2^{n-1}$, and $x_1, \dots, x_n \notin Z_n$.

For an element $a = (a_1, a_2, \dots)$ of ℓ_∞ , let $T(a) = \sum_{n=1}^{\infty} a_n x_n$. Since

$$\sum \|a_n x_n\| \leq \sum |a_n| \|x_n\| \leq \|a\|_\infty \sum \|x_n\| = 2\|a\|_\infty,$$

the series $\sum a_n x_n$ is convergent and T is a bounded linear map from ℓ_∞ into X . It is easy to see that $T(a) = 0$ if and only if $a = 0$. So, T is injective. Thus T is an algebraic isomorphism of ℓ_∞ onto its range.

12. Corollary. The algebraic dimension of any infinite-dimensional Banach space is at least c .

13. An *isometric isomorphism* is a map of one normed linear space onto another that preserves norms and is a linear isomorphism.

Proposition. Every separable Banach space X is isometrically isomorphic to a subspace of ℓ_∞ .

Proof. Let $D = \{x_1, x_2, \dots\}$ be a countable dense subset of X . By the H.B.T. there exists linear functionals f_n on X such that $\|f_n\| = 1$ and $f_n(x_n) = \|x_n\|$. For each x in X let

$$Tx = (f_1(x), f_2(x), \dots).$$

Since $|f_n(x)| \leq \|x\|$, $Tx \in \ell_\infty$. Thus T is a linear map from X into ℓ_∞ and $\|Tx\| \leq \|x\|$. It remains to show that $\|Tx\| = \|x\|$ for all x . Given any x choose a sequence x_m in D such that $x_m \rightarrow x$. Then $\|x_m\| \rightarrow \|x\|$ and $\|Tx_m\| \rightarrow \|Tx\|$. But for each m , $\|Tx_m\| = \sup_n |f_n(x_m)| = \|x_m\|$. So $\|Tx\| = \|x\|$. ■

14. The sequence spaces ℓ_p , $1 \leq p \leq \infty$, and c_0 seem more familiar than abstract Banach spaces since we can “see” sequences. Proposition 13 says every separable Banach space is (upto an isometric isomorphism) a subspace of ℓ_∞ . For long functional analysts sought to know whether every infinite dimensional separable Banach space contains a subspace that is isometrically isomorphic to either c_0 or to some ℓ_p , $1 \leq p < \infty$. In 1974, B. Tsirelson showed that this is not always so.

Dual Spaces

The idea of duality, and the associated notion of adjointness, are important in functional analysis. We will identify the spaces X^* for some of the standard Banach spaces.

The dual of $\mathbb{C}^n$

1. Let f be a linear functional on $\mathbb{C}^n$. If $e_1, \dots, e_n$ is the standard basis for $\mathbb{C}^n$, then the numbers $\eta_j = f(e_j)$ completely characterise f . The action of f on any element $x = (x_1, x_2, \dots, x_n)$ of $\mathbb{C}^n$ is given by the formula

$$f(x) = \sum_{j=1}^n x_j \eta_j. \quad (7.1)$$

Any vector $\eta = (\eta_1, \dots, \eta_n)$ gives rise to a linear functional f on $\mathbb{C}^n$ via this formula. Thus the vector space dual to $\mathbb{C}^n$ is $\mathbb{C}^n$ itself.

Every linear functional on $\mathbb{C}^n$ is continuous (no matter what norm we choose on $\mathbb{C}^n$). However, its norm will, of course, depend on the norm we choose for $\mathbb{C}^n$.

2. Consider the space $\ell_p^n, 1 \leq p \leq \infty$. We will calculate the norm of a linear functional f on this space in terms of the vector η with which f can be identified as in (7.1).

(i) Let $1 < p < \infty$. By Hölder's inequality $|f(x)| = |\sum x_j \eta_j| \leq \|x\|_p \|\eta\|_q$. So, $\|f\| \leq \|\eta\|_q$. Show that this is an equality by considering the special vector x defined

by

$$x_i = \begin{cases} 0 & \text{if } \eta_i = 0 \\ |\eta_i|^q / \eta_i & \text{if } \eta_i \neq 0. \end{cases}$$

(ii) Let $p = 1$. Note that

$$\begin{aligned} |f(x)| &\leq \sum |x_j| |\eta_j| \\ &\leq \max |\eta_j| \sum |x_j| = \|\eta\|_\infty \|x\|_1. \end{aligned}$$

So, $\|f\| \leq \|\eta\|_\infty$. Show that this is an equality by considering the vector x defined by

$$x_i = \begin{cases} |\eta_i| / \eta_i & \text{if } |\eta_i| = \|\eta\|_\infty \\ 0 & \text{otherwise.} \end{cases}$$

(iii) Let $p = \infty$. Once again note that $|f(x)| \leq \|\eta\|_1 \|x\|_\infty$. So $\|f\| \leq \|\eta\|_1$. Show that $\|f\| = \|\eta\|_1$.

The conclusion is

$$(\ell_p^n)^* = \ell_q^n \quad \text{for } 1 \leq p \leq \infty.$$

Of course this equality is to be understood in the sense that we do not distinguish between isometrically isomorphic spaces.

The dual of ℓ_p

3. The arguments we have given can be pushed to study sequence spaces. A little extra care is required to handle infinite sums.

Let $1 < p < \infty$. The standard basis e_j is a Schauder basis for ℓ_p . For $f \in \ell_p^*$, let $\eta_j = f(e_j)$. For each natural number n ,

$$\left(\sum_{j=1}^n |\eta_j|^q \right)^{\frac{1}{q}} = \sup_{x \in \ell_p^n, \|x\|_p=1} |f(x)| \leq \|f\|.$$

This is true for all n . So, $\eta \in \ell_q$ and $\|\eta\|_q \leq \|f\|$.

Let x be any element of ℓ_p . By Hölder's inequality $\sum |x_j \eta_j| \leq \|\eta\|_q \|x\|_p$. We

can write $x = \sum x_j e_j$ (a convergent series in ℓ_p). Then since f is continuous,

$$|f(x)| = \left| \sum x_j \eta_j \right| \leq \|\eta\|_q \|x\|_p.$$

So, $\|f\| \leq \|\eta\|_q$; and hence $\|f\| = \|\eta\|_q$.

Thus ℓ_p^* can be identified as a subspace of ℓ_q . But every element η of ℓ_q acts as a linear functional on ℓ_p , by sending an element x of ℓ_p to $\sum x_j \eta_j$. This linear functional f is bounded, the correspondence between η and f is linear, and $\|f\| = \|\eta\|_q$. (Hölder's inequality again.) Thus

$$\ell_p^* = \ell_q \quad \text{for } 1 < p < \infty.$$

Exercise. Show that $\ell_1^* = \ell_\infty$. (The argument is similar to the one used above.)

Thus

$$\ell_p^* = \ell_q \quad \text{for } 1 \leq p < \infty. \quad (7.2)$$

4. How about the remaining case $p = \infty$? Since ℓ_∞ does not have a Schauder basis we can not imitate the earlier reasoning. In any case ℓ_∞^* could not be ℓ_1 : we have seen that a nonseparable space can not have a separable dual.

What then is the dual of ℓ_∞ ? We will not calculate it here. There is a very general theorem called the *Riesz Representation Theorem* that describes the dual of the space of bounded continuous functions on a locally compact Hausdorff space X . (See W. Rudin, *Real and Complex Analysis*.) When $X = \mathbb{N}$, this space is ℓ_∞ .

5. Well, if ℓ_1 is not the dual of ℓ_∞ , is it the dual of *some* other space? (Incidentally, not every Banach space is the dual of another. The space $C[0,1]$ is not the dual of any Banach space. We will not prove this fact here.)

Proposition. $\ell_1 = c_0^*$.

Proof. Once again, the standard basis e_j is a Schauder basis for c_0 . Let $f \in c_0^*$

and let $\eta_j = f(e_j)$. If $\eta_j = |\eta_j| \exp(i\theta_j)$, then for every n we have

$$\begin{aligned} \sum_{j=1}^n |\eta_j| &= \sum_{j=1}^n \exp(-i\theta_j) \eta_j = \sum_{j=1}^n f(\exp(-i\theta_j) e_j) \\ &= f(e^{-i\theta_1}, e^{-i\theta_2}, \dots, e^{-i\theta_n}, 0, 0, \dots) \leq \|f\|. \end{aligned}$$

Hence $\eta \in \ell_1$ and $\|\eta\|_1 \leq \|f\|$. If x is a vector in the space c_{00} and $\|x\|_\infty \leq 1$, then $|f(x)| \leq \|\eta\|_1$. Since c_{00} is dense in c_0 , this shows that $\|f\| \leq \|\eta\|_1$. Hence $\|f\| = \|\eta\|_1$. Thus we have an isometric isomorphism between the spaces c_0^* and ℓ_1 . ■

6. (i) Show that ℓ_1 is the dual of the space c as well. [Hint: Let e_j , $j = 1, 2, \dots$, be the standard unit vectors and $e = (1, 1, 1, \dots)$. Then $\{e, e_1, e_2, \dots\}$ is a Schauder basis for the space c .]

(ii) Show that the spaces c_0 and c are not isometrically isomorphic. [Hint: the unit ball in c has two extreme points ($x_n \equiv 1$ and $x_n \equiv -1$), the unit ball in c_0 has none.]

7. The dual of the space $L_p[0, 1]$ for each $1 \leq p < \infty$ is $L_q[0, 1]$. The proof uses Hölder's inequality for integrals and very standard measure theory arguments.

This statement is true for $L_p(X, \mathcal{S}, \mu)$ with some restrictions on the measure space (like σ -finiteness).

The dual of $C[0, 1]$

8. Let g be a function on $[0, 1]$. Let $\mathcal{P}$ be a partition of $[0, 1]$ as $0 = t_0 < t_1 < \dots < t_n = 1$. Let

$$v(g; \mathcal{P}) = \sum_{j=1}^n |g(t_j) - g(t_{j-1})|,$$

and

$$V(g) = \sup v(g; \mathcal{P}), \tag{7.3}$$

where the supremum is taken over all possible partitions $\mathcal{P}$. If $V(g)$ is finite we say

that g is of *bounded variation*, and then $V(g)$ is called the *total variation* of g . The space $BV[0, 1]$ consisting of all such functions is a vector space. Every absolutely continuous function is in this space. There exists a continuous function that is not of bounded variation (consider $t \sin(\pi/t)$ near zero). A function of bounded variation need not be continuous (consider characteristic functions of intervals).

If we put $\|g\| = V(g)$ we get a *pseudonorm*: all properties of a norm are satisfied except one; $\|g\|$ could be zero without g being zero. Every constant function has zero total variation; the converse is also true. To get over this we could consider two functions of bounded variation to be *equivalent* if their difference is a constant. The space $BV[0, 1]$ then consists of equivalence classes of functions with respect to this relation. Alternately, we could modify the definition of $\|g\|$ by putting $\|g\| = |g(0)| + V(g)$. In either case we get a normed linear space which is in fact a Banach space. (Try to prove this.)

9. Every g in $BV[0, 1]$ gives rise to a linear functional g^* on $C[0, 1]$:

$$g^*(f) = \int f dg,$$

where the integral is the Riemann–Stieltjes integral. We have

$$|g^*(f)| \leq \int |f| |dg| \leq \|f\|_\infty \|g\|.$$

Thus $\|g^*\| \leq \|g\|$.

10. We will show that conversely, *every* bounded linear functional on $C[0, 1]$ arises in this way.

Let $\varphi \in (C[0, 1])^*$. The space $C[0, 1]$ is contained in the space $B[0, 1]$ consisting of all bounded functions with the $\|\cdot\|_\infty$ norm. By the Hahn-Banach Theorem φ can be extended to a linear functional on $B[0, 1]$ and the extension has the same norm as φ .

Let χ_0 be the zero function, and for $0 < t \leq 1$ let χ_t be the characteristic function

of the interval $[0, t]$. Let $g(t) = \varphi(\chi_t)$. We will show that g is of bounded variation.

For any partition $0 = t_0 < t_1 < \cdots < t_n = 1$,

$$\begin{aligned} \sum_{i=1}^n |g(t_i) - g(t_{i-1})| &= \sum_{i=1}^n [g(t_i) - g(t_{i-1})] \operatorname{sgn} [g(t_i) - g(t_{i-1})] \\ &= \sum_{i=1}^n [\varphi(\chi_{t_i}) - \varphi(\chi_{t_{i-1}})] \operatorname{sgn} [g(t_i) - g(t_{i-1})] \\ &= \varphi \left(\sum_{i=1}^n (\chi_{t_i} - \chi_{t_{i-1}}) \operatorname{sgn} [g(t_i) - g(t_{i-1})] \right). \end{aligned}$$

We have here the linear functional φ acting on an element of $B[0, 1]$. The $\|\cdot\|_\infty$ norm of this element is 1 (at any point t only one of the $\chi_{t_i}(t) - \chi_{t_{i-1}}(t)$ is nonvanishing; $\operatorname{sgn} x$ is a number of modulus 1). Thus $v(g; \mathcal{P}) \leq \|\varphi\|$ for every partition $\mathcal{P}$, and hence

$$V(g) \leq \|\varphi\|.$$

Note that $g(0) = \varphi(\chi_0) = 0$ and for each t

$$\varphi(\chi_t) = g(t) - g(0) = \int \chi_t dg.$$

Now follow the usual path in constructing integrals. Extend this relation to all step functions by linearity, and then to all continuous functions by taking limits. We thus have

$$\varphi(f) = \int f dg \quad \text{for all } f \in C[0, 1],$$

and further,

$$\|\varphi\| = \|g\| = V(g).$$

11. The space $BV[0, 1]$ contains some ill-behaved functions. For example, consider for $0 < c < 1$ the delta function

$$\delta_c(t) = \begin{cases} 1 & \text{if } t = c \\ 0 & \text{if } t \neq c. \end{cases}$$

We have $V(\delta_c) = 2$, but $\int f d\delta_c = 0$ for all continuous f . In other words

$$\|\delta_c^*\| = 0, \quad \text{but } \|\delta_c\| = 2.$$

How does one get over this? One needs to do a little more work than that for the pseudonorm problem in Section 8.

12. Let $g \in BV[0, 1]$ and suppose $g(0) = 0$. Then there exists a unique function $\tilde{g}$ in $BV[0, 1]$ satisfying the conditions $\tilde{g}(0) = 0$, $\tilde{g}$ is right continuous on $(0, 1)$ and

$$\int f d\tilde{g} = \int f dg \quad \text{for all } f \in C[0, 1].$$

This is an exercise in epsilonics. Every function of bounded variation has a countable number of discontinuities. We can choose $\tilde{g}(t) = g(t^+)$ for all $t > 0$.

Check that $\tilde{g}$ satisfies the requirements and also that $\|\tilde{g}\| \leq \|g\|$.

13. We say that a function g in $BV[0, 1]$ is *normalised* if $g(0) = 0$ and g is right continuous on $(0, 1)$. The collection of all such functions is denoted as $BVN[0, 1]$. This is a Banach space with the total variation norm.

14. **The Riesz Representation Theorem.** The dual of the space $C[0, 1]$ is the space $BVN[0, 1]$. Each element g of $BVN[0, 1]$ can be identified with a linear functional g^* on $C[0, 1]$ by the relation

$$g^*(f) = \int f dg, \quad f \in C[0, 1].$$

This gives an isometric isomorphism between $BVN[0, 1]$ and $(C[0, 1])^*$.

We have seen all the essential details of the proof.

15. Every real function of bounded variation is the difference of two monotonically increasing functions.

Let us consider the space $C_{\mathbb{R}}[0, 1]$ consisting of real continuous functions. We want to know for what bounded linear functionals φ on this space the function g associated to it by the Riesz Representation Theorem is monotonically increasing. (The measure corresponding to a monotonically increasing function is positive; that corresponding to the difference of two such functions is a signed measure.)

Positive Linear Functionals

16. A function (on any domain) is called *positive* if it takes nonnegative values; i.e., $f(x) \geq 0$ for all x . We write this briefly as $f \geq 0$.

A linear functional φ on $C[0, 1]$ is said to be *positive* if $\varphi(f) \geq 0$ whenever $f \geq 0$. (Note $\varphi(f)$ is a number.)

The study of maps that preserve positivity (in different senses) is an important topic in analysis.

17. Let φ be a positive linear functional on $C_{\mathbb{R}}[0, 1]$. Then $\|\varphi\| = \varphi(\mathbf{1})$, where $\mathbf{1}$ denotes the function taking the value 1 everywhere. This fact is easy to prove. Just note that for every f

$$\|f\| \mathbf{1} \pm f \geq 0.$$

Since φ is positive and linear

$$\|f\| \varphi(\mathbf{1}) \pm \varphi(f) \geq 0.$$

So,

$$|\varphi(f)| \leq \|f\| \varphi(\mathbf{1}).$$

Thus $\|\varphi\| \leq \varphi(\mathbf{1})$. Since $\varphi(\mathbf{1}) \leq \|\varphi\| \|\mathbf{1}\| = \|\varphi\|$, this means $\|\varphi\| = \varphi(\mathbf{1})$.

A corollary of this is that any extension $\tilde{\varphi}$ of φ to $B_{\mathbb{R}}[0, 1]$ obtained via the Hahn-Banach Theorem is also positive. If not, there would exist an f with $0 \leq f \leq \mathbf{1}$ such that $\tilde{\varphi}(f) < 0$. Then

$$\|\tilde{\varphi}\| \geq \tilde{\varphi}(\mathbf{1} - f) = \tilde{\varphi}(\mathbf{1}) - \tilde{\varphi}(f) > \tilde{\varphi}(\mathbf{1}).$$

But we know $\|\tilde{\varphi}\| = \tilde{\varphi}(\mathbf{1})$.

18. A linear functional φ on $C_{\mathbb{R}}[0, 1]$ is positive if and only if

$$\varphi(f) = \int f dg$$

for some monotonically increasing function g on $[0, 1]$.

To prove this choose the g given by the Riesz Representation Theorem. Let $0 \leq t_1 < t_2 \leq 1$. Then

$$g(t_2) - g(t_1) = \tilde{\varphi}(\chi_{t_2} - \chi_{t_1}) \geq 0.$$

A linear functional φ on $C_{\mathbb{R}}[0, 1]$ is called *unital* if $\varphi(\mathbf{1}) = 1$. We have proved that a linear functional φ on $C_{\mathbb{R}}[0, 1]$ is positive and unital if and only if there exists a probability measure μ on $[0, 1]$ such that

$$\varphi(f) = \int f d\mu.$$

(A probability measure on X is a measure such that $\mu(X) = 1$.)

Exercises

19. For $0 \leq t \leq 1$ let φ_t be the linear functional on $C[0, 1]$ defined as

$$\varphi_t(f) = f(t).$$

Find the function g in $BVN[0, 1]$ that corresponds to φ_t according to the Riesz Representation Theorem.

20. Show that the space $BVN[0, 1]$ is not separable. This is another example of a situation where a separable Banach space has a nonseparable dual.

Some Applications

The Montel–Helly Selection Principle

1. **Theorem.** Let μ_n be a sequence of probability measures on $[0, 1]$. Then there exists a subsequence μ_m and a probability measure μ such that

$$\int f d\mu_m \rightarrow \int f d\mu \quad \text{as } m \rightarrow \infty$$

for all $f \in C[0, 1]$.

(In a terminology that we will learn later, this says that the set of probability measures on $[0, 1]$ is weak* compact)

Proof. Let $\{f_j\}$ be a dense set in $C[0, 1]$. Since $|\int f_1 d\mu_n| \leq \|f_1\|_\infty$ for all n , the sequence $\{\int f_1 d\mu_n : n \in \mathbb{N}\}$ is a bounded sequence of complex numbers. Hence there exists a subsequence μ_{n_1} of μ_n such that $\int f_1 d\mu_{n_1}$ converges. By the same argument, there exists a subsequence μ_{n_2} of μ_{n_1} such that $\int f_2 d\mu_{n_2}$ converges. By the diagonal procedure we get a subsequence μ_m of μ_n such that for each j , the sequence $\int f_j d\mu_m$ converges as $m \rightarrow \infty$.

Using an $\varepsilon/3$ -argument, one can see that for every f in $C[0, 1]$ the sequence $\int f d\mu_m$ is Cauchy and hence convergent. Let $\Lambda(f) = \lim_{m \rightarrow \infty} \int f d\mu_m$. Then Λ is a linear functional on $C[0, 1]$, $|\Lambda(f)| \leq \|f\|_\infty$, and $\Lambda(1) = 1$. By the Riesz Representation Theorem, there exists a probability measure μ such that $\Lambda(f) = \int f d\mu$.

■

Positive definite sequences

2. Let μ be any positive finite measure on the interval $[-\pi, \pi]$. Let

$$a_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-inx} d\mu(x). \quad (8.1)$$

The sequence $\{a_n\}_{n \in \mathbb{Z}}$ is called the *Fourier-Stieltjes sequence* corresponding to μ .

Let N be any natural number, and pick up any complex numbers $z_0, \dots, z_{N-1}$.

Then

$$\begin{aligned} & \sum_{0 \leq r, s \leq N-1} a_{r-s} z_r \bar{z}_s \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{0 \leq r, s \leq N-1} e^{i(s-r)x} z_r \bar{z}_s d\mu(x) \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{r=0}^{N-1} e^{-irx} z_r \right|^2 d\mu(x). \end{aligned}$$

Hence,

$$\sum_{0 \leq r, s \leq N-1} a_{r-s} z_r \bar{z}_s \geq 0. \quad (8.2)$$

3. A doubly infinite sequence $\{a_n\}_{n \in \mathbb{Z}}$ is said to be *positive definite* if the inequality (8.2) is satisfied for all N and for all choices of N complex numbers $z_0, \dots, z_{N-1}$.

4. The condition (8.2) can be expressed by saying that for all N , the matrices

$$\begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & a_{-(N-1)} \\ a_1 & a_0 & a_{-1} & & \cdot \\ a_2 & a_1 & a_0 & \cdot & \cdot \\ \cdot & a_2 & a_1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{N-1} & a_{N-2} & & & a_0 \end{bmatrix}$$

are positive semidefinite.

Thus the terms of a positive definite sequence must satisfy the following condi-

tions

- (i) $a_{-n} = \bar{a}_n$ for all n ,
- (ii) $a_0 \geq 0$,
- (iii) $|a_n| \leq a_0$ for all n .

5. We have seen that the Fourier–Stieltjes sequence associated with a positive finite measure on $[-\pi, \pi]$ is a positive definite sequence. One of the basic theorems of harmonic analysis says that every positive definite sequence arises in this way.

The Herglotz Theorem

6. **Theorem.** Let $\{a_n\}_{n \in \mathbb{Z}}$ be a positive definite sequence such that

$$a_0 = 1. \tag{8.3}$$

Then there exists a probability measure μ on $[-\pi, \pi]$ such that

$$a_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-inx} d\mu(x).$$

(Note that (8.3) is just a convenient normalisation.)

Proof. The inequalities (8.2) are valid for all N , and for all complex numbers $z_0, \dots, z_{N-1}$. Make a special choice $z_r = e^{irx}$, $0 \leq r \leq N-1$, where x is any real number. We have

$$\sum_{0 \leq r, s \leq N-1} a_{r-s} e^{i(r-s)x} \geq 0.$$

This inequality can be stated in an equivalent form

$$\sum_{k=-N+1}^{N-1} (N - |k|) a_k e^{ikx} \geq 0.$$

For each natural number N , let

$$f_N(x) = \sum_{k=-N+1}^{N-1} \left(1 - \frac{|k|}{N}\right) a_k e^{ikx}.$$

Then $f_N(x) \geq 0$ for all x , and

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} f_N(x) dx = a_0 = 1.$$

If E is any measurable subset of $[-\pi, \pi]$, let $\mu_N(E) = \frac{1}{2\pi} \int_E f_N(x) dx$. Then μ_N is a probability measure. By the Montel–Helly Principle there exists a subsequence μ_N and a probability measure μ such that for all f in $C[-\pi, \pi]$, $\int f d\mu_N$ converges to $\int f d\mu$ as $N \rightarrow \infty$.

In particular

$$\begin{aligned} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-inx} d\mu(x) &= \lim_{N \rightarrow \infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-inx} d\mu_N(x) \\ &= \lim_{N \rightarrow \infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-inx} f_N(x) dx \\ &= \lim_{N \rightarrow \infty} \left(1 - \frac{|n|}{N}\right) a_n \\ &= a_n. \end{aligned}$$

This proves the theorem. ■

Holomorphic maps of the disk into a half-plane

7. Let $\{a_n\}_{n \in \mathbb{Z}}$ be a positive definite sequence. Consider the power series

$$f(z) = \frac{a_0}{2} + a_1 z + a_2 z^2 + \cdots \quad (8.4)$$

Since $|a_n| \leq a_0$ for all n , this series converges in the unit disk $D = \{z : |z| < 1\}$. For every z in D we have

$$\begin{aligned} \frac{2 \operatorname{Re} f(z)}{1 - |z|^2} &= \frac{f(z) + \overline{f(z)}}{1 - z\bar{z}} \\ &= \sum_{m=0}^{\infty} z^m \bar{z}^m \left\{ \sum_{k=0}^{\infty} a_k z^k + \sum_{k=1}^{\infty} a_{-k} \bar{z}^k \right\} \\ &= \sum_{m=0}^{\infty} \sum_{k=0}^{\infty} a_k z^{m+k} \bar{z}^m + \sum_{m=0}^{\infty} \sum_{k=1}^{\infty} a_{-k} z^m \bar{z}^{m+k} \end{aligned}$$

$$\begin{aligned}
&= \sum_{s=0}^{\infty} \sum_{r=s}^{\infty} a_{r-s} z^r \bar{z}^s + \sum_{r=0}^{\infty} \sum_{s=r+1}^{\infty} a_{r-s} z^r \bar{z}^s \\
&= \sum_{r,s=0}^{\infty} a_{r-s} z^r \bar{z}^s.
\end{aligned}$$

This last sum is positive because the sequence $\{a_n\}$ is positive definite. Thus the function f defined by (8.4) is a holomorphic map of D into the right half plane (RHP).

It is a remarkable fact of complex analysis that conversely, if the function f maps D into the RHP then the coefficients of its power series lead to a positive definite sequence.

8. Theorem. Every holomorphic function f mapping D into the RHP can be represented as

$$f(z) = i\nu + \int_{-\pi}^{\pi} \frac{e^{it} + z}{e^{it} - z} d\alpha(t), \quad (8.5)$$

where $\nu = \operatorname{Im} f(0)$, and α is a monotonically increasing function on $[-\pi, \pi]$.

The expression (8.5) is called the *Riesz-Herglotz integral representation*.

9. What does this theorem say? Let $\mathcal{C}$ be the collection of all holomorphic functions mapping D into the RHP. Constant functions (with values in the RHP) are in $\mathcal{C}$. It is easy to check that for each t in $[-\pi, \pi]$ the function

$$H_t(z) = \frac{e^{it} + z}{e^{it} - z} \quad (8.6)$$

is in $\mathcal{C}$. Positive linear combinations of functions in $\mathcal{C}$ are again in $\mathcal{C}$. So are limits of functions in $\mathcal{C}$. The formula (8.5) says that all functions in $\mathcal{C}$ can be obtained from the family (8.6) by performing these operations. (An integral is a limit of finite sums.)

10. The theorem can be proved using standard complex analysis techniques like contour integration. See D. Sarason, *Notes on Complex Function Theory*, TRIM,

Hindustan Book Agency, p.161–162. The proof we give uses the Hahn–Banach and the Riesz Representation Theorems.

A *trigonometric polynomial* is a function

$$g(\theta) = \frac{a_0}{2} + \sum_{n=1}^N (a_n \cos n\theta + b_n \sin n\theta), \quad a_n, b_n \in \mathbb{R}. \quad (8.7)$$

The numbers a_n, b_n are called the *Fourier coefficients* of g , and are uniquely determined by g . The collection of all such functions is a vector space, and is dense in $C_{\mathbb{R}}[-\pi, \pi]$. For brevity, we will write

$$u_n(\theta) = \cos n\theta, \quad v_n(\theta) = \sin n\theta.$$

11. Proof of the Theorem. Let f be a holomorphic function on D . Let $f(z) = \sum_{n=0}^{\infty} c_n z^n$ be its power series expansion. Let α_n, β_n be the real and imaginary parts of c_n , and let $z = re^{i\theta}$ be the polar form of z . Then

$$\operatorname{Re} f(z) = \alpha_0 + \sum_{n=1}^{\infty} r^n (\alpha_n u_n(\theta) - \beta_n v_n(\theta)). \quad (8.8)$$

If g is a trigonometric polynomial as in (8.7), let

$$\Lambda(g) = \frac{\alpha_0 a_0}{2} + \frac{1}{2} \sum_{n=1}^N (\alpha_n a_n - \beta_n b_n).$$

Then Λ is a linear functional on the space of trigonometric polynomials, and

$$\Lambda(1) = \alpha_0, \quad \Lambda(u_n) = \frac{\alpha_n}{2}, \quad \Lambda(v_n) = -\frac{\beta_n}{2}. \quad (8.9)$$

Note that

$$\|r^n(\alpha_n u_n - \beta_n v_n)\|_{\infty} \leq 2|c_n|r^n.$$

Since $\sum_{n=0}^{\infty} |c_n|r^n$ is convergent, this shows that the series in (8.8) is uniformly convergent on $[-\pi, \pi]$. So, from (8.7) and (8.8), integrating term by term and using orthogonality of the trigonometric functions, we obtain

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} g(\theta) \operatorname{Re} f(re^{i\theta}) d\theta = \frac{\alpha_0 a_0}{2} + \frac{1}{2} \sum_{n=1}^N r^n (\alpha_n a_n - \beta_n b_n).$$

Hence

$$\Lambda(g) = \lim_{r \rightarrow 1} \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\theta) \operatorname{Re} f(re^{i\theta}) d\theta.$$

This shows that $\Lambda(g) \geq 0$ if $g \geq 0$ (recall f maps D into the RHP). By continuity, Λ can be extended to a positive linear functional on all of $C_{\mathbb{R}}[-\pi, \pi]$. We have

$$\|\Lambda\| = \Lambda(\mathbf{1}) = \alpha_0.$$

By the Riesz Representation Theorem, there exists a monotonically increasing function α on $[-\pi, \pi]$ such that

$$\Lambda(g) = \int_{-\pi}^{\pi} g(t) d\alpha(t) \quad \text{for all } g \in C_{\mathbb{R}}[-\pi, \pi].$$

We can define a linear functional $\tilde{\Lambda}$ on the space $C[-\pi, \pi]$ of complex functions by putting

$$\tilde{\Lambda}(g_1 + ig_2) = \Lambda(g_1) + i\Lambda(g_2), \quad g_1, g_2 \in C_{\mathbb{R}}[-\pi, \pi].$$

We then have

$$\tilde{\Lambda}(g) = \int_{-\pi}^{\pi} g(t) d\alpha(t) \quad \text{for all } g \in C[-\pi, \pi].$$

Now for each $z \in D$ look at the function

$$\begin{aligned} H_z(t) &:= \frac{e^{it} + z}{e^{it} - z} = 1 + \frac{2ze^{-it}}{1 - ze^{-it}} = 1 + 2 \sum_{n=1}^{\infty} z^n e^{-int} \\ &= 1 + 2 \sum_{n=1}^{\infty} z^n \{u_n(t) - iv_n(t)\}. \end{aligned} \tag{8.10}$$

Use (8.9) to get

$$\tilde{\Lambda}(H_z) = \alpha_0 + \sum_{n=1}^{\infty} (\alpha_n + i\beta_n) z^n = \sum_{n=0}^{\infty} (\alpha_n + i\beta_n) z^n - i\beta_0 = f(z) - i \operatorname{Im} f(0).$$

So,

$$f(z) = i \operatorname{Im} f(0) + \tilde{\Lambda}(H_z) = i \operatorname{Im} f(0) + \int_{-\pi}^{\pi} \frac{e^{it} + z}{e^{it} - z} d\alpha(t).$$

■

12. Corollary. Let $f(z) = c_0 + c_1 z + c_2 z^2 + \dots$ be a holomorphic function mapping D into the RHP. Let $\{a_n\}_{n \in \mathbb{Z}}$ be the sequence in which $a_0 = 2 \operatorname{Re} c_0$, $a_n = c_n$, $a_{-n} = \bar{c}_n$ for $n \geq 1$. Then $\{a_n\}$ is a positive definite sequence.

Proof. The integral formula (8.5) shows that

$$f(z) = \frac{1}{2}(c_0 - \bar{c}_0) + \int_{-\pi}^{\pi} \frac{e^{it} + z}{e^{it} - z} d\alpha(t).$$

Expanding the integrand as the (first) series in (8.10), this gives

$$f(z) = \left[\frac{1}{2}(c_0 - \bar{c}_0) + \int_{-\pi}^{\pi} d\alpha(t) \right] + 2 \sum_{n=1}^{\infty} \left[\int_{-\pi}^{\pi} e^{-int} d\alpha(t) \right] z^n.$$

By the uniqueness of the coefficients of a power series

$$\begin{aligned} a_0 &= 2 \int_{-\pi}^{\pi} d\alpha(t) \\ a_n &= 2 \int_{-\pi}^{\pi} e^{-int} d\alpha(t). \end{aligned}$$

Thus the sequence $\{a_n\}_{n \in \mathbb{Z}}$ is positive definite. ■

13. The Riesz–Herglotz Integral Representation plays a central role in the theory of matrix monotone functions. See R. Bhatia, *Matrix Analysis*, Chapter V.

The Weak Topology

When we say that a sequence f_n in the space $C[0, 1]$ converges to f , we mean that $\|f_n - f\| \rightarrow 0$ as $n \rightarrow \infty$; and this is the same as saying f_n converges to f uniformly. There are other notions of convergence that are weaker, and still very useful in analysis. This is the motivation for studying different topologies on spaces of functions, and on general Banach spaces.

The weak topology

1. Let S be any set and let $(T, \mathcal{U})$ be a topological space. Let $\mathcal{F}$ be a family of maps from S into T . The *weak topology on S generated by $\mathcal{F}$* (or the $\mathcal{F}$ -weak topology) is the weakest (i.e., the smallest) topology on S for which all $f \in \mathcal{F}$ are continuous.

Exercise. The collection

$$\left\{ \bigcap_{j=1}^k f_j^{-1}(U_j) : U_j \in \mathcal{U}, f_j \in \mathcal{F}, 1 \leq j \leq k, k = 1, 2, \dots \right\}.$$

is a base for this topology.

2. **Examples.** 1. Let $C[a, b]$ be the space of all continuous functions on $[a, b]$. For each $x \in [a, b]$ the map $E_x(f) = f(x)$ is a map from $C[a, b]$ to $\mathbb{C}$, called the *evaluation map*. The weak topology generated by $\{E_x : x \in [a, b]\}$ is called the *topology of pointwise convergence* on $C[a, b]$.

2. The *product topology* on $\mathbb{R}^n$ or $\mathbb{C}^n$ is the weak topology generated by the projection maps π_j defined as $\pi_j(x_1, \dots, x_n) = x_j, 1 \leq j \leq n$.

3. More generally, if X_α is any family of topological spaces the *product topology* on the Cartesian product $\prod_\alpha X_\alpha$ is the weak topology generated by the projections π_α onto the components X_α .

3. Now let X be any Banach space and let X^* be its dual space. The weak topology on X generated by X^* is called *the* weak topology on X . For this topology, the sets

$$N(f_1, \dots, f_k; \varepsilon) = \{x : |f_i(x)| < \varepsilon, 1 \leq i \leq k\},$$

where $\varepsilon > 0$, $k = 1, 2, \dots$, and $f_1, f_2, \dots, f_k$ are in X^* , form a neighbourhood base at the point 0. A base at any other point can be obtained from this by a translation.

4. For brevity, members of the weak topology on X are called *weakly open* sets. Phrases such as *weak neighbourhood*, *weak closure* etc. are used to indicate neighbourhoods and closures in the weak topology.

The topology on X given by its norm is called the *norm topology* or the *strong topology* or the *usual topology* on X ; the adjective chosen depends on the point of view to be emphasized at a particular moment.

A sequence x_n in X converges to x in the norm /strong/usual topology if $\|x_n - x\| \rightarrow 0$. We write this as $x_n \rightarrow x$. The sequence x_n converges to x in the weak topology if and only if $f(x_n)$ converges to $f(x)$ for all $f \in X^*$. We write this as $x_n \xrightarrow{w} x$, and say x_n *converges weakly* to x .

5. If $x_n \rightarrow x$ it is clear that $x_n \xrightarrow{w} x$. The converse is not always true.

Example. Let $X = L_2[-\pi, \pi]$. Then $X^* = X$. Let $v_n(t) = \sin nt$. Then for all f in X , we have $\lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} f(t) \sin nt \, dt = 0$ by the *Riemann-Lebesgue Lemma*. So, the sequence v_n converges weakly to the function 0. On the other hand

$$\|v_n\|^2 = \int_{-\pi}^{\pi} |\sin^2 nt| \, dt = \pi.$$

So v_n can not converge to 0 in norm.

6. **Exercise.** Show that the norm topology on X is stronger than the weak topology

(i.e., every weakly open set is open in the usual topology).

If X is finite-dimensional, then its weak topology is the same as the norm topology.

7. Exercise. The weak topology on X is a Hausdorff topology. (Hint: Use the Hahn–Banach Theorem.)

8. If a sequence $\{x_n\}$ in X is convergent, then it is bounded; i.e., there exists a positive number C such that $\|x_n\| \leq C$ for all n . This happens to be true even when $\{x_n\}$ is weakly convergent.

The proof that follows uses the Uniform Boundedness Principle, and a simple idea with far reaching consequences—turning duality around by regarding elements of X as linear functionals on X^* . Every element x of X induces a linear functional F_x on X^* defined as

$$F_x(f) = f(x) \quad \text{for all } f \in X^*.$$

It is clear that F_x is a linear functional on X^* , and the map $x \mapsto F_x$ is linear. It follows from the definition that $\|F_x\| \leq \|x\|$. The Hahn–Banach theorem implies the stronger assertion that $\|F_x\| = \|x\|$. (We can find an f in X^* with $\|f\| = 1$, and $f(x) = \|x\|$.)

Now suppose $\{x_n\}$ is a weakly convergent sequence. Then for each f in X^* , the sequence $\{f(x_n)\}$ is convergent, and hence bounded. This means that there exists a positive number C_f such that

$$\sup_n |f(x_n)| \leq C_f.$$

In the notation introduced above this says

$$\sup_n |F_{x_n}(f)| \leq C_f.$$

Hence by the Uniform Boundedness Principle, there exists a positive number C such that

$$\sup_n \|F_{x_n}\| \leq C,$$

which is the same as saying $\sup_n \|x_n\| \leq C$.

9. We will use this to show that the weak topology on ℓ_p , $1 < p < \infty$, can not be obtained from *any* metric.

Let e_n be the standard basis for ℓ_p , and let

$$S = \{n^{1/q}e_n : n = 1, 2, \dots\}.$$

This is the collection of all vectors of the form $(0, 0, \dots, n^{1/q}, 0, \dots)$, $n = 1, 2, \dots$.

We will show that the set S intersects every weak neighbourhood of 0 in ℓ_p . If V is such a neighbourhood, then it contains a basic open set

$$N(f^{(1)}, \dots, f^{(k)}; \varepsilon) = \{x \in \ell_p : |f^{(j)}(x)| < \varepsilon, 1 \leq j \leq k\},$$

where ε is a positive number, and $f^{(j)}$ are elements of ℓ_q . If $f^{(j)} = (f_1^{(j)}, f_2^{(j)}, \dots)$, then by definition, $f^{(j)}(x) = \sum_{n=1}^{\infty} f_n^{(j)} x_n$, for every $x \in \ell_p$. In particular,

$$f^{(j)}(n^{1/q}e_n) = n^{1/q}f_n^{(j)}, \quad \text{for all } n.$$

So, if the set S does not intersect V , then for some j we have $|n^{1/q}f_n^{(j)}| > \varepsilon$ for all n .

This implies that

$$\sum_{j=1}^k |f_n^{(j)}| > \frac{\varepsilon}{n^{1/q}}, \quad \text{for all } n.$$

If $y = (y_1, y_2, \dots)$ is any vector, let us use the notation $|y|$ for the vector $(|y_1|, |y_2|, \dots)$.

Clearly, if y is in ℓ_q , then so is $|y|$. For $1 \leq j \leq k$, each $|f^{(j)}|$ is in ℓ_q , and hence so is their sum $f = \sum_{j=1}^k |f^{(j)}|$. But if the last inequality were true we would have

$$\sum_{n=1}^{\infty} |f_n|^q \geq \sum_{n=1}^{\infty} \frac{\varepsilon^q}{n},$$

and that implies f cannot be in ℓ_q . This contradiction shows that S intersects V .

This is true for every weak neighbourhood V of 0. Hence 0 is a *weak accumulation point* of the set S .

Now if the weak topology of ℓ_p arose from a metric there should be a sequence

of elements of S converging (weakly) to 0. Such a sequence has to be norm bounded. However,

$$\|n^{1/q}e_n\|_p = n^{1/q}$$

and hence *no* sequence from S can be norm bounded.

10. A topology (a collection $\mathcal{U}$ of open sets) on a given space X is called *metrisable* if there exists a metric on X such that the open sets generated by this metric are exactly those that are members of $\mathcal{U}$.

We have seen that the weak topology on ℓ_p , $1 < p < \infty$, is not metrisable. In fact, the weak topology on *any* infinite-dimensional Banach space is not metrisable. We will prove this a little later.

Nets

11. We have seen that in a topological space that is not metrisable, sequences might not be adequate to detect accumulation points. The remedy lies in the introduction of *nets*. Reasoning with nets is particularly useful in problems of functional analysis.

A partially ordered set I , with partial order $\prec$, is called a *directed set* if for all $\alpha, \beta \in I$, there exists $\gamma \in I$ such that $\alpha \prec \gamma$ and $\beta \prec \gamma$.

The sets $\mathbb{N}$ and $\mathbb{R}$ with their usual orders are directed sets. The collection of all subsets of a given set with set inclusion as the partial order is a directed set. Let I be the collection of all neighbourhoods of a point x in a topological space X . Say $N_1 \prec N_2$ if $N_2 \subset N_1$. Then I is a directed set.

Let X be a topological space. A *net* in X is a map $\alpha \mapsto x_\alpha$ from a directed set I to X . (When $I = \mathbb{N}$ this is just a sequence in X .) Sometimes we denote the net by $\{x_\alpha\}_{\alpha \in I}$ or simply by x_α .

12. We say that a net $\{x_\alpha\}_{\alpha \in I}$ *eventually* satisfies a property P , if there exists $\gamma \in I$

such that the property P is satisfied by all x_α with $\gamma \prec \alpha$. We say that $\{x_\alpha\}_{\alpha \in I}$ *frequently* satisfies P , if for each $\gamma \in I$, there exists an α such that $\gamma \prec \alpha$ and x_α satisfies the property P .

We say that the net x_α *converges* to x ($x_\alpha \rightarrow x$) if for each neighbourhood N of x , x_α is eventually in N . A point x is called a *cluster point* (or an *accumulation point*) of the net $\{x_\alpha\}$ if x_α is frequently in each neighbourhood of x .

13. Proposition. Let E be a subset of a topological space X . Then a point x is in the closure of E if and only if there exists a net $\{x_\alpha\}$ in E that converges to x .

Proof. If a net x_α in E converges to x , then each neighbourhood of x contains an element of E . So $x \in \bar{E}$, the closure of E . To see the converse, suppose $x \in \bar{E}$ and let I be the collection of all neighbourhoods of x with the partial order $N_1 \prec N_2$ defined to mean $N_2 \subset N_1$. Given $N \in I$, there exists a point x_N in $E \cap N$. Then $\{x_N\}_{N \in I}$ is a net that converges to x . ■

14. Exercises. 1. If X is a Hausdorff space then a net x_α in X can converge to at most one limit. (The converse is also true.)

2. A map f from a topological space X into another topological space Y is continuous if and only if the net $f(x_\alpha)$ converges to $f(x)$ in Y whenever the net x_α converges to x in X .

15. Let $\{x_\alpha\}_{\alpha \in I}$ and $\{y_\beta\}_{\beta \in J}$ be two nets. We say $\{x_\alpha\}$ is a *subnet* of $\{y_\beta\}$, if there exists a function $F : I \rightarrow J$ such that

$$(i) \quad x_\alpha = y_{F(\alpha)} \quad \text{for all } \alpha \in I.$$

$$(ii) \quad \text{For each } \beta \in J, \text{ there exists } \alpha \in I, \text{ such that } \beta \prec F(\alpha') \text{ if } \alpha \prec \alpha'.$$

(The second condition says that $F(\alpha)$ is eventually larger than each β in J .)

16. Exercises. 1. Every subsequence of a sequence is also a subnet of it. (But every subnet need not be a subsequence.)

2. A point x is a cluster point of a net $\{x_\alpha\}$ if and only if a subnet of $\{x_\alpha\}$ converges to x .

17. **Theorem** (Bolzano–Weierstrass Theorem). A topological space X is compact if and only if every net in X has a convergent subnet.

18. **The Tychonoff Theorem.** If $\{X_\alpha\}$ is any family of compact topological spaces, then the product topological space $\prod_\alpha X_\alpha$ is compact.

19. **Warning.** All this might suggest that everything is simple. We have to merely replace the subscript n in x_n by α and pretend nothing else has changed. This is not so. Here are two of the pitfalls.

(i) A net in a normed space may be convergent without being bounded. (Have we seen an example already?)

(ii) A sequence may have a convergent subnet without having any convergent subsequence. (We will soon see an example.)

20. Though the weak topology on an infinite-dimensional Banach space X is not metrisable, it is possible that some useful *subsets* of X could be metrisable. For example, if X^* is separable, then the unit ball of X with the weak topology is metrisable. We will prove this in a special case later.

The Second Dual and the Weak* Topology

The Second Dual and Reflexivity

1. The dual of X^* is another Banach space X^{**} . This is called the *second dual* or the *bidual* of X . Let J be the map from X into X^{**} that associates with $x \in X$ the element $F_x \in X^{**}$ defined as

$$F_x(f) = f(x) \text{ for all } f \in X^*.$$

Then J is a linear map and $\|Jx\| = \|x\|$. (See (9.2).) Thus J is an *isometric imbedding* and we can regard X as a subspace of X^{**} .

2. If the map J is *surjective*, then X is isomorphic to X^{**} via the map J , and we say that X is *reflexive*.

Note that we are demanding not just that X be isomorphic to X^{**} ; we want the natural map J to be an isomorphism. There is an example where the spaces X and X^{**} are isomorphic but the natural map J is not an isomorphism. Such spaces are not reflexive.

Every finite-dimensional space is reflexive. The ℓ_p spaces are reflexive for $1 < p < \infty$, but not for $p = 1, \infty$.

3. Show that a Banach space X is reflexive if and only if X^* is reflexive.

The weak* topology

4. Let X^* be the dual of a Banach space X . The usual topology on X^* is the one generated by its norm. Its weak topology is the weak topology generated by its dual X^{**} . There is one more topology on X^* that is useful. This is the weak topology on X^* generated by the subspace X of X^{**} ; i.e., the weakest (smallest) topology on X^* for which every element of X , acting as a linear functional on X^* , is continuous. This is called the *weak* topology* on X^* .

5. Note that a net f_α in X^* converges to f in the weak* topology if and only if $f_\alpha(x) \rightarrow f(x)$ for all $x \in X$. So, this is the *topology of pointwise convergence*.

The weak* topology is weaker than the weak topology on X^* .

If X is reflexive, then the weak topology on X^* is the same as the weak* topology.

6. **The Banach–Alaoglu Theorem.** Let X be any Banach space. Then the unit ball $\{f \in X^* : \|f\| \leq 1\}$ in the space X^* is compact in the weak* topology.

(This is the most important theorem on weak* topology.)

Proof. For each $x \in X$ consider the set $B_x = \{z \in \mathbb{C} : |z| \leq \|x\|\}$. This is a compact subset of the complex plane. Consider the space

$$B := \prod_{x \in X} B_x$$

with the product topology. By Tychonoff's Theorem B is compact.

What are elements of B ? They are maps b from X into $\cup_x B_x$ such that $b(x)$ is in B_x for each $x \in X$; i.e., they are maps $b : X \rightarrow \mathbb{C}$ such that $|b(x)| \leq \|x\|$. Among these the linear maps are exactly the elements of the unit ball $\mathcal{B}$ of X^* .

If we show $\mathcal{B}$ is a closed subset of B it will follow that $\mathcal{B}$ too is compact in the topology it inherits from B . But this inherited topology is the topology of pointwise convergence; this is the same as the weak* topology.

Let f_α be any net in $\mathcal{B}$ and suppose f_α converges to an element f of B . We have to show that $f \in \mathcal{B}$. Note that

$$\begin{aligned} f(a_1x + a_2y) &= \lim f_\alpha(a_1x + a_2y) \\ &= \lim(a_1f_\alpha(x) + a_2f_\alpha(y)) \\ &= a_1f(x) + a_2f(y). \end{aligned}$$

Thus f is linear. Since $f \in B$, we already know $|f(x)| \leq \|x\|$. Thus $\|f\| \leq 1$. So $f \in \mathcal{B}$. ■

7. If X is reflexive, the unit ball of X^* is weakly compact. (The weak topology and the weak* topology are the same in this case.)

If $X = X^*$ (as is the case when X is ℓ_2 or L_2) then the unit ball of X is weakly compact.

Recall that the unit ball of any infinite-dimensional space can not be compact in the strong (usual) topology. This weaker compactness can still be very useful.

It can be proved that a Banach space is reflexive if and only if the weak and the weak* topologies coincide.

8. The Montel–Helly Selection Principle is a special instance of the Banach–Alaoglu Theorem.

9. **Theorem.** Every Banach space is isometrically isomorphic to a closed linear subspace of the space $C(\mathcal{X})$ of continuous functions on a compact Hausdorff space $\mathcal{X}$.

Proof. Let $\mathcal{X}$ be the closed unit ball of the dual space X^* with the weak* topology. We have seen that $\mathcal{X}$ is compact. Every element x of X can be thought of as a continuous function on $\mathcal{X}$. ■

Earlier we saw that every *separable* Banach space is isomorphic to a subspace of ℓ_∞ . In this theorem the condition of separability has been dropped. If the Banach space X is separable, then the space $\mathcal{X}$ in Theorem 9 is the Stone-Cech compactification of $\mathbb{N}$.

Exercises

10. Show that the only linear functionals on X^* that are weak* continuous are the elements of X .

The only linear functionals on X that are weakly continuous are the elements of X^* . (Thus a linear functional on X is weakly continuous if and only if it is strongly continuous.)

11. A subset of X whose linear span is dense in X is called a *fundamental set*.

Show that $x_n \xrightarrow{w} x$ if and only if $\{\|x_n\|\}$ is bounded and $f(x_n) \rightarrow f(x)$ for every element f of a fundamental set in X^* .

12. Let $1 < p < \infty$. Show that a sequence $\{x_n\}$ in ℓ_p converges weakly to x if and only if $\{\|x_n\|_p\}$ is bounded and x_n converges coordinatewise to x ; i.e., if $x_n = (\xi_1^{(n)}, \xi_2^{(n)}, \dots)$ and $x = (\xi_1, \xi_2, \dots)$, then for each j the sequence $\xi_j^{(n)}$ converges to ξ_j as $n \rightarrow \infty$.

13. A sequence $\{f_n\}$ in X^* is weak* convergent if and only if $\{\|f_n\|\}$ is bounded and $\{f_n(x)\}$ is a Cauchy sequence for each x in some fundamental set in X .

Annihilators

14. Let S be any subset of a Banach space X , and let

$$S^\perp = \{f \in X^* : f(x) = 0 \text{ for all } x \in S\}.$$

Then $S^\perp$ is a (closed linear) subspace of X^* . This is called the *annihilator* of S (the collection of all linear functionals that kill every element of S). If $[S]$ denotes the closed linear space spanned by the set S , then $S^\perp = [S]^\perp$. The notation $S^\perp$ suggests orthogonality, and indeed there are several similarities with that notion.

It is easy to see that $S^\perp = \{0\}$ if and only if S is a fundamental set in X .

15. Let X/M be the quotient of X by M . The dimension of this space is called the *codimension* of M in X . In symbols

$$\text{codim } M = \dim X/M.$$

Exercise. Show that if X is finite-dimensional, then $\dim X = \dim M + \text{codim } M$.

In the proof of the next theorem we use the following:

Proposition. Let X be any normed linear space and M a closed subspace of X . If N is any finite-dimensional subspace, then the sum $M + N$ is a closed subspace of X .

Proof. Let X/M be the quotient space, and $Q : X \rightarrow X/M$ the quotient map. The image $\tilde{N} = Q(N)$ is finite-dimensional, and hence closed in X/M . Since Q is continuous, $Q^{-1}(\tilde{N})$ is closed in X . But $Q^{-1}(\tilde{N}) = M + N$. ■

16. **Theorem.** Let M be any closed subspace of a Banach space X . Then

$$\text{codim } M = \dim M^\perp, \tag{10.1}$$

in the sense that either both sides are infinite, or they are finite and equal.

Proof. Suppose $\text{codim } M$ is a finite number m . Let $\tilde{X} = X/M$. This is an m -dimensional space; choose a basis $\tilde{x}_1, \dots, \tilde{x}_m$ in it. Let u be any element of X , and

let $\tilde{u}$ be the element of $\tilde{X}$ corresponding to u . We can write

$$\tilde{u} = \xi_1 \tilde{x}_1 + \cdots + \xi_m \tilde{x}_m.$$

Then we must have

$$u = \xi_1 x_1 + \cdots + \xi_m x_m + v$$

for some $v \in M$. Thus the space M , together with the vectors $x_1, \dots, x_m$, spans X .

Let M_j be the subspace spanned by M and the vectors $x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_m$. Then $M_j, 1 \leq j \leq m$, are closed subspaces of X containing M . By the Hahn-Banach Theorem we can find f_j in X^* such that $f_j(M_j) = 0$ and $f_j(x_j) = 1$. We thus have a collection $f_j, 1 \leq j \leq m$, such that

$$f_j(M) = 0, \quad f_j(x_k) = \delta_{jk}. \quad (10.2)$$

This last condition shows that f_j are linearly independent. We will see that they form a basis for $M^\perp$.

Let f be any element of $M^\perp$. For $k = 1, 2, \dots, m$, we can write

$$f(x_k) = \sum_{j=1}^m f(x_j) f_j(x_k)$$

because of (10.2). Since M and the x_k span X , this shows

$$f = \sum_{j=1}^m f(x_j) f_j.$$

Thus $f_j, 1 \leq j \leq m$ form a basis for $M^\perp$. So $\dim M^\perp = m$.

Now suppose $\text{codim } M = \infty$. Choose a vector x_1 in X not in M . Let M_1 be the space spanned by M and x_1 . Choose a vector x_2 in X not in M_1 . Let M_2 be the space spanned by M_1 and x_2 . Since $\text{codim } M = \infty$, this process can be continued indefinitely leading to a strictly increasing sequence of closed subspaces

$$M \subset M_1 \subset M_2 \subset \cdots \subset X.$$

The inclusions

$$X^* \supset M^\perp \supset M_1^\perp \supset \cdots \supset \{0\}$$

are strict inclusions because of the Hahn–Banach Theorem. Thus $M^\perp$ is infinite dimensional.

We have proved the equality (10.1). ■

17. **Exercise.** Prove that

$$\dim M = \operatorname{codim} M^\perp. \quad (10.3)$$

18. **Corollary.** If M is a finite-dimensional subspace of X , then $M^{\perp\perp} = M$.

Proof. Note that M is a subspace of X , whereas $M^{\perp\perp}$ is a subspace of X^{**} . So the asserted equality is to be interpreted as an isomorphism. To prove it note that $M \subset M^{\perp\perp}$. By the equalities (10.1) and (10.3)

$$\dim M = \operatorname{codim} M^\perp = \dim M^{\perp\perp}.$$

If $\dim M < \infty$, this is sufficient to conclude that $M = M^{\perp\perp}$. ■

19. **Exercise.** If $f_1, \dots, f_n$ are linearly independent elements of X^* , then there exist $x_1, \dots, x_n$ in X such that $f_i(x_j) = \delta_{ij}$.

20. **Exercise.** The kernel of a linear functional is a vector space of codimension one. So, if X is infinite-dimensional and $f_1, \dots, f_k$ are linear functionals on it, then the set

$$\{x : f_j(x) = 0, 1 \leq j \leq k\}$$

is an infinite-dimensional subspace of X . Thus any weak neighbourhood of 0 contains an infinite-dimensional subspace (and hence a line through the origin).

Use this to show that in an infinite-dimensional Banach space the weak closure of the sphere $\{x : \|x\| = 1\}$ contains the point 0 and the ball $\{x : \|x\| \leq 1\}$. This shows how weak the weak topology is.

21. **Exercise.** The observation in Exercise 20 can be used to show that if X is an infinite-dimensional Banach space, then its weak topology is not metrisable. The weak closure of every sphere $S_n = \{x : \|x\| = n\}$ contains 0. If the weak topology arose from a metric d , then for each n , there would exist x_n in S_n such that $d(x_n, 0) < 1/n$. The sequence $\{x_n\}$ is then unbounded (in norm) and, at the same time, weakly convergent (to 0). That is not possible.

Hilbert Spaces

To each vector in the familiar Euclidean space we assign a length, and to each pair of vectors an angle between them. The first notion has been made abstract in the definition of a norm. What is missing in the theory so far is an appropriate concept of angle and the associated notion of *orthogonality*. These ideas depend on the introduction of an *inner product*. Hilbert spaces are special kinds of Banach spaces whose norms arise from inner products.

They are important for several reasons. Since they have a special structure, more can be said about them; so we have a richer theory. They are important for applications: quantum mechanics, signal processing, wavelets. Our intuition works better with them. Functional analysis leads us to use our geometric intuition to understand the behaviour of spaces of functions. Typically, to understand convergence of a sequence of functions an analyst might draw a picture on paper where the functions are represented by points; then argue by analogy to extend her reasoning from $\mathbb{R}^2$ to a Hilbert space and finally see whether the special nature of Hilbert space can be dispensed with.

Hilbert space is named after David Hilbert, one of the great mathematicians of the twentieth century. He used the space ℓ_2 in his study of problems concerned with integral equations. The general theory of Hilbert spaces was developed later by John von Neumann in 1928 to provide a mathematical framework for quantum mechanics.

Basic notions

1. A complex vector space X is called an *inner product space* if to each pair of

elements x, y of X is associated a complex number $\langle x, y \rangle$, called the *inner product* of x and y , that satisfies the following four conditions

- (i) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ for all $x, y, z \in X$.
- (ii) $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$ for all $\alpha \in \mathbb{C}$.
- (iii) $\langle x, y \rangle = \overline{\langle y, x \rangle}$ (the bar denotes complex conjugation).
- (iv) $\langle x, x \rangle \geq 0$, and $\langle x, x \rangle = 0$ if and only if $x = 0$.

Note that conditions (i), (ii) and (iii) imply that

- (v) $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$.
- (vi) $\langle x, \alpha y \rangle = \bar{\alpha} \langle x, y \rangle$.

Thus the function $\langle \cdot, \cdot \rangle$ is *linear* in the first variable and *conjugate linear* in the second.

This convention is followed by almost all functional analysis books, and violated by almost all physics and matrix theory books. The latter take the inner product to be conjugate linear in the first variable and linear in the second.

2. A real vector space X is called an inner product space if there is defined a *real* function $\langle \cdot, \cdot \rangle$ on $X \times X$ satisfying the properties (i)–(iv). The complex conjugation in (iii) and (vi) is redundant in this case. We will talk of complex inner product spaces most of the time. What we say is often true for real inner product spaces. (The latter parts, involving spectral theory, require the underlying field to be complex.)

3. The space $\mathbb{C}^n$ is an inner product space with the usual definition

$$\langle x, y \rangle = \sum_{j=1}^n x_j \bar{y}_j. \quad (11.1)$$

The space $C[a, b]$ is an inner product space if we define

$$\langle f, g \rangle = \int_a^b f(x) \overline{g(x)} dx. \quad (11.2)$$

4. Given an inner product on X , put

$$\|x\| := \langle x, x \rangle^{1/2}. \quad (11.3)$$

Then $\|\cdot\|$ defines a norm on X . To verify the triangle inequality one needs:

The Cauchy–Schwarz inequality.

$$|\langle x, y \rangle| \leq \|x\| \|y\|. \quad (11.4)$$

This can be proved as follows. If $y = 0$, both sides of (11.4) are zero. If $y \neq 0$, we may replace y by $y/\|y\|$, and thereby assume $\|y\| = 1$. Then

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

(Some times (11.4) is called just the *Schwarz inequality*.) Thus X is a normed linear space. The map $\langle x, y \rangle$ from $X \times X$ into $\mathbb{C}$ is continuous.

5. If the inner product space X with the norm (11.3) (*induced* by the inner product) is a complete metric space, we say it is a *Hilbert space*. An inner product space is some times called a *pre Hilbert space*. If it is not complete, then its completion is a Hilbert space.

We will use the symbol $\mathcal{H}$ for a Hilbert space .

6. The space ℓ_2 is a Hilbert space with the inner product

$$\langle x, y \rangle = \sum_{j=1}^{\infty} x_j \bar{y}_j. \quad (11.5)$$

The space $L_2[a, b]$ is a Hilbert space with the inner product

$$\langle f, g \rangle = \int_a^b f(x) \overline{g(x)} dx. \quad (11.6)$$

The finiteness of the sum in (11.5) and the integral in (11.6) are consequences of the Schwarz inequality. The completeness of $L_2[a, b]$ is the Riesz–Fischer theorem.

The space $C[a, b]$ is a dense subspace of $L_2[a, b]$. It is not a closed subspace.

Exercises

7. The norm in any inner product space satisfies the *parallelogram law*

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

8. Say that x is *orthogonal* to y (written $x \perp y$) if $\langle x, y \rangle = 0$. Show that in this case

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

This is called the *Pythagorean Theorem*.

9. The norm and the inner product are related by the *polarisation identity*

$$4\langle x, y \rangle = \|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2. \quad (11.7)$$

This can be written compactly as

$$\langle x, y \rangle = \frac{1}{4} \sum_{p=0}^3 i^p \|x + i^p y\|^2.$$

The polarisation identity allows us to recover the inner product from the norm.

In the case of a real inner product space the polarisation identity is

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

10. Suppose X is any normed linear space. From the norm on X we can define a function $\langle \cdot, \cdot \rangle$ on $X \times X$ by (11.7). Then $\langle x, x \rangle = \|x\|^2$. However, $\langle \cdot, \cdot \rangle$ defines an inner product on X if and only if the norm $\|\cdot\|$ satisfies the parallelogram law. (A bit of work is required for the proof of this statement.)

11. Let ω be a primitive n th root of unity, where $n > 2$. Show that

$$\langle x, y \rangle = \frac{1}{n} \sum_{p=0}^{n-1} \omega^p \|x + \omega^p y\|^2.$$

This is a generalisation of the polarisation identity.

12. For x, y, z in an inner product space

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

This is the *Appolonius Theorem*. It generalises the theorem with this name in plane geometry: if ABC is a triangle, and D is the mid-point of the side BC , then

$$(AB)^2 + (AC)^2 = 2[(AD)^2 + (BD)^2].$$

13. Let S be any subset of $\mathcal{H}$. Let

$$S^\perp = \{x \in \mathcal{H} : x \perp y \text{ for all } y \in S\}.$$

Show that

- (i) $S \cap S^\perp \subset \{0\}$.
- (ii) $S^\perp$ is a closed linear subspace of $\mathcal{H}$.
- (iii) $\{0\}^\perp = \mathcal{H}, \mathcal{H}^\perp = \{0\}$.
- (iv) If $S_1 \subset S_2$, then $S_2^\perp \subset S_1^\perp$.
- (v) $S \subset S^{\perp\perp}$.

Subspaces, direct sums and projections

14. **Theorem.** Let S be any closed convex subset of $\mathcal{H}$. Then for each x in $\mathcal{H}$ there exists a unique point x_0 in S such that

$$\|x - x_0\| = \text{dist}(x, S) := \inf_{y \in S} \|x - y\|.$$

Proof. Let $d = \text{dist}(x, S)$. Then there exists a sequence y_n in S such that $\|x - y_n\| \rightarrow d$. By the Appolonius Theorem

$$\begin{aligned} & \|x - y_n\|^2 + \|x - y_m\|^2 \\ &= 2 \left(\|x - \frac{1}{2}(y_n + y_m)\|^2 + \|\frac{1}{2}(y_n - y_m)\|^2 \right) \\ &\geq 2d^2 + \frac{1}{2}\|y_n - y_m\|^2. \end{aligned}$$

(We have used the convexity of S to conclude $\frac{1}{2}(y_n + y_m) \in S$.) As $n, m \rightarrow \infty$, the left hand side goes to $2d^2$. This shows $\{y_n\}$ is a Cauchy sequence. Since S is closed $x_0 := \lim y_n$ is in S and

$$\|x - x_0\| = \lim \|x - y_n\| = d.$$

If there is another point x_1 in S for which $\|x - x_1\| = d$, the same argument with the Appolonius Theorem shows that $x_1 = x_0$. ■

The theorem says that each point of $\mathcal{H}$ has a *unique best approximant* from any given closed convex set S . This is not true in all Banach spaces. Approximation problems in Hilbert spaces are generally easier because of this theorem.

15. Especially interesting is the case when S is a closed linear subspace. For each x in $\mathcal{H}$ let

$$P_S(x) = x_0, \tag{11.8}$$

where x_0 is the unique point in S closest to x . Then P_S is a well defined map with range S . If $x \in S$, then $P_S(x) = x$. Thus P_S is *idempotent*; i.e.,

$$P_S^2 = P_S. \tag{11.9}$$

For each y in S and t in $\mathbb{R}$, we have

$$\|x - (x_0 + ty)\|^2 \geq \|x - x_0\|^2.$$

From this we get

$$\|x - x_0\|^2 + t^2\|y\|^2 - 2t \operatorname{Re} \langle x - x_0, y \rangle \geq \|x - x_0\|^2,$$

i.e.,

$$t^2 \|y\|^2 \geq 2t \operatorname{Re} \langle x - x_0, y \rangle.$$

Since this is true for all real t we must have

$$\operatorname{Re} \langle x - x_0, y \rangle = 0.$$

Replacing y by iy , we get

$$\operatorname{Im} \langle x - x_0, y \rangle = 0.$$

Hence

$$\langle x - x_0, y \rangle = 0.$$

Thus $x - x_0$ is in the subspace $S^\perp$. Since $S \cap S^\perp = \{0\}$, we have a *direct sum decomposition*

$$\mathcal{H} = S \oplus S^\perp. \quad (11.10)$$

Recall that a vector space X is said to have a direct sum decomposition

$$X = V \oplus W, \quad (11.11)$$

if V, W are subspaces of X that have only the zero vector in common, and whose linear span is X . Then every vector x has a unique decomposition $x = v + w$ with $v \in V$, $w \in W$.

16. Show that the map P_S defined by (11.8) is linear, $\operatorname{ran} P_S = S$, and $\ker P_S = S^\perp$. (The symbols ran and $\ker$ stand for the *range* and the *kernel* of a linear operator.)

By the Pythagorean Theorem

$$\|x\|^2 = \|P_S x\|^2 + \|(I - P_S)x\|^2.$$

This shows that $\|P_S\| \leq 1$. Since $P_S x = x$ for all x in S , we have

$$\|P_S\| = 1. \quad (11.12)$$

(The obvious trivial exception is the case $S = \{0\}$. We do not explicitly mention such trivialities.)

The map P_S is called the *orthogonal projection* or the *orthoprojector* onto S . The space $S^\perp$ is called the *orthogonal complement* of the (closed linear) space S . In this case $S^{\perp\perp} = S$.

A problem with Banach spaces

17. The notion of direct sum in (11.11) is purely algebraic. If V is a linear subspace of a vector space X , then we can always find a subspace W such that X is the direct sum of V and W . (Hint: use a Hamel basis.)

When X is a Banach space it is natural to ask for a decomposition like (11.11) with the added requirement that both V and W be *closed* linear spaces.

Let us say that a closed linear subspace V of a Banach space X is a *direct summand* if there exists another closed linear subspace W of X such that we have the decomposition (11.11).

In a Hilbert space every closed linear subspace is a direct summand; we just choose $W = V^\perp$. In a general Banach space no obvious choice suggests itself. Indeed, there may *not* be any. There is a theorem of Lindenstrauss and Tzafriri that says that a Banach space in which every closed subspace is a direct summand is isomorphic to a Hilbert space.

The subspace c_0 in the Banach space ℓ_∞ is not a direct summand. This was proved by R.S. Phillips in 1940. A simple proof (that you can read) is given in R.J. Whitley, Projecting m onto c_0 , *American Mathematical Monthly*, 73 (1966) 285-286.

18. Let X be any vector space with a decomposition as in (11.11). We define a linear map $P_{V,W}$ called the *projection on V along W* by the relation $P_{V,W}(x) = v$, where $x = v + w$, $v \in V$, $w \in W$. Show that

- (i) $P_{V,W}$ is idempotent.
- (ii) $\text{ran } P_{V,W} = V$, $\ker P_{V,W} = W$.
- (iii) $I - P_{V,W} = P_{W,V}$.

Conversely suppose we are given an idempotent linear map P of X into itself. Let $\text{ran } P = V$, $\ker P = W$. Show that we have $X = V \oplus W$, and $P = P_{V,W}$.

19. Now assume that the space X in Section 18 is a Banach space. If the operator $P_{V,W}$ is *bounded* then V, W must be closed. (The kernel of a continuous map is closed.)

Show that if V is a direct summand in X , then the projection $P_{V,W}$ is a bounded operator. (Use the Closed Graph Theorem.) Show that $\|P_{V,W}\| \geq 1$.

Show that every finite-dimensional subspace V of a Banach space X is a direct summand. (Let $v_1, v_2, \dots, v_n$ be a basis for V . Every element x of V can be written as $\sum_{j=1}^n f_j(x)v_j$. The f_j define (bounded) linear functionals on V . By H.B.T. they can be extended to bounded linear functionals $\tilde{f}_j$ on X . For each $x \in X$ let $Px = \sum_{j=1}^n \tilde{f}_j(x)v_j$.)

20. If V is a direct summand in a Banach space X , then there exist infinitely many subspaces W such that $X = V \oplus W$. (You can see this in $\mathbb{R}^2$.) In a Hilbert space, there is a very special choice $W = V^\perp$.

In a Hilbert space by a direct sum decomposition we always mean a decomposition into a subspace and its orthogonal complement.

We will see later that among projections, orthogonal projections are characterised by one more condition: selfadjointness.

Self-duality

21. To every vector y in $\mathcal{H}$, there corresponds a linear functional f_y defined by

$$f_y(x) = \langle x, y \rangle \quad \text{for all } x \in \mathcal{H}.$$

This can be turned around. Let f be any (nonzero bounded) linear functional on $\mathcal{H}$. Let $S = \ker f$ and let z be any unit vector in $S^\perp$. Note that $x - (f(x)/f(z))z$ is in

S. So

$$\langle x - \frac{f(x)}{f(z)}z, z \rangle = 0,$$

i.e.,

$$\langle x, z \rangle = \frac{f(x)}{f(z)}.$$

So, if we choose $y = \overline{f(z)}z$, we have $f(x) = \langle x, y \rangle$.

Note that $\|f_y\| = \|y\|$. Thus the correspondence $y \mapsto f_y$ between $\mathcal{H}$ and $\mathcal{H}^*$ is isometric. There is just one minor irritant. This correspondence is *conjugate linear* and not linear:

$$f_{\alpha y} = \bar{\alpha} f_y.$$

The fact that $\mathcal{H}$ and $\mathcal{H}^*$ can be identified via the correspondence $y \mapsto f_y$ is sometimes called the *Riesz Representation Theorem* (for Hilbert spaces).

22. The Hahn–Banach Theorem for Hilbert spaces is a simple consequence of the above representation theorem.

23. A complex-valued function $B(\cdot, \cdot)$ on $\mathcal{H} \times \mathcal{H}$ is called a *sesquilinear form* if it is linear in the first and conjugate linear in the second variable. Its norm is defined to be

$$\|B\| = \sup_{\|x\|=\|y\|=1} |B(x, y)|. \text{ If this number is finite we say } B \text{ is } \textit{bounded}.$$

Let B be a bounded sesquilinear form. For each vector y let $f_y(x) := B(x, y)$. This is a bounded linear functional on $\mathcal{H}$. Hence, there exists a unique vector y' such that $f_y(x) = \langle x, y' \rangle$ for all x . Put $y' = Ay$. Now fill in the details of the proof of the following statement:

To every bounded sesquilinear form $B(\cdot, \cdot)$ on $\mathcal{H} \times \mathcal{H}$ there corresponds a unique linear operator A on $\mathcal{H}$ such that

$$B(x, y) = \langle x, Ay \rangle.$$

We have $\|B\| = \|A\|$.

24. Earlier on, we had defined the annihilator of any subset S of a Banach space X . This was a subset $S^\perp$ of X^* . When X is a Hilbert space, this set is the same as $S^\perp$ defined in Section 13.

25. Note that x_α converges to x in the weak topology of $\mathcal{H}$ if and only if $\langle x_\alpha, y \rangle \rightarrow \langle x, y \rangle$ for all $y \in \mathcal{H}$.

Supplementary Exercises

26. Let f be a nonzero bounded linear functional on a Banach space X and let $S = \{x \in X : f(x) = 1\}$. Show that S is a closed convex subset of X . Show that

$$\inf_{x \in S} \|x\| = \frac{1}{\|f\|}.$$

So, if there is no vector x in X for which $\|f\| = |f(x)| / \|x\|$, then the point 0 has no best approximant from S .

27. Let $X = C[0, 1]$ and let Y be its subspace consisting of all functions that vanish at 0. Let $\varphi(f) = \int_0^1 t f(t) dt$. Then φ is a bounded linear functional. Find its norm on X , and on Y . What are the points f in X and in Y for which $\|\varphi\| = |\varphi(f)| / \|f\|$.

28. Combine Exercises 26 and 27 to show that (the existence part of) Theorem 14 is not always true in all Banach spaces.

29. Let $S = \{x \in \mathbb{R}^2 : x_1, x_2 \geq 0, x_1 + x_2 = 1\}$. This is the line segment joining the points $(1, 0)$ and $(0, 1)$. Each point of S is at ℓ_1^2 distance 1 from the point $(0, 0)$. Thus the uniqueness part of Theorem 14 is violated in this Banach space.

30. Let V, W be any two subspaces of $\mathbb{R}^2$ not orthogonal to each other. Show that $\|P_{V,W}\| > 1$.

31. A function f on $\mathcal{H}$ is called a *quadratic form* if there exists a sesquilinear form B on $\mathcal{H} \times \mathcal{H}$ such that $f(x) = B(x, x)$. Show that a pointwise limit of quadratic forms is a quadratic form.

32. A sesquilinear form B is said to be *symmetric* if $B(x, y) = \overline{B(y, x)}$ for all x and y , *positive* if $B(x, x) \geq 0$ for all x , and *definite* if $B(x, x) = 0$ implies $x = 0$. Show that a positive, symmetric, sesquilinear form satisfies the Schwarz inequality

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

(If B is definite, then it is an inner product and we have proved the inequality in that case.) Hint : Consider $B(x, y) + \varepsilon \langle x, y \rangle$.

Orthonormal Bases

1. A subset E in a Hilbert space is said to be an *orthonormal set* if $\langle e_1, e_2 \rangle = 0$ for all e_1, e_2 in E ($e_1 \neq e_2$), and $\|e\| = 1$ for all e in E .

A maximal orthonormal set is called a *complete* orthonormal set, or an *orthonormal basis*. By Zorn's Lemma every Hilbert space has an orthonormal basis.

2. It follows from the Pythagorean Theorem that every orthonormal set is linearly independent.

3. Let $\{e_i : 1 \leq i \leq n\}$ be any finite orthonormal set. For each x in $\mathcal{H}$, $\langle x, e_j \rangle e_j$ is the *component of x in the direction of e_j* . One can see that $x - \sum_{j=1}^n \langle x, e_j \rangle e_j$ is orthogonal to each e_j , and hence to the sum $\sum_{j=1}^n \langle x, e_j \rangle e_j$. The Pythagorean Theorem then shows that

$$\sum_{j=1}^n |\langle x, e_j \rangle|^2 \leq \|x\|^2. \quad (12.1)$$

This is called *Bessel's inequality*.

4. Let $\{x_\alpha\}_{\alpha \in I}$ be a family of vectors in a Banach space. (The set I may be uncountable). We say that this family is *summable* and its sum is x , if for every $\varepsilon > 0$, there exists a finite subset J_0 of I such that

$$\left\| \sum_{\alpha \in J} x_\alpha - x \right\| < \varepsilon$$

for every finite subset J of I that contains J_0 . In this case we write

$$x = \sum_{\alpha \in I} x_\alpha.$$

Show that a sequence $\{x_n\}$ is summable if $\{\|x_n\|\}$ is summable.

5. Bessel's Inequality. Let $\{e_\alpha\}_{\alpha \in I}$ be any orthonormal set in $\mathcal{H}$. Then for all x

$$\sum_{\alpha \in I} |\langle x, e_\alpha \rangle|^2 \leq \|x\|^2. \quad (12.2)$$

Corollary. For each x , the set

$$E = \{e_\alpha : \langle x, e_\alpha \rangle \neq 0\} \quad (12.3)$$

is countable.

Proof. Let

$$E_n = \{e_\alpha : |\langle x, e_\alpha \rangle|^2 > \|x\|^2/n\}.$$

Then $E = \cup_{n=1}^{\infty} E_n$. By Bessel's inequality the set E_n can have no more than $n - 1$ elements. ■

6. Parseval's Equality. Let $\{e_\alpha\}_{\alpha \in I}$ be an orthonormal basis in $\mathcal{H}$. Then for each $x \in \mathcal{H}$

$$x = \sum_{\alpha \in I} \langle x, e_\alpha \rangle e_\alpha. \quad (12.4)$$

$$\|x\|^2 = \sum_{\alpha \in I} |\langle x, e_\alpha \rangle|^2. \quad (12.5)$$

Proof. Given an x , let E be the set given by (12.3). Enumerate its elements as $\{e_1, e_2, \dots\}$. For each n , let

$$y_n = \sum_{i=1}^n \langle x, e_i \rangle e_i.$$

If $n > m$, we have

$$\|y_n - y_m\|^2 = \sum_{i=m+1}^n |\langle x, e_i \rangle|^2.$$

By Bessel's inequality this sum goes to zero as $n, m \rightarrow \infty$. So y_n is a Cauchy sequence. Let y be its limit. Note that for all j

$$\begin{aligned} \langle x - y, e_j \rangle &= \langle x, e_j \rangle - \lim_{n \rightarrow \infty} \left\langle \sum_{i=1}^n \langle x, e_i \rangle e_i, e_j \right\rangle \\ &= \langle x, e_j \rangle - \langle x, e_j \rangle = 0. \end{aligned}$$

If e_β is any element of the given set $\{e_\alpha\}_{\alpha \in I}$ outside E , then $\langle x, e_\beta \rangle = 0$, and once again $\langle x - y, e_\beta \rangle = 0$. Thus $x - y$ is orthogonal to the maximal orthonormal family $\{e_\alpha\}_{\alpha \in I}$. Hence $x = y$. Thus

$$x = \sum_{\alpha \in I} \langle x, e_\alpha \rangle e_\alpha.$$

Only countably many terms in this sum are nonzero. (However, *this countable set depends on x* .) Further note that

$$\begin{aligned} \|x\|^2 - \sum_{\alpha \in I} |\langle x, e_\alpha \rangle|^2 &= \lim_{n \rightarrow \infty} \left[\|x\|^2 - \sum_{i=1}^n |\langle x, e_i \rangle|^2 \right] \\ &= \lim_{n \rightarrow \infty} \|x - \sum_{i=1}^n \langle x, e_i \rangle e_i\|^2 \\ &= 0. \end{aligned}$$

This proves (12.5). ■

Separable Hilbert spaces

7. Let $\{u_1, u_2, \dots\}$ be a finite or countable linearly independent set in $\mathcal{H}$. Then there exists an orthonormal set $\{e_1, e_2, \dots\}$ having the same cardinality and the same linear span as the set $\{u_n\}$. This is constructed by the familiar *Gram-Schmidt Process*.

8. **Theorem.** A Hilbert space is separable if and only if it has a countable orthonormal basis.

Proof. A countable orthonormal basis for $\mathcal{H}$ is also a Schauder basis for it. So, if such a basis exists, $\mathcal{H}$ must be separable.

Conversely, let $\mathcal{H}$ be separable and choose a countable dense set $\{x_n\}$ in $\mathcal{H}$. We can obtain from this a set $\{u_n\}$ that is linearly independent and has the same (closed) linear span. From this set $\{u_n\}$ we get an orthonormal basis by the Gram–Schmidt process. ■

9. A linear bijection U between two Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$ is called an *isomorphism* if it preserves inner products; i.e.,

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

10. **Theorem.** Every separable infinite-dimensional Hilbert space is isomorphic to ℓ_2 .

Proof. If $\mathcal{H}$ is separable, it has a countable orthonormal basis $\{e_n\}$. Let $U(x) = \{\langle x, e_n \rangle\}$. Show that for each x in $\mathcal{H}$ the sequence $\{\langle x, e_n \rangle\}$ is in ℓ_2 , and U is an isomorphism. ■

We will assume from now on that all our Hilbert spaces are separable.

11. Let $\mathcal{H} = L_2[-\pi, \pi]$. The functions $e_n(t) = \frac{1}{\sqrt{2\pi}} e^{int}$, $n \in \mathbb{Z}$, form an orthonormal basis in $\mathcal{H}$. It is easy to see that the family $\{e_n\}$ is orthonormal. Its completeness follows from standard results in Fourier series.

There are other orthonormal bases for $\mathcal{H}$ that have been of interest in classical analysis. In recent years there has been renewed interest in them because of the recent theory of wavelets.

12. Exercises. (i) Let $\{e_n\}$ be an orthonormal basis in $\mathcal{H}$. Any orthonormal set $\{f_n\}$ that satisfies

$$\sum_{n=1}^{\infty} \|e_n - f_n\|^2 < 1$$

is an orthonormal basis. (Hint: If x is orthogonal to $\{f_n\}$ show $\sum |\langle x, e_n \rangle|^2 < \|x\|^2$, violating Parseval's equality.)

(ii) More generally, show that if

$$\sum_{n=1}^{\infty} \|e_n - f_n\|^2 < \infty$$

then $\{f_n\}$ is an orthonormal basis. (Hints: Choose N such that $\sum_{n=N+1}^{\infty} \|e_n - f_n\|^2 < 1$. Let S be the closed linear span of $\{f_{N+1}, f_{N+2}, \dots\}$. For $1 \leq n \leq N$, the vectors

$$g_n = e_n - \sum_{m=N+1}^{\infty} \langle e_n, f_m \rangle f_m$$

are in $S^\perp$. Show that $\dim S^\perp = N$. The space $S^\perp$ is spanned by $\{g_1, \dots, g_N\}$ and by $\{f_1, \dots, f_N\}$. So, if a vector x is orthogonal to the family $\{f_n\}$, then it is orthogonal to the family $\{f_n : n \geq N+1\}$ and to $\{g_1, \dots, g_N\}$. Show that it is orthogonal to $\{e_1, \dots, e_N\}$. Use this and Part (i) to show that Parseval's equality forbids such behaviour.)

13. Metrisability of the unit ball with the weak topology. We have seen that the weak topology of ℓ_2 is not metrisable. However, its restriction to the unit ball is metrisable.

(i) Let $\mathcal{H}$ be any separable Hilbert space and let $\{e_n\}$ be an orthonormal basis for $\mathcal{H}$. Let $B = \{x \in \mathcal{H} : \|x\| \leq 1\}$. For $x, y \in B$, let

$$d(x, y) := \frac{1}{2^n} \sum_{n=1}^{\infty} |\langle x - y, e_n \rangle|.$$

Show that d is a metric on B .

(ii) Show that the topology generated by d is the same as the one given by the weak

topology; i.e., $d(x_n, x) \rightarrow 0$ if and only if $x_n \xrightarrow{w} x$.

(iii) Show that the metric space (B, d) is compact.

14. Let $\mathcal{H} = L_2[-1, 1]$. Apply the Gram–Schmidt process to the sequence of functions $\{1, t, t^2, \dots\}$. The resulting orthogonal functions are

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dt^n} (t^2 - 1)^n.$$

These are called the *Legendre polynomials*. Show that the family $\{\sqrt{n+1/2} P_n\}$ is an orthonormal basis for $\mathcal{H}$. (For proving the completeness of this system, the Weierstrass approximation theorem may be useful.)

15. Let $\mathcal{H} = L_2(\mathbb{R})$. Apply the Gram–Schmidt process to the family

$$\{e^{-t^2/2}, te^{-t^2/2}, t^2 e^{-t^2/2}, \dots\}.$$

This gives the functions

$$f_n(t) = (-1)^n e^{-t^2/2} \frac{d^n}{dt^n} e^{-t^2} =: H_n(t) e^{-t^2/2}, \quad n = 0, 1, 2, \dots$$

The functions $H_n(t)$ are called *Hermite polynomials*. Show that the members of $\{f_n(t)\}$ are pairwise orthogonal, and normalise them. Show that the resulting family is an orthonormal basis for $\mathcal{H}$.

(Hint: To show completeness, we need to show that if

$$\int_{-\infty}^{\infty} g(t) e^{-t^2/2} t^n dt = 0, \quad n = 0, 1, 2, \dots, \quad (12.6)$$

then $g = 0$. Introduce the complex function

$$G(z) = \int_{-\infty}^{\infty} g(t) e^{-t^2/2} e^{itz} dt.$$

This is an entire function. Use (12.6) to see that G and its derivatives of all orders vanish at 0. Hence G is zero everywhere. In particular

$$\int_{-\infty}^{\infty} g(t) e^{-t^2/2} e^{itx} dt = 0 \quad \text{for all } x \in \mathbb{R}.$$

Multiply this equality by e^{-ixy} , where y is a real number, then integrate with respect to x from $-a$ to a . This gives

$$\int_{-\infty}^{\infty} g(t)e^{-t^2/2} \frac{\sin a(t-y)}{t-y} dt = 0, \quad \text{for all } a, y \in \mathbb{R}.$$

Conclude that $g = 0$.)

16. Let $\mathcal{H} = L_2(0, \infty)$. The functions

$$L_n(t) = e^t \frac{d^n}{dt^n}(e^{-t}t^n), \quad n = 0, 1, 2, \dots$$

are called the *Laguerre polynomials*. Show that the family

$$f_n(t) = \frac{1}{n!} e^{-t/2} L_n(t)$$

is an orthonormal basis for $\mathcal{H}$.

17. Let $\mathcal{H} = L_2[0, 1]$. Let

$$r_k(t) = \text{sgn} \sin(2^k \cdot 2\pi t), \quad k = 0, 1, 2, \dots,$$

where the value of $r_k(t)$ at a discontinuity is taken as the right hand limit. Equivalently, on the dyadic intervals $[j/2^{k+1}, (j+1)/2^{k+1})$, $0 \leq j < 2^{k+1}$, $r_k(t)$ takes the value 1 if j is even and -1 if j is odd. The constant function 1 and the functions r_k together are called *Rademacher functions*. They form an orthonormal family but not a complete family. (The function $\cos 2\pi t$ is orthogonal to all of them.)

This system is included in another family called *Walsh functions* defined as follows. Let $w_0(t) = 1$. For $n \geq 1$, let

$$n = \sum_{k=0}^m n_k \cdot 2^k \quad \text{where } n_k = 0 \text{ or } 1$$

be the binary expansion of n . Let

$$w_n(t) = \prod_{k=0}^m [r_k(t)]^{n_k}.$$

The functions w_n together with the constant function 1 are called the Walsh functions. They are step functions that take the values ± 1 only. Note that if $n = 2^k$, then $w_n = r_k$. So this family includes the Rademacher functions. In fact it consists of all finite products of distinct Rademacher functions. Show that the Walsh functions form an orthonormal basis for $\mathcal{H}$. (Hint: To check orthogonality, observe that if at least two of the integers $k_1, k_2, \dots, k_n$ are distinct, then

$$\int_0^1 r_{k_1}(t)r_{k_2}(t)\cdots r_{k_n}(t)dt = 0.$$

To prove completeness, let $f \in \mathcal{H}$ and define $F(x) = \int_0^x f(t)dt$. Then $F'(x) = f(x)$ almost everywhere. Show that the conditions $\int_0^1 f(t)w_n(t)dt = 0$ lead successively to the conclusions $F(x) = 0$ if $x = k/2^m$, $m = 0, 1, 2, \dots$, $k = 1, \dots, 2^m$. Since F is continuous, this implies F is zero everywhere; hence f is zero almost everywhere.)

18. Gram matrices. Let $x_1, \dots, x_n$ be any vectors in a Hilbert space $\mathcal{H}$. The $n \times n$ matrix $G(x_1, \dots, x_n)$ whose i, j entry is $\langle x_i, x_j \rangle$ is called the *Gram matrix* of the given set of vectors. Its determinant is called the *Gram determinant*.

(i) Every Gram matrix is positive semidefinite; it is positive definite if and only if the vectors x_j are linearly independent. [Calculate $\langle Gu, u \rangle$.]

(ii) Every positive semidefinite matrix is a Gram matrix. [Hint: write $a_{ij} = \langle Ae_i, e_j \rangle = \langle A^{1/2}e_i, A^{1/2}e_j \rangle$.]

(iii) Let $\lambda_j, 1 \leq j \leq n$ be any positive numbers. Then the matrix whose i, j entry is $\frac{1}{\lambda_i + \lambda_j}$ is positive semidefinite. [Hint: $\frac{1}{\lambda_i + \lambda_j} = \int_0^\infty e^{-(\lambda_i + \lambda_j)t} dt$.]

(iv) Calculate, by induction on n or by some other argument, the determinant of the matrix in (iii); it has the value

$$\frac{\prod_{1 \leq i < j \leq n} (\lambda_i - \lambda_j)^2}{\prod_{1 \leq i, j \leq n} (\lambda_i + \lambda_j)}. \quad (12.7)$$

19. Let $x_1, \dots, x_n$ be linearly independent vectors in $\mathcal{H}$, and let $\mathcal{M}$ be their linear

span. Show that for every y in $\mathcal{H}$

$$[\text{dist}(y, \mathcal{M})]^2 = \frac{\det G(y, x_1, \dots, x_n)}{\det G(x_1, \dots, x_n)}.$$

(Hint: Let $y = x + z$, where $x \in \mathcal{M}, z \in \mathcal{M}^\perp$. Calculate $G(y, x_1, \dots, x_n)$ by substituting $y = x + z$ and using the fact that a determinant is a linear function of each of its columns.)

20. We know that the family $f_n(t) = t^n, n = 0, 1, 2, \dots$ is fundamental in $L_2[0, 1]$.

The following remarkable theorem tells us there is a lot of room here; much smaller subfamilies of this family are also fundamental.

Müntz's Theorem. Let $1 \leq n_1 < n_2 < \dots$ be any sequence of integers. Then the family $\{t^{n_k}\}$ is fundamental in $L_2[0, 1]$ if and only if

$$\sum_{j=1}^{\infty} \frac{1}{n_j} = \infty. \quad (12.8)$$

Proof. Let $\mathcal{M}_k$ be the linear span of the functions $t^{n_1}, \dots, t^{n_k}$. The set $\{t^{n_j}\}$ would be fundamental if and only if $\text{dist}(f, \mathcal{M}_k)$ goes to zero as $k \rightarrow \infty$. Since the family $\{t^m\}$ is fundamental, this is so if and only if for each m $\text{dist}(t^m, \mathcal{M}_k)$ goes to zero as $k \rightarrow \infty$. By Exercise 19, this is so if and only if

$$\lim_{k \rightarrow \infty} \frac{\det G(t^m, t^{n_1}, \dots, t^{n_k})}{\det G(t^{n_1}, \dots, t^{n_k})} = 0 \quad \text{for all } m. \quad (12.9)$$

Note that

$$\langle t^i, t^j \rangle = \int_0^1 t^{i+j} dt = \frac{1}{i+j+1}.$$

Hence the ratio of the two Gram determinants occurring above can be evaluated using (12.7). The answer is

$$\frac{1}{2m+1} \prod_{j=1}^k \frac{(n_j - m)^2}{(n_j + m + 1)^2} = \frac{1}{2m+1} \prod_{j=1}^k \frac{(1 - m/n_j)^2}{(1 + (m+1)/n_j)^2}.$$

So, the condition (12.9) becomes

$$\lim_{k \rightarrow \infty} \sum_{j=1}^k \left[\log\left(1 - \frac{m}{n_j}\right) - \log\left(1 + \frac{m+1}{n_j}\right) \right] = -\infty. \quad (12.10)$$

Since

$$\lim_{x \rightarrow 0} \frac{\log(1+x)}{x} = 1$$

the series $\sum \log(1+x_n)$ and $\sum x_n$ are convergent or divergent simultaneously. Use this to show that (12.10) is true if and only if (12.8) is. ■

Corollary. The family $\{t^p : p \text{ a prime number}\}$ is fundamental in $L_2[0, 1]$.

Linear Operators

Let X, Y be Banach spaces. For a while we will study bounded linear operators from X to Y . These will just be called operators.

Topologies on Operators

1. **The norm topology.** We denote the space of operators from X to Y by $\mathcal{B}(X, Y)$. This is a Banach space with the norm $\|A\| := \sup_{\|x\|=1} \|Ax\|$. The topology given by this norm is called the *usual topology*, the *norm topology* or the *uniform operator topology* on $\mathcal{B}(X, Y)$.

2. **The strong operator topology.** We say that a net A_α in $\mathcal{B}(X, Y)$ converges *strongly* to A if for each x in X , $A_\alpha x$ converges to Ax ; i.e., if $\|A_\alpha x - Ax\|$ converges to zero for each x . We write $A_\alpha \xrightarrow{s} A$ to indicate this convergence. The associated topology is called the *strong operator topology*. It is the weak topology generated by the family of maps

$$\begin{aligned} F_x : \mathcal{B}(X, Y) &\longrightarrow Y, \\ A &\longmapsto Ax, \end{aligned}$$

where x varies over X .

3. **The weak operator topology.** We say a net A_α converges to A in the weak operator topology if $f(A_\alpha x) \rightarrow f(Ax)$ for all $f \in Y^*$, $x \in X$. We write this as

$A_\alpha \xrightarrow{w} A$. This is the weak topology generated by the family

$$\begin{aligned} F_{x,f} : \mathcal{B}(X, Y) &\longrightarrow \mathbb{C}, \\ A &\mapsto f(Ax), \end{aligned}$$

where x varies over X and f over Y^* . If X, Y are Hilbert spaces, then $A_\alpha \xrightarrow{w} A$ if and only if $\langle A_\alpha x, y \rangle \rightarrow \langle Ax, y \rangle$ for all $x \in X, y \in Y$.

4. Caution. In Lecture 9, we defined the strong and the weak topologies for any Banach space. The adjectives strong and weak are now used in a different sense. (The “strong” topology of the Banach spaces $\mathcal{B}(X, Y)$ is its “usual” topology). For spaces of operators the words strong and weak will be used in the new sense introduced here; unless it is stated otherwise.

5. Examples. Clearly convergence in the norm topology implies convergence in the strong operator topology, which in turn implies convergence in the weak operator topology.

In the following examples, X and Y are the space ℓ_2 .

(i) Let $A_n = \frac{1}{n}I$; i.e., $A_n x = \frac{1}{n}x$ for all x . Then A_n converges to zero in the norm topology.

(ii) Let $e_1, e_2, \dots$ be the standard orthonormal basis for ℓ_2 . Let P_n be the orthogonal projection onto the linear span of $\{e_1, \dots, e_n\}$. Then $I - P_n$ is the orthogonal projection onto the orthogonal complement of this space. Here $P_n \rightarrow I$ in the strong operator topology. But $\|I - P_n\| = 1$ for all n . So P_n does not converge to I in the norm topology.

(iii) The *right shift operator* S on ℓ_2 is defined as follows. Let $x = (x_1, x_2, \dots)$ be any element of ℓ_2 . Then

$$S((x_1, x_2, \dots)) = (0, x_1, x_2, \dots).$$

Then for all x, y in ℓ_2 , and for all positive integers n

$$\langle S^n x, y \rangle = \sum_{i=1}^{\infty} x_i \bar{y}_{n+i}.$$

So,

$$|\langle S^n x, y \rangle| \leq \left(\sum_{i=1}^{\infty} |x_i|^2 \right)^{1/2} \left(\sum_{i=1}^{\infty} |y_{n+i}|^2 \right)^{1/2}.$$

As $n \rightarrow \infty$, the last sum goes to zero. So the sequence $\{S^n\}$ converges to zero in the weak operator topology. However, $\|S^n x\| = \|x\|$ for all x and n . So $\{S^n\}$ does not converge to zero in the strong operator topology. Hence it does not converge to *any* limit in the strong operator topology, because if it did, then the strong limit would also be a weak limit, and that can only be zero.

6. The strong operator topology and the weak operator topology are not metrisable. While convergence of sequences does not reveal all the features of these topologies, we may still be interested in sequences and their convergence. The Uniform Boundedness Principle is the useful tool in these situations.

Exercise. Let $\{A_n\}$ be a sequence of operators. Suppose $\{A_n x\}$ converges for each x . Then there exists an operator A such that $A_n \xrightarrow{s} A$.

Is this true for a net instead of a sequence?

7. **Lemma.** Let $\{A_n\}$ be a sequence of operators in a Hilbert space $\mathcal{H}$. Suppose $\{A_n\}$ is a weakly Cauchy sequence. Then there exists an operator A such that $A_n \xrightarrow{w} A$.

Proof. The sequence $\{A_n\}$ is weakly Cauchy if for each x, y in $\mathcal{H}$ the sequence $\{\langle A_n x, y \rangle\}$ is a Cauchy sequence (of complex numbers). Let

$$B(x, y) = \lim_{n \rightarrow \infty} \langle A_n x, y \rangle.$$

It is clear that B is a sesquilinear form. If we could show it is bounded, then we would know from the Riesz Representation Theorem that there exists an operator

A such that

$$B(x, y) = \langle Ax, y \rangle.$$

Then clearly $A_n \xrightarrow{w} A$. Since

$$|\langle A_n x, y \rangle| \leq \|A_n\| \|x\| \|y\|,$$

the boundedness of B would follow from that of the sequence $\{\|A_n\|\}$. This is proved by appealing to the Uniform Boundedness Principle.

First note that for each x, y , the sequence $\langle A_n x, y \rangle$ is bounded. Regard, for each fixed x , $A_n x$ as a linear functional on $\mathcal{H}$ acting as

$$(A_n x)(y) = \overline{\langle A_n x, y \rangle}.$$

By the U.B.P.,

$$\sup_n \|A_n x\| < \infty \quad \text{for all } x.$$

Once again by the U.B.P.,

$$\sup_n \|A_n\| < \infty.$$

■

Operator Multiplication

8. Consider the space $\mathcal{B}(X)$. Let $A_n \rightarrow A$ and $B_n \rightarrow B$ in the norm topology. Then show that $\|A_n B_n - AB\| \rightarrow 0$. This shows that multiplication of operators is *jointly continuous* in the norm topology of $\mathcal{B}(X)$.

9. Let A_n and B_n be sequences in $\mathcal{B}(X)$ converging in the strong operator topology to A and B , respectively. Use the U.B.P. to show the sequence $\{\|A_n\|\}$ is bounded; and then show that the product $A_n B_n$ converges to AB in the strong operator topology.

This argument fails for nets. Hence, it does not follow that multiplication of operators is jointly continuous in the strong operator topology. In fact, it is *not*.

Exercise. Let $\mathcal{H}$ be any infinite-dimensional Hilbert space. Let $\mathcal{N} = \{A \in \mathcal{B}(\mathcal{H}) : A^2 = 0\}$. Elements of $\mathcal{N}$ are called nilpotent operators of index 2.

(i) Let A_0 be any element of $\mathcal{B}(\mathcal{H})$. Then sets of the form

$$\{A : \|(A - A_0)x_i\| < \varepsilon, 1 \leq i \leq n\},$$

where $\varepsilon > 0$, $n \in \mathbb{N}$, and $x_1, \dots, x_n$ are *linearly independent*, form a neighbourhood base at A_0 in the strong operator topology.

(ii) Let $\{x_1, \dots, x_n, y_1, \dots, y_n\}$ be a linearly independent set in $\mathcal{H}$ such that $\|y_i - A_0x_i\| < \varepsilon$ for all i . Define an operator A by putting $Ax_i = y_i$, $Ay_i = 0$ for all i , and $Au = 0$ for all u orthogonal to $\{x_1, \dots, x_n, y_1, \dots, y_n\}$. Then $A^2 = 0$. Show that A belongs to the basic neighbourhood in (i).

(iii) This shows that the set $\mathcal{N}$ is dense in $\mathcal{B}(\mathcal{H})$ in the strong operator topology. So, if squaring of operators were a continuous operation, then $\mathcal{N}$ would equal $\mathcal{B}(\mathcal{H})$. That can't be.

Exercise. Here is one more proof of the same fact. Consider the set of all ordered pairs (M, u) where M is a finite-dimensional subspace of $\mathcal{H}$ and u a unit vector orthogonal to M . Define a partial order on this set by saying $(M, u) \prec (N, v)$ if N contains M and u . Now define two nets of operators as follows

$$\begin{aligned} A_{(M,u)}x &= (\dim M) \langle x, u \rangle x_0, \\ B_{(M,u)}x &= \frac{1}{\dim M} \langle x, x_0 \rangle u, \end{aligned}$$

where x_0 is a fixed unit vector. Show that both these nets converge to 0 in the strong operator topology; but their product does not.

10. Let $X = \ell_2$. We defined the right shift operator S in Section 5(iii). The *left shift* is the operator T defined as

$$T((x_1, x_2, \dots)) = (x_2, x_3, \dots).$$

Note that for each x , $\|T^n x\| \rightarrow 0$. Thus $\{T^n\}$ converges to 0 in the strong, and therefore also in the weak, operator topology. We have seen earlier that $\{S^n\}$ also converges to 0 in the weak operator topology. Note that $T^n S^n = I$ for all n . This example shows that operator multiplication is not continuous (even on sequences) in the weak operator topology.

11. However operator multiplication is *separately continuous* in both the strong and the weak topology; i.e., if a net A_α converges, strongly or weakly to A , then for each B , $A_\alpha B$ converges to AB in the same sense; and if B_α converges, strongly or weakly to B , then AB_α converges to AB in the same sense. It is easy to prove these statements.

12. **Exercise.** Let $\{e_n\}$ be an orthonormal basis for $\mathcal{H}$. Let

$$d_S(A, B) := \sum_{n=1}^{\infty} \frac{1}{2^n} \|(A - B)e_n\|,$$

$$d_W(A, B) := \sum_{m,n=1}^{\infty} \frac{1}{2^{m+n}} |\langle (A - B)e_m, e_n \rangle|.$$

Show that these are metrics on $\mathcal{B}(\mathcal{H})$. On each **bounded** set of $\mathcal{B}(\mathcal{H})$ the topology given by them is the strong (weak) operator topology.

Inverses

13. Let $A \in \mathcal{B}(X)$. If A is bijective, then by the inverse mapping theorem, A^{-1} is also in $\mathcal{B}(X)$. Let $\mathcal{G}$ be the collection of all invertible elements of $\mathcal{B}(X)$. This set is a multiplicative group. We have $(AB)^{-1} = B^{-1}A^{-1}$.

14. **Theorem.** If $\|I - A\| < 1$, then A is invertible and

$$A^{-1} = I + (I - A) + (I - A)^2 + \cdots. \quad (13.1)$$

Proof. To see that the series is convergent, let

$$S_n = \sum_{j=0}^n (I - A)^j.$$

Then note that

$$\|S_{n+m} - S_n\| \leq \sum_{j=n+1}^{n+m} \|I - A\|^j.$$

This goes to zero as $n, m \rightarrow \infty$. So $\{S_n\}$ is a Cauchy sequence. Hence the series in (13.1) is convergent. Let T denote its sum. Note that

$$AS_n = S_n - (I - A)S_n = I - (I - A)^{n+1}.$$

So, by continuity of operator multiplication $AT = I$. A similar argument shows $TA = I$. Hence $T = A^{-1}$. ■

15. If $\|A\| < 1$, then $I - A$ is invertible and

$$(I - A)^{-1} = I + A + A^2 + \dots \quad (13.2)$$

Note that

$$\|(I - A)^{-1}\| \leq \frac{1}{1 - \|A\|}.$$

The series (13.1) or (13.2) is called the *Neumann Series*.

16. The theorem just proved shows that $\mathcal{G}$ contains an open neighbourhood of I ; hence it contains an open neighbourhood of each of its points. Thus $\mathcal{G}$ is an open subset of $\mathcal{B}(X)$.

More precisely, show that if $A \in \mathcal{G}$ and $\|A - B\| < 1/\|A^{-1}\|$, then $B \in \mathcal{G}$ and

$$\|B^{-1}\| \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\| \|A - B\|}.$$

17. Show that

$$\|B^{-1} - A^{-1}\| \leq \frac{\|A^{-1}\|^2 \|A - B\|}{1 - \|A^{-1}\| \|A - B\|}.$$

This shows that operator inversion is continuous in the norm topology. Thus $\mathcal{G}$ is a *topological group*.

18. If X is finite-dimensional, $\mathcal{G}$ is dense in $\mathcal{B}(X)$. (Matrices with nonzero eigenvalues are dense in the space of all matrices.)

19. This is not true in infinite-dimensional spaces. Let $X = \ell_2$ and let S be the right shift operator. Then S is left-invertible (because $TS = I$) but not right-invertible (if it were it would be invertible). We will show that no operator in a ball of radius one around S is invertible. If $\|S - A\| < 1$ then

$$\|I - TA\| = \|T(S - A)\| \leq \|T\| \|S - A\| < 1.$$

So TA is invertible. If A were invertible, so would be T ; but that is not the case.

Exercise. The set of right invertible operators (a set that includes $\mathcal{G}$) is not dense in $\mathcal{B}(X)$. Nor is the set of left invertible operators.

If A is a linear operator on a finite-dimensional vector space, then one of the two conditions, injectivity and surjectivity, implies the other. This is not so for operators on infinite-dimensional spaces.

Adjoint Operators

Every operator A from X to Y gives rise, in a natural way to an operator A^* from the dual space Y^* to X^* . Many properties of A can be studied through this operator called the *adjoint* of A .

1. Let A be an operator from X to Y . For $f \in Y^*$ let

$$(A^*f)(x) = f(Ax) \quad \text{for all } x \in X. \quad (14.1)$$

Then A^*f is a bounded linear functional on X ; i.e., $A^*f \in X^*$. It is obvious from the definition that A^* is a linear map from Y^* to X^* . The equation (14.1) is some times written as

$$\langle A^*f, x \rangle = \langle f, Ax \rangle, \quad x \in X, f \in Y^*. \quad (14.2)$$

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

2. If $f \in Y^*$, and $\|f\| = 1$, then

$$\|A^*f\| = \sup_{\|x\|=1} |(A^*f)(x)| = \sup_{\|x\|=1} |f(Ax)| \leq \sup_{\|x\|=1} \|Ax\| = \|A\|.$$

Thus $\|A^*\| \leq \|A\|$, and A^* is a bounded linear operator from Y^* to X^* . We can say more:

$$\|A^*\| = \|A\|. \quad (14.3)$$

To prove this we need to show $\|A\| \leq \|A^*\|$. Let x be any element of X . By the Hahn-Banach Theorem, there exists a linear functional f on Y such that $\|f\| = 1$

and $f(Ax) = \|Ax\|$. Thus

$$\|Ax\| = f(Ax) = (A^*f)(x) \leq \|A^*\| \|f\| \|x\| = \|A^*\| \|x\|.$$

This shows that $\|A\| \leq \|A^*\|$.

3. Exercise.

(i) Let $A, B \in \mathcal{B}(X, Y)$. Then

$$(\alpha A + \beta B)^* = \alpha A^* + \beta B^* \quad \text{for } \alpha, \beta \in \mathbb{C}.$$

(ii) Let $A \in \mathcal{B}(X, Y)$, $B \in \mathcal{B}(Y, Z)$. Then

$$(BA)^* = A^*B^*.$$

(iii) The adjoint of the identity operator on X is the identity operator on X^* ; i.e.,

$$I^* = I.$$

(iv) If A is an invertible operator from X to Y then A^* is an invertible operator from Y^* to X^* , and

$$(A^*)^{-1} = (A^{-1})^*.$$

4. The conclusion of (i) above is that the map $A \mapsto A^*$ from $\mathcal{B}(X, Y)$ to $\mathcal{B}(Y^*, X^*)$ is linear; that of (ii) is some times expressed by saying this map is contravariant. The equation (14.3) says this map is an isometry. It is, in general, not surjective.

5. **Example.** Let $X = Y = \ell_p$ where $1 \leq p < \infty$. Let S be the right shift operator; i.e., if $x = (x_1, x_2, \dots)$, then $Sx = (0, x_1, x_2, \dots)$. Let $T = S^*$. This is an operator on

ℓ_q . What is it? Let $f \in \ell_q$ and let $g = S^*f$. The definition (14.1) says $g(x) = f(Sx)$ for all x in ℓ_p ; i.e.,

$$g_1x_1 + g_2x_2 + \cdots = f_2x_1 + f_3x_2 + \cdots.$$

This is true for all x . Hence

$$(g_1, g_2, \dots) = (f_2, f_3, \dots).$$

Thus T is the *left shift* operator on ℓ_q . It maps $(f_1, f_2, \dots)$ to $(f_2, f_3, \dots)$.

Adjoint of Hilbert Space Operators

6. Let $\mathcal{H}$ be a Hilbert space. Recall that $\mathcal{H}$ is isomorphic to $\mathcal{H}^*$ via a conjugate linear map R that associates to $y \in \mathcal{H}$ the linear functional f_y defined as $f_y(x) = \langle x, y \rangle$ for all $x \in \mathcal{H}$. (See Section 21, Lecture 11.) So, for every $A \in \mathcal{B}(\mathcal{H})$ its adjoint A^* can be identified with an operator on $\mathcal{H}$. Call this operator $A^\dagger$ for the time being. We have $A^\dagger = R^{-1}A^*R$ (as shown in the diagram).

$$\begin{array}{ccc}
 \mathcal{H}^* & \xrightarrow{A^*} & \mathcal{H}^* \\
 \uparrow R & & \downarrow R^{-1} \\
 \mathcal{H} & \xrightarrow{A^\dagger} & \mathcal{H}
 \end{array}$$

If $A^*f_y = f_z$, then $A^\dagger y = z$. We have

$$\langle Ax, y \rangle = f_y(Ax) = (A^*f_y)(x) = f_z(x) = \langle x, z \rangle = \langle x, A^\dagger y \rangle$$

for all x, y . Thus

$$\langle Ax, y \rangle = \langle x, A^\dagger y \rangle \quad \text{for all } x, y \in \mathcal{H}.$$

This equation determines $A^\dagger$ uniquely; i.e., if there is another linear operator B on $\mathcal{H}$ such that

$$\langle Ax, y \rangle = \langle x, By \rangle \quad \text{for all } x, y,$$

then $B = A^\dagger$. It is customary to call this operator $A^\dagger$ the adjoint of A . We will do so too and use the symbol A^* for *this* operator. Thus A^* is the unique operator associated with A by the condition

$$\langle Ax, y \rangle = \langle x, A^*y \rangle \quad \text{for all } x, y \in \mathcal{H}. \quad (14.4)$$

The correspondence $A \mapsto A^*$ is conjugate linear.

7. If $\mathcal{H}, \mathcal{K}$ are Hilbert spaces and A is a linear operator from $\mathcal{H}$ to $\mathcal{K}$, then A^* is a linear operator from $\mathcal{K}$ to $\mathcal{H}$ defined by (14.4) with $x \in \mathcal{H}, y \in \mathcal{K}$.

8. **Theorem.** The map $A \mapsto A^*$ on $\mathcal{B}(\mathcal{H})$ has the following properties :

- (i) it is conjugate linear.
- (ii) it is isometric, $\|A^*\| = \|A\|$ for all A .
- (iii) it is surjective.
- (iv) $A^{**} = A$ for all A .
- (v) $(AB)^* = B^*A^*$ for all A, B .
- (vii) $I^* = I$.
- (vii) If A is invertible, then so is A^* and $(A^*)^{-1} = (A^{-1})^*$.

Thus the map $A \mapsto A^*$ has properties very similar to the complex conjugation $z \mapsto \bar{z}$ on $\mathbb{C}$. A new feature is the relation (v) arising out of non-commutativity of operator multiplication.

9. **Theorem.** For all A in $\mathcal{B}(\mathcal{H})$ we have

$$\|A^*A\| = \|A\|^2. \quad (14.5)$$

Proof. The submultiplicativity of the norm, and the property (14.3) show

$$\|A^*A\| \leq \|A^*\| \|A\| = \|A\|^2.$$

On the other hand we have

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

for all vectors x . Hence $\|A\|^2 \leq \|A^*A\|$. ■

It is clear from this proof that

$$\|AA^*\| = \|A\|^2 = \|A^*A\| \quad (14.6)$$

as well.

10. The property (14.5) is very important. A Banach algebra (see Lecture 3) with an involution (a star operation $A \mapsto A^*$) whose norm satisfies (14.5) is called a C^* -algebra. Study of such algebras is an important area in functional analysis.

Continuity Properties

11. Since $\|A^*\| = \|A\|$, the map $A \mapsto A^*$ from $\mathcal{B}(X)$ to $\mathcal{B}(X^*)$ is continuous in the usual (norm) topology.

Let T be the left shift operator on ℓ_2 . Then for every vector x , $\lim_{n \rightarrow \infty} \|T^n x\| = 0$. So the sequence $\{T^n\}$ converges strongly to the zero operator. On the other hand $(T^n)^* = S^n$, where S is the right shift. We know that $\{S^n\}$ does not converge

strongly. (Section 5, Lecture 13). So, the map $A \mapsto A^*$ is not strongly continuous on ℓ_2 .

From the equation (14.4) it is clear that the map $A \mapsto A^*$ is continuous in the weak operator topology of $\mathcal{B}(\mathcal{H})$. This is true, more generally, when $\mathcal{H}$ is replaced by a reflexive Banach space.

Examples

12. Matrices. Let $\mathcal{H}$ be an n -dimensional Hilbert space and choose an orthonormal basis for $\mathcal{H}$. Every operator A on $\mathcal{H}$ has a matrix representation $A = [a_{ij}]$ with respect to this basis. Show that A^* is the operator corresponding to the matrix $[\bar{a}_{ji}]$ in this basis. This is the usual conjugate transpose of A .

13. Integral Operators. Let K be a square integrable kernel on $[0, 1] \times [0, 1]$ and let A_K be the integral operator induced by it on $L_2[0, 1]$, i.e.

$$(A_K f)(x) = \int_0^1 K(x, y) f(y) dy, \quad f \in L_2[0, 1].$$

Let $K^*(x, y) = \overline{K(y, x)}$. Show that the adjoint operator $(A_K)^*$ is the integral operator induced by the kernel K^* . (Use Fubini's Theorem.)

Exercise. Let A be the operator on $L_2[0, 1]$ defined as

$$(Af)(x) = \int_0^x f(t) dt.$$

Show that its adjoint is the operator

$$(A^* f)(x) = \int_x^1 f(t) dt.$$

14. Composition Operators. Let φ be a continuous map of $[0, 1]$ into itself. This

induces a map Φ of $C[0, 1]$ into itself defined as

$$(\Phi f)(t) = f(\varphi(t)), \quad f \in C[0, 1], \quad t \in [0, 1].$$

Show that Φ is a bounded linear operator on $C[0, 1]$, and $\|\Phi\| = 1$.

Recall that by the Riesz Representation Theorem, the dual of the space $C[0, 1]$ is the space of measures on $[0, 1]$. Show that the dual operator Φ^* is the operator defined by the relation

$$(\Phi^* \mu)(E) = \mu(\varphi^{-1}(E)),$$

for every measure μ and every measurable set $E \subset [0, 1]$.

Exercises

15. Let A be an operator on a Banach space X . Then A^{**} is an operator on X^{**} . We identify X as a subspace of X^{**} . Show that the restriction of A^{**} to X is the operator A .

16. We have seen that if A is an invertible operator from X to Y , then A^* is an invertible operator from Y^* to X^* . The converse is also true. The proof is outlined below.

(i) Let A^* be invertible. Then A^* is an open map. So the image of the unit ball $\{g : \|g\| \leq 1\}$ in Y^* under this map contains some ball $\{f : \|f\| \leq c\}$ in X^* .

(ii) For each $x \in X$ we have

$$\begin{aligned} \|Ax\| &= \sup\{|g(Ax)| : g \in Y^*, \|g\| = 1\} \\ &= \sup\{|(A^*g)(x)| : g \in Y^*, \|g\| = 1\} \\ &\geq \sup\{|f(x)| : f \in X^*, \|f\| \leq c\} \end{aligned}$$

$$= c\|x\|.$$

This says that A is *bounded below* and implies that A is one-to-one and its range $\text{ran } A$ is closed.

(iii) It is easy to see that for any $A \in \mathcal{B}(X, Y)$ we have $(\text{ran } A)^\perp = \ker A^*$. So if $\ker A^* = \{0\}$, then $\text{ran } A$ is dense.

(iv) Thus from (ii) we see A is bijective.

Some Special Operators in Hilbert Space

The additional structure in a Hilbert space and its self-duality make the adjoint operation especially interesting. All Hilbert spaces that we consider are over complex scalars except when we say otherwise.

1. Let $\mathcal{H}$ be a Hilbert space. If $\langle x, y \rangle = 0$ for all $y \in \mathcal{H}$, then $x = 0$. Thus an operator A on $\mathcal{H}$ is the zero operator if and only if $\langle Ax, y \rangle = 0$ for all $x, y \in \mathcal{H}$.

Exercise. Let $\mathcal{H}$ be a complex Hilbert space and let $A \in \mathcal{B}(\mathcal{H})$. Show that $A = 0$ iff $\langle Ax, x \rangle = 0$ for all x . (Use polarization.) Find an operator A on $\mathbb{R}^2$ for which $\langle Ax, x \rangle = 0$ for all x and $\|A\| = 1$.

Self-adjoint Operators

2. An operator A on $\mathcal{H}$ is said to be *self-adjoint*, or *Hermitian*, if $A = A^*$.
3. If A is self-adjoint, then for all $x \in \mathcal{H}$

$$\langle Ax, x \rangle = \langle x, Ax \rangle = \overline{\langle Ax, x \rangle}.$$

So, $\langle Ax, x \rangle$ is real. Conversely if $\mathcal{H}$ is a complex Hilbert space and $\langle Ax, x \rangle$ is real for all x , then A is self-adjoint.

4. For every operator A on $\mathcal{H}$, we have

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

and hence,

$$\sup_{\|x\|=1, \|y\|=1} |\langle Ax, y \rangle| = \sup_{\|x\|=1} \|Ax\| = \|A\|. \quad (15.1)$$

5. **Theorem.** If A is self-adjoint, then

$$\|A\| = \sup_{\|x\|=1} |\langle Ax, x \rangle|. \quad (15.2)$$

Proof. Let $M = \sup_{\|x\|=1} |\langle Ax, x \rangle|$. Then for each $y \in \mathcal{H}$, $|\langle Ay, y \rangle| \leq M \|y\|^2$. If x, y are any two vectors, we have

$$\begin{aligned} \langle A(x \pm y), (x \pm y) \rangle &= \langle Ax, x \rangle \pm \langle Ax, y \rangle \pm \langle Ay, x \rangle + \langle Ay, y \rangle \\ &= \langle Ax, x \rangle \pm \langle Ax, y \rangle \pm \langle y, Ax \rangle + \langle Ay, y \rangle \\ &= \langle Ax, x \rangle \pm 2 \operatorname{Re} \langle Ax, y \rangle + \langle Ay, y \rangle. \end{aligned}$$

There are two equations here. Subtract the second of them from the first to get

$$\begin{aligned} 4 \operatorname{Re} \langle Ax, y \rangle &= \langle A(x + y), x + y \rangle - \langle A(x - y), x - y \rangle \\ &\leq M (\|x + y\|^2 + \|x - y\|^2) \\ &= 2M (\|x\|^2 + \|y\|^2). \end{aligned}$$

Replacing x by $e^{i\theta}x$ does not change the right hand side. Choose θ such that $e^{i\theta} \langle Ax, y \rangle \geq 0$. The inequality above then becomes

$$4|\langle Ax, y \rangle| \leq 2M (\|x\|^2 + \|y\|^2).$$

Now take suprema over $\|x\| = \|y\| = 1$ and use (15.1) to get from this $\|A\| \leq M$, and hence $\|A\| = M$. ■

6. **Exercise.** Find an operator on the space $\mathbb{C}^2$ for which the equality (15.2) is not true.

7. If A_1 and A_2 are self-adjoint, then so is $\alpha A_1 + \beta A_2$ for any real numbers α, β . Thus the collection of all self-adjoint operators on $\mathcal{H}$ is a *real* vector space.

8. If A_1, A_2 are self-adjoint, then their product $A_1 A_2$ is self-adjoint if and only if $A_1 A_2 = A_2 A_1$.

Positive Operators

9. Let A be a self-adjoint operator. If for all x , $\langle Ax, x \rangle \geq 0$, we say that A is *positive semidefinite*. If $\langle Ax, x \rangle > 0$ for all nonzero vectors x we say A is *positive definite*. For brevity we will call positive semidefinite operators just *positive* operators; if we need to emphasize that A is positive definite we will say A is strictly positive.

If A is any operator on a complex Hilbert space, then the condition $\langle Ax, x \rangle \geq 0$ for all x implies that A is self-adjoint. The operator A on $\mathbb{R}^2$ defined by the matrix $A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ shows that this is not the case in real Hilbert spaces.

10. We write $A \geq 0$ to mean A is positive. If $A \geq 0$ then $\alpha A \geq 0$ for all positive real numbers α . If A, B are self-adjoint, we say $A \geq B$ if $A - B \geq 0$. This defines a partial order on the collection of self-adjoint operators. If $A_1 \geq B_1$ and $A_2 \geq B_2$, then $A_1 + A_2 \geq B_1 + B_2$.

11. Let A be any operator. Then A^*A and AA^* are positive.

12. Let A, B be operators on $\mathbb{R}^2$ represented by matrices $A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Then $A \geq B$. Is it true that $A^2 \geq B^2$?

Normal Operators

13. An operator A is said to be *normal* if $A^*A = AA^*$. Self-adjoint operators are a very special class of normal operators.

If A is normal, then so is zA for every complex number z . If A_1 and A_2 are normal, then $A_1 + A_2$ is not always normal. The collection of normal operators is a closed subset of $\mathcal{B}(\mathcal{H})$.

14. **Lemma.** A is normal if and only if

$$\|Ax\| = \|A^*x\| \quad \text{for all } x. \quad (15.3)$$

Proof. For any vector x we have the following chain of implications

$$\begin{aligned} \|Ax\|^2 = \|A^*x\|^2 &\Leftrightarrow \langle Ax, Ax \rangle = \langle A^*x, A^*x \rangle \\ &\Leftrightarrow \langle A^*Ax, x \rangle = \langle AA^*x, x \rangle \Leftrightarrow \langle (A^*A - AA^*)x, x \rangle = 0. \end{aligned}$$

The last statement is true for all x if and only if $A^*A = AA^*$. ■

The condition (15.3) is a weakening of the condition $Ax = A^*x$ that defines a self-adjoint operator.

15. **Lemma.** If A is normal, then

$$\|A^2\| = \|A\|^2. \quad (15.4)$$

Proof. By the preceding lemma $\|A(Ax)\| = \|A^*(Ax)\|$ for every x . Hence $\|A^2\| = \|A^*A\|$, and this is equal to $\|A\|^2$ by (14.5). ■

The operator A on $\mathbb{C}^3$ defined by the matrix $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ is not normal but the equality (15.4) is still true for this A .

16. Let A be any operator, and let

$$B = \frac{A + A^*}{2}, \quad C = \frac{A - A^*}{2i}. \quad (15.5)$$

Then B and C are self-adjoint, and

$$A = B + iC. \quad (15.6)$$

This is some times called the *Cartesian decomposition* of A , in analogy with the decomposition $z = x + iy$ of a complex number. B and C are called the *real and imaginary parts* of A .

Exercise. A is normal if and only B and C commute.

Unitary Operators

17. An operator U is *unitary* if

$$U^*U = UU^* = I. \quad (15.7)$$

Clearly unitary operators are normal.

Exercise. Let U be a linear operator on $\mathcal{H}$. Then the following conditions are equivalent:

(i) U is unitary.

(ii) U is invertible and $U^{-1} = U^*$.

(iii) U is surjective and

$$\langle Ux, Uy \rangle = \langle x, y \rangle \quad \text{for all } x \text{ and } y. \quad (15.8)$$

(iv) If $\{e_n\}$ is an orthonormal basis for $\mathcal{H}$, then $\{Ue_n\}$ is also an orthonormal basis.

18. **Exercise.** Show that the condition (15.8) is equivalent to the condition

$$\|Ux\| = \|x\| \quad \text{for all } x. \quad (15.9)$$

In other words U is an *isometry*.

19. The properties listed in (iii) in Exercises 17, say that U preserves all the structures that go into defining a Hilbert space : U is linear, bijective, and preserves inner products. Thus we can say U is an *automorphism* of $\mathcal{H}$. If $\mathcal{H}, \mathcal{K}$ are two Hilbert spaces and if there exists a bijective linear map U from $\mathcal{H}$ to $\mathcal{K}$ that satisfies (15.8) we say $\mathcal{H}$ and $\mathcal{K}$ are *isomorphic* Hilbert spaces.

20. An isometry (on any metric space) is always one-to-one. A linear operator on a finite-dimensional vector space is one-to-one if and only if it is onto. This is not the case if the vector space is infinite-dimensional. For example, the right shift operator S on ℓ_2 is one-to-one but not onto while the left shift T is onto but not one-to-one.

Thus if $\mathcal{H}$ is finite-dimensional and U is a linear operator satisfying (15.8), or the equivalent condition (15.9), then U is unitary. In other words a linear isometry is the same thing as a unitary operator. If $\mathcal{H}$ is infinite-dimensional, then a linear isometry is a unitary operator if and only if it is an onto map.

If $\mathcal{H}$ is finite-dimensional and U any operator on it, then the condition $U^*U = I$ is equivalent to $UU^* = I$. This is not always the case in infinite-dimensional— consider the shift S . So, it is necessary to have the two separate conditions in the definition (15.7).

21. **Lemma.** An operator A on $\mathcal{H}$ is an isometry if and only if

$$A^*A = I. \quad (15.10)$$

Proof. We have the implications

$$\begin{aligned} \|Ax\|^2 = \|x\|^2 &\Leftrightarrow \langle Ax, Ax \rangle = \langle x, x \rangle \Leftrightarrow \langle A^*Ax, x \rangle = \langle x, x \rangle \\ &\Leftrightarrow \langle (A^*A - I)x, x \rangle = 0. \end{aligned}$$

■

If

$$AA^* = I \quad (15.11)$$

we say A is *co-isometry*. This is equivalent to saying A^* is an isometry. An operator is unitary if it is both an isometry and a co-isometry.

Projections and Subspaces

22. Recall our discussion of projections in Lecture 11, Sections 18, 19. A linear map P on $\mathcal{H}$ is called a projection if it is idempotent ($P^2 = P$). If $\mathcal{S} = \text{ran } P$ and $\mathcal{S}' = \ker P$, then $\mathcal{H} = \mathcal{S} + \mathcal{S}'$, and P is the projection on $\mathcal{S}$ along $\mathcal{S}'$. The operator $I - P$ is also a projection, its range is $\mathcal{S}'$ and kernel $\mathcal{S}$. For example, the operator P on $\mathbb{C}^2$ corresponding to the matrix $P = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ is idempotent. Its range is the space $\mathcal{S} = \{(x, 0) : x \in \mathbb{C}\}$, and its kernel $\mathcal{S}' = \{(x, -x) : x \in \mathbb{C}\}$. A special property characterises *orthogonal* projections: those for which $\mathcal{S}' = \mathcal{S}^\perp$.

Proposition. An idempotent operator P on $\mathcal{H}$ is an orthogonal projection if and only if it is self-adjoint.

Proof. Let $x \in \mathcal{S}$, $y \in \mathcal{S}'$. Then $Px = x$, $Py = 0$. So, if $P^* = P$, we have $\langle x, y \rangle = \langle Px, y \rangle = \langle x, Py \rangle = 0$. This shows $\mathcal{S}' = \mathcal{S}^\perp$.

Conversely let z be any vector in $\mathcal{H}$, and split it as $z = x + y$ with $x \in \mathcal{S}$, $y \in \mathcal{S}^\perp$. Let $Pz = x$. Then for any two vectors z_1, z_2

$$\begin{aligned} \langle Pz_1, z_2 \rangle &= \langle x_1, x_2 + y_2 \rangle = \langle x_1, x_2 \rangle = \langle x_1 + y_1, x_2 \rangle \\ &= \langle z_1, Pz_2 \rangle. \end{aligned}$$

This shows $P^* = P$. ■

23. When we talk of Hilbert spaces we usually mean an orthogonal projection when we say a projection. To each closed linear subspace $\mathcal{S}$ in $\mathcal{H}$ there corresponds a unique (orthogonal) projection P and vice versa. There is an intimate connection between (geometric) properties of subspaces and the (algebraic) properties of projections corresponding to them.

24. **Exercise.** Every orthogonal projection is a positive operator.

25. Let A be an operator on $\mathcal{H}$. A subspace $\mathcal{M}$ of $\mathcal{H}$ is said to be *invariant* under A if A maps $\mathcal{M}$ into itself. If both $\mathcal{M}$ and $\mathcal{M}^\perp$ are invariant under A , we say $\mathcal{M}$ *reduces* A , or $\mathcal{M}$ is a *reducing subspace* for A .

Exercise. A closed subspace $\mathcal{M}$ is invariant under A if and only if $\mathcal{M}^\perp$ is invariant under A^* . Thus $\mathcal{M}$ reduces A if and only if it is invariant under both A and A^* .

26. Let A be the operator on $\mathbb{C}^2$ corresponding to the matrix $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. Then the space $\mathcal{M} = \{(x, 0) : x \in \mathbb{C}\}$ is invariant under A but does not reduce A .

Let $\mathcal{M}$ be the orthogonal complement of the 1-dimensional space spanned by the

basis vector e_1 in ℓ_2 . Then $\mathcal{M}$ is invariant under the right shift operator S but not under its adjoint S^* . So $\mathcal{M}$ does not reduce A .

27. Theorem. Let P be the orthogonal projection onto the subspace $\mathcal{M}$ of $\mathcal{H}$. Then $\mathcal{M}$ is invariant under an operator A , if and only if $AP = PAP$; and $\mathcal{M}$ reduces A if and only if $AP = PA$.

Proof. For each $x \in \mathcal{H}$, $Px \in \mathcal{M}$. So, if $\mathcal{M}$ is invariant under A , then $A(Px) \in \mathcal{M}$, and hence $PAPx = APx$. In other words $PAP = AP$. Conversely, if $PAP = AP$, then for every x in $\mathcal{M}$ we have $Ax = APx = PAPx$, and this is a vector in $\mathcal{M}$. This proves the first part of the theorem. Use this to prove the second part as follows :

$$\begin{aligned} \mathcal{M} \text{ reduces } A &\Leftrightarrow AP = PAP \text{ and } A^*P = PA^*P \\ &\Leftrightarrow AP = PAP \text{ and } PA = PAP \\ &\Leftrightarrow AP = PA. \end{aligned}$$

We have used the property $P^* = P$ at the second step here, and $P^2 = P$ at the third.

Exercises

28. Let P_1, P_2 be (orthogonal) projections. Show that P_1P_2 is a projection if and only if $P_1P_2 = P_2P_1$. In this case $\text{ran } P_1P_2 = \text{ran } P_1 \cap \text{ran } P_2$.

29. If $P_1P_2 = 0$, we say the projections P_1 and P_2 are *mutually orthogonal*. Show that this condition is equivalent to saying that the ranges of P_1 and P_2 are mutually orthogonal subspaces. If P_1 and P_2 are projections, then $P_1 + P_2$ is a projection if and only if P_1 and P_2 are mutually orthogonal. In this case $\text{ran } (P_1 + P_2) = \text{ran } P_1 \oplus \text{ran } P_2$.

30. Let P_1, P_2 be projections. Show that the following conditions are equivalent

(i) $\text{ran } P_1 \subseteq \text{ran } P_2$.

(ii) $P_1 \leq P_2$.

(iii) $P_1 P_2 = P_2 P_1 = P_1$.

(iv) $\|P_1 x\| \leq \|P_2 x\|$ for all x .

31. If P_1 and P_2 , are projections, then $P_1 - P_2$ is a projection if and only if $P_2 \leq P_1$.

In this case $\text{ran } (P_1 - P_2) = \text{ran } P_1 \cap (\text{ran } P_2)^\perp$.

32. Show that the Laplace transform operator $\mathcal{L}$ defined in Section 19 of Lecture 3 is a self-adjoint operator on $L_2(\mathbb{R}_+)$.

33. The *Hilbert–Hankel operator* H is the integral kernel operator on $L_2(0, \infty)$ defined as

$$Hf(x) = \int_0^\infty \frac{f(y)}{x+y} dy.$$

Show that $H = \mathcal{L}^2$, where $\mathcal{L}$ is the Laplace transform operator. This shows that $\|H\| = \pi$.

The Resolvent and The Spectrum

A large, and the most important, part of operator theory is the study of the spectrum of an operator. In finite dimensions, this is the set of eigenvalues of A . In infinite dimensions there are complications that arise from the fact that an operator could fail to be invertible in different ways. Finding the spectrum is not an easy problem even in the finite-dimensional case; it is much more difficult in infinite dimensions.

Banach space-valued maps

1. Let $x(t)$ be a map from an interval $[a, b]$ of the real line into a Banach space X . It is obvious how to define continuity of this map. If $\|x(t) - x(t_0)\| \rightarrow 0$ as $t \rightarrow t_0$, we say $x(t)$ is *continuous* at t_0 .

If $x(t)$ is continuous at t_0 , then clearly for each $f \in X^*$, the (complex-valued) function $f(x(t))$ is continuous at t_0 . We say that $x(t)$ is *weakly continuous* at t_0 if $f(x(t))$ is continuous at t_0 for all $f \in X^*$. (If emphasis is needed we call a continuous map *strongly continuous*.)

Strong and weak differentiability can be defined in the same way. If t_0 is a point in (a, b) we consider the limits

$$\lim_{h \rightarrow 0} \frac{x(t_0 + h) - x(t_0)}{h}$$

and

$$\lim_{h \rightarrow 0} \frac{f(x(t_0 + h)) - f(x(t_0))}{h}, \quad f \in X^*.$$

If the first limit exists, we say $x(t)$ is (*strongly*) *differentiable* at t_0 . If the second limit exists for every $f \in X^*$, we say $x(t)$ is *weakly differentiable* at t_0 . Clearly strong differentiability implies weak differentiability. The converse is not always true when X is infinite-dimensional.

2. Example. Let $X = L_2(\mathbb{R})$. Choose and fix a nonzero element g of X . Define a map $t \mapsto f(t)$ from $(-1, 1)$ into X as follows. Let $f(0)$ be the zero function and for $t \neq 0$ let

$$f(t)(u) = t e^{-iu/t} g(u).$$

Let φ be any element of $X^*(= L_2(\mathbb{R}))$. Then

$$\frac{\varphi(f(t)) - \varphi(f(0))}{t} = \int e^{-iu/t} g(u) \overline{\varphi(u)} du. \quad (16.1)$$

The integral on the right is the Fourier transform of the function $g\bar{\varphi}$ at the point $1/t$. Since g and φ are in $L_2(\mathbb{R})$, the function $g\bar{\varphi}$ is in $L_1(\mathbb{R})$. Hence, by the Riemann-Lebesgue Lemma, its Fourier transform has limit 0 at $\pm\infty$; i.e.,

$$\lim_{t \rightarrow 0} \widehat{g\bar{\varphi}}(1/t) = 0.$$

So from (16.1) we see that $f(t)$ is weakly differentiable at $t = 0$, and the weak derivative is the zero function. If the map $f(t)$ had a strong derivative at 0, it would have to be equal to the weak derivative. But for all $t \neq 0$,

$$\left\| \frac{f(t) - f(0)}{t} \right\| = \|g\| \neq 0.$$

So the map is not strongly differentiable at $t = 0$.

3. Let G be any open connected set of the complex plane and let $x(z)$ be a map from G into X . If for every point z in G the limit

$$\lim_{h \rightarrow 0} \frac{x(z+h) - x(z)}{h}$$

exists we say $x(z)$ is *strongly analytic* on G . If for every $z \in G$ and $f \in X^*$, the limit

$$\lim_{h \rightarrow 0} \frac{f(x(z+h)) - f(x(z))}{h}$$

exists we say $x(z)$ is *weakly analytic* on G .

As for ordinary complex functions, this analyticity turns out to be a much stronger property than in the real case. Here the strong and the weak notions coincide. So questions of analyticity of the Banach space-valued map $x(z)$ are reduced to those about the family of complex-valued maps $f(x(z))$, $f \in X^*$.

4. Theorem. Let $x(z)$ be a weakly analytic map from a complex region G into a Banach space X . Then $x(z)$ is strongly analytic.

Proof. Let f be any element of X^* . Then $(f \circ x)(z) = f(x(z))$ is an analytic function on G . Let $(f \circ x)'(z)$ be its derivative. Let z_0 be any point in G and Γ a closed curve in G with winding number 1 around z_0 and winding number 0 around any point outside G . By Cauchy's integral formula

$$f(x(z_0)) = \frac{1}{2\pi i} \int_{\Gamma} \frac{f(x(\zeta))}{\zeta - z_0} d\zeta.$$

Hence for small h ,

$$\begin{aligned} & \frac{f(x(z_0+h)) - f(x(z_0))}{h} - (f \circ x)'(z_0) \\ &= \frac{1}{2\pi i h} \int_{\Gamma} f(x(\zeta)) \left[\frac{1}{\zeta - z_0 - h} - \frac{1}{\zeta - z_0} \right] d\zeta - \frac{1}{2\pi i} \int_{\Gamma} \frac{f(x(\zeta))}{(\zeta - z_0)^2} d\zeta \\ &= \frac{h}{2\pi i} \int_{\Gamma} \frac{f(x(\zeta))}{(\zeta - z_0 - h)(\zeta - z_0)^2} d\zeta. \end{aligned} \tag{16.2}$$

Since Γ is a compact set and $f(x(\cdot))$ a continuous functions, the supremum

$$\sup_{\zeta \in \Gamma} |f(x(\zeta))| = C_f$$

is finite. Hence, by the uniform boundedness principle the supremum

$$\sup_{\|f\| \leq 1} \sup_{\zeta \in \Gamma} |f(x(\zeta))| = C$$

is finite. (Think of $x(\zeta)$ as linear functionals on X^* .) Hence the quantity in (16.2) is bounded by

$$\frac{C|h|}{2\pi} \int_{\Gamma} \frac{|d\zeta|}{|(\zeta - z_0 - h)(\zeta - z_0)^2|}$$

for all f with $\|f\| \leq 1$. As $h \rightarrow 0$ this goes to 0, and the convergence is uniform for $\|f\| \leq 1$. Hence the limit

$$\lim_{h \rightarrow 0} \frac{x(z_0 + h) - x(z_0)}{h}$$

exists in X (see (4.2)). Thus $x(z)$ is strongly analytic at z_0 . ■

Exercise. The space $\mathcal{B}(X)$ has three topologies that are of interest: norm topology, strong operator topology, and weak operator topology. Define analyticity of a map $z \mapsto A(z)$ from a complex G into $\mathcal{B}(X)$ with respect to these topologies. Show that the three notions of analyticity are equivalent.

Resolvents

6. Let $A \in \mathcal{B}(X)$ and let λ be any complex number. It is customary to write the operator $A - \lambda I$ as $A - \lambda$.

The *resolvent set* of A is the collection of all complex numbers λ for which $A - \lambda$ is invertible. Note that if $(A - \lambda)^{-1}$ exists, it is a bounded operator. (The Inverse Mapping Theorem, Lecture 6.) We write $\rho(A)$ for the resolvent set of A . The operator

$$R_\lambda(A) = (A - \lambda)^{-1}, \quad \lambda \in \rho(A)$$

is called the *resolvent* of A at λ .

If $|\lambda| > \|A\|$, then $\|A/\lambda\| < 1$. Hence $I - A/\lambda$ is invertible. (See Chapter 13, Theorem 14.) Hence the operator $A - \lambda = \lambda(A/\lambda - 1)$ is also invertible. We have

$$(A - \lambda)^{-1} = -\frac{1}{\lambda} \sum_{n=0}^{\infty} \left(\frac{A}{\lambda}\right)^n \quad \text{for } |\lambda| > \|A\|. \quad (16.3)$$

Thus $\rho(A)$ is a nonempty set.

7. The Resolvent Identity. Let λ, μ be any two points in $\rho(A)$. Then

$$R_\lambda(A) - R_\mu(A) = (\lambda - \mu)R_\lambda(A)R_\mu(A). \quad (16.4)$$

Proof. A simple algebraic manipulation using the definition of the resolvent shows that

$$\begin{aligned} R_\lambda(A) - R_\mu(A) &= R_\lambda(A) [I - (A - \lambda)R_\mu(A)] \\ &= R_\lambda(A) [I - \{(A - \mu) - (\lambda - \mu)\}R_\mu(A)] \\ &= R_\lambda(A) [(\lambda - \mu)R_\mu(A)]. \end{aligned}$$

■

8. Corollary. The family $\{R_\lambda(A) : \lambda \in \rho(A)\}$ is a commuting family; i.e., any two elements of this family commute with each other.

Exercise. Show that $R_\lambda(A)$ and A commute for all $\lambda \in \rho(A)$.

9. Theorem. For each $A \in \mathcal{B}(X)$ the set $\rho(A)$ is an open subset of $\mathbb{C}$, and the map $\lambda \mapsto R_\lambda(A)$ is an analytic map from $\rho(A)$ into $\mathcal{B}(X)$.

Proof. The argument that was used to show that the set of invertible operators is open in $\mathcal{B}(X)$ can be modified to show $\rho(A)$ is an open set. Let $\lambda_0 \in \rho(A)$. We want to show that $\lambda \in \rho(A)$ if λ is close to λ_0 . We have the identity

$$A - \lambda = (A - \lambda_0) \left[I - (\lambda - \lambda_0)(A - \lambda_0)^{-1} \right]$$

$$= (A - \lambda_0) [I - (\lambda - \lambda_0)R_{\lambda_0}(A)].$$

The term inside the square brackets is invertible provided

$$\|(\lambda - \lambda_0)R_{\lambda_0}(A)\| < 1,$$

i.e., $|\lambda - \lambda_0| < 1/\|R_{\lambda_0}(A)\|$. Thus if λ satisfies this inequality, then it belongs to $\rho(A)$. Hence $\rho(A)$ is open. Further, this shows

$$R_\lambda(A) = \sum_{n=0}^{\infty} (\lambda - \lambda_0)^n [R_{\lambda_0}(A)]^{n+1}.$$

Thus $R_\lambda(A)$ is represented by a convergent power series in $(\lambda - \lambda_0)$. Hence it is analytic. ■

10. From the series (16.3) it is clear that

$$\lim_{|\lambda| \rightarrow \infty} \|R_\lambda(A)\| = 0.$$

So, by Liouville's Theorem $\rho(A)$ can not be the entire complex plane. (A bounded entire function is a constant.)

The Spectrum

11. The complement of the resolvent set in the complex plane is called the *spectrum* of A , and is denoted by $\sigma(A)$.

We have seen that this is a nonempty compact subset of $\mathbb{C}$. We know that

$$\sigma(A) \subset \{\lambda : |\lambda| \leq \|A\|\}.$$

12. If X is a finite-dimensional space, then $\sigma(A)$ is a finite set. Its elements are the eigenvalues of A . Every operator on an n -dimensional space has at least one and at most n eigenvalues.

13. Let S be the right shift operator on ℓ_p , $1 \leq p \leq \infty$. For any complex number λ the equation $Sx = \lambda x$, i.e.,

$$(0, x_1, x_2) = \lambda(x_1, x_2, \dots)$$

can never be satisfied by any nonzero vector x . So, S does not have any eigenvalue. At the same time we do know $\sigma(S)$ is not an empty set. So, a point can be in the spectrum of an operator A without being an eigenvalue. This is because $A - \lambda$ can be injective without being invertible.

Spectral Radius

14. The *spectral radius* of A is the number

$$\text{spr}(A) = \sup \{|\lambda| : \lambda \in \sigma(A)\}.$$

This is the radius of the smallest disk centered at the origin that contains the spectrum of A . We know that

$$\text{spr}(A) \leq \|A\|. \quad (16.5)$$

The spectral radius of a nilpotent matrix is 0; so the two sides of (16.5) need not be equal.

15. Consider a power series $\sum_{n=0}^{\infty} A_n z^n$, where $A_n \in \mathcal{B}(X)$, and $z \in \mathbb{C}$. It is easy to see (following the usual arguments for the ordinary power series $\sum_{n=0}^{\infty} a_n z^n$) that the series converges uniformly on every closed subset of an open disk of radius R centred at the origin, where

$$\frac{1}{R} = \overline{\lim} \|A_n\|^{1/n}. \quad (16.6)$$

The series diverges for all z outside this disk, and also for at least one point on the boundary of the disk.

16. Consider the series (16.3) — a power series in $1/\lambda$. This series converges when

$$|\lambda| > \overline{\lim} \|A^n\|^{1/n}$$

and then defines $(A - \lambda)^{-1}$. It does not converge for at least one point λ with $|\lambda| = \overline{\lim} \|A^n\|^{1/n}$. Hence

$$\text{spr}(A) = \overline{\lim} \|A^n\|^{1/n}. \quad (16.7)$$

Much more interesting is the fact that $\overline{\lim}$ here is actually the limit of the (convergent) sequence $\|A^n\|^{1/n}$.

17. The Spectral Radius Formula. For every $A \in \mathcal{B}(X)$, the sequence $\|A^n\|^{1/n}$ converges, and

$$\lim \|A^n\|^{1/n} = \text{spr}(A). \quad (16.8)$$

Proof. For each $n > 1$ we have the factorings

$$\begin{aligned} A^n - \lambda^n &= (A - \lambda)(A^{n-1} + \lambda A^{n-2} + \cdots + \lambda^{n-1}) \\ &= (A^{n-1} + \lambda A^{n-2} + \cdots + \lambda^{n-1})(A - \lambda). \end{aligned}$$

So, if $A^n - \lambda^n$ were invertible, then $A - \lambda$ would have a left inverse and a right inverse, and would therefore be invertible. By contraposition if $A - \lambda$ is not invertible, then nor is $A^n - \lambda^n$. In other words, if $\lambda \in \sigma(A)$, then $\lambda^n \in \sigma(A^n)$. Hence $|\lambda^n| \leq \|A^n\|$; i.e., $|\lambda| \leq \|A^n\|^{1/n}$ for all n . This shows that $\text{spr}(A) \leq \underline{\lim} \|A^n\|^{1/n}$. But we have already obtained the equality (16.7). ■

18. Our proof shows that

$$\text{spr}(A) = \lim \|A^n\|^{1/n} = \inf \|A^n\|^{1/n}.$$

This may lead one to believe that the sequence $\|A^n\|^{1/n}$ is monotonically decreasing.

This is, however, not always true. Consider the operator A on the Hilbert space $\mathbb{C}^2$

given by the matrix $A = \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}$. In this case $\|A^3\|^{1/3}$ is bigger than $\|A^2\|^{1/2}$.

19. **Exercise.** If A is a normal operator on a Hilbert space. Then $\text{spr}(A) = \|A\|$. (Use Lemma 15 of Lecture 15. In a finite-dimensional space prove this using the spectral theorem for normal operators.) Find an operator A that is not normal but has $\text{spr}(A) = \|A\|$.

20. **Spectral Mapping Theorem for Polynomials.** Let p be any polynomial, and A any operator. Then

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

Proof. Let $\lambda \in \sigma(A)$. If p is a polynomial of degree $n \geq 1$, then $p(z) - p(\lambda)$ is a polynomial of degree n with λ as a root and we can factor $p(z) - p(\lambda)$ as $(z - \lambda)q(z)$ where q is a polynomial of degree $n - 1$. Then

$$p(A) - p(\lambda) = (A - \lambda)q(A) = B, \text{ say.}$$

If B were invertible, then the equation $BB^{-1} = B^{-1}B = I$ can be written as

$$(A - \lambda)q(A)B^{-1} = B^{-1}q(A)(A - \lambda).$$

This would mean $A - \lambda$ is invertible, which is not possible if $\lambda \in \sigma(A)$. Thus B is not invertible; i.e., $p(\lambda) \in \sigma(p(A))$. So $p(\sigma(A)) \subset \sigma(p(A))$.

Let $\lambda \in \sigma(p(A))$. Factorise the polynomial $p(z) - \lambda$ into linear factors, and write

$$p(A) - \lambda = c(A - \lambda_1)(A - \lambda_2) \dots (A - \lambda_n).$$

Since the operator $p(A) - \lambda$ is not invertible, one of the factors $A - \lambda_j$ is not invertible. Thus $\lambda_j \in \sigma(A)$ and also $p(\lambda_j) - \lambda = 0$. This shows $\lambda = p(\lambda_j)$ for some $\lambda_j \in \sigma(A)$. Hence $\sigma(p(A)) \subset p(\sigma(A))$. ■

21. **Exercise.** If A is an invertible operator, then

$$\sigma(A^{-1}) = [\sigma(A)]^{-1} := \{1/\lambda : \lambda \in \sigma(A)\}.$$

22. **Exercise.** For every $A \in \mathcal{B}(X)$, we have

$$\sigma(A) = \sigma(A^*).$$

$$[R_\lambda(A)]^* = R_\lambda(A^*) \quad \text{for all } \lambda \in \rho(A).$$

If X is a Hilbert space, then

$$\sigma(A^*) = \overline{\sigma(A)}.$$

$$[R_\lambda(A)]^* = R_{\overline{\lambda}}(A^*) \quad \text{for all } \lambda \in \rho(A).$$

Here the bar denotes complex conjugation.

Subdivision of the Spectrum

Let S be the right shift operator on the space ℓ_1 . Since $\|S\| = 1$ the spectrum $\sigma(S)$ is contained in the closed unit disk D . We have seen that S has no eigenvalue. The adjoint of S is the left shift operator T on the space ℓ_∞ . If λ is any complex number with $|\lambda| \leq 1$, then the vector $x_\lambda = (1, \lambda, \lambda^2, \dots)$ is in ℓ_∞ and $Tx_\lambda = \lambda x_\lambda$. Thus *every* point λ in the disk D is an eigenvalue of T . This shows also that $\sigma(S) = \sigma(T) = D$.

To understand how a point λ gets into the spectrum of an operator A it is helpful to divide the spectrum into different parts, and to study A and A^* together.

1. The Point Spectrum. A number λ is an *eigenvalue* of A if there exists a nonzero vector x such that $(A - \lambda)x = 0$. The set of all eigenvalues of A is called the *point spectrum* of A , and is written as $\sigma_p(A)$.

We have seen an example where $\sigma_p(A) = \emptyset$, and another where $\sigma_p(A) = \sigma(A)$.

2. We say an operator A is *bounded below* if there exists a positive real number α such that $\|Ax\| \geq \alpha\|x\|$ for all $x \in X$.

If A is bounded below, then A is one-to-one. The operator A on ℓ_p , $1 \leq p \leq \infty$ defined by $Ae_n = e_n/n$ is one-to-one but is not bounded below.

If A is invertible, then $\|Ax\| \geq \frac{1}{\|A^{-1}\|}\|x\|$. Thus A is bounded below.

3. **Lemma.** If A is bounded below, then its range, $\text{ran } A$, is closed.

Proof. Let $\{Ax_n\}$ be a Cauchy sequence in $\text{ran } A$. Since A is bounded below, the sequence $\{x_n\}$ is also a Cauchy sequence. Let x be the limit of this sequence. Then Ax is the limit of $\{Ax_n\}$ and is a point in $\text{ran } A$. ■

4. **Theorem.** An operator A on the Banach space X is invertible if and only if it is bounded below and its range is dense in X .

Proof. If A is invertible, then it is bounded below, and its range is all of X , not just dense in X .

If A is bounded below, then it is one-to-one, and by Lemma 3 its range is closed. So, if the range is dense it has to be all of X . Hence A is invertible. ■

5. This simple theorem leads to a useful division of the spectrum into two parts (not always disjoint).

Theorem 4 tells us that $\lambda \in \sigma(A)$ if *either* $A - \lambda$ is not bounded below *or* $\text{ran } (A - \lambda)$ is not dense. (The possibilities are not mutually exclusive.) The set

$$\sigma_{\text{app}}(A) := \{\lambda : A - \lambda \text{ is not bounded below}\}$$

is called the *approximate point spectrum* of A . Its members are called *approximate eigenvalues* of A .

Note that λ is an approximate eigenvalue if and only if there exists a sequence of unit vectors $\{x_n\}$ such that $(A - \lambda)x_n \rightarrow 0$. Every eigenvalue of A is also an approximate eigenvalue.

The set

$$\sigma_{\text{comp}}(A) := \{\lambda : \text{ran } (A - \lambda) \text{ is not dense in } X\}$$

is called the *compression spectrum* of A .

6. Finer subdivisions are sometimes useful. The set

$$\sigma_{\text{res}}(A) := \sigma_{\text{comp}}(A) \setminus \sigma_p(A),$$

called the *residual spectrum* of A , is the set of those points in the compression spectrum that are not eigenvalues. The set

$$\sigma_{\text{cont}}(A) := \sigma_{\text{app}}(A) \setminus [\sigma_p(A) \cup \sigma_{\text{res}}(A)]$$

is called the *continuous spectrum* of A . It consists of those approximate eigenvalues that are neither eigenvalues nor points of the compression spectrum.

Warning: This terminology is unfortunately not standardised. In particular, the term continuous spectrum has a different meaning in other books. The books by Yosida, Hille and Phillips, and Halmos use the word in the same sense as we have done. Those by Kato, Riesz and Nagy, and Reed and Simon use it in a different sense (that we will see later).

7. We have observed that for every operator A on a Banach space $\sigma(A) = \sigma(A^*)$. This equality does not persist for parts of the spectrum.

Theorem. (i) $\sigma_{\text{comp}}(A) \subset \sigma_p(A^*)$. (ii) $\sigma_p(A) \subset \sigma_{\text{comp}}(A^*)$.

Proof. Let M be the closure of the space $\text{ran}(A - \lambda)$. If $\lambda \in \sigma_{\text{comp}}(A)$, then M is a proper subspace of X . Hence there exists a nonzero linear functional f on X that vanishes on M . Write this in the notation (14.2) as

$$\langle f, (A - \lambda)x \rangle = 0 \quad \text{for all } x \in X.$$

Taking adjoints this says

$$\langle (A^* - \lambda)f, x \rangle = 0 \quad \text{for all } x \in X.$$

Thus f is an eigenvector and λ an eigenvalue of A^* . This proves (i).

If $\lambda \in \sigma_p(A)$, then there exists a nonzero vector x in X such that $(A - \lambda)x = 0$.

Hence

$$\begin{aligned}\langle f, (A - \lambda)x \rangle &= 0 \quad \text{for all } f \in X^*, \text{ i.e.,} \\ \langle (A^* - \lambda)f, x \rangle &= 0 \quad \text{for all } f \in X^*.\end{aligned}$$

This says that $g(x) = 0$ for all $g \in \text{ran } (A^* - \lambda)$. If the closure of $\text{ran } (A^* - \lambda)$ were the entire space X^* , this would mean $g(x) = 0$ for all $g \in X^*$. But the Hahn-Banach Theorem guarantees the existence of at least one linear functional g that does not vanish at x . So $\text{ran } (A^* - \lambda)$ can not be dense. This proves (ii). ■

8. **Exercise.** If A is an operator on a Hilbert space $\mathcal{H}$, then

$$\begin{aligned}\sigma_p(A^*) &= \overline{\sigma_{\text{comp}}(A)} \\ \sigma(A^*) &= \sigma_{\text{app}}(A^*) \cup \overline{\sigma_{\text{app}}(A)}.\end{aligned}$$

Here the bar denotes complex conjugation operation. (Recall that we identified $\mathcal{H}$ with $\mathcal{H}^*$ and A^{**} with A ; in this process linearity was replaced by conjugate linearity.)

The set $\sigma_p(A)$ consists of eigenvalues—objects familiar to us; the set $\sigma_{\text{app}}(A)$ is a little more complicated but still simpler than the remaining part of the spectrum. The relations given in Theorem 7 and Exercise 8 are often helpful in studying the more complicated parts of the spectrum of A in terms of the simpler parts of the spectrum of A^* .

9. **Exercise.** Let A be any operator on a Banach space. Then $\sigma_{\text{app}}(A)$ is a closed set.

10. **Proposition.** Let $\{\lambda_n\}$ be a sequence in $\rho(A)$ and suppose λ_n converges to λ . If the sequence $\{R_{\lambda_n}(A)\}$ is bounded in $\mathcal{B}(X)$, then $\lambda \in \rho(A)$.

Proof. By the Resolvent Identity

$$\|R_{\lambda_n}(A) - R_{\lambda_m}(A)\| = |\lambda_n - \lambda_m| \|R_{\lambda_n}(A)R_{\lambda_m}(A)\|.$$

Hence under the given conditions $R_{\lambda_n}(A)$ is a Cauchy sequence. Let R be the limit of this sequence. Then

$$R(A - \lambda) = \lim_{n \rightarrow \infty} R_{\lambda_n}(A)(A - \lambda_n) = I.$$

In the same way $(A - \lambda)R = I$. So $A - \lambda$ is invertible, and $\lambda \in \rho(A)$. ■

11. Theorem. The boundary of the set $\sigma(A)$ is contained in $\sigma_{\text{app}}(A)$.

Proof. If λ is on the boundary of $\sigma(A)$, then there exists a sequence $\{\lambda_n\}$ in $\rho(A)$ converging to λ . So, by Proposition 10, $\{\|(A - \lambda_n)^{-1}\|\}$ is an unbounded sequence. So, it contains a subsequence, again denoted by $\{\lambda_n\}$, such that for every n , there exists a unit vector x_n for which $\|(A - \lambda_n)^{-1}x_n\| \geq n$. Let

$$y_n = \frac{(A - \lambda_n)^{-1}x_n}{\|(A - \lambda_n)^{-1}x_n\|}.$$

Then $\|y_n\| = 1$, and $\|(A - \lambda_n)y_n\| \leq \frac{1}{n}$. Since $\|(A - \lambda)y_n\| \leq \|(A - \lambda_n)y_n\| + |\lambda - \lambda_n|$, this shows $(A - \lambda)y_n \rightarrow 0$. Hence $\lambda \in \sigma_{\text{app}}(A)$. ■

12. Exercise. (The shift operator again)

Let T be the left shift on ℓ_1 . Then $T^* = S$ the right shift on ℓ_∞ . Since $\|T\| = 1$, we know that $\sigma(T)$ is contained in the closed unit disk D . From Exercise 16.22 we know that $\sigma(S) = \sigma(T)$. Fill in the details in the statements that follow.

(i) If $|\lambda| < 1$, then $x_\lambda := (1, \lambda, \lambda^2, \dots)$ is in ℓ_1 and is an eigenvector of T for eigenvalue λ . Thus the interior D° is contained in $\sigma_p(T)$.

(ii) This shows that $\sigma(T) = \sigma_{\text{app}}(T) = D$.

(iii) If $|\lambda| = 1$, then there does not exist any vector x in ℓ_1 for which $Tx = \lambda x$.

Thus no point on the boundary of D is in $\sigma_p(T)$.

(iv) The point spectrum $\sigma_p(S)$ is empty. Hence the compression spectrum $\sigma_{\text{comp}}(T)$ is empty. (Theorem 7.)

(v) $\sigma_{\text{cont}}(T) = \text{Bdry } D$ (the boundary of D).

(vi) $D^\circ \subset \sigma_{\text{comp}}(S) = \sigma_{\text{res}}(S)$.

(vii) Let $|\lambda| = 1$. Then $u = (\bar{\lambda}, \bar{\lambda}^2, \dots)$ is in ℓ_∞ . Let y be any element of ℓ_∞ and let $x = (S - \lambda)y$. From the relation

$$(x_1, x_2, x_3, \dots) = (-\lambda y_1, y_1 - \lambda y_2, y_2 - \lambda y_3, \dots)$$

calculate y_n inductively to see that

$$y_n = -\bar{\lambda}^{n+1} \sum_{j=1}^n \lambda^j x_j.$$

If $\|x - u\|_\infty \leq 1/2$, then

$$\text{Re } \lambda^j x_j = \text{Re } \lambda^j u_j - \text{Re } \lambda^j (u_j - x_j) \geq 1 - \|u - x\| \geq 1/2. \quad (17.1)$$

Hence $|y_n| \geq n/2$. But that cannot be true if $y \in \ell_\infty$. So we must have $\|x - u\|_\infty > 1/2$ for every $x \in \text{ran } (S - \lambda)$. Hence $\lambda \in \sigma_{\text{comp}}(S)$.

(viii) $D = \sigma_{\text{comp}}(S) = \sigma_{\text{res}}(S)$.

The conclusion of this exercise is summarised in the table :

Space	Operator	σ	σ_p	σ_{app}	σ_{comp}	σ_{res}	σ_{cont}
ℓ_1	T	D	D°	D	ϕ	ϕ	Bdry D
ℓ_∞	S	D	ϕ	Bdry D	D	D	ϕ

13. Exercise. Find the various parts of the spectra of the right and left shift operators on ℓ_p , $1 \leq p \leq \infty$.

14. **Exercise.** Let P be a projection operator in any Banach space. What is the spectrum of P , and what are the various parts of $\sigma(P)$?

Exercise. (Spectrum of a product)

(i) Suppose $I - AB$ is invertible and let $X = (I - AB)^{-1}$. Show that

$$(I - BA)(I + BXA) = I = (I + BXA)(I - BA).$$

Hence $I - BA$ is invertible.

(ii) Show that the sets $\sigma(AB)$ and $\sigma(BA)$ have the same elements with one possible exception: the point zero.

(iii) The statement (ii) is true if σ is replaced by σ_p .

(iv) Give an example showing that the point 0 is exceptional.

(v) If A, B are operators on a finite-dimensional space, then $\sigma(AB) = \sigma(BA)$.

More is true in this case. Each eigenvalue of AB is an eigenvalue of BA with the same multiplicity.

16. **Exercise.** Let $X = C[0, 1]$ and let A be the operator on X defined as

$$(Af)(x) = \int_0^x f(t)dt \quad \text{for all } f \in X.$$

Show that $\|A\| = 1$, $\text{spr}(A) = 0$, $\sigma_{\text{res}}(A) = \{0\}$.

Spectra of Normal Operators

In Lecture 15 we studied normal operators in Hilbert spaces. For this class the spectrum is somewhat simpler.

1. **Theorem.** Every point in the spectrum of a normal operator is an approximate eigenvalue.

Proof. If A is a normal operator, then so is $A - \lambda$ for every complex number λ . So $\|(A - \lambda)x\| = \|(A - \lambda)^*x\| = \|(A^* - \bar{\lambda})x\|$ for all vectors x . Thus λ is an eigenvalue of A if and only if $\bar{\lambda}$ is an eigenvalue of A^* . By Exercise 8 in Lecture 17, this means that $\sigma_p(A) = \sigma_{\text{comp}}(A)$. In other words the residual spectrum of A is empty. The rest of the spectrum is just $\sigma_{\text{app}}(A)$. ■

2. This theorem has an important corollary:

Theorem. The spectrum of every self-adjoint operator is real.

Proof. Let λ be any complex number and write $\lambda = \mu + i\nu$, where μ and ν are real. If A is self-adjoint, then for every vector x

$$\begin{aligned} \|(A - \lambda)x\|^2 &= \langle (A - \lambda)x, (A - \lambda)x \rangle \\ &= \langle (A - \bar{\lambda})(A - \lambda)x, x \rangle \end{aligned}$$

$$\begin{aligned}
&= \|(A - \mu)x\|^2 + \nu^2\|x\|^2 \\
&\geq \nu^2\|x\|^2.
\end{aligned}$$

So if $\nu \neq 0$, then $A - \lambda$ is bounded below. This means λ is not an approximate eigenvalue of A . Thus only real numbers can enter $\sigma(A)$. ■

Exercise. Find a simpler proof for the more special statement that every *eigenvalue* of a self-adjoint operator is real.

Diagonal Operators

3. Let $\mathcal{H}$ be a separable Hilbert space and let $\{e_n\}$ be an orthonormal basis. Let $\alpha = (\alpha_1, \alpha_2, \dots)$ be a bounded sequence of complex numbers. Let $A_\alpha e_n = \alpha_n e_n$. This gives a linear operator on $\mathcal{H}$ if we do the obvious: let $A_\alpha(\sum \xi_n e_n) = \sum \xi_n \alpha_n e_n$. It is easy to see that A_α is bounded and

$$\|A_\alpha\| = \sup |\alpha_n| = \|\alpha\|_\infty. \quad (18.1)$$

We say A_α is the *diagonal operator* on $\mathcal{H}$ induced by the sequence α . We think of it as the operator corresponding to the infinite diagonal matrix

$$A_\alpha = \begin{bmatrix} \alpha_1 & & \\ & \alpha_2 & \\ & & \ddots \end{bmatrix}$$

4. Let α, β be any two elements of ℓ_∞ . It is easy to see that

$$\begin{aligned}
A_\alpha + A_\beta &= A_{\alpha+\beta}, \\
A_\alpha A_\beta &= A_{\alpha\beta}, \\
A_\alpha^* &= A_{\bar{\alpha}}.
\end{aligned}$$

Thus the map $\alpha \mapsto A_\alpha$ is a $\star$ -algebra homomorphism of ℓ_∞ into $\mathcal{B}(\mathcal{H})$. The relation (18.1) shows that this map is an isometry. Note that the family $\{A_\alpha : \alpha \in \ell_\infty\}$ consists of mutually commuting normal operators.

The sequence $\mathbf{1} = (1, 1, \dots)$ is the identity element for the algebra ℓ_∞ . An element α is invertible in ℓ_∞ if there exists β in ℓ_∞ such that $\alpha\beta = \mathbf{1}$. This happens if and only if $\{\alpha_n\}$ is bounded away from zero; i.e., $\inf |\alpha_n| > 0$. The diagonal operator A_α is invertible (with inverse A_β) if and only if α is invertible (with inverse β).

5. Proposition. The spectrum of A_α contains all α_n as eigenvalues, and all limit points of $\{\alpha_n\}$ as approximate eigenvalues.

Proof. It is obvious that each α_n is an eigenvalue of A_α , and easy to see that there are no other eigenvalues. Let λ be any complex number different from all α_n . The operator $A_\alpha - \lambda$ is not invertible if and only if the sequence $\{\alpha_n - \lambda\}$ is not bounded away from zero. This is equivalent to saying that a subsequence α_n converges to λ ; i.e., λ is a limit point of the set $\{\alpha_n\}$.

■

Multiplication Operators

6. Let $(X, \mathcal{S}, \mu)$ be a σ -finite measure space. For each $\varphi \in L_\infty(\mu)$ let M_φ be the linear operator on the Hilbert space $\mathcal{H} = L_2(\mu)$ defined as $M_\varphi f = \varphi f$ for all $f \in \mathcal{H}$. We have then

$$\begin{aligned} \|M_\varphi\| &= \|\varphi\|_\infty, \\ M_\varphi + M_\psi &= M_{\varphi+\psi}, \\ M_\varphi M_\psi &= M_{\varphi\psi}, \\ M_\varphi^\star &= M_{\bar{\varphi}}. \end{aligned}$$

The operator M_φ is called the *multiplication operator* on $L_2(\mu)$ induced by φ . It is a normal operator.

The map $\varphi \mapsto M_\varphi$ is an isometric $\star$ -homomorphism of the algebra L_∞ into $\mathcal{B}(\mathcal{H})$.

A diagonal operator is a multiplication operator: the space $X = \mathbb{N}$ in this case.

7. The function $\mathbf{1}$ that is equal to 1 almost everywhere is an identity for the algebra L_∞ . An element φ of L_∞ is invertible if and only if there exists $\psi \in L_\infty$ such that $\varphi\psi = \mathbf{1}$ a.e. This happens if and only if φ is bounded away from zero; i.e., there exists $\delta > 0$ such that $|\varphi(x)| \geq \delta$ a.e. The multiplication operator M_φ is invertible (with inverse M_ψ) if and only if φ is invertible (with inverse ψ).

8. Let φ be a complex measurable function on $(X, \mathcal{S}, \mu)$. The *thick range* of φ , written as $\text{tran } \varphi$, is the collection of all $\lambda \in \mathbb{C}$ for which

$$\mu\left(\varphi^{-1}(\{\lambda\})\right) > 0.$$

Thus $\lambda \in \text{tran } \varphi$ if φ assumes the value λ on a set of positive measure in X . The *essential range* of φ , written as $\text{ess ran } \varphi$, is the collection of all $\lambda \in \mathbb{C}$ such that for every neighbourhood E of λ

$$\mu\left(\varphi^{-1}(E)\right) > 0.$$

Clearly $\text{tran } \varphi \subset \text{ess ran } \varphi$. Let $\varphi(n) = 1/n$ for every $n \in \mathbb{N}$. Then the range of φ and its thick range are the set $\{1/n : n \in \mathbb{N}\}$. The essential range is the union of this set and $\{0\}$. Let $\varphi(t) = t$ for each t in $[0, 1]$. Then the range of φ and its essential range are equal to $[0, 1]$, while the thick range is empty.

9. Proposition. Let M_φ be the multiplication operator on $L_2(\mu)$ induced by the function $\varphi \in L_2(\mu)$. Then

$$\sigma(M_\varphi) = \text{ess ran } \varphi, \quad \sigma_p(M_\varphi) = \text{tran } \varphi.$$

Proof. The operator $M_\varphi - \lambda$ is not invertible if and only if the function $\varphi - \lambda \mathbf{1}$ is not invertible. This is so if and only if

$$\mu(\{x : |\varphi(x) - \lambda| < \delta\}) > 0$$

for every $\delta > 0$. This is the same as saying $\lambda \in \text{ess ran } \varphi$. This proves the first assertion. Let $\lambda \in \sigma_p(M_\varphi)$. Then there exists a nonzero function f such that $(\varphi(x) - \lambda)f(x) = 0$. So $\varphi(x) = \lambda$ for all x where $f(x) \neq 0$. Such x constitute a set of positive measure. So $\lambda \in \text{tran } \varphi$. Conversely, if $\lambda \in \text{tran } \varphi$, then the set $E = \{x : \varphi(x) = \lambda\}$ has nonzero (possibly infinite) measure. Choose a subset F of E that has a finite positive measure. Then the characteristic function χ_F is in $L_2(\mu)$ and is an eigenvector of M_φ for the eigenvalue λ . Thus $\lambda \in \sigma_p(M_\varphi)$. ■

10. One of the highlights of Functional Analysis is the Spectral Theorem. This says that every normal operator A on any Hilbert space $\mathcal{H}$ is unitarily equivalent to a multiplication operator; i.e., there exists a measure space $(X, \mathcal{S}, \mu)$ a unitary operator $U : \mathcal{H} \rightarrow L_2(\mu)$ and a function $\varphi \in L_\infty$ such that $A = U^* M_\varphi U$. If A is Hermitian the function φ is real, and if A is positive φ is positive.

Two sided shifts

11. Let $\ell_2(\mathbb{Z})$ be the space of all doubly infinite sequences $\{x_n\}_{n=-\infty}^\infty$ such that $\sum_{n=-\infty}^\infty |x_n|^2 < \infty$. The standard basis for this space is the collection $\{e_n\}_{n=-\infty}^\infty$ of vectors that have all entries zero except an entry 1 in the n th place. The *right shift* or the *forward shift* on this space is the operator S defined as $Se_n = e_{n+1}$ for all n . Its inverse is the *left shift* or the *backward shift* T defined as $Te_n = e_{n-1}$ for all n . The operators S and T are unitary. To distinguish them from the shift operators on $l_2 = l_2(\mathbb{N})$ these are called two sided shifts.

12. A *weighted shift* is a composition of a shift and a diagonal operator. More

precisely a bounded two-sided sequence α is called a *weight sequence*. The *weighted backward shift* with weight sequence α is the operator T defined as

$$Te_n = \alpha_{n-1}e_{n-1} \quad \text{for all } n.$$

If α is bounded away from zero, then T is invertible, and its inverse is the operator S acting as

$$Se_n = \frac{1}{\alpha_{n+1}}e_{n+1} \quad \text{for all } n.$$

This is a *weighted forward shift*.

13. Exercise. Let T be a weighted backward shift with weight sequence α . Show that

(i) $\|T\| = \|\alpha\|_\infty$.

(ii) $\text{spr } T = \lim_{k \rightarrow \infty} \sup_n |\alpha_{n-1}\alpha_{n-2} \cdots \alpha_{n-k}|^{1/k}$.

[Use the spectral radius formula.]

(iii) If $\inf |\alpha_n| = r > 0$, then $\|T^{-1}\| = 1/r$.

(iv) If α is bounded above by R and below by r , then $\sigma(T)$ is contained in the annulus $\{\lambda : r \leq |\lambda| \leq R\}$.

[See Exercise 21 in Lecture 16.]

Discontinuity of the spectrum

14. Let T be the weighted backward shift on $l_2(\mathbb{Z})$ with weight sequence α in which $\alpha_{-1} = 0$ and $\alpha_n = 1$ for all $n \neq -1$. By Exercise 13 (ii) the spectral radius of T is 1. For each λ with $|\lambda| < 1$, the vector $x_\lambda = \sum_{n=0}^{\infty} \lambda^n e_n$ is an eigenvector of T with eigenvalue λ . So $\sigma(T) = D$, the closed unit disk.

Consider another weighted backward shift T' with weights α' in which $\alpha'_{-1} = 1$ and $\alpha'_n = 0$ for all $n \neq -1$. For every real number ε , let $T_\varepsilon = T + \varepsilon T'$. This is a weighted backward shift with weight sequence $\alpha(\varepsilon)$ in which $\alpha_{-1}(\varepsilon) = \varepsilon$ and $\alpha_n(\varepsilon) = 1$ for all $n \neq -1$. Thus $\text{spr}(T_\varepsilon) = 1$. If $\varepsilon \neq 0$, then T_ε is invertible, and by Exercise 13(ii) the spectral radius of T_ε^{-1} also is 1. This means that $\sigma(T_\varepsilon)$ is contained in the boundary of the disk D .

This example shows something striking. The spectrum of $T = T_0$ is the unit disk D . Adding a *small* operator $\varepsilon T'$ to T makes the spectrum shrink to the boundary of D . (The operator $\varepsilon T'$ has rank 1 and norm ε .)

15. Thus the map $A \mapsto \sigma(A)$ that associates to an operator A its spectrum is a *discontinuous* map. Let us make this statement more precise.

Exercise Let (X, d) be any metric space and let E, F be any compact subsets of X . Let

$$s(E, F) := \sup_{x \in E} \text{dist}(x, F) = \sup_{x \in E} \inf_{y \in F} d(x, y),$$

and

$$h(E, F) := \max(s(E, F), s(F, E)).$$

Show that $h(E, F)$ is a metric on the collection of all compact subsets of X . This is called the *Hausdorff distance* between E and F . It is the smallest number δ such that any point of E is within distance δ of some point of F , and vice versa.

The space $\mathcal{B}(\mathcal{H})$ is a metric space with its usual norm and the collection of compact subsets of $\mathbb{C}$ is a metric space with the Hausdorff distance. The example in Section 14 shows that the map $A \mapsto \sigma(A)$ between these two spaces is discontinuous (when $\mathcal{H}$ is infinite-dimensional).

16. If a map is not continuous, one looks for some weaker regular behaviour it may

display. It turns out that the spectrum can *shrink* drastically with a small change in the operator (as our example above shows) but it can not *expand* in this wild a manner. The appropriate way to describe this is to say that the map $A \mapsto \sigma(A)$ is *upper semicontinuous*. By definition, this means that for every open set G that contains $\sigma(A)$ there exists an $\varepsilon > 0$ such that

$$\|A - B\| < \varepsilon \Rightarrow \sigma(B) \subset G.$$

Exercise. Prove this as follows.

(i) For $\lambda \in G'$, let $\varphi(\lambda) = \|R_\lambda(A)\| = \|(A - \lambda)^{-1}\|$. This function is continuous on G' and it goes to 0 as λ goes to ∞ . So $\varphi(\lambda)$ is bounded on G' by some number K . Let $\varepsilon = 1/K$.

(ii) Let $\|A - B\| < \varepsilon$. If $\lambda \in G'$, then

$$\|(A - \lambda) - (B - \lambda)\| = \|A - B\| < \varepsilon < \frac{1}{\|(A - \lambda)^{-1}\|}.$$

This shows $B - \lambda$ is invertible; i.e., $\lambda \notin \sigma(B)$.

Continuity of the spectrum in special cases

On the set of normal operators the spectrum is continuous.

17. Theorem. Let A, B be normal operators. Then

$$h(\sigma(A), \sigma(B)) \leq \|A - B\|. \quad (18.2)$$

Proof. Let $\varepsilon = \|A - B\|$. It suffices to show $s(\sigma(A), \sigma(B)) \leq \varepsilon$ and then invoke symmetry. For this we have to show that for each $\lambda \in \sigma(A)$ there is a $\mu \in \sigma(B)$ such

that $|\lambda - \mu| \leq \varepsilon$. If we replace A and B by $A - \lambda$ and $B - \lambda$, then neither the left nor the right hand side of (18.2) changes. So we may assume $\lambda = 0$ and then prove that there exists $\mu \in \sigma(B)$ with $|\mu| \leq \varepsilon$. If this is not the case, then B is invertible and $\text{spr}(B^{-1}) < 1/\varepsilon$. Since B is normal $\|B^{-1}(A - B)\| \leq \|B^{-1}\| \|A - B\| < 1$. This implies $I + B^{-1}(A - B)$ is invertible, and hence so is $B(I + B^{-1}(A - B)) = A$. But this is contrary to our assumption that the point $\lambda = 0$ is in $\sigma(A)$. ■

18. When the space $\mathcal{H}$ is finite-dimensional the spectrum is continuous on all of $\mathcal{B}(\mathcal{H})$. More is true in this case. Let $A \mapsto \text{Eig } A$ be the map that assigns to A the (unordered) n -tuple $\{\lambda_1, \dots, \lambda_n\}$ whose elements are eigenvalues of A counted with multiplicities. Then this map is continuous. (The set $\sigma(A)$ gives no information about multiplicities of the eigenvalues.)

19. When $\mathcal{H}$ is infinite-dimensional, the spectral radius is discontinuous on $\mathcal{B}(\mathcal{H})$. Study the example in P. R. Halmos, *A Hilbert Space Problem Book* that shows this.

Square Roots and the Polar Decomposition

One of the most important and useful theorems of linear algebra is the spectral theorem. This says that every normal operator on an n -dimensional Hilbert space $\mathcal{H}$ can be diagonalised by a unitary conjugation: there exists a unitary operator U such that $U^*AU = \Lambda$, where Λ is the diagonal matrix with the eigenvalues of A on its diagonal. Among other things, this allows us to define *functions* of a normal matrix A in a natural way. Let f be any functions on $\mathbb{C}$. If $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ is a diagonal matrix with λ_j as its diagonal entries, define $f(\Lambda)$ to be the diagonal matrix $\text{diag}(f(\lambda_1), \dots, f(\lambda_n))$, and if $A = U\Lambda U^*$, put $f(A) = Uf(\Lambda)U^*$.

If A is a positive operator, then all its eigenvalues are positive. Each of them has a unique positive square root. Thus A has a unique positive square root, written as $A^{1/2}$.

One of the important consequences of this is the *polar decomposition theorem*. This says that every operator A on $\mathcal{H}$ can be written as $A = UP$, where U is unitary and P is positive. The operator P , called the *positive part* of A is the positive square root of the positive operator A^*A .

The spectral theorem for infinite-dimensional Hilbert spaces will be proved later in this course. It says that a normal operator A is unitarily equivalent to a multiplication operator M_φ . If A is positive, then we define $A^{1/2}$ as the operator equivalent to $M_{\varphi^{1/2}}$.

However, the existence of the square root $A^{1/2}$ can be proved by more elementary arguments. Though less transparent, they are useful in other contexts.

1. Let A be a positive operator. Then $(x, y)_A = \langle Ax, y \rangle$ is a symmetric positive sesquilinear form. It is not always a definite form. The Schwarz inequality for such forms (Exercise 32, Lecture 11) tells us

$$|\langle Ax, y \rangle|^2 \leq \langle Ax, x \rangle \langle Ay, y \rangle. \quad (19.1)$$

2. **A convergence theorem.** Let A_n be an increasing sequence of self-adjoint operators that is bounded from above; i.e.,

$$A_1 \leq A_2 \leq \cdots \leq A_n \leq \cdots \leq \alpha I$$

for some real number α . Then A_n is strongly convergent.

Proof. We prove first that A_n is weakly convergent. For each vector x , the sequence $\langle A_n x, x \rangle$ is an increasing sequence of real numbers bounded from above by $\alpha \langle x, x \rangle$. So the limit $f(x) = \lim_{n \rightarrow \infty} \langle A_n x, x \rangle$ exists. Being a limit of quadratic forms, this is again a quadratic form; i.e., there exists a sesquilinear form $B(x, y)$ on $\mathcal{H}$ such that $f(x) = B(x, x)$. (See Exercise 31, Lecture 11). Clearly B is bounded. So, by the result proved in Section 23 of Lecture 11, there exists a self-adjoint operator A such that $f(x) = \langle Ax, x \rangle$. This operator A is the weak limit of A_n . We will show that, in fact, A_n converges strongly to A . There is some simplification, and no loss of generality, if we assume $A_1 \geq 0$. (Add $\|A_1\|I$ to all the A_n .) Then for $n > m$ we have $0 \leq A_n - A_m \leq \alpha I$. This shows that $\|A_n - A_m\| \leq \alpha$. (Recall that $\|A_n - A_m\| = \sup_{\|x\|=1} \langle (A_n - A_m)x, x \rangle$.) Using the Schwarz inequality (19.1) we get for every x

$$\begin{aligned} & \langle (A_n - A_m)x, (A_n - A_m)x \rangle^2 \\ & \leq \langle (A_n - A_m)x, x \rangle \langle (A_n - A_m)^2 x, (A_n - A_m)x \rangle \\ & \leq \langle (A_n - A_m)x, x \rangle \alpha^3 \|x\|^2. \end{aligned}$$

Since A_n is weakly convergent, the inner product in the last line goes to zero as $n, m \rightarrow \infty$. So, the left hand side of this inequality goes to zero. This shows that for every vector x , $\|(A_n - A_m)x\|$ goes to zero as $n, m \rightarrow \infty$. Hence A_n is strongly convergent; and its strong limit is A . ■

We remark here that the proof above can be simplified considerably if we assume that every positive operator has a positive square root: The weak limit A is bigger than all A_n , so $A - A_n$ is positive and hence equal to P_n^2 for some positive P_n . For every x

$$\|P_n x\|^2 = \langle P_n^2 x, x \rangle = \langle (A - A_n)x, x \rangle$$

converges to zero. Thus P_n converges strongly to 0, and hence so does P_n^2 .

Existence of square roots

3. Theorem. Let A be a positive operator. Then there exists a unique positive operator B such that $B^2 = A$.

Proof. We may assume that $A \leq I$. (Divide A by $\|A\|$.) Consider the sequence X_n defined inductively as

$$X_0 = 0, \quad X_{n+1} = \frac{I - A + X_n^2}{2}.$$

Each X_n is a polynomial in $(I - A)$ with positive coefficients. So $X_n \geq 0$. It is easy to see that $X_1 \leq X_2 \leq \dots \leq X_n \leq \dots \leq I$. Hence, by Theorem 2, X_n converges strongly to a positive operator X . So $X_n^2 \xrightarrow{s} X^2$, and we have

$$X = \text{s-lim} \frac{I - A + X_n^2}{2} = \frac{I - A + X^2}{2},$$

where s-lim stands for strong limit. The last equality shows that

$$A = I - 2X + X^2 = (I - X)^2.$$

Let $B = I - X$. Then B is positive and $B^2 = A$. It remains to show that B is the unique positive square root of A . Note that the operator B was obtained as a strong limit of polynomials in A . Now suppose that C is any positive operator such that $C^2 = A$. Then $C^3 = AC = CA$. Thus C commutes with A , and hence with B . Choose any vector x , and let $y = (B - C)x$. Then

$$\begin{aligned}\langle By, y \rangle + \langle Cy, y \rangle &= \langle (B + C)y, y \rangle \\ &= \langle (B + C)(B - C)x, y \rangle \\ &= \langle (B^2 - C^2)x, y \rangle = 0.\end{aligned}$$

Hence $\langle By, y \rangle$ and $\langle Cy, y \rangle$ both are zero. (They are nonnegative quantities.) Thus

$$\begin{aligned}0 &= \langle (B - C)y, y \rangle = \langle (B - C)^2 x, (B - C)x \rangle \\ &= \langle (B - C)^3 x, x \rangle.\end{aligned}$$

Since x is an arbitrary vector, this shows $(B - C)^3 = 0$. But then $B - C$ must be zero. (Why?). Hence $B = C$. ■

Exercise. If T is a self-adjoint operator and $T^m = 0$ for some positive integer m , then $T = 0$. (This answers the question at the end of the preceding proof.).

The Polar Decomposition

Let us recall how this decomposition is derived in the finite-dimensional case, and then see the modifications needed in infinite dimensions. We use the notation $|A|$ for the positive operator $(A^*A)^{1/2}$.

4. Exercise. For any linear operator A on $\mathcal{H}$ let $\text{ran } A$ and $\ker A$ stand for the range and the kernel of A . Show that

$$(i) \quad \ker A^* = (\text{ran } A)^\perp.$$

- (ii) $\ker (A^*A) = \ker A$.
- (iii) If $\mathcal{H}$ is finite-dimensional, then A , A^*A and $|A|$ have the same rank.
- (iv) $(\ker A)^\perp = \overline{\text{ran } A^*}$ (the closure of $\text{ran } A^*$).

5. Theorem. Let A be any operator on a finite-dimensional Hilbert space. Then there exist a unitary operator U and a positive operator P such that $A = UP$. In this decomposition $P = (A^*A)^{1/2}$, and is thus uniquely determined. If A is invertible then U is uniquely determined.

Proof. Let $P = (A^*A)^{1/2} = |A|$. If A is invertible, then so is P . Let $U = AP^{-1}$. Then for all x

$$\begin{aligned}\langle Ux, Ux \rangle &= \langle AP^{-1}x, AP^{-1}x \rangle \\ &= \langle P^{-1}A^*AP^{-1}x, x \rangle = \langle x, x \rangle.\end{aligned}$$

This shows that U is unitary and $A = UP$.

If A is not invertible, then $\text{ran } A$ is a proper subspace of $\mathcal{H}$ and its dimension equals that of the space $\text{ran } P$. Define a linear map $U : \text{ran } P \rightarrow \text{ran } A$ by putting $UPx = Ax$ for every $x \in \mathcal{H}$. Note that

$$\|Px\|^2 = \langle P^2x, x \rangle = \langle A^*Ax, x \rangle = \|Ax\|^2.$$

This shows U is well-defined and is an isometry. We have defined U on a part of $\mathcal{H}$. Extend U to the whole space by choosing it to be an arbitrary isometry from $(\text{ran } P)^\perp$ onto $(\text{ran } A)^\perp$. Such an isometry exists since these two spaces have the same dimension. The equation $A = UP$ remains valid for the extended U . Suppose $A = U_1P_1 = U_2P_2$ are two polar decompositions of A . Then $A^*A = P_1^2 = P_2^2$. But the positive square root of A^*A is unique. So $P_1 = P_2$. This proves the theorem. ■

6. Exercise. Show that every operator A on a finite-dimensional space can be

written as $A = P'U'$ where $P' = |A^*|$, and U' is unitary. Note that $|A^*| = |A|$ if and only if A is normal.

7. **Exercise.** An operator $A = UP$ on a finite-dimensional space is normal if and only if $UP = PU$.

8. **Exercise.** Use the polar decomposition to prove the singular value decomposition: every linear operator A on an n -dimensional space can be written as $A = USV$, where U and V are unitary and S is diagonal with nonnegative diagonal entries $s_1 \geq \cdots \geq s_n$.

9. Let S be the right shift on the space l_2 . Then $S^*S = I$, and hence $|S| = I$. Since S is not unitary we can not have $S = U|S|$ for any unitary operator U . Thus the polar decomposition theorem for infinite-dimensional spaces has to be different from Theorem 5. The difference is small.

Partial isometries

10. An operator W on $\mathcal{H}$ is called *partial isometry* if $\|Wx\| = \|x\|$ for every $x \in (\ker W)^\perp$.

Every isometry is a partial isometry. Every (orthogonal) projection is a partial isometry.

The space $(\ker W)^\perp$ is called the *initial space* of W , and $\text{ran } W$ is called its *final space*. Both these spaces are closed. The map $W : (\ker W)^\perp \rightarrow \text{ran } W$ is an isometry of one Hilbert space onto another.

Exercise. (i) If W is a partial isometry, then so is W^* . The initial space of W^* is

$\text{ran } W$ and its final space is $(\ker W)^\perp$. The operators $P_i = W^*W$ and $P_f = WW^*$ are the projection operators on the initial and the final spaces of W , respectively.

11. **Exercise.** Let W be any linear operator on $\mathcal{H}$. Then the following conditions are equivalent :

- (i) W is a partial isometry.
- (ii) W^* is a partial isometry.
- (iii) W^*W is a projection.
- (iv) WW^* is a projection.
- (v) $WW^*W = W$.
- (vi) $W^*WW^* = W^*$.

(Recall W is an isometry if and only if $W^*W = I$. This condition is *not* equivalent to $WW^* = I$. If $WW^* = I$, then W is called a co-isometry.)

12. **Theorem.** Let A be any operator on $\mathcal{H}$. Then there exists a partial isometry W such that $A = W|A|$. The initial space of W is $(\ker A)^\perp$ and its final space is $\overline{\text{ran } A}$. This decomposition is unique in the following sense: if $A = UP$, where P is positive and U is a partial isometry with $\ker U = \ker P$, then $P = |A|$ and $U = W$.

Proof. Define $W : \text{ran } |A| \rightarrow \text{ran } A$ by putting $W|A|x = Ax$ for all $x \in \mathcal{H}$. It is easy to see that W is an isometry. The space $\text{ran } |A|$ is dense in $(\ker A)^\perp$ (Exercise!) and hence, W extends to an isometry $W : (\ker A)^\perp \rightarrow \overline{\text{ran } A}$. Put $Wx = 0$ for all $x \in \ker A$. This gives a partial isometry on $\mathcal{H}$, and $A = W|A|$. To prove uniqueness note that $A^*A = PU^*UP = PEP$, where E is the projection onto the initial space of E . This space is $(\ker U)^\perp = (\ker P)^\perp = \overline{\text{ran } A}$. So $A^*A = P^2$, and hence $P = |A|$,

the unique positive square root of A^*A . This shows $A = W|A| = U|A|$. So W and U are equal on $\text{ran } |A|$ and hence on $(\ker A)^\perp$, their common initial space. ■

13. **Exercise.** Let $A = W|A|$ be the polar decomposition of A . Show that

- (i) $W^*A = |A|$.
- (ii) W is an isometry if and only if A is one-to-one.
- (iii) W and $|A|$ commute if and only if A commutes with A^*A .

Compact Operators

This is a special class of operators and for several reasons it is good to study them in some detail at this stage. Their spectral theory is much simpler than that of general bounded operators, and it is just a little bit more complicated than that of finite-dimensional operators. Many problems in mathematical physics lead to integral equations, and the associated integral operators are compact. For this reason these operators were among the first to be studied, and in fact, this was the forerunner to the general theory.

1. We say that a subset E of a complete metric space X is *precompact* if its closure $\bar{E}$ is compact. If X is a finite-dimensional normed space, then every bounded set is precompact. The unit ball in an infinite-dimensional space is not precompact.

A set E is precompact if and only if for every $\varepsilon > 0$, E can be covered by a finite number of balls of radius ε .

2. Let X, Y be Banach spaces. A linear operator A from X to Y is called a *compact operator* if it maps the unit ball of X onto a precompact subset of Y . Since A is linear this means that A maps every bounded set in X to a precompact subset of Y .

The sequence criterion for compactness of metric spaces tells us that A is compact if and only if for each bounded sequence $\{x_n\}$ the sequence $\{Ax_n\}$ has a convergent subsequence.

If either X or Y is finite-dimensional, then every $A \in \mathcal{B}(X, Y)$ is compact. The identity operator I on any infinite-dimensional space is not compact.

3. If the range of A is finite-dimensional, we say that A has finite rank. Every finite-rank operator is compact. We write $\mathcal{B}_0(X, Y)$ for the collection of all compact operators from X to Y and $\mathcal{B}_{00}(X, Y)$ for all finite-rank operators. Each of them is a vector space.

4. **Example.** Let $X = C[0, 1]$. Let $K(x, y)$ be a continuous kernel on $[0, 1] \times [0, 1]$ and let A be the integral operator induced by it

$$(Af)(x) = \int_0^1 K(x, y)f(y)dy.$$

Then A is a compact operator. To prove this we show that whenever $\{f_n\}$ is a sequence in X with $\|f_n\| \leq 1$ for all n , the sequence $\{Af_n\}$ has a convergent subsequence. For this we use Ascoli's Theorem. Since $\|Af_n\| \leq \|A\|$, the family $\{Af_n\}$ is bounded. We will show that it is equicontinuous. Since K is uniformly continuous, for each $\varepsilon > 0$ there exists $\delta > 0$ such that whenever $|x_1 - x_2| < \delta$ we have $|K(x_1, y) - K(x_2, y)| < \varepsilon$ for all y . This shows that whenever $|x_1 - x_2| < \delta$ we have

$$\begin{aligned} |Af_n(x_1) - Af_n(x_2)| &\leq \int_0^1 |K(x_1, y) - K(x_2, y)| |f_n(y)| dy \\ &\leq \varepsilon \int_0^1 |f_n(y)| dy \leq \varepsilon. \end{aligned}$$

Thus the family $\{Af_n\}$ is equicontinuous. So by Ascoli's Theorem it has a convergent subsequence. Thus the operator A is compact.

The condition that $K(x, y)$ be continuous in (x, y) is too stringent. If

$$\lim_{x_n \rightarrow x} \int_0^1 |K(x_n, y) - K(x, y)| dy = 0,$$

then the operator A induced by K is a compact operator on $C[0, 1]$.

5. **Theorem.** $\mathcal{B}_0(X, Y)$ is a closed subspace of $\mathcal{B}(X, Y)$.

Proof. Let A_n be a sequence of compact operators converging in norm to a bounded operator A . Given $\varepsilon > 0$ we can find an n such that $\|A_n - A\| < \varepsilon/2$. Let S be the unit ball of X . Since A_n is compact the set $A_n(S)$ in Y can be covered by a finite number of balls of radius $\varepsilon/2$. Keeping the same centres and increasing the radii to ε we get a finite collection of balls that covers $A(S)$. Thus $A(S)$ is a precompact set.

Corollary. If $A \in \mathcal{B}(X, Y)$ and there exists a sequence $A_n \in \mathcal{B}_0(X, Y)$ such that $\|A_n - A\| \rightarrow 0$, then $A \in \mathcal{B}_0(X, Y)$.

6. **Exercise.** Show that a strong limit of finite-rank operators is not always compact.

7. **Exercise.** Let α be a bounded sequence and let A_α be the diagonal operator on l_2 with diagonal α . Show that A_α is compact if and only if α_n converges to zero.

8. **Theorem.** Let A and B be bounded operators. If either A or B is compact, then the product AB is compact.

Proof. Let $\{x_n\}$ be a bounded sequence. Then $\{Bx_n\}$ is bounded and if A is compact, then $\{ABx_n\}$ has a convergent subsequence. If B is compact, then $\{Bx_n\}$ has a convergent subsequence. The image of this subsequence under A is convergent. ■

Exercise. Let A and B be bounded operators. If AB is compact, then is it necessary that either A or B is compact?

Exercise. If $A^2 = 0$, then is it necessary that A is a compact operator?

We have seen that the space $\mathcal{B}_0(X)$ is a vector space of $\mathcal{B}(X)$. Theorem 8 says

that $\mathcal{B}_0(X)$ is a *two-sided ideal* in the algebra $\mathcal{B}(X)$. By Theorem 5 this ideal is closed.

9. Complete Continuity. A linear operator from X into Y is bounded if and only if it is continuous. This can be expressed in another way: A is bounded if and only if it maps every convergent sequence in X to a convergent sequence in Y . The convergence we are talking of is the usual (strong) convergence in the respective norms of X and Y . To emphasize this let us say A is bounded (continuous) iff

$$x_n \xrightarrow{s} x \Rightarrow Ax_n \xrightarrow{s} Ax. \quad (20.1)$$

Now suppose $\{x_n\}$ is a sequence in X converging weakly to x . Then for every $g \in Y^*$

$$\langle g, A(x_n - x) \rangle = \langle A^*g, x_n - x \rangle \rightarrow 0;$$

i.e. Ax_n converges weakly to Ax . Thus for every bounded operator A

$$x_n \xrightarrow{w} x \Rightarrow Ax_n \xrightarrow{w} Ax. \quad (20.2)$$

We say that A is *completely continuous* if it satisfies the stronger requirement

$$x_n \xrightarrow{w} x \Rightarrow Ax_n \xrightarrow{s} Ax. \quad (20.3)$$

10. Theorem. Every compact operator A is completely continuous.

Proof. Let $x_n \xrightarrow{w} x$. Then the sequence $\{\|x_n\|\}$ is bounded. (Lecture 9, Section 8.) If Ax_n does not converge strongly to Ax , then there exists an $\varepsilon > 0$ and a subsequence $\{x_m\}$ such that $\|Ax_m - Ax\| \geq \varepsilon$ for all m . Since $\{x_m\}$ is bounded and A compact, $\{Ax_m\}$ has a convergent subsequence. Suppose y is the limit of this sequences. Then y is also its weak limit. But by (20.2) we must have $y = Ax$. This leads to a contradiction. ■

Exercise. Let A be a compact operator on $\mathcal{H}$ and let $\{e_n\}$ be an orthonormal basis. Then the sequence $\{Ae_n\}$ converges to 0.

11. **Theorem.** If A is a completely continuous operator on a Hilbert space $\mathcal{H}$, then A is compact.

Proof. Let $\{x_n\}$ be any sequence in $\mathcal{H}$ with $\|x_n\| \leq 1$. If we show $\{x_n\}$ has a weakly convergent subsequence $\{x_m\}$, the complete continuity of A would imply that Ax_m is (strongly) convergent and hence A is compact. In a compact metric space every sequence has a convergent subsequence. So, if the unit ball $\{x : \|x\| \leq 1\}$ in $\mathcal{H}$ with the *weak* topology were a compact metric space, then $\{x_n\}$ would surely have a convergent subsequence. In Section 13 of Lecture 12 we constructed exactly such a metric. ■

12. It can be shown, more generally, that if X is a reflexive Banach space then every completely continuous operator on it is compact. In some books the terms “compact operator” and “completely continuous operator” are used interchangeably.

Warning. The condition (20.3) is phrased in terms of *sequences*. These are enough to capture everything about the strong topology but not about the weak topology. If X is given its weak topology and Y its strong topology, then a map $A : X \rightarrow Y$ is continuous if for every *net* x_α converging weakly to x , the net Ax_α converges strongly to Ax . It can be shown that the only such linear operators are finite-rank operators.

13. **Theorem.** If A is compact, then its adjoint A^* is also compact.

Proof. Let $A \in \mathcal{B}_0(X, Y)$. Let $\{g_n\}$ be a sequence in Y^* with $\|g_n\| \leq 1$. We have to show that the sequence $\{A^*g_n\}$ in X^* has a convergent subsequence. Let S be the unit ball in X . Then $\overline{A(S)}$ the closure of $A(S)$ is a compact metric space. Regard

g_n as elements of the space $C(\overline{A(S)})$. Note that for all n

$$\sup_{y \in A(S)} |g_n(y)| = \sup_{y \in A(S)} |g_n(y)| \leq \sup_{y \in A(S)} \|g_n\| \|y\| \leq \|A\|.$$

Thus the family $\{g_n\}$ is uniformly bounded in $C(\overline{A(S)})$. Next note that for all $y_1, y_2 \in Y$

$$|g_n(y_1) - g_n(y_2)| \leq \|y_1 - y_2\|.$$

So $\{g_n\}$ is an equicontinuous family. Hence by Ascoli's Theorem a subsequence $\{g_m\}$ converges to a limit g in the space $C(\overline{A(S)})$. This convergence means that the sequence $g_m(Ax)$ converges to $g(Ax)$ uniformly for x in S . That is the same thing as saying that the sequence A^*g_m converges to A^*g in X^* . ■

Exercise. Show that if $A \in \mathcal{B}(X, Y)$ and A^* is compact, then A is compact.

14. For Hilbert space operators Theorem 13 can be proved easily using the polar decomposition.

When $\mathcal{H}$ is a Hilbert space $\mathcal{B}_0(\mathcal{H})$ is a closed, two-sided, $\star$ -closed ideal in $\mathcal{B}(\mathcal{H})$. It can be proved (using the spectral theorem) that this is the only ideal in $\mathcal{B}(\mathcal{H})$ with this property.

15. **Theorem.** Let $\mathcal{H}$ be a separable Hilbert space. Then $\mathcal{B}_{00}(\mathcal{H})$ is dense in $\mathcal{B}_0(\mathcal{H})$. In other words, every compact operator on $\mathcal{H}$ is a norm limit of finite rank operators.

Proof. Let $\{e_n\}$ be an orthonormal basis for $\mathcal{H}$. Let $\mathcal{H}_n$ be the subspace spanned by the vectors $e_1, \dots, e_n$. Let P_n be the orthogonal projection onto $\mathcal{H}_n$. Then AP_n is a finite-rank operator and

$$\|A - AP_n\| = \|A(I - P_n)\| =: \alpha_n, \text{ say.}$$

Note that α_n is a decreasing sequence of nonnegative numbers. So, α_n converges to a limit $\alpha \geq 0$. By the definition of α_n , there exists a unit vector x_n in $\mathcal{H}_n^\perp$ such that

$\|Ax_n\| \geq \alpha_n/2$. Since the sequence $\mathcal{H}_n$ increases to $\mathcal{H}$, the sequence x_n converges weakly to 0. So, if A is compact $Ax_n \rightarrow 0$. Hence $\alpha_n \rightarrow 0$. Thus A is the norm limit of the sequence AP_n . ■

16. Is the assertion of Theorem 15 valid for all separable Banach spaces? This question turns out to be difficult. In 1973, P. Enflo answered it in the negative. There exists a separable Banach space on which some compact operator is not a norm limit of finite rank operators. Our proof of Theorem 15 suggests that if X has a Schauder basis, then $\mathcal{B}_{00}(X)$ is dense in $\mathcal{B}_0(X)$. This is indeed the case. So the space X in Enflo's example does not have a Schauder basis.

The Spectrum of a Compact Operator

Most of the spectral properties of a compact operator in a Banach space were discovered by F. Riesz, and appeared in a paper in 1918 (several years *before* Banach's book). These results were augmented and simplified by the work of Schauder. What follows is an exposition of these ideas.

Unless stated otherwise, X and Y will stand for *infinite-dimensional* Banach spaces.

1. Recall Riesz's lemma from Lecture 2. If M is a proper closed subspace of X , then for each $\varepsilon > 0$, there exists a unit vector x in X such that $\text{dist}(x, M) \geq 1 - \varepsilon$. If M is finite-dimensional then there exists a unit vector x such that $\text{dist}(x, M) = 1$.

2. **Theorem.** Let A be a compact operator from X to Y . Then the range of A is separable. Further, if $\text{ran } A$ is closed, then it is finite-dimensional.

Proof. For each n , let $S_n = \{x \in X : \|x\| < n\}$. Then $A(S_n)$ is precompact. Every compact metric space is separable. So $A(S_n)$ is separable. Hence so is the countable union

$$\bigcup_{n=1}^{\infty} A(S_n) = \text{ran } A.$$

The Open Mapping Theorem tells us that if $\text{ran } A$ is closed, then A is an open map. So $A(S_n)$ is an open precompact set in $\text{ran } A$. Every point in $\text{ran } A$ belongs to some

$A(S_n)$. So $\text{ran } A$ is locally compact. Hence it is finite-dimensional. ■

3. **Example.** The diagonal operator on l_2 with diagonal $1, 1/2, 1/3, \dots$ is compact and its range is not closed. (Lecture 6, Remark 3.)

4. **Corollary.** Let $A \in \mathcal{B}_0(X)$ and let λ be a nonzero complex number. Then the space $\ker(A - \lambda)$ is finite-dimensional.

Proof. For each linear operator A and complex number λ , the space $N = \ker(A - \lambda)$ is closed. It is easy to see that if $\lambda \neq 0$, then A maps N onto itself. So if A is compact, then by Theorem 2, N is finite-dimensional. ■

5. If A is a compact operator on X ($\dim X = \infty$), then A cannot be invertible. So the point 0 is always in $\sigma(A)$. It is a special point in the spectrum, as we will see.

6. **Proposition.** Let $A \in \mathcal{B}_0(X)$. Then the point spectrum $\sigma_p(A)$ is countable and has only one possible limit point 0.

Proof. We need to prove that for each $\varepsilon > 0$ the set

$$\sigma_p(A) \cap \{\lambda : |\lambda| \geq \varepsilon\}$$

is finite. If this is not the case, then there exists an ε , an infinite set $\{\lambda_n\}$ with $|\lambda_n| \geq \varepsilon$ and vectors x_n such that $\|x_n\| = 1$ and $Ax_n = \lambda_n x_n$. The vectors x_n , being eigenvectors corresponding to distinct eigenvalues of A , are linearly independent. So for each n , the space M_n spanned by $\{x_1, \dots, x_n\}$ is an n -dimensional space. By Riesz's Lemma, for each $n > 1$, there exists $y_n \in M_n$ such that $\|y_n\| = 1$ and $\text{dist}(y_n, M_{n-1}) = 1$. Since $y_n \in M_n$ we can write

$$\begin{aligned} y_n &= \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n, \\ Ay_n &= \alpha_1 \lambda_1 x_1 + \alpha_2 \lambda_2 x_2 + \cdots + \alpha_n \lambda_n x_n. \end{aligned}$$

This shows that $Ay_n - \lambda_n y_n$ is in M_{n-1} . For $n > m$ the vector $Ay_n - Ay_m$ has the form $\lambda_n y_n - z$ where $z \in M_{n-1}$. Since $\text{dist}(y_n, M_{n-1}) = 1$, this shows that

$$\|Ay_n - Ay_m\| \geq |\lambda_n| \geq \varepsilon.$$

But then no subsequence of $\{Ay_n\}$ can converge and A cannot be compact. ■

7. Proposition. Let $A \in \mathcal{B}_0(X)$. If $\lambda \neq 0$ and $\lambda \in \sigma(A)$, then $\lambda \in \sigma_p(A)$.

Proof. Let $\lambda \neq 0$ and suppose that λ is an approximate eigenvalue of A . Then there exists a sequence x_n of unit vectors such that $(A - \lambda)x_n \rightarrow 0$. Since A is compact, a subsequence $\{Ax_m\}$ of $\{Ax_n\}$ converges to some limit y . Hence $\{\lambda x_m\}$ converges to y . Since $\lambda \neq 0$, y is not the zero vector. Note that $Ay = \lambda y$. So $\lambda \in \sigma_p(A)$. We have shown that every nonzero point of the approximate point spectrum $\sigma_{\text{app}}(A)$ is in $\sigma_p(A)$. Hence by Proposition 6 the set $\sigma_{\text{app}}(A)$ is countable. This set contains the boundary of $\sigma(A)$ (Lecture 17, Theorem 11.). Thus $\sigma(A)$ is a compact subset of the complex plane with a countable boundary. Hence $\sigma(A)$ is equal to its boundary. (Exercise). This shows that $\sigma(A) = \sigma_{\text{app}}(A)$. Every nonzero point of this set is in $\sigma_p(A)$. ■

8. Let λ be an eigenvalue of any operator A . The dimension of the space $\ker(A - \lambda)$ is called the *multiplicity* of the eigenvalue λ .

The results of Sections 4-8 together can be summarised as the following.

9. Theorem. (Riesz) Let A be a compact operator. Then

- (i) $\sigma(A)$ is a countable set containing 0.
- (ii) No point other than 0 can be a limit point of $\sigma(A)$.
- (iii) Each nonzero point of $\sigma(A)$ is an eigenvalue of A and has finite multiplicity.

10. The behaviour of 0

If A is compact, then $\sigma(A) = \sigma_{\text{app}}(A)$ and $0 \in \sigma(A)$. The following examples show that the point 0 can act in different ways. In all these examples the underlying space X is l_2 .

- (i) Let A be a projection onto a k -dimensional subspace. Then 0 is an eigenvalue of infinite multiplicity. The only other point in $\sigma(A)$ is 1, and this is an eigenvalue with multiplicity k .
- (ii) Let A be the diagonal operator with diagonal entries $1, 0, 1/2, 0, 1/3, 0, \dots$. Then 0 is an eigenvalue of A with infinite multiplicity. Each point $1/n$ is an eigenvalue of A with multiplicity one.
- (iii) Let $A = D$ the diagonal operator with diagonal entries $1, 1/2, 1/3, \dots$. Then 0 is not an eigenvalue. The points $1/n$ are eigenvalues of A and 0 is their limit point.
- (iv) Let T be the left shift operator and $A = DT$; i.e.,

$$Ax = (x_2, \frac{x_3}{2}, \frac{x_4}{3}, \dots).$$

If $Ax = \lambda x$, then

$$x_n = (n-1)! \lambda^{n-1} x_1 \quad \text{for all } n.$$

If $\lambda \neq 0$ such an x can be in l_2 only if $x = 0$. So λ cannot be an eigenvalue of A . A vector x is mapped to 0 by A if and only if x is a scalar multiple of e_1 . So 0 is an eigenvalue of A with multiplicity one, and is the only point in $\sigma(A)$.

- (v) Let S be the right shift operator and $A = SD$; i.e.,

$$Ax = \left(0, x_1, \frac{x_2}{2}, \dots\right).$$

It is easy to see that A has no eigenvalue. So in this case 0 is the only point in $\sigma(A)$, and is not an eigenvalue. Note that the operators in (iii) and (iv) are

adjoints of each other. If we represent these two operators by infinite matrices, then

$$DT = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 1/2 & 0 & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 1/3 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix},$$

and SD is the transpose of this matrix. The first matrix has entries $(1, 1/2, 1/3, \dots)$ on its first superdiagonal, and the second on its first subdiagonal. If we take the top left $n \times n$ block of either of these matrices, it has zero as an eigenvalue of multiplicity n . One may naively expect that DT and SD have 0 as an eigenvalue with infinite multiplicity. This fails, in different ways, in both the cases.

11. Theorem. Let A be a compact operator on X and λ any nonzero complex number. Then $\text{ran}(A - \lambda)$ is closed.

Proof. By Corollary 4, the space $\ker(A - \lambda)$ is finite-dimensional. Hence it is a direct summand; i.e., there exists a closed subspace W such that

$$X = \ker(A - \lambda) \oplus W.$$

(See Lecture 11, Section 19.) Note that

$$\text{ran}(A - \lambda) = (A - \lambda)X = (A - \lambda)W.$$

If $A - \lambda$ were not bounded below on W , then λ would be an approximate eigenvalue, and hence an eigenvalue of A . This is not possible as $\ker(A - \lambda) \cap W = \{0\}$. So $A - \lambda$ is bounded below on W ; i.e., there exists $\alpha > 0$ such that $\|(A - \lambda)w\| \geq \alpha\|w\|$ for all $w \in W$. Let w_n be any sequence in W , and suppose $(A - \lambda)w_n$ converges to y .

For all n and m

$$\|(A - \lambda)(w_n - w_m)\| \geq \alpha \|w_n - w_m\|,$$

and hence w_n is a Cauchy sequence. Since W is closed w_n converges to a limit $w \in W$. Hence $y = (A - \lambda)w$ is in $(A - \lambda)W$. This shows that $\text{ran}(A - \lambda)$ is closed. ■

12. We know that A is compact if and only if A^* is compact. We know also that $\sigma(A) = \sigma(A^*)$. In Section 10 we have seen an example where 0 is an eigenvalue of A but not of A^* . The nonzero points in the set $\sigma(A) = \sigma(A^*)$ can only be eigenvalues of finite multiplicity for either operator. More is true: each point $\lambda \neq 0$ has the same multiplicity as an eigenvalue for A as it has for A^* .

Theorem. Let $A \in \mathcal{B}_0(X)$ and let $\lambda \neq 0$. Then

$$\dim \ker (A^* - \lambda) = \dim \ker (A - \lambda). \quad (21.1)$$

Proof. Let m^* and m be the numbers on the left and the right hand sides of (21.1). We show first that $m^* \leq m$. Let $x_1, \dots, x_m$ be a basis for the space $\ker (A - \lambda)$. Choose linear functionals $f_1, \dots, f_m$ on X such that $f_i(x_j) = \delta_{ij}$. (Use the H.B.T.) If $m^* > m$, there exist $m + 1$ linearly independent elements $g_1, \dots, g_{m+1}$ in the space $\ker (A^* - \lambda) \subset X^*$. Choose $y_1, \dots, y_{m+1}$ in X such that $g_i(y_j) = \delta_{ij}$. (See Exercise 19 in Lecture 10.) For each $x \in X$ let $Bx = \sum_{i=1}^m f_i(x)y_i$. This is a linear operator of finite rank, and hence is compact. Note that

$$\langle Bx, g_j \rangle = \begin{cases} f_j(x) & \text{if } 1 \leq j \leq m \\ 0 & \text{if } j = m + 1. \end{cases}$$

Since $g_j \in \ker (A^* - \lambda)$,

$$\langle (A - \lambda)x, g_j \rangle = \langle x, (A^* - \lambda)g_j \rangle = 0 \quad \text{for all } j.$$

Adding these two equations we get, for all $x \in X$,

$$\langle (A + B - \lambda)x, g_j \rangle = \begin{cases} f_j(x) & \text{if } 1 \leq j \leq m \\ 0 & \text{if } j = m + 1. \end{cases} \quad (21.2)$$

Thus g_{m+1} annihilates $\text{ran } (A + B - \lambda)$. Since $g_{m+1}(y_{m+1}) = 1$, this shows $\text{ran } (A + B - \lambda) \neq X$.

Hence $\lambda \in \sigma(A + B)$ and since $A + B$ is compact λ has to be an eigenvalue. This is possible only if there exists a nonzero vector x such that $(A + B - \lambda)x = 0$. If x is such a vector, then from (21.2) $f_j(x) = 0$ for all $1 \leq j \leq m$, and hence by the definition of B we have $Bx = 0$. So $x \in \ker(A - \lambda)$. The vectors x_j are a basis for this space, and hence

$$x = \alpha_1 x_1 + \cdots + \alpha_m x_m.$$

Using the relations $f_j(x_i) = \delta_{ij}$ we get from this $\alpha_j = f_j(x) = 0$ for all $1 \leq j \leq m$. But then $x = 0$. This is a contradiction. Hence we must have $m^* \leq m$. Applying the same argument to A^* in place of A , we see that m^{**} , the dimension of $\ker(A^{**} - \lambda)$ is bounded as $m^{**} \leq m^* \leq m$. On the other hand, if J is the canonical embedding of X in X^{**} , then $JA = A^{**}J$. Hence $\ker(A - \lambda I) \subset \ker(A^{**} - \lambda I)$ and $m \leq m^{**}$. Thus $m = m^*$. ■

13. Corollary. Let $A \in \mathcal{B}_0(X)$ and $\lambda \neq 0$. Then

$$\dim \ker(A - \lambda) = \text{codim } \text{ran } (A - \lambda).$$

Proof. From the relation defining adjoints

$$\langle (A - \lambda)x, y \rangle = \langle x, (A^* - \lambda)y \rangle$$

we see that $\ker(A^* - \lambda) = [\text{ran } (A - \lambda)]^\perp$. Since $\text{ran } (A - \lambda)$ is closed (Theorem 11), $\dim [\text{ran } (A - \lambda)]^\perp = \text{codim } \text{ran } (A - \lambda)$ by Theorem 16 of Lecture 10. ■

14. Fredholm Operators. Let $A \in \mathcal{B}(X, Y)$. The quotient space $Y/\text{ran } A$ is called the *cokernel* of A , and written as $\text{coker } A$. If either $\ker A$ or $\text{coker } A$ has finite

dimension, we define the *index* of A as the extended integer.

$$\text{ind } A = \dim \ker A - \dim \text{coker } A.$$

If $\ker A$ and $\text{coker } A$ both are finite-dimensional, we say that A is a *Fredholm operator*.

The index of such an operator is a finite integer.

We have shown that if A is a compact operator on X and λ a nonzero complex number, then $A - \lambda$ is a Fredholm operator and its index is zero.

15. The Fredholm Alternative. From Theorems 9 and 12 we can extract the following statement, a special case of which for certain integral equations was obtained by Fredholm.

Let A be a compact operator on X . Then exactly one of the following alternatives is true

- (i) For every $y \in X$, there is a unique $x \in X$ such that $Ax - x = y$.
- (ii) There exists a nonzero x such that $Ax - x = 0$.

If the alternative (ii) is true, then the homogeneous equation $Ax - x = 0$ has only a finite number of linearly independent solutions.

The homogeneous equation $Ax - x = 0$ has a nonzero solution in X if and only if the transposed equation $A^*y - y = 0$ has a nonzero solution in X^* . The number of linearly independent solutions of these two equations is the same.

Compact Operators and Invariant Subspaces

Continuing the analysis of the previous lecture we obtain more information about compact operators.

1. Let $A \in \mathcal{B}_0(X)$ and let $\lambda \neq 0$. For brevity let us write N_j for the closed linear space $\ker (A - \lambda)^j$, $j = 0, 1, 2, \dots$. We have a nested chain of subspaces

$$N_0 \subset N_1 \subset N_2 \subset \dots \subset N_j \subset \dots \subset X. \quad (22.1)$$

Note that $(A - \lambda) N_{j+1} \subset N_j$ for all j . Suppose for some p , $N_p = N_{p+1}$, then $N_p = N_{p+m}$ for all m . This is an easy exercise. Using Riesz's Lemma one can see that the chain (22.1) is finite; i.e. there exists p such that

$$N_{p+m} = N_p \quad \text{for all } m. \quad (22.2)$$

If this is not the case, then there exists a sequence y_j of unit vectors such that $y_j \in N_j$ and $\text{dist}(y_j, N_{j-1}) > 1/2$. For $n > m$

$$Ay_n - Ay_m = \lambda y_n + (A - \lambda)y_n - (A - \lambda)y_m - \lambda y_m$$

The last three terms in this sum are elements of N_{n-1} . So

$$\|Ay_n - Ay_m\| \geq |\lambda|/2.$$

Thus the sequence $\{Ay_n\}$ has no Cauchy subsequence. Since $\|y_j\| = 1$ and A is compact, this is a contradiction. Therefore, the condition (22.2) must hold.

2. **Exercise.** Let A and λ be as above. Let R_j be the closed linear space $\text{ran}(A - \lambda)^j$. We have a decreasing chain of subspaces

$$R_0 \supset R_1 \supset R_2 \supset \cdots \supset R_j \supset \cdots \supset \{0\}. \quad (22.3)$$

Note that $(A - \lambda)R_j = R_{j+1}$.

Show that there exists q such that

$$R_{q+m} = R_q \quad \text{for all } m. \quad (22.4)$$

3. **The Riesz Decomposition Theorem.** Let A be a compact operator on X and let $\lambda \neq 0$. Then there exists a positive integer n such that $\ker(A - \lambda)^{n+1} = \ker(A - \lambda)^n$ and $\text{ran}(A - \lambda)^{n+1} = \text{ran}(A - \lambda)^n$. We have

$$X = \ker(A - \lambda)^n \oplus \text{ran}(A - \lambda)^n, \quad (22.5)$$

and each of the spaces in this decomposition is invariant under A .

Proof. Choose indices p and q , not both zero, satisfying (22.2) and (22.4). Let $n = \max(p, q)$. Let $y \in \ker(A - \lambda)^n \cap \text{ran}(A - \lambda)^n$. Then there exists x such that $y = (A - \lambda)^n x$, and $(A - \lambda)^n y = 0$. But then $(A - \lambda)^{2n} x = 0$; i.e., $x \in \ker(A - \lambda)^{2n}$. Since $\ker(A - \lambda)^{2n} = \ker(A - \lambda)^n$ this means $y = 0$. Thus the two subspaces on the right hand side of (22.5) have zero intersection. Let x be any element of X . Then $(A - \lambda)^n x$ is in $\text{ran}(A - \lambda)^n = \text{ran}(A - \lambda)^{2n}$. So there exists a vector y such that $(A - \lambda)^n x = (A - \lambda)^{2n} y$. We have

$$x = (x - (A - \lambda)^n y) + (A - \lambda)^n y.$$

The first summand in this sum is in $\ker(A - \lambda)^n$ and the second is in $\text{ran}(A - \lambda)^n$. This proves (22.5). It is clear that each of the spaces is invariant under A . ■

4. **Corollary.** Let A be a compact operator and suppose a nonzero number λ is an

eigenvalue of A . Let n be an integer as in the Theorem above. Let

$$\begin{aligned} N_\lambda &= \ker (A - \lambda)^n, \\ R_\lambda &= \operatorname{ran} (A - \lambda)^n. \end{aligned}$$

Then the restriction of A to N_λ has a single point spectrum $\{\lambda\}$ and the restriction of A to R_λ has spectrum $\sigma(A) \setminus \{\lambda\}$.

Proof. The space N_λ is finite-dimensional and is invariant under A . The restriction of $A - \lambda$ to this space is nilpotent. So $\sigma(A - \lambda|_{N_\lambda}) = \{0\}$. Hence $\sigma(A|_{N_\lambda}) = \{\lambda\}$. The spectrum of the direct sum of two operators is the union of their spectra. The point λ can not be in $\sigma(A|_{R_\lambda})$ as $Ax = \lambda x$ only if $x \in N_\lambda$. ■

Note that the space N_λ is the linear span of the spaces $\ker (A - \lambda)^j$, $j = 1, 2, \dots$. Likewise R_λ is the intersection of the spaces $\operatorname{ran} (A - \lambda)^j$, $j = 1, 2, \dots$. So, the integer n plays no essential role in the statement of this corollary.

5. The Riesz Projection. In the decomposition

$$X = N_\lambda \oplus R_\lambda$$

obtained above, let P_λ be the projection on N_λ along R_λ . This is called the *Riesz projection* of A corresponding to the eigenvalue λ . Since λ is an isolated point of $\sigma(A)$ we can find a closed curve Γ in the plane with winding number 1 around λ and 0 around any other point of $\sigma(A)$. It turns out that P_λ has a representation

$$P_\lambda = \frac{1}{2\pi i} \int_\Gamma (A - \zeta)^{-1} d\zeta.$$

Invariant subspaces

The Riesz decomposition theorem *seems* to give a decomposition of X into a direct sum of generalised eigenspaces of a compact operator A . However, this is not

the case. A may have no nonzero eigenvalue and then the Riesz theory does not even tell us whether A has any nontrivial invariant subspaces. Our next theorem says such a space does exist.

Let $A \in \mathcal{B}(X)$. Let M be a (closed linear) subspace of X and let M be neither the null space $\{0\}$ nor the whole space X . Recall that the space M is said to be *invariant* under A if $A(M) \subset M$. Let $\mathcal{A}$ be the set of all operators T that commute with A . This is called the *commutant* of A and is a sub algebra of $\mathcal{B}(X)$. We say that M is a *hyperinvariant subspace* for A if $T(M) \subset M$ for all $T \in \mathcal{A}$.

6. Lomonosov's Theorem. Every nonzero compact operator has a nontrivial hyperinvariant subspace.

Proof Let $A \in \mathcal{B}_0(X)$, $A \neq 0$, and let $\mathcal{A}$ be the commutant of A . If there exists a nonzero point λ in $\sigma(A)$, then the eigenspace $\ker(A - \lambda)$ is invariant under all $T \in \mathcal{A}$. So, we need to prove the theorem only when $\sigma(A) = \{0\}$. Replacing A by $\frac{1}{\|A\|}A$, we may assume $\|A\| = 1$. Let x_0 be any vector such that $\|Ax_0\| > 1$. Then $\|x_0\| > 1$. Let $D = \{x : \|x - x_0\| < 1\}$ be the open ball of radius 1 centred at x_0 . Since $\|A\| = 1$ and $\|Ax_0\| > 1$, the closure $\overline{A(D)}$ does not contain the vector 0. For each nonzero vector $y \in X$ consider the set $\mathcal{A}y = \{Ty : T \in \mathcal{A}\}$. This is a nonzero linear subspace of X and is invariant under $\mathcal{A}$. If we show that for some y the space $\mathcal{A}y$ is not dense in X , then its closure is a nontrivial hyperinvariant subspace for A . Suppose, to the contrary, that for every $y \neq 0$ the space $\mathcal{A}y$ is dense in X . Then, in particular, for every $y \neq 0$ there exists $T \in \mathcal{A}$ such that $\|Ty - x_0\| < 1$. In other words, $y \in T^{-1}(D)$ for some $T \in \mathcal{A}$. Note that the set $T^{-1}(D)$ is open since D is open. So the family $\{T^{-1}(D) : T \in \mathcal{A}\}$ is an open cover for $X \setminus \{0\}$, and hence for the set $\overline{A(D)}$. Since this set is compact (because A is compact) there is a finite set $\{T_1, T_2, \dots, T_n\}$ in $\mathcal{A}$ such that

$$\overline{A(D)} \subset \bigcup_{i=1}^n T_i^{-1}(D).$$

In particular, $Ax_0 \in T_{i_1}^{-1}(D)$ for some $1 \leq i_1 \leq n$. This means that $T_{i_1}Ax_0 \in D$ and $AT_{i_1}Ax_0 \in A(D)$. So $AT_{i_1}Ax_0 \in T_{i_2}^{-1}(D)$ for some $1 \leq i_2 \leq n$. This means that $T_{i_2}AT_{i_1}Ax_0 \in D$. Continuing this process m times we see that

$$T_{i_m}AT_{i_{m-1}} \cdots T_{i_2}AT_{i_1}Ax_0$$

is in D , and since A commutes with the T 's

$$T_{i_m} \cdots T_{i_1}A^m x_0 \in D. \quad (22.6)$$

All the T_{i_j} here are from the finite set $\{T_1, \dots, T_n\}$. Let $c = \max \{\|T_i\| : 1 \leq i \leq n\}$. Then

$$\|T_{i_m} \cdots T_{i_1}A^m\| \leq c^m \|A^m\| = \|(cA)^m\|.$$

The operator cA has spectral radius 0. So, by the spectral radius formula $\|(cA)^m\|^{1/m}$ converges to 0, and hence $\|(cA)^m\|$ converges to 0. Thus

$$\|T_{i_m} \cdots T_{i_1}A^m x_0\| \rightarrow 0$$

as $m \rightarrow \infty$. So from (22.6) the point 0 is in the closure of the set D . This is a contradiction. ■

7. Each of the following statements is an easy corollary of Lomonosov's theorem.

1. Every compact operator has an invariant subspace. (This was proved by Aronszajn and Smith.)
2. A commuting family of compact operators has a common invariant subspace.
3. Every operator that commutes with a nonzero compact operator has an invariant subspace.

Compact Operators in Hilbert spaces

The case of Hilbert space, as in most problems, is simpler. The case of normal operators is especially simple and interesting. Before Riesz did it for Banach space operators, Hilbert had made an analysis of the spectrum of compact self-adjoint integral operators in the space L_2 . These ideas were extended by E. Schmidt to general Hilbert spaces—a term that came into existence later.

Let us recall that all our Hilbert spaces are assumed to be separable.

8. Hilbert–Schmidt Theorem. (The Spectral Theorem for Compact Operators.) Let $\mathcal{H}$ be an infinite dimensional Hilbert space and let A be a compact self-adjoint operator on $\mathcal{H}$. Then there exist an orthonormal basis $\{e_n\}$ and a sequence of real numbers $\{\lambda_n\}$ such that $Ae_n = \lambda_n e_n$ for all n , and $\lambda_n \rightarrow 0$ as $n \rightarrow \infty$.

Proof. Most of the work for the proof has already been done. We know that $\sigma(A)$ is real, and each nonzero point in $\sigma(A)$ is an eigenvalue of finite multiplicity. It is easy to see that eigenvectors corresponding to distinct eigenvalues are mutually orthogonal. (If $Ax = \lambda x$, and $Ay = \mu y$, then $(\lambda - \mu)\langle x, y \rangle = \langle \lambda x, y \rangle - \langle x, \mu y \rangle = \langle Ax, y \rangle - \langle x, Ay \rangle = 0$.) For each eigenvalue of A choose an orthonormal basis for the corresponding eigenspace. Let $\{e_n\}$ be the collection of all these eigenvectors for all the eigenvalues. This is an orthonormal set whose closed linear span $\mathcal{M}$ is invariant under A . Suppose the space $\mathcal{M}^\perp$ is nonzero. Since A is self-adjoint $\mathcal{M}^\perp$ is also invariant under A . Let A_0 be the restriction of A to $\mathcal{M}^\perp$. Then A_0 is self-adjoint and compact. If $\sigma(A_0)$ contains a nonzero point λ , then λ is an eigenvalue of A_0 and hence of A . (Because $\lambda x = A_0 x = Ax$.) Since all eigenvectors of A are in $\mathcal{M}$, this is not possible. Hence $\sigma(A_0) = \{0\}$, which means $\text{spr}(A_0) = 0$. Since A_0 is self-adjoint, this means $\|A_0\| = 0$, and hence $A_0 = 0$. Thus for every $x \in \mathcal{M}^\perp$ we have $Ax = A_0 x = 0$, which implies $x \in \mathcal{M}$. Hence $\mathcal{M}^\perp = \{0\}$ and $\mathcal{M} = \mathcal{H}$.

We have shown that $\{e_n\}$ is an orthonormal basis for $\mathcal{H}$ and there exist real numbers λ_n such that $Ae_n = \lambda_n e_n$. We have seen earlier (see Sections 7 and 10 of Lecture 20) that under these circumstances λ_n converges to 0. ■

9. With just one change—the λ_j are complex numbers—all assertions of the Hilbert–Schmidt theorem are valid for compact normal operators. The proof is essentially the same.

Thus every compact normal operator A has a special form

$$A = \sum_n \lambda_n \langle \cdot, e_n \rangle e_n \quad (22.7)$$

in which e_n is an orthonormal basis and $\{\lambda_n\}$ is a sequence of complex numbers converging to zero. This is also written as

$$A = \sum_n \lambda_n e_n e_n^*. \quad (22.8)$$

Here $e_n e_n^*$ is the orthogonal projection onto the one-dimensional space spanned by the vector e_n . The expression 22.8 is called the *spectral decomposition* of A .

If f is any bounded function on the set $\sigma(A)$ we define $f(A)$ as

$$f(A) = \sum_n f(\lambda_n) e_n e_n^*.$$

This is a bounded operator.

In particular, if A is compact and positive, we can define its positive square root $A^{1/2}$ using the spectral decomposition.

10. The spectral theorem shows that every compact normal operator A has a reducing subspace—a closed subspace $\mathcal{M}$ such that $\mathcal{M}$ and $\mathcal{M}^\perp$ both are invariant under A .

11. **The Singular Value Decomposition.** Let A be any compact operator on $\mathcal{H}$. Then there exist two orthonormal sets $\{e_n\}$ and $\{f_n\}$ in $\mathcal{H}$, and a sequence of

positive numbers $\{s_n\}$ converging to 0 such that

$$A = \sum_n s_n \langle \cdot, e_n \rangle f_n. \quad (22.9)$$

Proof. The operator A^*A is compact and positive. So there exists an orthonormal set $\{e_n\}$ and positive numbers s_n such that $A^*Ae_n = s_n^2 e_n$. The s_n^2 are all the nonzero eigenvalues of A^*A ; the operator A^*A vanishes on the orthogonal complement of the space spanned by the $\{e_n\}$. Let $f_n = \frac{1}{s_n}(Ae_n)$. Then

$$\langle f_n, f_m \rangle = \frac{1}{s_n s_m} \langle Ae_n, Ae_m \rangle = \frac{1}{s_n s_m} \langle A^*Ae_n, e_m \rangle = \delta_{nm},$$

i.e., the set $\{f_n\}$ consists of orthonormal vectors. Every vector x in $\mathcal{H}$ can be expanded as

$$x = \sum_n \langle x, e_n \rangle e_n + y,$$

where $y \in \ker A^*A = \ker |A|$. Using the polar decomposition $A = U|A|$ we see that $Ay = 0$. Thus $Ax = \sum_n \langle x, e_n \rangle Ae_n = \sum_n s_n \langle x, e_n \rangle f_n$. ■

We may expand the sequence $\{s_n\}$ to include the zero eigenvalues of A^*A and the sets $\{e_n\}$ and $\{f_n\}$ to orthonormal bases. The numbers s_n are called the *singular values* of A . They are the eigenvalues of the operator $|A|$. It is customary to arrange s_n in decreasing order. We have then an enumeration

$$s_1 \geq s_2 \geq \cdots \geq s_n \geq \cdots \geq 0$$

in which each s_j is repeated as often as its multiplicity as an eigenvalue of $|A|$. Whenever we talk of the singular value decomposition we assume that the s_j are arranged decreasingly.

12. Exercise. Let M_φ be a multiplication operator on the space $L_2[0, 1]$. Then M_φ is compact if and only if $\varphi = 0$ almost everywhere.

The Invariant Subspace Problem

Let X be any Banach space and let A be any (bounded linear) operator on it. Does there exist a (proper closed) subspace Y in X that is invariant under A ? This question is called the *Invariant subspace problem* and has been of much interest in functional analysis.

If A has an eigenvalue, then the subspace spanned by any eigenvector is an invariant subspace for A . If X is finite-dimensional, then every operator A on it has an eigenvalue and hence an invariant subspace. For the same reason every compact normal operator in a Hilbert space has an invariant subspace. The spectral theorem (to be proved later in this course) shows that every normal operator (whether compact or not) has an invariant subspace.

In 1949 von Neumann proved that every compact operator on a Hilbert space has an invariant subspace. In 1954 Aronszajn and Smith extended this result to all Banach spaces. For many years after that there was small progress on this problem. (Sample result: if there exists a polynomial p such that $p(A)$ is compact, then A has an invariant subspace.) Lomonosov's Theorem announced in 1973 subsumed most of the results then known, had a simple proof, and seemed to be valid for almost all operators. (One needs to ensure that A commutes with *some* nonzero compact operator.) The proof of Theorem 6 given here is due to H. M. Hilden.

Around 1980 P. Enflo constructed an example of a Banach space and an operator on it that has *no* invariant subspace. The same result was proved by C. J. Read, who also gave an example of an operator with no invariant subspace on the more familiar space l_1 .

The problem for Hilbert spaces remains unsolved.

Trace Ideals

Let A be a compact operator on (an infinite-dimensional) Hilbert space $\mathcal{H}$ and let

$$s_1(A) \geq s_2(A) \geq \cdots \geq 0 \quad (23.1)$$

be the singular values of A . The sequence $s_n(A)$ converges to 0. In this lecture we study special compact operators for which this sequence belongs to the space ℓ_1 or the space ℓ_2 .

Extremal Properties of Singular Values

The singular values have many interesting characterisations as solutions of some extremal problems. One of them is the following.

1. **Lemma.** Let A be a compact operator with singular values $\{s_n(A)\}$ counted as in (23.1). Then

$$s_n(A) = \min \{ \|A - F\| : \text{rank } F \leq n - 1 \}. \quad (23.2)$$

Proof. For brevity we write s_n for $s_n(A)$. Let A have the singular value decomposition

$$A = \sum s_n \langle \cdot, e_n \rangle f_n. \quad (23.3)$$

Let F be any operator with $\text{rank } F \leq n - 1$. Then we may choose a unit vector x in

the span of the vectors $\{e_1, \dots, e_n\}$ such that $Fx = 0$. We have

$$\|A - F\| \geq \|(A - F)x\| \geq \|Ax\| = \left\| \sum_{j=1}^n s_j \langle x, e_j \rangle f_j \right\|.$$

Using the Pythagoras Theorem, the last quantity above is equal to

$$\left(\sum_{j=1}^n s_j^2 |\langle x, e_j \rangle|^2 \right)^{1/2}.$$

Since $\sum_{j=1}^n |\langle x, e_j \rangle|^2 = 1$, this quantity is bounded below by s_n . So $\|A - F\| \geq s_n$. If we choose

$$F = \sum_{j=1}^{n-1} s_j \langle \cdot, e_j \rangle f_j, \quad (23.4)$$

then $\text{rank } F = n - 1$ and

$$A - F = \sum_{j=n}^{\infty} s_j \langle \cdot, e_j \rangle f_j.$$

This shows that $\|A - F\| = s_n$. ■

2. Corollary. Let A be a compact and B a bounded operator. Then

$$s_n(AB) \leq s_n(A)\|B\|,$$

$$s_n(BA) \leq s_n(A)\|B\|.$$

Proof. Let A and F be as in (23.3) and (23.4). Since $\text{rank } FB \leq n - 1$, we have from Lemma 1

$$s_n(AB) \leq \|AB - FB\| \leq \|A - F\|\|B\| = s_n(A)\|B\|.$$

This proves the first assertion. The second has a similar proof. ■

3. Corollary. (Continuity of singular values) Let A and B be compact operators. Then for all n

$$|s_n(A) - s_n(B)| \leq \|A - B\|.$$

Proof. From (23.2) we have

$$\begin{aligned} s_n(A) &= \min \|A - F\| = \min \|B - F + A - B\| \\ &\leq \min \|B - F\| + \|A - B\| = s_n(B) + \|A - B\|. \end{aligned}$$

Here the minimum is taken over all operators F with $\text{rank } F \leq n - 1$. Thus

$$s_n(A) - s_n(B) \leq \|A - B\|.$$

The right hand side of this inequality is symmetric in A and B . Hence we have also

$$s_n(B) - s_n(A) \leq \|A - B\|.$$

This proves the assertion. ■

Trace Class Operators

Let A be a compact operator such that

$$\sum_{n=1}^{\infty} s_n(A) < \infty.$$

Then we say that A belongs to $\mathcal{C}_1$, or that A is a *trace class operator*. In this case we define $\|A\|_1$ as

$$\|A\|_1 = \sum_{n=1}^{\infty} s_n(A). \quad (23.5)$$

The norm symbol is used in anticipation of what will be proved shortly.

4. Lemma. Let A be a trace class operator. Then for any two orthonormal sets $\{x_m\}$ and $\{y_m\}$ we have

$$\sum_{m=1}^{\infty} |\langle Ax_m, y_m \rangle| \leq \|A\|_1. \quad (23.6)$$

Proof. Represent A as in (23.3). Then

$$\begin{aligned} \sum_{m=1}^{\infty} |\langle Ax_m, y_m \rangle| &= \sum_{m=1}^{\infty} \left| \left\langle \sum_{n=1}^{\infty} s_n \langle x_m, e_n \rangle f_n, y_m \right\rangle \right| \\ &\leq \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} s_n |\langle x_m, e_n \rangle| |\langle f_n, y_m \rangle|. \end{aligned}$$

Since all the summands are positive, the two sums may be interchanged, and this double sum is equal to

$$\sum_{n=1}^{\infty} s_n \sum_{m=1}^{\infty} |\langle x_m, e_n \rangle| |\langle f_n, y_m \rangle|.$$

Using the Cauchy–Schwarz inequality, this is bounded by

$$\sum_{n=1}^{\infty} s_n \left(\sum_{m=1}^{\infty} |\langle x_m, e_n \rangle|^2 \right)^{1/2} \left(\sum_{m=1}^{\infty} |\langle f_n, y_m \rangle|^2 \right)^{1/2}.$$

Since e_n and f_n are unit vectors, by Bessel's inequality this expression is bounded by

$$\sum_{n=1}^{\infty} s_n = \|A\|_1.$$

■

5. The trace. Let $A \in \mathcal{C}_1$ and let $\{x_m\}$ be any orthonormal basis for $\mathcal{H}$. Let

$$\operatorname{tr} A = \sum_{m=1}^{\infty} \langle Ax_m, x_m \rangle. \quad (23.7)$$

Lemma 4 implies that this series converges absolutely and its terms may, therefore, be rearranged. We show that the sum in (23.7) does not depend on the orthonormal basis $\{x_m\}$.

Theorem. Let A be a trace class operator with singular value decomposition (23.3).

Then for every orthonormal basis $\{x_m\}$ we have

$$\sum_{m=1}^{\infty} \langle Ax_m, x_m \rangle = \sum_{n=1}^{\infty} s_n \langle f_n, e_n \rangle. \quad (23.8)$$

Proof. Using (23.3) we have

$$\sum_{m=1}^{\infty} \langle Ax_m, x_m \rangle = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} s_n \langle x_m, e_n \rangle \langle f_n, x_m \rangle.$$

The order of summation can be changed by the argument in the proof of Lemma 4 and we have

$$\begin{aligned}
 \sum_{m=1}^{\infty} \langle Ax_m, x_m \rangle &= \sum_{n=1}^{\infty} s_n \sum_{m=1}^{\infty} \langle x_m, e_n \rangle \langle f_n, x_m \rangle \\
 &= \sum_{n=1}^{\infty} s_n \sum_{m=1}^{\infty} \langle f_n, \langle e_n, x_m \rangle x_m \rangle \\
 &= \sum_{n=1}^{\infty} s_n \langle f_n, \sum_{m=1}^{\infty} \langle e_n, x_m \rangle x_m \rangle.
 \end{aligned}$$

Since $\{x_m\}$ is an orthonormal basis, we have $e_n = \sum_{m=1}^{\infty} \langle e_n, x_m \rangle x_m$. This proves the theorem. ■

The number $\text{tr } A$ defined by (23.7) is called the *trace* of A .

From Lemma 4 it follows that

$$|\text{tr } A| \leq \|A\|_1. \quad (23.9)$$

Warning. If A is an operator and

$$\sum_{m=1}^{\infty} |\langle Ax_m, x_m \rangle| < \infty. \quad (23.10)$$

for *some* orthonormal basis, then it is not necessary that A is trace class. (The right shift S on ℓ_2 is an example.) For A to be trace class the series (23.10) must converge for *every* orthonormal basis.

6. Theorem. The collection $\mathcal{C}_1$ is a vector space and $\|\cdot\|_1$ is a norm on it.

Proof. Let A and B be two elements of $\mathcal{C}_1$. Then $A + B$ is compact. Let

$$A + B = \sum_{n=1}^{\infty} s_n (A + B) \langle \cdot, e_n \rangle f_n$$

be the singular value decomposition. Then

$$s_n(A + B) = \langle (A + B)e_n, f_n \rangle \leq |Ae_n, f_n| + |Be_n, f_n|.$$

Hence

$$\begin{aligned} \sum_{n=1}^{\infty} s_n(A+B) &\leq \sum_{n=1}^{\infty} |\langle Ae_n, f_n \rangle| + \sum_{n=1}^{\infty} |\langle Be_n, f_n \rangle| \\ &\leq \|A\|_1 + \|B\|_1. \end{aligned}$$

This shows that $A+B$ is in $\mathcal{C}_1$ and $\|A+B\|_1 \leq \|A\|_1 + \|B\|_1$. The rest of the proof is easy. ■

7. Theorem. The space $\mathcal{C}_1$ with norm $\|\cdot\|_1$ is complete.

Proof. Clearly $\|A\| \leq \|A\|_1$ for every $A \in \mathcal{C}_1$. So if $\{A_n\}$ is a Cauchy sequence in $\mathcal{C}_1$, then it is a Cauchy sequence in the usual norm $\|\cdot\|$ as well. Since A_n are compact, there exists a compact operator A such that $\|A_n - A\|$ goes to 0. We will show that A is in $\mathcal{C}_1$, and $\|A_n - A\|_1$ converges to 0. By Corollary 3, for each j we have

$$\lim_{n \rightarrow \infty} s_j(A_n) = s_j(A).$$

Let ε be any positive number. By the diagonal procedure we can obtain a subsequence $\{A_k\}$ such that

$$|s_j(A_k) - s_j(A)| < \varepsilon/2^j \quad \text{for all } j.$$

This implies that

$$\sum_{j=1}^{\infty} s_j(A) < \sum_{j=1}^{\infty} \left(s_j(A_k) + \frac{\varepsilon}{2^j} \right) = \sum_{j=1}^{\infty} s_j(A_k) + \varepsilon,$$

and hence $A \in \mathcal{C}_1$. Now choose any n and let $(A_n - A) = \sum_{j=1}^{\infty} s_j(A_n - A) \langle \cdot, e_j \rangle f_j$ be the singular value decomposition of the operator $A_n - A$. Then for each positive integer N

$$\begin{aligned} \sum_{j=1}^N s_j(A_n - A) &= \sum_{j=1}^N \langle (A_n - A)e_j, f_j \rangle \\ &= \lim_{k \rightarrow \infty} \sum_{j=1}^N |\langle (A_n - A_k)e_j, f_j \rangle| \quad (\text{since } A_k \rightarrow A) \\ &\leq \lim_{k \rightarrow \infty} \|A_n - A_k\|_1 \quad (\text{by Lemma 4}). \end{aligned}$$

(The last limit exists because $\{\|A_n - A_k\|_1\}$ is a Cauchy sequence.) It follows that $\|A_n - A\|_1 \leq \lim_{k \rightarrow \infty} \|A_n - A_k\|_1$. Taking the limit as $n \rightarrow \infty$ one sees that $\|A_n - A\|_1$ converges to 0. ■

8. Theorem. Let A be a trace class operator and B any bounded operator. Then AB and BA are trace class and

$$\|AB\|_1 \leq \|A\|_1 \|B\|, \quad (23.11)$$

$$\|BA\|_1 \leq \|A\|_1 \|B\|. \quad (23.12)$$

Proof. Since A is compact, AB is also compact. Use Corollary 2 to complete the proof. ■

9. One of the important properties of trace in finite dimensions is that $\text{tr } AB = \text{tr } BA$ for any two matrices A and B . This remains true for trace class operators.

Theorem. Let A be a trace class operator and B any bounded operator. Then

$$\text{tr } AB = \text{tr } BA. \quad (23.13)$$

Proof. We prove this for a special case first. Let U be any unitary operator, and let $\{x_n\}$ be any orthonormal basis. Then the vectors $y_n = U^*x_n$ form another orthonormal basis. We have

$$\begin{aligned} \text{tr } UA &= \sum \langle UA x_n, x_n \rangle = \sum \langle A x_n, U^* x_n \rangle \\ &= \sum \langle AU y_n, y_n \rangle = \text{tr } AU. \end{aligned}$$

So, the equality (23.13) is true when B is a unitary operator. The general case follows from this because of the following lemma and the obvious fact that the trace is a linear functional on $\mathcal{C}_1$. ■

Lemma. Every bounded operator is a complex linear combination of four unitary operators.

Proof. First, let B be a self-adjoint operator with $\|B\| \leq 1$. Let

$$U_{\pm} = B \pm i(I - B^2)^{1/2}. \quad (23.14)$$

It is easy to see that $U_{\pm}$ are unitary operators. Clearly

$$B = \frac{1}{2}(U_+ + U_-).$$

If B is any self-adjoint operator, we may divide it by $\|B\|$ and reduce it to the special case above. Thus every self-adjoint operator is a linear combination of two unitary operators. Since every bounded operator is a linear combination of two self-adjoint operators the lemma is proved. ■

If b is a real number with $|b| \leq 1$, then $b = \cos \theta$ for some θ in $[0, \pi]$. In this case the equation (23.14) defines two numbers $\exp(\pm i\theta)$ whose sum is $2 \cos \theta$.

10. Summary. We have proved (most of) the following statements.

- (i) The collection $\mathcal{C}_1$ consisting of trace class operators is a 2-sided, $*$ -closed ideal in $\mathcal{B}(\mathcal{H})$.
- (ii) There is a natural norm $\|\cdot\|_1$ on $\mathcal{C}_1$ under which $\mathcal{C}_1$ is complete.
- (iii) Finite-rank operators are dense in $(\mathcal{C}_1, \|\cdot\|_1)$.
- (iv) $\mathcal{C}_1$ is not closed under the operator norm $\|\cdot\|$.
- (v) The formula (23.7) defines a linear functional called *trace* on $\mathcal{C}_1$. This has the property $\text{tr } AB = \text{tr } BA$.
- (vi) If $A \in \mathcal{C}_1$ and X, Y are any two bounded operators, then

$$\|XAY\|_1 \leq \|X\| \|A\|_1 \|Y\|. \quad (23.15)$$

If U and V are unitary operators, then

$$\|UAV\|_1 = \|A\|_1. \quad (23.16)$$

11. **Exercise.** Let A be a self-adjoint, trace class operator. Show that $\text{tr } A$ is the sum of the eigenvalues of A each counted as often as its multiplicity.

(This statement is true for all trace class operators, not necessarily self-adjoint. This is called Lidskii's Theorem and its proof is somewhat intricate.)

Hilbert-Schmidt Operators

A compact operator A is said to be a *Hilbert-Schmidt operator*, or to belong to the class $\mathcal{C}_2$ if

$$\sum_{n=1}^{\infty} s_n^2(A) < \infty.$$

Thus $A \in \mathcal{C}_2$ if and only if $A^*A \in \mathcal{C}_1$.

12. **Exercise.** Prove the following assertions.

(i) $\mathcal{C}_2$ is a two-sided, $*$ -closed ideal in $\mathcal{B}(\mathcal{H})$.

(ii) If A and B are in $\mathcal{C}_2$, then

$$\langle A, B \rangle := \text{tr } AB^* = \sum \langle AB^* x_n, x_n \rangle \quad (23.17)$$

is finite for every orthonormal basis $\{x_n\}$ and is independent of the choice of the orthonormal basis. This gives an inner product on $\mathcal{C}_2$.

(iii) The norm associated with this inner product is

$$\|A\|_2 = (\text{tr } AA^*)^{1/2} = \left(\sum_{n=1}^{\infty} s_n^2(A) \right)^{1/2}. \quad (23.18)$$

This is called the *Hilbert-Schmidt norm*. With this norm $\mathcal{C}_2$ is complete. Thus $\mathcal{C}_2$ is a Hilbert space.

(iv) If A is in $\mathcal{C}_1$, then it is in $\mathcal{C}_2$ and $\|A\| \leq \|A\|_2 \leq \|A\|_1$.

(v) Finite-rank operators are dense in $\mathcal{C}_2$.

(vi) $\mathcal{C}_2$ is not closed in $\mathcal{B}(\mathcal{H})$ with the $\|\cdot\|$ norm.

(vii) If $A \in \mathcal{C}_2$ and X, Y are bounded operators, then

$$\|XAY\|_2 \leq \|X\| \|A\|_2 \|Y\|. \quad (23.19)$$

If U and V are unitary operators, then

$$\|UAV\|_2 = \|A\|_2. \quad (23.20)$$

13. Example. The integral kernel operator K defined in Section 10 of Lecture 3 is a Hilbert-Schmidt operator on $L_2[0, 1]$. To see this choose an orthonormal basis $\{\varphi_n\}$ for $L_2[0, 1]$. Then the family

$$\psi_{mn}(x, y) = \varphi_m(x) \overline{\varphi_n(y)}$$

is an orthonormal basis for $L_2([0, 1] \times [0, 1])$. We have

$$\langle K\varphi_m, \varphi_n \rangle = \int_0^1 \int_0^1 K(x, y) \varphi_m(y) \overline{\varphi_n(x)} dx dy.$$

It follows that

$$\|K\|_2 = \text{tr } K^* K = \int_0^1 \int_0^1 |K(x, y)|^2 dx dy.$$

Schatten Classes

Let A be a compact operator and let $1 \leq p < \infty$. We say A is in the class $\mathcal{C}_p$ if

$$\sum_{n=1}^{\infty} s_n^p(A) < \infty.$$

For such an operator we define

$$\|A\|_p = \left(\sum_{n=1}^{\infty} s_n^p(A) \right)^{1/p}.$$

For every compact operator A we define

$$\|A\|_\infty = s_1(A) = \|A\|.$$

Each $\mathcal{C}_p$ is a 2-sided, $*$ -closed ideal in $\mathcal{B}(\mathcal{H})$. It is complete under the norm $\|\cdot\|_p$.

There is an interesting analogy between the sequence spaces ℓ_p and the *Schatten spaces* $\mathcal{C}_p$. This goes as follows. The class of all compact operators is thought of as being analogous to the space c_0 consisting of sequences that converge to 0. (The singular values of a compact operator are in the space c_0 .) The space $\mathcal{C}_p$, $1 \leq p < \infty$, is thought of being analogous to the space ℓ_p . We state without proof some facts.

- (i) For $1 < p < \infty$ the space $\mathcal{C}_p$ is the Banach space dual of $\mathcal{C}_q$, where $\frac{1}{p} + \frac{1}{q} = 1$.
- (ii) The space $\mathcal{C}_1$ is the Banach space dual of the space $\mathcal{B}_0(X)$ consisting of all compact operators.
- (iii) The space $\mathcal{B}(\mathcal{H})$ is the Banach space dual of $\mathcal{C}_1$.

Because of the noncommutativity of operator multiplication the spaces $\mathcal{C}_p$ are sometimes called noncommutative ℓ_p spaces.

Exercises. Here are two exercises arising from our discussion of singular values.

14. Let $\mathcal{F}_n$ be the collection of all operators whose rank is at most n . Show that this is a norm closed subset of $\mathcal{B}(\mathcal{H})$.

15. Let A be an $n \times n$ Hermitian matrix and let its eigenvalues be listed as

$$\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A).$$

Use Corollary 3 to show that if A and B are $n \times n$ Hermitian matrices, then

$$\max |\lambda_j(A) - \lambda_j(B)| \leq \|A - B\|.$$

The Spectral Theorem -I

Let A be a Hermitian operator on the space $\mathbb{C}^n$. Then there exists an orthonormal basis $\{e_j\}$ of $\mathbb{C}^n$ each of whose elements is an eigenvector of A . We thus have the representation

$$A = \sum_{j=1}^n \lambda_j \langle \cdot, e_j \rangle e_j, \quad (24.1)$$

where $Ae_j = \lambda_j e_j$. We can express this in other ways. Let $\lambda_1 > \lambda_2 > \dots > \lambda_k$ be the distinct eigenvalues of A and let $m_1, m_2, \dots, m_k$ be their multiplicities. Then there exists a unitary operator U such that

$$U^*AU = \sum_{j=1}^k \lambda_j P_j, \quad (24.2)$$

where $P_1, P_2, \dots, P_k$ are mutually orthogonal projections and

$$\sum_{j=1}^k P_j = I. \quad (24.3)$$

The range of P_j is the m_j -dimensional eigenspace of A corresponding to the eigenvalue λ_j . This is called the *spectral theorem* for finite-dimensional operators.

In Lecture 22 we saw how this theorem may be extended to compact self-adjoint operators in an infinite-dimensional Hilbert space $\mathcal{H}$. The extension seemed a minor step: the finite sum in (24.1) was replaced by an infinite sum. It is time now to go beyond compact operators and to consider all bounded self-adjoint operators. The spectral theorem in this case is a more substantial extension of the finite-dimensional theorem. It has several different formulations, each of which emphasizes a different

viewpoint and each of which is useful in different ways. We will study some of these versions.

In Lecture 18 we studied multiplication operators. Let $(X, \mathcal{S}, \mu)$ be a σ -finite measure space. Every bounded measurable function φ on X induces an operator M_φ on the Hilbert space $L_2(\mu)$ by the action $M_\varphi f = \varphi f$ for every $f \in L_2(\mu)$. If φ is a real function, then M_φ is self-adjoint.

If $\{\mathcal{H}_n\}$ is a countable family of Hilbert spaces we define their direct sum

$$\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \cdots = \bigoplus_{n=1}^{\infty} \mathcal{H}_n$$

as follows. Its elements consist of sequences

$$\mathbf{x} = (x_1, x_2, \dots)$$

where $x_j \in \mathcal{H}_j$ and $\sum \|x_j\|^2 < \infty$. The inner product on $\mathcal{H}$ is defined as

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n=1}^{\infty} \langle x_n, y_n \rangle,$$

and this makes $\mathcal{H}$ into a Hilbert space.

If $\{\mu_n\}$ is a sequence of measures on $(X, \mathcal{S})$ we may form the Hilbert space $\bigoplus_n L_2(\mu_n)$. Each bounded measurable function φ on X induces a multiplication operator M_φ on this space by the action

$$M_\varphi \mathbf{f} = \varphi \mathbf{f} = (\varphi f_1, \varphi f_2, \dots).$$

A very special and simple situation is the case when X is an interval $[a, b]$ and $\varphi(t) = t$. The induced multiplication operator M_φ is then called a *canonical multiplication operator*. For brevity we write this operator as M . One version of the spectral theorem says that every self-adjoint operator on a Hilbert space is equivalent to a canonical multiplication operator.

1. The Spectral Theorem (Multiplication operator form). Let A be a self-adjoint operator on a Hilbert space $\mathcal{H}$. Then there exist a sequence of probability

measures $\{\mu_n\}$ on the interval $X = [-\|A\|, \|A\|]$, and a unitary operator U from $\mathcal{H}$ onto the Hilbert space $\oplus_n L_2(\mu_n)$ such that $UAU^* = M$, the canonical multiplication operator on $\oplus_n L_2(\mu_n)$.

The theorem is proved in two steps. First we consider a special case when A has a cyclic vector. The proof in this case is an application of the Riesz Representation Theorem or Helly's Theorem proved in Lectures 7 and 8. We follow arguments that lead from the finite-dimensional case to the infinite-dimensional one, thereby reducing the mystery of the proof to some extent.

2. Cyclic spaces and vectors. Let A be any operator on $\mathcal{H}$. Given a vector x let S be the closure of the linear span of the family $\{x, Ax, A^2x, \dots\}$. We say that S is a *cyclic subspace* of $\mathcal{H}$ with x as a *cyclic vector*. If there exists a vector x_0 such that the cyclic subspace corresponding to it is the entire space $\mathcal{H}$ we say that A has a *cyclic vector* x_0 in $\mathcal{H}$.

3. Proposition. Suppose A is a self-adjoint operator with a cyclic vector in $\mathcal{H}$. Then there exist a probability measure μ on the interval $X = [-\|A\|, \|A\|]$, and a unitary operator U from $\mathcal{H}$ onto $L_2(\mu)$ such that $UAU^* = M$, the canonical multiplication operator in $L_2(\mu)$.

Proof. Let x_0 be a cyclic vector for A . We may assume $\|x_0\| = 1$. Using the Gram-Schmidt procedure obtain from the set $\{x_0, Ax_0, A^2x_0, \dots\}$ an orthonormal basis $\{y_0, y_1, y_2, \dots\}$ for $\mathcal{H}$. Let S_n be the subspace spanned by the first n vectors in this basis, and P_n the orthogonal projection onto S_n . The sequence $\{P_n\}$ is an increasing sequence that converges strongly to I . Let $A_n = P_n A P_n$. Then $\|A_n\| \leq \|A\|$ and A_n converges strongly to A . The operator A_n annihilates $S_n^\perp$ and maps S_n into itself. Let $\tilde{A}_n$ be the restriction of A_n to S_n .

We apply the known finite-dimensional spectral theorem to the Hermitian oper-

ator $\tilde{A}_n$ on the n -dimensional space S_n . Let $\lambda_{n1} > \lambda_{n2} > \cdots > \lambda_{nk_n}$ be the distinct eigenvalues of $\tilde{A}_n$. Then $|\lambda_{nj}| \leq \|\tilde{A}_n\| \leq \|A\|$, and there exist mutually orthogonal projections with ranges contained in S_n such that

$$\tilde{A}_n = \sum_{j=1}^{k_n} \lambda_{nj} P_{nj}. \quad (24.4)$$

There is no harm in thinking of P_{nj} as projections on $\mathcal{H}$; all of them annihilate $S_n^\perp$. Then the right hand side of (24.4) is equal to A_n .

Given a measurable subset E of the interval $X = [-\|A\|, \|A\|]$ let

$$\mu_n(E) = \sum_{j: \lambda_{nj} \in E} \langle P_{nj} x_0, x_0 \rangle. \quad (24.5)$$

It is easy to see that μ_n is a probability measure on X concentrated at the points $\{\lambda_{n1}, \lambda_{n2}, \dots, \lambda_{nk_n}\}$. (Use the properties of the projections P_{nj} to check that μ_n is nonnegative and countably additive, and $\mu_n(X) = 1$.)

This gives us a sequence $\{\mu_n\}$ of probability measures on X . By the Montel-Helly Selection Principle (Lecture 8), there exists a subsequence $\{\mu_n\}$ and a probability measure μ such that for every continuous function f on X

$$\lim_{n \rightarrow \infty} \int f d\mu_n = \int f d\mu.$$

Since the measure μ_n is concentrated at the finite set $\{\lambda_{n1}, \dots, \lambda_{nk_n}\}$ we have

$$\int f d\mu_n = \sum_{j=1}^{k_n} f(\lambda_{nj}) \langle P_{nj} x_0, x_0 \rangle.$$

Applying this to the functions $f(t) = t^r$, $r = 0, 1, 2, \dots$ we see that

$$\int t^r d\mu_n(t) = \sum_{j=1}^{k_n} \lambda_{nj}^r \langle P_{nj} x_0, x_0 \rangle. \quad (24.6)$$

From the representation (24.4) we see that the right hand side of (24.6) is equal to $\langle A_n^r x_0, x_0 \rangle$. Since $A_n \xrightarrow{s} A$, we have

$$\langle A^r x_0, x_0 \rangle = \lim_{n \rightarrow \infty} \langle A_n^r x_0, x_0 \rangle = \lim_{n \rightarrow \infty} \int t^r d\mu_n(t) = \int t^r d\mu(t), \quad (24.7)$$

for $r = 0, 1, 2, \dots$

For $r = 0, 1, 2, \dots$, let $\varphi_r(t) = t^r$. The collection $\{\varphi_r(t)\}$ is a fundamental set in $L_2(X, \mu)$ while the collection $\{A^r x_0\}$ is a fundamental set in $\mathcal{H}$. Define a map U between these two sets as follows

$$U(A^r x_0) = \varphi_r, \quad r = 0, 1, 2, \dots \quad (24.8)$$

By the definition of the inner product in $L_2(X, \mu)$ we have

$$\langle \varphi_r, \varphi_s \rangle = \int t^{r+s} d\mu(t).$$

From (24.7) we have, therefore

$$\langle \varphi_r, \varphi_s \rangle = \langle A^{r+s} x_0, x_0 \rangle = \langle A^r x_0, A^s x_0 \rangle.$$

In other words

$$\langle U(A^r x_0), U(A^s x_0) \rangle = \langle A^r x_0, A^s x_0 \rangle.$$

Thus the map U preserves inner products. Since $\{A^r x_0\}$ is a fundamental set in $\mathcal{H}$ we can extend U uniquely to a linear isometry from $\mathcal{H}$ into $L_2(X, \mu)$. The range of an isometry is always closed. In this case the range contains all polynomial functions, and hence is equal to $L_2(X, \mu)$. Thus U is a unitary operator. From the equation (24.8) defining U we have

$$(U A U^*)(\varphi_r) = U A(A^r x_0) = U(A^{r+1} x_0) = \varphi_{r+1}.$$

In other words

$$(U A U^* \varphi_r)(t) = \varphi_{r+1}(t) = t \varphi_r(t).$$

Since the set $\{\varphi_r\}$ is fundamental in $L_2(X, \mu)$ we have

$$(U A U^* f)(t) = t f(t) \quad \text{for all } f \in L_2(X, \mu).$$

This proves the proposition ■

4. Proof of Theorem 1. Let x_1 be any unit vector in $\mathcal{H}$ and let S_1 be the closed linear span of the set $\{x_1, A x_1, A^2 x_1, \dots\}$. If $S_1 = \mathcal{H}$, the theorem reduces to the

case considered above. If $S_1 \neq \mathcal{H}$, then S_1 is an invariant space for A , and so is $S_1^\perp$. Let x_2 be a unit vector in $S_1^\perp$ and let S_2 be the closed linear span of the set $\{x_2, Ax_2, A^2x_2, \dots\}$. An application of Zorn's Lemma shows that $\mathcal{H}$ can be written as a countable direct sum

$$\mathcal{H} = S_1 \oplus S_2 \oplus \cdots \oplus S_n \oplus \cdots,$$

in which each S_n is a cyclic subspace and the S_n are mutually orthogonal.

Proposition 3 can be applied to each cyclic space S_n to get a measure μ_n . The theorem follows from this. ■

Examples

5. Let $\mathcal{H} = \mathbb{C}^n$ and let A be a Hermitian operator on $\mathcal{H}$ with distinct eigenvalues $\lambda_1, \dots, \lambda_n$. Define a probability measure μ on $[-\|A\|, \|A\|]$ by the rule $\mu(\{\lambda_j\}) = \frac{1}{n}$ and $\mu(E) = 0$ for every set that contains none of the points λ_j . Then the space $L_2(\mu)$ is $\mathbb{C}^n$. A typical element of this space may be written as $f(\lambda) = (f(\lambda_1), \dots, f(\lambda_n))$. The spectral theorem tells us that A is equivalent to the operator that sends $f(\lambda)$ to $\lambda f(\lambda) = (\lambda_1 f(\lambda_1), \dots, \lambda_n f(\lambda_n))$.

Note that in this example we could have chosen any probability measure μ such that $\mu(\{\lambda_j\}) = a_j > 0$ and $\sum_{j=1}^n a_j = 1$.

6. Let A be a Hermitian operator on $\mathbb{C}^n$ with some multiple eigenvalues. Arrange the distinct eigenvalues as $\lambda_1, \dots, \lambda_k$ in such a way that their respective multiplicities are $m_1 \geq m_2 \geq \cdots \geq m_k$. Pick up for each λ_j one eigenvector x_j . Let $\mathcal{H}_1$ be the k -dimensional space spanned by these vectors. Let μ_1 be the probability measure on $[-\|A\|, \|A\|]$ such that $\mu_1(\{\lambda_j\}) = 1/k$, $1 \leq j \leq k$. Then $L_2(\mu_1) = \mathbb{C}^k$, the space $\mathcal{H}_1$ is isomorphic to this, and the operator A restricted to $\mathcal{H}_1$ is equivalent to the canonical multiplication operator on $L_2(\mu_1)$. Now consider the restriction of A to the orthogonal complement $\mathcal{H}_1^\perp$. The eigenvalues of this operator are $\lambda_1, \dots, \lambda_k$ with multiplicities $(m_1 - 1) \geq (m_2 - 1) \geq \cdots \geq (m_k - 1)$. If ℓ is the largest index

for which $m_\ell > 1$, then the eigenvalues $\lambda_{\ell+1}, \dots, \lambda_k$ no longer occur in this listing. Pick up vectors $y_1, \dots, y_\ell$ in $\mathcal{H}_1^\perp$ such that y_j is an eigenvector of A corresponding to λ_j . Let $\mathcal{H}_2$ be the ℓ -dimensional space spanned by these vectors. Let μ_2 be the probability measure on $[-\|A\|, \|A\|]$ such that $\mu_2(\{\lambda_j\}) = 1/\ell$ for $1 \leq j \leq \ell$. The space $\mathcal{H}_2$ is isomorphic to $L_2(\mu_2)$, and A restricted to $\mathcal{H}_2$ is equivalent to the canonical multiplication operator on $L_2(\mu_2)$. This process can be repeated, and we get measures $\mu_1, \mu_2, \dots, \mu_r$ such that A is equivalent to the canonical multiplication operator on $L_2(\mu_1) \oplus L_2(\mu_2) \cdots \oplus L_2(\mu_r)$.

7. Let A be a compact operator on an infinite-dimensional Hilbert space. Let $\{0, \lambda_1, \lambda_2, \dots\}$ be its spectrum. The idea of the above examples may be modified to get a family of measures $\{\mu_n\}$ concentrated on the spectrum of A . The spectral theorem for compact operators proved earlier in Lecture 22 is equivalent to Theorem 1.

8. Let $\mathcal{H} = L_2[-1, 1]$ and let A be the operator on $\mathcal{H}$ defined as $(Af)(x) = x^2 f(x)$. Then A is a positive operator and $\sigma(A)$ is a subset of $[0, 1]$. Fill in the gaps in arguments that follow.

- (i) The operator A has no cyclic vector. This can be proved as follows. Let $\text{sgn}(x)$ be the function that takes the value 1 when $x \geq 0$ and -1 when $x < 0$. For any function f in $\mathcal{H}$ let $g(x) = \text{sgn}(x) \overline{f(-x)}$. Then

$$\int_{-1}^1 g(x) x^{2n} \overline{f(x)} dx = \int_{-1}^1 \text{sgn}(x) \overline{f(-x)} x^{2n} \overline{f(x)} dx.$$

The integrand is an odd function and so the integral is zero. This shows that g is orthogonal to $\{f, Af, A^2 f, \dots\}$.

- (ii) Let $\mathcal{H}_{\text{even}}$ and $\mathcal{H}_{\text{odd}}$ be subspaces of $\mathcal{H}$ consisting of even and odd functions, respectively. These two spaces are mutually orthogonal and each of them is a cyclic subspace for A .

(iii) Define a map U from $\mathcal{H}_{\text{even}}$ onto $L_2[0, 1]$ as follows. For $\varphi \in \mathcal{H}_{\text{even}}$

$$(U\varphi)(t) = \frac{\varphi(t^{1/2})}{t^{1/4}}, \quad t \in [0, 1]. \quad (24.9)$$

The inverse of this map takes a function f in $L_2[0, 1]$ to the function

$$(U^{-1}f)(x) = |x|^{1/2}f(x^2), \quad x \in [-1, 1].$$

Show that U is unitary. Check that UAU^* is the canonical multiplication operator on $L_2[0, 1]$.

(iv) Use the formula (24.9) to define a map U from $\mathcal{H}_{\text{odd}}$ to $L_2[0, 1]$. In this case the inverse of this map is

$$(U^{-1}f)(x) = \begin{cases} x^{1/2}f(x^2) & \text{if } x \geq 0, \\ -|x|^{1/2}f(x^2) & \text{if } x < 0. \end{cases}$$

Show that U is unitary and UAU^* is again the canonical multiplication operator on $L_2[0, 1]$.

(v) Thus we have shown

$$L_2[-1, 1] = L_2[0, 1] \oplus L_2[0, 1]$$

and the multiplication operator $f(x) \mapsto x^2f(x)$ in $L_2[-1, 1]$ is equivalent to the canonical multiplication operator in $L_2[0, 1] \oplus L_2[0, 1]$.

9. The spectral theorem, another form. One can replace the family $\{\mu_n\}$ occurring in Theorem 1 by a single measure. The price to be paid is that the underlying space $[-\|A\|, \|A\|]$ is replaced by a more complicated space. One way of doing this is as follows. Let $X_n = [-\|A\|, \|A\|]$ for all $n = 1, 2, \dots$. Let $X = \cup_{n=1}^{\infty} X_n$, where this union is understood to mean a union of different copies of the same space. Let μ be the probability measure on X defined by the requirement that its restriction to the n th copy in the union above is the measure $\mu_n/2^n$. Then μ is a probability measure on X and the space $L_2(\mu)$ is isomorphic to the space $\oplus L_2(\mu_n)$. The operator A is

now equivalent to a multiplication operator M_φ in $L_2(\mu)$, where φ is a real-valued bounded measurable function on X .

Support of the spectral measures

The measures $\{\mu_n\}$ associated with A by the spectral theorem are called *spectral measures*. They are measures on the interval $X = [-\|A\|, \|A\|]$. In the familiar situation of Examples 5, 6 and 7 we saw that these measures are concentrated on $\sigma(A)$ and vanish on the rest of X . This is, in fact, true always.

Let μ be a measure on a second countable Hausdorff topological space X with its Borel σ -algebra. Let E be the union of all open sets G in X for which $\mu(G) = 0$. The set $X \setminus E$ is called the *support* of μ , and written as $\text{supp } \mu$. In other words $\text{supp } \mu$ is the smallest closed set F such that $\mu(F^c) = 0$.

10. Exercise. (i) Let M be the canonical multiplication operator in $L_2(X, \mu)$. Show that $\sigma(M) = \text{supp } \mu$. [Hint: If φ is any bounded measurable function, then $\sigma(M_\varphi)$ is the essential range of φ . See Lecture 18.]

(ii) Let A be a self-adjoint operator with a cyclic vector and let μ be a spectral measure associated with it (as in Proposition 3). Then

$$\text{supp } \mu = \sigma(A).$$

[If $B = UAU^*$, then A and B have the same spectrum.]

11. Theorem. Let A be a self-adjoint operator and let $\{\mu_n\}$ be its spectral measures. Then

$$\sigma(A) = \overline{\cup_n (\text{supp } \mu_n)}$$

(The set on the right hand side is called the support of the family $\{\mu_n\}$ and is written as $\text{supp } \{\mu_n\}$.)

The uniqueness question

We saw that spectral measures associated with A are not unique. This is less serious than it seems at first.

Two measures μ and ν on X are said to be *equivalent* if they have the same null sets; i.e., $\mu(E) = 0 \Leftrightarrow \nu(E) = 0$. If $\mu(E) = 0$ whenever $\nu(E) = 0$ we say $\mu \ll \nu$. (μ is *absolutely continuous* with respect to ν .)

12. Exercise. (i) If μ and ν are equivalent measures, then the Hilbert spaces $L_2(\mu)$ and $L_2(\nu)$ are isomorphic. The operator U defined by

$$(Uf)(x) = \sqrt{\frac{d\mu}{d\nu}(x)} f(x)$$

is a unitary operator from $L_2(\mu)$ onto $L_2(\nu)$. ($d\mu/d\nu$ is the Radon-Nikodym derivative; it exists whenever $\mu \ll \nu$.)

(ii) Let M_μ and M_ν be the canonical multiplication operators on $L_2(\mu)$ and $L_2(\nu)$. Then

$$UM_\mu U^* = M_\nu.$$

(iii) Conversely, suppose μ and ν are two measures on X , and there exists a unitary operator U from $L_2(\mu)$ onto $L_2(\nu)$ such that $UM_\mu U^* = M_\nu$. Then μ and ν are equivalent. [To prove this proceed as follows. Show that $UM_\mu^k U^* = M_\nu^k$ for all $k = 0, 1, 2, \dots$. Let $\varphi_k(x) = x^k$ and let $g_0 = U(\varphi_0)$. Show that $U(\varphi_k) = \varphi_k g_0$. Since U is unitary this shows that

$$\int x^k d\mu(x) = \int x^k g_0^2(x) d\nu(x), \quad k = 0, 1, 2, \dots$$

It follows that

$$\int f(x) d\mu(x) = \int f(x) g_0^2(x) d\nu(x),$$

for all $f \in L_2(\mu)$. This shows that $\mu \ll \nu$.]

This leaves the question of “multiplicities” of the spectral measures: how often

does a measure μ_n occur in the sequence $\{\mu_n\}$ of spectral measures. This too has a natural and pleasing answer, and we leave it at that for now.

The Spectral Theorem -II

Look at the formulation of the finite-dimensional spectral theorem given in (24.2) and (24.3). For infinite dimensional compact operators the finite sum is replaced by an infinite sum. Going further we may replace the sum by an integral. The second version of the spectral theorem says that every self-adjoint operator can be represented by such an integral. The integration is now with respect to a projection-valued measure (instead of an ordinary positive measure) and the resulting definite integral is a self-adjoint operator (instead of a number.)

Projection-valued measures

Let $(X, \mathcal{S})$ be a measurable space (any set X with a σ -algebra of subsets $\mathcal{S}$). Let $\mathcal{P}(\mathcal{H})$ be the collection of all orthogonal projection operators in a Hilbert space $\mathcal{H}$. A *projection-valued measure* on X is a map $E \mapsto P(E)$ from $\mathcal{S}$ into $\mathcal{P}(\mathcal{H})$ with the following properties:

- (i) If $\{E_n\}$ is a countable family of mutually disjoint elements of $\mathcal{S}$, then

$$P(\cup_n E_n) = \sum_n P(E_n),$$

where the series on the right converges in the strong operator topology.

- (ii) $P(X) = I$, the identity operator in $\mathcal{H}$.

1. **Exercise.** Some other properties of a projection-valued measure can be derived

from properties of sums, differences and products of projections stated in Lecture 15. Prove the following:

- (i) $P(\phi) = 0$.
- (ii) $P(E - F) = P(E) - P(F)$ if $F \subset E$.
- (iii) $P(E)$ is orthogonal to $P(F)$ if E and F are disjoint.
- (iv) $P(E)P(F) = P(F)P(E) = P(E \cap F)$.

The multiplicative property (iv) is important. Let us indicate its proof. Write $E \cup F$ as a disjoint union

$$E \cup F = (E \cap F') \cup (F \cap E') \cup (E \cap F).$$

This shows that

$$P(E \cup F) = P(E \cap F') + P(F \cap E') + P(E \cap F).$$

Multiply on the left by $P(E)$ and on the right by $P(F)$ to get

$$\begin{aligned} P(E)P(E \cup F)P(F) &= P(E)P(E \cap F')P(F) + P(E)P(F \cap E')P(E) \\ &\quad + P(E)P(E \cap F)P(F). \end{aligned}$$

The first two terms on the right hand side vanish (as $P(F) \perp P(E \cap F')$ and $P(E) \perp P(F \cap E')$.) Since E is a subset of $E \cup F$ we have $P(E)P(E \cup F) = P(E)$. So, the left hand side is equal to $P(E)P(F)$. Similarly the fourth term on the right hand side is equal to $P(E \cap F)$.

If $P(\cdot)$ is projection-valued measure on $(X, \mathcal{S})$ with values in $\mathcal{P}(\mathcal{H})$ we say P is a projection-valued measure on X in $\mathcal{H}$. For brevity, we will write pvm for a projection-valued measure.

2. **Exercise.** Let P_1 be a pvm on X in a Hilbert space $\mathcal{H}_1$. Let $\mathcal{H}_2$ be another Hilbert space and U a unitary operator from $\mathcal{H}_1$ onto $\mathcal{H}_2$. For all measurable sets E , let

$$P_2(E) = UP_1(E)U^*.$$

Then P_2 is a pvm on X in $\mathcal{H}_2$. We say that P_1 and P_2 are *unitarily equivalent* pvm's.

3. **Exercise.** Let $\{\mathcal{H}_n\}$ be a family of Hilbert spaces and let $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \cdots$ be their direct sum. Let P_n be a pvm on X in $\mathcal{H}_n$. If $\mathbf{x} = (x_1, x_2, \dots)$ is an element of $\mathcal{H}$, let

$$P(E)\mathbf{x} = (P_1(E)x_1, P_2(E)x_2, \dots).$$

Then P is a pvm on X in $\mathcal{H}$. We say that P is the *direct sum* of $P_1, P_2, \dots$ and write this as $P = \oplus_n P_n$.

4. The canonical pvm.

Let μ be a measure on $(X, \mathcal{S})$ and let $\mathcal{H}$ be the space $L_2(\mu)$. For every $f \in L_2(\mu)$ and $E \in \mathcal{S}$ let

$$(P^\mu(E))f = \chi_E f,$$

where χ_E is the characteristic function of the set E . Then P^μ is a pvm on X in the space $\mathcal{H} = L_2(X, \mathcal{S}, \mu)$. This is called a *canonical pvm*.

A little more generally, let $\{\mu_n\}$ be a family of measures on $(X, \mathcal{S})$ and let $\mathcal{H} = \oplus_n L_2(\mu_n)$. Let P^{μ_n} be the canonical pvm on X in $L_2(\mu_n)$. Then the pvm $P = \oplus_n P^{\mu_n}$ is called a *canonical pvm*.

5. **Proposition.** Let $(X, \mathcal{S})$ be a measurable space and let $E \mapsto P(E)$ be a map from $\mathcal{S}$ into $\mathcal{P}(\mathcal{H})$. For each unit vector v in $\mathcal{H}$ let

$$\mu_v(E) := \langle P(E)v, v \rangle = \|P(E)v\|^2. \quad (25.1)$$

Then P is a pvm if and only if for every v μ_v is a probability measure on $(X, \mathcal{S})$.

Proof. It is easy to see that each μ_ν is a probability measure if P is a pvm. Conversely, suppose each μ_ν is a probability measure. If E and F are two disjoint sets, then for all v

$$\langle P(E \cup F)v, v \rangle = \mu_\nu(E \cup F) = \mu_\nu(E) + \mu_\nu(F) = \langle P(E)v, v \rangle + \langle P(F)v, v \rangle.$$

Hence

$$P(E \cup F) = P(E) + P(F).$$

Since $P(\cdot)$ is a projection operator, this means that $P(E) \perp P(F)$ whenever E and F are disjoint.

Now let $\{E_n\}$ be a family of mutually disjoint sets. Then $\{P(E_n)\}$ is a family of mutually orthogonal projections. Hence the series $\sum_{n=1}^{\infty} P(E_n)$ converges strongly to a projection. (The sequence of partial sums of this series is an increasing sequence of projections.) Hence we have

$$\begin{aligned} \left\langle \sum_{n=1}^{\infty} P(E_n)v, v \right\rangle &= \sum_{n=1}^{\infty} \langle P(E_n)v, v \rangle = \sum_{n=1}^{\infty} \mu_\nu(E_n) = \mu_\nu(\cup_{n=1}^{\infty} E_n) \\ &= \left\langle P(\cup_{n=1}^{\infty} E_n)v, v \right\rangle. \end{aligned}$$

Thus $P(\cdot)$ is countably additive on $\mathcal{S}$. This shows that P is a pvm. ■

6. Exercise. Let P be a pvm on X in $\mathcal{H}$. Given vectors u, v in $\mathcal{H}$, let

$$\mu_{u,v}(E) = \langle P(E)u, v \rangle.$$

Then $\mu_{u,v}$ is a complex measure on X .

Integration

Following familiar ideas of Lebesgue integration we may define an integral $\int f dP$ in which f is a complex function and P a pvm on X . This is done in the exercises and propositions that follows.

Start, as usual, with a simple function

$$s = \sum_{i=1}^k a_i \chi_{E_i}$$

in which $\{E_i\}$ are disjoint measurable sets that cover X and a_i are complex numbers.

It is natural to define

$$\int s \, dP = \sum_{i=1}^k a_i P(E_i).$$

This integral is written also as $\int s(x) dP(x)$, or as $\int s(x) P(dx)$. This is an operator on $\mathcal{H}$.

7. Exercise. Verify the following properties:

(i) For all simple functions s_1 and s_2 , and complex numbers α_1 and α_2 we have

$$\int (\alpha_1 s_1 + \alpha_2 s_2) dP = \alpha_1 \int s_1 dP + \alpha_2 \int s_2 dP.$$

(ii) $\int (s_1 s_2) dP = \int s_1 dP \cdot \int s_2 dP$.

(iii) $(\int s \, dP)^* = \int \bar{s} \, dP$.

(iv) If v is a unit vector in $\mathcal{H}$ and μ_v the measure associated with P according to the equation (25.1), then

$$\|(\int s \, dP)v\|^2 = \int |s|^2 d\mu_v.$$

(v) $\|\int s \, dP\| \leq \sup_{x \in X} |s(x)|$.

8. Now let f be any bounded measurable function on X . Then there exists a sequence of simple functions $\{s_n\}$ such that s_n converges uniformly to f . From the Property

(v) in Exercise 7, we see that for any m and n

$$\|\int s_n dP - \int s_m dP\| \leq \sup |s_n(x) - s_m(x)|.$$

This goes to 0 as n and m go to ∞ . Thus $\{\int s_n dP\}$ is a Cauchy sequence in $\mathcal{B}(\mathcal{H})$. Hence it has a limit

$$A = \lim_{n \rightarrow \infty} \int s_n dP.$$

Show that this limit does not depend on the sequence $\{s_n\}$ that was chosen to approximate f . Thus we may define the integral $\int f dP$ by putting

$$\int f dP = \lim_{n \rightarrow \infty} \int s_n dP,$$

where $\{s_n\}$ is any sequence of simple functions converging uniformly to f . The limit here is the limit in the norm topology of $\mathcal{B}(\mathcal{H})$.

9. Exercise. Verify that $\int f dP$ defined above satisfies the five properties analogous to those proved in Exercise 7. (The supremum in Property (v) is now replaced by the essential supremum of f .)

10. Note that $\int f dP$ so defined is a normal operator on $\mathcal{H}$. If f is real, this operator is self-adjoint and if $|f(x)| = 1$, then it is unitary.

What we have called $\int f dP$ is the integral over the entire space X . If E is any measurable subset of X we put

$$\int_E f dP = \int (\chi_E f) dP.$$

Other conventions of ordinary integration are adopted in an obvious way.

11. The Lebesgue Dominated Convergence Theorem

This most useful theorem of Lebesgue integration is true in this context:

Theorem. Let $\{f_n\}$ be a sequence of bounded measurable functions on X and suppose there exists a number k such that

$$\text{ess sup } |f_n(x)| \leq k \text{ for all } n.$$

Suppose $f_n \rightarrow f$ almost everywhere (with respect to the pvm P). Then the sequence of operators $\{\int f_n dP\}$ converges strongly to the operator $\int f dP$.

Proof. We use the measures μ_v defined in (25.1) to reduce the problem to one about ordinary measures.

The assumption that $f_n \rightarrow f$ except possibly on a set E with $P(E) = 0$ implies that for every unit vector v , $f_n \rightarrow f$ almost everywhere with respect to μ_v . Hence by the (ordinary) Lebesgue dominated convergence theorem the integral $\int |f_n - f|^2 d\mu_v$ converges to 0. By property (iv) of Exercises 7 and 9

$$\|(\int f_n dP - \int f dP)v\|^2 = \int |f_n - f|^2 d\mu_v.$$

To say that this goes to 0 for all v is to say that $\int f_n dP$ converges strongly to $\int f dP$. ■

12. Exercise. Under the hypotheses of the theorem above it is not necessary that $\int f_n dP$ converges in norm to $\int f dP$. To see this let $X = [0, 1]$, $\mathcal{H} = L_2[0, 1]$, and let P be the canonical pvm. For each n let f_n be the characteristic function of the interval $[0, 1 - 1/n]$. Observe that $\int f_n dP$ is not a Cauchy sequence in $\mathcal{B}(\mathcal{H})$.

13. Exercise. Let P_1 and P_2 be two unitarily equivalent measures on X . Let $P_2(E) = UP_1(E)U^*$ for all E . Then for all bounded measurable functions f on X

$$\int f dP_2 = U(\int f dP_1)U^*.$$

Prove this first when f is a characteristic function, then a simple function, and finally the general case.

14. Proposition. Let μ be a measure on X and P^μ the associated canonical pvm in $L_2(\mu)$. Then for every bounded measurable function φ the integral $\int \varphi dP^\mu$ is the multiplication operator M_φ .

Proof. It is to be proved that for all $f \in L_2(\mu)$

$$\left(\int \varphi \, dP^\mu \right) f = M_\varphi f. \quad (25.2)$$

When φ is equal to a characteristic function χ_E , then

$$\int \varphi \, dP^\mu = P^\mu(E)$$

by the definition of the integral, and

$$P^\mu(E)f = \chi_E f = M_\varphi f$$

by the definition of P^μ . Thus (25.2) is true when φ is a characteristic function. Therefore it is true for simple functions (by linearity) and for bounded measurable functions (by continuity). ■

Corollary. Let μ be the Lebesgue measure on $X = [a, b]$ and P^μ the associated canonical pvm. Let $\varphi(t) = t$ for all t in X . Then the operator $\int \varphi \, dP^\mu$ is the canonical multiplication operator in $L_2(\mu)$. In other words

$$\left[\left(\int t \, dP^\mu(t) \right) f \right] (s) = sf(s) \text{ a.e. } (\mu).$$

A similar assertion can be made for a family of measures $\{\mu_n\}$. The operator $\int t \, dP^\mu(t)$ on the space $\oplus_n L_2(\mu_n)$ acts as

$$\left[\left(\int t \, dP^\mu(t) \right) \mathbf{f} \right] (s) = s\mathbf{f}(s).$$

We now have all the machinery to prove another form of the spectral theorem.

15. The Spectral Theorem (integral form). Let A be a self-adjoint operator on $\mathcal{H}$. Then there exists a unique pvm on the interval $X = [-\|A\|, \|A\|]$ with values in $\mathcal{P}(\mathcal{H})$ such that

$$A = \int_X \lambda \, dP(\lambda). \quad (25.3)$$

Proof. Recall the multiplication operator form of the spectral theorem. This says that there exist a sequence of probability measures $\{\mu_n\}$ on X and a unitary operator U from $\mathcal{H}$ onto the space $\mathcal{H}_0 = \oplus L_2(\mu_n)$ such that $UAU^* = M$, the canonical multiplication operator on $\mathcal{H}_0$. By the Corollary in Section 14 $M = \int t \, dP_0(t)$, where P_0 is the canonical pvm on X in $\mathcal{H}_0$. In other words,

$$UAU^* = \int t \, dP_0(t).$$

Let P be the pvm on X in $\mathcal{H}$ defined by

$$P(\cdot) = U^*P_0(\cdot)U.$$

Then, by Exercise 13, we see that the representation (25.3) is valid. (We have used the variable λ here to show the theorem as a generalisation of the finite-dimensional expression (24.2).)

It remains to be shown that the pvm P occurring in (25.3) is unique. Suppose Q is another pvm on X such that

$$A = \int_X \lambda \, dQ(\lambda).$$

By the Property (ii) of Exercise 7 and Exercise 9 we have then

$$\int_X \lambda^n dP(\lambda) = \int_X \lambda^n dQ(\lambda), \quad n = 0, 1, 2, \dots$$

Hence for all unit vectors v ,

$$\int_X \lambda^n d\langle P(\lambda)v, v \rangle = \int_X \lambda^n d\langle Q(\lambda)v, v \rangle. \quad (25.4)$$

Now the integrals involved are with respect to ordinary probability measures. The equality (25.4) shows that

$$\langle P(\cdot)v, v \rangle = \langle Q(\cdot)v, v \rangle \quad \text{for all } v.$$

Hence $P(\cdot) = Q(\cdot)$. ■

16. Exercise. Let $P(\cdot)$ be the pvm associated with A via (25.3). Then the family $P(\cdot)$ commutes with A . [Let f be a characteristic function χ_F . Then $\int f(\lambda)dP(\lambda) = P(F)$ and this commutes with all $P(E)$. Extend this to all f by the familiar routine.]

Support of the pvm

Let P be a pvm on a Hausdorff topological space with its Borel σ -algebra. Let E be the union of all open sets G in X for which $P(G) = 0$. The set $X \setminus E$ is called the *support of P* and is written as $\text{supp } P$.

17. Proposition. Let P be the pvm associated with a self-adjoint operator A via the spectral theorem. Then

$$\text{supp } P = \sigma(A). \quad (25.5)$$

Proof. Suppose $\lambda \notin \text{supp } P$. Then there exists $\varepsilon > 0$ such that $P(\lambda - \varepsilon, \lambda + \varepsilon) = 0$. Let v be any unit vector and μ_v the measure defined by (25.1). Then μ_v is concentrated on the complement of the interval $(\lambda - \varepsilon, \lambda + \varepsilon)$. Hence $|t - \lambda| \geq \varepsilon$ almost everywhere with respect to μ_v . Since

$$\|(A - \lambda)v\|^2 = \int_X |t - \lambda|^2 d\mu_v(t),$$

this shows that $\|(A - \lambda)v\|^2 \geq \varepsilon^2$. This shows that $A - \lambda$ is bounded below by ε . So λ cannot be an approximate eigenvalue of A , and hence cannot be in $\sigma(A)$.

Now suppose $\lambda \in \text{supp } P$. Then for every positive integer n , the projection $P(\lambda - \frac{1}{n}, \lambda + \frac{1}{n}) \neq 0$. Let v_n be any unit vector in the range of this projection. Then for any set E contained in the complement of the interval $(\lambda - \frac{1}{n}, \lambda + \frac{1}{n})$ we have $\mu_{v_n}(E) = 0$. Hence

$$\|(A - \lambda)v_n\|^2 = \int_X |t - \lambda|^2 d\mu_{v_n}(t) = \int_{\lambda - 1/n}^{\lambda + 1/n} |t - \lambda|^2 d\mu_{v_n}(t) \leq \frac{1}{n^2}.$$

Thus $\{v_n\}$ is a sequence of approximate eigenvectors of A , and hence $\lambda \in \sigma(A)$. ■

18. Exercise. Show that λ is an eigenvalue of A if and only if the point λ is an atom of the measure P ; i.e., the single-point set $\{\lambda\}$ has nonzero measure $P(\{\lambda\})$. It follows that every isolated point of $\sigma(A)$ is an eigenvalue of A .

The Spectral Theorem -III

This lecture is a quick review of some matters related to the spectral theorem.

The spectral measures $\{\mu_n\}$ of Lecture 24 and the projection-valued measure P of Lecture 25 associated with a self-adjoint operator A have as their support the spectrum $\sigma(A)$. This set is contained in $[-\|A\|, \|A\|]$. A smaller interval that contains $\sigma(A)$ is the *numerical range* of A defined as

$$W(A) = \{\langle Ax, x \rangle : \|x\| = 1\}.$$

1. **Proposition.** Let A be a self-adjoint operator and let

$$a = \min_{\|x\|=1} \langle Ax, x \rangle, \quad b = \max_{\|x\|=1} \langle Ax, x \rangle.$$

Then the spectrum of A is contained in the interval $[a, b]$ and contains the points a and b .

Proof. It is enough to prove the proposition in the special case when $a = 0$; i.e. when the operator A is positive. (Consider the operator $A - a$ instead of A .) In this case for every real number λ we have

$$\langle (A - \lambda)x, x \rangle \geq -\lambda\|x\|^2.$$

So if $\lambda < 0$, then $A - \lambda$ is bounded below and hence invertible. Thus $\sigma(A)$ does not contain any negative number. Since $a = 0$, the operator A is not invertible. Hence

$\sigma(A)$ contains the point a . We know also that

$$\text{spr}(A) = \|A\| = \max_{\|x\|=1} \langle Ax, x \rangle.$$

So $\sigma(A)$ is contained in $[a, b]$. Since $\sigma(A)$ is a closed set it contains the point b . ■

Functions of A

The spectral theorem makes it easy to define a *function* $f(A)$ of the operator A corresponding to every bounded measurable function f defined on $\sigma(A)$.

Let A be a self-adjoint operator with representation

$$A = \int_{\sigma(A)} \lambda \, dP(\lambda) \quad (26.1)$$

given to us by the spectral theorem. Let f be any bounded measurable function on $\sigma(A)$. Then we define $f(A)$ as

$$f(A) = \int_{\sigma(A)} f(\lambda) \, dP(\lambda). \quad (26.2)$$

We could also have used the first form of the spectral theorem. If A is equivalent to the multiplication operator M_φ , then $f(A)$ is equivalent to the multiplication operator $M_{f \circ \varphi}$.

If A is a positive operator, $\sigma(A)$ is contained in $[0, \infty)$. Every point of this set has a unique positive square root. So, we get from the prescription (26.2) a unique positive operator $A^{1/2}$, the square root of A . In the other picture, the function φ representing A takes only nonnegative values. The operator $A^{1/2}$ is then the multiplication operator corresponding to the function $\varphi^{1/2}$.

Operators commuting with A

Let A be a self-adjoint operator and let $P(\cdot)$ be the pvm associated with it. Suppose B is any operator that commutes with $P(E)$ for all measurable sets E . Then

B commutes with $\int f dP$ for all bounded measurable functions f . (Prove this first for characteristic functions, then for simple functions, and then for all f .) Conversely, suppose B commutes with A . Then B commutes with all powers A^n . Let x and y be any two vectors. Since $A^n = \int \lambda^n dP(\lambda)$, we have

$$\int \lambda^n d\langle P(\lambda)x, B^*y \rangle = \langle A^n x, B^*y \rangle = \langle BA^n x, y \rangle = \langle A^n Bx, y \rangle = \int \lambda^n d\langle P(\lambda)Bx, y \rangle.$$

Since this is true for all n , we must have

$$\langle P(E)x, B^*y \rangle = \langle P(E)Bx, y \rangle,$$

$$\text{i.e.,} \quad \langle BP(E)x, y \rangle = \langle P(E)Bx, y \rangle,$$

for every measurable set E . This is true for all x, y . Hence $BP(E) = P(E)B$ for all E .

The functional calculus

The spectral theorem is often stated as the “existence of a functional calculus”. This means the following statements, all of which may be derived from what we have proved.

Let A be a bounded self-adjoint operator on $\mathcal{H}$ and let $X = [-\|A\|, \|A\|]$. Then there exists a unique homomorphism φ of the algebra $L_\infty(X)$ into the algebra $\mathcal{B}(\mathcal{H})$ that satisfies the following properties:

1. $\varphi(\mathbf{1}) = I$, i.e. φ is unital.
2. If g is the “identity function” $g(x) = x$, then $\varphi(g) = A$.
3. If f_n is a uniformly bounded sequence of functions and f_n converge pointwise to f , then the operators $\varphi(f_n)$ converges strongly to $\varphi(f)$.
4. $\varphi(\bar{f}) = \varphi(f)^*$.
5. $\|\varphi(f)\| \leq \|f\|_\infty$.
6. If B is an operator that commutes with A , then $\varphi(f)$ commutes with B for all f .

The essential and the discrete spectrum

In Proposition 17 of Lecture 25 we have seen that a point λ is in the spectrum of a self-adjoint operator A if and only if the projection $P(\lambda - \varepsilon, \lambda + \varepsilon)$ is not zero for every $\varepsilon > 0$. This leads to a subdivision of the spectrum that is useful.

The *essential spectrum* $\sigma_{\text{ess}}(A)$ consists of those points λ for which the range of the projection $P(\lambda - \varepsilon, \lambda + \varepsilon)$ is infinite-dimensional for every $\varepsilon > 0$. If for some $\varepsilon > 0$, this range is finite-dimensional we say that λ is in $\sigma_{\text{disc}}(A)$, the *discrete spectrum* of A . Thus the spectrum $\sigma(A)$ is decomposed into two disjoint parts, the essential and the discrete spectrum.

2. Exercise. Let A be any self-adjoint operator. Prove the following statements:

- (i) $\sigma_{\text{ess}}(A)$ is a closed subset of $\mathbb{R}$.
- (ii) $\sigma_{\text{disc}}(A)$ is not always a closed set. (e.g. in the case of a compact operator for which 0 is not in the spectrum but is a limit point of the spectrum.)
- (iii) A point λ is in the set $\sigma_{\text{disc}}(A)$ if and only if λ is an isolated point of $\sigma(A)$ and is an eigenvalue of finite multiplicity. Thus λ is in $\sigma_{\text{ess}}(A)$ if it is either an eigenvalue of infinite multiplicity or is a limit point of $\sigma(A)$.

There is another characterisation of the essential spectrum in terms of approximate eigenvectors. By Theorem 1 in Lecture 18 every point λ in $\sigma(A)$ is an approximate eigenvalue; i.e. there exists a sequence of unit vectors $\{x_n\}$ such that $(A - \lambda)x_n$ converges to 0. A point in $\sigma_{\text{ess}}(A)$ has to meet a more stringent requirement:

3. Proposition. A point λ is in the essential spectrum of a self-adjoint operator A if and only if there exists an infinite sequence of orthonormal vectors $\{x_n\}$ such that $(A - \lambda)x_n$ converges to 0.

Proof. If $\lambda \in \sigma_{\text{ess}}(A)$, then for every n the space $\text{ran } P\left(\lambda - \frac{1}{n}, \lambda + \frac{1}{n}\right)$ is infinite-dimensional. Choose an orthonormal sequence $\{x_{nk} : k = 1, 2, \dots\}$ in this space. Then

$$\|(A - \lambda)x_{nk}\|^2 \leq \frac{1}{n^2} \quad \text{for all } k.$$

(See the proof of Proposition 17 in Lecture 25.) By the diagonal procedure we may pick up a sequence $\{x_n\}$ such that $\|(A - \lambda)x_n\|^2 \leq 1/n^2$ for $n = 1, 2, \dots$.

If $\lambda \in \sigma_{\text{disc}}(A)$, then for some $\varepsilon > 0$ the space $\text{ran } P(\lambda - \varepsilon, \lambda + \varepsilon)$ is finite-dimensional. So, if $\{x_n\}$ is any orthonormal sequence, then this space can contain only finitely many terms of this sequence, say $x_1, x_2, \dots, x_N$. For $n > N$ we have, therefore, $\|(A - \lambda)x_n\|^2 \geq \varepsilon^2$. Thus $(A - \lambda)x_n$ cannot converge to 0. ■

In the finite-dimensional case the spectrum of every operator consists of a finite number of eigenvalues. So, in the infinite-dimensional case we may think of the discrete spectrum as an object familiar to us from linear algebra. The essential spectrum is not so familiar. If A is a compact operator, then 0 is the only point it may have in its essential spectrum. But, in general, a self-adjoint operator A can have a large essential spectrum. Think of an example where $\sigma(A) = \sigma_{\text{ess}}(A)$.

The following theorem says that adding a compact operator to a bounded self-adjoint operator does not change its essential spectrum.

4. Weyl's Perturbation Theorem. Let A and B be self-adjoint operators in $\mathcal{H}$. If $A - B$ is compact, then $\sigma_{\text{ess}}(A) = \sigma_{\text{ess}}(B)$.

Proof. Let $\lambda \in \sigma_{\text{ess}}(A)$. By Proposition 3 there exists an infinite sequence of orthonormal vectors $\{x_n\}$ such that $(A - \lambda)x_n$ converges to 0. If y is any vector in $\mathcal{H}$, then $\langle x_n, y \rangle$ converges to zero as $n \rightarrow \infty$. (Consider first the two special cases when y is in the space spanned by $\{x_n\}$ and when it is in the orthogonal complement of this space.) In other words $x_n \xrightarrow{w} 0$. Since $A - B$ is compact, $(A - B)x_n \rightarrow 0$.

(Theorem 10, Lecture 20.) Since

$$\|(B - \lambda)x_n\| \leq \|(A - \lambda)x_n\| + \|(B - A)x_n\|,$$

this shows that $(B - \lambda)x_n \rightarrow 0$, and hence $\lambda \in \sigma_{\text{ess}}(B)$. Thus $\sigma_{\text{ess}}(A) \subset \sigma_{\text{ess}}(B)$.

By symmetry the reverse inclusion is also true. ■

One may note here that the spectral theorem for a compact self-adjoint operator follows from this. (Choose $B = 0$.)

This theorem is important in applications where a compact operator is considered “small” compared to a noncompact operator. The theorem says that the essential spectrum is unaffected by such “small changes”.

Spectral Theorem for normal operators

If $\{A_m\}$ is a family of pairwise commuting self-adjoint operators on a finite-dimensional Hilbert space, then there exists a unitary operator U such that all the operators UA_mU^* are diagonal. This has an infinite-dimensional analogue that we state without proof.

5. Theorem. Let $A_1, A_2, \dots, A_k$ be pairwise commuting self-adjoint operators on $\mathcal{H}$. Then there exists a projection valued measure on the product space $X = \prod_{j=1}^k [-\|A_j\|, \|A_j\|]$ with values in $\mathcal{P}(\mathcal{H})$ such that each operator A_j has the representation

$$A_j = \int_X \lambda_j \, dP(\lambda_1, \dots, \lambda_k).$$

A consequence of this is the spectral theorem for normal operators. If A is normal, then we have $A = A_1 + iA_2$ where A_1 and A_2 are commuting self-adjoint operators. We get from Theorem 5, the following.

6. Theorem. Let A be a normal operator on $\mathcal{H}$. Then there exists a pvm P on $\mathbb{C}$

with values in $\mathcal{P}(\mathcal{H})$ such that

$$A = \int z \, dP(z). \quad (26.3)$$

The support of P is the spectrum of A .

The multiplication operator form of this theorem says that A is unitarily equivalent to an operator of the form M_φ in some space $L_2(\mu)$.

Spectral Theorem for unitary operators

Unitary operators constitute an important special class of normal operators. A proof of the spectral theorem for this class is outlined below. The ideas are similar to the ones used in Lectures 24 and 25.

Let U be a unitary operator. Then $\sigma(U)$ is contained in the unit circle. We may identify the unit circle with the interval $[-\pi, \pi]$ as usual.

Let x be any vector in $\mathcal{H}$ and for $n \in \mathbb{Z}$, let

$$a_n = \langle U^n x, x \rangle.$$

Then for any sequence of complex numbers $z_1, z_2, \dots$, we have

$$\begin{aligned} \sum_{j,k} a_{j-k} z_j \bar{z}_k &= \sum_{j,k} \langle U^{j-k} x, x \rangle z_j \bar{z}_k \\ &= \sum_{j,k} \langle U^j x, U^k x \rangle z_j \bar{z}_k \\ &= \left\| \sum_j z_j U^j x \right\|^2 \geq 0. \end{aligned}$$

Thus the sequence $\{a_n\}$ is a positive-definite sequence. By the Herglotz Theorem (Lecture 8) there exists a positive measure μ_x on $[-\pi, \pi]$ such that

$$\langle U^n x, x \rangle = \int_{-\pi}^{\pi} e^{int} d\mu_x(t). \quad (26.4)$$

Using the polarisation identity we can express $\langle U^n x, y \rangle$ for any pair of vectors x, y as a sum of four such terms. This leads to the relation

$$\langle U^n x, y \rangle = \int_{-\pi}^{\pi} e^{int} d\mu_{x,y}(t), \quad (26.5)$$

where $\mu_{x,y}$ is the complex measure given by

$$\mu_{x,y} = \frac{1}{4} (\mu_{x+y} - \mu_{x-y} + i\mu_{x+iy} - i\mu_{x-iy}).$$

7. Exercise. The measures $\mu_{x,y}$ satisfy the following properties

- (i) Each $\mu_{x,y}$ is linear in x and conjugate linear in y .
- (ii) $\mu_{x,y} = \bar{\mu}_{y,x}$.
- (iii) The total mass of $\mu_{x,y}$ is bounded by $\|x\| \|y\|$.

For any measurable set E of $[-\pi, \pi]$ let

$$\langle P(E)x, y \rangle = \mu_{x,y}(E). \quad (26.6)$$

From the properties in Exercise 7 it follows that $P(E)$ is self-adjoint and countably additive. To prove that it is a pvm we need to show that $P(E)^2 = P(E)$ for all E . We prove a stronger statement.

8. Proposition. The operator function $P(\cdot)$ defined by (26.6) satisfies the relation

$$P(E \cap F) = P(E)P(F) \text{ for all } E, F. \quad (26.7)$$

Proof. Let n, k be any two integers. Then

$$\langle U^{n+k}x, y \rangle = \langle U^n U^k x, y \rangle.$$

So from (26.5) and (26.6)

$$\int_{-\pi}^{\pi} e^{int} e^{ikt} d\langle P(t)x, y \rangle = \int_{-\pi}^{\pi} e^{int} d\langle P(t)U^k x, y \rangle.$$

This is true for all n . Hence

$$e^{ikt} d\langle P(t)x, y \rangle = d\langle P(t)U^k x, y \rangle. \quad (26.8)$$

(If $\int e^{int} d\mu(t) = \int e^{int} d\nu(t)$ for all n , then the measures μ and ν on $[-\pi, \pi]$ are equal.)

Integrate the two sides of (26.8) over the set E . This gives

$$\begin{aligned} \int_{-\pi}^{\pi} \chi_E(t) e^{ikt} d\langle P(t)x, y \rangle &= \langle P(E)U^k x, y \rangle \\ &= \langle U^k x, P(E)y \rangle \quad (\text{since } P(E) \text{ is self-adjoint}) \\ &= \int_{-\pi}^{\pi} e^{ikt} d\langle P(t)x, P(E)y \rangle \quad (\text{from (26.5) and (26.6)}). \end{aligned}$$

This is true for all k . Hence,

$$\chi_E(t) d\langle P(t)x, y \rangle = d\langle P(t)x, P(E)y \rangle.$$

Integrate the two sides over the set F . This gives

$$\int_{-\pi}^{\pi} \chi_F(t) \chi_E(t) d\langle P(t)x, y \rangle = \langle P(F)x, P(E)y \rangle.$$

Since $\chi_F \chi_E = \chi_{E \cap F}$, this shows that

$$\langle P(E \cap F)x, y \rangle = \langle P(F)x, P(E)y \rangle = \langle P(E)P(F)x, y \rangle.$$

This is true for all x and y . Hence we have the assertion (26.7). ■

Thus $P(\cdot)$ is a pvm on the unit circle (identified with $[-\pi, \pi]$). The relations (26.5) and (26.6) show that

$$\langle U^n x, y \rangle = \int_{-\pi}^{\pi} e^{int} d\langle P(t)x, y \rangle \quad \text{for all } x, y.$$

This shows that the operator U may be represented as

$$U = \int_{-\pi}^{\pi} e^{it} dP(t), \tag{26.9}$$

where P is a pvm on the unit circle. The integral exists in the norm topology; the proof given for self-adjoint operators in Lecture 25 works here too.

9. Exercise (von Neumann's ergodic theorem). A proof of this theorem, also called the L_2 ergodic theorem or the mean ergodic theorem, is outlined in this exercise. Fill in the details.

Let $(X, \mathcal{S}, \mu)$ be a measure space. A bijection T of X such that T and T^{-1} are measurable is called an *automorphism* of $(X, \mathcal{S})$. If $\mu T^{-1}(E) = \mu(E)$ for all $E \in \mathcal{S}$, then T is called a *measure-preserving* map.

Let T be a measure-preserving automorphism. The operator U on $L_2(\mu)$ defined as $(Uf)(x) = f(Tx)$ is called the *Koopman operator* associated with T . Show that U is a unitary operator.

Use the representation (26.9) to show that

$$\frac{f + fT + \cdots + fT^{n-1}}{n} = \frac{(I + U + \cdots + U^{n-1})f}{n} = \left(\int_{-\pi}^{\pi} \frac{1 - e^{int}}{n(1 - e^{it})} dP(t) \right) f.$$

The integrand is interpreted to be equal to 1 at $t = 0$. As n goes to ∞ , the integrand converges to the characteristic function of the set $\{1\}$. So, by the Dominated Convergence Theorem, the integral converges to $P(\{1\})$. This is the projection onto the set $\{f : Uf = f\}$. Another description of this set is $\{f : fT = f\}$. Elements of this set are called T -invariant functions. The mean ergodic theorem is the statement

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} fT^j = P_0 f \quad \text{for all } f \in L_2(\mu),$$

where P_0 is the projection onto the subspace consisting of T -invariant functions.

10. Exercise. The aim of this exercise is to show that the set of compact operators $\mathcal{B}_0(\mathcal{H})$ is the only closed 2-sided (proper) ideal in $\mathcal{B}(\mathcal{H})$. Fill in the details.

- (i) Let $\mathcal{I}$ be any 2-sided ideal in $\mathcal{B}(\mathcal{H})$. Let $T \in \mathcal{I}$ and let u, v be any two vectors such that $Tu = v$. Let A be any rank-one operator. Then there exist vectors x and y such that $A = \langle \cdot, x \rangle y$. Let $B = \langle \cdot, x \rangle u$ and let C be any operator such that $Cv = y$. Show that $A = CTB$. Thus $\mathcal{I}$ contains all rank-one operators, and hence it contains all operators of finite rank.
- (ii) Suppose $\mathcal{I}$ contains a positive operator A that is not compact. Then there exists an $\varepsilon > 0$ such that the range of the projection $P(\varepsilon, \infty)$ is infinite-dimensional.

(Here P is the pvm associated with A .) Let $\mathcal{M}$ be this range and let V be a unitary operator from $\mathcal{H}$ onto $\mathcal{M}$. Since $A(\mathcal{M}) = \mathcal{M}$, we have

$$V^*AV(\mathcal{H}) = V^*A(\mathcal{M}) = V^*(\mathcal{M}) = \mathcal{H}.$$

Show that for every $x \in \mathcal{H}$ we have

$$\|V^*AVx\| \geq \varepsilon\|x\|.$$

Thus V^*AV is invertible. Since $V^*AV \in \mathcal{I}$, this means that $\mathcal{I} = \mathcal{B}(\mathcal{H})$.

- (iii) Thus if $\mathcal{I}$ is any proper 2-sided ideal in $\mathcal{B}(\mathcal{H})$ then every element of $\mathcal{I}$ is a compact operator and every finite-rank operator is in $\mathcal{I}$. Since $\mathcal{B}_0(\mathcal{H})$ is the norm closure of finite-rank operators, if $\mathcal{I}$ is closed, then it is equal to $\mathcal{B}_0(\mathcal{H})$.

Index

- $A^{1/2}$, 155
- $A^\dagger$, 113
- A^* , 111
- $A_\alpha \xrightarrow{s} A$, 103
- $A_\alpha \xrightarrow{w} A$, 104
- $BV[0, 1]$, 53
- $C(X)$, 3
- $C[0, 1]$, 3
- $C^r[0, 1]$, 4
- $L_p(X, \mathcal{S}, \mu)$, 7
- $L_p[0, 1]$, 7
- $L_\infty[0, 1]$, 7
- $R_\lambda(A)$, 132
- $S^\perp$, 76
- $S^\perp$, 85
- $W(A)$, 219
- X/M , 19
- X^{**} , 73
- X^* , 25
- $[S]$, 77
- $\mathcal{B}(X, Y)$, 21
- $\mathcal{B}(X)$, 23
- $\mathcal{H}$, 83
- $\dim X$, 13
- ℓ_p^n , 2
- ℓ_∞ , 5
- ℓ_p , 5
- $\int f dP$, 214
- ∞ -norm, 2
- $\langle x, y \rangle$, 82
- codim, 77
- ess ran φ , 149
- ind A , 177
- ker, 87
- ker A , 158
- ran, 87
- ran A , 158
- spr (A) , 135
- supp P , 218
- supp μ , 206
- tran φ , 149
- tr A , 190
- $\mu_v(E)$, 211
- $\mu_{u,v}(E)$, 212
- $\rho(A)$, 132
- $\sigma(A)$, 134
- $\sigma_p(A)$, 139
- $\sigma_{\text{app}}(A)$, 140
- $\sigma_{\text{comp}}(A)$, 140
- $\sigma_{\text{disc}}(A)$, 222

- $\sigma_{\text{ess}}(A)$, 222
- $\sigma_{\text{res}}(A)$, 141
- $\varepsilon/3$ argument, 4
- c , 5
- c_{00} , 5
- p -norm, 2
- s_n , 185
- $s_n(A)$, 187
- $x \perp y$, 84
- $x_n \xrightarrow{w} x$, 67
- $\mathcal{B}_0(X, Y)$, 164
- $\mathcal{B}_{00}(X, Y)$, 164
- $\mathcal{C}_1$, 189, 191
- $\mathcal{C}_2$, 195
- $\mathcal{C}_p$, 196
- $\mathcal{P}(\mathcal{H})$, 209
- absolutely continuous, 9
- absolutely summable sequence, 20
- adjoint, 111
 - of a matrix, 116
 - of an integral operator, 116
 - of Hilbert space operator, 113
- algebra, 24
- algebraic dimension, 46
- algebraic dual, 25
- analyticity
 - strong, 131
 - weak, 131
- annihilator, 77
- Appolonius Theorem, 85
- approximate eigenvalues, 140
- approximate point spectrum, 140
- arithmetic-geometric mean inequality, 2
- automorphism, 124
- backward shift, 150
- Baire Category Theorem, 36
- Banach–Alaoglu Theorem, 74
- Banach–Steinhaus Theorem, 36
- Banach algebra, 24
- Banach limit, 34
- Banach space, 1
- basis
 - algebraic, 11
 - Hamel, 11
 - Schauder, 13
 - topological, 13
- Bessel’s inequality, 93
- bidual, 73
- Bolzano–Weierstrass Theorem, 72
- bounded below, 118, 139
- bounded linear functional, 22
- bounded linear operator, 21
- bounded variation, 53
- C^* -algebra, 115
- canonical multiplication operator, 199
- canonical pvm, 211
- Cartesian decomposition, 123
- Cauchy–Schwarz inequality, 3, 83

- Closed Graph Theorem, 44
- co-isometry, 125
- codimension, 77
- coker A , 176
- cokernel, 176
- commutant, 181
- compact operator, 163, 228
 - adjoint of, 167
 - invariant subspace, 181
 - product, 165
 - Riesz decomposition, 179
 - spectral theorem, 183
 - spectrum of, 172
- completely continuous, 166
- composition operators, 116
- compression spectrum, 140
- condensation of singularities, 39
- conjugate index, 2
- conjugate linear functional, 25
- continuity
 - of adjoint, 115
 - of inverse, 108
 - of operator multiplication, 106
 - strong, 129
 - weak, 129
- continuous spectrum, 141
- convergence, 67
 - strong, 67
 - weak, 67
- cyclic subspace, 200
- cyclic vector, 200
- diagonal operator, 147, 171
 - compact, 165
- differentiability
 - strong, 129
 - weak, 129
- dilation, 42
- dimension, 13
- directed set, 70
- direct sum decomposition, 87, 89
- direct summand, 88
- discrete spectrum, 222
- dual
 - of ℓ_p , 50
 - of ℓ_∞ , 51
 - of $C[0, 1]$, 52
 - of c_0 , 51
- dual space, 25, 33
- eigenvalue, 134, 139
- Enflo's example, 169, 186
- essentially bounded, 6
- essential range, 149
- essential spectrum, 222
- essential supremum, 6
- eventually, 70
- final space, 160
- finite-rank operator, 164

- first category, 40
- forward shift, 150
- Fourier-Stieltjes sequence, 59
- Fourier coefficients, 39
- Fourier kernel, 26
- Fourier series, 39, 96
- Fourier transform, 26
- Fredholm alternative, 177
- Fredholm operator, 177
- frequently, 71
- functional calculus, 221
- fundamental set, 76

- Gram–Schmidt Process, 95
- Gram determinant, 100
- Gram matrix, 100
- graph, 44

- Hölder inequality, 2, 6
- Hahn–Banach Theorem, 53, 68, 79
 - (H.B.T.), 28
 - for Hilbert spaces, 90
- Hausdorff distance, 152
- Helly’s Theorem, 200
- Herglotz Theorem, 60
- Hermite polynomials, 98
- Hermitian, 119
- Hilbert–Hankel operator, 128
- Hilbert–Schmidt norm, 195
- Hilbert–Schmidt operator, 195
- Hilbert space, 83
 - separable, 95
- hyperinvariant subspace, 181

- ideal
 - compact operators, 228
 - Schatten, 197
 - trace class operators, 194
- idempotent, 86
- index, 177
- initial space, 160
- inner product, 82
- inner product space, 81
- integral kernel operator, 23
- integral operator, 164
 - compactness, 164
- invariant subspace, 126, 181
- Invariant subspace problem, 186
- Inverse Mapping Theorem, 43
- isometric isomorphism, 47
- isometry, 124
- isomorphism
 - between Hilbert spaces, 96

- Laguerre polynomials, 99
- Laplace transform, 26
- Lebesgue Dominated Convergence Theorem, 214
- left shift, 107, 113, 139, 143, 150, 173
- Legendre polynomials, 98
- Lidskii’s Theorem, 195
- linear functional

- positive, 56
- unital, 57
- linear operator, 21
- locally compact, 17
- Lomonosov's Theorem, 181
- Müntz's Theorem, 101
- measure
 - absolutely continuous, 207
 - equivalent, 207
 - projection-valued, 209
 - support of, 206
- Minkowski inequality, 3
- Montel–Helly Selection Principle, 58, 75
- multiplication operator, 149
 - canonical, 199
 - compact, 185
- multiplicity, 172, 173
- nets, 70
- Neumann series, 109
- norm, 1
 - equivalent, 15, 16
 - induced by inner product, 83
- normal operator, 122
 - polar decomposition, 160
- normed algebra, 24
- normed linear space, 1
- normed vector space, 1
- norm topology, 103
- numerical range, 219
- open mapping theorem, 42
- operator
 - compact, 163, 167
 - completely continuous, 166, 167
 - function of, 220
 - Hermitian, 119
 - positive, 121
 - positive definite, 121
 - real and imaginary parts of, 123
 - self-adjoint, 119
 - unitary, 123
- orthogonal, 84
- orthogonal complement, 88
- orthogonal projection, 88, 125
- orthonormal basis, 93
- orthonormal set, 93
 - complete, 93
- orthoprojector, 88
- parallelogram law, 84
- Parseval's equality, 94
- partial isometry, 160
- partially ordered set, 12
- partial order, 11
- point spectrum, 139
- polar decomposition, 155, 158
- polarisation identity, 84
- positive operator
 - square root of, 155
- positive part, 155

- positive semidefinite, 121
- precompact, 163
- pre Hilbert space, 83
- probability measure, 57
- product topology, 66
- projection, 44, 88
- projection-valued measure, 209
 - canonical, 211
 - support of, 218
- pvm, 210
- Pythagorean Theorem, 84
- quadratic form, 92
- quotient, 19
- Rademacher functions, 99
- Radon-Nikodym derivative, 207
- reducing subspace, 126, 184
- reflexive, 73
- resolvent, 132
- resolvent identity, 133
- resolvent set, 132
- Riemann–Lebesgue Lemma, 67
- Riesz’s Lemma, 17
- Riesz–Fischer Theorem, 7
- Riesz–Herglotz integral representation, 62
- Riesz Decomposition Theorem, 179
- Riesz Projection, 180
- Riesz Representation Theorem, 55, 58, 64, 200
 - for Hilbert spaces, 90
- right shift, 104, 112, 135, 139, 143, 150, 160, 173
- Schatten spaces, 197
- Schauder basis, 14, 169
- Schwarz inequality, 83
- second dual, 73
- self-adjoint, 119
- separable, 8
- sequence
 - positive definite, 59
- sesquilinear form, 90
- shift
 - backward, 150
 - forward, 150
 - left, 150
 - right, 150
 - weighted, 150
- singular value decomposition, 160, 184
- singular values, 185, 187
 - continuity of, 188
 - of a product, 188
- Sobolev spaces, 9
- Spectral Mapping Theorem, 137
- spectral measure, 206
 - integration, 212
- spectral radius, 135
- spectral radius formula, 136
- spectral theorem, 155, 198

- for compact operators, 183
- for normal operators, 224
- for unitary operators, 225
- in finite dimensions, 198
- integral form, 216
- multiplication operator form, 199
- spectrum, 129, 134, 141
 - approximate point, 140
 - boundary of, 143
 - compression, 140
 - continuous, 141
 - discontinuity of, 152
 - of a diagonal operator, 148
 - of adjoint, 141
 - of a multiplication operator, 149
 - of a normal operator, 153
 - of normal operator, 146
 - of product, 145
 - of self-adjoint operator, 146
 - residual, 141
 - upper semicontinuity of, 153
- square integrable kernel, 22
- square root, 155
- strongly analytic, 131
- strongly differentiable, 130
- strong operator topology, 103
- sublinear functional, 28
- subnet, 71
- subspace
 - invariant, 126
 - reducing, 126
- summable family, 93
- summable sequence, 20
- support, 206
- thick range, 149
- topological dual, 25
- topology
 - norm, 67
 - of pointwise convergence, 66, 74
 - strong, 67
 - usual, 67
 - weak, 67
 - weak*, 74
- topology on operators, 103
 - norm, 103
 - strong, 103
 - uniform, 103
 - usual, 103
 - weak, 103
- totally ordered, 12
- trace, 190, 191, 194
- trace class operator, 189
- translation, 42
- triangle inequality, 1
- trigonometric polynomial, 63
- two-sided ideal, 166
- Tychonoff Theorem, 72, 74
- Uniform Boundedness Principle, 68, 105

(U.B.P.), 36

von Neumann's Ergodic Theorem, 227

Walsh functions, 99

weak* compact, 58

weak* continuous, 76

weak* topology, 74

weakly analytic, 131

weakly differentiable, 130

weak operator topology, 103

weak topology, 66, 74, 79

 metrisability of unit ball, 97

 not metrisable, 69

weighted shift, 150

weight sequence, 151

Weyl's Perturbation Theorem, 223

Young's inequality, 2

Zorn's Lemma, 12, 29, 30

Texts and Readings in Mathematics

1. R. B. Bapat: Linear Algebra and Linear Models (Second Edition)
2. Rajendra Bhatia: Fourier Series (Second Edition)
3. C. Musili: Representations of Finite Groups
4. H. Helson: Linear Algebra (Second Edition)
5. D. Sarason: Complex Function Theory (Second Edition)
6. M. G. Nadkarni: Basic Ergodic Theory (Second Edition)
7. H. Helson: Harmonic Analysis (Second Edition)
8. K. Chandrasekharan: A Course on Integration Theory
9. K. Chandrasekharan: A Course on Topological Groups
10. R. Bhatia (ed.): Analysis, Geometry and Probability
11. K. R. Davidson: C^* – Algebras by Example
12. M. Bhattacharjee et al.: Notes on Infinite Permutation Groups
13. V. S. Sunder: Functional Analysis — Spectral Theory
14. V. S. Varadarajan: Algebra in Ancient and Modern Times
15. M. G. Nadkarni: Spectral Theory of Dynamical Systems
16. A. Borel: Semisimple Groups and Riemannian Symmetric Spaces
17. M. Marcolli: Seiberg – Witten Gauge Theory
18. A. Bottcher and S. M. Grudsky: Toeplitz Matrices, Asymptotic
Linear Algebra and Functional Analysis
19. A. R. Rao and P. Bhimasankaram: Linear Algebra (Second Edition)
20. C. Musili: Algebraic Geometry for Beginners
21. A. R. Rajwade: Convex Polyhedra with Regularity Conditions
and Hilbert's Third Problem
22. S. Kumaresan: A Course in Differential Geometry and Lie Groups
23. Stef Tijs: Introduction to Game Theory
24. B. Sury: The Congruence Subgroup Problem
25. R. Bhatia (ed.): Connected at Infinity
26. K. Mukherjee: Differential Calculus in Normed Linear Spaces
(Second Edition)
27. Satya Deo: Algebraic Topology: A Primer (Corrected Reprint)
28. S. Kesavan: Nonlinear Functional Analysis: A First Course
29. S. Szabó: Topics in Factorization of Abelian Groups
30. S. Kumaresan and G. Santhanam: An Expedition to Geometry
31. D. Mumford: Lectures on Curves on an Algebraic Surface (Reprint)
32. J. W. Milnor and J. D. Stasheff: Characteristic Classes (Reprint)
33. K. R. Parthasarathy: Introduction to Probability and Measure
(Corrected Reprint)
34. A. Mukherjee: Topics in Differential Topology
35. K. R. Parthasarathy: Mathematical Foundations of Quantum
Mechanics
36. K. B. Athreya and S. N. Lahiri: Measure Theory
37. Terence Tao: Analysis I
38. Terence Tao: Analysis II

- 39. W. Decker and C. Lossen: Computing in Algebraic Geometry
- 40. A. Goswami and B. V. Rao: A Course in Applied Stochastic Processes
- 41. K. B. Athreya and S. N. Lahiri: Probability Theory
- 42. A. R. Rajwade and A. K. Bhandari: Surprises and Counterexamples in Real Function Theory
- 43. G. H. Golub and C. F. Van Loan: Matrix Computations (Reprint of the Third Edition)
- 44. Rajendra Bhatia: Positive Definite Matrices
- 45. K. R. Parthasarathy: Coding Theorems of Classical and Quantum Information Theory
- 46. C. S. Seshadri: Introduction to the Theory of Standard Monomials
- 47. Alain Connes and Matilde Marcolli: Noncommutative Geometry, Quantum Fields and Motives
- 48. Vivek S. Borkar: Stochastic Approximation: A Dynamical Systems Viewpoint
- 49. B. J. Venkatachala: Inequalities: An Approach Through Problems