CONTENTS

Foreword xvi

1. Linear Spaces 1
   Axioms for linear spaces—Infinite-dimensional examples—
   Subspace, linear span—Quotient space—Isomorphism—Convex
   sets—Extreme subsets

2. Linear Maps 8
   2.1 Algebra of linear maps, 8
   Axioms for linear maps—Sums and composites—Invertible
   linear maps—Nullspace and range—Invariant subspaces
   2.2 Index of a linear map, 12
   Degenerate maps—Pseudoinverse—Index—Product formula for
   the index—Stability of the index

3. The Hahn-Banach Theorem 19
   3.1 The extension theorem, 19
   Positive homogeneous, subadditive functionals—Extension of
   linear functionals—Gauge functions of convex sets
   3.2 Geometric Hahn-Banach theorem, 21
   The hyperplane separation theorem
   3.3 Extensions of the Hahn-Banach theorem, 24
   The Agnew-Morse theorem—The
   Bohnenblust-Sobczyk-Souhilmov theorem

4. Applications of the Hahn-Banach theorem 29
   4.1 Extension of positive linear functionals, 29
   4.2 Banach limits, 31
4.3 Finitely additive invariant set functions, 33
Historical note, 34

5. Normed Linear Spaces 36
5.1 Norms, 36
Norms for quotient spaces—Complete normed linear spaces—
The spaces $C, B—L^p$ spaces and Hölder's inequality—Sobolev
spaces, embedding theorems—Separable spaces
5.2 Noncompactness of the unit ball, 43
Uniform convexity—The Mazur-Ulam theorem on isometries
5.3 Isometries, 47

6. Hilbert Space 52
6.1 Scalar product, 52
Schwarz inequality—Parallelogram identity—Completeness,
closure—$l^2, L^2$
6.2 Closest point in a closed convex subset, 54
Orthogonal complement of a subspace—Orthogonal
decomposition
6.3 Linear functionals, 56
The Riesz-Fréchet representation theorem—Lax-Milgram lemma
6.4 Linear span, 58
Orthogonal projection—Orthonormal bases, Gram-Schmidt
process—Isometries of a Hilbert space

7. Applications of Hilbert Space Results 63
7.1 Radon-Nikodym theorem, 63
7.2 Dirichlet's problem, 65
Use of the Riesz-Fréchet theorem—Use of the Lax-Milgram
theorem—Use of orthogonal decomposition

8. Duals of Normed Linear Spaces 72
8.1 Bounded linear functionals, 72
Dual space
8.2 Extension of bounded linear functionals, 74
Dual characterization of norm—Dual characterization of
distance from a subspace—Dual characterization of the closed
linear span of a set

8.3 Reflexive spaces, 78
 Reflexivity of $L^p, 1 < p < \infty$—Separable spaces—Separability
 of the dual—Dual of $C(Q), Q$ compact—Reflexivity of
 subspaces
8.4 Support function of a set, 83
Dual characterization of convex hull—Dual characterization of
distance from a closed, convex set

9. Applications of Duality 87
9.1 Completeness of weighted powers, 87
9.2 The Müntz approximation theorem, 88
9.3 Runge's theorem, 91
9.4 Dual variational problems in function theory, 91
9.5 Existence of Green's function, 94

10. Weak Convergence 99
10.1 Uniform boundedness of weakly convergent sequences, 101
 Principle of uniform boundedness—Weakly sequentially
closed convex sets
10.2 Weak sequential compactness, 104
Compactness of unit ball in reflexive space
10.3 Weak* convergence, 105
Helly's theorem

11. Applications of Weak Convergence 108
11.1 Approximation of the $\delta$ function by continuous functions, 108
Toeplitz's theorem on summability
11.2 Divergence of Fourier series, 109
11.3 Approximate quadrature, 110
11.4 Weak and strong analyticity of vector-valued functions, 111
11.5 Existence of solutions of partial differential equations, 112
Galerkin's method
11.6 The representation of analytic functions with positive real part, 115
Herglotz-Riesz theorem

12. The Weak and Weak* Topologies 118
Comparison with weak sequential topology—Closed convex sets
in the weak topology—Weak compactness—A Faàglu's theorem
13. Locally Convex Topologies and the Krein-Milman Theorem 122
   13.1 Separation of points by linear functionals, 123
   13.2 The Krein-Milman theorem, 124
   13.3 The Stone-Weierstrass theorem, 126
   13.4 Choquet’s theorem, 128

14. Examples of Convex Sets and Their Extreme Points 133
   14.1 Positive functionals, 133
   14.2 Convex functions, 135
   14.3 Completely monotone functions, 137
   14.4 Theorems of Carathéodory and Bochner, 141
   14.5 A theorem of Krein, 147
   14.6 Positive harmonic functions, 148
   14.7 The Hamburger moment problem, 150
   14.8 G. Birkhoff’s conjecture, 151
   14.9 De Finetti’s theorem, 156
   14.10 Measure-preserving mappings, 157
   Historical note, 159

15. Bounded Linear Maps 160
   15.1 Boundedness and continuity, 160
   Norm of a bounded linear map—Transpose
   15.2 Strong and weak topologies, 165
   Strong and weak sequential convergence
   15.3 Principle of uniform boundedness, 166
   15.4 Composition of bounded maps, 167
   15.5 The open mapping principle, 168
   Closed graph theorem
   Historical note, 172

16. Examples of Bounded Linear Maps 173
   16.1 Boundedness of integral operators, 173
   Integral operators of Hilbert-Schmidt type—Integral operators of
   Holmgren type
   16.2 The convexity theorem of Marcel Riesz, 177
   16.3 Examples of bounded integral operators, 180
   The Fourier transform, Parseval’s theorem and Hausdorff-Young
   inequality—The Hilbert transform—The Laplace transform—
   The Hilbert-Hankel transform

16.4 Solution operators for hyperbolic equations, 186
16.5 Solution operator for the heat equation, 188
16.6 Singular integral operators, pseudodifferential operators and
Fourier integral operators, 190

17. Banach Algebras and their Elementary Spectral Theory 192
   17.1 Normed algebras, 192
   Invertible elements—Resolvent set and spectrum—Resolvent—
   Spectral radius
   17.2 Functional calculus, 197
   Spectral mapping theorem—Projections

18. Gelfand’s Theory of Commutative Banach Algebras 202
   Homomorphisms into C—Maximal ideals—Mazur’s lemma—
The spectrum as the range of homomorphisms—The spectral
mapping theorem revisited—The Gelfand representation—
Gelfand topology

19. Applications of Gelfand’s Theory of Commutative Banach Algebras 210
   19.1 The algebra C(S), 210
   19.2 Gelfand compactification, 210
   19.3 Absolutely convergent Fourier series, 212
   19.4 Analytic functions in the closed unit disk, 213
   Analytic functions in the closed polydisk
   19.5 Analytic functions in the open unit disk, 214
   19.6 Wiener’s Tauberian theorem, 215
   19.7 Commutative B*-algebras, 221
   Historical note, 224

20. Examples of Operators and Their Spectra 226
   20.1 Invertible maps, 226
   Boundary points of the spectrum
   20.2 Shifts, 229
   20.3 Volterra integral operators, 230
   20.4 The Fourier transform, 231

21. Compact Maps 233
   21.1 Basic properties of compact maps, 233
   Compact maps form a two-sided ideal—Identity plus
   compact map has index zero
21.2 The spectral theory of compact maps, 238
   The transpose of a compact operator is compact—The Fredholm alternative
   Historical note, 244

22. Examples of Compact Operators
   22.1 Compactness criteria, 245
       Arzela-Ascoli compactness criterion—Rellich compactness criterion
   22.2 Integral operators, 246
       Hilbert-Schmidt operators
   22.3 The inverse of elliptic partial differential operators, 249
   22.4 Operators defined by parabolic equations, 250
   22.5 Almost orthogonal bases, 251

23. Positive compact operators
   23.1 The spectrum of compact positive operators, 253
   23.2 Stochastic integral operators, 256
       Invariant probability density
   23.3 Inverse of a second order elliptic operator, 258

24. Fredholm’s Theory of Integral Equations
   24.1 The Fredholm determinant and the Fredholm resolvent, 260
       The spectrum of Fredholm operators—A trace formula for Fredholm operators
   24.2 The multiplicative property of the Fredholm determinant, 268
   24.3 The Gelfand-Levitan-Marchenko equation and Dyson’s formula, 271

25. Invariant Subspaces
   25.1 Invariant subspaces of compact maps, 275
       The von Neumann-Aronszajn-Smith theorem
   25.2 Nested invariant subspaces, 277
       Ringrose’s theorem—Unicellular operators: the Brodsky-Donoghue theorem—The Robinson-Bernstein and Lomonosov theorems—Enflo’s example

26. Harmonic Analysis on a Halfline
   26.1 The Pragmén-Lindelöf principle for harmonic functions, 284

26.2 An abstract Pragmén-Lindelöf principle, 285
   Interior compactness

26.3 Asymptotic expansion, 297
   Solutions of elliptic differential equations in a half-cylinder

27. Index Theory
   27.1 The Noether index, 301
       Pseudo-inverse—Stability of index—Product formula—Hörmander’s stability theorem
   Historical note, 305
   27.2 Toeplitz operators, 305
       Index-winding number—The inverson of Toeplitz operators—Discontinuous symbols—Matrix Toeplitz operators
   27.3 Hankel operators, 312

28. Compact Symmetric Operators in Hilbert Space
   Variational principle for eigenvalues—Completeness of eigenfunctions—The variational principles of Fisher and Courant—Functional calculus—Spectral theory of compact normal operators—Unitary operators

29. Examples of Compact Symmetric Operators
   29.1 Convolution, 323
   29.2 The inverse of a differential operator, 326
   29.3 The inverse of partial differential operators, 327

30. Trace Class and Trace Formula
   30.1 Polar decomposition and singular values, 329
   30.2 Trace class, trace norm, and trace, 330
       Matrix trace
   30.3 The trace formula, 334
       Weyl’s inequalities—Lidskii’s theorem
   30.4 The determinant, 341
   30.5 Examples and counterexamples of trace class operators, 342
       Mercer’s theorem—The trace of integral operators—A Volterra integral operator—The trace of the powers of an operator
30.6 The Poisson summation formula, 348
Convolution on $S^1$ and the convergence of Fourier series—The Selberg trace formula

30.7 How to express the index of an operator as a difference of traces, 349

30.8 The Hilbert-Schmidt class, 352
Relation of Hilbert-Schmidt class and trace class

30.9 Determinant and trace for operator in Banach spaces, 353

31. Spectral Theory of Symmetric, Normal, and Unitary Operators 354

31.1 The spectrum of symmetric operators, 356
Reality of spectrum—Upper and lower bounds for the spectrum—Spectral radius

31.2 Functional calculus for symmetric operators, 358
The square root of a positive operator—Polar decomposition of bounded operators

31.3 Spectral resolution of symmetric operators, 361
Projection-valued measures

31.4 Absolutely continuous, singular, and point spectra, 364

31.5 The spectral representation of symmetric operators, 364
Spectral multiplicity—Unitary equivalence

31.6 Spectral resolution of normal operators, 370
Functional calculus—Commutative $B^*$-algebras

31.7 Spectral resolution of unitary operators, 372
Historical note, 375

32. Spectral Theory of Self-Adjoint Operators 377
The Hellinger-Toeplitz theorem—Definition of self-adjointness—Domain

32.1 Spectral resolution, 378
Sharpening of Herglotz’s theorem—Cauchy transform of measures—The spectrum of a self-adjoint operator—
Representation of the resolvent as a Cauchy transform—
Projection-valued measures

32.2 Spectral resolution using the Cayley transform, 389

32.3 A functional calculus for self-adjoint operators, 390

33. Examples of Self-Adjoint Operators 394

33.1 The extension of unbounded symmetric operators, 394
Closure of a symmetric operator

33.2 Examples of the extension of symmetric operators; deficiency indices, 397
The operator $i(d/dx)$ on $C^1_0(R), C^1_0(R_+), \text{and } C^1_0(0, 1)$—
Deficiency indices and von Neumann’s theorem—Symmetric operators in a real Hilbert space

33.3 The Friedrichs extension, 402
Semibounded symmetric operators—Symmetric ODE—
Symmetric elliptic PDE

33.4 The Rellich perturbation theorem, 406
Self-adjointness of Schrödinger operators with singular potentials

33.5 The moment problem, 410
The Hamburger and Stieltjes moment problems—Uniqueness, or not, of the moment problem
Historical note, 414

34. Semigroups of Operators 416

34.1 Strongly continuous one-parameter semigroups, 418
Infinitesimal generator—Resolvent—Laplace transform

34.2 The generation of semigroups, 424
The Hille-Yosida theorem

34.3 The approximation of semigroups, 427
The Lax equivalence theorem—Trotter’s product formula—
Strang’s product formula

34.4 Perturbation of semigroups, 432
Lumer-Phillip’s theorem—Trotter’s perturbation theorem

34.5 The spectral theory of semigroups, 434
Phillip’s spectral mapping theorem—Adjoint semigroups—
Semigroups of eventually compact operators

35. Groups of Unitary Operators 440

35.1 Stone’s theorem, 440
Generation of unitary groups—Positive definiteness and
Bochner’s theorem

35.2 Ergodic theory, 443
von Neumann’s mean ergodic theorem
### CONTENTS

36. **Examples of Strongly Continuous Semigroups**

36.1 Semigroups defined by parabolic equations, 461
36.2 Semigroups defined by elliptic equations, 462
36.3 Exponential decay of semigroups, 465
36.4 The Lax-Phillips semigroup, 470
36.5 The wave equation in the exterior of an obstacle, 472

37. **Scattering Theory**

37.1 Perturbation theory, 477
37.2 The wave operators, 480
37.3 Existence of the wave operators, 482
37.4 The invariance of wave operators, 490
37.5 Potential scattering, 490
37.6 The scattering operator, 491
37.7 The Lax-Phillips scattering theory, 493
37.8 The zeros of the scattering matrix, 499
37.9 The automorphic wave equation, 500
Faddeev and Pavlov’s theory—The Riemann hypothesis

38. **A Theorem of Beurling**

38.1 The Hardy space, 513
38.2 Beurling’s theorem, 515
Inner and outer factors—Factorization in the algebra of bounded analytic functions

---

### A. Riesz-Kakutani representation theorem

A.1 Positive linear functionals, 529
A.2 Volume, 532
A.3 $L$ as a space of functions, 535
A.4 Measurable sets and measure, 538
A.5 The Lebesgue measure and integral, 541

### B. Theory of distributions

B.1 Definitions and examples, 543
B.2 Operations on distributions, 544
B.3 Local properties of distributions, 547
B.4 Applications to partial differential equations, 554
B.5 The Fourier transform, 558
B.6 Applications of the Fourier transform, 568
B.7 Fourier series, 569

### C. Zorn’s Lemma

**Author Index**

**Subject Index**
LINEAR SPACES

A linear space $X$ over a field $F$ is a mathematical object in which two operations are defined: addition and multiplication by scalars.

Addition, denoted by $+$, as in

$$x + y$$

is assumed to be commutative,

$$x + y = y + x,$$  \hspace{1cm} (1)

associative,

$$x + (y + z) = (x + y) + z,$$  \hspace{1cm} (2)

and to form a group, with the neutral element denoted as 0:

$$x + 0 = x.$$  \hspace{1cm} (4)

The inverse of addition is denoted by $-$:

$$x + (-x) = x - x = 0.$$  \hspace{1cm} (5)

The second operation is the multiplication of elements of $X$ by elements $k$ of the field $F$:

$$kx.$$  

The result of this multiplication is again an element of $X$. Multiplication by elements of $F$ is assumed to be associative,

$$k(ax) = (ka)x,$$  \hspace{1cm} (6)

and distributive,

$$k(x + y) = kx + ky$$  \hspace{1cm} (7)
as well as
\[(a + b)x = ax + bx.\] \hfill (8)

We assume that multiplication by the unit of \(F\), denoted as 1, acts as the identity:
\[1x = x.\] \hfill (9)

These are the axioms of linear algebra. From them proceed to draw some deductions.

Set \(b = 0\) in (8). It follows that for all \(x\),
\[0x = 0.\] \hfill (10)

Set \(a = 1, b = -1\) in (8). Using (9) and (10), we deduce that for all \(x\),
\[(-1)x = -x.\] \hfill (11)

The finite-dimensional linear spaces are dealt with in courses on linear algebra. In this book the emphasis is on the infinite-dimensional ones—those that are not finite-dimensional. The field \(F\) will be either the real numbers \(\mathbb{R}\) or the complex numbers \(\mathbb{C}\). Here are some examples.

**Example 1.** \(X\) is the space of all polynomials in a single variable \(s\), with real coefficients, here \(F = \mathbb{R}\).

**Example 2.** \(X\) is the space of all polynomials in \(N\) variables \(s_1, \ldots, s_N\), with real coefficients, here \(F = \mathbb{R}\).

**Example 3.** \(G\) is a domain in the complex plane, and \(X\) the space of all functions complex analytic in \(G\), here \(F = \mathbb{C}\).

**Example 4.** \(X\) = space of all vectors
\[x = (a_1, a_2, \ldots)\]
with infinitely many real components, here \(F = \mathbb{R}\).

**Example 5.** \(Q\) is a Hausdorff space, \(X\) the space of all continuous real-valued functions on \(Q\), here \(F = \mathbb{R}\).

**Example 6.** \(M\) is a \(C^\infty\) differentiable manifold, \(X = C^\infty(M)\), the space of all differentiable functions on \(M\).

**Example 7.** \(Q\) is a measure space with measure \(m\), \(X = L^1(Q, m)\).

**LINEAR SPACES**

**Example 8.** \(X = L^p(Q, m)\).

**Example 9.** \(X\) = harmonic functions in the upper half-plane.

**Example 10.** \(X\) = all solutions of a linear partial differential equation in a given domain.

**Example 11.** All meromorphic functions on a given Riemann surface; \(F = \mathbb{C}\).

We start the development of the theory by giving the basic constructions and concepts. Given two subsets \(S\) and \(T\) of a linear space \(X\), we define their sum, denoted as \(S + T\) to be the set of all points \(x\) of the form \(x = y + z\), \(y\) in \(S\), \(z\) in \(T\). The negative of a set \(S\), denoted as \(-S\), consists of all points \(x\) of the form \(x = -y\), \(y\) in \(S\).

Given two linear spaces \(Z\) and \(U\) over the same field, their direct sum is a linear space denoted as \(Z \oplus U\), consisting of ordered pairs \((z, u)\), \(z\) in \(Z\), \(u\) in \(U\). Addition and multiplication by scalars is componentwise.

**Definition.** A subset \(Y\) of a linear space \(X\) is called a linear subspace of \(X\) if sums and scalar multiples of \(Y\) belong to \(Y\).

**Theorem 1.**

(i) The sets \([0]\) and \(X\) are linear subspaces of \(X\).

(ii) The sum of any collection of subspaces is a subspace.

(iii) The intersection of any collection of subspaces is a subspace.

(iv) The union of a collection of subspaces totally ordered by inclusion is a subspace.

**Exercise 1.** Prove theorem 1.

Let \(S\) be some subset of the linear space \(X\). Consider the collection \(\{Y_s\}\) of all linear subspaces that contain the set \(S\). This collection is not empty, since it certainly contains \(X\).

**Definition.** The intersection \(\bigcap Y_s\) of all linear subspaces \(Y_s\) containing the set \(S\) is called the linear span of the set \(S\).

**Theorem 2.**

(i) The linear span of a set \(S\) is the smallest linear subspace containing \(S\).

(ii) The linear span of \(S\) consists of all elements \(x\) of the form
\[
x = \sum_{i=1}^{r} a_i x_i, \quad x_i \in S, \quad a_i \in F, \quad r \text{ any natural number.}\] 

\hfill (12)
Proof. Part (i) is merely a rephrasing of the definition of linear span. To prove part (ii), we remark that on the one hand, the elements of the form (12) form a linear subspace of $X$; on the other hand, every $x$ of form (12) is contained in any subspace $Y$ containing $S$.

REMARK 1. An element $x$ of form (12) is called a linear combination of the points $x_1, \ldots, x_n$. So theorem 1 can be restated as follows:

The linear span of a subset $S$ of a linear space consists of all linear combinations of elements of $S$.

Definition. $X$ a linear space, $Y$ a linear subspace of $X$. Two points $x_1$ and $x_2$ of $X$ are called equivalent modulo $Y$, denoted as $x_1 \equiv x_2 \pmod{Y}$, if $x_1 - x_2$ belongs to $Y$.

It follows from the properties of addition that equivalence mod $Y$ is an equivalence relation, meaning that it is symmetric, reflexive, and transitive. That being the case, we can divide $X$ into distinct equivalence classes mod $Y$. We denote the set of equivalence classes as $X/Y$. The set $X/Y$ has a natural linear structure; the sum of two equivalence classes is defined by choosing arbitrary points in each equivalence class, adding them and forming the equivalence class of the sum. It is easy to check that the last equivalence class is independent of the representatives we picked; put differently, if $x_1 \equiv z_1, x_2 \equiv z_2$, then $x_1 + x_2 \equiv z_1 + z_2 \pmod{Y}$. Similarly we define multiplication by a scalar by picking arbitrary elements in the equivalence class. The resulting operation does not depend on the choice. Since, if $x_1 \equiv z_1$, then $kx_1 \equiv kz_1 \pmod{Y}$. The quotient set $X/Y$ endowed with this natural linear structure is called the quotient space of $X$ mod $Y$. We define $\text{codim } Y = \dim X/Y$.

Exercise 2. Verify the assertions made above.

As with all algebraic structures, so with linear structures we have the concept of isomorphism.

Definition. Two linear spaces $X$ and $Z$ over the same field are isomorphic if there is a one-to-one correspondence $T$ carrying one into the other that maps sums into sums, scalar multiples into scalar multiples; that is,

$$
T(x_1 + x_2) = T(x_1) + T(x_2),
$$

$$
T(kx) = kT(x).
$$

We define similarly homomorphism, called in this context a linear map.

Definition. $X$ and $U$ are linear spaces over the same field. A mapping $M : X \rightarrow U$ is called linear if it carries sums into sums, and scalar multiples into scalar multiples.

that is, if for all $x, y$ in $X$ and all $k$ in $F$

$$
M(x + y) = M(x) + M(y),
$$

$$
M(kx) = kM(x).
$$

X is called the domain of $M$, $U$ its target.

REMARK 2. An isomorphism of linear spaces is a linear map that is one-to-one and onto.

Theorem 3.

(i) The image of a linear subspace $Y$ of $X$ under a linear map $M : X \rightarrow U$ is a linear subspace of $U$.

(ii) The inverse image under $M$ of a linear subspace $V$ of $U$ is a linear subspace of $X$.


A very important concept in a linear space over the reals is convexity:

Definition. $X$ is a linear space over the reals; a subset $K$ of $X$ is called convex if, whenever $x$ and $y$ belong to $K$, the whole segment with endpoints $x, y$, meaning all points of the form

$$
ax + (1 - a)y, \quad 0 \leq a \leq 1,
$$

also belong to $K$.

Examples of convex sets in the plane are the circular disk, triangle, and semicircular disk. The following property of convex sets is an immediate consequence of the definition:

Theorem 4. Let $K$ be a convex subset of a linear space $X$ over the reals. Suppose that $x_1, \ldots, x_n$ belong to $K$; then so does every $x$ of the form

$$
x = \sum_{j=1}^{n} a_j x_j, \quad a_j \geq 0,
$$

$$
\sum_{j=1}^{n} a_j = 1.
$$


An $x$ of form (16) is called a convex combination of $x_1, x_2, \ldots, x_n$. 
Theorem 5. Let $X$ be a linear space over the reals.

(i) The empty set is convex.

(ii) A subset consisting of a single point is convex.

(iii) Every linear subspace of $X$ is convex.

(iv) The sum of two convex subsets is convex.

(v) If $K$ is convex, so is $-K$.

(vi) The intersection of an arbitrary collection of convex sets is convex.

(vii) Let $\{K_j\}$ be a collection of convex subsets that is totally ordered by inclusion. Then their union $\bigcup K_j$ is convex.

(viii) The image of a convex set under a linear map is convex.

(ix) The inverse image of a convex set under a linear map is convex.

Exercise 5. Prove theorem 5.

Definition. Let $S$ be any subset of a linear space $X$ over the reals. The convex hull of $S$ is defined as the intersection of all convex sets containing $S$. The hull is denoted as $\hat{S}$.

Theorem 6.

(i) The convex hull of $S$ is the smallest convex set containing $S$.

(ii) The convex hull of $S$ consists of all convex combinations (16) of points of $S$.


Definition. A subset $E$ of a convex set $K$ is called an extreme subset of $K$ if:

(i) $E$ is convex and nonempty.

(ii) whenever a point $x$ of $E$ is expressed as

\[ x = \frac{y+z}{2}, \quad y, z \in K, \]

then both $y$ and $z$ belong to $E$.

An extreme subset consisting of a single point is called an extreme point of $K$.

Example 1. $K$ is the interval $0 \leq x \leq 1$; the two endpoints are extreme points.

Example 2. $K$ is the closed disk $x^2 + y^2 \leq 1$.

Every point on the circle $x^2 + y^2 = 1$ is an extreme point.

Example 3. The open disk

\[ x^2 + y^2 < 1 \]

has no extreme points.

Example 4. $K$ a polyhedron, including faces. Its extreme subsets are its faces, edges, vertices, and of course $K$ itself.

Theorem 7. Let $K$ be a convex set, $E$ an extreme subset of $K$, and $F$ an extreme subset of $E$. Then $F$ is an extreme subset of $K$.


Theorem 8. Let $M$ be a linear map of the linear space $X$ into the linear space $U$. Let $K$ be a convex subset of $U$, $E$ an extreme subset of $K$. Then the inverse image of $E$ is either empty or an extreme subset of the inverse image of $K$.


Exercise 9. Give an example to show that the image of an extreme subset under a linear map need not be an extreme subset of the image.

Taking $U$ to be one dimensional, we get

Corollary 8'. Denote by $H$ a convex subset of a linear space $X$, $\ell$ a linear map of $X$ into $\mathbb{R}$, $H_{\text{min}}$ and $H_{\text{max}}$ the subsets of $H$, where $\ell$ achieves its minimum and maximum, respectively.

Assertion. When nonempty, $H_{\text{min}}$ and $H_{\text{max}}$ are extreme subsets of $H$. 
2

LINEAR MAPS

2.1 ALGEBRA OF LINEAR MAPS

We recall from chapter 1 that a linear map from one linear space \( X \) into another, \( U \), both over the same field of scalars, is a mapping of \( X \) into \( U \),

\[ M : X \rightarrow U, \]

that is an algebraic homomorphism:

\[ M(x + y) = M(x) + M(y), \]
\[ M(kx) = kM(x). \]

In this section we explore those properties of linear maps that depend on the purely algebraic properties (1), without any topological restrictions imposed on the spaces \( X, U \).

The sum of two linear maps \( M \) and \( N \) of \( X \) into \( U \), and the scalar multiple is defined as

\[ (M + N)(x) = M(x) + N(x), \]
\[ (kM)(x) = kM(x). \]

This makes a linear space out of the set of linear maps of \( X \) into \( U \). The space is denoted as \( L(X, U) \). Given two linear maps, one, \( M \) from \( X \rightarrow U \), the other, \( N \) from \( U \rightarrow W \), we can define their product as the composite map

\[ (NM)(x) = N(M(x)). \]

Since composition of maps in general is associative, so is in particular the composition of linear maps. As we will see, composition is far from being commutative.

From now on we omit the bracket and denote the action of a linear map on \( x \) as

\[ M(x) = Mx. \]

This notation suggests that the action of \( M \) on \( x \) is a kind of multiplication; indeed (1) and (2) give the distributive property of this kind of multiplication.

ALGEBRA OF LINEAR MAPS

Exercise 1. Verify that the composite of two linear maps is linear, and that the distributive law holds:

\[ M(N + K) = MN + MK, \]
\[ (M + K)N = MN + KN. \]

Definition. A mapping is invertible if it maps \( X \) one-to-one and onto \( U \).

If \( M \) is invertible, it has an inverse, denoted as \( M^{-1} \), that satisfies

\[ M^{-1}M = I, \quad MM^{-1} = I, \]

where \( I \) on the left is the identity mapping in \( X \), on the right on \( U \). If \( M \) is linear, so is \( M^{-1} \).

Definition. The nullspace of \( M \), denoted by \( N_M \), is the set of points mapped into zero.

The range of \( M \), denoted by \( R_M \), is the image of \( X \) under \( M \) in \( U \).

Theorem 1. Let \( M \) be a linear map of \( X \rightarrow U \).

(i) The nullspace \( N_M \) is a linear subspace of \( X \), the range \( R_M \) a linear subspace of \( U \).

(ii) \( M \) is invertible iff \( N_M = \{0\} \) and \( R_M = U \).

(iii) \( M \) maps the quotient space \( X/N_M \) one-to-one into \( R_M \).

(iv) If \( M : X \rightarrow U \) and \( K : U \rightarrow W \) are both invertible, so is their product, and

\[ (KM)^{-1} = M^{-1}K^{-1}. \]

(v) If \( KM \) is invertible, then

\[ N_M = \{0\}, \quad R_K = W. \]

Exercise 2. Prove Theorem 1.

We remark that when \( x = U = W \) are finite dimensional, then the invertibility of the product \( NM \) implies that \( N \) and \( M \) separately are invertible. This is not so in the infinite-dimensional case; take, for instance, \( X \) to be the space of infinite sequences

\[ x = (a_1, a_2, \ldots) \]

and define \( R \) and \( L \) to be right and left shift: \( Rx = (0, a_1, a_2, \ldots) \), \( Lx = (a_2, a_3, \ldots) \). Clearly, \( LR \) is the identity map, but neither \( R \) nor \( L \) are invertible; nor is \( RL \), the identity.
We formulate now a number of useful notions and results concerning mappings of a linear space into itself:

$$M : X \rightarrow X.$$  

We denote by $N_j$ the nullspace of the $j$th power of $M$:

$$N_j = N_{M^j}. \quad (5)$$

**Theorem 2.** The subspaces $N_j$ defined in (5) have these properties:

$$N_j \subseteq N_{j+1} \quad \text{for all } j \quad (6)$$

and

$$\dim \left( \frac{N_j}{N_{j+1}} \right) \geq \dim \left( \frac{N_{j+1}}{N_j} \right) \quad \text{for all } j. \quad (7)$$

**Proof.** Equation (6) is an immediate consequence of (5). To show (7), we claim that $M$ maps $N_{j+1}/N_j$ into $N_{j}/N_{j-1}$ in a one-to-one fashion. To see this, note that a nonzero element of $N_{j+1}/N_j$ is represented by a point $z$ in $N_{j+1}$ that does not lie in $N_j$. Clearly, $Mz$ lies in $N_j$ but not in $N_{j-1}$, this shows the one-to-oneness. It follows that $N_{j+1}/N_j$ is isomorphic to a subspace of $N_j/N_{j-1}$, from which the statement (7) about dimension follows. When $N_{j+1}/N_j$ is infinite-dimensional, so is $N_j/N_{j-1}$. \( \square \)

The following is an immediate corollary of equation (7):

**Theorem 2'.** Suppose that for some $i$ the subspaces defined by (5) satisfy

$$N_i = N_{i+1}; \quad (8)$$

then

$$N_i = N_k \quad \text{for all } k > i. \quad (8')$$

**Definition.** A subspace $Y$ of $X$ is called an *invariant subspace* of a linear map $M : X \rightarrow X$ if $M$ maps $Y$ into $Y$.

**Theorem 3.** Suppose that $Y$ is an invariant subspace of $X$ for a mapping $M : X \rightarrow X$. Then

(i) there is a natural interpretation of $M$ as a mapping $X/Y \rightarrow X/Y$.

(ii) if both maps

$$M : Y \rightarrow Y \text{ and } M : X/Y \rightarrow X/Y$$

are invertible, so is $M : X \rightarrow X$.

**Proof.** We leave part (i) to the reader. In (ii) we show first that the null space of $M$ on $X$ is trivial. To see this, suppose that

$$Mz = 0;$$

then, since the null space of $M$ on $X/Y$ is assumed to be trivial, it follows that $z$ belongs to $Y$. But since the null space of $M$ on $Y$ is also trivial, it follows that $z = 0$.

Next we show that $M : X \rightarrow X$ is onto, meaning that

$$Mx_0 = u_0 \quad (9)$$

has a solution $x_0$ for every $u_0$ in $X$. To this end we solve equation (9) in two stages. First we solve the congruence

$$Mx = u_0 \pmod{Y},$$

which is possible since $M$ maps $X/Y$ onto itself. Let $x_1$ be an element of the solution class; then $x_1$ satisfies

$$Mx_1 = u_0 + z, \quad z \in Y.$$ 

Therefore the solution $x_0$ of (9) is

$$x_0 = x_1 - y,$$

where $y$ is the solution in $Y$ of

$$My = z.$$ 

Such a solution exists since $M$ is assumed to map $Y$ onto $Y$. \( \square \)

We remark that whereas invertibility of $M$ on $X$ and $X/Y$ guarantees the invertibility of $M$ on $X$, the converse by no means holds in spaces of infinite dimension. For example, let $X$ be the space of all bounded continuous functions on $\mathbb{R}$, $S$ the shift operator

$$(Sx)(t) = x(t-1),$$

and $Y$ the subspace of functions $x(t)$ that vanish on the negative axis. Clearly, $Y$ is shift invariant, and equally clearly, $S$ is invertible on $X$, its inverse being the left unit shift. But $S$ is not invertible on either $Y$ or $X/Y$; on $Y$ its range consists of functions $x(t)$ that are zero for $t \leq 1$, and on $X/Y$ it has a nontrivial nullspace.

**Exercise 3.** What is the null space of $S$ on $X/Y$?

The construction of invariant subspaces will be taken up in chapter 25. Here we gather the following useful observations:
Theorem 4. Let $M$ be a linear map: $X \rightarrow X$.

(i) For any $y$ in $X$, the set $(p(M)y)$, where $p$ represents any polynomial, is an invariant subspace of $M$.

(ii) Let $T$ be a linear map: $X \rightarrow X$ that commutes with $M$: $TM = MT$. Then the nullspace of $T$ is an invariant subspace of $M$.

Proof. Part (i) rests on the observation that if $p(M)$ is a polynomial, so is $M p(M)$. Part (ii) follows from the observation that if $M$ and $T$ commute, and if $z$ is in the nullspace of $T$: $Tz = 0$, then $TMz = MTz = M0 = 0$.

2.2 INDEX OF A LINEAR MAP

The next group of theorems describe an important special class of mappings.

Definition. A linear map $G$ is called degenerate if its range is finite dimensional:

$$\dim R_G < \infty$$  \hspace{1cm} (10)

Theorem 5. The degenerate maps form an ideal in the following sense:

(i) The sum of two degenerate maps is degenerate.

(ii) The product of a degenerate map with any linear map, in either order, is degenerate; that is, if $G$ is degenerate, so are $MG$ and $GN$, provided of course that the products can be defined.

Exercise 4. Prove theorem 5.

Definition. The linear maps $M : X \rightarrow U$ and $L : U \rightarrow X$ are pseudoinverse to each other if

$$LM = I + G, \quad ML = I + G,$$  \hspace{1cm} (11)

where $I$ denotes the identity, $G$ degenerate maps of $X \rightarrow X$, and $U \rightarrow U$, respectively.

Exercise 5. Prove that the right shift and the left shift described after theorem 1 are pseudoinverses of each other on the space of all sequences.

Theorem 6.

(i) If $L$ and $M$ are pseudoinverses of each other, so are $L + G_1$ and $M + G_2$, where $G_1$, $G_2$ are arbitrary degenerate maps.

(ii) Suppose that $M : X \rightarrow U$ and $A : U \rightarrow W$ have pseudoinverses $L$ and $B$, respectively. Then $AM$ and $LB$ are pseudoinverse to each other.


We recall the definition of codimension of a subspace $R$ of a linear space $U$:

$$\text{codim } R = \dim(U / R).$$

Theorem 7. A linear map $M : X \rightarrow U$ has a pseudoinverse if and only if

$$\dim N_M < \infty, \quad \text{codim } R_M < \infty.$$  \hspace{1cm} (12)

Proof. For the "only if" part we use a lemma:

Lemma 8. If $G$ is a degenerate map of $X \rightarrow X$, then

$$\dim N_{I+G} < \infty, \quad \text{codim } R_{I+G} < \infty.$$  \hspace{1cm} (13)

Proof. For $x$ in $N_{I+G}$,

$$x + Gx = 0.$$

This shows that

$$N_{I+G} \subset R_G;$$

combined with (10) this shows the first part of (13).

According to theorem 1 (iii), $G$ maps $X/N_G$ one-to-one onto $R_G$; so

$$\dim N_G = \dim R_G.$$  \hspace{1cm} (14)

Obviously $I + G$ maps every $x$ in $N_G$ into itself; this shows that $R_{I+G} \supset N_G$. It follows from this relation that

$$\text{codim } R_{I+G} \leq \text{codim } N_G.$$  \hspace{1cm} (14')

Combining (14) and (14'), we conclude that $\text{codim } R_{I+G} \leq \dim R_G$; using (10), we deduce the second part of (13).

Suppose now that $M$ has a pseudoinverse; then (11) holds. From the first relation in (11) we deduce that $N_M \subset N_{I+G}$ and therefore $\dim N_M \leq \dim N_{I+G}$; combining this with the first part of (13), we obtain the first part of (12). It follows from the second relation in (11) that $R_M \supset R_{I+G}$. Therefore

$$\text{codim } R_M \leq \text{codim } R_{I+G}.$$  \hspace{1cm} (12')

Combining this with the second relation in (13), we deduce the second part of (12).
For the "if" part we need:

**Lemma 9.** Every subspace \( N \) of a linear space has a complementary subspace \( Y \), namely a linear subspace \( Y \) of \( X \) such that

\[
X = N \oplus Y,
\]

meaning that every \( x \) in \( X \) can be decomposed uniquely as

\[
x = n + y, \quad n \in N, \ y \in Y.
\]  \hfill (15)

**Proof.** Consider all subspaces \( Y \) of \( X \) whose intersection with \( N \) is \( \{0\} \), partially ordered by inclusion. Every totally ordered collection of \( Y_j \) has as upper bound the union of the \( Y_j \). Zorn's lemma shows that there is a maximal \( Y \); this \( Y \) clearly has the property stated in the lemma. Now, if some \( x \) cannot be expressed of form (15), we could enlarge \( Y \) by adjoining \( x \), contradicting the maximality of \( Y \). \( \square \)

Note that the complementary subspace \( Y \) is in no way uniquely determined. Having determined a particular \( Y \), we define the projection \( P \) onto \( N \) from the decomposition (15):

\[
P_x = n.
\]

**Exercise 7.** Prove that \( P \) is a linear map.

**Exercise 8.** Show that when \( N \) has finite codimension, \( \dim Y = \text{codim} \ N \).

We return now to the proof of the "if" part of theorem 7: it follows from (15) that every equivalence class of \( X \mod N \) contains exactly one element belonging to \( Y \), and that this correspondence is an isomorphism:

\[
Y \leftrightarrow X/N.
\]

Suppose that \( M : X \rightarrow U \) satisfies conditions (12); we choose complementary subspaces \( Y \) and \( V \) for the nullspace and range of \( M \):

\[
X = N_M \oplus Y, \quad U = R_M \oplus V.
\]  \hfill (16)

According to theorem 1 (iii), \( M \) maps \( X/N_M \) one-to-one onto \( R_M \). Since \( X/N_M \) is isomorphic with \( Y \), we conclude that

\[
M : Y \rightarrow R_M
\]

is invertible. Denote its inverse by \( M^{-1} \) and define the map \( K \) as follows:

\[
K = M^{-1} \text{ on } R_M, \quad K = 0 \text{ on } V.
\]  \hfill (17)

**INDEX OF A LINEAR MAP**

Using (16), we can extend \( K \) to all of \( U \). Clearly,

\[
KM = \begin{cases} 1 & \text{on } Y \\ 0 & \text{on } N_M \end{cases}, \quad MK = \begin{cases} 1 & \text{on } R_M \\ 0 & \text{on } V \end{cases}
\]  \hfill (17')

We can rewrite (17') as follows:

\[
KM = 1 - P, \quad MK = 1 - Q,
\]

where \( P \) is projection onto \( N \), \( Q \) is projection onto \( V \). It follows from this that \( K \) and \( M \) are pseudoinverse to each other in the sense of (11). Since \( P \) and \( Q \) are degenerate, the proof of theorem 7 is complete. \( \square \)

**Definition.** Let \( M : X \rightarrow U \) be a linear map with a pseudoinverse. We define the index of such an \( M \) as

\[
\text{ind} M = \dim N_M - \text{codim} R_M.
\]  \hfill (18)

It follows from theorem 7 that this definition makes sense.

**Theorem 10.** \( M : X \rightarrow U \) and \( L : U \rightarrow W \) are linear maps with pseudoinverse. Then the product \( LM \) has pseudoinverse, and

\[
\text{ind} (LM) = \text{ind} L + \text{ind} M.
\]  \hfill (19)

**Proof.** By theorem 6 (ii), \( LM \) has a pseudoinverse. To prove (19), we want to use as a counting device the notion of an exact sequence:

**Definition.** A sequence of linear spaces \( V_0, V_1, \ldots, V_n \) and a sequence of linear maps \( T_j : V_j \rightarrow V_{j+1} \),

\[
V_0 \xrightarrow{T_0} V_1 \xrightarrow{T_1} \cdots \xrightarrow{T_n} V_n,
\]

is called exact if the range of \( T_j \) is the nullspace of \( T_{j+1} \).

**Lemma 11.** Suppose that all the \( V_j \) in the exact sequence above are finite dimensional and that

\[
\text{dim } V_0 - 0 = \text{dim } V_1.
\]  \hfill (20)

Then

\[
\sum_j (-1)^j \text{dim } V_j = 0.
\]  \hfill (20')

**Proof.** Decompose \( V_j \) as

\[
V_j = N_j \oplus Y_j.
\]

We use the fact that \( V_0 = N_0 \oplus N_1 \oplus \cdots \oplus N_n \) and, consequently,

\[
\text{dim } V_0 = \text{dim } N_0 + \text{dim } N_1 + \cdots + \text{dim } N_n.
\]

Since \( N_0 \) is the nullspace of \( T_0 \), we have

\[
\text{dim } N_0 = \text{dim } \text{range } T_0 - \text{nullspace } T_0.
\]

Since \( V_1 = N_1 \oplus Y_1 \), we have

\[
\text{dim } V_1 = \text{dim } N_1 + \text{dim } Y_1.
\]

Since \( Y_1 \) is the range of \( T_1 \), we have

\[
\text{dim } Y_1 = \text{nullspace } T_1.
\]

Continuing in this manner, we find

\[
\text{dim } V_j = \text{dim } N_j + \text{dim } Y_j.
\]

Since \( Y_{j+1} \) is the nullspace of \( T_{j+1} \), we have

\[
\text{dim } Y_{j+1} = \text{nullspace } T_{j+1}.
\]

Since \( V_{j+1} = N_{j+1} \oplus Y_{j+1} \), we have

\[
\text{dim } V_{j+1} = \text{dim } N_{j+1} + \text{dim } Y_{j+1}.
\]

Continuing in this manner, we find

\[
\sum_j (-1)^j \text{dim } V_j = 0.
\]  \hfill (20')
where \( N_j \) is the nullspace of \( T_j \) and \( Y_j \) complementary to \( N_j \). The condition of exactness requires that \( T_j \) be an isomorphism of \( Y_j \) with \( N_{j+1} \). Since \( \dim V_j = \dim N_j + \dim Y_j \), it follows that
\[
\dim V_j = \dim N_j + \dim N_{j+1}, \quad 0 \leq j < n - 1. \tag{21}
\]

By \( 20' \),
\[
\dim N_0 = 0 \quad \text{and} \quad \dim V_{n-1} = \dim N_{n-1}. \tag{21'}
\]

Setting \( 21 \) and \( 21' \) in the left side of \( 20' \) shows that the alternating sum is zero. \( \Box \)

To prove theorem 10, we construct the following exact sequence:
\[
0 \to N_M \xrightarrow{L_0} N_{LM} \xrightarrow{Q} U/R_M \xrightarrow{L} W/R_{LM} \xrightarrow{E} W/R_L \to 0. \tag{22}
\]

The mapping \( L_0 \) identifies \( N_M \) as a subspace of \( N_{LM} \). \( Q \) is the natural map of points of \( U \) into the equivalence classes of \( U \mod R_M \) containing them. \( E \) is the mapping of equivalence classes of \( W \mod R_{LM} \) into equivalence classes \( \mod R_L \).

**Exercise 9.** Verify that \( 22 \) is an exact sequence.

We apply relation \( 20' \) to the exact sequence \( 22 \), with
\[
V_0 = 0, \quad V_1 = N_M, \quad V_2 = N_{LM}, \quad V_3 = N_L, \\
V_4 = U/R_M, \quad V_5 = W/R_{LM}, \quad V_6 = W/R_L, \quad V_7 = 0.
\]

Using the definition of codimension, we can write \( 20' \) as follows:
\[
\dim N_M - \dim N_{LM} + \dim N_L - \text{codim} \ R_M + \text{codim} \ R_{LM} - \text{codim} \ R_L = 0.
\]

Using the definition \( 18 \) of the index, we deduce the product formula \( 19 \) for the index. \( \Box \)

The next result is called the *stability of index*:

**Theorem 12.** Let \( M : X \to U \) be a linear map with a pseudoinverse, and \( G : X \to U \) a degenerate linear map. Then \( M + G \) has a pseudoinverse, and
\[
\text{ind} (M + G) = \text{ind} M. \tag{23}
\]

**Proof.** We first verify \( 23 \) for \( U = X \) and \( M = I \). For this we need a lemma:

**Lemma 13.** Let \( X \) be a linear space, and \( K : X \to U \) a linear map of \( X \) into \( U \) that has a pseudoinverse. Let \( X_0 \) be a linear subspace of \( X \) that has finite codimension.

Then \( K_0 : X_0 \to U \), the restriction of \( K \) to \( X_0 \), has a pseudoinverse, and
\[
\text{ind} K_0 = \text{ind} K - \text{codim} X_0. \tag{24}
\]

**Proof.** Factor \( K_0 \) as
\[
K_0 = KL_0, \tag{24'}
\]

where \( L_0 : X_0 \to X \) is the identification map. Clearly \( N_{L_0} = \{0\} \), \( R_{L_0} = X_0 \), so
\[
\text{ind} L_0 = -\text{codim} X_0. \tag{25}
\]

Now we apply the product formula \( 19 \) to \( 24' \) and deduce \( 24 \). \( \Box \)

Let \( G : X \to X \) be a degenerate map; take \( K : X \to X \) to be
\[
K = I + G. \tag{26}
\]

Clearly, \( I \) is a pseudoinverse to \( K \). Take \( X_0 \) to be the nullspace of \( G \):
\[
X_0 = N_G. \tag{27}
\]

By \( 14 \), \( X_0 \) has finite codimension. Since \( G \) is zero on \( X_0 \), \( K_0 \), the restriction of \( K \) to \( X_0 \), is the identification map \( L_0 \). So by \( 25 \),
\[
\text{ind} K_0 = \text{ind} L_0 = -\text{codim} X_0.
\]

We apply now lemma 13 to \( K \). By \( 24 \),
\[
\text{ind} K_0 = \text{ind} K - \text{codim} X_0.
\]

We deduce from the last two relations that
\[
\text{ind} K = 0 \tag{28}
\]

for every \( K \) of form \( 26 \). This proves \( 23 \) for \( M = I \).

We take now \( M \) as any map with a pseudoinverse; denote by \( L : U \to X \) a pseudoinverse of \( M \). By definition,
\[
LM = K = I + G',
\]

\( G' \) degenerate. So by \( 28 \),
\[
\text{ind} (LM) = \text{ind} (I + G') = 0. \tag{29}
\]

Using the product formula \( 19 \), we get from \( 29 \) that
\[
\text{ind} L = -\text{ind} M. \tag{30}
\]
As we saw in theorem 6 (i), for degenerate \( G, L \) is also a pseudoinverse of \( M + G \). Therefore, using (30), once more we deduce that:

\[
\text{ind } L = -\text{ind } (M + G). \tag{30'}
\]

Combining (30) and (30'), we get (23).

Notes

(i) The notion of the index of linear maps that have a pseudoinverse, theorem 7.
(ii) The product formula for the index, theorem 10.
(iii) The invariance of the index under perturbation by degenerate maps, theorem 12.

Strange to say, these results of linear algebra were first discovered in the setting of bounded maps of normed linear spaces. That they hold without any topological assumptions has remained a folk theorem. The first statement and proof of the multiplicative property in print is due to Donald Sarason. The proof presented here, using exact sequences, is due to Sergiu Klainerman.

BIBLIOGRAPHY


3
THE HAHN-BANACH THEOREM

3.1 THE EXTENSION THEOREM

The result named in the title of this chapter is remarkable for its simplicity and for its far-reaching consequences. It deals with the extension of linear functionals.

Definition. A linear functional \( \ell \) is a mapping of a linear space \( X \) over a field \( F \) into \( F \), that is linear:

\[
\ell(x + y) = \ell(x) + \ell(y)
\]

for all \( x, y \) in \( X \) and

\[
\ell(kx) = k\ell(x)
\]

for all \( k \) in \( F \).

In this section we will mainly deal with linear spaces over the field of reals, and real number valued linear functionals.

Theorem 1 (Hahn-Banach Theorem). Let \( X \) be a linear space over the reals, and \( p \) a real-valued function defined on \( X \), which has the following properties:

(i) Positive homogeneity,

\[
p(ax) = ap(x) \quad \text{for all } a > 0
\]

for every \( x \) in \( X \).

(ii) Subadditivity,

\[
p(x + y) \leq p(x) + p(y)
\]

for all \( x, y \) in \( X \).
THE HAHN-BANACH THEOREM

$Y$ denotes a linear subspace of $X$ on which a linear functional $\ell$ is defined that is dominated by $p$:

$$\ell(y) \leq p(y) \quad \text{for all } y \in Y.$$  \hspace{1cm} (3)

**Assertion.** $\ell$ can be extended to all of $X$ as a linear functional dominated by $p$:

$$\ell(x) \leq p(x) \quad \text{for all } x \in X.$$  \hspace{1cm} (3')

**Proof.** Suppose that $Y$ is not all of $X$; then there is some $z$ in $X$ that is not in $Y$. Denote by $Z$ the linear span of $Y$ and $z$, meaning all points of the form

$$y + az, \quad y \in Y, a \in \mathbb{R}.$$  

Our aim is to extend $\ell$ as a linear functional to $Z$ so that (3') is satisfied for $x$ in $Z$.

that is,

$$\ell(y + az) = \ell(y) + a\ell(z) \leq p(y + az)$$

holds for all $y$ in $Y$ and all real $a$. By (3), the inequality holds for $a = 0$. Since $p$ is positive homogeneous, it suffices to verify it for $a = \pm 1$:

$$\ell(y) + \ell(z) \leq p(y + z), \quad \ell(y') - \ell(z) \leq p(y' - z).$$

Thus for all $y, y'$ in $Y$,

$$\ell(y') - p(y' - z) \leq p(y + z) - \ell(y)$$  \hspace{1cm} (4)

must hold. Such an $\ell(z)$ exists iff for all pairs $y, y'$,

$$\ell(y') - p(y' - z) \leq p(y + z) - \ell(y).$$  \hspace{1cm} (5)

This is the same as

$$\ell(y') + \ell(y) = \ell(y' + y) \leq p(y + z) + p(y' - z).$$  \hspace{1cm} (5')

Since $y + y'$ lies in $Y$, (3) holds:

$$\ell(y' + y) \leq p(y + y').$$  \hspace{1cm} (6)

By subadditivity,

$$p(y + y') = p(y + z + y' - z) \leq p(y + z) + p(y' - z).$$  \hspace{1cm} (7)

Combining (6) and (7) gives (5'), proving the possibility of extending $\ell$ from $Y$ to $Z$.

So (3') remains satisfied.

Consider all extensions of $\ell$ to linear spaces $Z$ containing $Y$ on which inequality (3') continues to hold. We order these extensions by defining

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

to mean that $Z'$ contains $Z$, and that $\ell'$ agrees with $\ell$ on $Z$.

GEOMETRIC HAHN-BANACH THEOREM

Let $(Z_n, \ell_n)$ be a totally ordered collection of extensions of $\ell$. Then we can define $\ell$ on the union $Z = \bigcup Z_n$ as being $\ell_n$ on $Z_n$. Clearly, $\ell$ on $Z$ satisfies (3'); equally clearly, $(Z, \ell)$ is $(Z, \ell_n)$ for all $n$. This shows that every totally ordered collection of extensions of $\ell$ has an upper bound. So the hypothesis of Zorn's lemma is satisfied, and we conclude that there exists a maximal extension. But according to the foregoing, a maximal extension must be to the whole space $X$.

3.2 GEOMETRIC HAHN-BANACH THEOREM

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

**Definition.** $X$ is a linear space over the reals, $S$ a subset of $X$. A point $x_0$ is called an interior point of $S$ if for any $y$ in $X$ there is an $\epsilon$, depending on $y$, such that

$$x_0 + t y \in S \quad \text{for all } t, |t| < \epsilon.$$  

Let $K$ be a convex set that has an interior point, which we take to be the origin. We denote the gauge $p_K$ of $K$ with respect to the origin as follows:

$$p_K(x) = \inf a \quad a > 0, \frac{x}{a} \in K.$$  \hspace{1cm} (8)

Since the origin is assumed to be an interior point of $K$,

$$p_K(x) < \infty$$

for every $x$.

**Theorem 2.** The gauge $p_K$ of a convex set $K$ in a linear space over the reals is positive homogeneous and subadditive.

**Proof.** Positive homogeneity follows from the definition (8), even when $K$ is not convex. To prove subadditivity, let $x$ and $y$ be any pair of points in $X$, and $a$ and $b$ positive numbers such that

$$\frac{x}{a} \in K, \quad \frac{y}{b} \in K.$$  \hspace{1cm} (9)

Convexity, as defined in chapter 1, means that any convex combination of points of $K$ belongs to $K$. We take the convex combination of $x/c$ and $y/b$ with weights $a/(a+b)$ and $b/(a+b)$. These are nonnegative numbers whose sum is 1. We conclude that

$$\frac{a}{a+b} x + \frac{b}{a+b} y = \frac{x+y}{a+b} \in K.$$
Since \((x + y)/(a + b)\) is in \(K\), by definition (8), \(p_K(x + y) \leq a + b\). Since this holds for all \(a\) and \(b\) satisfying (9),

\[
p_K(x + y) \leq \inf(a + b) = \inf a + \inf b = p_K(x) + p_K(y),
\]

where in the last step we have again used (8). This proves subadditivity of \(p_K\). \(\square\)

**Theorem 3.** For any convex set \(K\),

\[
p_K(x) \leq 1 \quad \text{if} \ x \in K, \tag{10}
\]

\[
p_K(x) < 1 \text{ iff } x \text{ is an interior point of } K. \tag{10'}
\]

**Proof.** (10) is an immediate consequence of definition (8) of \(p_K\). \(\square\)

**Exercise 1.** Prove (10').

The converse of theorem 3 also is true:

**Theorem 4.** Let \(p\) denote a positive homogeneous, subadditive function defined on a linear space \(X\) over the reals.

(i) The set of points \(x\) satisfying \(p(x) < 1\) is a convex subset of \(X\), and 0 is an interior point of it.

(ii) The set of points \(x\) satisfying \(p(x) \leq 1\) is a convex subset of \(X\).

**Exercise 2.** Prove theorem 4.

We turn now to the notion of a hyperplane. Suppose that \(\ell\) is a linear functional not \(= 0\); for any real \(c\), all points of \(X\) belong to one, and only one, of the following three sets:

\[
\ell(x) < c, \quad \ell(x) = c, \quad \ell(x) > c.
\]

The set of \(x\) that satisfies \(\ell(x) = c\)

is called a hyperplane; the sets where \(\ell(x) < c\), respectively \(\ell(x) > c\) are called open halfspaces. The sets where

\[
\ell(x) \geq c, \quad \text{or } \ell(x) \leq c,
\]

are called closed halfspaces.

**Theorem 5 (Hyperplane Separation Theorem).** Let \(K\) be a nonempty convex subset of a linear space \(X\) over the reals; suppose that all points of \(K\) are interior. Any point \(y\) not in \(K\) can be separated from \(K\) by a hyperplane \(\ell(x) = c\); that is, there is a linear functional \(\ell\), depending on \(y\), such that

\[
\ell(x) < c \quad \text{for all } x \in K, \quad \ell(y) = c. \tag{11}
\]

**Proof.** Assume that \(0 \in K\), and denote by \(p_K\) the gauge of \(K\). Since all points of \(K\) are interior, it follows from theorem 3 that \(p_K(x') < 1\) for every \(x\) in \(K\). We set

\[
\ell(y) = 1. \tag{12}
\]

Then \(\ell\) is defined for all \(x\) of the form \(ay\),

\[
\ell(ay) = a. \tag{12'}
\]

We claim that for all such \(z\),

\[
\ell(z) \leq p_K(z).
\]

This is obvious for \(a \leq 0\), for then \(\ell(z) \leq 0\) while \(p_K \geq 0\). Since \(y\) is not in \(K\), by (8), \(p_K(y) \geq 1\). So, by positive homogeneity, \(p_K(y) \geq a\) for \(a > 0\).

Having shown that \(\ell\), as defined on the above one-dimensional subspace, is dominated by \(p_K\), we conclude from the HB theorem that \(\ell\) can be so extended to all of \(X\). We deduce from this and (10') that for any \(x\) in \(K\),

\[
\ell(x) \leq p_K(x) < 1
\]

This gives the first part of (11), with \(c = 1\); the second part is (12). \(\square\)

**Corollary 5.** Let \(K\) denote a convex set with at least one interior point. For any \(y\) not in \(K\) there is a nonzero linear functional \(\ell\) that satisfies

\[
\ell(x) \leq \ell(y) \quad \text{for all } x \in K. \tag{13}
\]

**Theorem 6 (Extended Hyperplane Separation).** \(X\) is a linear space over \(\mathbb{R}\), \(H\), and \(M\) disjoint convex subsets of \(X\), at least one of which has an interior point. Then \(H\) and \(M\) can be separated by a hyperplane \(\ell(x) = c\); that is, there is a nonzero linear functional \(\ell\), and a number \(c\), such that

\[
\ell(u) \leq c \leq \ell(v)
\]

for all \(u \in H\), all \(v \in M\).
Proof. According to theorem 5 of chapter I., the difference set $H - M = K$ is convex; since either $H$ or $M$ contains an interior point, it does $K$.

Since $H$ and $M$ are disjoint, $0 \not\in K$; according to (13) of corollary $S'$ applied to $y = 0$, there is a linear functional $\ell$ such that

$$\ell(x) \leq \ell(0) = 0 \text{ for all } x \in K. \quad (15)$$

Since all $x$ in $K = H - M$ is of the form $x = u - v$, $u$ in $H$, $v$ in $M$, (15) means that

$$\ell(u) \leq \ell(v); \quad (14)$$

(14) follows from this, with $c = \sup_{a \in H} \ell(u)$. \hfill \Box

3.3 Extensions of the Hahn-Banach Theorem

The following extension of the H-B theorem, due to R. E. Agnew and A. P. Morse, is both useful and beautiful:

**Theorem 7.** Let $X$ denote a linear space over the reals and $A$ be a collection of linear maps $A_x : X \to X$ that commute; that is,

$$A_{y}A_{x} = A_{x}A_{y} \quad (16)$$

for all pairs in the collection. Let $p$ denote a real-valued, positive homogeneous, subadditive function on $X$—see (1) and (2)—that is invariant under each $A_x$:

$$p(\lambda A_x x) = p(x). \quad (17)$$

Let $Y$ denote a linear subspace of $X$ on which a linear functional $\ell$ is defined, with the following properties:

(i) $\ell$ is dominated by $p$, namely

$$\ell(y) \leq p(y) \quad (18)$$

for every $y$ in $Y$.

(ii) $Y$ is invariant under each mapping $A_x$, namely

$$A_x y \in Y \quad (19)$$

for $y$ in $Y$.

(iii) $\ell$ is invariant under each mapping $A_x$, namely

$$\ell(A_x y) = \ell(y) \quad (19')$$

for $y$ in $Y$.

Assertion. $\ell$ can be extended to all of $X$ so that $\ell$ is dominated by $p$ in the sense of (18), and is invariant under each mapping $A_x$.

**Extensions of the Hahn-Banach Theorem**

Proof. If (17) holds for two mappings $A$ and $B$ of the collection $A$, it also holds for their product $AB$, defined as their composite. Similarly, if (19) and (19') hold for $A$ and $B$, they hold for the product $AB$. Likewise, if $A$ and $B$ commute with all $A_x$, so does their product. Thus we may adjoin to the collection $A$ any finite products and the identity $I$. This enlarged collection will now form a semigroup. Then, if $A$ and $B$ belong to it, so does their product $AB$. From now on we assume that the collection $A$ is a semigroup under multiplication.

We define a new function $g$ on $X$ as follows:

$$g(x) = \inf p(Cx), \quad (20)$$

with $C$ a convex combination of mappings in $A$, namely maps of the form

$$C = \sum a_j A_j, \quad a_j \geq 0, \sum a_j = 1, A_j \in A.$$

Since $A$ is a semigroup, the product of two convex combinations of mappings in $A$ is also a convex combination.

Using subadditivity, homogeneity, and invariance (17), we deduce that

$$p(Cx) = p\left(\sum a_j A_j x\right) \leq \sum a_j p(A_j x) = p(x). \quad (21)$$

Since in (20) we may take $C$ to be the identity, it follows that

$$g(x) \leq p(x). \quad (21')$$

Since $p$ is positive homogeneous, it follows from (20) that so is $g$. We show next that $g$ is subadditive.

Let $x$ and $y$ be arbitrary elements of $X$. By definition (20), for any $\epsilon > 0$ there are maps $C$ and $D$ in the convex hull of $A$ such that

$$p(Cx) \leq g(x) + \epsilon, \quad p(Dy) \leq g(y) + \epsilon. \quad (22)$$

Applying (20) to the map $CD$, we get, since $C$ and $D$ commute, that

$$g(x + y) \leq p(CD(x + y)) = p(DC x + CD y). \quad (23)$$

Using subadditivity, and (21), the right side of (23) is seen to be less than

$$p(DC x) + p(CD y) \leq p(Cx) + p(Dy). \quad (24)$$

Using (22) to estimate (24), we conclude that

$$g(x + y) \leq g(x) + g(y) + 2\epsilon;$$

since $\epsilon$ is arbitrary, subadditivity of $g$ follows.

Since, by (19'), $\ell$ on $Y$ is invariant under each $A$, for any convex combination $C$ of mappings in $A$ and for any $y$ in $Y$,
THE HAHN-BANACH THEOREM

\[ \ell(Cy) = \ell\left( \sum a_j A_j y \right) = \sum a_j \ell(A_j y) = \sum a_j \ell(y) = \ell(y). \]

It follows from (19) that if \( y \) belongs to \( Y \), so does \( Cy \). Applying (18) to \( Cy \), we get that for \( y \) in \( Y \),

\[ \ell(Cy) \leq \rho(Cy). \]

Since we have shown that \( \ell(Cy) = \ell(y) \),

\[ \ell(y) \leq \rho(Cy); \]

by definition (20) of \( \rho \), it follows from this that for all \( y \) in \( Y \),

\[ \ell(y) \leq \rho(y). \quad (25) \]

We apply now the Hahn-Banach theorem to conclude that \( \ell \) can be extended to all of \( X \) so that (25) holds. We claim that \( \ell \) thus extended is invariant under all mappings \( A \) in \( A \) in the sense of (19). For any \( A \) in \( A \), and any natural number \( n \), we define \( C_n \) by \( C_n = \frac{1}{n} \sum_{0}^{n-1} A^j \). Since \( A \) is a semigroup, \( C_n \) belongs to the convex hull of \( A \).

According to the basic formula for geometric series, \( C_n(I - A) = \frac{1}{n}(I - A^n) \).

Let \( x \) be any point in \( X \); by definition (20) of \( \rho \),

\[ g(x - Ax) \leq \rho(C_n(x - Ax)) = \rho(C_n(I - A)x) = \frac{1}{n} \rho(x - A^n x). \quad (26) \]

In the last step we used the formula for geometric series, and the positive homogeneity of \( \rho \). Using subadditivity and (17), we deduce that

\[ \frac{1}{n} \rho(x - A^n x) \leq \frac{1}{n} [\rho(x) + \rho(-A^n x)] = \frac{1}{n} [\rho(x) + \rho(-x)]. \]

Combining this with (26), we get

\[ g(x - Ax) \leq \frac{1}{n} [\rho(x) + \rho(-x)]. \quad (26') \]

Now we let \( n \to \infty \); since the right side of (26') tends to 0,

\[ g(x - Ax) \leq 0. \quad (27) \]

Since \( g \) dominates \( \ell \), we deduce from (27) that

\[ \ell(x - Ax) \leq 0. \]

Since \( \ell \) is linear, this implies that for all \( x \),

\[ \ell(x) \leq \ell(Ax). \quad (27') \]

Replacing \( x \) by \( -x \), we get

\[ \ell(-x) \leq \ell(-Ax). \]

EXTENSIONS OF THE HAHN-BANACH THEOREM

which is the opposite of inequality (27'). So equality must hold, meaning that \( \ell \) is invariant under each \( A \).

By construction, \( \ell \) is dominated by \( g \). It follows then from (21') that it is dominated by \( p \).

\[ \Box \]

Exercise 3. Show that theorem 7 remains true if condition (17) is replaced by \( \rho(Ax) \leq \rho(x) \).

We conclude by a version of HB for complex linear space due to Bohnenblust and Sobczyk, and Soukhomlinoff:

Theorem 8. Let \( X \) be a linear space over \( \mathbb{C} \), and let \( \rho \) a real valued function that satisfies

(i)

\[ \rho(ax) = |a| \rho(x) \]

for all complex \( a \), all \( x \) in \( X \);

(ii) subadditivity,

\[ \rho(x + y) \leq \rho(x) + \rho(y). \]

Let \( Y \) be a linear subspace of \( X \) over \( \mathbb{C} \), and let \( \ell \) be a linear functional on \( Y \) that satisfies

\[ |\ell(y)| \leq \rho(y) \]

for \( y \) in \( Y \).

Assertion. \( \ell \) can be extended to all of \( X \) so that (25) holds over \( X \).

Proof. Split \( \ell \) into its real and imaginary part:

\[ \ell(y) = \ell_1(y) + i \ell_2(y). \]

(30)

Clearly, \( \ell_1 \) and \( \ell_2 \) are linear over \( \mathbb{R} \), and are related by

\[ \ell_1(iy) = -\ell_2(y). \]

(31)

Conversely, if \( \ell_1 \) is a linear functional over \( \mathbb{R} \),

\[ \ell(x) = \ell_1(x) - i \ell_1(ix) \]

(31')

is linear over \( \mathbb{C} \).

We turn now to the task of extending \( \ell \). It follows from (29) and (30) that

\[ \ell_1(y) \leq \rho(y). \]

(32)
Therefore by the real H-B theorem, \( \ell_1 \) can be extended to all of \( X \) so that (32) holds. We define \( \ell \) on \( X \) by (31). Clearly, \( \ell \) is linear over \( \mathbb{C} \) and we claim that (29) holds. To see this, write

\[
\ell(x) = \alpha r, \quad r \in \mathbb{R}, \quad |\alpha| = 1.
\]

Then

\[
|\ell(x)| = r = \alpha^{-1} \ell(\alpha^{-1} x) = \ell_1(\alpha^{-1} x) \leq p(\alpha^{-1} x) = p(x).
\]

This completes the proof of the complex H-B theorem.

A historical review and a modern update is given by Gerard Buskes in his survey article.

**BIBLIOGRAPHY**


---

**4**

**APPLICATIONS OF THE HAHN-BANACH THEOREM**

**4.1 EXTENSION OF POSITIVE LINEAR FUNCTIONALS**

\( S \) denotes any abstract set, and \( B = B(S) \) the collection of all real-valued functions \( x \) on \( S \) that are bounded, that is, satisfy

\[
|x(s)| \leq c.
\]

\( B \) is a linear space over the reals.

There is a natural partial order for the elements of \( B : x \leq y \) means that \( x(s) \leq y(s) \) for all \( s \in S \). A function \( x \) satisfying \( 0 \leq x \) is called nonnegative.

Let \( Y \) be a linear subspace of \( B \) that contains some nonnegative functions. A linear functional \( \ell \) defined on \( Y \) is called positive on \( Y \) if \( \ell(y) \geq 0 \) for all nonnegative \( y \in Y \). Every positive linear functional \( \ell \) is monotone:

\[
y_1 \leq y_2 \quad \text{implies} \quad \ell(y_1) \leq \ell(y_2).
\]

**Theorem 1.** Let \( Y \) be a linear subspace of \( B \) that contains a function \( y_0 \) greater than some positive constant, say \( 1 \):

\[
1 \leq y_0(s) \quad \text{for all } s \in S.
\]

Let \( \ell \) be a positive linear functional defined on \( Y \).

**Assertion.** \( \ell \) can be extended to all of \( B \) as a positive linear functional.

**Proof.** We define the function \( p \) on \( B \) as follows: for any \( x \) in \( B \),

\[
p(x) = \inf_{y \in Y} \ell(y), \quad x \leq y; \ y \in Y.
\]

This function \( p \) is well defined; for it follows from (1) and (3) that

\[
-cy_0 \leq x \leq cy_0.
\]
which shows that the inf in (4) is over a nonempty set, and that \( p(x) \leq c \ell(y_0) \) where \( c \) is any constant satisfying (1). The smallest such constant is \( c = \sup_{y \in X} \varepsilon(x, y) \).

It follows from (5) that any \( y, x \) satisfies \(-c y_0 \leq x \leq y\). Since \( \ell \) is linear and positive, for such \( y \) it follows from (2) that \(-\varepsilon(y_0) \leq \ell(y) \), and so by (4) \n
\[ -c \ell(y_0) \leq p(x). \]  

(6)

**Lemma 2.** The function \( p \) defined by (4) is

(i) positive homogeneous.

(ii) subadditive.

(iii) negative: \( p(x) \leq 0 \) for \( x \leq 0 \).

(iv) \( p(x) = \ell(x) \) for \( x \) in \( Y \).

**Proof.**

(i) It follows from the definition that \( x \leq y \) implies \( ax \leq ay, \ a > 0 \). Positive homogeneity follows from definition (4).

(ii) Let \( x_1 \) and \( x_2 \) be any two functions in \( B \), \( \gamma_1 \) and \( \gamma_2 \) any two functions in \( Y \) satisfying

\[ x_1 \leq \gamma_1, \ x_2 \leq \gamma_2. \]

Adding the two we obtain \( x_1 + x_2 \leq \gamma_1 + \gamma_2 \; \) so by definition (4) of \( p \),

\[ p(x_1 + x_2) = \inf_{x_1 + x_2 \leq \gamma} \ell(x_1) + \inf_{x_1 + x_2 \leq \gamma} \ell(x_2) \]

\[ = \inf_{x_1 \leq \gamma} \ell(x_1) + \inf_{x_2 \leq \gamma} \ell(x_2) = p(x_1) + p(x_2) \]

(7)

This proves subadditivity.

(iii) Suppose that \( x \leq 0 \); then \( y = 0 \) is admissible in the inf on the right in (4), giving \( p(x) \leq \ell(0) = 0 \), as asserted in (iii).

(iv) Suppose that \( x \) belongs to \( Y \); then by (2), \( x \leq y \) implies \( \ell(x) \leq \ell(y) \), equality holding for \( y = x \). Setting this into (4) gives \( p(x) = \ell(x) \), as asserted in (iv).

\[ \square \]

It follows from lemma 2 that we can apply the Hahn-Banach theorem to extend \( \ell \) from \( Y \) to all of \( B \) so that \( \ell \) remains dominated by \( p \):

\[ \ell(x) \leq p(x). \]  

(8)

Suppose that \( x \) is nonpositive. Then by (iii), \( p(x) \leq 0 \), so by (8),

\[ \ell(x) \leq 0 \quad \text{for} \ x \leq 0. \]  

(9)

This shows that \( \ell \) is positive, as asserted in theorem 1.

\[ \square \]

**4.2 BANACH LIMITS**

\( B \) denotes the space of bounded infinite sequences \( x \) of real numbers,

\[ x = (a_1, a_2, \ldots). \]

(10)

\( B \) is a linear space over the reals when vector addition and multiplication by a scalar are defined componentwise. We define the function \( p \) on \( B \) as follows:

\[ p(x) = \lim_{n \to \infty} \sup_{a_n} a_n, \]

(11)

where \( x \) is given by (10). It follows from this definition that \( p \) is a positive homogeneous function of \( x \); we leave it as an exercise for the reader to prove that \( p \) is subadditive.

Define \( A \) as left translation, that is,

\[ Ax = (a_2, a_3, \ldots). \]  

(12)

It is an immediate consequence of definition (11) that \( p \) is translation invariant, namely that

\[ p(Ax) = p(x). \]  

(13)

We define \( Y \) as the space of convergent sequences of real numbers. Clearly, \( Y \) is a linear subspace of \( B \). On \( Y \), we define the linear functional \( \ell \) by

\[ \ell(y) = \lim_{n \to \infty} b_n, \]

(14)

where

\[ y = (b_1, b_2, \ldots). \]  

(14')

Clearly, \( \ell \) is linear. Comparing definitions (11) and (14), we conclude that

\[ \ell(y) = p(y) \quad \text{for} \ y \ in \ Y. \]  

(15)

Clearly, \( Y \) is mapped into itself by translation; equally clearly, \( \ell \) is invariant on \( Y \) under translation:

\[ \ell(Ay) = \ell(y) \quad \text{for} \ y \ in \ Y. \]  

(16)

We apply now theorem 2 in chapter 3 to conclude that \( \ell \) can be extended to all bounded sequences \( x \) in \( B \) so that
(i) \( \ell \) is linear 
(ii) \( \ell \) is invariant under translation 
(iii) \( \ell \) is dominated by \( p \).

**Theorem 3.** To each bounded sequence \((a_n)\) we can assign a generalized limit (or Banach limit), denoted as

\[
\lim_{n \to \infty} a_n
\]

so that

(i) For convergent sequences the generalized limit agrees with the usual limit.

(ii) 
\[
\lim_{n \to \infty} (a_n + b_n) = \lim_{n \to \infty} a_n + \lim_{n \to \infty} b_n.
\]

(iii) For any \( k \)
\[
\lim_{n \to \infty} a_{n+k} = \lim_{n \to \infty} a_n.
\]

(iv) 
\[
\liminf_{n \to \infty} a_n \leq \lim_{n \to \infty} a_n \leq \limsup_{n \to \infty} a_n.
\]

**Proof.** We set, in the notation of (10),

\[
\lim_{n \to \infty} a_n = i(x).
\]

Part (i) follows from (14), (14'), part (ii) expresses the linearity of \( \ell \); part (iii) is the translation invariance of \( \ell \); part (iv) expresses the domination of \( \ell \) by \( p \), as defined by (11), and applied to \( \ell(x) \) and \( \ell(-x) \):

\[-p(-x) \leq \ell(x) \leq p(x).\]

\( \square \)

**Exercise 1.** Show that if in section 4.1 we take \( S = \{ \text{positive integers} \} \), \( Y \) the space of convergent sequences, \( \ell \) defined by (14), the function \( p \) given by (4) is the same as defined by (11).

**Exercise 2.** Show that a Banach limit can be so chosen that for any bounded sequence \((c_1, c_2, \ldots)\) that is Cesaro summable, namely the arithmetic means of the partial sums converge to \( c \),

\[
\lim_{n \to \infty} c_n = c.
\]

**Exercise 3.** Show that a generalized limit as \( t \to \infty \) can be assigned to all bounded functions \( x(t) \) defined on \( t \geq 0 \) that has properties (i) to (iv) in theorem 3.

### 4.3 FINITELY ADDITIVE INARIANT SET FUNCTIONS

The Lebesgue measure on the unit circle is invariant under rotation. This measure can be extended to a considerably larger \(\sigma\)-algebra than the Lebesgue measurable sets on the unit circle so that rotational invariance is retained. However it is well known, and easy to show, that if we accept the axiom of choice, then there is no rotationally invariant countably additive measure defined for all subsets of the circle. We show now

**Theorem 4.** One can define a nonnegative finitely additive set function \( m(P) \), for all subsets \( P \) of the circle, that is invariant under rotation.

**Proof.** We take \( S \) to be the unit circle, and \( B \) the set of all bounded real-valued functions on \( S \). We take \( Y \) to be the space of bounded, Lebesgue measurable functions on \( S \), and take \( \ell(y) \) to be the Lebesgue integral of \( y \):

\[
\ell(y) = \int_S y(\theta) \, d\theta.
\]

(17)

The space \( Y \) contains the function \( y_0 = 1 \), so condition (3) of theorem 1 of section 4.1 is fulfilled. Therefore the function \( p \) described there by equation (4) is well defined.

We denote by \( \{ A_{\rho} \} \) the action on function of rotations \( \rho \) of the circle. As remarked above, \( \ell \) is invariant under rotation:

\[
(A_{\rho}) y(\theta) = y(\theta + \rho), \quad \ell(A_{\rho} y) = \ell(y).
\]

(18)

Since the relation \( x \leq y \) also is invariant under rotation, it follows that \( p \) as defined by (4) is rotation invariant:

\[
p(A_{\rho} x) = p(x).
\]

(18')

Rotations of the circle commute, and so the linear maps \( \{ A_{\rho} \} \) form a commuting group of maps. We apply now theorem 7 of chapter 3 to conclude that \( \ell \) can be extended to all of \( B \) so that \( \ell \)

(i) linear.

(ii) invariant under rotation.

(iii) dominated by \( p \).

Let \( P \) be any set of points of the circle \( S \); denote by \( c_P \) its characteristic function:

\[
c_P(\theta) = \begin{cases} 1 & \text{if } \theta \text{ is in } P \\
0 & \text{otherwise.} \end{cases}
\]

(19)

We define the set function \( m \) by setting,

\[
m(P) = \ell(c_P).
\]

(19')
APPLICATIONS OF THE HAHN-BANACH THEOREM

As shown in theorem 1, it follows from \( \ell(x) \leq p(x) \) that \( \ell \) is positive. Since \( c_P \) is a nonnegative function, it follows from definition (19') of \( m \) that \( m \) is nonnegative:

\[
m(P) \geq 0.
\]

Let \( \rho \) be any rotation; denote the set \( P \) rotated by \( \rho \) as \( P + \rho \). It follows from the definition (19) of \( c_P \) that

\[
c_{P + \rho} = A_\rho c_P.
\]

Since \( \ell \) is rotation invariant, it follows from the definition (19') of \( m \) that

\[
m(P + \rho) = m(P),
\]

meaning that \( m \) is rotationally invariant.

Let \( P_1 \) and \( P_2 \) be disjoint subsets. Then, by definition (19),

\[
c_{P_1 \cup P_2} = c_{P_1} + c_{P_2}.
\]

Setting this into the definition (19) of \( m \), and using the linearity of \( \ell \), we deduce that

\[
m(P_1 \cup P_2) = m(P_1) + m(P_2).
\]

This proves that \( m \) is finitely additive. \( \square \)

NOTE. Rotations of the circle commute with each other, and so the operators \( A_\rho \) commute; this was needed in invoking theorem 7 of chapter 3. Rotations of the three-dimensional sphere do not commute, and neither do the corresponding operators \( A_\rho \). Therefore the above proof cannot be used to extend theorem 4 to three dimensions.

In fact Hausdorff has shown that the three-dimensional analogue of theorem 4 is false; there is no rotational invariant, finitely additive set function on the 2-sphere.

The proof is based on a finite decomposition of the 2-sphere, sometimes called the Banach-Tarski paradox.

In conclusion, we point out that the duality theory of Banach spaces constitutes the richest applications of the Hahn-Banach theorem. These are described in chapters 8 and 9.

HISTORICAL NOTE. His name is etched into the foundations of modern analysis: Hausdorff space, Hausdorff maximality principle, and Hausdorff measure are household concepts. He was a German mathematician, born in 1868; as a young man he published several volumes of poetry and aphorisms. He spent most of his professional life as professor in Bonn. Because he was Jewish, in 1942 he was ordered deported, part of the "Final Solution" to kill all the Jews in Europe. Knowing what awaited them, Hausdorff, his wife, and sister-in-law committed suicide.

BIBLIOGRAPHY


5

NORMED LINEAR SPACES

5.1 NORMS

Let $X$ denote a linear space over $\mathbb{R}$ or $\mathbb{C}$. A norm in $X$ is a real-valued function $X \to \mathbb{R}$, denoted as $|x|$, with the following properties:

(i) **Positivity**, 

$$ |x| > 0 \quad \text{for } x \neq 0; |0| = 0. \quad (1) $$

(ii) **Subadditivity**, 

$$ |x + y| \leq |x| + |y|. \quad (2) $$

(iii) **Homogeneity.** For all scalars $a$,

$$ |ax| = |a| \cdot |x|. \quad (3) $$

With the aid of a norm we can introduce a metric in $X$, by defining the distance of two points to be

$$ d(x, y) = |x - y|. \quad (4) $$

It is easy to verify that this has all properties of a metric. Conversely, it is easy to show that every metric in a linear space that is translation invariant and homogeneous:

$$ d(x + z, y + z) = d(x, y), \quad d(ax, ay) = |a| \cdot d(x, y) \quad (4') $$

comes from a norm via (4).

With a metric (4) we can employ topological notions such as convergent series, open sets, closed sets, and compact sets. Those notions turn out to be crucial.

**Definition.** Two different norms, $|x|_1$ and $|x|_2$, defined on the same space $X$ are called equivalent if there is a constant $c$ such that

$$ c|x|_1 \leq |x|_2 \leq c^{-1}|x|_1 \quad (5) $$

for all $x$ in $X$.

The significance of this notion is that equivalent norms induce the same topology.

In chapter 4 we looked at various ways of building new linear spaces; the same constructions can be used to build new normed linear spaces. Specifically we observed the following:

(i) A subspace $Y$ of a normed linear space $X$ is again a normed linear space.

(ii) Given two linear spaces $Z$ and $U$, their Cartesian product, denoted as a direct sum $Z \oplus U$, consists of all ordered pairs $(z, u), z \in Z, u \in U$. When $Z$ and $U$ are normed, $Z \oplus U$ can be normed, such as by setting

$$ |(z, u)| = |z| + |u|, \quad |(z, u)|' = \max(|z|, |u|) \quad \text{or} \quad |(z, u)|'' = (|z|^2 + |u|^2)^{1/2}. \quad (6) $$

**Exercise 1.**

(a) Show that (6) are norms.

(b) Show that they are equivalent norms in the sense of (5).

Let $X$ be a normed linear space, $Y$ a subspace. We saw in chapter 4 that we can define their quotient $X/Y$ as a linear space. We raise now the question: is there a natural way to introduce a norm in the quotient space? The answer is yes, provided that $Y$ is closed:

**Theorem 1.** Let $Y$ be a closed subspace of a normed linear space $X$. Let $\{x_j\}$ be an equivalence class of elements of $X$ mod $Y$. We define

$$ ||x_j|| = \inf_{x_j \in [x]} |x_j|. \quad (7) $$

**Assertion.** (7) has all properties of a norm in the quotient space $X/Y$.

**Proof.** Property (3), homogeneity, holds trivially. To verify subadditivity, let $\{x_j\}$ and $\{z_j\}$ denote two equivalence classes. For any $\epsilon > 0$ we can, by definition (7), choose representatives so that

$$ |x_j| < ||x_j|| + \epsilon, \quad |z_j| < ||z_j|| + \epsilon. \quad (8) $$

By definition of addition in $X/Y$, $x_j + z_j$ belongs to $\{x_j\} + \{z_j\}$; therefore, by definition (7),

$$ ||x_j + z_j|| \leq ||x_j|| + ||z_j||. $$

37
which by subadditivity in $X$, and (8), is
\[ |x_j| + |z_j| < |(x_j)| + |(z_j)| + 2\varepsilon. \]

Since this is true for all $\varepsilon > 0$, subadditivity of the norm (7) follows.

Clearly, (7) is nonnegative. To show positivity, suppose that $|(x_j)| = 0$. By definition (7), there is a sequence of elements $x_n$ in $(x_j)$ such that
\[ \lim_{n \to \infty} |x_n| = 0. \]  
(9)

By definition of equivalence, the equivalent elements $x_n$ differ from each other by elements that belong to $Y$. In particular, we can write
\[ x_n = x_1 - y_n, \quad n = 2, 3, \ldots, \quad y_n \text{ in } Y. \]

Setting this into (9), we see that
\[ \lim_{n \to \infty} |x_1 - y_n| = 0, \]

which by (4) means in the language of metric spaces that
\[ \lim_{n \to \infty} y_n = x_1. \]  
(9')

In a metric space, the limit of a sequence of elements in a subset $Y$ belongs to the closure of $Y$. Now, since $Y$ is assumed to be closed, (9') implies that $x_1$ belongs to $Y$. But then the whole equivalence class $(x_j)$ consists of elements of $Y$, which is the zero element in $X/Y$. \qed

**Theorem 2.** Let $X$ be a normed linear space, $Y$ a subspace of $X$. The closure of $Y$ is a linear subspace of $X$.

**Exercise 2.** Prove theorem 2.

For purposes of analysis, in the construction of objects with desirable properties through limiting processes, we need metric spaces that are complete in the sense that every Cauchy sequence has a limit. So it is with normed linear spaces:

**Definition.** A Banach space is a normed linear space that is complete.

We recall the process of completion of a metric space whereby any metric space $S$ is embedded in a complete metric space denoted by $\bar{S}$, consisting of equivalence classes of Cauchy sequences. $S$ is a dense subset of $\bar{S}$, i.e. the closure of $S$ is $\bar{S}$.

**Theorem 3.** The completion $\bar{X}$ of a normed linear space $X$ under the metric (4) has a natural linear structure that makes $\bar{X}$ a complete normed linear space.

**Proof.** Recall that the points of the completion of a metric space are equivalence classes of Cauchy sequences. The term-by-term sum of two Cauchy sequences is again a Cauchy sequence, and sums of equivalent Cauchy sequences are equivalent. \qed

**Exercise 3.** Show that if $X$ is Banach space, $Y$ a close subspace of $X$, the quotient space $X/Y$ is complete. (Hint: Use a Cauchy sequence $(q_n)$ in $X/Y$ that satisfies $|q_n - q_{n+1}| < 1/n^2$.)

The process of completion of a normed linear space is one of the royal roads to obtaining complete normed linear spaces. This is extremely important for the success of functional analysis. We describe now a number of the most important normed linear spaces. These are the household items of modern analysis.

(a) The space of all vectors with infinite number of components
\[ x = (a_1, a_2, \ldots), \quad a_j \text{ complex}, \]

where the $|a_j|$ are bounded. The norm is
\[ |x|_\infty = \sup_j |a_j|. \]  
(10)

This space is denoted as $\ell^\infty$; it is complete.

(b) The space of all vectors with infinitely many components such that $\sum |a_j|^p < \infty$, $p$ some fixed number $\geq 1$. The norm is
\[ |x|_p = \left( \sum |a_j|^p \right)^{1/p}. \]  
(11)

This space is denoted as $\ell^p$; it is complete.

(c) $S$ an abstract set, $X$ the space of all complex-valued functions $f$ that are bounded. The norm is
\[ |f|_\infty = \sup_S |f(s)|. \]  
(12)

This space is complete.

(d) $Q$ a topological space, $X$ the space of all complex valued, continuous, bounded functions $f$ on $Q$. The norm is
\[ |f| = \sup_{Q} |f(q)|. \]  
(13)

This space is complete.
(e) $\mathcal{O}$ a topological space, $X$ the space of all complex-valued, continuous functions $f$ with compact support. The norm is

$$|f|_{\text{max}} = \max_{g} |f(g)|.$$  

This space is not complete unless $\mathcal{O}$ is compact.

(f) $D$, some domain in $\mathbb{R}^n$, $X$ the space of continuous functions $f$ with compact support. The norm is

$$|f|_p = \left(\int_{D} |f(x)|^p \, dx \right)^{1/p}, \quad 1 \leq p.$$  

This space is not complete; its completion is denoted by $L^p$.

(g) $D$, some domain in $\mathbb{R}^n$, the space of all $C^\infty$ functions $f$ in $D$ with the following property: for some integer $k$ and $p \geq 1$,

$$\int_{D} |\partial^\alpha f|^p \, dx < \infty \quad \text{for all } |\alpha| \leq k,$$

where $\partial^\alpha$ is any partial derivative:

$$\partial^\alpha = \partial_1^{\alpha_1} \cdots \partial_n^{\alpha_n}, \quad \partial_j = \frac{\partial}{\partial x_j}, \quad |\alpha| = \alpha_1 + \cdots + \alpha_n.$$

The norm is

$$|f|_{k,p} = \left(\sum_{|\alpha| \leq k} \int_{D} |\partial^\alpha f|^p \, dx \right)^{1/p}.$$  

This space is not complete; its completion is denoted as $W^{k,p}$, and is called a Sobolev space.

Theorem 4. The norms defined in examples (a) through (g) have properties (1) through (3) imposed on a norm.

Proof. Properties (1) and (3)—positivity and homogeneity—are obviously satisfied. We turn now to property (2), subadditivity. For the sake of brevity we consider only examples (a) and (b). Note that (a) can be regarded as a limiting case of (b), with $p = \infty$.

Define $x$ and $y$ as

$$x = (a_1, a_2, \ldots), \quad y = (b_1, b_2, \ldots).$$

Then

$$x + y = (a_1 + b_1, \ldots)$$

NORMS

We take first $p = \infty$. By (10),

$$|x + y|_\infty = \sup_j |a_j + b_j| \leq \sum_j |a_j| + |b_j| \leq \sum_j |a_j| + \sum_j |b_j| = |x|_\infty + |y|_\infty.$$

Next we turn to $p = 1$. By (11),

$$|x + y|_1 = \sum_j |a_j + b_j| \leq \sum_j |a_j| + |b_j| = |x|_1 + |y|_1.$$

For $1 < p < \infty$ we need Hölder's inequality. To state it, we introduce vectors $u$ with finite $q$-norm:

$$u = (c_1, c_2, \ldots), \quad \left(\sum_j |c_j|^q\right)^{1/q} = |u|_q < \infty,$$

where $q$ is conjugate to $p$, in the sense

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

We define now a scalar product between vectors in $\ell^p$ and $\ell^q$ as follows:

$$(x, u) = \sum_j a_j c_j.$$  

Hölder's Inequality. For $x$ in $\ell^p$, $u$ in $\ell^q$ the series defining the scalar product (18) converges, and

$$|(x, u)| \leq |x|_p |u|_q,$$

provided that $p$ and $q$ are conjugate in the sense of (17).

For a proof we refer to Courant's Calculus, Vol 2. The sign of equality holds in (19) iff

$$\arg a_j c_j \quad \text{and} \quad |a_j|^p / |c_j|^q$$

are independent of $j$.

Since for given $x$ in $\ell^p$ we can always choose $u$ in $\ell^q$ so that (20) is satisfied, and so that $|u|_q = 1$, we can restate Hölder's inequality thus:

Theorem 5. For any $x$ in $\ell^p$,

$$|x|_p = \max_{|u|_q = 1} |(x, u)|.$$  

Note that the scalar product (18) is bilinear as a function of $x$ and $u$. Applying (21) to $x + y$ in place of $x$ and using the linear dependence, we get
NORMED LINEAR SPACES

\[ |x + y|_p = \max_{\|u\| = 1} \left| (x + y, u) \right| \leq \max_{\|u\| = 1} \left| (x, u) \right| + \left| (y, u) \right|. \]  \hfill (22)

By Hölder's inequality (19), for \( |u|_q = 1 \),

\[ |(x, u)| \leq |x|_p, \quad |(y, u)| \leq |y|_p. \]

Setting this into (22) gives

\[ |x + y|_p \leq |x|_p + |y|_p, \]

as asserted in theorem 4. \( \Box \)

The self-conjugate case \( p = q = 2 \) is an instance of a supremely important class of norms, to be discussed in the next chapter.

The norms defined in examples (i) and (ii) satisfy important inequalities due to Sobolev: If

\[ mp \leq n \quad \text{and} \quad p \leq q \leq \frac{np}{n - kp} \]  \hfill (23)

and if \( Q \) is a cube, then

\[ \|f\|_q \leq \text{const.} \|f\|_{k,p}. \]  \hfill (23')

where the constant depends only on \( p, q, k, s \). These inequalities hold of course for all \( Q \) that are the images of cubes under a smooth mapping. Even more generally, they hold for all domains \( Q \) that satisfy a cone condition. For a proof, see Adams or Mazya.

Since the spaces \( L^q \) and \( W^{k,p} \) are constructed by completing the space of smooth functions in the appropriate norms, it follows that if condition (23) is fulfilled, \( W^{m,p} \) is contained in \( L^q \).

The norms linear spaces studied and used in analysis are infinite-dimensional. According to Cantor's theory of sets, there is a gradation among infinites; the least of them are the countable sets.

Definition. A normed linear space is called separable if it contains a countable set of points that is dense, namely, whose closure is the whole space.

Most, but not all, spaces that are used in analysis are separable. Here is an important example that is not:

(h) The space of all signed measures \( m \) on, say, the interval \([0, 1]\), of finite total mass. We define the norm to be the total mass:

\[ |m| = \int_0^1 |dm|. \]

NONCOMPACTNESS OF THE UNIT BALL

Denote by \( m_y \) the unit mass located at the point \( y \). Clearly, for \( y \neq z \), \( |m_y - m_z| = 2 \). Since there are noncountably many points \( y \) in the interval \([0, 1]\), this shows that the space of measures is not separable.

5.2 NONCOMPACTNESS OF THE UNIT BALL

Many existence theorems in finite-dimensional spaces rest on the fact that the closed unit ball, meaning that the set of points

\[ B_1 = \{ x : \|x\| \leq 1 \}, \]  \hfill (24)

is compact, that is to say, that any sequence of points in \( B_1 \) has a convergent subsequence. F. Riesz has shown that this property characterizes finite-dimensional spaces:

Theorem 6. Let \( X \) be an infinite-dimensional normed linear space; then the unit ball \( B_1 \) defined by (24) is not compact.

Proof. We require first a lemma:

Lemma 7. Let \( Y \) be a closed, proper subspace of the normed linear space \( X \). Then there is a vector \( z \) in \( X \) of length \( 1 \), \( |z| = 1 \),

\[ |z| = 1, \]  \hfill (25)

and that satisfies

\[ |z - y| > \frac{1}{2} \quad \text{for all } y \in Y. \]  \hfill (25')

Proof. Since \( Y \) is a proper subspace of \( X \), some point \( x \) of \( X \) does not belong to \( Y \). Since \( Y \) is closed, \( x \) has a positive distance to \( Y \):

\[ \inf_{y \in Y} |x - y| = d > 0. \]  \hfill (26)

There is then a \( y_0 \) in \( Y \) such that

\[ |x - y_0| < 2d. \]  \hfill (27)

Denote \( z' = x - y_0 \); we can then write (27) as

\[ |z'| < 2d. \]  \hfill (27')

It follows from (26) that

\[ |z' - y| \leq d \quad \text{for all } y \in Y. \]  \hfill (28)
NORMED LINEAR SPACES

We set

\[ z = \frac{z'}{|z'|}. \]

Clearly, (25) holds, and (25') follows from combining (27) and (28).

Remark 1. Clearly, the number \( \frac{1}{2} \) on the right of (25') can be replaced by any number \( < 1 \).

We turn now to the proof of theorem 6. We construct a sequence \( \{y_n\} \) of unit vectors recursively as follows: \( y_1 \) is chosen arbitrarily. Suppose that \( y_1, \ldots, y_{n-1} \) have been chosen; denote by \( Y_n \) the linear space spanned by them. Since \( Y_n \) is finite-dimensional, it is closed; since \( X \) is infinite-dimensional, \( Y_n \) is a proper subspace of \( X \). So lemma 7 is applicable and a \( z \) with properties (23), (25') exists. We set

\[ y_n = z. \]

Since \( y_j, j < n \) belongs to \( Y_n \),

\[ |y_n - y_j| > \frac{1}{2}, \quad j < n. \]

This shows that the distance of any two distinct \( y_j \) exceeds \( \frac{1}{2} \). Therefore no subsequence can form a Cauchy sequence. Since all \( y_j \) belong to the unit ball \( B_1 \), it follows that \( B_1 \) is not compact.

Exercise 4. Prove that every finite-dimensional subspace of a normed linear space is closed. (Hint: Use the fact that all norms are equivalent on finite-dimensional spaces to show that every finite-dimensional subspace is complete.)

Next we describe a kind of a substitute for the compactness that is lacking in the unit ball.

Definition. A norm is called strictly subadditive if in (2) strict inequality holds except when \( x \) or \( y \) is a nonnegative multiple of the other.

Exercise 5. Show that the sup norms of examples (a), (c), (d), and (e) are not strictly subadditive.

Exercise 6. Show that the norms in examples (b) and (f) are not strictly subadditive for \( p = 1 \).

All the norms in examples (b) and (f) are strictly subadditive when \( 1 < p < \infty \). Furthermore for each of these norms the condition holds uniformly, in the following sense:

NONCOMPACTNESS OF THE UNIT BALL

For any pair of unit vectors \( x, y \), the norm of \( (x + y)/2 \) is strictly less than 1 by an amount that depends only on \( |x - y| \). More explicitly, there is an increasing function \( \epsilon(r) \) defined for positive \( r \),

\[ \epsilon(r) > 0, \quad \lim_{r \to 0} \epsilon(r) = 0, \tag{29} \]

such that for all \( x, y \) in the unit ball \( |x| \leq 1, \quad |y| \leq 1 \), the inequality

\[ \left| \frac{x + y}{2} \right| \leq 1 - \epsilon(|x - y|). \tag{30} \]

holds.

Definition. A normed linear space whose norm satisfies (30) for all vectors \( x, y \) of unit length, where \( \epsilon(r) \) is some function satisfying (29), is called uniformly convex.

Theorem 8. Let \( X \) be a uniformly convex Banach space. Let \( K \) be a closed, convex subset of \( X \), \( z \) any point of \( X \). Then there is a unique point \( y \) of \( K \) which is closer to \( z \) than any other point of \( K \).

Proof. We may take \( z = 0 \), provided that we assume that \( 0 \) does not lie in \( K \). Denote by \( s \) the distance of \( 0 \) to \( K \), that is,

\[ s = \inf \{|y|, \quad y \in K \}. \tag{31} \]

Since \( 0 \) does not lie in \( K \), and since \( K \) is closed, \( s > 0 \). Let \( \{y_n\} \) be a minimizing sequence for (31), that is,

\[ y_n \text{ is in } K, \quad |y_n| = s_n \to s. \tag{31'} \]

Define the unit vectors \( x_n \) as

\[ x_n = \frac{y_n}{s_n}. \tag{31''} \]

we can write

\[ \frac{x_n + x_m}{2} = \frac{1}{2s_n} y_n + \frac{1}{2s_m} y_m \]

\[ = \left( \frac{1}{2s_n} + \frac{1}{2s_m} \right) (c_n y_n + c_m y_m). \tag{32} \]

Clearly, \( c_n \) and \( c_m \) are positive, and \( c_n + c_m = 1 \). Since \( K \) is convex, it follows that \( c_n y_n + c_m y_m \) belongs to \( K \). Therefore, by (31),

\[ |c_n y_n + c_m y_m| \geq s. \]
NORMED LINEAR SPACES

Setting this into (32), we get that

\[
\frac{|x_n + x_m|}{2} \geq \frac{s}{2s_n} + \frac{s}{2s_m}.
\]

(33)

Since \( \{y_n\} \) is a minimizing sequence for (31), \( s_n \rightarrow s \); therefore the right side of (33) tends to 1. So it follows from (33), (30), and (29) that \( \lim_{n,m \rightarrow \infty} |x_n - x_m| = 0 \).

It follows then from (31') that also \( \lim_{n,m \rightarrow \infty} |y_n - y_m| = 0 \), meaning that the minimizing sequence \( \{y_n\} \) is a Cauchy sequence. Since \( X \) is complete and \( K \) is closed, the sequence \( \{y_n\} \) converges to an element \( y \) of \( K \). Clearly, \( |y| = s \).

The power of theorem 8 lies in the fact that it asserts the existence of a minimum, when the set \( K \) over which we wish to minimize is not compact; according to theorem 6, a Banach space has many closed, bounded sets \( K \) that are not compact.

The notion of uniform convexity is due to Clarkson, as is the result that the \( L^p \) spaces are uniformly convex for \( 1 < p < \infty \).

We give now an example that shows that in the space \( C \), which is not uniformly convex—indeed, the maximum norm is not even strictly subadditive—the conclusion of theorem 8 fails.

We take \( X \) to be \( C[-1, 1] \), the space of continuous real-valued functions defined on the closed interval \(-1 \leq t \leq 1\). We take \( K \) to consist of all functions \( k(t) \) that satisfy

\[
\int_{-1}^{0} k \, dt = 0, \quad \int_{0}^{1} k \, dt = 0.
\]

(34)

\( K \) is a linear subspace and therefore convex, and clearly closed.

We take for \( z(t) \) any function in \( C \) for which

\[
\int_{-1}^{0} z \, dt = 1, \quad \int_{0}^{1} z \, dt = -1.
\]

It follows from (34) that for any \( k \) in \( K \)

\[
\int_{-1}^{0} (z - k) \, dt = 1, \quad \int_{0}^{1} (z - k) \, dt = -1.
\]

From this it follows that

\[
\max_{-1 \leq t \leq 0} [z(t) - k(t)] \geq 1.
\]

(35)

equality holding iff

\[
z(t) - k(t) = 1 \quad \text{for} \quad -1 \leq t \leq 0,
\]

(35')

and similarly that

\[
\min_{0 \leq t \leq 1} [z(t) - k(t)] \leq -1,
\]

equality holding iff

\[
z(t) - k(t) \equiv -1 \quad \text{for} \quad 0 \leq t \leq 1.
\]

(36)

Conditions (35') and (36') cannot both hold at \( t = 0 \), so it follows that in at least one of (35) or (36) inequality holds. This proves that

\[
|z - k|_{\max} > 1
\]

(37)

for any \( k \) in \( K \). On the other hand, one could choose \( k \) in \( K \) so that the max and min in (35) and (36) are as close to 1 and -1, respectively, as one wishes. So

\[
\inf_{k \in K} |z - k|_{\max} = 1.
\]

(37')

This combined with (37) shows that there is no closest point to \( z \) in \( K \).

5.3 ISOMETRIES

We turn now to isometries of a Banach space \( X \) onto itself, meaning mappings \( M \) of \( X \) onto \( X \) which preserve the distance of any pair of points:

\[
|M(x) - M(y)| = |x - y| \quad \text{for all} \quad x, y \text{ in } X.
\]

(38)

Clearly, translations \( M(x) = x + u, u \text{ fixed}, \) are isometries. Also the isometries of \( X \) form a group. We want to investigate those isometries that map 0 into 0; all others can be obtained by composing these with a translation.

**Theorem 9.** Let \( X \) be a linear space over the reals with a strictly subadditive norm. Let \( M \) be an isometric mapping of \( X \) into itself that maps the origin into itself. Then \( M \) is linear.

**Proof.** Denote for simplicity \( M(x) \) by \( x' \). Take any pair of points \( x \) and \( y \) and define

\[
\varepsilon = \frac{x + y}{2}.
\]

(39)

Using isometry as stated in (38), and the definition of \( \varepsilon \), we have

\[
|x' - \varepsilon'| = |x - \varepsilon| = \frac{|x - y|}{2},
\]

\[
|\varepsilon' - y'| = |\varepsilon - y| = \frac{|x - y|}{2},
\]

(40)
and

\[ |x' - y'| = |x - y|. \tag{40'} \]

These imply that

\[ |x' - y'| = |x' - z' + z' - y'| = |x' - z'| + |z' - y'|. \]

Since the norm is strictly subadditive, \( x' - z' \) and \( z' - y' \) must be positive multiples of each other. Since by (40) they have the same norm, they must be equal: \( x' = z' = z' - y' \). Hence

\[ 2z' = z' + y'. \tag{41} \]

\[ \square \]

**Exercise 7.** Deduce from (41) that \( M \) is linear.

It is a fact of life that some Banach spaces are very rich in isometries; others are very poor. Among the rich ones are the Hilbert spaces discussed in chapter 6; among the poor ones are function spaces with the max norm. Here is an example due to Schur:

Denote by \( X \) the space of null sequences of complex numbers

\[ x = [a_n], \quad \lim_{n \to \infty} a_n = 0 \tag{42} \]

normed by

\[ |x| = \max_n |a_n|. \tag{42'} \]

**Exercise 8.** Show that \( X \) is complete.

Let \( \{b_n\} \) be an arbitrary sequence of complex numbers of absolute value 1:

\[ |b_n| = 1. \]

Define the mapping \( U \) by

\[ Ux = [b_na_n]. \tag{43} \]

Clearly, \( U \) is a linear map of \( X \) onto \( X \), and satisfies \(|UX| = |x|\); thus \( U \) is an isometry.

Let \( p \) be a permutation of the positive integers. Define the map \( P \) by

\[ Px = [a'_n], \quad a'_n = a_{p(n)}. \tag{44} \]

Clearly, \( P \) is a linear map of \( X \) onto \( X \), and an isometry.

**Theorem 10.** Every linear isometry of the Banach space \( X \) defined by (42), (42') is the composite of an isometry of type (43) and (44).

**Proof.** Let \( u_j \) be a \( j \)th unit vector, i.e., a vector whose \( j \)th component has absolute value 1, all others are zero. Denote by \( T_j \) the linear subspace of \( X \) consisting of all vectors whose \( j \)th component is zero. Clearly,

\[ T_j \text{ is closed and codim } T_j = 1, \tag{45} \]

\[ |u_j + t| = 1 \quad \text{for all } t \in T_j, |t| \leq 1. \tag{46} \]

Conversely, we have

**Lemma 11.** Let \( u \) be a vector in \( X \), \(|u| = 1\), and \( T \) a subspace of \( X \) of codimension 1 so that (45) and (46) hold. Then \( u \) is a unit vector and \( T \) the corresponding subspace \( T_j \).

**Proof.** By definition (42') of the norm

\[ 1 = |u| = |u_m| \quad \text{for some index } m. \]

It follows from (46) that no vector \( t \) in \( T \) can have an \( m \)th component \( \neq 0 \). Since \( T \) is assumed to have codimension 1, it follows that \( T \) consists of all null sequences whose \( m \)th component is zero. From this it follows, by (46), that all components of \( u \) other than the \( m \)th must be zero.

Let \( M \) be a linear isometry of \( X \) onto \( X \), let \( u_j \) be any unit vector and \( T_j \) the corresponding subspace. Since \( M \) is linear, isometric, and onto, it follows that \( u'_j \) and \( T'_j \), the image of \( u_j \) and \( T_j \) under \( M \), satisfy (45) and (46). Then by Lemma 11, \( u' \) is a unit vector; from this and the linearity of isometry theorem 10 follows readily.

We conclude this chapter with the following result due to Mazur and Ulam:

**Theorem 12.** Let \( X \) and \( X' \) be two normed linear spaces over their reals, \( M \) an isometric mapping of \( X \) onto \( X' \) that carries 0 into 0. Then \( M \) is linear.

**Proof.** The case where the norms are strictly subadditive is covered in theorem 9. In the general case we take, as before, \( x \) and \( y \) to be any pair of points, and \( z \) their midpoint:

\[ z = \frac{x + y}{2}. \]

As before, in (40), \( z \) is halfway between \( x \) and \( y \), but when the norm in \( X \) is not strictly subadditive, this no longer characterizes the midpoint \( z \). There may be other points \( u \) also halfway between \( x \) and \( y \):

\[ |x - u| = |y - u| = \frac{|x - y|}{2}. \tag{47} \]
We denote the set of all such $u$ by $A$. We claim that this set $A$ is symmetric with respect to the midpoint $z$. That is, that if $u$ belongs to $A$, then so does

$$v = 2z - u.$$  (48)

To see this, we note that $2z = x + y$, and so

$$v = x + y - u,$$

and

$$v - y = x - u.$$

It follows from (47) that $v$ is halfway between $x$ and $y$.

We define the diameter $d_A$ of $A$ as the greatest distance between pairs of points of $A$:

$$d_A = \sup_{u, w \in A} |u - w|.$$  (49)

Since $A$ is symmetric with respect to $z$, for all $u$ in $A$,

$$|u - z| \leq \frac{1}{2} d_A.$$  (50)

Of course, there may be other points $p$ in $A$ with this property:

$$|u - p| \leq \frac{1}{2} d_A \quad \text{for all } u \in A.$$  (50')

We denote the set of all such $p$ by $A_1$. We claim that $A_1$ is symmetric with respect to the midpoint $z$. That is, if $p$ belongs to $A_1$, so does

$$q = 2z - p.$$  (51)

For using (48), we can write for any $u$ in $A$,

$$q - u = 2z - u - p = v - p.$$  (51')

We conclude from (51') that $|q - u| = |v - p|$. Since $v$ belongs to $A$ when $u$ does, it follows from (50) that $|u - q| \leq \frac{1}{2} d_A$.

It follows from (50) that the diameter of $A_1$ does not exceed half the diameter of $A$:

$$d_{A_1} \leq \frac{1}{2} d_A.$$  (52)

We now repeat this construction, obtaining a nested sequence of sets $A \supset A_1 \supset A_2 \cdots$, each containing the midpoint $z$, each symmetric with respect to $z$, and their diameters satisfying

$$d_{A_{n+1}} \leq \frac{1}{2} d_{A_n}.$$  (53)

Clearly, $d_{A_n}$ tends to zero; it follows that the intersection of all the sets $A_n$ consist of the single point $z$. This characterizes the midpoint $z$ of $x$, $y$ purely in terms of the metric structure of $X$.

BIBLIOGRAPHY

Let $M$ be an isometric mapping of $X$ onto $X'$. Then the inverse of $M$ maps $X'$ isometrically onto $X$. Denote by $x'$ and $y'$ the images of $x$, $y$ under $M$, and denote by $A'$, $A'_1$, 


It is the first property of linearity, see equation (1) in chapter 2. From this we deduce that $M(kx) = kM(x)$ for all rational $k$. Since $M$ is an isometry, it is continuous, and so the relation holds for all real $k$. \[\Box\]

BIBLIOGRAPHY


6

HILBERT SPACE

6.1 SCALAR PRODUCT

A scalar product in a linear space X over R is a real valued function of two points x and y in X, denoted as \( (x, y) \), having the following properties:

(i) Bilinearity. For fixed y, \( (x, y) \) is a linear function of x, for fixed x a linear function of y.

(ii) Symmetry, \( (y, x) = (x, y) \).

(iii) Positivity, \( (x, x) > 0 \) for \( x \neq 0 \).

When the field of scalars is C, \( (x, y) \) is complex valued, and properties (i) and (ii) are altered as follows:

(i) Sesquilinearity. For fixed y, (1) is a linear function of x and for x fixed (1) is a skewlinear function of y, that is,

\[
(ax, y) = a(x, y), \quad (x, ay) = \overline{a}(x, y).
\]

(ii) Skew symmetry,

\[
(y, x) = \overline{(x, y)}. \tag{2}
\]

Given a scalar product, we can define a norm, denoted by \( \| \| \), as follows:

\[
\|x\| = (x, x)^{1/2}. \tag{3}
\]

We claim that \( \| \| \) has the obvious properties of a norm:
Positivity follows from (iii), and homogeneity from (1). To show subadditivity we need

**Theorem 1 (Schwarz Inequality).** A scalar product satisfying (i), (ii), and (iii) satisfies

\[
|(x, y)| \leq \|x\| \|y\|, \tag{4}
\]

where the norm is defined by (3). Equality holds for \( x = ay \) or \( y = 0 \).

**Proof.** Let \( t \) be a real scalar and \( y \neq 0 \). Using bilinearity and skew symmetry, we can write

\[
\|x + ty\|^2 = \|x\|^2 + 2t \text{Re}(x, y) - t^2 \|y\|^2. \tag{5}
\]

By (iii), this is nonnegative. Set \( t = -\text{Re}(x, y)/\|y\|^2 \) and multiply by \( \|y\|^2 \). We get

\[
(\text{Re}(x, y))^2 \leq \|x\|^2 \|y\|^2.
\]

Replacing \( x \) by \( a x \), |a| = 1 so chosen that \( a(x, y) \) is real, we deduce (4). Note that equality holds in (4) if \( x \) and \( y \) are scalar multiples of one another.

**Corollary 1.** For every vector \( x \) in a scalar product space

\[
\|x\| = \max_{\|y\|=1} |(x, y)|.
\]

Now we are ready to prove that the norm is subadditive. Set \( t = 1 \) in (5) and estimate the middle term by (4). We get

\[
\|x + y\|^2 \leq (\|x\| + \|y\|)^2,
\]

which is subadditivity of the norm.

We set \( t = \pm 1 \) in (5) and add to obtain the parallelogram identity:

\[
\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2. \tag{6}
\]

**Exercise 1.** Show that a norm that satisfies (6) comes from a scalar product, an observation due to von Neumann.

**Exercise 2.** Show that the scalar product depends continuously on its factors; that is, if \( x_n \to x \), \( y_n \to y \) in the sense of \( \|x_n - x\| \to 0 \), \( \|y_n - y\| \to 0 \), then \( (x_n, y_n) \) tends to \( (x, y) \). (Use the Schwarz inequality.)

**Definition.** Two vectors \( x \) and \( y \) are called orthogonal if \( (x, y) = 0 \).

**Definition.** A linear space with a scalar product that is complete with respect to the induced norm is called a Hilbert space.

Given a linear space with a scalar product, it can be completed with respect to the norm derived from the scalar product. It follows from the Schwarz inequality that the
scalar product is a continuous function of its factors; therefore it can be extended to the completed space. Thus the completion is a Hilbert space.

We give some examples of linear spaces with inner product:

**Example 1.** The space of continuous functions \( x(t) \) on the interval \([0, 1]\), with
\[
(x, y) = \int_0^1 x(t) \overline{y(t)} \, dt.
\]

This space is incomplete.

**Example 2.** The space \( \ell^2 \) of vectors with infinitely many components:
\[
x = (a_1, a_2, \ldots), \quad y = (b_1, b_2, \ldots)
\]
subject to the restriction
\[
\sum |a_j|^2 < \infty, \quad \sum |b_j|^2 < \infty.
\]
We define the scalar product as
\[
(x, y) = \sum a_j \overline{b_j}.
\]

**Exercise 3.** Show that \( \ell^2 \) is complete.

**Example 3.** The space \( L^2 \) of all functions square integrable according to Lebesgue on some domain in \( \mathbb{R}^n \). This space is complete.

Many other examples will come up in the applications presented in subsequent chapters.

### 6.2 Closest Point in a Closed Convex Subset

**Theorem 2.** Given a nonempty closed, convex subset \( K \) of a Hilbert space \( H \), and a point \( x \in H \), there is a unique point \( y \) in \( K \) that is closer to \( x \) than any other point of \( K \).

**Proof.** Define
\[
\inf_{z \in K} \| x - z \| = d.
\]
Let \( y_n \) in \( K \) be a minimizing sequence:
\[
\lim d_n = d, \quad d_n = \| x - y_n \|.
\]

We apply the parallelogram identity (6) to \( x = (x - y_n)/2, y = (x - y_m)/2 \):
\[
\| x - \frac{y_n + y_m}{2} \|^2 + \frac{1}{4} \| y_n - y_m \|^2 = \frac{1}{2} (d_n^2 + d_m^2).
\]
Since \( K \) is convex, \( (y_n + y_m)/2 \) belongs to \( K \), and so by (7), \( \| x - (y_n + y_m)/2 \| = \| y_n \| \geq d \). Using this and (8) in (9), we deduce that \( y_n \) is a Cauchy sequence. Since \( H \) is complete and \( K \) closed, \( y = \lim y_n \) belongs to \( K \). Since \( \| x - y \| = \lim \| x - y_n \| = d \), \( y \) minimizes the distance from \( x \). That there is only one minimizer follows from (6); suppose that \( y' \) is another minimum, and apply (6) to \( x - y, x - y' \).

**Definition.** Let \( Y \) be a linear subspace; its **orthogonal complement** consists of all vectors \( v \) orthogonal to \( Y \), that is, satisfying \( (v, y) = 0 \). It is denoted as \( Y^\perp \).

**Theorem 3.** Let \( H \) be a Hilbert space, \( Y \) a closed subspace of \( H \), \( Y^\perp \) the orthogonal complement of \( Y \). We claim that

(i) \( Y^\perp \) is a closed linear subspace of \( H \);
(ii) \( Y \) and \( Y^\perp \) are complementary subspaces, meaning that every \( x \) can be decomposed uniquely as a sum of a vector in \( Y \) and in \( Y^\perp \);
(iii) \( (Y^\perp)^\perp = Y \).

**Proof.** It follows from the bilinearity of scalar product that the set of vectors \( v \) orthogonal to all vectors of any set \( Y \) form a linear space. This shows that \( Y^\perp \) is a linear space. Let \( \{v_j\} \) be a convergent sequence of elements of \( Y^\perp \):
\[
\lim v_j = v.
\]
We claim that \( v \) belongs to \( Y^\perp \), namely, that
\[
(v, z) = 0 \quad \text{when } z \text{ is in } Y.
\]
Since \( v_j \) belongs to \( Y^\perp \),
\[
(v, z) = (v - v_j, z) = (v - v_j, z).
\]
By the Schwarz inequality applied on the right,
\[
|v, z| \leq \| v - v_j \| \| z \|.
\]
by (10), \( \| v - v_j \| \) tends to zero. So (11) shows that \( (v, z) = 0 \), meaning that \( Y^\perp \) is closed, as asserted in (i).
We turn now to (ii). Given any \( x \) in \( H \), there is, according to theorem 2, a vector \( y \) in \( Y \) closest to \( x \). Set
\[
v = x - y.
\]
(12)

The minimum property of \( y \) means that for any \( z \) in \( Y \) and any real \( t \),
\[
\|v\|^2 \leq \|v + tz\|^2.
\]

Using (5), we can rewrite the right side as \( \|v\|^2 + 2t \text{Re}(v, z) + t^2\|z\|^2 \) and conclude that
\[
\text{Re}(v, z) = 0 \quad \text{for all } z \text{ in } Y.
\]
(13)

This shows that \( v \) belongs to \( Y^\perp \); (12) gives the decomposition of \( x \) as \( y + v \), the sum of a vector from \( Y \) and one from \( Y^\perp \).

This decomposition is unique, for if \( x = y + v = y' + v' \), then \( y - y' = v' - v \) would belong to both \( Y \) and \( Y^\perp \) and thus would be orthogonal to itself. But then, by positivity, \( y - y' = v' - v = 0 \). Thus (ii) is proved. Part (iii) is an immediate consequence of (ii).

\[\square\]

**COMMENT.** It follows from theorem 3 that every closed linear subspace of a Hilbert space has a closed complement. This is not true for all Banach spaces; examples will be given later.

### 6.3 LINEAR FUNCTIONALS

A Hilbert space comes equipped with a whole set of built-in linear functionals. For \( y \) fixed, \( \ell(x) = \langle x, y \rangle \) is a linear functional of \( x \), that is, a linear mapping of \( H \) into \( \mathbb{C} \). Furthermore, according to the Schwarz inequality (4), \( \ell(x) \) is bounded by a constant multiple of \( \|x\| \). It turns out that conversely:

**Theorem 4.** Let \( \ell(x) \) be a linear functional on a Hilbert space \( H \) that is bounded:
\[
|\ell(x)| \leq \text{const.} \|x\|.
\]
(14)

Then \( \ell \) is of the form
\[
\ell(x) = \langle x, y \rangle, \quad y \text{ in } H.
\]
(15)

The point \( y \) is uniquely determined.

**Proof.** We will use the following facts:

**Lemma 5.**

(i) The nullspace of a linear functional that is not \( \equiv 0 \) is a linear subspace of codimension 1.

(ii) If two linear functionals \( \ell \) and \( m \) have the same nullspace, they are constant multiples of each other:
\[
\ell = cm.
\]
(16)

(iii) The nullspace of a linear functional that is bounded in the sense of (14) is a closed subspace.

**Exercise 4.** Prove lemma 5.

Note that lemma 5 holds in any Banach space, and parts (i) and (ii) in any linear space.

Suppose now that \( \ell \equiv 0 \). Then its nullspace is a closed subspace \( Y \) of \( H \) of codimension 1. Its orthogonal complement \( Y^\perp \) (see theorem 3) is one dimensional. Let \( p \) be any nonzero vector in \( Y^\perp \), and define the linear functional \( m \) by
\[
m(x) = \langle x, p \rangle.
\]

Clearly, the nullspace of \( m \) is \( Y \). So by (16) of part (ii) of lemma 5,
\[
\ell(x) = cm(x) = \langle x, \overline{p} \rangle.
\]

**Theorem 4** is called the Riesz-Fréchet representation theorem. The following useful generalization has been given by Milgram and Lax:

**Theorem 6 (Lax-Milgram Lemma).** Let \( H \) be a Hilbert space, and \( B(x, y) \) a function of two vectors with the following properties:

(i) \( B(x, y) \) is for fixed \( y \) a linear function of \( x \), for fixed \( x \) a skewlinear function of \( y \).

(ii) \( B \) is bounded: there is a constant \( c \) so that for all \( x \) and \( y \) in \( H \)
\[
|B(x, y)| \leq c\|x\|\|y\|.
\]
(17)

(iii) There is a positive constant \( b \) such that
\[
|B(y, y)| \geq b\|y\|^2.
\]
(18)

for all \( y \) in \( H \).

**Assertion.** Every linear functional \( \ell \) on \( H \) that is bounded in the sense of (14) is of the form
\[
\ell(x) = B(x, y), \quad y \text{ a uniquely determined vector in } H.
\]
(19)
HILBERT SPACE

Proof. By (i) and (ii), for y fixed B(x, y) is a bounded linear functional of x. Therefore by theorem 4 it can be written as

\[ B(x, y) = (x, z). \quad z \text{ in } H. \]  

(20)

Since z is uniquely determined by y, it is a function of y. It follows from (20) that the relation of z to y is linear; it follows from this that the set of z appearing in (20) as y takes on all values in H is a linear subspace of H. We claim that it is a closed linear subspace. To see this, set x = y in (20):

\[ B(y, y) = (y, z). \]  

(20')

Using (18) on the left and the Schwarz inequality on the right, we get after dividing by \( \|y\| \) that

\[ b\|y\| \leq \|z\|. \]  

(21)

Let \( \{z_n\} \) be a sequence of vectors appearing in (20), with corresponding \( y_n \):

\[ B(x, y_n) = (x, z_n). \]  

(22)

Subtraction and skew linearity gives \( B(x, y_n - y_m) = (x, z_n - z_m) \). By (21), \( \|y_n - y_m\| \leq \|z_n - z_m\|. \) From this it follows that if \( z_n \) converges to \( z \), the corresponding \( y_n \) form a Cauchy sequence. Since H is complete, the sequence \( \{y_n\} \) converges to a limit \( y \). It follows from (17) that the left side of (22) converges to \( B(x, y) \), and it follows from (4) that the right side converges to \( (x, z) \). So

\[ B(x, y) = (x, z), \]

which proves that the set of \( z \) appearing in (20) form a closed subspace of H.

We claim that this closed subspace is all of H; for if not, then according to theorem 3 there would be a nonzero vector x orthogonal to all \( z \). It follows from (20) that such an x satisfies \( B(x, y) = 0 \) for all y. Setting \( y = x \) gives \( B(x, x) = 0 \); using (18), we get \( \|x\| = 0 \), contrary to \( x \neq 0 \).

According to theorem 4, all linear functionals \( \ell(x) \) can be represented as \( (x, z) \), \( z \) in H. Combined with (20), this establishes (19). It follows from (18) that y is uniquely determined. \( \square \)

6.4 LINEAR SPAN

We recall from chapter 1 that the linear span of a collection of points \{y_j\} is the smallest linear subspace containing them. The closed linear span of a collection of points S in a Hilbert space H is defined to be the smallest closed linear subspace containing S, that is, the intersection of all such subspaces.

Exercise 5. Show that the closed linear span of a set is the closure of its linear span.

LINEAR SPAN

Theorem 7. The point y of a Hilbert space H belongs to the closed linear span Y of the set \{y_j\} if every vector z that is orthogonal to all \( y_j \) is orthogonal also to y:

\[ (y, z) = 0 \quad \text{for all } z \text{ that satisfy } (y_j, z) = 0 \quad \text{for all } j. \]  

(23)

Proof. We claim that the set Z of vectors z orthogonal to all \( y_j \) form the orthogonal complement of Y. Since every vector z orthogonal to all \( y_j \) is orthogonal to all linear combinations of \( y_j \) and by continuity to limits of linear combinations, \( Z \subset Y^\perp \). Conversely, every vector in \( Y^\perp \) is orthogonal to all the \( y_j \), and so belongs to Z. This shows that \( Z = Y^\perp \). We appeal to (iii) of theorem 3 to conclude that \( Y = (Y^\perp)^\perp = Z^\perp \), as asserted in theorem 7. \( \square \)

We stated in chapter 5 that every isometry of a Banach space onto itself that maps 0 into 0 is linear. We give now a new proof of this in a Hilbert space:

Denote by \( x \rightarrow x' \) an isometry of a Hilbert space that maps 0 \rightarrow 0. Let \( x, y \) be any pair of vectors, \( x' \) and \( y' \) their images. Since distances are preserved, \( d(0, x) = d(0, x') \), \( d(0, y) = d(0, y') \), and \( d(x, y) = d(x', y') \), which can be expressed as

\[ \|x\| = \|x'\|, \quad \|y\| = \|y'\|. \]  

(24)

\[ \|x - y\|^2 = \|x' - y'\|^2. \]  

(24')

Expanding both sides in (24') and using (24), we get

\[ (x, y) = (x', y'). \]  

(25)

Now denote \( x + y \) by \( z \), and let \( u \) be any vector in H. Using (25) we have

\[ (z', u') = (z, u) = (x + y, u) = (x, u) + (y, u) = (x', u') + (y', u') = (x' + y', u'). \]

Thus

\[ (z' - x' - y', u') = 0 \]

for all \( u' \). This can be only if \( z' = x' + y' \). \( \square \)

The virtue of this proof is that it applies even when the scalar product is not positive, as long as it is nondegenerate, meaning that no \( u \) is orthogonal to all points.

We turn now to orthonormal sets:

Definition. A collection of vectors in an inner product space, \( \{x_j\} \), is called orthonormal if

\[ (x_j, x_k) = 0 \quad \text{for } j \neq k, \quad \text{and} \quad \|x_j\| = 1 \quad \text{for all } j. \]  

(26)

Definition. A collection of vectors \( \{x_j\} \) is called an orthonormal base if the vectors are orthonormal, and if the closed linear span of \( \{x_j\} \) is the whole space.
Lemma 8. Let $H$ denote a Hilbert space, $\{x_j\}$ an orthonormal set in $H$. The closed linear span of $\{x_j\}$ consists of all vectors of the form

$$x = \sum a_j x_j,$$

(27)

where the $a_j$ are complex numbers so chosen that

$$\sum |a_j|^2 < \infty.$$

(27')

The sum (27) converges in the sense of the Hilbert space norm. Furthermore

$$\|x\|^2 = \sum |a_j|^2,$$

(28)

and

$$a_j = (x, x_j).$$

(28')


Theorem 9. Every Hilbert space contains an orthonormal basis.

Proof. Consider all orthonormal sets, partially ordered by inclusion. Given a totally ordered collection, the union of all vectors contained in the sets in the collection includes all of them. Therefore, by Zorn's lemma, there is an orthonormal set that is maximal. We claim that the closed linear span $X$ of a maximal orthonormal set $\{x_j\}$ is the whole space. We argue indirectly: suppose that there is a $y$ that does not belong to the closed linear span $X$. Define $a_j$ by

$$a_j = (y, x_j).$$

(29)

We claim that Bessel's inequality holds:

$$\sum |a_j|^2 \leq \|y\|^2;$$

(30)

for consider

$$\|y - \sum_F a_j x_j\|^2,$$

(31)

where $F$ means a finite collection of $j$. Using the orthonormality of $\{x_j\}$, we find that (31) equals

$$\|y\|^2 - \sum_F a_j (y, x_j) - \sum_F a_j (x_j, y) + \sum_F |a_j|^2,$$

which by (29) equals

$$\|y\|^2 - \sum_F |a_j|^2.$$

Since (31) is nonnegative, (30) follows for every finite collection $F$, and therefore for the infinite sum.

It follows then from lemma 8 that we can define a vector $x$ by (27), and that $x$ belongs to $X$. Now using (29) and (28'), we have

$$(y - x, x_j) = (y, x_j) - (x, x_j) = a_j - a_j = 0,$$

meaning that the $y - x$ is orthogonal to all $x_j$. The difference $y - x$ is not zero, since by assumption $y$ does not belong to $X$, while $x$ does; so

$$\frac{y - x}{\|y - x\|}$$

could be joined to the orthonormal set $\{x_j\}$, enlarging it, contradicting maximality.

Suppose that $H$ is a separable Hilbert space; that is, it contains a denumerable set of points that is dense. In this case every orthogonal basis is denumerable, and the basis elements can be constructed without appealing to transcendental arguments such as Zorn's lemma:

Theorem 9'. Let $\{y_j\}$ be a sequence of vectors in a Hilbert space whose closed linear span is all of $H$. Then there exists an orthonormal basis $\{x_j\}$ such that the linear span of $\{x_1, \ldots, x_n\}$ contains $y_1, \ldots, y_n$.

Exercise 7. Prove theorem 9'.

The construction of the orthonormal basis $\{x_j\}$ in theorem 9' is called the Gram-Schmidt process.

Exercise 8. Let $H$ be a Hilbert space; prove that any two orthonormal bases in $H$ have the same cardinality.

Theorem 10. Let $H$ denote a Hilbert space, $\{x_j\}$ and $\{y_j\}$ two orthonormal bases. According to theorem 8, every $x$ can be written as

$$x = \sum a_j x_j, \quad a_j = (x, x_j).$$

Then the mapping

$$x \rightarrow y = \sum a_j y_j$$

is an isometry of $H$ onto $H$, mapping $0 \rightarrow 0$. Furthermore every isometry of $H$ onto $H$ mapping $0 \rightarrow 0$ can be obtained in this fashion.

Exercise 10. Show that every infinite-dimensional separable Hilbert space is isomorphic with the space $\ell^2$ consisting of all vectors with infinitely many components: $x = (a_1, a_2, \ldots)$, subject to the restriction $\|x\|^2 = \sum |a_i|^2 < \infty$.

NOTES. The abstract notion of Hilbert space described in this chapter is due to von Neumann in 1929. Earlier, Hilbert and his school had used the concrete spaces described in Examples 2 and 3; hence the name.

Theorem 2 is essentially due to Beppo Levi, in a concrete context.

### BIBLIOGRAPHY


### 7

#### APPLICATIONS OF HILBERT SPACE RESULTS

7.1 RADON-NIKODYM THEOREM

Let $\nu$ and $\mu$ be finite nonnegative measures on the same $\sigma$-algebra, $\nu$ is said to be absolutely continuous with respect to $\mu$ if every set that has $\mu$-measure zero has $\nu$-measure zero. The Radon-Nikodym theorem asserts that such a $\nu$-measure can be expressed as

$$\nu(E) = \int_E g \, d\mu, \quad (1)$$

where $g$ is a nonnegative integrable function with respect to $\mu$.

Von Neumann showed how to derive this from the Riesz representation theorem for linear functionals in Hilbert space:

Let $H$ be the real Hilbert space $L^2(\mu + \nu)$, with the norm

$$\|x\|^2 = \int x^2 \, d(\mu + \nu). \quad (2)$$

Assume, for simplicity, that the $\mu$ and $\nu$ measure of the whole space is finite; then it follows, via the Schwarz inequality, that every square integrable function is integrable. The linear functional

$$\ell(x) = \int x \, d\mu \quad (3)$$

is bounded with respect to the $L^2(\mu)$-norm, so even more with respect to the $L^2(\mu + \nu)$-norm. Then, by theorem 4 of chapter 6, $\ell(x)$ can be represented as a scalar product $(x, y)$ for some $y$ in $L^2(\mu + \nu)$:

$$\int x \, d\mu = \int x y \, d(\mu + \nu);$$

63
y depends only on the measures \( \mu \) and \( \nu \). We rewrite this as

\[ \int x(1 - y) \, d\mu = \int x \, d\nu. \tag{4} \]

We claim that

\[ 0 < y \leq 1 \tag{5} \]

except for a set of \( \mu \)-measure zero. To show this, we denote by \( F \) the set on which \( y \leq 0 \), and claim that

\[ \mu(F) = 0. \tag{6} \]

Set \( x = 1 \) on \( F \), \( x = 0 \) off \( F \); with this choice, (4) becomes

\[ \int_F (1 - y) \, d\mu = \int_F y \, d\nu. \tag{7} \]

Since \( y \leq 0 \) on \( F \), the right side of (7) is \( \leq 0 \), while the left side is \( \geq \mu(F) \); this proves (6).

Denote by \( G \) the set where \( y > 1 \); suppose that \( \mu(G) > 0 \). Set \( x = 1 \) on \( G \), \( x = 0 \) off \( G \); with this choice (4) becomes

\[ \int_G (1 - y) \, d\mu = \int_G y \, d\nu. \tag{8} \]

Since \( y > 1 \) on \( G \), the left side of (8) is negative, and the right side is positive, a contradiction. This completes the proof of (5).

We modify, if necessary, the function \( y \) on a set of \( \mu \)-measure 0 so that (5) holds everywhere. Since \( \nu \) is absolutely continuous with respect to \( \mu \), this does not affect (4).

We claim that the function \( g \) in (1) is given by \( g = (1 - y)/y \). To see this, denote \( u = xy \), and rewrite (4) as

\[ \int u \, g \, d\mu = \int u \, d\nu. \tag{9} \]

Let \( E \) be any measurable set; we choose \( x \) so that \( u = 1 \) on \( E \), \( 0 \) off \( E \). Then (9) gives

\[ \int_E g \, d\mu = \nu(E). \tag{10} \]

This is relation (1).

**Exercise 1.** Prove the Radon-Nikodym theorem for measures that are only \( \sigma \)-finite.

### 7.2 Dirichlet's Problem

First, let \( D \) be a bounded domain in \( \mathbb{R}^n \). Denote by \( C^\infty_0(D) \) the space of real-valued infinitely differentiable functions \( f \) whose support is contained in a compact subset of \( D \). On the space \( C^\infty_0(D) \) we introduce two scalar products:

\[ \langle f, g \rangle_0 = \int_D f g \, d\rho \quad \text{and} \quad \langle f, g \rangle_1 = \int_D \sum f_j g_j \, d\rho, \tag{11} \]

where \( f_j = \frac{\partial f}{\partial x_j}, \) \( j = 1, \ldots, n \).

**Exercise 2.** Verify that \( C^\infty_0(D) \) is an inner product space under each of these scalar products.

The following inequality connecting these two norms is due to Zaremba:

**Lemma 1.** For all \( f \) in \( C^\infty_0(D) \),

\[ \| f \|_0 \leq d \| f \|_1, \tag{12} \]

where \( d \) is the width of \( D \).

**Proof.** Since \( f \) is zero on the boundary of \( D \), at any point \( x \) in \( D \),

\[ f(x) = \int_{x^b}^x f_1 \, dx_1 \]

where \( x^b \) is a boundary point of \( D \) with the same \( x_2, \ldots, x_n \) coordinate as \( x \). Applying the Schwarz inequality above gives

\[ f^2(x) \leq d \int |f_1|^2 \, dx_1. \]

Integrating this over \( D \) gives (12). \( \square \)

Denote by \( H^1_0 \) the completion of \( C^\infty_0(D) \) with respect to the norm \( \| \|_1 \), by \( H_0 \) its completion with respect to the norm \( \| \|_0 \).

**Lemma 2.** Every element \( v \) of \( H^1_0 \) belongs to \( H_0 \) and has partial derivatives \( v_j \) of first order that belong to \( H_0 \); these partial derivatives satisfy

\[ (z, v_j)_0 = -\langle \partial z/\partial x_j, v \rangle_0 \]

for any \( C^\infty_0 \) function \( z \). Furthermore formula (11) holds for \( f, g \) in \( H^1_0 \).
Proof. Let \( \{v^{(n)}\} \) be a sequence of \( C_0^\infty \) functions that tends to \( v \) in the \( \| \cdot \|_1 \) norm. That means that the first derivatives \( v^{(n)}_j \) converge in the \( \| \cdot \|_0 \) norm; we call these limits \( v_j \). By lemma 1, \( \{v^{(n)}\} \) converges in the \( \| \cdot \|_1 \) to a limit in \( H_0 \), which we identify with the limit \( v \) in \( H_0^1 \). Integration by parts gives relation (13) for \( v^{(n)} \) in place of \( v \); letting \( n \to \infty \) gives (13); relation (11) follows similarly. □

We claim that the identification of elements \( v \) in \( H_1^0 \) with elements of \( H_0 \) is one-to-one, that is, an embedding of \( H_0^1 \) in \( H_0 \). We have to show that if \( v \) is zero in \( H_0 \), then it is zero in \( H_0^1 \). Clearly, it follows from (12) that if \( v = 0 \) in \( H_0 \), then \( v_j = 0 \) in \( H_0 \) for all \( j \). This makes \( \nu = 0 \) in \( H_0^1 \).

Relation (13) asserts that \( v_j \) are the first partial derivatives of \( v \) in the sense of distributions (see Appendix B).

Let \( f \) be any element of \( H_0 \); define the linear functional \( \ell \) by

\[
\ell(u) = (u, f_0).
\]

By the Schwarz inequality, and inequality (12) of lemma 1,

\[
|\ell(u)| \leq \|f_0\| \|u\|_0 \leq d \|f_0\| \|u\|_1
\]

for all \( u \) in \( H_0^1 \). According to the Riesz-Frechet representation theorem, theorem 4 of chapter 6, the functional (14) can be represented as an inner product. That is, there exists \( v \) in \( H_0^1 \) such that

\[
(u, f_0) = (u, v)_1
\]

for all \( u \) in \( H_0^1 \). By definition (11) and lemma 2, with \( u_j = \partial u / \partial x_j \),

\[
(u, v)_1 = \sum (u_j, v_j)_0.
\]

Take now \( u \) to be \( C_0^\infty \). We can, using the theory of distributions, rewrite the right side of (17) as

\[
- \sum (u_j, v_j)_0 = -(u, \Delta v)_0,
\]

where \( u_j \) are the second partial derivatives of \( u \), and \( \Delta \) the Laplace operator, acting in the sense of distributions. Combining (17), (17'), and (16), we deduce that

\[
(u, f)_0 = -(u, \Delta v)_0
\]

for all \( u \) in \( C_0^\infty \). From this it follows that in the sense of distribution theory,

\[
f = -\Delta v.
\]

Thus \( v \) is a distribution solution of the inhomogeneous equation (18).

Next we show that by virtue of belonging to \( H_0^1 \), \( v(x) \) tends to zero in an average sense as \( x \) tends to the boundary of \( D \). The precise statement is lemma 3:

**Lemma 3.** Suppose that \( D \) is a domain in \( \mathbb{R}^2 \) whose boundary \( \partial D \) is a \( C^1 \) curve. For any point \( p \) on the boundary of \( D \), choose a coordinate system \( x_1, x_2 \), with \( p \) as the origin, and the positive \( x_1 \) axis perpendicular to the boundary of \( D \) and pointing inward. Denote by \( R(p, d) \) all points of \( D \) where \( x_1 < d \) and \( |x_2| < d \). Let \( v \) be any function in \( H_1^1 \). We claim that the mean value of \( |v| \) over \( R(p, d) \) tends to zero as \( d \) tends to zero.

**Proof.** Since the area of \( R(p, d) \) is proportional to \( d^2 \), the claim is that

\[
\int_R |v| \, dx \leq o(d^2).
\]

(19)

To deduce this, we need to estimate

\[
\int_R |f| \, dx
\]

for functions \( f \) in \( C_0^\infty (D) \). Integrate by parts with respect to \( x_1 \):

\[
\int_R |f| \, dx = \int_R |f_1| |d - x_1| \, dx \leq \int_R |f_1| |d - x_1| \, dx
\]

\[
\leq \left( \int_R (d - x_1)^2 \, dx \right)^{1/2} \leq d^2 \left( \int_R f_1^2 \, dx \right)^{1/2};
\]

(20)

in the second step we have used the Schwarz inequality. Now approximate \( v \) in the 1-norm by a sequence \( v^{(n)} \) of \( C_0^\infty (D) \) functions. The limit of (20) for \( f = v^{(n)} \) is

\[
\int_R |v| \, dx \leq d^2 \left( \int_R v_1^2 \, dx \right)^{1/2}.
\]

(21)

Since the integral of \( v_1 \) over \( R(d) \) tends to zero as \( d \) tends to zero, (19) follows from (21). □

We have thus succeeded in constructing a generalized function \( v \) that solves the differential equation (18) in the sense of distribution theory, and that vanishes on the boundary in a mean value sense.

The argument above can be extended to solve the Dirichlet problem in this generalized sense for any second-order partial differential equation that is self-adjoint and positive. We show now how the Lax-Milgram lemma can be used to extend the argument above to non-self-adjoint partial differential operators. For example, consider for any pair \( u, v \) in \( H_1^1 \) the functional \( B \) defined by

\[
B(u, v) = \int_D \left( \sum u_j v_j + \sum u v_j + u v \right) \, dx \, dy.
\]

(22)

Clearly, \( B \) is bilinear; it follows from lemma 3 that it is bounded. Estimating the middle term by the Schwarz inequality, we see that \( B \) is positive in the sense of
theorem 6 of chapter 6. It follows then from that theorem that the linear functional (14) can be represented in terms of \( B \), meaning that there exists a \( v \) in \( H^0 \) such that

\[
(u, f)_0 = B(u, v)
\]  (23)
for all \( u \) in \( H^0 \). Integrating by parts on the right in (23), we deduce that

\[
f = -\Delta v + \sum j_j + v
\]  (24)
in the sense of distributions.

We present now another method for solving the Dirichlet problem for the homogeneous Laplace equation

\[
\Delta v = 0 \quad \text{in } D
\]  (25)
whose value on the boundary is prescribed. This method exploits some special properties of harmonic functions and yields genuine solutions that satisfy the boundary condition in the usual sense. We assume that \( \partial D \) is once differentiable, and that the boundary values of \( u \) are also once differentiable. We can then construct a \( C^1 \) function \( f \) on \( D \cup \partial D \) that has the prescribed value on \( \partial D \), and we state the boundary condition thus:

\[
v = f \quad \text{on } \partial D.
\]  (26)

We reformulate the boundary value problem (25), (26): decompose \( f \) as

\[
f = v + \zeta,
\]  (27)
where \( v \) is harmonic and \( \zeta \) vanishes on \( \partial D \). When \( v \) and \( \zeta \) have continuous first partial derivatives up to the boundary, we can apply Green's formula:

\[
(v, \zeta)_1 = \int_D \sum_j v_j \zeta_j \, dx \, dy = - \int_D (\Delta v) \zeta \, dx \, dy + \int_{\partial D} \frac{\partial v}{\partial n} \zeta \, ds.
\]

Since by (25), \( \Delta v = 0 \) in \( D \), and by (26), \( \zeta = 0 \) on \( \partial D \), the right side = 0. In words, the space of harmonic functions and the space of functions that vanish on the boundary are orthogonal to each other in the \(( , )_1 \) scalar product.

Performing the decomposition (27) thus appears as the task of splitting \( f \) into the sum of two functions from two orthogonal function spaces. We show how this can be accomplished by appealing to theorem 3 of chapter 6. The scalar product defined by (11) for all \( f \) in \( C^\infty(D) \) is not positive because \((f, f)_1 = 0 \), not only for \( f \equiv 0 \) but for all constant functions \( f \). We overcome this slight blemish by considering two functions as equivalent if they differ by a constant.

Denote by \( H^1 \) the completion in the \( \| \|_1 \) norm of all \( C^1 \) functions on \( D \cup \partial D \); the space \( H^1 \) is a closed subspace of \( H^1 \). We apply now the orthogonal decomposition theorem, theorem 3 of chapter 6, to conclude that every \( f \) in \( H^1 \) can be decomposed uniquely as in (27), where \( \zeta \) in \( H^0 \), \( v \perp H^0 \). The condition \( v \perp H^0 \) asserts that

\[
(u, \zeta)_1 = \sum (u_j, v_j)_0 = 0
\]
for all \( u \) in \( H^0 \). Taking \( \zeta \) to be \( C^\infty \), we can integrate by parts in the sense of distributions to get

\[
0 = \sum (u_j, v_j)_0 = - \sum (u, v_j)_0 = -(u, \Delta v)_0,
\]
which implies that

\[
\Delta v = 0
\]
in the sense of distributions. It is a well-known result of Hermann Weyl that a function harmonic in the distribution sense is harmonic in the classical sense; a proof is provided in section 4 of Appendix B.

We claim that when \( f \) is continuous up to the boundary, \( \zeta(\partial) \) tends to zero as \( \partial \) approaches the boundary. Denote by \( d \) the distance of \( \partial \) to \( \partial D \), and let \( C \) be the circular disc with center \( q \) and radius \( r = d/2 \) and denote mean values over \( C \) bars: taking the mean value of (27), we obtain

\[
\bar{f}(\partial) = \bar{\zeta}(\partial) + \bar{v}(\partial).
\]  (28)

Since \( f \) is continuous up to the boundary, its mean value on \( C \) differs from its value at the center \( q \) of \( C \) by an amount \( w \) that tends to zero as \( d \rightarrow 0 \). According to the elementary theory of partial differential equations, the mean value of the harmonic function \( v \) on a circular disc \( C \) is equal to its value at the center of \( C \). So (28) can be rewritten as

\[
\bar{f}(\partial) = \bar{\zeta}(\partial) + v(\partial).
\]

Subtracting (27) from this, we conclude that

\[
w = \bar{\zeta}(\partial) - \zeta(\partial).
\]  (29)

We appeal now to lemma 3, according to which the mean value over \( R(p, d) \) of the absolute value of a function \( \zeta \) in \( H^1 \), tends to zero as the distance \( d \) to \( \partial D \) tends to zero. It follows that the mean value of \( \zeta \) over the disc \( C \) tends to zero. Combined with (29), we conclude that \( \zeta(\partial) \) itself tends to zero as \( \partial \) approaches the boundary. Using (27), we see that the harmonic function \( v \) is continuous up to the boundary, and its boundary value equals \( f \). Thus we succeed in constructing not a generalized, but a genuine solution of the Dirichlet boundary value problem.

Here is another way of looking at (27). We saw that the boundary value problem (25), (26) for the Laplace equation amounts to decomposing \( f \) as a sum of a harmonic function and of one vanishing on the boundary. We showed that these spaces are orthogonal to each other in the \(( , )_1 \) scalar product. The decomposition was accomplished by appealing to the orthogonal decomposition theorem, theorem 3 of
chapter 6. Here we show how to perform an orthogonal decomposition with respect to the subspace $V$ consisting of harmonic functions whose first derivatives are square integrable in $D$. It is an easy fact in the theory of harmonic functions that $V$ is complete in the $\| \cdot \|_1$-norm. So, according to theorem 3 of chapter 6, we can decompose any $f$ in $H_1$ as

$$f = v + z,$$

(30)

where $v$ is in the space $V$ of harmonic functions with square integrable first derivatives, and $z$ is orthogonal to $V$. Our aim is to show that $z$ vanishes on the boundary.

We take $D$ to lie in $\mathbb{R}^2$, and assume that the boundary of $D$ is twice differentiable; we assume $f$ to be twice differentiable. For any pair of points $p$ and $q$ in the plane, we define $k(p, q)$ as the fundamental singular solution of the Laplace equation:

$$k(p, q) = -\frac{1}{2\pi} \log |p - q|.$$

(31)

Suppose that $q$ lies in $D$. If we knew that the function $z$ vanished on the boundary of $D$, then by Green's formula, see section 4 of Appendix B, we would deduce that

$$z(q) = \int_D (z_x x + z_y y) \, dx \, dy,$$

where $(x, y)$ are the coordinates of $p$. However, we do not at this point know that $z$ vanishes on $\partial D$, and so we denote the function defined by the integral above as $u$:

$$u(q) = \frac{1}{2\pi} \int_D \left( z_x \frac{x - x}{|p - q|^2} + z_y \frac{y - y}{|p - q|^2} \right) \, dx \, dy,$$

(32)

where $x', y'$ denote the coordinates of $q$.

**Lemma 4.** If $\partial D$ is twice differentiable, $u(q)$ defined by (32) is continuous up to the boundary and vanishes there.

**Proof.** It is easy to show that $u$ is continuous inside $D$. Let $q$ be a point of $D$ near the boundary; denote the nearest boundary point by $b$. Since $\partial D$ is assumed to be twice differentiable, there are two circular disks $S$ and $\overline{S}$ with the same radius, $d$, tangent to $\partial D$ at $b$, $S$ contained in $D$ and $\overline{S}$ exterior to $D$. Denote by $\overline{q} = (\overline{x}, \overline{y})$ the image of $q$ under inversion across the circle bounding $S$. For $q$ near enough to $\partial D$, $\overline{q}$ lies in $\overline{S}$.

Since $\overline{q}$ lies outside $D$, $k(p, \overline{q})$ is a regular harmonic function in $D$. In particular, $k(p, \overline{q})$ belongs to $V$, and thus is orthogonal to $z$ in the $(\cdot, \cdot)_1$ scalar product. Therefore $u(q) = 0$, and we can write

$$u(q) = u(q) - u(\overline{q}) = \frac{1}{2\pi} \int_D \left( z_x \frac{x - x}{|p - q|^2} - \frac{\overline{x} - x}{|\overline{p} - \overline{q}|^2} \right) \, dx \, dy + z_y \left( \frac{y - y}{|p - q|^2} - \frac{\overline{y} - y}{|\overline{p} - \overline{q}|^2} \right) \, dx \, dy.$$

(33)

**BIBLIOGRAPHY**

As $q$ approaches the boundary point $b$, so does the point $\overline{q}$. Therefore the integrand on the right in (33) tends to zero uniformly at all points of $D$ whose distance from $b$ exceeds any positive quantity $r$. It can be shown (see Lax for details) that also the integral over the remaining portion of $D$ tends to zero. This completes the proof of lemma 4.

**Lemma 5.** The function $u$ defined in $D$ by (32) is twice differentiable in $D$, and

$$\Delta u = \Delta z.$$

(34)

**Proof.** Since $f$ was assumed twice differentiable, and since by (30), $z$ differs from $f$ by a harmonic function $v$, it follows that $z$ is twice differentiable in $D$ and that $\Delta z = \Delta f$. Let $D'$ be a subdomain of $D$ that contains $q$, and whose closure is contained in $D$. We split the integral on the right (32) and integrate by parts over $D'$:

$$u(q) = z(q) + \int_{\partial D'} z_x \frac{\partial}{\partial n} ds + \int_{D' - \overline{D}} (z_x x + z_y y) \, dx \, dy.$$

(35)

The two integrals on the right are harmonic functions in $D'$; therefore, since $D'$ is arbitrary,

$$u = z + h,$$

(36)

$h$ harmonic. This proves lemma 5.

Express $z$ from (36) as $u - h$ and set into (30):

$$f = v - h + z.$$

Since $v - h$ is harmonic, and $u = 0$ on $\partial D$, $v - h$ solves the boundary value problem (25), (26).

The preceding proof is a reworking of an argument of Garabedian and Schiffer.

**BIBLIOGRAPHY**


8

DUALS OF NORMED LINEAR SPACES

8.1 BOUNDED LINEAR FUNCTIONALS

In this chapter we deal with normed linear spaces \( X \) over the real or the complex numbers. We will study linear functionals, namely mappings \( \ell \) of \( X \) into \( \mathbb{R} \) or \( \mathbb{C} \) satisfying

\[
\ell(ax) = a\ell(x), \quad \ell(x + y) = \ell(x) + \ell(y),
\]

that are in addition continuous. They are continuous in that they satisfy

\[
\lim_{n \to \infty} \ell(x_n) = \ell(x) \quad \text{when} \quad \lim_{n \to \infty} |x_n - x| = 0.
\]

Definition. The collection of all continuous linear functionals is called the dual of \( X \). It is denoted by \( X' \).

Clearly, the sum and constant multiple of continuous linear functionals is continuous and linear; thus \( X' \) is a linear space.

Definition. A linear functional \( \ell \) on \( X \) is called bounded if there is a positive number \( c \) such that

\[
|\ell(x)| \leq c|x| \quad \text{for all} \ x \in X,
\]

where \(| | \) on the left denotes the absolute value.

Theorem 1. A linear functional \( \ell \) on \( X \) is continuous if and only if it is bounded.

Proof. Set \( x_n - x = y_n \) in (2); using (1) and (3), we get

\[
|\ell(x_n) - \ell(x)| = |\ell(y_n)| \leq c|y_n|,
\]

this shows that boundedness implies continuity.

Suppose that \( \ell \) is not bounded; then for any choice of \( c = n \), (2) is violated by some \( x_n \):

\[
\ell(x_n) > n|x_n|.
\]

Clearly, \( x_n \) can be replaced by any multiple of \( x_n \); if we normalize \( x_n \) so that

\[
|x_n| = \frac{1}{\sqrt{n}},
\]

then \( x_n \to 0 \) but \( \ell(x_n) \to \infty \). This shows that lack of boundedness implies lack of continuity.

\[
\square
\]

Theorem 2. The nullspace of a bounded linear functional \( \ell \) on a normed linear space is a closed linear subspace. For \( \ell \) nontrivial, meaning \( \ell \neq 0 \), the nullspace has codimension 1.

Proof. The nullspace of any linear map is a linear subspace. Since a bounded linear functional is continuous, it follows that the inverse image of 0 is closed. That for \( \ell \neq 0 \) the nullspace has codimension 1 is immediate.

Definition. The norm of a bounded linear functional is the smallest \( c \) for which (3) holds; it is denoted as \(|\ell|\):

\[
|\ell| = \sup_{x \neq 0} \frac{|\ell(x)|}{|x|};
\]

by homogeneity, we may take \( x \) to have norm equal to 1.

Theorem 3. The dual \( X' \) of any normed linear space is a complete normed linear space under the norm defined by (4).

Proof. Homogeneity and positivity are obvious. For subadditivity, consider two bounded linear functionals \( \ell \) and \( m \):

\[
|\ell + m| = \sup_{|x| = 1} |(\ell + m)(x)| \leq \sup_{|x| = 1} |\ell(x)| + |m(x)|
\]

\[
\leq \sup_{|x| = 1} |\ell(x)| + \sup_{|x| = 1} |m(x)| = |\ell| + |m|.
\]

We show now completeness: let \( \{\ell_n\} \) be a Cauchy sequence in \( X' \):

\[
|\ell_n - \ell_m| \to 0 \quad \text{as} \ n, m \to \infty.
\]

According to definition (4) of the norm for functionals and (5),

\[
|(\ell_n - \ell_m)(x)| = |\ell_n(x) - \ell_m(x)| \leq |\ell_n - \ell_m||x| \to 0 \quad \text{as} \ n, m \to \infty.
\]
for every \( x \) in \( X \). Since the field of scalars, \( \mathbb{R} \) or \( \mathbb{C} \), is complete,

\[
\lim_{n \to \infty} \epsilon_n(x) = \epsilon(x)
\]

exists. It is easy to show that \( \epsilon(x) \) is linear and bounded, and it is not hard to deduce from (5) that if \( |\epsilon_n - \epsilon_m| \leq \epsilon \) for \( m > n \), then also \( |\epsilon_n - \epsilon| \leq \epsilon \). Therefore

\[
\lim_{n \to \infty} |\epsilon_n - \epsilon| = 0. \quad \square
\]

### 8.2 Extension of Bounded Linear Functionals

So far we have not shown the existence of a single linear functional except \( \epsilon \equiv 0 \). Certainly there are lots of them in a Hilbert space; we show now that there are just as many in a Banach space. The tool needed is the Hahn-Banach theorem, specialized to the case

\[
p(x) = c|x|.
\]

**Theorem 4.** Let \( X \) be a normed linear space over the real or complex numbers, \( Y \) a subspace, and \( \epsilon \) a linear functional defined on \( Y \) and bounded there:

\[
|\epsilon(y)| \leq c|y|, \quad y \in Y.
\]

Then \( \epsilon \) can be extended as a bounded linear functional to all of \( X \) so that its bound on \( X \) equals its bound on \( Y \).

This theorem is a special case of theorem 8 of chapter 3. We give now some applications.

**Theorem 5.** Say that \( y_1, \ldots, y_N \) are \( N \) linearly independent vectors in a normed linear space \( X \), \( a_1, \ldots, a_N \) arbitrary complex numbers. Then there exists a bounded linear functional \( \epsilon \) such that

\[
\epsilon(y_j) = a_j, \quad j = 1, \ldots, N. \quad (6)
\]

**Proof.** Denote by \( Y \) the linear space spanned by \( y_1, \ldots, y_m \): it consists of vectors of the form

\[
y = \sum b_j y_j.
\]

Since the \( y_j \) are linearly independent, this representation of \( y \) is unique. Now define \( \epsilon \) on \( Y \) by

\[
\epsilon(y) = \sum b_j a_j.
\]

Clearly, \( \epsilon \) is linear and bounded on \( Y \), and satisfies (6); by theorem 4, \( \epsilon \) can be boundedly extended to all of \( X \). \quad \square

### Extension of Bounded Linear Functionals

**Corollary 4.** Every finite-dimensional subspace \( Y \) of a normed linear space \( X \) has a closed complement.

**Proof.** Choose a basis \( y_1, \ldots, y_N \) in \( Y \). According to theorem 5, there exist \( N \) bounded linear functionals \( \epsilon_j, j = 1, \ldots, N \), such that

\[
\epsilon_j(y_k) = \delta_{jk};
\]

according to theorem 2, the nullspace \( Z_j \) of \( \epsilon_j \) is closed. So then is their intersection

\[
Z = Z_1 \cap \ldots \cap Z_N.
\]

It is easy to check that \( Z \) and \( Y \) are complementary, namely that \( X = Y \oplus Z \). \quad \square

**Theorem 6.** For every \( y \) in a normed linear space \( X \) over the real or complex field,

\[
|y| = \max_{|\epsilon| = 1} |\epsilon(y)|. \quad (7)
\]

**Proof.** By definition (4) of \( |\epsilon|, |\epsilon(y)| \leq |\epsilon||y| \). It follows that the right side of (7) is \( \leq \) the left side. Therefore to prove the result, we have to exhibit for every \( y \) in \( X \) an \( \epsilon \) in \( X' \) such that

\[
\epsilon(y) = |y|, \quad |\epsilon| = 1.
\]

To accomplish this, we note that this defines \( \epsilon \) on all scalar multiples \( Y \) of \( y \) as \( \epsilon(ay) = a|y| \). Clearly, on this one dimensional space \( Y \), \( \epsilon \) has norm 1. By theorem 4, \( \epsilon \) can be extended to all of \( X \) so that \( |\epsilon| = 1 \).

**Corollary 5.** When the field of scalars is \( \mathbb{R} \), for every \( x \) in \( X \)

\[
|x| = \max_{|\epsilon| \leq 1} \epsilon(x). \quad (8)
\]

The following is a far-reaching generalization of theorem 6.

**Theorem 7.** \( X \) is a normed linear space over \( \mathbb{C} \), \( Y \) a linear subspace of \( X \). For any \( z \) in \( X \), denote by \( m(z) \) its distance from \( Y \):

\[
m(z) = \inf_{y \in Y} |z - y| \quad (9)
\]

**We claim that for every \( z \) in \( X \)**

\[
m(z) = M(z).
\]

**where**

\[
M(z) = \max_{|\epsilon| \leq 1, \epsilon = 0 \text{ on } Y} |\epsilon(z)|. \quad (11)
\]
DUALS OF NORMED LINEAR SPACES

Proof. Since the functionals \( \ell \) entering the maximum problem (11) vanish on \( Y \), and since \( |\ell| \leq 1 \), \( |\ell(z) - \ell(y)| \leq |z - y| \) holds for all \( y \) in \( Y \); therefore

\[
|\ell(z)| \leq \inf_{y \in Y} |z - y| = m(z).
\]

It follows from this and the definition (11) of \( M(z) \) that

\[
M(z) \leq m(z). \tag{12}
\]

To show equality, we look at the linear space \( Y_0 \) consisting of all vectors of the form \( y + az, y \in Y, a \) complex, and define on \( Y_0 \) the linear functional \( \ell_0 \):

\[
\ell_0(y + az) = \text{ann}(z). \tag{13}
\]

By definition (9) of \( m \), it follows that \( \ell_0 \) is bounded on \( Y_0 \) by 1; so by theorem 4, it can be extended to all of \( X \) so that \( |\ell_0| = 1 \). Set \( y = 0, a = 1 \) in (13):

\[
\ell_0(z) = m(z).
\]

Combined with (12) this shows that \( \ell_0 \) solves the maximum problem (11), and that (10) holds.

Remark 1. In case \( Y \) is the trivial subspace consisting of \( 0 \), theorem 7 reduces to theorem 6.

Exercise 3. Show that \( Y' \) is isometrically isomorphic with \( X'/Y' \).

Definition. The set of linear functionals \( \ell \) that vanish on a subspace \( Y \) of \( X \) is called the annihilator of \( Y \), and is denoted by \( Y^\perp \).

Exercise 1. Show that \( Y^\perp \) is a closed linear subspace of \( X' \).

Exercise 2. Let \( Y \) be a closed subspace of a normed linear space \( X \). Show that the dual of \( (X'/Y') \) is isometrically isomorphic with \( Y^\perp \).

Theorem 7'. \( X \) is a normed linear space over \( C \), \( Y \) a subspace of \( X \). For any \( \ell \) in \( X' \), define

\[
|\ell|_Y = \text{sup}_{y \in Y} |\ell(y)|. \tag{14}
\]

We claim that

\[
|\ell|_Y = \min_{m \in Y^\perp} |\ell - m|. \tag{15}
\]

EXTENSION OF BOUNDED LINEAR FUNCTIONALS

Proof. For any \( m \) in \( Y^\perp \), and any \( y \) in \( Y \) with \( |y| = 1 \),

\[
|\ell(y)| = |(\ell - m)(y)| \leq |\ell - m|.
\]

It follows that \( |\ell|_Y \) is \( \leq \) the right side of (15).

According to theorem 4, the restriction of \( \ell \) to \( Y \) has an extension to \( X \), call it \( \ell_0 \), whose norm on \( X \) equals its norm on \( Y' \):

\[
|\ell_0| = |\ell|_Y. \tag{17}
\]

Since \( \ell_0 \) and \( \ell \) are equal on \( Y \), \( \ell - \ell_0 = m \) belongs to \( Y^\perp \); furthermore by (17),

\[
|\ell - m| = |\ell_0| = |\ell|_Y. \tag{17'}
\]

This combined with (16) proves that equality holds in (15).

This is another example of dual variational problems.

Exercise 3. Show that \( Y' \) is isometrically isomorphic with \( X'/Y' \).

Definition. The closed linear span of a subset \( \{y_j\} \) of a normed linear space is the smallest closed linear space containing all \( y_j \); that is, the intersection of all closed linear spaces containing all \( y_j \).

Exercise 4. Show that the closed linear span of \( \{y_j\} \) is the closure of the linear span \( Y' \) of \( \{y_j\} \), consisting of all finite linear combinations of the \( y_j \):

\[
y = \sum_{j \in J} a_j y_j. \tag{18}
\]

The following result, called the spanning criterion, is one of the workhorses of functional analysis.

Theorem 8. A point \( z \) of a normed linear space \( X \) belongs to the closed linear span \( Y' \) of a subset \( \{y_j\} \) of \( X \) iff every bounded linear functional \( \ell \) that vanishes on the subset vanishes at \( z \); that is,

\[
\ell(y_j) = 0 \quad \text{for all} \quad y_j \tag{19}
\]

implies that \( \ell(z) = 0 \).

Proof. Since \( \ell \) is linear, (19) implies that \( \ell(y) = 0 \) for \( y \) of form (18); since \( \ell \) is continuous, it vanishes on all limits of points of form (18). Conversely, suppose that \( z \) does not belong to the closed linear span \( Y' \); then

\[
\inf_{y \in Y} |z - y| = d > 0. \tag{20}
\]
Define the subspace \( Z \) to consist of all points of the form
\[
y + az, \quad y \in Y,
\]
and define on \( Z \) the linear functional \( \ell_0 \) by
\[
\ell_0(y + az) = a.
\]
It follows from (20) that
\[
|y + az| \geq d|a|.
\]
Combining this with the definition of \( \ell_0 \), we deduce that on \( Z \), \( \ell_0 \) is bounded by \( d^{-1} \).

So by theorem 4, \( \ell_0 \) can be extended boundedly to all of \( X \). By definition,
\[
\ell_0(y_j) = 0 \quad \text{for all } y_j, \ell_0(z) = 1.
\]

\textbf{NOTE.} Theorem 8 is a generalization to Banach spaces of theorem 7 in chapter 6 on Hilbert spaces.

### 8.3 Reflexive Spaces

The dual \( X' \) of a normed linear space has its own dual, denoted as \( X'' \). Since \( \ell(x) \) is a bilinear function of \( \ell \) and \( x \), and is bounded, definition (4), it follows that, for fixed \( x \), \( \ell(x) \) is a bounded linear functional of \( \ell \). It follows from theorem 6 that the norm of this linear functional is \( |x| \). Thus the space \( X \) is, in this natural way, isometrically embedded in \( X'' \). It is a basic result of the theory of finite dimensional vector spaces that \( X'' = X \). This is no longer true for all Banach spaces.

\textbf{Definition.} A Banach space is called reflexive if \( X'' = X \), that is, if \( X \) is all of \( X'' \).

\textbf{Theorem 9.} Every Hilbert space is reflexive.

\textbf{Proof.} This is an immediate consequence of theorem 4 in chapter 6.

The following result is due to Milman.

\textbf{Theorem 10.} A uniformly convex Banach space is reflexive.

For proof we refer to Milman.

We stated in chapter 5 that the \( L^p \) spaces, \( 1 < p < \infty \), are uniformly convex. Combining this result of Clarkson's with the preceding result of Milman, we conclude that \( L^p, \ 1 < p < \infty \), are reflexive.

\textbf{REFLEXIVE SPACES}

\textbf{Theorem 11.} The dual of \( L^p \) is \( L^q \),
\[
\frac{1}{p} + \frac{1}{q} = 1.
\]

\textbf{Proof.} We saw in chapter 5 that for any \( u \) in \( L^q \) we can define a bounded linear functional \( \ell \) on \( L^p \) by
\[
\ell(f) = (f, u) = \int f(s) u(s) \, dm.
\]

Furthermore we showed, as theorem 5 of chapter 5, that the norm of this linear functional is \( |u|_q \). Thus \( L^q \) is isometrically embedded in \( (L^p)' \). We claim that \( L^q \) is all of \( (L^p)' \); for if not, there would be some \( z \) in \( (L^p)' \) not in \( L^q \). Since \( L^q \) is closed, it would follow from the spanning criterion, theorem 8, that there is an \( \ell 
eq 0 \) in \( (L^p)' \) such that \( \ell(u) = 0 \) for all \( u \) in \( L^q \). Since \( L^p \) is reflexive, \( \ell \) lies in \( L^p \), and it follows that \( \ell(u) = 0 \) for all \( u \) in \( L^q \). By theorem 5 of chapter 5, this implies that \( \ell = 0 \), a contradiction.

We give now a second proof that for \( p < 2 \) the dual of \( L^p \) is \( L^q \), without appealing to uniform convexity.

We assume for simplicity that the total measure with respect to which we form the \( L^p \) norms equals 1. Then, by Hölder's inequality with \( p' = 2/p, q' = 2/(2-p) \),
\[
\|f\|_p = \int |f|^p \, dm \leq \|1\|_{2/(2-p)} \|f\|_{2/p} \|2/p = \|f\|_2^p.
\]

So \( \|f\|_p \leq \|f\|_2 \) for all \( f \) in \( L^2 \).

Let \( \ell \) be a linear functional defined on all \( L^2 \) functions that is bounded in the \( L^p \) norm:
\[
|\ell(f)| \leq \text{const.} \|f\|_p.
\]

Since the \( p \)-norm is less than the \( 2 \)-norm, \( \ell \) is also bounded in the \( L^2 \)-norm. According to the representation theorem, theorem 4 in chapter 6, we can express \( \ell \) as
\[
\ell(f) = \int f u \, dm, \quad u \text{ in } L^2.
\]

(22)

We claim that in fact \( u \) lies in \( L^q \). To see this, we choose \( f = f_k \) as follows:
\[
f_k(x) = |u_k|^{q-1}(x) \operatorname{sgn} u(x),
\]

where
\[
|u_k|(x) = \min\{|u(x)|, k\}.
\]
k a constant. Setting \( f = f_k \) in (22) gives

\[
\ell(f_k) = \int f_k \, u \, d\mu = \int |u_k^{(q-1)}| \, |u| \, d\mu \geq \int |u_k|^q \, d\mu.
\]

On the other hand,

\[
\|f_k\|^p = \int |u_k|^{(q-1)p} \, d\mu = \int |u_k|^q \, d\mu.
\]

Since, by assumption, \( |\ell(f_k)| \leq c \|f_k\|\), the last two inequalities imply that

\[
\int |u_k|^q \, d\mu \leq c \left( \int |u_k|^q \, d\mu \right)^{1/p}.
\]

Dividing both sides by the right side gives

\[
\|u_k\|_q \leq c.
\]

Letting \( k \to \infty \) we deduce that \( u \) is in \( L_q \). This completes the proof. \( \square \)

**Exercise 5.** Show that if the total measure equals 1, then \( \|f\|_p \) is an increasing function of \( p \).

**Theorem 12.** \( C[-1, 1] \), normed by the maximum norm, is not reflexive.

**Proof.** If it were, \( C \) would be the dual of \( C' \). According to theorem 6 applied to \( X = C' \), for every \( \ell \) in \( C' \) there is an \( f \) in \( C'' = C \) such that

\[
|\ell| = \ell(f), \quad |f|_{\text{max}} = 1.
\]

Now define

\[
\ell(g) = \int_{-1}^0 g(t) \, dt - \int_0^1 g(t) \, dt.
\]

Clearly, for every \( g \) in \( C[-1, 1] \),

\[
|\ell(g)| < 2 |g|_{\text{max}}.
\]

But given any \( \varepsilon > 0 \), we can choose \( g \) so that

\[
|\ell(g)| > (2 - \varepsilon) |g|_{\text{max}}.
\]

This result shows that \( |\ell| = 2 \). Along with (23'), it contradicts (23) for \( g = f \). \( \square \)

**Theorem 13.** Let \( Z \) be a normed linear space over \( C \). If \( Z' \) is separable, so is \( Z \).

**Proof.** Separability means that \( Z' \) contains a dense denumerable set \( \{\xi_n\} \). By definition of the norm in \( Z' \), there is a \( \zeta_n \) in \( Z \) such that

\[
|\zeta_n| = 1, \quad \ell_n(\zeta_n) > \frac{1}{2} |\xi_n|.
\]

We claim that the denumerable set \( \{\zeta_n\} \) has \( Z \) as its closed linear span. According to theorem 8 this means that a linear functional \( \ell \) that vanishes on every \( \zeta_n \) vanishes everywhere. Suppose, on the contrary, that there is an \( \ell \) such that

\[
\ell(\zeta_n) = 0 \quad \text{for all } n, \quad \text{and } |\ell| = 1.
\]

Since \( \{\xi_n\} \) are dense in \( Z' \), we can find an \( \xi_n \) such that

\[
|\ell - \xi_n| < \frac{1}{2}.
\]

Since \( |\ell| = 1 \), it follows that

\[
|\xi_n| > \frac{1}{2}.
\]

Since \( \ell(\zeta_n) = 0 \), it follows from (26), (24) that

\[
\frac{1}{2} > |(\ell - \xi_n)(\zeta_n)| = |\xi_n(\zeta_n)| > \frac{1}{2} |\xi_n|.
\]

This contradicts (26'), and shows that no \( \ell \) satisfying (25) exists. It proves, by theorem 8, that finite linear combinations of the \( \zeta_n \) are dense in \( Z \). But then finite linear combinations of the \( \zeta_n \) with rational coefficients also are dense in \( Z' \); since these are denumerable, \( Z \) is separable. \( \square \)

Theorem 13 furnishes another proof of theorem 12. For \( C[-1, 1] \) is separable: every continuous function can be approximated by piecewise linear functions with rational nodes and rational ordinates. On the other hand, \( C' \) is not separable; the linear functionals \( \ell_x \) defined by

\[
\ell_x(f) = f(s), \quad -1 \leq s \leq 1,
\]

are clearly each bounded by 1, and equally clearly

\[
|\ell_x - \ell_t| = 2 \quad \text{for } t \neq s.
\]

Since the \( \{\ell_x\} \) form a nondenumerable collection, \( C' \) cannot contain a dense denumerable subset. It follows now that \( C'' \neq C \). If it were, we could apply theorem 13 to \( Z = C' \) and conclude that since \( C'' = C \) is separable, so is \( C' \), but \( C' \) is not separable. \( \square \)

The conclusion of theorem 12 is applicable to the space \( C(Q) \), \( Q \) any Hausdorff space containing more than a discrete set of points. This is the precise state of affairs:
Theorem 14. Let $Q$ be a compact Hausdorff space, $C(Q)$ the space of continuous real-valued functions on $Q$, normed by the max norm.

(i) $C'$ consists of all signed measures $m$ of finite total mass, defined over all Borel sets. That is, every bounded linear functional $\ell$ on $C(Q)$ can be written as

$$\ell(f) = \int_{Q} f \, dm. \quad (27)$$

The norm of $\ell$ is

$$|\ell| = \int_{Q} |dm|. \quad (28)$$

The measure $m$ is uniquely determined by $\ell$.

(ii) $C''$ is $L^\infty(Q)$, the space of all bounded, Borel-measurable functions on $Q$.

The prototype of this basic result is due to F. Riesz; the general result is due to Kakutani. A functional analytic proof for metric is supplied in Appendix A.

NOTE. Theorem 14 is emphatically false when $Q$ is not compact and $C(Q)$ is the space of all bounded continuous functions on $Q$, normed by the sup norm. Here is what happens:

Take $Q$ to be the real line $\mathbb{R}$, $\{x_k\}$ a sequence of points $\to \infty$. We take $Y$ to be the subspace of $C(\mathbb{R})$ consisting of all functions $f$ for which

$$\lim f(x_k) = f_0$$

exists. For $f$ in $Y$ we define the functional $\ell$ by

$$\ell(f) = f_0.$$ 

Clearly, $\ell$ is linear and $\ell$ is bounded on $Y$: $|\ell| \leq 1$. By the Hahn-Banach theorem, $\ell$ can be extended to all of $C(\mathbb{R})$ as a bounded linear functional.

We claim that this $\ell$ cannot be of the form (27). If it were, the value of $\ell(f)$ would depend on values of $f$ on any compact interval $I$ on which

$$\int_{I} |dm| \neq 0.$$

But clearly, we can alter the values of $f$ on $I$ without changing the value of $\ell(f)$; so there cannot be such a dependence.

The following result is of some interest:

Theorem 15. A closed linear subspace $Y$ of a reflexive Banach space $X$ is reflexive.

Support Function of a Set

Proof. Every bounded linear functional $\ell$ on $X$, when restricted to $Y$, becomes a bounded linear functional on $Y$; we denote this functional by $\ell_0$. Since by Hahn-Banach every bounded linear functional on $Y$ can be extended to $X$, this restriction map $\ell \to \ell_0$,

$$X' \to Y'$$

maps $X'$ onto $Y'$. The restriction map induces the following mapping from $Y''$ to $X''$:

For any $\eta$ in $Y''$ we define $\zeta$ in $X''$ by setting, for any $\ell$ in $X'$,

$$\zeta(\ell) = \eta(\ell_0), \quad (29)$$

where $\ell_0$ is the restriction of $\ell$ to $Y$. Since $X$ is reflexive, $\zeta$ can be identified with an element $z$ of $X$:

$$\zeta(\ell) = \ell(z);$$

setting this into (29) gives

$$\ell(z) = \eta(\ell_0). \quad (29')$$

We claim that $z$ belongs to $Y$. To show this, we note that if $\ell$ belongs to $Y'$, meaning it vanishes on $Y'$, then $\ell_0 = 0$, and so by (29'), $\ell(z) = 0$. We appeal now to theorem 8 to conclude that $z$ belongs to the closure of $Y$. But since $Y$ is closed, $z$ belongs to $Y$. So we can rewrite (29') as

$$\ell_0(z) = \eta(\ell_0). \quad (30)$$

Since every functional in $Y'$ occurs as $\ell_0$, (30) shows that every $\eta$ in $Y''$ can be identified with some $z$ in $Y$. \hfill \square

8.4 Support Function of a Set

We recall from chapter 1 the notion of the convex hull of a point set $M$ in a linear space $X$ over the reals as the smallest convex set in $X$ containing $M$, that is, the intersection of all convex sets that contain $M$. The convex hull of $M$ is denoted by $\hat{M}$.

As remarked in theorem 6 of chapter 1, $\hat{M}$ consists of all convex combinations of points of $M$. These are points of the form

$$x = \sum_{F} a_j x_j, \quad x_j \in M \quad (31)$$

$$a_j \geq 0, \quad \sum_{F} a_j = 1. \quad (31')$$
**Definition.** The closed convex hull of a subset $M$ of a normed linear space $X$ is the smallest closed convex set containing $M$, that is the intersection of all closed convex sets containing $M$. We denote this set by $\breve{M}$.

**Exercise 6.** Show that the closed convex hull of $M$ is the closure of the convex hull of $M$.

**Definition.** For any bounded subset $M$ of a normed linear space $X$ over $\mathbb{R}$, we define the support function $S_M$ as the following function on $X'$:

$$S_M(\ell) = \sup_{y \in M} \ell(y).$$  \hspace{1cm} (32)

**Theorem 16.** Support functions have the following properties:

(i) Subadditivity, for all $\ell, m$ in $X'$, $S_M(\ell + m) \leq S_M(\ell) + S_M(m)$.

(ii) $S_M(0) = 0$.

(iii) Positive homogeneity, $S_M(a\ell) = aS_M(\ell)$ for $a > 0$.

(iv) Monotonicity, for $M \subset N$, $S_M(\ell) \leq S_N(\ell)$.

(v) Additivity, $S_{M+N} = S_M + S_N$.

(vi) $S_{-\ell} = S_M(-\ell)$.

(vii) $S_{\breve{M}} = S_M$.

(viii) $S_{M'} = S_M$.

**Exercise 7.** Prove theorem 16.

We give now some examples.

(a) $M$ consists of a single point $x_0$,

$$S_{\{x_0\}}(\ell) = \ell(x_0).$$

(b) $M$ is the ball $B_R$ of radius $R$ around $0$: $\{ |x| \leq R \}$,

$$S_{B_R}(\ell) = R|\ell|.$$  \hspace{1cm} (33)

(c) $M$ is the ball $B_R(x_0) : |x-x_0| \leq R$, using examples (a) and (b), and part (v) of theorem 16, we get

$$S_{B_R(x_0)}(\ell) = \ell(x_0) + R|\ell|.$$  \hspace{1cm} (33')

**Theorem 17.** $X$ is a normed linear space over $\mathbb{R}$, $M$ a bounded subset of $X$. A point $z$ of $X$ belongs to the closed, convex hull of $M$ if and only if for all $\ell$ in $X'$,

$$\ell(z) \leq S_M(\ell).$$  \hspace{1cm} (34)

**SUPPORT FUNCTION OF A SET**

**Proof.** By definition (32) of support function, for all $\ell$ in $X'$ and any $z$ in $\breve{M}$, $\ell(z) \leq S_M(\ell)$. By parts (vii) and (viii) of theorem 16, $S_M = S_M$, so that (34) is satisfied for all $z$ in $\breve{M}$.

Conversely, suppose that $z$ does not belong to $\breve{M}$. Since $\breve{M}$ is closed, some open ball $B_R(z)$ centered at $z$ does not intersect $\breve{M}$. By the extended hyperplane separation theorem, theorem 6 of chapter 3, there is a nonzero linear functional $\ell_0$ and a real number $c$ such that

$$\ell_0(u) \leq c \leq \ell_0(v)$$  \hspace{1cm} (35)

for all $u$ in $\breve{M}$, all $v$ in $B_R(z)$. It follows from the right half of inequality (35) that $\ell_0$ is a bounded linear functional.

The points $v$ of $B_R(z)$ are of the form $v = z + Rx$, $|x| < 1$. By the right half of inequality (35),

$$c \leq \ell_0(z) + R\ell_0(x).$$

It follows from the definition of the norm of a linear functional that

$$\inf_{|x|<1} \ell_0(x) = -\ell_0(z).$$

It follows from the inequality above that

$$c \leq \ell_0(z) - R\ell_0(z).$$

(36)

From the left half of inequality (35) and the definition (32) of $S_M$, we conclude that

$$S_M(\ell_0) \leq c.$$  \hspace{1cm} (36')

Combining (36) and (36') gives

$$S_M(\ell_0) + R|\ell_0| \leq \ell(z).$$  \hspace{1cm} (37)

Since $\ell_0 \neq 0$, $|\ell_0| > 0$; thus (37) shows that if $z$ does not belong to $\breve{M}$, (34) fails for some $\ell_0$, as asserted in theorem 17.

**Theorem 18.** $K$ denotes a closed, convex subset of a real linear space $X$, $z$ a point of $X$ not in $K$. Then

$$\inf_{u \text{ in } K} |z - u| = \sup_{|\ell| = 1} [\ell(z) - S_K(\ell)].$$  \hspace{1cm} (38)

**Proof.** By definition (32) of support function,

$$S_K(\ell) \geq \ell(u)$$

for all $\ell$, all $u \in K$.

So for $|\ell| = 1$,
DUALS OF NORMED LINEAR SPACES

\[ S_K(\ell) \geq \ell(u) = \ell(z) + \ell(u - z) \geq \ell(z) - |u - z|, \]

which is the same as \(|u - z| \geq \ell(z) - S_K(\ell)|. It follows from this that

\[ \inf_{u \in K} |u - z| \geq \sup_{|\ell| = 1} \{ \ell(z) - S_K(\ell) \}. \quad (39) \]

To show the opposite inequality, let \( R \) be any positive number less than the inf on the left in (38). Denote by \( B R \) the ball of radius \( R \) around the origin; then the set \( K + B R \) has a positive distance from \( z \). So it follows from theorem 17, with \( K + B R \) in place of \( M \), that for some \( \ell_0 \) in \( X' \),

\[ S_{K + B R}(\ell_0) < \ell_0(z). \quad (40) \]

We use now additivity and example (b):

\[ S_{K + B R}(\ell_0) = S_K(\ell_0) + R|\ell_0|. \]

Since we may choose \( \ell_0 \) to have norm = 1, it follows from (40) and (40') that

\[ R < \ell_0(z) - S_K(\ell_0) \]

for some \( \ell_0 \) with \(|\ell_0| = 1\); it follows from this that the sup on the right of (38) is \( \geq R \). Since \( R \) can be any number less than the inf on the left of (38), it follows that

\[ \inf_{u \in K} |u - z| \leq \sup_{|\ell| = 1} \{ \ell(z) - S_K(\ell) \}. \quad (39') \]

Combined with (39) this proves (38).

Theorem 18 presents another example of dual variational problems. Theorem 7 is a special case of theorem 18, if we extend the definition of support function to a linear space \( Y \).

\[ S_Y(\ell) = \begin{cases} 0 & \text{if } \ell \text{ in } Y^\perp \\ \infty & \text{if } \ell \text{ not in } Y^\perp \end{cases} \]

BIBLIOGRAPHY


9

APPLICATIONS OF DUALITY

9.1 COMPLETENESS OF WEIGHTED POWERS

Let \( w(t) \) be a given positive function defined on \( \mathbb{R} \) that decays exponentially as \(|t| \to \infty\):

\[ 0 < w(t) < ae^{-ct}, \quad c > 0. \quad (1) \]

Denote by \( C \) the set of continuous functions on \( \mathbb{R} \) that vanish at \( \infty \):

\[ \lim_{|t| \to \infty} x(t) = 0. \quad (2) \]

\( C \) is a Banach space under the maximum norm.

Theorem 1. The functions \( t^n w(t) \) belong to \( C \); their closed linear span is all of \( C \). That is, every function in \( C \) can be approximated uniformly on \( \mathbb{R} \) by weighted polynomials.

Proof. We will use theorem 8 of chapter 8. Let \( \ell \) be any bounded linear functional over \( C \) that vanishes on the functions \( t^n w \):

\[ \ell(t^n w) = 0, \quad n = 0, 1, \ldots \quad (3) \]

Let \( z \) be a complex variable, \(|\Im z| < c\). Then \( w(t) e^{zt} \) belongs to \( C \), and so

\[ f(z) = \ell(w e^{zt}) \quad (4) \]

is defined in the strip \(|\Im z| < c\). We claim that \( f(z) \) is analytic there. For the complex difference quotients of \( w e^{zt} \) tend to \( \ell w e^{zt} \) in the norm of \( C \), and so

\[ f'(z) = \lim_{\Delta t \to 0} \frac{f(z + \Delta t) - f(z)}{\Delta t} = \lim_{\Delta t \to 0} \ell \left( \frac{e^{z(t+\Delta t)} - e^{zt}}{\Delta t} \right) = \ell(\ell w e^{zt}). \]
Similarly for the higher derivatives, in particular, using (3).

\[ \frac{d^n f}{dx^n} \bigg|_{x=0} = i^n \ell(\pi n) = 0, \quad n = 0, 1, \ldots \]

Since \( f \) is analytic, the vanishing of all its derivatives at \( x = 0 \) means that \( f(\xi) = 0 \) in the strip, in particular,

\[ f(\xi) = \ell(w e^{i\xi t}) = 0 \quad \text{for all} \ \xi \ \text{real.} \]

By theorem 8, chapter 8, it follows that all functions \( w e^{i\xi t} \) belong to the closed linear span of \( r^n w \).

According to the Weierstrass approximation theorem, every continuous period function \( h(t) \) is the uniform limit of trigonometric polynomials. It follows that \( wh \) belongs to the closed linear span of the functions \( w e^{i\xi t} \), \( \xi \) real, hence of the functions \( r^n w \). Let \( y \) be any continuous function of compact support; define \( x \) by

\[ x = \frac{y}{w}. \quad (5) \]

Denote by \( h \) a 2-periodic function such that

\[ x(t) = h(t) \quad \text{for} \ |t| < p. \quad (5') \]

\( p \) chosen so large that the support of \( x \) is contained in the interval \( |t| < p \). Then

\[ |x - h|_{\max} \leq |x|_{\max}. \]

and so, by (5), (5'), and (1),

\[ |y - wh| \leq 2 e^{-2p} |x|_{\max}. \]

This shows that as \( p \to \infty, wh \to y \). Since \( wh \) belongs to the closed linear span of the functions \( r^n w \), so does \( y \). The functions \( y \) of compact support are dense in \( C \), and the proof is complete. \( \square \)

9.2 THE MÜNTZ APPROXIMATION THEOREM

According to the Weierstrass approximation theorem, any continuous function \( x(t) \) on the interval \([0, 1]\) can be approximated uniformly by polynomials in \( t \). Let \( n \) be any integer. Clearly, if \( x(t) \) is continuous on \([0, 1]\), so is

\[ y(s) = x(s^{1/n}). \]

Now \( y(s) \) can be approximated arbitrarily closely in the maximum norm by polynomials \( p(s) \). Setting \( s = r^n \), we conclude that \( x(t) \) can be approximated arbitrarily closely by linear combination of \( r^n, j = 0, 1, \ldots \). Thus in the Weierstrass approximation theorem not all powers of \( t \) are needed.

Serge Bernstein posed the following question: What sequences of positive numbers \( \{\lambda_j\} \) tending to \( \infty \) have the property that the closed linear span of the functions

\[ 1, \quad r^{\lambda_j}, \quad j = 1, 2, \ldots \]

is the space \( C \) of all continuous functions \([0, 1]\)? After some preliminary results were obtained by Bernstein, Müntz proved the following:

**Theorem 2.** Let \( \lambda_j \) be a sequence of positive numbers tending to \( \infty \). The functions \( (6) \) span the space of all continuous functions \( C \) on \([0, 1]\) that vanish at \( t = 0 \) iff

\[ \sum \frac{1}{\lambda_j} = \infty. \quad (7) \]

**Proof.** We will use the spanning criterion, theorem 8, chapter 8. Let \( \ell \) be a bounded linear functional on \( C \) that vanishes on all the functions \( (6) \):

\[ \ell(r^{\lambda_j}) = 0, \quad j = 1, 2, \ldots. \quad (8) \]

Let \( \xi \) be a complex variable, \( \text{Re} \ \xi > 0 \). For such \( \xi \), \( r^\xi \) belongs to \( C \) and depends analytically on \( \xi \), in the sense that

\[ \lim_{\delta \to 0} \frac{r^{\xi + \delta} - r^\xi}{\delta} = (\log r) r^\xi \]

exists in the sense of the norm in \( C \), the maximum norm. Define

\[ f(\xi) = \ell(r^\xi). \quad (9) \]

It follows that \( f \) is an analytic function of \( \xi \). Furthermore, since \( |r^\xi| \leq 1 \) when \( 0 \leq t < 1 \) and \( \text{Re} \ \xi > 0 \), and since \( \ell \) is bounded, say \( |\ell| \leq 1 \), it follows from (9) that

\[ |f(\xi)| \leq 1 \quad \text{for} \ \text{Re} \ \xi > 0. \quad (10) \]

Relation (8) can be expressed as

\[ f(\lambda_j) = 0. \quad (11) \]

We define a Blaschke product \( B_N(\xi) \) as follows:

\[ B_N(\xi) = \prod_{\lambda_j < 1} \frac{\xi - \lambda_j}{\xi + \lambda_j}. \quad (12) \]

It has the following properties:

\[ B_N(\lambda_j) = 0, \quad j = 1, \ldots, N. \quad (13a) \]
APPLIEDS OF DUALITY

\[ B_N(\zeta) \neq 0 \quad \text{for } \zeta \neq \lambda_j. \quad (13b) \]
\[ |B_N(\zeta)| \to 1 \quad \text{as } \Re \zeta \to 0. \quad (13c) \]
\[ |B_N(\zeta)| \to 1 \quad \text{as } |\zeta| \to \infty. \quad (13d) \]

Since the zeros of \( B_N(\zeta) \) are shared by \( f(\zeta) \)

\[ g_N(\zeta) = \frac{f(\zeta)}{B_N(\zeta)} \quad (14) \]

is regular analytic in \( \Re \zeta > 0 \). We claim that

\[ |g_N(\zeta)| \leq 1 \quad \text{for } \Re \zeta > 0. \quad (15) \]

For combining (10) and (13c), (13d) we conclude that for any \( \epsilon > 0 \), \( |g_N(\zeta)| \leq 1 + \epsilon \) for \( \Re \zeta = \delta \), and for \( |\zeta| = \delta^{-1} \), \( \delta \) small enough. By the maximum principle for the analytic function \( g_N \) on the domain \( \Re \zeta \geq \delta, |\zeta| \leq \delta^{-1}, |g_N(\zeta)| \leq 1 + \epsilon \) there.

Letting \( \delta, \epsilon \to 0 \), we obtain (15). Let \( k \) be a positive number such that \( f(k) \neq 0 \);

from (14) and (15) we conclude that

\[ \prod_{1}^{N} \frac{\lambda_j + k}{\lambda_j - k} \leq \frac{1}{|f(k)|}. \quad (16) \]

We can write the factors on the left in (16) as

\[ 1 + \frac{2k}{\lambda_j - k}. \]

Since \( \lambda_j \to \infty \), all but a finite number of the factors above are \( > 1 \). So from the uniform boundedness of the product (16) for all \( N \), we conclude the uniform boundedness, for all \( N \), of the sum

\[ \sum_{1}^{N} \frac{1}{\lambda_j - k}. \]

This contradicts (7); so we conclude that \( f(k) = 0 \) for all \( k \). In view of the definition (9) of \( f \), and property (8) of \( \ell \), this says that any linear functional that vanishes on the functions \( \tau^{1/2} \) vanishes on \( \tau \), \( k \) positive. So by theorem 8, chapter 8, the spanning criterion, we conclude that all functions \( \tau^{k} \) can be approximated uniformly by linear combination of the functions \( \{\tau^{1/2}\} \). Taking in particular \( k = 1, 2, 3, \ldots \) and appealing to the Weierstrass approximation theorem, we conclude that the functions (6) span \( C \).

We omit the proof of the necessity of condition (7).

\[ \square \]

NOTE. Szász has extended Müntz's theorem to complex \( \lambda_j \).

Exercise 1. Formulate and prove the extension of theorem 2 to complex exponents.

9.3 RUNGE'S THEOREM

Theorem 3. Let \( D \) be a bounded simply connected domain in \( \mathbb{C} \). Every analytic function \( f(\zeta) \) in \( D \) can be approximated, uniformly on compact subsets \( K \), by polynomials in \( \zeta \).

Proof. Since \( D \) is simply connected, every compact subset of \( D \) is contained in a simply connected compact subset \( K \) of \( D \). Choose a closed smooth curve in \( D - K \) that winds once around every point of \( K \), and express \( f(\zeta), \zeta \in K \), by Cauchy's integral formula. This integral can be approximated, uniformly for all points \( \zeta \) of \( K \), by a sum. This sum is a linear combination of functions of the form \( (\chi - \zeta)^{-1}, \chi \) on the curve. Therefore to prove the theorem, it suffices to show that all functions of \( \zeta \) of form \( (\chi - \zeta)^{-1}, \chi \) not in \( K \), can be approximated on \( K \) by polynomials in \( \zeta \).

This is clear when \( |\chi| > R, R = \max |\zeta|, \zeta \) in \( K \). Then the geometric series

\[ (\chi - \zeta)^{-1} = \sum_{0}^{\infty} \frac{\zeta^{n}}{\chi^{n+1}} \]

converges uniformly on \( K \). To show it for all \( \zeta \), we will use the spanning criterion. Let \( \ell \) be any bounded linear functional on \( C(K) \) that vanishes on all polynomials:

\[ \ell(p) = 0. \]

We claim that \( \ell \) vanishes on all functions of the form \( (\chi - \zeta)^{-1}, \chi \) not in \( K \). Define

\[ g(\chi) = \ell((\chi - \zeta)^{-1}). \]

Since \( (\chi - \zeta)^{-1} \), as element of \( C(K) \), depends analytically on \( \chi \), it follows that \( g(\chi) \) is an analytic function of \( \chi \) in the exterior of \( K \). Since for \( |\chi| > R, (\chi - \zeta)^{-1} \) belongs to the closure of polynomials \( p \), and since \( \ell(p) = 0 \), it follows by continuity that \( \ell((\chi - \zeta)^{-1}) = 0 \) for such \( \chi \), and so

\[ g(\chi) = 0 \quad \text{for } |R| < |\chi|. \]

Since \( g \) is analytic in the exterior, and since the exterior of a simply connected set \( K \) is connected, it follows that \( g(\chi) = 0 \) for all \( \chi \) not in \( K \). Then, by the spanning criterion, theorem 8 of chapter 8, for all \( \chi \) outside \( K \), \( (\chi - \zeta)^{-1} \) is in the closure of the space of polynomials.

This beautiful proof is due to Lars Hörmander.

9.4 DUAL VARIATIONAL PROBLEMS IN FUNCTION THEORY

Theorem 4. Let \( D \) be a bounded domain in \( \mathbb{C} \), whose boundary consists of a finite number of \( C^1 \) arcs. Denote by \( A \) the space of functions analytic in \( D \) and continuous
Applications of Duality

up to the boundary, and by \( \xi_0 \) any point in \( D \). Define

\[
M = \sup_{\xi_0} |f'(\xi_0)|, \quad f \text{ in } A, |f|_{\max} \leq 1. \tag{17}
\]

Denote by \( u_0 \) the function

\[
u_0(\xi) = \frac{1}{2\pi i} \frac{1}{(\xi - \xi_0)^2}
\]

on \( \partial D \), and define

\[
m = \inf_{g \in A} \int_{\partial D} |u_0(\xi) - g(\xi)| \, d\xi. \tag{19}
\]

We claim that \( m = M \), and that the supremum \( M \) in (17) is attained.

Proof. We show first that \( M \leq m \). Using the Cauchy integral formula, and the Cauchy–integral theorem, we can represent \( f'(\xi_0) \) as follows:

\[
f'(\xi_0) = \int_{\partial D} f u_0 \, d\xi = \int_{\partial D} f(u_0 - g) \, d\xi, \tag{20}
\]

where \( g \) is any function in \( A \). Since \( f \) in (17) is \( \leq 1 \) in absolute value, we deduce from (20) that

\[
|f'(\xi_0)| \leq \int_{\partial D} |u_0 - g| \, d\xi.
\]

Choose \( g \) so that it nearly minimizes (19). We get

\[
|f'(\xi_0)| \leq m + \epsilon.
\]

In view of (17) this implies \( M \leq m + \epsilon \), and since \( \epsilon > 0 \) is arbitrary, \( M \leq m \).

To prove the converse inequality, we look at the space \( C \) of continuous functions on \( \partial D \); \( u_0 \) belongs to \( C \), and \( A \) is a linear subspace of \( C \). We impose the \( L^1 \) norm \( | \cdot |_1 \) on \( C \), with respect to arclength \( |d\xi| \) along \( \partial D \). The infimum (19) can be written as

\[
m = \inf_{g \in A} |u_0 - g|_1.
\]

According to theorem 7 of chapter 8,

\[
m = \max_{|\xi| \leq 1} |\ell(u_0)|, \quad \ell = 0 \text{ on } A. \tag{21}
\]

We denote by \( \ell_0 \) an \( \ell \) that maximizes (21). We define the function \( f_0 \) for \( \chi \) not on \( \partial D \) by

\[
f_0(\chi) = \frac{1}{2\pi i} \ell_0 \left( \frac{1}{\xi - \chi} \right). \tag{22}
\]

Dual Variational Problems in Function Theory

We claim that \( f_0 \) has the following properties:

(i) \( f_0(\chi) \) is analytic for \( \chi \) not on \( \partial D \).

(ii) \( f_0(\chi) = 0 \) for \( \chi \) not in \( D \).

(iii) \( |f_0(\chi)| \leq 1 \).

(iv) \( f_0(\xi_0) = m \).

Property (i) follows from the fact that \( 1/(\xi - \chi) \) as an element of \( C \) depends analytically on \( \chi \). Property (ii) follows since for \( \chi \) not in \( D \), \( 1/(\xi - \chi) \) lies in \( A \), and \( \ell_0 \) vanishes on \( A \). To prove (iii), choose \( \chi \) near \( \partial D \) and take \( \chi' \) to be the reflection of \( \chi \) across \( D \). \( \chi' \) lies outside \( D \). We will use (ii) to write

\[
f_0(\chi) = f_0(\chi) - f_0(\chi') = \frac{1}{2\pi i} \ell_0 \left( \frac{1}{\xi - \chi} - \frac{1}{\xi - \chi'} \right).
\]

Since \( \ell_0 \) is \( \leq 1 \),

\[
|f_0(\chi)| \leq \frac{1}{2\pi} \left| \frac{1}{\xi - \chi} - \frac{1}{\xi - \chi'} \right| = \frac{1}{2\pi} \int_{\partial D} \frac{|x - \chi'|}{|\xi - \chi||\xi - \chi'|} \, d\xi.
\]

A simple estimate shows that the integral on the right is \( \leq 1 + \epsilon \) for \( \chi \) near the boundary. So \( |f_0(\chi)| \leq 1 + \epsilon \) for \( \chi \) near the boundary; from this (iii) follows by the maximum principle.

Differentiate (22) and set \( \xi = \xi_0 \); using (18) and (21), we get

\[
f_0'(\xi_0) = \ell_0(u_0) = m. \tag{23}
\]

Since we have already shown \( |f_0'(\xi_0)| \leq m \) for \( |f| \leq 1 \) in \( D \), it follows that \( f_0 \) solves the maximum problem (17), and that \( M = m \).

Since \( D \) is simply connected, we claim that \( f_0 \), the function maximizing (17), maps \( D \) conformally onto the unit disc. Here is a sketch of a proof:

It can be shown that the minimum problem (19) has a solution, call it \( g \). Setting \( f = f_0 \) and \( g = g_0 \) into (20), and using (23), we get

\[
m = \int_{\partial D} f_0(u_0 - g_0) \frac{d\xi}{ds} \, ds,
\]

where \( s \) is arclength. Using the fact that \( |f_0(\xi)| \leq 1 \), and that, by (19),

\[
m = \int_{\partial D} |u_0 - g_0| \frac{|d\xi|}{|d\xi|} \, ds,
\]

we conclude that \( |f_0(\xi)| = 1 \) on \( \partial D \) and that

\[
f_0(u_0 - g_0) \frac{d\xi}{ds} > 0 \quad \text{on } \partial D. \tag{24}
\]
Denote by $2\pi|\hbar|$ the change of argument of a non-zero complex valued function $h$ around $\partial D$. From (24) we deduce that

$$[f_0] + [u_0 - g_0] + \left[\frac{d\zeta}{ds}\right] = 0.$$  

(25)

According to the argument principle, for the boundary values $h$ of functions meromorphic in $D$,

$$[\hbar] = \# \text{zeros} - \# \text{poles in } D.$$  

But $f_0$ and $g_0$ have no poles, and $u_0$ has a single pole of order 2. For $D$ simply connected, $[d\zeta/ds] = 1$; so we deduce from (25) that

$$\# \text{zeros of } f_0 + \# \text{zeros of } (u_0 - g_0) - 2 + 1 = 0.$$  

(26)

It follows from (26) that $f_0$ has at most one zero in $D$. We claim that it has at least one zero; otherwise, $f_0^{-1}$ would be analytic in $D$. Since $|f_0| = 1$ on $\partial D$, it would follow from the maximum principle that $|f_0(\xi)| \equiv 1$ in $D$, which implies that $f_0 \equiv \text{const.}$, contrary to $f'(\xi_0) = m$. Combining the two statements above, we conclude that $f_0$ has exactly one zero in $D$.

According to the argument principle, $[f_0]$ equals 1. Since $|f_0(\xi)| = 1$ for $\xi$ on $\partial D$, $[f_0 - w] = 1$ for all $w$ inside the unit disc; it follows that $f(\xi)$ takes on every value $w$ exactly once in $D$. This shows that $f_0$ maps $D$ one-to-one onto the open unit disc.

The arguments used in this section combine methods introduced by Rogosinski and Shapiro with results of Garabedian and Schiffer.

9.5 EXISTENCE OF GREEN’S FUNCTION

Definition. Let $D$ be a plane domain whose boundary $B$ is once continuously differentiable. Green’s function $G(p, q)$ of the domain $D$ is defined for $p, q$ in $D$ by the requirements that

(i) $\Delta_p G = \delta(p - q)$, where $\Delta_p$ is the Laplace operator with respect to $p$, and $\delta$ is the Dirac distribution, see Appendix B.

(ii) $G(p, q) = 0$ for $p$ on $B$.

In this definition, the variable $p$ and $q$ play unsymmetric roles, $q$ appearing merely as a parameter in a boundary value problem.

The significance of Green’s function is that it can be used to represent every harmonic function $h$ in $D$ in terms of the boundary values of $h$ by using Green’s formula:

$$h(q) = \int_{\partial D} k(p, q) G(p, q) \, dp,$$  

(30)

where $G(p, q)$ is the derivative of $G$ with respect to $p$ in the direction normal to the boundary $B$, and $dp$ is arc-length. Furthermore, for $D$ simply connected, $G$ is the logarithm of the absolute value of the analytic function $f$ mapping $D$ onto the unit disk, carrying $q$ into the origin.

Green’s function can be split into its singular and regular part:

$$G(p, q) = -\frac{1}{2\pi} \log |p - q| + g_0(p, q).$$  

(27)

The function $g_0$ is called the regular part of Green’s function. Referring to (i) and (ii) above, we can characterize $g_0$ as the solution of the following boundary value problem:

$$\Delta_p g_0 = 0 \quad \text{in } D,$$  

(28)

$$g_0(p, q) = \log |p - q| \quad \text{for } p \text{ on } B,$$  

(29)

where $\log r$ is an abbreviation for $(1/2\pi) \log r$.

Clearly, the definition of Green’s function rests on the fact that the boundary value problem (28), (29) can be solved. Classically this is deduced from the solvability of the Dirichlet boundary value problem for $\Delta$ with arbitrarily prescribed boundary values. Here we show how to solve (28), (29) without appealing to the general theory. We denote by $C$ the space of continuous functions on the boundary $B$ normed by the max norm. We denote by $H$ the subspace consisting of the boundary values of functions $h$ that are harmonic in $D$ and continuous up to the boundary. In what follows the point $q$ is fixed once and for all in $D$. We define the functional $\ell_q$ for functions $h$ in $H$ by setting

$$\ell_q(h) = h(q),$$  

(30)

where $h(q)$ denotes the value at $q$ of the harmonic function whose value on $B$ is denoted by $h$. It is well known in the theory of harmonic functions that $h$ at $q$ is uniquely determined by $h$ on the boundary $B$, and that the maximum principle holds:

$$h(q) \leq \max_{z \in B} |h(z)| = |h|.$$  

This inequality can be expressed so: as defined on $H$, the norm of $\ell_q$ is $\leq 1$. It follows then from the Hahn-Banach theorem that $\ell_q$ can be extended from $H$ to all of $C$ so that

$$|\ell_q| \leq 1.$$  

(31)

We denote by $w$ any point of the plane not on the boundary $B$ of $D$, and define the element $k(w)$ of $C$ by

$$k(p, w) = \log |p - w|, \quad p \in B.$$  

(32)

Two observations on the manner of dependence of $k$ on the parameter $w$:
APPLICATIONS OF DUALITY

(i) $k(w)$ is a differentiable function of $w$, and satisfies

$$\Delta w k = 0 \quad (33)$$

in each component of the complement of $B$.

(ii) For $w$ in the exterior of $D$, $k(w)$ belongs to $H$.

We define now the function $g(w, q)$ by

$$g(w, q) = \ell_q(k(w)). \quad (34)$$

where $k$ is defined by (32).

**Lemma 5.**

(i) $g(w, q)$ is a harmonic function of $w$ in each component of the complement of $B$.

(ii) For $w'$ in the exterior of $D$,

$$g(w', q) = \log|q - w'|. \quad (35)$$

**Proof.** Since $\ell_q$ is linear, by (34),

$$\ell_q\left(\frac{k(w + du) - k(w)}{du}\right) = \frac{g(w + du, q) - g(w, q)}{du}. \quad (36)$$

We let $du$ tend to 0. Since $\ell_q$ is bounded, we deduce that

$$\ell_q(\nabla w k) = \nabla w g(w, q). \quad (37)$$

Applying this to second derivatives and using (33), we get

$$\Delta w g = \ell_q(\Delta w k) = 0,$nabla$$

as asserted in part (i).

For $w'$ in the exterior of $D$, $k(w')$ belongs to $H$. Applying the original definition (30) of $\ell_q$ in (34) yields (35).

**Lemma 6.** $g(w, q)$ depends continuously on $w$ as $w$ crosses the boundary.

To see this, let $w$ be a point in $D$ close to the boundary, and $w'$ the reflection of $w$ across the boundary. Reflection $w'$ is obtained by drawing a straight line from $w$ to the nearest boundary point $p_0$, and choosing $w'$ so that

$$\frac{w + w'}{2} = p_0.$$
APPLIEDS OF DUALITY

REMARK 2. In case of a boundary curve $B$ that is twice differentiable, relation (37) can be sharpened to

$$\left| \frac{p - w}{p - w} \right| = 1 + O(d).$$

where $d = |w - p|$ is the distance of $w$ to the boundary. Using this, we can sharpen (38):

$$\log\left| \frac{p - w}{p - w} \right| = O(d).$$

We set this into (36). Since by (35), $g(w', q') = \log|q - w|$, it differs from $g(p, q) = \log|q - p|$ by $O(d)$. So we conclude that as $w$ in $D$ tends to the nearest boundary point $p$,

$$|g(w, q) - g(p, q)| \leq O(d).$$

We claim that the first derivatives of $g$ are uniformly bounded in $D$ up to the boundary. This is because we can express the first derivatives of the harmonic function $g$ at $w$ as integrals over the circle of radius $d$ centered at $w$:

$$2\pi \text{ grad } g(w) = \frac{1}{d} \int \left[ g(w + de(\theta)) - g(p) \right] e(\theta) d\theta,$$

where $e(\theta) = (\cos \theta, \sin \theta)$. It follows from (41) that the integrand in (42) is $O(d)$; from this the uniform boundedness of grad $g$ follows.

We know from the Cauchy-Riemann equations that the conjugate harmonic function to $g$ also has uniformly bounded first derivatives in $D$. This shows that for $D$ simply connected, the analytic function mapping $D$ onto a disk is uniformly Lipschitz continuous.

Green's functions of more than two variables were constructed in Lax.

BIBLIOGRAPHY


10

WEAK CONVERGENCE

Definition. A sequence $\{x_n\}$ in a normed linear space $X$ is said to converge weakly to $x$ if

$$\lim_{n \to \infty} \ell(x_n) = \ell(x)$$

for every $\ell$ in $X'$. This relation is indicated by a half arrow:

$$x_n \rightharpoonup x.$$

Another notation is

$$w = \lim_{n \to \infty} x_n = x.$$

The notion is to be contrasted with convergence in norm of the sequence:

$$\lim_{n \to \infty} |y_n - y| = 0.$$

In this case we say that $\{y_n\}$ tends strongly to $y$, and denote it as

$$y_n \to y.$$

Another suggestive notation for strong convergence is

$$s = \lim_{n \to \infty} y_n = y.$$

Clearly, a sequence that converges to $x$ strongly also converges weakly to $x$, but in general not vice versa. Here are some examples:

Example 1. $X = \ell^2$; its points are vectors

$$x = (a_1, a_2, \ldots)$$

with denumerably many components, such that

$$\|x\|^2 = \sum |a_j|^2 < \infty.$$
Since $\ell^2$ is a Hilbert space, according to theorem 4 of chapter 6 all bounded linear functionals on $\ell^2$ are of the form
\[ \ell(x) = (x, y) = \sum a_jb_j, \quad \sum |b_j|^2 < \infty. \tag{4} \]

Define $x_n$ as the nth unit vector, that is, the vector whose n'th component is 1, all others zero: $x_n = (0, \ldots, 0, 1, 0, \ldots)$. It is easy to show, and left as an exercise, that the sequence $\{x_n\}$ tends weakly to zero, but not strongly.

**Example 2.** $H$ any Hilbert space, $\{x_n\}$ an orthonormal sequence; such a sequence tends weakly, but not strongly, to zero.

**Proof.** It follows from Bessel's inequality (31) of chapter 6, that for any $y$ in $H$
\[ \sum |(x_n, y)|^2 \leq \|y\|^2, \tag{5} \]
from which it follows that
\[ \ell(x_n) = (x_n, y) \to 0. \tag{5'} \]

Since according to the Riesz representation theorem, theorem 4 in chapter 6, all linear functionals of the form (5'), weak convergence to zero follows. Since $\|x_n\| = 1$ for all $n$, $\{x_n\}$ does not tend to zero strongly.

**Example 3.** $X = C[0, 1]$.
\[ x_n(t) = \begin{cases} nt & \text{for } 0 \leq t \leq \frac{1}{n} \\ 2 - nt & \text{for } \frac{1}{n} \leq t \leq \frac{2}{n} \\ 0 & \text{for } \frac{2}{n} \leq t \leq 1. \end{cases} \]

**Claim.** $x_n$ tends to zero weakly but not strongly.

**Proof.** Let $\ell$ be a bounded linear functional; we claim that $\lim \ell(x_n) = 0$. Now suppose not; then there would be infinitely many $n$ such that $|\ell(x_n)| > \delta > 0$, say
\[ \ell(x_n) > \delta. \tag{6} \]

Choose a subsequence $\{n_k\}$, $n_{k+1} > 2n_k$, for which (6) holds. It is not hard to show that for all $t$ in $[0,1]$
\[ y_K = \sum_{i=1}^{K} x_{n_k}(t) < 4, \tag{7} \]

which implies that $|y_K| < 4$ for all $K$. From (6) it follows that
\[ \ell(y_K) = \sum_{i=1}^{K} \ell(x_{n_k}) > K\delta. \]

Since this holds for all $K$, and since $|y_K| < 4$ for all $K$, the boundedness of $\ell$ is contradicted. Since $|x_n| = \max x_n(t) = 1$, $\{x_n\}$ does not tend to zero strongly.

**Exercise 1.** Prove inequality (7). (Draw a picture of the graph of $x_n(t)$).

**Example 4.** $X = \ell^1$, consisting of vectors $x = (a_1, a_2, \ldots)$, $|x| = \sum|a_k| < \infty$. As we saw in chapter 8 (exercise after theorem 11), the dual of $\ell^1$ is $\ell^\infty$. The following observation is due to Schur:

If a sequence $\{x_n\}$ in $\ell^1$ converges weakly, it converges strongly.

**Exercise 2.** Prove the preceding statement.

### 10.1 Uniform Boundedness of Weakly Convergent Sequences

The following result is useful in proving weak convergence:

**Theorem 1.** Suppose that a sequence $\{x_n\}$ of points in a normed linear space satisfies

(i) $\{|x_n|\}$ are uniformly bounded:
\[ |x_n| \leq c. \]

(ii) $\lim \ell(x_n) = \ell(x)$ for a set of $\ell$ dense in $X'$. Then
\[ w = \lim x_n = x. \]

**Exercise 3.** Prove theorem 1.

Surprisingly, the converse of theorem 1 holds in Banach spaces. To see this, we appeal to the principle of uniform boundedness for a complete metric space $S$: if a collection $\{f_x\}$ of continuous real-valued functions $f_x$ on $S$ is bounded at each point $x$ of $S$,
\[ |f_x(x)| \leq M(x) \quad \text{for all } x, \tag{8} \]

then the functions $f_x$ are uniformly bounded,
\[ |f_x(u)| \leq M, \tag{8'} \]
for all \( u \) in some nonempty open set \( O \). We specialize this to the case where \( S \) is a Banach space \( X \) and each \( f_v \) is subadditive and absolutely homogeneous:

\[
f(x + y) \leq f(x) + f(y), \quad f(ax) = |a|f(x).
\]  

**Theorem 2.** Let \( X \) be a Banach space, \( \{f_v\} \) a collection of real-valued continuous subadditive and absolutely homogeneous functions on \( X \), bounded at each point \( x \) of \( X \) as in (8). Then the \( \{f_v\} \) are uniformly bounded, that is, there is a number \( c \) such that

\[
|f_v(x)| \leq c|x|
\]  

for all \( f_v \) and all \( x \) in \( X \).

**Proof.** By the principle of uniform boundedness for metric spaces, \( |f_v(u)| \leq M \) for all \( f_v \) and all \( u \) in some open ball \( u = z + y, \ |y| < r \). Using subadditivity, we have that

\[
|f_v(y)| = |f_v(u - z)| \leq |f_v(u)| + |f_v(z)| \leq 2M
\]

(11)

holds for all \( y \) with \( |y| = r/2 \). For any \( x \) in \( X \) define \( y \) by \( y = rx/2|x| \). By construction, \( |y| = r/2 \), so (11) holds. Using absolute homogeneity we get from (11) that

\[
|f_v(x)| = \left| f_v \left( \frac{2|x|}{r} \right) \right| = \frac{2|x|}{r} f(y) \leq \frac{4M}{r} |x|;
\]

this proves (10), with \( c = 4M/r \).

An immediate consequence of theorem 2 is

**Theorem 3.** \( X \) is a Banach space, \( \{\ell_v\} \) a collection of bounded linear functionals such that at every point \( x \) of \( X \)

\[
|\ell_v(x)| \leq M(x) \quad \text{for all } \ell_v.
\]  

(12)

Then there is a constant \( c \) such that

\[
|\ell_v| \leq c \quad \text{for all } \ell_v.
\]  

(13)

**Proof.** \( \ell(x) \) is a continuous subadditive and absolutely homogeneous function of \( x \). Therefore theorem 2 is applicable; its conclusion (10) yields (13).

Another immediate consequence is

**Theorem 4.** \( X \) is a normed linear space, \( \{x_v\} \) a collection of points in \( X \) such that for every bounded linear functional \( \ell \)

\[
|\ell(x_v)| \leq M(\ell) \quad \text{for all } \ell_v.
\]  

(12')

Then there is a constant \( c \) such that

\[
|x_v| \leq c \quad \text{for all } x_v.
\]  

(13')

**Proof.** Theorem 4 follows from theorem 3 applied to the Banach space \( X' \), on which the elements \( x_v \) of \( X \) act as bounded linear functionals.

An immediate consequence of theorem 4 is

**Theorem 4'.** A weakly convergent sequence \( \{x_v\} \) in a normed linear space \( X \) is uniformly bounded in norm.

**Proof.** Weak convergence means that \( \ell(x_v) \) converges for every \( \ell \) in \( X' \). Since a convergent sequence of numbers is bounded, hypothesis (12') of theorem 4 is satisfied; therefore (13') holds.

Theorems 2, 3, and 4 are called the principle of uniform boundedness.

**Theorem 5.** Let \( \{x_v\} \) be a sequence in a normed linear space converging weakly to \( x \). Then

\[
|x| = \lim \inf |x_v|.
\]  

(14)

**Proof.** According to theorem 6 of chapter 8, there is an \( \ell \) in \( X' \), such that

\[
|x| = |\ell(x)|, \quad |\ell| = 1.
\]

Since weak convergence means that

\[
\ell(x_v) = \lim \ell(x_v),
\]

and since

\[
|\ell(x_v)| \leq |\ell| |x_v| = |x_v|,
\]

(14) follows.

The following far reaching generalization of theorem 5 is due to Mazur:

**Theorem 6.** Let \( K \) be a closed, convex subset of a normed linear space \( X \), \( \{x_v\} \) a sequence of points in \( K \), converging weakly to a point \( x \). Then \( x \) belongs to \( K \).

**Proof.** Let \( S_K \) be the support function of \( K \), defined by equation (32) of chapter 8 as \( \sup_{x \in K} \ell(x) \). It follows from that definition that for any \( \ell \) in \( X' \)
Since \( \ell(x_n) \) tends to \( \ell(x) \), it follows that also
\[
\ell(x) \leq S_K(\ell).
\]

But according to theorem 17 of chapter 5, this guarantees that \( x \) belongs to \( K \).

\( \square \)

**Exercise 4.** Deduce theorem 5 from theorem 6 applied to balls centered at the origin \( K = B_R : \{x \mid |x| \leq R\} \).

### 10.2 WEAK SEQUENTIAL COMPACTNESS

**Definition.** A subset \( C \) of a Banach space \( X \) is called **weakly sequentially compact** if any sequence of points in \( C \) has a subsequence weakly convergent to a point of \( C \).

**Exercise 5.** Show that a weakly sequentially compact set is bounded.

The importance of weak sequential compactness is the same as that of compactness in the sense of strong convergence. Weak compactness is a valuable tool in constructing, as weak limits, mathematical objects of interest. To wield this tool, we need simple, easily verifiable criteria for weak compactness; the following is such a criterion:

**Theorem 7.** In a reflexive Banach space \( X \) the closed unit ball is weakly sequentially compact.

**Proof.** Let \( \{y_n\} \) be a sequence of points in the unit ball, that is, \( |y_n| \leq 1 \). Denote by \( Y \) the closed linear subspace spanned by the set \( \{y_n\} \). \( Y \) is separable. Since \( X \) is assumed reflexive, it follows from theorem 15 of chapter 8 that \( Y \) is reflexive. Since \( Y = Y'' \) is separable as well, it follows from theorem 13 of chapter 8 that \( Y'' \) is separable, meaning that it contains a dense, denumerable subset \( \{m_j\} \). Using the classical diagonal process, we can select a subsequence \( \{z_n\} \) of \( \{y_n\} \) such that
\[
\lim_{k \to \infty} m_j(z_n) = \ell(z)
\]
exists for every \( m_j \). Since all \( z_n \) satisfy \( |z_n| \leq 1 \), and since the \( \{m_j\} \) are dense, it follows from (16) and theorem 1 that for all \( m \) in \( Y' \), \( m(z_n) \) tends to a limit as \( n \to \infty \). This limit is a linear functional of \( m \):
\[
\lim_{n \to \infty} m'(z_n) = y(m) = m(y).
\]

Since \( |m(z_n)| \leq |m| \cdot |z_n| \leq |m| \), it follows from (16') that the linear functional \( y(m) = m(y) \), \( |y| \leq 1 \), and so (16) says that for all \( m \) in \( Y' \), \( m(z_n) \) tends to \( m(y) \) as \( n \to \infty \). Since the restriction of any \( \ell \) in \( X' \) to \( Y \) is an \( m \) in \( Y' \), this proves that \( z_n \) converges weakly to a point \( y \) in the unit ball.

\( \square \)

Note the sharp contrast between theorem 7 and theorem 6 of chapter 5, according to which the unit ball is never compact in the norm topology. Compactness is gained by replacing strong with weak convergence.

Eberlein has proved the converse of theorem 7:

**Theorem 8.** The closed unit ball in a Banach space \( X \) is weakly sequentially compact only if \( X \) is reflexive.

Combining theorems 6 and 7 gives the following useful result:

**Theorem 9.** In a reflexive Banach space every bounded, closed, convex set is weakly sequentially compact.

Here is a useful application of theorem 9.

**Theorem 10.** Let \( X \) be a reflexive Banach space. \( K \) a closed, convex subset of \( X \), \( z \) any point of \( X \). Then there is a point \( y \) of \( K \) which is as close to \( z \) as any other point of \( K \).

**Proof.** We may take \( z = 0 \), and assume that \( 0 \not\in K \). Denote by \( s \) the distance of 0 to \( K \), that is,
\[
s = \inf \{y \mid y \in K\}, \quad y \in K.
\]

(17)

Let \( \{y_n\} \) be a minimizing sequence for (17). We may assume that each \( y_n \) lies in the intersection of \( K \) and that the ball of radius is \( 2s \) around the origin. This is a bounded, closed, convex set, therefore, by theorem 9, a subsequence \( \{z_n\} \) of \( \{y_n\} \) converges weakly to some point \( z \) of \( K \). According to theorem 5,
\[
|z| \leq \lim \inf |z_n|.
\]

(18)

Since \( \{z_n\} \) is the subsequence of a minimizing sequence, \( \lim |z_n| = s \). Combining this with (17) and (18) gives \( |z| = s \); that is, \( z \) is a point of \( K \) closest to \( 0 \).

**Theorem 10** is a generalization of theorem 8 of chapter 5. There we assumed that \( X \) is uniformly convex; here we assume only that \( X \) is reflexive.

### 10.3 WEAK* CONVERGENCE

In a Banach space \( U \) that is the dual \( X' \) of another Banach space \( X \), there is a subclass of linear functionals associated with elements \( x \) of \( X \):
One can define sequential convergence in $U$ with respect to this subclass of linear functionals:

**Definition.** A sequence $\{u_n\}$ in a Banach space $U$ that is the dual of another Banach space $X$ is said to be weak* convergent to $u$ if

$$\lim_{n \to \infty} u_n(x) = u(x)$$

for all $x$ in $X$. We denote this relation as

$$w^* - \lim_{n \to \infty} u_n = u.$$

**Remark 1.** Of course, if $X$ is reflexive, weak* convergence is no different than weak convergence.

**Example 5.** $U$ is the space of all signed Borel measures $\mu$ on $[-1, 1]$, of finite total mass. According to theorem 14 of chapter 8, $U$ is the dual of $C([-1, 1])$.

Consider the sequence $\{m_n\}$:

$$m_n(h) = \int h \, d\mu_n = \frac{n}{2} \int_{-1/n}^{1/n} h(t) \, dt.$$  \hspace{1cm} (21)

Clearly, for any continuous $h$,

$$\lim_{n \to \infty} m_n(h) = h(0).$$  \hspace{1cm} (22)

This shows that $m_n$ is weak* convergent to the unit mass at the origin. The dual of $U$ is $L^\infty([-1, 1])$, the space of all bounded measurable functions. Since (22) is not true for some $h$ discontinuous at 0, $m_n$ does not converge weakly.

**Theorem 11.** A weak* convergent sequence $\{u_n\}$ of points in a Banach space $U = X'$ is uniformly bounded.

**Proof.** Weak* convergence implies that the boundedness condition (12) holds at every point $x$ of $X$. So theorem 3 implies (13), uniform boundedness. $\square$

**Exercise 6.** Show that if the sequence $\{u_n\}$ is weak* convergent to $u$,

$$|u| \leq \lim \inf |u_n|.$$  \hspace{1cm} (23)

**Definition.** A subset $C$ of a Banach space $U$ that is the dual of another Banach space $X$ is called weak* sequentially compact if every sequence of points in $C$ has a subsequence that is weak* convergent to a point of $C$.

---

**BIBLIOGRAPHY**

The following important result is due to Helly

**Theorem 12.** Let $X$ be a separable Banach space. $U = X'$. The closed unit ball in $U$ is weak* sequentially compact.

**Proof.** Given a sequence $\{u_n\}$ in $U$,

$$|u_n| \leq 1,$$  \hspace{1cm} (23)

and a denumerable set $\{x_n\}$ in $X$, we can, by the diagonal process, select a subsequence $\{u_{n_k}\}$ of $\{u_n\}$ such that

$$\lim_{n \to \infty} u_{n_k}(x_k)$$  \hspace{1cm} (24)

exists for all $x_k$. It follows from (23) and (24) that $u_{n_k}(x)$ tends to a limit for all $x$ that lie in the closure of the set $\{x_k\}$. So, if we take for $\{x_k\}$ a set dense in $X$, $u_n(x)$ tends to a limit for all $x$ in $X$. It is easy to see that this limit is a linear function of $x$, and, using (23), that it is bounded by 1. $\square$

**BIBLIOGRAPHY**


11
APPLICATIONS OF WEAK CONVERGENCE

11.1 APPROXIMATION OF THE $\delta$ FUNCTION
BY CONTINUOUS FUNCTIONS

Definition. A sequence $\{k_n\}$ of continuous functions on $[-1, 1]$ tends to the $\delta$ function if

$$\lim_{n \to \infty} \int_{-1}^{1} f(t) k_n(t) \, dt = f(0)$$

for all continuous functions $f$ on $[-1, 1]$.

Theorem 1 (Toeplitz). The sequence $\{k_n\}$ of continuous functions on $[-1, 1]$ tends to the $\delta$ function in the sense of (1) if and only if it satisfies the following conditions:

(i) \[ \lim_{n \to \infty} \int_{-1}^{1} k_n(t) \, dt = 1. \]

(ii) For every $C^\infty$ function $g$ whose support does not contain 0,

\[ \lim_{n \to \infty} \int_{-1}^{1} g(t) k_n(t) \, dt = 0. \] (3)

(iii) There is a constant $c$ for which

\[ \int_{-1}^{1} |k_n(t)| \, dt \leq c \] (4)

holds for all $n$.

Proof. Suppose that $f(0) = 0$; let $g$ be a $C^\infty$ function that differs from $f$ by less than $\varepsilon$ at all points $t$ in $[-1, 1]$, and that is zero in some interval around $t = 0$. Then by (4),

$$\left| \int_{-1}^{1} (f - g) k_n \, dt \right| \leq \varepsilon \int |k_n| \, dt \leq c \varepsilon. \tag{5}$$

By assumption (3), $\int g k_n \, dt$ tends to zero, so it follows from (5) that

$$\limsup \left| \int f k_n \, dt \right| \leq \varepsilon. \tag{6}$$

Since $\varepsilon$ is arbitrary, (1) is verified in case $f(0) = 0$. Every function can be decomposed as $b + f$, $b$ some constant and $f(0) = 0$, so (1) follows from (2) for every continuous $f$.

Now to the converse: condition (2) is clearly necessary, for it is a special case of (1) when $f(t) = 1$; the same goes for (3).

We can regard $\{k_n\}$ as a sequence in $C^0([-1, 1])$, and state (1) as

$$w^* - \lim k_n = \delta.$$

According to theorem 3 of chapter 10, the norms

$$|k_n| = \int |k_n(t)| \, dt$$

must be uniformly bounded. This proves the necessity of (4), and even more:

Corollary 1. If (4) is violated, there exists a continuous function $f$ for which the left side of (1) tends to infinity. \[ \square \]

11.2 DIVERGENCE OF FOURIER SERIES

Theorem 2. There exists a periodic continuous function $f(\theta)$ whose Fourier series diverges at one point.

Proof. The Fourier series of a continuous function $f$ on the unit circle $S^1$ is

$$f(\theta) \approx \sum_{-\infty}^{\infty} a_n e^{in\theta}, \tag{7}$$

where

$$a_n = \int_{-\pi}^{\pi} f(\theta) e^{-in\theta} \, d\theta \quad \text{where} \quad d\theta = \frac{d\theta}{2\pi}. \tag{7'}$$
The convergence of the series at, say, \(\theta = 0\) means that
\[
f(0) = \lim_{N \to \infty} \sum_{-N}^{N} a_n.
\] (8)

Using (7'), we can write
\[
\sum_{-N}^{N} a_n = \int_{-\pi}^{\pi} f(\theta) k_N(\theta) d\theta,
\] (9)

where
\[
2\pi k_N(\theta) = \sum_{-N}^{N} e^{-in\theta}.
\] (10)

Using the formula for the sum of a finite geometric series, we readily get, for \(\theta \neq 0\),
\[
2\pi k_N(\theta) = \frac{\sin(\pi + 1/2)\theta}{\sin \theta/2}.
\] (10')

Thus the convergence of the Fourier series of every continuous function is equivalent to the sequence \(k_n\) defined by (10) approximating the \(\delta\) function. According to theorem 1 this is the case iff conditions (2), (3), and (4) are satisfied. We show now that condition (4) fails! To see this, we use the inequality \(|\sin \phi| \leq |\phi|\), which implies that \(1/(\sin \theta/2) \geq 2/|\theta|\). Using (10') and a little calculus, we get
\[
\int_{-\pi}^{\pi} |k_N(\theta)| d\theta \geq \frac{1}{\pi} \int_{-\pi}^{\pi} |\sin \left(\frac{N + 1/2}{2}\right)| d\theta = \frac{2}{\pi} \int_{0}^{(N+1/2)\pi} |\sin \phi| d\phi.
\]
This last integral is, it is easy to show, \(\geq\) const. \(\log N\). Thus condition (4) fails, and so, by Corollary 1', for some function \(f\), the Fourier series of \(f\) diverges at \(\theta = 0\) to infinity.

Exercise 1. Show that there exists a continuous periodic function whose Fourier series diverges at \(n\) arbitrarily given points.

11.3 APPROXIMATE QUADRATURE

An approximate quadrature formula is an approximation to the integral of a continuous function \(f\) on, say, \([-1, 1]\). Take \(N\) points \(t_j\) in \([-1, 1]\), called nodes, and \(N\) numbers \(w_j\) called weights; define \(q(f)\) by
\[
q(f) = \sum_{j=1}^{N} w_j f(t_j).
\] (11)

We regard \(q\) as an approximation to
\[
\int_{-1}^{1} f(t) dt.
\] (11')

Theorem 3. Let \(q_N\) be a sequence of quadrature formulas of form (11), satisfying the following conditions:

(i) For every nonnegative integer \(k\),
\[
\lim_{N \to \infty} q_N(t^k) = \int_{-1}^{1} t^k dt.
\] (12)

(ii) For all \(N\),
\[
\sum_{j=1}^{N} |w_j(N)| \leq c,
\] (13)

\(c\) a constant. Then
\[
\lim_{N \to \infty} q_N(f) = \int_{-1}^{1} f(t) dt
\] (14)

for all continuous \(f\). Conversely, if (14) holds for all continuous \(f\), (12) and (13) must be satisfied.

Proof. It follows from (12) that (14) holds for all polynomials \(f\). Inequality (13) asserts that the linear functionals \(q_N\) on \(C[-1, 1]\) have uniformly bounded norms. Since the polynomials are dense in \(C[-1, 1]\), (14) follows for all continuous \(f\). The converse follows from theorem 3 of chapter 10.

Exercise 2. Prove that if the weights \(w_j\) are positive, (13) follows from (12).

11.4 WEAK AND STRONG ANALYTICITY OF VECTOR-VALUED FUNCTIONS

Let \(f\) be a function defined in some domain \(G\) of the complex \(\xi\) plane, whose values lie in some complex Banach space \(X\).

Definition. \(f(\xi)\) is strongly analytic in \(G\) if the limit
\[
\lim_{h \to 0} \frac{f(\xi + h) - f(\xi)}{h}
\]
exists in the norm topology at every point of \(G\).
**Definition.** \( f(\xi) \) is weakly analytic in \( G \) if for every bounded linear functional \( \ell \), \( \ell(f(\xi)) \) is an analytic function of \( \xi \) in the classical sense.

N. Dunford has proved the following surprising result:

**Theorem 4.** A weakly analytic function is strongly analytic.

**Proof.** If \( \ell(f(\xi)) \) is analytic in \( G \), we can represent it by the Cauchy integral formula

\[
\ell(f(\xi)) = \int_{C} \frac{\ell(f(\chi))}{\chi - \xi} d\chi, \quad \text{where} \quad d\chi = \frac{d\chi}{2\pi i}.
\]

(15)

\( C \) some rectifiable curve winding around \( \xi \). Similar formulas hold when \( \xi \) is replaced by \( \xi + h \) and \( \xi + k \) and \( h, k \) small enough. Assume that \( k \neq 0, h \neq 0, h \neq k \); then we can express the difference quotient of difference quotients thus:

\[
\frac{1}{h-k} \left[ \frac{\ell(f(\xi+h)) - \ell(f(\xi))}{h} - \frac{\ell(f(\xi+k)) - \ell(f(\xi))}{k} \right]
\]

\[
= \int_{C} \frac{\ell(f(\chi))}{(\chi-\xi-h)(\chi-\xi-k)(\chi-\xi)} d\chi.
\]

(16)

For fixed \( \ell \), and \( |h|, |k| \) small enough, the right side of (16) is bounded by a constant \( M \) independent of \( h \) and \( k \). We can rewrite the left side as \( \ell(x_{h,k}) \) where

\[
x_{h,k} = \frac{1}{h-k} \left[ \frac{f(\xi+h) - f(\xi)}{h} - \frac{f(\xi+k) - f(\xi)}{k} \right].
\]

(17)

Weak analyticity thus implies that for each \( \ell \) and all \( h \) and \( k \) sufficiently small, \( |\ell(x_{h,k})| \leq M(\ell) \). We appeal now to the principle of uniform boundedness, theorem 4, chapter 10, and conclude that \( |x_{h,k}| \leq c \) for all \( h, k \) sufficiently small. By definition (17) of \( x_{h,k} \) this implies the norm inequality

\[
\left| \frac{f(\xi+h) - f(\xi)}{h} - \frac{f(\xi+k) - f(\xi)}{k} \right| \leq c|h-k|.
\]

(18)

Since \( X \) is complete, it follows that the difference quotients of \( f(\xi) \) tend to a limit in the strong sense. \( \square \)

### 11.5 Existence of Solutions of Partial Differential Equations

We denote by \( L \) a first-order partial differential operator of the following form, acting on vector-valued functions:

\[
L = \sum_{j=1}^{m} A_{j} \partial_{j} + B.
\]

(19)

Here \( A_{j} \) and \( B \) are square matrix valued functions of the independent variables \( s \), \( A_{j}(s) \) being once differentiable, \( B(s) \) continuous, and

\[
\partial_{j} = \frac{\partial}{\partial s_{j}}.
\]

(19')

We assume, for simplicity, that \( A_{j} \) and \( B \), as well as the functions on which \( L \) acts, are periodic in all variables \( s \) and that they are real valued. We denote, as customary, the formal adjoint of \( L \) by \( L^{*} \):

\[
L^{*} = -\sum \partial_{j} A_{j}^{T} + B^{T}.
\]

(19'')

where \( A^{T}, B^{T} \) denote transposes. Integration by parts shows that for any pair of \( C^{1} \) vector-valued periodic functions \( u \) and \( v \),

\[
\int (v, Lu) = \langle L^{*} v, u \rangle,
\]

(20)

Here the bracket denotes the \( L^{2} \) scalar product over a period cube:

\[
\langle v, w \rangle = \int_{F} v(s) \cdot w(s) \, ds;
\]

the dot denotes the dot product between vectors, and \( F \) a period cube.

Suppose that every \( A_{j} \) is symmetric: \( A_{j}^{T} = A_{j} \). Then comparing (19) and (19''), we deduce that

\[
L^{*} = -L - \sum A_{j,j} + B + B^{T}.
\]

Here \( A_{j,j} \) denotes the partial derivative of \( A_{j} \) with respect to \( s_{j} \). Setting this into (20), and choosing \( v = u \), we get

\[
2(u, Lu) = \langle (L + L^{*}) u, u \rangle = \left( [B + B^{T} - \sum A_{j,j}] u, u \right).
\]

(20')

Using the language of distributions, see Appendix B, we state

**Theorem 5.** Suppose the matrix on the right in (20') is positive definite:

\[
B + B^{T} - \sum A_{j,j} > kI, \quad k > 0.
\]

(21)

Then for every periodic, square integrable \( u \) the equation

\[
Ly = f
\]

(22)

has a solution \( y \) in the sense of distributions that is periodic and square integrable.

**Proof.** It follows from (20') and (21) that every periodic \( C^{1} \) function satisfies the inequality

\[
(u, Lu) \geq \frac{k}{2} \| u \|^{2}.
\]

(21')
APPLICATIONS OF WEAK CONVERGENCE

where \( \| u \| \) denotes the \( L^2(F) \)-norm. Denote the Hilbert space \( L^2(F) \) by \( H \). Let \( Y \) be any finite-dimensional subspace of \( H \) consisting of periodic \( C^1 \) functions; denote the orthogonal complement of \( Y \) in \( H \) by \( Y^\perp \). Consider the equation

\[
Ly - f \in Y^\perp \quad (22'N)
\]

for \( y \) in \( Y \). These are \( N \) linear equations for \( y \) in \( Y \), \( N = \dim Y \). According to linear algebra, such a system of linear equations has a solution for every \( f \) iff the homogeneous equation

\[
Lz \in Y^\perp, \quad z \in Y, \quad (23)
\]

is satisfied only by \( z = 0 \). Take the scalar product of (23) with \( z \); using (21'), we get

\[
0 = (z, Lz) \geq \frac{k}{2} \| z \|^2,
\]

which implies that \( z = 0 \). So it follows that (22') has a unique solution \( y \). Take the scalar product of (22'N) with \( y \); using (21) and the Schwarz inequality, we get

\[
\frac{k}{2} \| y \|^2 \leq (y, Ly) = (y, f) \leq \| y \| \| f \|.
\]

This implies that

\[
\| y \| \leq \frac{2}{k} \| f \|. \quad (24)
\]

Now let \( Y_N \) be an increasing sequence of subspaces of \( C^1 \) functions whose union is dense in \( H \). Denote by \( y_N \) the solution of (22'N). It follows from (24) that \( \| y_N \| \) is a uniformly bounded sequence. Therefore, since \( H \) is reflexive, we appeal to theorem 7 of chapter 10 to conclude that a subsequence of \( \{ y_N \} \), also denoted as \( \{ y_N \} \), converges weakly:

\[
w \in \lim y_N = y.
\]

Let \( v \) belong to \( \cup Y_N \); since each \( Y_N \) consists of differentiable functions, \( v \) is differentiable, for it belongs to some \( Y_N \). For \( y_N \) in \( Y_N \),

\[
Ly_N - f \in Y_N^\perp.
\]

Take the scalar product of this with \( v \); for \( N > M \) we get

\[
(v, Ly_N) - (v, f) = 0.
\]

Since \( v \) is differentiable, this can be rewritten by (20), as

\[
(L^* v, y_N) - (v, f) = 0.
\]

THE REPRESENTATION OF ANALYTIC FUNCTIONS WITH POSITIVE REAL PART

Since the sequence \( y_N \) converges weakly to \( y \), we conclude that for every \( v \) in \( \cup Y_N \)

\[
(L^* v, y) - (v, f) = 0.
\]

(25)

We can choose the spaces \( Y_N \) so that their union is dense not only in \( H \) but also in \( H_1 \), the space of all periodic \( L^2 \) functions whose first derivatives belong to \( L^2 \). That means that given any periodic \( C^1 \) function \( v \), there is a sequence \( \{ v_k \} \) of functions in \( \cup Y_N \) such that \( v_k \) converges in \( L^2 \)-norm to \( v \), and the first derivatives of \( v_k \) converge to the first derivatives of \( v \) in the \( L^2 \)-norm. Since \( L^* \) is a first-order operator, it follows that \( L^* v_k \) tends to \( L^* v \) in the \( L^2 \)-norm. Setting \( v = v_k \) in (25), we can pass to the limit and conclude that (25) holds for all \( v \) in \( C^0 \).

A function \( y \) that satisfies (25) for all \( C^1 \) functions \( v \) is said to satisfy the differential equation (22) in the weak sense. Clearly, such a \( v \) is a solution of (22) in the sense of distributions, which requires (25) to hold for all \( C^0 \) functions \( v \).

Friedrichs has shown that a weak solution \( y \) of (22) is a strong solution in the following sense: there is a sequence of \( C^1 \) functions \( \varepsilon_n \) that converge to \( y \) in the \( L^2 \) sense, and at the same time \( L \varepsilon_n \) converges to \( f \) in the \( L^2 \) sense. It is easy to show, using (21'), that equation (22) has only one strong solution. It follows that not only a subsequence, but the whole sequence \( y_N \) converges.

The method described in this section to obtain the solution \( y \) of equation (22) as the weak limit of the solutions \( y_N \) of equation (22'N) is called Galerkin's method. It is more than a theoretical device for proving the existence of a solution of (22); it is also a practical method for constructing it.

11.6 THE REPRESENTATION OF ANALYTIC FUNCTIONS WITH POSITIVE REAL PART

Let \( f(\xi) \) be an analytic function in the unit disk \( |\xi| < 1 \) whose real part is positive:

\[
h(\xi) = \text{Re } f(\xi) \geq 0, \quad |\xi| < 1.
\]

Every analytic function defined in a disk and continuous up to the boundary can be expressed—up to an imaginary constant—in terms of its real part on the boundary by the Poisson integral. On the disk of radius \( R < 1 \) we have for \( |\xi| < R \)

\[
f(\xi) = \int_0^{2\pi} \frac{R^2 + \xi e^{-i\theta}}{R^2 - \xi e^{-i\theta}} h(Re^{i\theta}) \, d\theta + ic. \quad (26)
\]

Setting \( \xi = 0 \), we see that

\[
h(0) = \int_0^{2\pi} h(Re^{i\theta}) \, d\theta. \quad (26')
\]

Let \( R \to 1 \) through a sequence \( R_n \to 1 \). The functions \( h(R_n e^{i\theta}) \) are nonnegative functions of \( \theta \) whose integrals over the whole circle are, by (26'), all equal to \( h(0) \).
We associate with each \( R_n \) a linear functional
\[
\ell_n(u) = \int_0^{2\pi} h(R_n e^{i\theta}) u(\theta) \, d\theta
\]
acting on the space \( C \) of continuous functions \( u \) on the circle \( S^1 \). It follows from \( h \geq 0 \) and (26') that
\[
|\ell_n| = h(0).
\]
Since \( C(S^1) \) is separable, we can appeal to Heil's theorem, theorem 12 of chapter 10, and conclude that a subsequence of \( \ell_n \) is weak* convergent to some limit \( \ell \):
\[
\lim_{n \to \infty} \ell_n(u) = \ell(u) \quad (28)
\]
for all continuous functions \( u \). It follows from (28) and the uniform boundedness of \( |\ell_n| \) that for any sequence \( u_n \) strongly convergent to \( u \)
\[
\lim_{n \to \infty} \ell_n(u_n) = \ell(u) \quad (28')
\]
We apply this now to
\[
u_n = \frac{R_n + \xi e^{-i\theta}}{R_n - \xi e^{-i\theta}}, \quad u = \frac{1 + \xi e^{-i\theta}}{1 - \xi e^{-i\theta}},
\]
\( \xi \) any complex number with \( |\xi| < 1 \); using (26), (27) and (28'), we get
\[
f(\xi) = \ell \left( \frac{1 + \xi e^{-i\theta}}{1 - \xi e^{-i\theta}} \right).
\]
The functionals \( \ell_n \) defined by (27) are clearly nonnegative; therefore so is their weak* limit \( \ell \). According to corollary 14 of the Riesz representation theorem, chapter 8, such a nonnegative functional acting on \( C(S^1) \) can be represented as an integral with respect to a positive measure \( m \). Thus we have proved the first part of

**Theorem 6 (Herglotz-Riesz).** Every analytic function \( f \) in the unit disk \( |\xi| < 1 \) whose real part is positive there can be expressed as
\[
f(\xi) = \int \frac{1 + \xi e^{-i\theta}}{1 - \xi e^{-i\theta}} \, dm + ic, \quad m \text{ a positive measure, } c \text{ real} \quad (29)
\]
Conversely every function \( f \) so represented is analytic in the unit disk and has positive real part there. The representation (29) is unique.

**Proof.** That (29) represents an analytic function with positive real part in the unit disk for any positive measure \( m \) is evident from formula (30) below. To see that the representation is unique, we note that the real part of (29) is
\[
h(\xi) = \int \frac{1 - r^2}{1 - 2r \cos(\phi - \theta) + r^2} \, dm, \quad \xi = r e^{i\phi} \quad (30)
\]
Take any continuous function \( u(\phi) \), multiply (30) by \( u(\phi) \), and integrate with respect to \( \phi \) over \( S^1 \). We get, after interchanging the order of integration on the right,
\[
\int h(r e^{i\phi}) u(\phi) \, d\phi = \int u_r(\theta) \, dm, \quad (31)
\]
where
\[
u_r(\theta) = \int \frac{1 - r^2}{1 - 2r \cos(\phi - \theta) + r^2} u(\phi) \, d\phi.
\]
Suppose that \( h(\xi) \) can be represented in the form (30) by two different measures \( m \) and \( m' \). Let \( r \to 1 \) in (31); by theorem 1 of this chapter, \( u_r \to u \) in the maximum norm. Since the left side of (31) does not depend on the representing measure, it follows that
\[
\int u(\theta) \, dm = \int u(\theta) \, dm'
\]
for all continuous functions \( u \). We appeal now to the uniqueness of measure in the Riesz representation theorem to conclude that \( m = m' \). This completes the proof of theorem 6.

From the uniqueness of the measure \( m \) it follows that the limit (28) exists not only for a subsequence but every sequence of \( R_n \).

**BIBLIOGRAPHY**


THE WEAK AND WEAK* TOPOLOGIES

**Definition.** The weak topology in a Banach space is the weakest topology in which all bounded linear functionals are continuous. Since bounded linear functionals are continuous in the norm (strong) topology, it follows that the weak topology is coarser than the strong topology. We show now that—except in finite-dimensional spaces—the weak topology is genuinely coarser than the strong topology:

The open sets in the weak topology are unions of finite intersections of sets of the form

\[ \{ x : a < \ell(x) < b \}. \]  

(1)

Clearly, in an infinite-dimensional space the intersection of a finite number of sets of form (1) is unbounded. This shows that every set that is open in the weak topology is unbounded. In particular, the balls

\[ \{ x : |x| < R \}. \]  

(2)

open in the strong topology, are not open in the weak topology.

Next we show that the weak topology is coarser than weak sequential convergence, in the following sense: define the weak sequential closure of a set \( S \) in a Banach space \( X \) as the weak limit of all weakly convergent sequences in \( S \).

**Theorem 1.**

(i) The weak sequential closure of any set \( S \) belongs to the closure of \( S \) in the weak topology.

(ii) In every infinite-dimensional Banach space there are sets weakly sequentially closed, but not closed in the weak topology.

**Proof.** Part (i) is an immediate consequence of the definitions of weak convergence and weak topology. Part (ii) is exemplified by the following set \( S \):

\[ S = S_2 \cup S_3 \cup \ldots. \]  

(3)

Each set \( S_k \) is finite, constructed as follows: Choose any sequence of subspaces \( X_k \) of \( X \), \( \dim X_k = k \). \( S_k \) consists of points \( x_{k,j} \), finite in number, so chosen that for every point \( x \) in \( X_k \) of norm \( k \), \( x \) in \( X_k \), \( |x| = k \), there is a point \( x_{k,j} \) in \( S_k \) such that

\[ |x - x_{k,j}| < \frac{1}{k}, \quad |x_{k,j}| = k. \]  

(4)

We claim that the origin belongs to the closure of \( S \) in the weak topology. Any open set containing the origin contains a subset of the form

\[ \{ x : |\ell_i(x)| < \epsilon, \quad i = 1, \ldots, n \}, \]  

(5)

where the \( \ell_i \) are linear functionals of norm 1. Since \( X_k \) is \( k \)-dimensional, for \( k > n \) it contains a nonzero vector \( x_k \) such that

\[ \ell_i(x_k) = 0, \quad i = 1, \ldots, n, \quad |x_k| = k. \]  

(6)

By construction, there is an \( x_{k,j} \) in \( S_k \) satisfying condition (4) for \( x = x_k \). Using (4), (5), and \( |\ell| = 1 \), we get for any \( \ell = \ell_i \), \( i = 1, \ldots, n \), that

\[ \ell(x_{k,j}) = \ell(x_{k,j} - x_k) \leq |x_{k,j} - x_k| \leq \frac{1}{k}. \]  

(7)

Therefore for \( k > 1/\epsilon \), the point \( x_{k,j} \) belongs to the subset (5). This proves that the origin belongs to the closure of \( S \) in the weak topology.

On the other hand, \( S \) contains only a finite number of points in any ball of radius \( R \). So, by the principle of uniform boundedness (see theorem 4 of chapter 10), \( S \) contains no weakly convergent sequences other than the trivial ones.

Despite the coarseness of the weak topology compared to the strong topology, the following is true:

**Theorem 2.** Every convex subset \( K \) of a Banach space \( X \) that is closed in the strong topology is closed in the weak topology.

**Proof.** We will show that if \( z \) in \( X \) does not belong to \( K \), then \( z \) is not in the weak closure of \( K \). Since \( K \) is closed in the strong topology, there is an open ball \( B_R(z) \) centered at \( z \) that is disjoint from \( K \). According to the hyperplane separation theorem, theorem 6 of chapter 3, there is a nonzero functional \( \ell \) and a constant \( c \) such that

\[ \ell(u) \leq c \leq \ell(u) \]  

(8)

for all \( u \) in \( K \) and all \( v \) in \( B_R(z) \). As explained in the proof of theorem 17 of chapter 8, the norm of \( \ell \) is bounded by \( 1/R \). The points \( E_R(z) \) are of the form \( v = z + x \),
topology from $P$ under the embedding. Each $I_x$ is a compact interval of $R$; it follows from Tychonov's theorem that $P$ is compact. Since a closed subset of a compact set is compact, it suffices to prove that (12) maps $B$ into a closed subset of $P$.

Let $p$ be a point in the closure of the image of $B$ under (12); we will show that then $p$ is the image of some $u$ in $B$, that is, $p_x = u(x)$ for all $x$ in $X$,

$$p_x = u(x) \quad \text{for all } x \in X, \quad (13)$$

where $\{p_x\}$ denote the components of $p$. Equation (13) defines a function $u(x)$ on $X$. We have to show that it is bounded by 1 and linear. Boundedness follows from the fact that $p_x$ belongs to $I_x = [-|x|, |x|]$. Linearity means that

$$p_{x+y} = p_x + p_y, \quad p_{\alpha x} = \alpha p_x. \quad (14)$$

For every $q$ that is the image of $B$ in $P$ under the mapping (12),

$$q_{x+y} = q_x + q_y, \quad q_{\alpha x} = \alpha q_x. \quad (15)$$

Since $p$ lies in the weak* closure of $\{q\}$, and since these relations involve only 3 (resp. 2) components of $q$, (14) follows from (15). \hfill \Box

**Theorem 3'.** A subset $S$ of $U = X'$ that is closed in the weak* topology is weak* compact iff it is bounded in norm.

**Proof.** If $S$ is bounded in norm, it belongs to some closed ball $B_R$. According to theorem 3, $B_R$ is weak*—compact, but then so is its weak*—closed subset $S$.

Conversely, suppose that $S$ is weak*—compact; then so is its image $\{u(x)\}$ under the continuous mapping $u \rightarrow u(x)$ for every $x$. A compact set in $R$ (or $C$) is bounded, so for every $x$ in $X$ \( \{|u(x)|\} \leq b(x) \) for every $u$ in $S$. But then by theorem 3 of chapter 10, the principle of uniform boundedness, $|u| \leq b$ for all $u$ in $S$. \hfill \Box

**Theorem 4.** The closed unit ball in a Banach space $Z$ is compact in the weak topology iff $Z$ is reflexive.

**Proof.** The "if" part follows from theorem 3. The "only if" is due to Eberlein and Smulyan; see Dunford and Schwartz.

---

**BIBLIOGRAPHY**


13

LOCALLY CONVEX TOPOLOGIES AND THE KREIN-MILMAN THEOREM

The weak and weak* topologies are the weakest in which certain linear functionals are continuous. If one demands the continuity of even fewer functionals, one gets even weaker topologies. All these topologies have the property that openness can be defined in terms of convex sets. In this chapter we develop, and apply, the theory of such topologies.

Definition. A locally convex topological (LCT) linear space is a linear space over the reals with a Hausdorff topology that has the following properties:

(i) Addition is continuous; that is, \((x, y) \mapsto x + y\) is a continuous mapping of \(X \times X\) into \(X\).
(ii) Multiplication by scalars is continuous; that is, \((k, x) \mapsto kx\) is a continuous mapping of \(\mathbb{R} \times X\) into \(X\).
(iii) There is a basis for the open sets at the origin consisting of convex sets; that is, every open set containing the origin contains a convex open set containing the origin.

Note that the norm topology of a Banach space is a locally convex topology; the convex, open sets containing the origin that form a basis for the topology are the open balls centered at the origin.

Exercise 1. Show that the weak and weak* topologies are locally convex.

Exercise 2. Let \(\{\ell_a\}\) be a collection of linear functions in a linear space \(X\) over \(\mathbb{R}\) that separates points; that is, for any two distinct points \(x\) and \(y\) of \(X\) there is an \(\ell_a\) such that \(\ell_a(x) \neq \ell_a(y)\).

Exercise 3. Show that a continuous function \(f(a, b)\) on a product of two topological spaces whose values lie in a topological space is a continuous function of \(a\) when \(b\) is fixed.

Theorem 1.

(i) In a LCT linear space \(X\) the collection of open sets is translation invariant; that is, if \(T\) is an open set, so \(T - x\), for any \(x\) in \(X\).
(ii) If \(T\) is an open set, so is \(kT\), for \(k \neq 0\); in particular, \(-T\) is open.
(iii) Every point of an open set \(T\) is interior to \(T\).

Proof. According to exercise 3, \(x + y\) is a continuous function of \(y\) for \(x\) fixed. The inverse image of the open set \(T\) under this mapping is \(T - x\); this proves part (i). Part (ii) follows similarly.
(iii) By exercise 3, \(kx\) is a continuous function of \(k\) for \(x\) fixed. So the set of \(k\) for which \(kx\) lies in some open set \(T\) is an open subset of \(\mathbb{R}\). Suppose that \(T\) contains the origin; then \(k = 0\) belongs to this set, and by the above observation, so does an open interval containing \(k = 0\). That means that for \(k\) small enough, \(kx\) belongs to \(T\), but that is what it means for the origin to be an interior point of \(T\). This proves (iii).

Theorem 2. The continuous linear functionals in a LCT linear space \(X\) separate points. That is, if \(y\) and \(z\) are distinct points of \(X\), there is a continuous linear functional \(\ell\) such that
\[
\ell(y) \neq \ell(z).
\]

Proof. We construct a linear functional separating \(y\) and \(z\). Without loss of generality we take \(y = 0\). Since the topology is Hausdorff, there is an open set \(T\) containing \(y = 0\) but not \(z\); by (iii) of the definition of LCT space, we may take \(T\) to be convex. By theorem 1, 0 is an interior point of \(T\), so the gauge function \(p_T\) of \(T\) is finite, and
\[
p_T(u) < 1 \quad \text{for all} \, u \in T.
\]

According to the hyperplane separation theorem, theorem 5 of chapter 3, there exists a linear functional \(\ell\) satisfying

\[
\ell(u) < 1 \quad \text{for all} \, u \in T.
\]

\[
\ell(u) > 0 \quad \text{for all} \, u \in T.
\]

\[
\ell(u) < 0 \quad \text{for all} \, u \in T.
\]

\[
\ell(u) = 0 \quad \text{for all} \, u \in T.
\]
THE KREIN-MILMAN THEOREM

An extreme set consisting of a single point is called an extreme point.

Exercise 5. Show that the nonempty intersection of extreme sets is extreme.

The elementary properties of extreme set are contained in theorems 7 and 8 of chapter 1. The basic result concerning convex sets in finite dimensional spaces is the following theorem of Carathéodory:

Every compact convex subset \( K \) in \( \mathbb{R}^n \) has extreme points, and every point of \( K \) can be written as a convex combination of \( N + 1 \) extreme points.

Exercise 6. Furnish a proof of Carathéodory's theorem by induction on \( N \).

M. G. Krein and D. P. Milman have given the following beautiful—and useful—generalization of this result:

**Theorem 3.** Let \( X \) be a LCT linear space, \( K \) a nonempty, compact, convex subset of \( X \).

(i) \( K \) has at least one extreme point.

(ii) \( K \) is the closure of the convex hull of its extreme points.

**Proof.** Consider the collection \( \{E_j\} \) of all nonempty closed extreme subsets of \( K \). This collection is nonempty, for it contains \( K \) itself. Partially order this collection by inclusion. We claim that every totally ordered subcollection \( \{E_j\} \) has a lower bound. That lower bound is the intersection \( \cap E_j \). To see this, we have to show that \( \cap E_j \) is nonempty, closed, and extreme.

We claim that every finite subset of the totally ordered collection \( \{E_j\} \) has a nonempty intersection. This is because in being totally ordered by inclusion, the intersection of a finite subset of the collection \( \{E_j\} \) is the smallest member of that subset. To conclude that \( \cap E_j \) is nonempty, we argue indirectly: suppose that the intersection is empty. Then the union of the complements of the \( E_j \) cover \( K \). Since \( K \) is compact, a finite collection of these already cover \( K \), but then the intersection of these finite number of \( E_j \) is empty, contrary to what we have already shown. Being the intersection of closed sets, \( \cap E_j \) is closed. By exercise 5 the nonempty intersection of extreme subsets of a convex set \( K \) is itself an extreme subset of \( K \).

We conclude from Zorn's lemma that \( K \) has a closed extreme subset \( E \) that is minimal with respect to inclusion. We claim that such an \( E \) consists of a single point. To see this, suppose, on the contrary, that \( E \) contains two distinct points. According to theorem 2, there exists a continuous linear functional \( \ell \) that separates these points. Since \( E \) is compact, and \( \ell \) continuous and not constant on \( E \), \( \ell \) achieves its maximum on some proper subset \( M \) of \( E \). Since \( \ell \) is continuous and \( E \) is closed, \( M \) is closed. Since the inverse image of an extreme subset is extreme (see corollary 8 in chapter 1), the set \( M \) where a linear functional \( \ell \) assu...
E is an extreme subset of E. It is easy to show further (see theorem 7 of chapter 1) that if E is an extreme subset of K, and M an extreme subset of E, then M is an extreme subset of K. Since E is a minimal extreme subset of K, and M an extreme subset smaller than E, we have a contradiction, into which we got by assuming that E contains more than one point. We conclude therefore that a minimal E consists of a single point. This single point is an extreme point of K. This completes the proof of part (i), and gives a little more:

(i') Every closed, extreme subset of K contains an extreme point.

We turn now to the proof of part (ii). Denote by K_e the set of extreme points of K, and by \( \bar{K}_e \) the convex hull of K_e. To show that every point of K belongs to the closure of K_e is the same as showing that a point z which does not belong to the closure of K_e does not belong to K. According to exercise 4, the closure of K_e is convex. So, if z does not belong to the closure, then according to theorem 2' there is a continuous linear functional \( \ell \) such that

\[
\ell(y) \leq c \quad \text{for all } y \text{ in } \bar{K}_e, \quad \ell(z) > c. \tag{8}
\]

Since K is compact and \( \ell \) continuous, \( \ell \) achieves its maximum over K on some closed subset E of K. According to corollary 8', chapter 1, E is an extreme subset of K. According to part (i') of theorem 3 noted above, E contains some extreme point p of K. Since p belongs to K_e, and so to \( \bar{K}_e \), it follows from (8) that \( \ell(p) \leq c \). Since by construction \( \ell(p) = \max_{x \in E} \ell(x) \), \( \ell(x) \leq \ell(p) \leq c \) for all x in K. Since by (8), \( \ell(z) > c \), this proves that z does not belong to K.

\[ \square \]

13.3 THE STONE-WEIERSTRASS THEOREM

Theorem 4. Let S be a compact Hausdorff space, C(S) the set of all real-valued continuous functions on S. Let E be a subalgebra of C(S), that is,

(i) E is a linear subspace of C(S).
(ii) The product of two functions in E belongs to E.

In addition we impose the following conditions on E:

(iii) E separates points of S, that is, given any pair of points p and q, p \( \neq \) q, there is a function f in E such that f(p) \( \neq \) f(q).
(iv) All constant functions belong to E.

Conclusion: E is dense in C(S) in the maximum norm.

The classical Weierstrass theorem is a special case of this proposition, with S an interval of the x axis, and E the set of all polynomials in x. We present Louis de Branges's elegant proof, based on the Krein-Milman theorem, of Stone's generalization of the Weierstrass theorem.

The Stone-Weierstrass theorem

Proof. According to the spanning criterion, theorem 8 of chapter 8, E is dense in C(S) if the only bounded linear functional \( \ell \) or C(S) that is zero on E is the zero functional. According to the Riesz-Kakutani representation theorem, theorem 14 of chapter 8, the bounded linear functionals on C(S) are of the form

\[
\ell(f) = \int_{S} f \, dv,
\]

\( v \) a signed measure of finite total variation \( \|v\| = \int |dv| \). So what we have to show is that if \( \int f \, dv = 0 \) for all \( f \) in E, \( v = 0 \).

Suppose not; denote by U the set of signed measures of finite total mass is \( \leq 1 \) that annihilate all functions in E. This is a convex set, and according to Alaoglu's theorem, theorem 3 in chapter 12, compact in the weak* topology. So according to the Krein-Milman theorem, if U contained a nonzero measure, it would contain a nonzero extreme point; call it \( \mu \). Since \( \mu \) is extreme, \( \|\mu\| = 1 \). Since E is an algebra, if \( f \) and \( g \) belong to E, so does \( fg \). Since \( \mu \) annihilates every function in E,

\[
\int (fg) \, d\mu = 0.
\]

It follows that the measure \( gd\mu \) also annihilates every function in E.

Let \( g \) be a function in E whose values lie between 0 and 1:

\[
0 < g(p) < 1 \quad \text{for all } p \text{ in } S.
\]

Denote

\[
a = \|g\mu\| = \int g \, d\mu, \quad b = \|(1-g)\mu\| = \int (1-g) \, d\mu.
\]

Clearly a and b are positive. Add them:

\[
a + b = \int |d\mu| = 1.
\]

The identity

\[
\mu = a \frac{g\mu}{a} + b (1-g) \mu
\]

represents \( \mu \) as a nontrivial convex combination of \( g\mu \) and \( (1-g)\mu \), both points in U. Since \( \mu \) is an extreme point, \( \mu \) must be equal to \( g\mu \).

Define the support of the measure \( \mu \) to be the set of points p that have the property that \( |d\mu| > 0 \) for any open set N containing p. If \( \mu = g\mu/a \), it follows that \( g \) has the same value at all points of the support of \( \mu \).

We claim that the support of \( g\mu/a \) consists of a single point. To see this, suppose that both p and q, p \( \neq \) q, belong to the support \( \mu \). Since the functions in E separate points of S, there is a function h in E, \( h(p) \neq h(q) \). Adding a large enough constant
to $h$ and dividing it by another large constant, we obtain a function $g$ whose values lie between 0 and 1, and $g(p) \neq g(q)$. This contradicts our previous conclusion.

A measure $\mu$ whose support consists of a single point $p$, and $\|\mu\| = 1$, is a unit point mass at $p$. Therefore

$$\int f d\mu = f(p) \text{ or } -f(p).$$

Since, by hypothesis, the constant 1 belong to $E$, $\int f d\mu \neq 0$ for $f \equiv 1$ in $E$, a contradiction. \hfill \Box

13.4 Choquet's Theorem

The following further extension of Caratheodory's theorem holds on locally convex linear spaces:

**Theorem 5.** Let $X$ be a LCT linear space, $K$ a nonempty compact, convex subset of $X$, $K_e$ the set of extreme points of $K$. For any point $u$ of $K$ there is a probability measure $m_u$ on $K_e$, the closure of $K_e$, that is a measure satisfying:

$$m_u \geq 0, \quad \int_{K_e} dm_u = 1, \quad (9)$$

such that in the weak sense

$$u = \int_{K_e} \ell e \, dm_u. \quad (10)$$

The weak sense of the integral representation above is that for every continuous linear functional $\ell$ over $X$,

$$\ell(u) = \int_{K_e} \ell(e) \, dm_u(e). \quad (10')$$

**Proof.** For any continuous $\ell$ and $K$ compact, $\ell$ achieves its minimum and maximum on $K$. According to corollary 8' of chapter 1, the sets where $\ell$ achieves its minimum and maximum are extreme subsets of $K$. According to part (i') of theorem 3, these extreme subsets contain extreme points. Therefore for any $u$ in $K$ and for every continuous linear functional $\ell$,

$$\min_{p \in K_e} \ell(p) \leq \ell(u) \leq \max_{p \in K_e} \ell(p). \quad (11)$$

It follows from (11) applied to $\ell_1 - \ell_2$ that if $\ell_1$ and $\ell_2$ are equal on $K_e$, they are equal on $K$; therefore $\ell$ on $K_e$ uniquely determines the value of $\ell(u)$ for every $u$ in $K$. Denote the restriction of $\ell$ to $K_e$ by $f$:

$$f(q) = \ell(e), \quad q \in K_e. \quad (12)$$

Since $\ell(u)$ is determined by $f$, we can write

$$\ell(u) = u(f), \quad (12')$$

clearly a linear functional of $f$. We rewrite (11) as

$$\min_{q \in K_e} f(q) \leq u(f) \leq \max_{q \in K_e} f(q). \quad (13)$$

The set $L$ of functions $f$ defined in (12) form a linear subspace of the space $C(K_e)$ of continuous functions on $K_e$. We claim that we can extend the linear functional $u(f)$ defined in (12') from $L$ to all of $C(K_e)$ so that property (13) is preserved. To see this, we adjoin the function $f_0 \equiv 1$ to $L$ and define $u(f_0) = 1$. Then we appeal to theorem 1 of chapter 4, according to which a positive linear functional can be extended positively to the space of all functions. Since (13) implies that $u(f)$ is positive, such an extension is possible; imagine it done.

$K_e$ is a closed subset of the compact set $K$ and is therefore compact. We appeal now to the Riesz-Kakutani representation theorem (see chapter 8, theorem 14) according to which a bounded linear functional $u$ on $C(K_e)$ can be represented as

$$u(f) = \int_{K_e} f \, dm. \quad (14)$$

Since the functional is positive, so is the representing measure $m$; it follows from $u(f_0) = 1$ that $m(K_e) = 1$. Setting (12) and (12') into (14), we obtain (10'). \hfill \Box

Theorem 5 asserts that every point $u$ of the compact convex set $K$ can be represented as a continuous convex combination of points of the closure of the set of extreme points. This proviso is needed because the set of extreme points may not be closed, not even in a finite-dimensional space. Say we take in $\mathbb{R}^3$ the convex hull of the circle: $x^2 + y^2 = 1$, $z = 0$, and the interval:

$$x = 1, \quad y = 0, \quad -1 \leq z \leq 1.$$

The extreme points of the convex hull are $x = 1, y = 0, z = \pm 1$, and all points of the circle $x^2 + y^2 = 1, z = 0$ except $x = 1, y = 0, z = 0$.

**Exercise 7.** Let $v$ be a point in $K_e$ that does not belong to $K_e$; show that $v$ can then be represented as

$$v = \int_{K_e} \ell \, dm,$$

where $m$ is a probability measure on $K_e$ such that $m(v) = 0$.

**Exercise 8.** Deduce part (ii) of theorem 3 from theorem 5.
Choquet gave the following sharpening of theorem 5:

**Theorem 6 (Choquet).** Let $K$ be a nonempty compact, convex subset of an LCT linear space, and assume in addition that $K$ is metrizable. Then every point $u$ of $K$ can be represented in the weak sense as

$$u = \int_K e d\mu_u,$$

where $\mu_u$ is a probability measure on the set of extreme points.

**Proof.** For proof, see Phelps.

We call (10') a Choquet-type representation. In the next chapter we give many examples of Choquet-type representations of convex sets.

We present now a useful result that extends to LCT spaces, the following intuitively clear property of convex hulls of compact sets $S$ in finite-dimensional spaces: the points of $S$ that are added to $S$ to make i. convex contain no extreme points.

**Theorem 7.** Let $X$ be a LCT linear space, $S$ a compact subset of $X$. Suppose $K$, the closure of the convex hull of $S$, is compact. Then every extreme point of $K$ belongs to $S$.

**Proof.** Let $N$ be any open convex set that contains the origin. The open sets $y + N$ for $y$ in $S$ form an open cover of $S$; since $S$ is compact, a finite number of them cover $S$:

$$\cup(y_i + N) \supset S.$$  \hspace{1cm} (15)

Denote by $S_i$ the intersection $(y_i + N) \cap S$; it follows from (15) that

$$\cup S_i = S.$$ \hspace{1cm} (16)

Denote the closure of the convex hull of $S_i$ by $K_i$. Since $S_i \subset S$, it follows that $K_i \subset K$; since $K_i$ is closed and $K$ assumed compact, it follows that each $K_i$ is compact. Next we need

**Lemma 8.** Let $K_1$ and $K_2$ be a pair of compact convex sets in a LCT linear space. Then the convex hull of their union is compact.

**Proof.** Since $K_1$ and $K_2$ themselves are convex, it is easy to see that the convex hull of their union consists of all points of the form

$$ay_1 + (1 - a)y_2, \quad y_1 \in K_1, \quad y_2 \in K_2,$$

$$0 \leq a \leq 1.$$ 

These points are images of the triple product

$$K_1 \times K_2 \times I, \quad I = [0, 1]$$

under the mapping (17). The triple product (18) is compact, and according to the definition of a LCT space, the mapping (17) is continuous. It follows that the image of the compact set (18) is compact, as claimed in lemma 8.

From lemma 8 we deduce inductively that the convex hull of the union of a finite number of compact sets is compact. We turn to the compact sets $K_i$ defined above and claim that the convex hull of their union, denoted as $CH[K_1 \cup \ldots \cup K_n]$, contains $K$:

$$K \subset CH[K_1 \cup \ldots \cup K_n].$$ \hspace{1cm} (19)

We note that $K_i$ contains $S_i$. Therefore, by (16), $K_1 \cup \ldots \cup K_n$ contains $S_1 \cup \ldots \cup S_n = S$. By lemma 8, the right side of (19) is compact, and therefore closed. Thus it is a closed, convex set that contains $S$. But $K$ is defined to be the smallest such set, and it therefore is contained in $CH[K_1 \cup \ldots \cup K_n]$. This proves (19). In words, every point of $K$ is a convex combination of points of $F_j$. Since each $K_j$ is contained in $K$, it follows from the definition of extreme point: that each extreme point $p$ of $K$ belongs to a $K_i$.

By definition, $S_i$ is contained in $y_i + N$. Since $N$ is convex, the convex hull of $S_i$ belongs to $y_i + N$: $S_i \subset y_i + N$. For any set $R$, $R + N$ contains the closure of $R$. Since $K_i$ is the closure of $S_i$, it follows that $K_i \subset y_i + N + N = y_i + 2N$. Therefore, since each $y_i$ belongs to $S$,

$$\cup K_i \subset S + 2N.$$ \hspace{1cm} (20)

We have shown that each extreme point of $K$ belongs to some $K_i$, so it follows from the result above that every extreme point $p$ of $K$ belongs to $S + 2N$. Now $N$ is arbitrary; since $S$ is closed, the intersection of all sets $S + 2N$ is $S$ itself. Thus every extreme point $p$ of $K$ is contained in $S$.

**Exercise 9.** Show that if $S$ is a compact set in a Banach space, its closed convex hull is compact. Is this true in every LCT space?

**Note.** The prime examples of LCT linear spaces are Banach spaces in the weak* and weak* topologies. Other important examples are spaces of distributions. In view of the enormous success of the theory of distributions in the theory of partial differential equations and in harmonic analysis, it was thought that other locally convex topologies might play a similarly fruitful role; this hope has not yet been realized.

Other applications of the Krein-Milman theorem and its generalizations are described in Diestel and Uhl.

**BIBLIOGRAPHY**

14
EXAMPLES OF CONVEX SETS AND THEIR EXTREME POINTS

In this chapter we present a great variety of examples of convex sets, their extreme points, and Choquet-type integral representations of points of the set in terms of the extreme points. In some examples the extreme points are determined by a direct argument. Then a locally convex topology is introduced so that the convex set in question is compact and the Choquet-type representation is then derived via Choquet's theorem. In other examples the Choquet-type representation is derived directly by an analytic argument. The representation is then used to identify the extreme points of the set. In most of these examples the representation is unique.

14.1 POSITIVE FUNCTIONALS

Let \( Q \) denote a compact Hausdorff space, and \( C(Q) \) the space of continuous functions on \( Q \) whose values are real. We denote a linear functional \( \ell \) defined on \( C(Q) \) as positive if \( \ell(f) \geq 0 \) for all nonnegative \( f \) in \( C(Q) \). Recall from chapter 8 that a positive linear functional is bounded. Denote by \( P \) the collection of all positive linear functionals \( \ell \) that satisfy

\[
\ell(1) = 1.
\]

**Theorem 1.** \( P \) is a convex set whose extreme points are the point evaluations \( \varepsilon_r \), defined as

\[
\varepsilon_r(f) = f(r),
\]

for any point of \( Q \).
Proof. The convexity of $P$ is obvious. To see that every $e_r$, defined by (2), is an extreme point, imagine $e_r$ represented as

$$e_r = am + (1 - a)e, \quad m \text{ and } e \in P, \quad 0 < a < 1. \quad (3)$$

Let $f$ be any function in $C(Q)$ that is nonnegative and satisfies

$$f(r) = 0. \quad (4)$$

Set $f$ into (3); using (2) and (4) obtains

$$e_r(f) = f(r) = 0 = am(f) + (1 - a)e(f).$$

We claim that $m(f)$ and $e(f)$ are both zero; otherwise, one of them would have to be positive, the other negative, in contradiction to $f \geq 0$ and $e, m$ both positive functionals.

Every continuous $f$ can be decomposed into its positive and negative part:

$$f = f_+ - f_- \quad f_+ = \max(f, 0).$$

Both $f_+$ and $f_-$ are nonnegative, and if $f(r) = 0$, $f_+(r) = f_-(r) = 0$. It follows from this and the foregoing that if $f(r) = 0$, $m(f) = e(f) = 0$. In other words, the nullspaces of $m$ and $e$ contain the nullspace of $e_r$. Since the nullspace of a nontrivial functional has codimension 1, it follows that $e$ and $m$ are constant multiples of $e_r$; since all functionals in $P$ satisfy (1), it follows that the constant multiplier is 1. This proves that $e = m = e_r$, so $e_r$ is extreme.

We show now that the $e_r$ are the only extreme points of $P$. Let $e$ be any positive linear functional on $C(Q)$, normalized by (1). According to the Riesz-Kakutani representation theorem, there exists a nonnegative measure $m$ on $Q$ such that for every continuous function $f$ on $Q$,

$$e(f) = \int f \, dm. \quad (5)$$

Because of the normalization (1), $m(Q) = 1$; the measure $m$ is uniquely determined by the functional $e$. We claim that the only extreme points of the set of positive linear functionals normalized by (1) are those where the measure $m$ is concentrated at a single point. Otherwise, $m$ can be split as $am_1 + (1 - a)m_2$, where both $m_1$ and $m_2$ are nonnegative measures of total mass 1, and $m_1 \neq m_2$. Setting

$$e_j(f) = \int f \, dm_j, \quad j = 1, 2,$$

we get $e = ae_1 + (1 - a)e_2$. If $e$ were an extreme point, $e_1 = e_2 = e$, and so $e$ can be represented in form (5) by the distinct measures $m_1$ and $m_2$. This contradicts uniqueness of the representing measure. This completes the proof of theorem 1. \qed

14.2 CONVEX FUNCTIONS

In this section we make use of the notions and results of the theory of distributions as explained in Appendix B.

Definition. A real-valued function $f$ in $\mathbb{R}^n$ is convex if it satisfies

$$f \left( \sum a_j x_j \right) \leq \sum a_j f(x_j) \quad (7)$$

for all choices of $x_1, \ldots, x_N$ in $\mathbb{R}^n$ and all $a_j$ satisfying $a_j \geq 0$, $\sum a_j = 1$.

Here we consider convex functions $f$ of a single variable. It suffices to assume (7) to hold for $N = 2$:

$$f(ax + (1 - a)x) \leq a f(x) + (1 - a) f(z), \quad 0 \leq a \leq 1, \quad (8)$$

for all $x, z$. Setting $ax + (1 - a)x = y$, condition (7) is easily seen to be equivalent with the following: for $x < y < z$,

$$\frac{f(y) - f(x)}{y - x} \leq \frac{f(z) - f(y)}{z - y}. \quad (9)$$

It follows from (9) that every convex function is continuous and has right and left derivatives.

The second difference quotients

$$\frac{f(x + h) - 2f(x) + f(x - h)}{h^2} \quad (10)$$

converge in the sense of distributions to $f''$ as $h$ tends to zero. It follows from (9) that the difference quotients (10) are nonnegative; since the limit in the sense of distributions of a nonnegative distribution is nonnegative, it follows that for convex $f$,

$$0 \leq f'' \quad (11)$$

in the sense of distributions.

Convexity in an interval is defined in the same way; note that a function convex in an interval need not be continuous at the endpoints.
EXAMPLES OF CONVEX SETS AND THEIR EXTREME POINTS

We denote by \( C \) the set of a functions \( f \) convex in the interval \([0, 1]\) and satisfying
\[
    f(0) = 0, \quad f(1) = 1, \quad f(x) \geq 0 \quad \text{for} \quad 0 \leq x \leq 1. \quad (12)
\]

**Theorem 2.** \( C \) is a convex set, whose extreme points are the functions

\[
e_r(x) = \begin{cases} 
0 & \text{for } x \leq r \\
\frac{1}{1-r} & \text{for } r \leq x,
\end{cases} \quad (13)
\]

where \( 0 \leq r < 1 \), and

\[
e_1(x) = \begin{cases} 
0 & \text{for } x < 1 \\
1 & \text{for } x = 1.
\end{cases} \quad (13')
\]

**Proof.** First we show that all the functions \( e_r \) defined above are extreme points of \( C \). Suppose that \( e_r \) is represented as
\[
e_r = af + (1-a)g. \quad (14)
\]

We claim that both \( f \) and \( g \) are zero on \([0, r]\); for \( e_r \) is, and so otherwise by (14) one of the functions \( f \) or \( g \) would be negative. Since this is contrary to (12),
\[
f(x) = g(x) = 0, \quad 0 \leq x \leq r,
\]

follows. Similarly we claim that \( f(x) \) and \( g(x) \) are both equal to \( e_r \) on \([r, 1]\). For if not, one of them would be \( e_r \) at some point \( y > r \); a short calculation shows that this contradicts (9) with \( x = r, z = 1 \). This shows that \( e_r \) is extreme.

Let \( f \) be any convex function in \([0, 1]\) satisfying (12). We set \( f(r) = 0 \) for \( r < 0 \). Clearly, \( f \) thus extended remains convex. Let \( \phi \) be any \( C_0^\infty \) test function that is zero for \( r \geq 0 \). Then according to the theory of distributions
\[
    \int f \phi'' \, dr = \int f'' \phi \, dr \quad (15)
\]

where \( \phi' \) denotes differentiation with respect to \( r \). We choose now \( x \) in \( 0 < x < 1 \), and define the function \( \phi_x(y) \) by
\[
    \phi_x(r) = \begin{cases} 
x-r & \text{for } r \leq x, \\
0 & \text{for } r \geq x.
\end{cases} \quad (16)
\]

The function \( \phi_x \) is piecewise linear and \( \phi_x'' = \delta(r-x) \). If we could substitute \( \phi = \phi_x \) in (15), we would obtain
\[
f(x) = \int \phi_x(r) f''(r) \, dr. \quad (17)
\]

Since \( \phi_x \) is not \( C_0^\infty \), this is not legitimate, so we approximate \( \phi_x \) by a sequence \( \phi_x^\varepsilon \) of \( C_0^\infty \) functions. Since \( f(r) \) is continuous for \( r < 1 \), the left side of (15) tends to the left side of (17). On the other hand, the nonnegative distribution \( f'' \) is a nonnegative

**Exercise 1.** Find a version of theorem 2 for convex functions of \( n \) variables.

**Exercise 2.** Prove theorem 2 without the theory of distributions, using the theorem of Krein-Milman.

### 14.3 COMPLETELY MONOTONE FUNCTIONS

The difference operator \( D_a \), acting on functions of a single variable, is defined by
\[
(D_a f)(t) = f(t + a) - f(t). \quad (21)
\]

For \( a > 0 \), \( D_a \) maps functions \( f \) defined on \( \mathbb{R}_+ \) into functions defined on \( \mathbb{R}_+ \).
Definition. A real-valued function $f$ defined on $\mathbb{R}_+$ is called completely monotone (c.m.) if

$$(-1)^n \left( \prod_{1}^{n} D_{a_j} \right) f \geq 0 \quad \text{on } \mathbb{R}_+$$  \hspace{1cm} (22)

for all $a_j > 0$ and all $n = 0, 1, \ldots$.

The following result is due to S. Bernstein:

Theorem 3. Every completely monotone function $f$ on $\mathbb{R}_+$ can be represented as

$$f(t) = \int_{0}^{\infty} e^{-\lambda t} \, d\mu(\lambda),$$  \hspace{1cm} (23)

$m$ some nonnegative measure, $m(\mathbb{R}_+) < \infty$. Conversely, every function of form (23) is completely monotone.

Proof. To show that $f$ of form (23) is completely monotone we write

$$D_a f = \int_{0}^{\infty} D_a e^{-\lambda t} \, d\mu(\lambda) = \int (e^{-a\lambda} - 1) e^{-\lambda t} \, d\mu(\lambda);$$

then

$$(-1)^n \left( \prod_{1}^{n} D_{a_j} \right) f = \int \prod (1 - e^{-a_j \lambda}) e^{-\lambda t} \, d\mu(\lambda)$$

is clearly nonnegative.

Turning to the direct part, we will make use of the following properties of c.m. functions.

Lemma 4.

(i) The sum of two c.m. functions is c.m.

Suppose that $f$ is a c.m. function; then

(ii) $f$ is nonnegative.

(iii) $af$ is c.m. for $a > 0$.

(iv) $-D_a f$ is c.m. for $a > 0$.

(v) $T_a f = f(t + a)$ is c.m. for $a > 0$.

(vi) $H_b f = f(bt)$ is c.m. for $b > 0$.

(vii) $f$ is nonincreasing.

(viii) $f$ is convex.

Completely Monotone Functions

Proof. Parts (i) and (iii) follow from the fact that the operators $D_a$ appearing in condition (22) characterizing c.m. functions are linear. Part (ii) is (22), parts (iv) and (v) can be deduced by applying the operators $D_a$, respectively $T_a$, to (22), and noting that these operators commute with $D_a$. Part (vi) follows by applying $H_b$ to (22), and noting that

$$H_b D_a = D_{a(b-1)} H_b.$$  \hspace{1cm}

Part (vii) follows from (22) for $n = 1$, and (viii) from (22) for $n = 2$.

We define $X$ to be the space of all real-valued functions on $\mathbb{R}_+$, and take $K$ to be the subset of all c.m. functions, in the sense of (22), normalized by

$$f(\infty) = 1.\hspace{1cm} (24)$$

It follows from parts (i) and (iii) of lemma 4 that $K$ is a convex set.

Lemma 5. The extreme points of $K$ are of the form

$$e_\lambda(t) = e^{-\lambda t}, \quad 0 \leq \lambda < \infty, \hspace{1cm} (25a)$$

and

$$e_\infty(t) = \begin{cases} 0 & \text{for } t > 0 \\ 1 & \text{for } t = 0. \end{cases} \hspace{1cm} (25b)$$

Proof. By parts (ii) and (vii) of lemma 4, every c.m. function is nonnegative and nonincreasing. It follows from (24) that every $f$ in $K$ satisfies

$$0 \leq f(t) \leq 1.\hspace{1cm} (26)$$

Let $e$ be an extreme point of $K$; then in particular,

$$0 \leq e(t) \leq 1.\hspace{1cm} (26')$$

Suppose that for all $a > 0$, strict inequality holds:

$$0 < e(a) < 1.\hspace{1cm} (26'')$$

We define two auxiliary functions as follows:

$$f(t) = \frac{e(t) - e(a) + a}{1 - e(a)},$$

$$g(t) = \frac{e(t + a)}{e(a)}.\hspace{1cm} (27)$$

It follows from part (iii), (iv), and (v) of lemma 4 and (26'') that $f$ and $g$ belong to $K$.\hspace{1cm}
Clearly,

\[ e = (1 - e(a)) f + e(a) g. \]  \hspace{1cm} (27')

By definition of extreme point given in the previous chapter, it follows from (27') that \( f = g = e \). In particular, by (27), this implies that for all \( t \) and \( a \),

\[ e(t) e(a) = e(t + a). \]  \hspace{1cm} (28)

All continuous solutions of this equation are exponential functions. It follows from part (vii) of lemma 4 that every \( f \) in \( K \) is convex, and so is continuous for \( t > 0 \). We can conclude that

\[ e(t) = e^{-\lambda t}. \]

That \( \lambda \) is \( \geq 0 \) follows from part (vii) of lemma 4.

The cases where (26) fails to hold for some \( a > 0 \) can be easily handled. When \( e(a) = 1 \) for some \( a > 0 \), \( e = e_0 \); when \( e(a) = 0 \) for some \( a > 0 \), \( e = e_{\infty} \).

We introduce now the topology for functions that is the coarsest in which all the linear functionals \( \ell_t \):

\[ \ell_t(f) = f(t), \quad 0 \leq t, \]  \hspace{1cm} (29)

are continuous. The topology is the product topology

\[ \prod_{0 \leq t} f(t). \]

According to (26), the values of \( f \) in \( K \) lie between 0 and 1. So \( K \) is a subset of

\[ \prod_{0 \leq t} [0, 1]. \]

which by Tychonov's theorem is compact. So to show \( K \) compact, it suffices to show that \( K \) is closed. But this is easy: for fixed \( a_j \) and \( t \), the set of \( f \) that satisfy (22) is clearly closed. \( K \), being the intersection of these sets for all \( a_j \) and all \( t \geq 0 \), is closed.

We showed in lemma 5 that the extreme points of \( K \) are contained in the set \( \{e_\lambda\} \), \( 0 \leq \lambda \leq \infty \) defined by (25a), (25b). The set \( \{e_\lambda\} \), it is easy enough to show, is closed and therefore contains the closure of the set of extreme points.

We appeal now to formula (10) in theorem 5 of chapter 13. That formula with \( e \) given by (25) and \( \ell \) by (29), is precisely the desired representation formula (23).

Some corollaries and addenda:

\[ H_b e_\lambda = e_{\lambda b}, \quad b > 0. \]

This completes part (iii) of theorem 6.

We remark that an analogue of Bernstein's theorem holds in \( n \)-dimensional space, namely for functions defined on \( \mathbb{R}^n_+ \). It also holds for functions on \( \mathbb{Z}^n_+ \).

14.4 THEOREMS OF CARATHÉODORY AND BOCHNER

Definition. A skew-symmetric doubly infinite sequence \( \{a_n\} \) of complex numbers:

\[ a_{-n} = \bar{a}_n, \]  \hspace{1cm} (30)
is called positive definite if
\[ \sum_{n,k} a_{n-k} \phi_n \overline{\phi_k} \geq 0 \]  
(31)
for all finite sets of complex numbers \( \phi_n, -N \leq n \leq N \).

The following result is due to Toeplitz, Carathéodory, and Herglotz:

**Theorem 7.** All positive definite sequences can be represented uniquely as
\[ a_n = \int_0^\pi e^{i n \theta} d \mu(\theta), \]  
(32)
where \( \mu \) is a nonnegative measure on \( S^1 \). Conversely, every sequence of form (32) is positive definite.

**Proof.** First we show that every sequence of form (32) is positive definite. Substitute (32) for \( a_{n-k} \) into the left side of (31):
\[
\sum_{n,k} \int e^{i(n-k)\theta} d \mu(\theta) \phi_n \overline{\phi_k} = \int \left( \sum_n e^{i n \theta} \phi_n \right) \left( \sum_k e^{-i k \theta} \overline{\phi_k} \right) d \mu(\theta)
\]
\[
= \int \left( \sum_n e^{i n \theta} \phi_n \right)^2 d \mu(\theta)
\]
\[
= \int \sum_n e^{i n \theta} |\phi_n|^2 d \mu(\theta)
\]
is clearly nonnegative.

We turn now to the direct part of the theorem. We claim that if \( \{a_n\} \) is a positive definite sequence, then
\[ |a_m| \leq a_0 \quad \text{for all integers } m. \]  
(33)
To see this, set \( \phi_0 = 1, \phi_m = \phi, \) and all other \( \phi_n = 0. \) Substituting this into (31), we get, using (30), that
\[ a_0 + a_m \phi + \overline{a_m} \phi + a_0 |\phi|^2 \geq 0 \]
for all complex \( \phi; \) this implies (33).

According to the theory of distributions it follows from (33) that there exists a distribution \( a \) whose Fourier coefficients are \( a_n. \)
\[ a_n = \int_{S^1} e^{i n \theta} a d \theta. \]  
(34)

For any \( C^\infty \) function \( \psi, \)
\[ \int \psi a d \theta = \sum_n \overline{\psi}_n a_n, \]  
(34')
where \( \psi_n \) are the Fourier coefficients of \( \psi. \) It follows from (33) that the right side converges. We claim that \( a \) is nonnegative; to see this, take any trigonometric polynomial \( q_N \) of degree \( N, \)
\[ q_N(\theta) = \sum_{-N}^N \phi_n e^{i n \theta}. \]
Then \( |q_N(\theta)|^2 = \sum_{n,k} \phi_n \overline{\phi_k} e^{i(n-k)\theta}. \) Set \( \psi = |q_N|^2 \) in (34'). We get an expression on the right that, by (31), is nonnegative:
\[ \int a |q_N(\theta)|^2 d \theta = \sum_{n-k} a_{n-k} \phi_n \overline{\phi_k} \geq 0. \]  
(35)
Let \( q(\theta) \) be any \( C^\infty \) function on \( S^1; \) it is easy and classical to show that \( q \) can be approximated by a sequence \( \{q_N\} \) of trigonometric polynomials in the \( C^\infty \) topology. By definition of distribution, as \( N \) tends to \( \infty, \) and (35) tends to
\[ \int |q(\theta)|^2 a d \theta \geq 0, \]  
(36)
for all \( C^\infty \) functions \( q. \) Let \( p(\theta) \) be any \( C^\infty \) function that is positive on \( S^1. \) Then
\[ q(\theta) = \sqrt{p(\theta)} \]
is a \( C^\infty \) function; therefore (36) implies that
\[ \int p(\theta) a d \theta \geq 0 \]  
(37)
for any positive \( C^\infty \) function \( p. \) A distribution \( c \) with this property is called nonnegative. It is a classical result of the theory of distributions, see Appendix B, that every nonnegative distribution is a nonnegative measure. Thus \( a d \theta = dm \) and formula (34) is the desired formula (32).

**Theorem 7** can be extended to functions \( a \) defined on \( \mathbb{Z}^k, k \) any positive integer; the proof is the same.

**Note.** Carathéodory's own proof made use of his theorem on convex sets in finite-dimensional spaces. In section 14.6 we will give yet another proof, using the theory of positive harmonic functions.
Exercise 3. Denote by $P$ the set of all positive-definite sequences normalized by
\[ a_0 = 1. \]  
(38)
(a) Show that $P$ is a convex subset of the space $\ell^\infty$ of all bounded sequences.
(b) Show that $P$ is a compact subset of $\ell^\infty$ in the product topology.
and deduce the representation (32) using theorem 4 of chapter 13.

An important extension of Carathéodory's theorem is due to Bochner:

Definition. A skew-symmetric complex-valued continuous function $a(s)$ on $\mathbb{R}$:
\[ a(-s) = \overline{a(s)}, \]  
(40)
is called positive-definite if
\[ \sum a(s_j - s_k)\phi_j\overline{\phi_k} \geq 0 \]  
(41)
for all choices of $s_1, \ldots, s_N$ on $\mathbb{R}$, and for all complex numbers $\phi_1, \ldots, \phi_N$.

Exercise 4. Show that condition (41) is equivalent to the requirement that
\[ \int \int a(s - t)\phi(s)\overline{\phi(t)} dt \, ds \geq 0 \]  
(41')
for all continuous, complex-valued functions $\phi$ with compact support.

Theorem 8. Every continuous positive-definite function $a$ can be represented uniquely as
\[ a(s) = \int e^{is\sigma}dm(\sigma), \]  
(42)
m a nonnegative measure on $\mathbb{R}$, $m(\mathbb{R}) < \infty$. Conversely, every function of form (12) is positive-definite.

Proof. We show first that every function of form (42) is positive definite. Setting (42) into the left side of (41') yields
\[ \int \int e^{is(t-t')}\phi(s)\overline{\phi(t)} dt \, ds \, dm(\sigma) = \int |\phi(\sigma)|^2 dm(\sigma), \]
where $\phi$ is the Fourier transform of $\phi$. Clearly, the right side is nonnegative.

To construct the measure $m$ for a given positive-definite function, we proceed as
in the discrete case. We deduce from (41), analogously to (33), that
\[ |a(s)| \leq a(0). \]  
(43)
We recall from section B.5 of Appendix B the Schwartz class of functions $S$, consisting of functions $f(s)$ all of whose derivatives $\partial_s^n f(s), n = 0, 1, \ldots$, tend to zero faster than $|s|^{-k}$ as $|s| \to \infty$, for any $k$. $S'$ is the dual of $S$; its elements are called tempered distributions. The function $a$, is, according to (43), bounded; therefore it belongs to $S'$. Therefore $a$ has a Fourier inverse $\hat{b}$ that also belongs to $S'$. The Parseval relation
\[ \int \hat{b}f \, ds = \int cf \, ds \]  
(44)
holds for all $f$ in $S$, where $\hat{f}$ denotes the Fourier transform of $f$.

According to exercise 4, (41') holds for all $C^\infty$ functions $\phi$ with compact support.
Introduce in (41') $s - t = r$ and $s$ as new variables:
\[ \int \int a(r)\phi(s)\overline{\phi(s-r)} ds \, dr \geq 0. \]  
(45)
Denote by $f$ the convolution
\[ \int \phi(s)\overline{\phi(s-r)} ds = f(r); \]  
(46)
f belongs to $C^\infty$ and has compact support, and (45) can be written as
\[ \int a(r)f \, dr \geq 0. \]  
(45')
Denote by $\psi$ the Fourier transform of $\phi$; taking the Fourier transform of (46) gives
\[ |\psi(\sigma)|^2 = \hat{f}(\sigma). \]  
(46')
Formula (44) expresses the left side of (45') in terms of $b$ and $\hat{f}$; using formula (46')
for $\hat{f}$, we get
\[ \int b(\sigma)|\psi(\sigma)|^2 d\sigma \geq 0. \]  
(45'')
Let $p(\sigma)$ be any nonnegative $C^\infty$ function with compact support on $\mathbb{R}$. Then $\psi = p^{1/2}$ too is $C^\infty$ with compact support, and so belongs to $S$. Setting $\psi^2 = p$ into
(45''), we get
\[ \int b(\sigma)p(\sigma) \, d\sigma \geq 0 \]
for every nonnegative $C^\infty$ function $p$ with compact support. Such a distribution $b$ is called nonnegative. According to theorem 13 of Appendix B, $b$ is a nonnegative measure: $b(\sigma) \, d\sigma = dm$.

We claim that the total mass of $m$ is finite:

$$\int dm = \int_{-\infty}^{\infty} b(\sigma) \, d\sigma < \infty. \quad (47)$$

For let $g$ be any nonnegative $C^\infty$ function of compact support that is equal 1 on $[-1, 1]$. Define $g_n(\sigma) = g(\sigma / n)$. Denote by $f$ the Fourier inverse of $g$. Then the Fourier inverse of $g_n$ is $f_n(\sigma) = nf(\sigma)$. Set $f_n$ into (44):

$$\int b(\sigma) \, g \left( \frac{\sigma}{n} \right) \, d\sigma = \int a(s) n \, f(ns) \, ds. \quad (48)$$

The measure $b$ and the function $g$ are nonnegative, and $g(\sigma) = 1$ for $|\sigma| \leq 1$. Therefore the left side of (48) is greater than

$$\int_{-\infty}^{\infty} b(\sigma) \, d\sigma. \quad (48')$$

On the other hand, according to (43), $|a(s)| \leq a(0)$. Therefore the right side of (48) is less than

$$a(0) \int |f(ns)| \, ds = a(0) \int |f(s)| \, ds,$$

a quantity independent of $n$. This shows that the integrals (48') are bounded independently of $n$, proving (47).

It follows from (47) that $a(s)$ can be represented pointwise as the Fourier transform of $b$:

$$a(s) = \int e^{is\tau} \, dm$$

for every $s$, as claimed in (42). The uniqueness of a representation of form (42) follows from the uniqueness of the Fourier transform. \hfill \Box

Denote by $P$ the set of positive definite functions $\varphi$ normalized by $a(0) = 1$. It follows from theorem 8 that the extreme points of $P$ are the exponentials $e^{is\varphi}$, $\sigma$ real. Thus (42) is seen as a Choquet-type representation of positive definite functions.

Theorem 8 is easily extended to $n$ dimensions; see Rudin's book, Fourier Analysis on Groups.

Laurent Schwartz has given the following extension of Bochner's theorem.

**Definition.** A skew-symmetric complex-valued tempered distribution $a(s)$ on $\mathbb{R}$ is called positive definite if

$$\int \int a(s - t) \phi(s) \phi(t) \, ds \, dt \geq 0$$

for all $C^\infty_c$ functions $\phi$.

**Theorem 8'.** Every positive definite tempered distribution is the Fourier transform of a nonnegative measure of class $\mathcal{S}'$.

Schwartz has extended his theorem to $\mathbb{R}^n$.

### 14.5 A THEOREM OF KREIN

**Definition.** Let $p$ be a continuous real-valued even function defined on $\mathbb{R}$:

$$p(-t) = p(t).$$

$p$ is called *evenly positive definite* if

$$\int \int p(s - t) \phi(s) \phi(t) \, ds \, dt \geq 0 \quad (49)$$

for all real-valued, continuous, even functions $\phi$ of compact support:

$$\phi(-s) = \phi(s).$$

Clearly, every even function of form (42) has this property. These functions can be written as

$$p(s) = \int_{-\infty}^{\infty} \cos \sigma s \, dm(\sigma). \quad dm \geq 0.$$

These are, however, not all. For all real $\lambda$, and all even, real-valued continuous functions $\phi$ with compact support

$$\int \int \cosh \lambda(s - t) \phi(s) \phi(t) \, ds \, dt = \int e^{\lambda s} \phi(s) \, ds \int e^{\lambda t} \phi(t) \, dt.$$

This shows that $\cosh \lambda s$ is evenly positive definite. But then so is

$$p(s) = \int \cosh \lambda s \, dm(\lambda),$$

$n$ any nonnegative measure for which the integral converges for all $s$.

Similarly an even, real-valued function on $\mathbb{R}$ is called *oddly positive definite* if (49) holds for all real-valued, continuous odd functions $\phi$. Examples of such a function are $-\cosh \lambda s$ and superpositions.
M. G. Krein has proved the following result:

**Theorem 9.** Every real, even, continuous function \( p \) on \( \mathbb{R} \) that is evenly positive-definite can be represented uniquely as

\[
p(s) = \int_0^\infty \cos \sigma s \, dm(\sigma) + \int_0^\infty \cosh \lambda s \, dn(\lambda),
\]

\( m \) and \( n \) nonnegative measures. Similarly, every oddly positive-definite function can be represented as

\[
q(s) = \int_0^\infty \cos \sigma s \, dm(\sigma) - \int_0^\infty \cosh \lambda s \, dn(\lambda).
\]

An easy consequence is

**Theorem 9'.** Denote by \( P \) the set of evenly positive-definite functions, normalized by \( p(0) = 1 \). \( P \) is a convex set, and its extreme points are

\[
\cos \sigma s, \quad \sigma \geq 0, \quad \text{and} \quad \cosh \lambda s, \quad \lambda \geq 0.
\]

For a proof we refer to Krein.

### 14.6 POSITIVE HARMONIC FUNCTIONS

In chapter 11, section 11.6, it was shown that every harmonic function \( h \) defined in the unit disk and positive there can be represented uniquely by Poisson’s formula

\[
h(re^{ix}) = \frac{1 - r^2}{1 - 2r \cos(x - \theta) + r^2} \int dm(\theta), \quad (50)
\]

\( m \) a nonnegative measure.

It is easy to verify that the Poisson kernel has for \( r < 1 \) the following Fourier expansion:

\[
\frac{1 - r^2}{1 - 2r \cos \lambda + r^2} = \sum_{k=0}^{\infty} r^{|k|} e^{ik\lambda}. \quad (51)
\]

Setting this into (50) gives the Fourier expansion of \( h \):

\[
h(re^{ix}) = \sum b_k r^{|k|} e^{ikx}, \quad (52)
\]

where

\[
b_k = \int e^{-i\theta k} \, dm(\theta). \quad (52')
\]

We show now how to deduce Carathéodory's theorem, theorem 7, from (50); this proof is due to Herglotz. Let \( \{a_n\} \) be a positive definite sequence in the sense of (31). It follows then from (33) that the sequence \( a_n \) is bounded; therefore the series

\[
k(re^{ix}) = \sum_{|k|\leq n} a_k r^{|k|} e^{ikx}
\]

converges for \( r < 1 \), uniformly for \( r < 1 - \delta \). Clearly, \( k \) is a harmonic function of \( x, y \), where \( x + iy = re^{ix} \) in the unit disk. We claim that \( k \) is positive. To see this, we rewrite the left side of (51) by introducing \( n - k = \ell \) as a new variable. We get

\[
\sum \alpha_\ell \sum_n \phi_n \phi_{n-\ell} \geq 0. \quad (53')
\]

We want to choose \( \{\phi_n\} \) so that for all \( \ell \) and \( r < 1 \), \( \ell \) given

\[
\sum_n \phi_n \phi_{n-\ell} = r^{|\ell|} e^{i\ell x}. \quad (54)
\]

To satisfy (54) we multiply it by \( e^{-i\theta \ell} \) and sum over \( \ell \). We get

\[
\sum \phi_n \phi_{n-\ell} e^{-i\theta \ell} = \sum r^{|\ell|} e^{i\ell(x-\theta)}. \quad (54')
\]

The left side can be written as

\[
\sum \phi_n e^{-i\theta (n-\ell)} = | \sum \phi_n e^{-i\theta n} |^2;
\]

the right side, by (51), is the Poisson kernel, which is positive. Therefore we can set

\[
\sum \phi_n e^{-i\theta n} = \left( \frac{1 - r^2}{1 - 2r \cos(x - \theta) + r^2} \right)^{1/2}. \quad (55)
\]

This choice of \( \{\phi_n\} \) satisfies (54'), from which (54) follows. Setting (54) into (53') shows that \( k(re^{ix}) \) is positive for \( r < 1 \).

Once \( k \) has been shown to be positive, we can appeal to the Herglotz-Riesz theorem, theorem 6 in chapter 11, and obtain a representation of form (50) for \( k \). As was shown above, this in turn gives formula (52') for the coefficients \( a_k \) in the Fourier expansion of \( k \). This is the desired formula (32). \( \square \)

Denote by \( H \) the linear space of real-valued harmonic functions in the open unit disk. Denote by \( P \) the subset of positive harmonic functions \( h \), normalized by

\[
h(0) = 1. \quad (56)
\]

Clearly, \( P \) is a convex subset of \( H \). From the uniqueness of the measure in representation (50), we deduce, as in earlier sections of this chapter, that all extreme points
EXAMPLES OF CONVEX SETS AND THEIR EXTREME POINTS

of $P$ are of the form

$$ e_z = \frac{1 - r^2}{1 - \beta r (\chi - \theta) + r^2}. \quad (57) $$

This shows that the Herglotz-Riesz representation of positive harmonic functions is a Choquet-type representation.

**Exercise 5.** We impose on $H$ the weakest topology that makes continuous all linear functionals

$$ h \to h(z), \quad |z| < 1. \quad (58) $$

Show that the convex set $P$ defined above is compact in this topology. (Hint: Use Harnack's theorem on positive harmonic functions.)

14.7 THE HAMBURGER MOMENT PROBLEM

A sequence of real numbers $a_0, a_1, \ldots$ is called positive in the sense of Hankel ($H$ positive) if

$$ \sum_{n,k} a_{n+k} \xi_n \xi_k \geq 0 \quad (59) $$

for all finite collection of real numbers $\xi_n, n = 0, 1, \ldots, N$.

Let $m$ be a nonnegative measure on $\mathbb{R}$ all of whose moments are finite.

$$ \int_{\mathbb{R}} t^{2n} \, dm(t) < \infty, \quad n = 0, 1, \ldots. \quad (60) $$

Define

$$ a_\ell = \int_{\mathbb{R}} t^{\ell} \, dm(t), \quad \ell = 0, 1, \ldots. \quad (61) $$

We claim that this sequence is $H$ positive, for

$$ \sum_{n,k} a_{n+k} \xi_n \xi_k = \int \sum_{n,k} t^{n+k} \xi_n \xi_k \, dm(t) = \int \left( \sum t^n \xi_n \right)^2 \, dm(t) \geq 0 \quad (62) $$

Conversely, Hamburger has proved:

**Theorem 10.** Every $H$ positive sequence $\{a_n\}$ can be represented in the form (61).

For proof, we refer to chapter 33. An interesting fact is that there are $H$ positive sequences that can be represented in form (61) in only one way, as well as others that

have several distinct representations. Why this is so will be explained in chapter 33 on self-adjoint operators.

Denote by $H_0$ the set of all $H$ positive sequences normalized by $a_0 = 1$. It follows from theorem 10 that every extreme point $e$ of $H_0$ is of the form

$$ e_k(t) = t^k, \quad k \in \mathbb{Z}_+; \; t \text{ real.} \quad (63) $$

It is not hard to show that conversely, every sequence $e(t)$ of form (63) is an extreme point of $H_0$. Thus (61) is a Choquet-type representation of the set $H_0$.

Examples of Hankel positive sequences:

Take in (61) as

$$ \frac{dm(t)}{dt} = \begin{cases} t^{\delta-1} & \text{for } \delta \leq t \leq 1, \; \delta > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (64) $$

Then

$$ a_\ell = \int_0^1 t^{\ell} t^{\delta-1} \, dt = \frac{1}{t + \delta} \quad (65) $$

Thus

$$ 0 \leq \sum_{n+k} \frac{\xi_n \xi_k}{n+k+\delta} \quad (66) $$

for all real $\xi_n$.

In conclusion, we note that theorem 10 fails to hold in more than one variable.

14.8 G. BIRKHOFF'S CONJECTURE

**Definition.** An $n \times n$ matrix $S = (s_{ij})$ is called doubly stochastic if

(i) all entries are nonnegative,

$$ s_{ij} \geq 0 \quad (66') $$

(ii) all row sums and column sums are equal to 1,

$$ \sum_i s_{ij} = 1, \quad \sum_j s_{ij} = 1 \quad \text{for all } i, \text{ resp. } j. \quad (66') $$

It is obvious that the set $D$ of all doubly stochastic matrices forms a convex set in $\mathbb{R}^{n^2}$.

A permutation $p$ of $n$ objects is a one-to-one map of the indices $1, \ldots, n$ onto themselves. The associated permutation matrix $P$ is defined by
EXAMPLES OF CONVEX SETS AND THEIR EXTREME POINTS

\[ p_{ij} = \begin{cases} 
1 & \text{if } j = p(i) \\
0 & \text{if } j \neq p(i). 
\end{cases} \]

Clearly, each row, and each column of \( P \) permutation matrix, contains exactly one entry equal to 1 and all others are zero. This shows that each \( P \) is doubly stochastic, that is, \( P \in D \). We claim that each \( P \) is an extreme point of the set \( D \). To see this, suppose that \( P \) were the midpoint of an interval, whose endpoints

\[ P = \frac{Q + P}{2} \]

both belong to \( D \). Clearly, it follows from (66) that if \( p_{ij} = 0 \), then \( q_{ij} = 0 \), and from (66') that if \( p_{ij} = 1 \), then \( q_{ij} = 0 \). Since all entries of \( P \) are either 0 or 1, it follows that \( Q = 0 \). This proves that \( P \) is extreme.

Conversely, D. König and G. Birkhoff have shown that all extreme points of \( D \) are permutation matrices \( P \).

**Exercise 6.** Prove the König-Birkhoff theorem.

By Carathéodory's theorem, it follows that all doubly stochastic matrices are convex combinations of permutation matrices. This representation, however, is not unique in general.

**Definition.** An \( n \times n \) matrix \( S = (s_{ij}) \) is called doubly stochastic if

(i) all entries are nonnegative,

\[ s_{ij} \geq 0. \quad (67) \]

(ii) all row sums and column sums are \( 1 \),

\[ \sum_j s_{ij} \leq 1, \quad \sum_i s_{ij} \leq 1 \quad (67') \]

for all \( i \), respectively \( j \).

We denote the set of all doubly stochastic matrices by \( D_0 \). Clearly, \( D_0 \) is a convex set that contains the set \( D \). We call a matrix \( P_0 \) a subpermutation matrix if its entries are either 0 or 1, and if each row and column contains at most a single entry 1. Every \( P_0 \) belongs to \( D_0 \). The argument used above to show that every \( P \) is an extreme point of \( D \) can be used to prove that every \( P_0 \) is an extreme point of \( D_0 \). Conversely,

**Exercise 7.** Show that all extreme points of \( D_0 \) are subpermutation matrices \( P_0 \).

We turn now to infinite matrices \( S = (s_{ij}), i, j \in \mathbb{Z}_+ \). The notions of doubly stochastic, doubly substochastic, permutation, and subpermutation matrices are de-
any \( S \),

\[
\ell(S) = \ell(S_n).
\] (73)

Denote the projection of \( Z \) by \( Z_n \). As remarked earlier, the extreme points of the set of doubly substochastic \( n \times n \) matrices are the \( n \times n \) subpermutation matrices. It follows from Carathéodory's theorem that on a compact convex set, a continuous linear functional takes its maximum one of the extreme points. We have that

\[
\ell(Z_n) \leq \sup_{P_n} \ell(P_n),
\] (74)

where the \( P_n \) are \( n \times n \) subpermutation matrices. Such a \( P_n \) is the projection of a subpermutation matrix \( P_0 \) of infinite order, whose elements not in the first \( n \) rows and columns are set to 0. By (73),

\[
\ell(P_n) = \ell(P_0) \quad \text{and} \quad \ell(Z_n) = \ell(Z).
\] (75)

Combining (74) with (75), we obtain

\[
\ell(Z) \leq \sup_{P_0} \ell(P_0).
\]

This combined with (71'') shows that (70) cannot hold. Therefore every \( Z \) in \( D_0 \) belongs to the closure of the convex hull of \( \{P_0\} \).

To show that, conversely, all points of the closure of the convex hull of \( \{P_0\} \) belong to \( D_0 \), we rewrite the criterion (67) and (67') for belonging to \( D_0 \) as follows:

\[
\ell_{ij}(S) \geq 0, \quad \sum_{j < a} \ell_{ij}(S) \leq 1, \quad \sum_{i < a} \ell_{ij}(S) \leq 1.
\] (76)

(76')

for all positive integers \( n \). Since by definition of the topology, the functionals (76), (76') are continuous, and since these inequalities hold on the convex hull of \( \{P_0\} \), it follows that they hold on its closure. This completes the proof of part (ii).

The proof of part (i) is based on theorem 6 of chapter 13, which says that the extreme points of the closure of the convex hull of a set \( S \) belong to \( S \), provided both sets are compact. In order to apply that theorem to \( S = \{P_0\} \), we have to verify that both \( D_0 \) and \( \{P_0\} \) are compact sets. To see this, we note that the topology we have imposed is the weak product topology

\[
\prod_{ij} s_{ij}.
\]

The entries of \( S \) in \( D_0 \) lie in \([0, 1]\), so \( D_0 \) is a subset of

\[
\prod_{ij} [0, 1].
\]

G. BIRKHOFF'S CONJECTURE

According to Tychonov's theorem, this is a compact set. We have shown already in part (ii) that \( D_0 \) is a closed set; being a closed subset of a compact set makes \( D_0 \) compact.

Similarly, in order to show that \( \{P_0\} \) is compact, it suffices to show that this set is closed. The matrices \( P_0 \) are characterized by the inequalities (76') and

\[
\ell_{ij}(S) \in \{0, 1\}.
\]

Each of these sets is closed; therefore so is their intersection \( \{P_0\} \).

We now appeal to theorem 6 of chapter 13. It states that given a compact set such as \( \{P_0\} \) whose closed convex hull—which by part (ii) of theorem 11 is \( D_0 \)—is also compact, then all extreme points of the closed convex hull belong to the original compact set. This completes the proof of part (i) of theorem 11.

\[ \square \]

Theorem 12.

(i) Every extreme point of the set \( D \) of doubly stochastic infinite matrices is a permutation matrix \( P \).

(ii) \( D \) is the closure of the convex hull of the set \( \{P\} \) of permutation matrices in the coarsest topology that makes continuous the linear functionals \( \ell_{ij}, \ell_i, \text{and} \ell^j \), where \( \ell_{ij} \) are defined by (69) and

\[
\ell_{ij}(S) = \sum_j s_{ij}, \quad \ell_i(S) = \sum_i s_{ij}.
\]

Proof. (i) We start by showing that \( D \) is an extreme subset of \( D_0 \). Suppose that \( S \) in \( D \) lies on an interval

\[
S = aT + bR, \quad T \text{ and } R \text{ in } D_0, \quad a + b = 1, \quad 0 < a, 0 < b.
\]

Form the row and column sum of both sides:

\[
\sum_j s_{ij} = a \sum_j t_{ij} + b \sum_j r_{ij}, \quad \sum_i s_{ij} = a \sum_i t_{ij} + b \sum_i r_{ij}.
\]

Since \( S \) belongs to \( D \), the sums on the left are \( = 1 \); since \( T \) and \( R \) belong to \( D_0 \), the sums on the right are \( \leq 1 \). Since the two sides are equal, the sums on the right must all be \( = 1 \). But this means that \( T \) and \( R \) belong to \( D \) and this proves that \( D \) is an extreme subset of \( D_0 \).

We noted in theorem 7 of chapter 1 that being an extreme subset is a transitive relation among convex sets. Thus an extreme point \( E \) of \( D \) is an extreme point of \( D_0 \). According to part (ii) of theorem 11, all extreme points of \( D_0 \) are subpermutation matrices \( P_0 \), so \( E = P_0 \). Since \( E \) belongs to \( D \), it follows that \( E \) is not a subbut a genuine permutation matrix \( P \).

This completes the proof of part (i) of theorem 12; part (ii) can be proved along the lines of the argument presented for part (iii) of theorem 11. Note, however, that
part (ii) cannot be proved by appealing to the Krein-Milman theorem, since the set \( D \) is not closed, and therefore not compact.

Theorem 12 was conjectured by Garrett Birkhoff. The preceding theorems and proofs are due to Kiefer and Kendall; see D. G. Kendall. \( \square \)

14.9 De Finetti's Theorem

The setting of probability theory is a space \( \Omega \), in which a \( \sigma \)-algebra \( \Sigma \) is specified. The sets in \( \Sigma \) represent all possible events; a probability measure on \( \Sigma \) represents the probability of their occurrence. An infinite sequence of occurrences is modeled by the direct product \( \mathbb{Z} \times \Omega \). The events in \( \mathbb{Z} \times \Omega \) form the smallest \( \sigma \)-algebra that contains all the cylinder sets, formed by the product sets

\[
\prod E_j,
\]

where \( E_j \) belongs to the \( \sigma \)-algebra \( \Sigma \) in \( \Omega \), and all but a finite number of the sets \( E_j \) are the whole space \( \Omega \).

A probability measure \( m \) on the cylinder sets of \( \mathbb{Z} \times \Omega \) is called \emph{invariant under permutations} if for all cylinder sets

\[
m \left( \prod E_j \right) = m \left( \prod E_{p(j)} \right);
\]

where \( j \to p(j) \) is a permutation of the indices, such that \( p(j) = j \) for all but a finite number of \( j \).

The set of probability measures on \( \mathbb{Z} \times \Omega \) that are invariant under permutations is clearly a \emph{convex set}, where convex combinations of measure is defined in the obvious way.

A probability measure \( m \) on the \( \sigma \)-algebra \( \Sigma \) in \( \Omega \) induces the \emph{product measure} on the cylinder sets of \( \mathbb{Z} \times \Omega \) by the formula

\[
m \left( \prod E_j \right) = \prod m(E_j).
\]

Since all but a finite number of the \( E_j \) are equal to \( \Omega \), all but a finite number of the factors on the right are equal to 1. Clearly, the product measure on \( \mathbb{Z} \times \Omega \) induced by a measure on \( \Omega \) is invariant under permutation.

De Finetti proved the following important result:

Theorem 13.

(i) The extreme points of the set of permutation invariant probability measures on \( \mathbb{Z} \times \Omega \) are the product measures.

(ii) Each measure on \( \mathbb{Z} \times \Omega \) invariant under permutation can be expressed in a unique fashion as an integral over the product measures.

Clearly this is a Choquet type of result. For a proof, see de Finetti or any advanced text on probability.

14.10 Measure-Preserving Mappings

In this section \( \Omega \) denotes a compact metric space, \( T \) a homeomorphism of \( \Omega \) onto \( \Omega \). It can be shown that there exists at least one probability measure on the Borel subsets of \( \Omega \) that is invariant under \( T \). There may be many. The collection of all invariant probability measures form a convex set. The following result of John Oxtoby sheds light on the structure of this collection:

Theorem 14.

(i) The extreme points of the convex set of probability measures invariant under \( T \) are those measures with respect to which \( T \) is ergodic.

(ii) Every invariant measure can be represented as an integral over the ergodic measures. This representation is unique.

\[ \begin{align*}
\Omega &= \Omega_1 \cup \Omega_2, \\
m(\Omega_1) &> 0, \quad m(\Omega_2) > 0,
\end{align*} \]

so that both \( \Omega_1 \) and \( \Omega_2 \) are invariant under \( T \).

Suppose now that \( m \) is invariant under \( T \) but that \( T \) is not ergodic with respect to \( m \). Then there is a decomposition of \( \Omega \) as above. We define two new measures \( m_1 \) and \( m_2 \) as the restrictions of \( m \) to \( \Omega_1 \), \( \Omega_2 \), respectively. That is, for any Borel set \( S \),

\[
m_1(S) = \frac{m(S \cap \Omega_1)}{m(\Omega_1)} , \quad m_2(S) = \frac{m(S \cap \Omega_2)}{m(\Omega_2)} .
\]

Clearly, \( m_1 \) and \( m_2 \) are probability measures, and they are invariant under \( T \). The measure \( m \) is a convex combination of them:

\[
m = m_1(\Omega_1)m_1 + m_2(\Omega_2)m_2 .
\]

Since \( m_1 \neq m_2 \), this shows that if \( m \) is not ergodic, it is not an extreme point.

Conversely, we show that if \( m \) is not an extremal point, it is not ergodic. Suppose that
EXAMPLES OF CONVEX SETS AND THEIR EXTREME POINTS

\[ m = am_1 + (1-a)m_2, \quad 0 < a < 1, \quad m_1 \neq m_2. \]

We first take the case that \( m_1 \) is absolutely continuous with respect to \( m_2 \). By the Radon-Nikodým theorem,

\[ m_1 = f m_2, \quad f \text{ nonnegative and in } L^1(m_2). \]

Since both \( m_1 \) and \( m_2 \) are invariant under \( T \), so is \( f \). Since \( m_1 \neq m_2 \), \( f \neq 1 \); therefore there exists a positive number \( c \) such that the sets \( \Omega_1 = \{ \omega | f(\omega) > c \} \) and \( \Omega_2 = \{ \omega | f(\omega) \leq c \} \), both having positive measure \( m_2 \). Since \( f \) is invariant under \( T \), \( T \) maps \( \Omega_1 \) and \( \Omega_2 \) onto themselves.

Substituting \( m_1 = f m_2 \) into the expression for \( m \) as convex combination of \( m_1 \) and \( m_2 \) gives

\[ m = [af + (1-a)m_2]. \]

It follows that \( \Omega_1 \) and \( \Omega_2 \) have positive \( m \) measures; this shows that \( T \) is not ergodic with respect to \( m \).

The case when \( m_1 \) is not absolutely continuous with respect to \( m_2 \) is just as simple. Then there exist sets \( E \) whose \( m_2 \) measure is zero, but \( m_1(E) > 0 \). Denote by \( s \) the quantity

\[ s = \sup m_1(E), \quad m_2(E) = 0. \]

Let \( E_n \) be a maximizing sequence:

\[ \lim m_1(E_n) = s, \quad m_2(E_n) = 0. \]

Denote by \( F \) the union of the \( E_n \). Clearly, \( m_1(F) = s, m_2(F) = 0 \). It follows that \( F \) is invariant under \( T \); for if not, the set \( (F \cup T F) \) would have \( m_1 \) measure greater than \( m(F) = s \) but \( m_2 \) measure 0 contrary to the definition of \( s \).

We claim that in the decomposition \( \Omega = F \cup F^c \), both pieces have positive \( m \)-measure. For, using the expression of \( m \) as linear combination of \( m_1 \) and \( m_2 \), we get

\[ m(F) \geq am_1(F) = as \]

and

\[ m(F^c) \geq (1-a)m_2(F^c) = (1-a)s. \]

This shows that \( T \) is not ergodic with respect to \( m \).

For the proof of part (ii), see the article by Oxtoby.

\[ \square \]

NOTE. In their work on X-ray crystallography, honored by the Nobel Prize in physics in 1986, H. Hauptman and J. Karle made crucial use of Toeplitz’s characterization (32) of the Fourier coefficients of positive measures.

BIBLIOGRAPHY

**Historical Note.** Dénes König (1884–1944) professor at the Technical University in Budapest, was the founding father of graph theory. He developed many of the basic concepts, and wrote the first book on the subject in 1936. His proof of the Birchhoff-König theorem is graph theoretical. The brilliant Hungarian school in graph theory is his legacy.

König supervised the Eötvös mathematical competitions for high school students. He was extremely kind and encouraging to budding young mathematicians, including the writer of these pages.

When the German army occupied Hungary in 1944, putting Hungarian Nazis in power, König saw what was coming and threw himself out the window of his apartment.

**BIBLIOGRAPHY**


15

BOUNDDED LINEAR MAPS

In chapter 2 we studied some rudimentary properties of linear maps $M$ of one linear space into another. Here we impose topological structures on the linear spaces and on the mappings themselves. Alternative names for maps, and synonymous with it, are operator and transformation.

15.1 BOUNDEDNESS AND CONTINUITY

Definition. $X$ and $U$ are a pair of Banach spaces. A linear map (actually any map)

$$M : X \rightarrow U$$

is called continuous if it maps convergent sequences into convergent ones, that is, if $x_n \rightarrow x$ implies $Mx_n \rightarrow Mx$. (1)

Here convergence is reckoned in the sense of the norm in $X$ and $U$, respectively.

Definition. A linear map $M : X \rightarrow U$ of one Banach space $X$ into another $U$ is called bounded if there is a constant $c$ such that for all $x$ in $X$

$$|Mx| \leq c|x|.$$ (2)

Theorem 1. A linear map $M : X \rightarrow U$ of one Banach space $X$ into another $U$ is continuous if and only if it is bounded.

Proof. It is easy to show that a bounded linear map is continuous, even Lipschitz continuous.

Conversely, if $M$ were not bounded, (2) fails for any $c$, say $n$, for some $x$, say $x_n$: $|Mx_n| > n|x_n|$. Normalize $x_n$ so that $|x_n| = 1/\sqrt{n}$; $x_n$ tends to zero but $Mx_n$ does not. Clearly, (1) is violated, so $M$ is not continuous.

Suppose that the spaces $X$ and $U$ on which and into which $M$ acts are merely normed linear spaces, not complete, and suppose that $M$ is bounded in the sense of (2). Then $M$ can be extended by continuity to a bounded mapping of the completion of $X$ into the completion of $U$. This observation is as trivial as it is important, since most maps of interest are constructed in the fashion described above, being first defined in an incomplete space and then extended by a flick of the wrist to the completed space. The incomplete space usually consists of smooth functions, the complete space of functions less smooth, or not at all smooth.

Definition. Let $M : X \rightarrow U$ be a bounded linear map of one Banach space into another. Its norm, denoted as $|M|$, is defined by

$$|M| = \sup_{x \neq 0} \frac{|Mx|}{|x|}.$$ (3)

Clearly, for any $x$ in $X$, (2) holds with $c = |M|$:

$$|Mx| \leq |M||x|.$$ (2')

Equally clearly, $|M|$ is the smallest value of $c$ for which (2) holds for all $x$ in $X$. A useful reformulation of (3) is

$$|M| = \sup_{|x|=1} |Mx|.$$ (3')

Theorem 2. Norm of bounded maps has the following properties:

(iii) Homogeneity, for any scalar, real, or complex, $|aM| = |a||M|$.

(ii) Positivity, $|M| \geq 0, |M| = 0$ if and only if $M = 0$.

(iii) Subadditivity, $|M + K| \leq |M| + |K|$.

Proof. Properties (i) and (ii) are obvious.

Exercise 1. Prove property (iii).

Definition. The set of all bounded maps of one Banach space $X$ into another $U$ is denoted by

$$\mathcal{L}(X, U).$$

Theorem 3. $\mathcal{L}(X, U)$ is a Banach space under the norm (3).

Proof. Properties (i), (ii), and (iii) in theorem 2 show that $\mathcal{L}(X, U)$ forms a normed linear space under the norm (3). What remains to be shown is the completeness of $\mathcal{L}(X, U)$.
Let \( \{M_n\} \) be a Cauchy sequence in \( \mathcal{L}(X, U) \):

\[
\lim_{n,k \to \infty} |M_n - M_k| = 0.
\]  

(4)

It follows from (4) that for any \( x \in X \)

\[
\lim_{n,k \to \infty} |M_n x - M_k x| = 0.
\]  

(4')

This shows that \( \{M_n x\} \) is a Cauchy sequence in \( U \); since \( U \) is complete, the limit \( \lim M_n x = u \) exists. We define the mapping \( M \) to be \( Mx = u \); clearly, \( M \) is linear.

By definition of a norm,

\[
|M_n - M| = \sup_{|x|=1} |M_n x - Mx| = \sup_{|x|=1} \lim_{k \to \infty} |M_n x - M_k x| = \sup_{|x|=1} |M_n - M_k|.
\]

Using (4) it follows that \( |M_n - M| \to 0 \). □

In the special case in which the target space \( U \) of the linear mappings is one-dimensional, that is, isomorphic with \( \mathbb{R} \) or \( \mathbb{C} \), the bounded linear maps are bounded linear functionals, and \( \mathcal{L}(X, U) \) is just the dual space \( X' \) of \( X \).

We recall from Chapter 2 the notion of the nullspace \( N_M \) of a linear map \( M : X \to U \); it consists of all points \( x \) of \( X \) mapped into \( 0 \) by \( M \).

\[
Mx = 0.
\]  

(5)

**Theorem 4.** Let \( X \) and \( U \) denote normed linear spaces, \( M : X \to U \) a bounded, linear map.

(i) \( N_M \), the nullspace of \( M \), is a closed linear subspace of \( X \).

(ii) \( M \), when regarded as a map

\[
M_0 : \left( \frac{X}{N_M} \right) \to U
\]

is one-to-one, bounded, with \( |M_0| = |M| \). The range of \( M_0 \) is the same as the range of \( M \).

**Proof.** (i) \( N_M \) is the inverse image in \( X \) of \( \{0\} \) in \( U \). Since \( \{0\} \) is a closed set, and \( M \) is continuous, \( N_M \) is closed.

(ii) \( x_1 \) and \( x_2 \) belong to the same equivalence class \( N_M \) if \( x_1 - x_2 \in N_M \). By (5), and linearity, \( Mx_1 = Mx_2 \); therefore the mapping \( M_0 \) is unequivocally defined. We recall from Chapter 5 that the norm in the quotient space \( X/N \) is defined as

\[
|\langle x \rangle| = \inf_{y \in N} |y|.
\]  

(6)

**Boundedness and Continuity**

We have shown there, in Theorem 1, that if \( N \) is closed, as \( N_M \) is, then the quantity \( |\langle x \rangle| \) is a norm. Using the definition (3) of the norm of a map, and some obvious manipulations, we have

\[
|M| = \sup_{x \neq 0} \frac{|Mx|}{|x|} = \sup_{\langle x \rangle \neq 0} \frac{|Mx|}{|y|} = \sup_{\langle x \rangle \neq 0} \frac{|Mx|}{\inf_{y \in N} |y|} = \sup_{\langle x \rangle \neq 0} \frac{|M(x)|}{|\langle x \rangle|} = |M_0|.
\]

We turn to defining the transpose of a bounded linear map \( M : X \to U \), \( X \) and \( U \) normed linear spaces. Let \( \ell \) be a point of \( U' \), the dual of \( U \); that is, \( \ell \) is a bounded linear functional on \( U \). The composite \( \ell(Mx) \) is a linear and bounded functional of \( x \):

\[
\ell(Mx) = \xi(x).
\]  

(7)

The linear functional \( \xi \in X' \) clearly depends linearly on \( \ell \):

\[
\xi = M' \ell.
\]  

(7')

\( M' : U' \to X' \) is called the transpose of \( M \).

The transpose of a bounded linear map is the infinite-dimensional generalization of the transpose of a matrix; it is an enormously useful concept. In studying and using the transpose, it is convenient to denote the action of the linear functionals by parentheses as follows:

\[
\ell(u) = (u, \ell), \quad \xi(x) = (x, \xi),
\]

where \( u \in U \), \( \ell \in U' \), \( x \in X \), \( \xi \in X' \). In this notation the relations (7), (7') defining the transpose can be rewritten as

\[
(Mx, \ell) = (x, M' \ell).
\]  

(8)

We recall from Chapter 8, (see theorem 7') the definition of the annihilator \( R^\perp \) of a subspace \( R \) of a normed linear space \( U \) as the subspace \( R^\perp \) of \( U' \) consisting of all bounded linear functionals \( \ell \) that vanish on \( R \). Similarly, for any subset \( S \) of \( X' \), we define \( S^\perp \) as the subset of those vectors in \( X \) that are annihilated by every vector \( \xi \) in \( S \). Clearly, \( S^\perp \) is a closed linear subspace of \( X \). The basic properties of transposition are summarized in

**Theorem 5.**

(i) The transpose \( M' \) of a bounded linear map \( M \) is bounded, and

\[
|M'| \leq |M|.
\]  

(9)

(ii) The nullspace of \( M' \) is the annihilator of the range of \( M \),

\[
N_{M'} = R_M^\perp.
\]  

(10)
BOUNDARY LINEAR MAPS

(iii) The nullspace of \( M \) is the annihilator of the range of \( M' \).

\[ N_M = R_{M'}^\perp. \] (11)

(iv) \((M + N)' = M' + N'\).

Proof. By (3'), applied to \( M' \),

\[ |M'| = \sup_{|\xi| = 1} |M'\xi|. \] (12)

By definition, the norm of \( \xi \) in \( X' \) is

\[ |\xi| = \sup_{|x| = 1} |(x, \xi)|. \] (13)

Setting \( \xi = M'\ell \) into (13) and combining this with (12), we get, using (8), that

\[ |M'| = \sup_{|\xi| = 1} \sup_{|\ell| = 1} |(x, M'\ell)| = \sup_{|\ell| = 1} |(Mx, \ell)|. \] (12')

According to theorem 6 of chapter 8, for every \( u \) in \( U \),

\[ |u| = \max_{|\ell| = 1} |(u, \ell)|. \] (14)

On the right side of (12') we maximize first with respect to \( \ell \). Using (14), with \( u = Mx \), we get

\[ |M'| = \sup_{|\ell| = 1} |Mx|, \]

which by (3') equals \( |M| \); this proves (9).

To prove (10), we note that for any \( x \) in \( X \) and any \( \ell \) in \( N_M \), the right side of (8) is zero. Therefore so is the left side; this shows that \( N_M \subseteq R_{M'}^\perp \). Conversely, if \( \ell \) annihilates the range of \( M \), the left side of (8) is zero for every \( x \). Therefore so is the right side, which can only be if \( M'\ell = 0 \). This shows that \( N_M \cap R_{M'}^\perp \); these two relations taken together prove (10).

To prove (11), we note that the left side of (8) is zero when \( x \) belongs to the nullspace of \( M \). This shows that every \( x \) in \( N_M \) belongs to the nullspace of every \( \xi = M'\ell, \ell \) in \( U' \). Conversely, suppose that \( x \) belongs to the nullspace of all such \( \xi \); then the right side of (8) is zero for all \( \ell \) in \( U' \). But then so is the left side; this can be only if \( Mx = u = 0 \), that is, if \( x \) belongs to \( N_M \). This proves (11).

Part (iv) is obvious. \( \square \)

Exercise 2. Let \( X \) and \( U \) be Banach spaces, \( U \) reflexive. Let \( M \) be a bounded linear map: \( X \rightarrow U \). Let \( x_n \) be a sequence in \( X \) weakly convergent to \( x \). Then \( Mx_n \) converges weakly to \( Mx \).

**Exercise 3.** Denote by \( I \) the identity map \( X \rightarrow X \). Show that \( I \) is the identity map: \( X' \rightarrow X' \).

In a complex Hilbert space the notion of transpose is replaced by adjoint, defined by the analogue of (8) and denoted by an asterisk:

\[ (Mx, y) = (x, M^*y). \]

For matrices, the adjoint is the conjugate transpose.

**Exercise 4.** Show that theorem 5 is valid for the adjoint operation.

### 15.2 STRONG AND WEAK TOPOLOGIES

The norm of linear maps \( X \rightarrow U \) defines a metric topology in \( L(X, U) \) that is sometimes called the uniform topology, in deference to two other topologies that are also useful and therefore important:

**Definition.** The strong topology in \( L(X, U) \) is the weakest topology in which all functions \( L \rightarrow U \) of the form \( M \rightarrow Mx \) are continuous, \( x \) being any point of \( X \).

**Definition.** The weak topology in \( L(X, U) \) is the weakest topology in which all linear functionals of the form \( M \rightarrow (Mx, \ell) \) are continuous, \( x \) being any point of \( X \) and \( \ell \) any point in \( U' \).

**Exercise 5.** Define the weak* topology in \( L(X, U') \), \( X, U \) Banach spaces. Show that there is a natural one-to-one correspondence between \( L(X, U') \) and \( L(U, X') \), and that this correspondence is continuous in the weak* topology.

Of equal importance are the corresponding notions of sequential convergence:

**Definition.** A sequence \( \{M_n\} \) of bounded linear maps: \( X \rightarrow U, X, U \) Banach spaces, is called strongly convergent if

\[ s - \lim_{n \to \infty} M_n x \]

exists for every \( x \) in \( X \).

\( \{M_n\} \) is called weakly convergent if

\[ w - \lim_{n \to \infty} M_n x \]

exists for all \( x \) in \( X \).
It is easy to show, and is left as an exercise, that a strongly or weakly convergent sequence of maps has a limit $M$, in the sense that (15), (15') are equal to $M$. We will denote these relations as $s - \lim M_n = M$ and $w - \lim M_n = M$.

**Exercise 6.** Prove that if $w - \lim M_n = M$, then $w - \lim M'_n = M'$ provided that $X$ is reflexive. (Hint: Use the definition of weak convergence; see (18) below.)

No such result holds for strong convergence; take $X$ and $U$ both to be the Hilbert space $\ell^2$ (see chapter 6) consisting of vectors

$$x = (a_1, a_2, \ldots), \quad \|x\|^2 = \sum |a_j|^2.$$ 

Define $M_n$ to be

$$M_n x = (a_1, 0, 0, \ldots).$$

It is easy to see that $s - \lim M_n = 0$. Since $\ell^2$ is a Hilbert space, it is self-dual; take $\ell = (b_1, b_2, \ldots)$; the relation $(M_n x, \ell) = a_1 b_1 = x, M'_n \ell, \ell)$ shows that $M'_n \ell = (0, b_1, b_2, \ldots)$. Clearly, $s - \lim M'_n \ell$ does not exist unless $b_1 = 0$.

The significance of these notions is that maps of interest are often—one is tempted to say usually—constructed as limits, uniform, strong, or weak of sequences of approximate maps. The following result, as important as it is trivial, is used all the time:

**Theorem 6.** Let $X, U$ be Banach spaces, $M_n$ a sequence of linear maps: $X \to U$, uniformly bounded in norm:

$$|M_n| \leq c \quad \text{for all } n. \quad (16)$$

Suppose further that

$$s - \lim M_n x$$

exists for a dense set of $x$ in $X$. Then $\{M_n\}$ converges strongly, i.e. the $s$-lim exists for all $x$ in $X$.

**Exercise 7.**

(a) Prove theorem 6.

(b) Formulate and prove an analogous theorem for weak convergence.

### 15.3 Principle of Uniform Boundedness

Uniform boundedness turns out to be not only convenient for proving strong or weak convergence, but it is necessary as well.

**Theorem 7.** Let $X$ and $U$ be Banach spaces, $\{M_n\}$ a collection of bounded maps $X \to U$, such that for each $x$ in $X$ and each $\ell$ in $U'$, $(M_n x, \ell)$ is bounded by a constant that only depends on $x$ and $\ell$:

$$|M_n x, \ell| \leq c(x, \ell) \quad \text{for all } M_n.$$ 

Conclusion: $\{M_n\}$ is uniformly bounded, meaning that (16) holds.

**Proof.** We appeal to the principle of uniform boundedness, theorem 4 of chapter 10: If $\{u_1\}$ is a collection of points in a normed linear space $U$ such that for every linear function $\ell$ in $U'$, $|\ell(u_1)| \leq c(\ell)$ for all $u_1$, then there is a constant $c$ such that $|u_1| \leq c$. We apply this result to $u_1 = M_n x$, and conclude that for all $v$,

$$|M_n x| \leq c(x). \quad (17')$$

Next we appeal to theorem 2 of chapter 10: If $\{f_\ell\}$ is a collection of real-valued, continuous, subadditive, and positive homogeneous functions defined on a Banach space $X$, and if at each point $x$ of $X$, $f_\ell(x) \leq c(x)$ for all $\ell$, then there is a number $c$ such that

$$f_\ell(x) \leq c|x| \quad \text{for all } x, \text{ all } \ell.$$ 

We identify the functions $f_\ell$ with $f_\ell(x) = |M_n x|$. Clearly the $f_\ell$ are homogeneous, subadditive, and continuous. According to (17') above, the $f_\ell(x)$ are bounded at every point. Therefore the $f_\ell$ are uniformly bounded, which in our case means that $|M_n x| \leq c|x|$ for all $x$ in $X$, as asserted in (16).

The relation $w - \lim M_n = M$ means that (15') holds for every $x$ in $X$, which in turn means (see definition 1 of chapter 10) that

$$\lim_{n \to \infty} (M_n x, \ell) = (M x, \ell)$$ 

for all $\ell$ in $U'$ and all $x$ in $X$. Since a convergent sequence is bounded, it follows that condition (17) of theorem 7 is satisfied; therefore by theorem 7, the sequence $M_n$ is uniformly bounded.

**Corollary 7.** A weakly convergent sequence of maps of one Banach space into another is uniformly bounded.

### 15.4 Composition of Bounded Maps

We turn now to the discussion of composition, called the product, of a map $M : X \to U$ with another map $N : U \to W$. This operation was studied in chapter 2 from the point of view of linear algebra. Here we study some further properties of it in case $X, U$ and $W$ are Banach spaces, and $M$ and $N$ are bounded linear maps.
Theorem 8. Let $X$, $U$, $W$ denote Banach spaces, $M$ and $N$ bounded linear maps,

\[ M : X \to U, \quad N : U \to W. \]

Then the composite $NM$ is a bounded linear map: $X \to W$ with the following properties:

(i) Submultiplicativity, $|NM| \leq |N||M|$.

(ii) $(NM)^* = M^*N^*$.

Proof. Applying inequality (4) twice, we get

\[ |NMx| \leq |N||M||x|. \]

Applying definition (3), we get

\[ |NM| = \sup \frac{|NMx|}{|x|} \leq |N||M|. \quad (19) \]

We turn to (ii): applying (8) twice, we get:

\[ (NMx, m) = (Mx, N^*m) = (x, M^*N^*m). \quad (20) \]

Exercise 8. Prove that multiplication of maps is a continuous operation on the strong topology on the unit balls of $L(X, U)$ and $L(U, W)$.

Definition. Two maps $A$ and $M$ of a linear space $X$ into itself are said to commute if $AM = MA$.

Exercise 9. Let $X$ denote a Banach space, $A$ a bounded map: $X \to X$ that commutes with each of a collection $\{M_n\}$ of bounded maps $X \to X$. Show that then $A$ commutes with every map $M$ that lies in the closed linear span of the set of maps $\{M_n\}$ in the weak topology.

Exercise 10. Show that in a complex Hilbert space $(NM)^* = M^*N^*$.

15.5 THE OPEN MAPPING PRINCIPLE

The next group of results, the open mapping principle, and the closed graph theorem, goes considerably deeper than the foregoing material. These ideas are due to Stefan Banach, their validity is far from being intuitively clear at first glance, or even a second one.

Theorem 9. $X$ and $U$ are Banach spaces, and $M : X \to U$ a bounded linear mapping of $X$ onto all of $U$. Then there is a $d > 0$ such that the image of the open unit ball in $X$ under $M$ contains the ball of radius $d$ in $U$:

\[ MB(0) \supseteq B_d(U). \quad (21) \]

Proof. Denote by $B_0$ the open ball of radius $n$ around the origin in either the space $X$ or $U$. Since $M$ is assumed to map $X$ onto $U$, and since the union of all the $B_0$ is all of $X$, it follows that $\cup MB_{\infty} = U$. Since the Banach space $U$ is complete, it follows from the Baire category principle that at least one of the sets $MB_0$ is dense in some open set. Some translate of this set is dense in some ball around the origin; since the range of $M$ is all of $U$, by linearity of $M$ we face that translate to be of the form $M(B_0 - x_0)$. The set $B_0 - x_0$ is contained in the ball of radius $n = |x_0|$ around the origin. So by homogeneity of $M$, we conclude that $MB(0)$ is dense in $B_r(0)$ for some $r > 0$. Consequently for any $c > 0$,

\[ MB(0) \text{ is dense in } B_{cr}(0). \quad (22) \]

We want to show now that any point $u$ in $B_r(0)$ is the image of some point $x$ in $B_2(0)$:

\[ Mx = u. \quad (23) \]

This point $x$ in $B_2(0)$ is constructed as an infinite series

\[ x = \sum_{j=1}^{\infty} x_j. \quad (23') \]

The terms $x_j$ are constructed recursively: $x_1$ is taken as a point satisfying

\[ |u - Mx_1| < \frac{r}{2}, \quad |x_1| < 1; \quad (24a) \]

by (22), with $c = 1$, there is such an $x_1$. We choose $x_2$ as a point satisfying

\[ |u - Mx_1 - Mx_2| < \frac{r}{4}, \quad |x_2| < \frac{1}{2}; \quad (24b) \]

it follows from (22), with $c = \frac{1}{2}$, and (24a) that such an $x_2$ exists. Generally, we choose $x_m$ to satisfy

\[ |u - \sum_{j=1}^{m} Mx_j| < \frac{r}{2^m}, \quad |x_m| < \frac{1}{2^{m-1}}; \quad (24c) \]

It follows from (22), with $c = 1/2^{m-1}$ and (24c) that there is such an $x_m$.

We noted in chapter 5 on the geometry of normed spaces that if the sum of the norms $\sum |x_j|$ of a series in a complete normed linear space $X$ converges, the series $\sum x_j$ converges strongly. Since by (24c), $|x_j| < 1/2^{j-1}$, it follows that $\sum_{j=1}^{\infty} x_j$
converges to a point \( x \) in \( X \), and

\[
|x| \leq \sum_{j=1}^{\infty} |x_j| < \sum_{j=1}^{\infty} \frac{1}{2^{j-1}} = 2.
\] (25)

Since \( M \) is a bounded map, letting \( m \to \infty \) in (24c), we conclude that \( Mx = \sum_{j=1}^{\infty} Mx_j = u. \) \( \square \)

Theorem 9 has a number of interesting and important consequences; the first one is the open mapping principle:

**Theorem 10.** \( X \) and \( U \) are Banach spaces, \( M : X \to U \) a bounded linear map onto all of \( U \). Then \( M \) maps open sets onto open sets.

This is an immediate corollary of theorem 9. \( \square \)

**Theorem 11.** \( X \) and \( U \) are Banach spaces, \( M : X \to U \) a bounded linear map that carries \( X \) one-to-one onto \( U \). Then the algebraic inverse of \( M \) is a bounded linear map of \( U \to X \).

**Proof.** It follows from (21) of theorem 9 that for every \( u \in U \) of norm \( \frac{d}{2} \), there is an \( x \) in the unit ball of \( X \) such that \( Mx = u \); note that \( |x| \leq 1 = 2|u|/d \). Since \( M \) is homogeneous, it follows that for every \( u \in U \) there is an \( x \) in \( X \), such that

\[
Mx = u, \quad |x| \leq 2|u|/d.
\] (26)

Since \( M \) is assumed one-to-one, \( x = M^{-1}u \). Clearly, from (26), \( |M^{-1}| \leq 2/d. \) \( \square \)

**Definition.** A map \( M : X \to U \) from one Banach space into another is called closed if whenever \( \{x_n\} \) is a sequence in \( X \) such that

\[
x_n \to x \quad \text{and} \quad Mx_n \to u
\] (27)

then

\[
Mx = u.
\] (27')

If \( M \) is continuous, it is obviously closed. It is surprising but true that conversely, a closed linear map of a Banach space into another is continuous:

**Theorem 12.** \( X \) and \( U \) are Banach space, \( M : X \to U \) is a closed linear map.

**Assertion.** \( M \) is continuous.

**Proof.** Define the linear space \( G \) to consist of all pairs \( g \) of form

\[
g = \{x, Mx\}, \quad x \in X.
\] (28)

The open mapping principle

We define the following norm for \( g \) in \( G \):

\[
|g| = |x| + |Mx|
\] (28')

Clearly, this is a norm. It follows from (27), (27') and the completeness of \( X \) and \( U \) that \( G \) is complete under this norm. Define the mapping \( P : G \to X \) to be the projection onto the first component, that is,

\[
g = \{x, Mx\}, \quad Pg = x.
\] (29)

By definition (28') of \( |g|, \ |Pg| \leq |g| \), meaning that \( P \) is a bounded operator, \( |P| \leq 1 \). Clearly, \( P \) is linear and maps \( G \) one-to-one onto \( X \). Therefore, by theorem 11, the inverse of \( P \) is bounded; that is, there is a constant \( c \) such that \( c |Pg| \geq |g| \). In view of the definition (29) of \( P \) and (28') of \( |g| \), it follows that \( (c - 1) |x| \geq |Mx| \), meaning that \( M \) is bounded. \( \square \)

The space \( G \) defined by (28) is called the graph of the mapping \( M \). Requiring \( M \) to be closed is the same as requiring its graph to be closed. Theorem 12 is known as the closed graph theorem. The closed graph theorem has many surprising applications.

**Theorem 13.** \( X \) is a linear space equipped with two norms \( |x|_1 \) and \( |x|_2 \) that are compatible in the following sense: If a sequence \( \{x_n\} \) converges in both norms, the two limits are equal.

Suppose that \( X \) is complete with respect to both norms; then the two norms are equivalent. That is to say, there is a constant \( c \) such that for all \( x \in X \),

\[
|x|_1 \leq c |x|_2, \quad |x|_2 \leq c |x|_1.
\]

**Proof.** Denote by \( X_1 \), resp. \( X_2 \) the space \( X \) under the \( 1 \)-, resp. \( 2 \)-norm. By hypothesis, both \( X_1 \) and \( X_2 \) are complete. Compatibility clearly means that the identity map between \( X_1 \) and \( X_2 \) is closed. Therefore, by the closed graph theorem, it is bounded in both directions. \( \square \)

**Theorem 14.** \( X \) and \( U \) are Banach spaces, \( M : X \to U \) a bounded linear map. Assume that the range \( R_M \) is a finite-dimensional subspace of \( U \); then \( R_M \) is closed.

**Exercise 11.** Prove theorem 14. (Hint: Extend \( M \) to \( X \oplus Z \) so that its range is all of \( U \).)

**Exercise 12.** Show that for every infinite-dimensional Banach space there are linear subspaces of finite codimension that are not closed. (Hint: Use Zorn's lemma.)

**Theorem 15.** \( X \) is a Banach space, \( Y \) and \( Z \) closed subspaces of \( X \) that complement each other: \( X = Y \oplus Z \), in the sense that every \( x \) in \( X \) can be decomposed uniquely
as \( x = y + z, \) \( y \) in \( Y, z \) in \( Z. \) Denote the two components \( y \) and \( z \) of \( x \) by

\[
 y = P_Y x, \quad z = P_Z x.
\]

(i) \( P_Y \) and \( P_Z \) are linear maps of \( X \) on \( Y \) and \( Z, \) respectively.

(ii) \( P_Y^2 = P_Y, \) \( P_Z^2 = P_Z, \) \( P_Y P_Z = 0. \)

(iii) \( P_Y \) and \( P_Z \) are continuous.

Proof. Parts (i) and (ii) are obvious. To prove part (iii) we observe that since \( Y \) and \( Z \) are closed, and the decomposition is unique, it follows that the graphs of \( P_Y \) and \( P_Z \) are closed. The closed graph theorem does the rest.

A map satisfying \( P^2 = P \) is called a projection.

We conclude this chapter by observing that complete metric spaces have proper subsets, called sets of second category, that are not unions of a denumerable number of nowhere dense sets. This allows a sharpening of the open mapping principle:

**Theorem 16.** \( X \) and \( U \) are Banach spaces, \( M : X \to U \) a bounded linear map whose range \( \operatorname{R} M \) is a subset of \( U \) of second category. Then the range of \( M \) is all of \( U. \)

**Exercise 13.** Prove theorem 16.

**HISTORICAL NOTE.** Stefan Banach (1892–1945), a Polish mathematician, was one of the founding fathers of functional analysis. Banach spaces are named in recognition of his numerous and deep contributions, and for having written the first monograph on the subject (1932). He was the inspiration of the brilliant Polish school of functional analysis.

During the Second World War, Banach was one of a group of people whose bodies were used by the Nazi occupiers of Poland to breed lice, in an attempt to extract an anti-typhoid serum. He died shortly after the conclusion of the war.

The Nazi attitude toward Poles is epitomized by the following story. After the conquest of France in 1940, when Hitler ruled most of Europe, a leading German mathematician, a member of the Nazi party, called on Elie Cartan, the dean of French mathematicians, to discuss the organization of mathematical life in the new European order. Cartan wanted to know how the Polish mathematicians would fit in. “Oh,” the German replied, “the Fuhrer has declared the Poles to be subhuman.”

**BIBLIOGRAPHY**


16

EXAMPLES OF BOUNDED LINEAR MAPS

An important class of linear maps is furnished by integral operators. The first part of this chapter is devoted to investigating their boundedness in various norms. Let \( T \) and \( S \) be Hausdorff spaces, equipped with measures \( n \) and \( m. \) \( K \) denotes an integral operator mapping complex-valued functions \( f \) on \( T \) into complex-valued functions \( g \) on \( S: \)

\[
g(s) = (Kf)(s) = \int_T K(s, t) f(t) \, dn(t). \tag{1}
\]

The complex-valued function \( K(s, t) \) is called the kernel of \( K; \) \( f, K \) are assumed to be measurable and restricted so that (1) defines a measurable function \( g. \) Each subsequent theorem reveals a natural class for \( f, K, \) and \( g. \) We recall from chapter 4 the \( L^p \)-norms:

\[
|f|_{L^p} = \left( \int_T |f(t)|^p \, dn(t) \right)^{1/p}, \quad 1 \leq p \leq \infty.
\]

The space \( L^p(T, n) \) is the completion of the space \( C_0(T) \) in the \( L^p \)-norm. The space \( L^p(S, m) \) is defined analogously. The space \( L^{\infty} \) is the space of essentially bounded, measurable functions.

16.1 BOUNDEDNESS OF INTEGRAL OPERATORS

We start with conditions that guarantee that (1) is a bounded map from \( L^1(T, n) \) or \( L^{\infty}(T, n) \) to \( L^1(S, m) \) or \( L^{\infty}(S, m). \)

**Theorem 1.**

(i) The map \( K \) defined by (1) is bounded as a mapping \( L^1 \to L^{\infty}, \) and
EXAMPLES OF BOUNDED LINEAR MAPS

\[ |K| \leq \sup_{s,t} |K(s, t)|, \quad (2_t) \]

provided that the quantity on the right is \( < \infty \).

(ii) \( K \) is bounded as a mapping \( L^\infty \to L^1 \), and

\[ |K| \leq \int \int |K(s, t)| \, dm(s) \, dn(t), \quad (2_{ii}) \]

provided that the quantity on the right is \( < \infty \).

(iii) \( K \) is bounded as a mapping \( L^\infty \to L^\infty \),

\[ |K| \leq \sup_{z} \int |K(s, t)| \, dn(t) \quad (2_{iii}) \]

if the quantity on the right is \( < \infty \).

(iv) \( K \) is bounded as a mapping \( L^1 \to L^1 \), and

\[ |K| \leq \sup_{t} \int |K(s, t)| \, dm(s), \quad (2_{iv}) \]

if the quantity on the right is \( < \infty \).

Proof. By (1), for any \( s \) in \( S \),

\[ |g(s)| \leq \int |K(s, t)| |f(t)| \, dn(t). \quad (3) \]

The right side is \( \leq \sup_{s} |K(s, t)| |f|_{L^1} \), so

\[ |g|_{L^1} = \sup_{s} |g(s)| \leq \sup_{s,t} |K(s, t)| |f|_{L^1}. \]

This proves (2_t).

Integrate (3) with respect to \( dm \) over \( S \):

\[ |g|_{L^1} = \int |g(s)| \, dm(s) = \int \int |K(s, t)| |f(t)| \, dn(t) \, dm(s) \]

\[ = \int \left[ \int |K(s, t)| \, dm(s) \right] |f(t)| \, dn(t). \quad (4) \]

The right side is

\[ \leq \int \left[ \int |K(s, t)| \, dm(s) \right] \, dn(t) \, |f|_{L^\infty} \]

this proves (2_{ii}).

BOUNDEDNESS OF INTEGRAL OPERATORS

The right side of (4) is also less than

\[ \sup_{t} \int |K(s, t)| \, dm(s) \, |f|_{L^1}; \]

this proves (2_{iv}).

The right side of (3) is less than

\[ \int |K(s, t)| \, dn(t) \, |f|_{L^\infty}; \]

this combined with (3) proves (2_{iii}). \( \square \)

Note that when \( K(s, t) \) and \( f(t) \) are both positive, the sign of equality holds in both (3) and (4). From this it is not hard to deduce

Corollary 2'. When the kernel \( K(s, t) \) in (1) is nonnegative, the sign of equality holds in (2_{ii}) and (2_{iv}).

The transpose of \( K \) is easily written down; denote by \((,)_S\) and \((,)_T\) the standard \( L^2 \) scalar product on \( S \) and \( T \) with respect to \( dm \) and \( dn \), respectively:

\[ (g, h)_S = \int g(s) h(s) \, dm(s), \quad (5) \]

\[ (k, f)_T = \int k(t) f(t) \, dn(t). \quad (5') \]

Multiplying (1) by \( h(s) \) and integrating gives

\[ (Kf, h)_S = \int \int K(s, t) f(t) h(s) \, dx(t) \, dm(s) = (f, K'h)_T, \quad (6) \]

where

\[ (K'h)(t) = \int S \, K(s, t) h(s) \, dm(s). \quad (6') \]

In words, the kernel of the transpose \( K' \) is the same as the kernel of \( K \), with the roles of the variables \( s \) and \( t \) interchanged.

We recall now theorem 5 of chapter 15, according to which the norm of \( K' \) equals the norm of \( K \). We verify this in the case when the kernel \( K \) is nonnegative, and \( K \) is regarded as mapping \( L^1(T) \) into \( L^1(S) \). According to corollary 1', \( |K| \) is given by formula (2_{iv}). \( K' \), on the other hand, maps \( L^\infty(S) \) into \( L^\infty(T) \), and its norm is given by formula (2_{iii}), with the roles of \( s \) and \( t \) reversed. Clearly, \( |K'| = |K| \), as it should be.
We turn now to the $L^2$-norms, which we denote as $\|\|$; the corresponding norm of $K$ is denoted as $\|K\|$.  

**Theorem 2.** The map $K$ defined by (1) is bounded as a map: $L^2 \rightarrow L^2$, and  

$$\|K\|^2 \leq \iint_{ST} |K^2(s, t)| \, dm \, dn,$$  

(7)  

provided that the quantity on the right is $< \infty$.  

**Proof.** Applying the Schwarz inequality (see chapter 6) to the integral on the right in (1), we get  

$$|g(s)|^2 \leq \int_T |K^2(s, t)| \, dn \int_T |f(t)|^2 \, dt.$$  

Integrating both sides $dm$ gives  

$$\|g\|^2 \leq \iint_{ST} |K(s, t)|^2 \, dn \, dm \|f\|^2,$$  

as asserted in (7).  

Inequality (7) is due to Hilbert and E. Schmidt. Another criterion has been given by Holmgren:  

**Theorem 3.** $K$ as defined by (1) is bounded: $L^2 \rightarrow L^2$, and  

$$\|K\| \leq \left( \sup_s \int |K(s, t)| \, dn \right)^{1/2} \left( \sup_t \int |K(s, t)| \, dm \right)^{1/2},$$  

(8)  

provided that the quantity on the right is $< \infty$.  

**Proof.** According to theorem 1 in chapter 6,  

$$\|g\| = \max_{\|h\|=1} (g, h)_S.$$

(9)  

We are going to use (9) to estimate $g = Kf$. By (6),  

$$(g, h)_S = \int \int K(s, \cdot) f(t) h(s) \, dn \, dm.$$  

(10)  

For any three positive numbers $f, h$ and $c, f h \leq c f^2/2 + h^2/2c$, so the right side of (10) is  

$$\leq \int \int |K(s, t)| \left\{ \frac{c}{2} f(t)^2 + \frac{1}{2c} |h(s)|^2 \right\} \, dm \, dn.$$  

In the first term we integrate first with respect to $s$, in the second with respect to $t$. We get the estimate  

$$\frac{c}{2} \sup_s \int |K(s, t)| \, dm \|f\|^2 + \frac{1}{2c} \sup_t \int |K(s, t)| \, dn \|h\|^2.$$  

(10')  

We take now $\|f\| = 1 = \|h\|$ and choose $c$ so that (10') is as small as possible. This minimum is  

$$\left( \sup_s \int \right)^{1/2} \left( \sup_t \int \right)^{1/2}.$$  

(10")  

Combining (9) and (10) with (10"), we conclude that for $\|f\| = 1$, $\|Kf\|$ is $\leq$ the quantity in (10'). In view of the definition $\|K\| = \sup \|Kf\|$, $\|f\| = 1$, this proves (8).  

□  

16.2 THE CONVEXITY THEOREM OF MARCEL RIESZ  

The two factors appearing on the right in (8) are the square roots of the quantities appearing on the right in (QIII) and (QIV). We noted in corollary 1 that for a positive kernel these quantities are not merely upper bounds for the norm of $K : L^\infty \rightarrow L^\infty$ and $L^1 \rightarrow L^1$ but equal to these norms.  

**Definition.** Denote by $M(p, q)$ the norm of  

$$K : L^p(T, n) \rightarrow L^q(S, m).$$  

(11)  

For integral operators whose kernel is $\geq 0$, we can restate inequality (8) as follows:  

$$M(2, 2) \leq M^{1/2}(1, 1) M^{1/2}(\infty, \infty).$$  

This turns out to be a special case of a far more general theorem due to M. Riesz.  

**Theorem 4.** Let $M$ be a linear map of complex-valued functions defined on $T$ into complex-valued functions defined on $S$. Suppose that $M$ carries functions measurable with respect to $n$ into functions measurable with respect to $m$. Suppose further that $M$ is bounded with respect to two pairs of norms:  

$$L^p(T, n) \rightarrow L^p(S, m) \quad \text{and} \quad L^q(T, n) \rightarrow L^q(S, m).$$  

Conclusion: then $M$ is a bounded map of $L^p(a)(T, a) \rightarrow L^q(a)(S, m)$ where  

$$\left( \frac{1}{p(a)}, \frac{1}{q(a)} \right) = (1 - a) \left( \frac{1}{p_0}, \frac{1}{q_0} \right) + a \left( \frac{1}{p_1}, \frac{1}{q_1} \right), \quad 0 \leq a \leq 1.$$  

(12)
Furthermore, $M(p, q)$ is a log-convex function of its arguments:

$$M(p(a), q(a)) \leq M^{1-a}(p_0, q_0) M^a(p_1, q_1);$$

(12')

where $M(p, q)$ is the norm of the operator $K$ defined in (11).

**Proof.** We sketch Thorin's beautiful proof of this theorem. The starting point is the following result due to Hadamard:

**Three Lines Theorem.** Let $\phi(\zeta)$ be a bounded analytic function in the strip $0 \leq \text{Re} \zeta \leq 1$. Denote

$$N(a) = \sup_{\eta} |\phi(\zeta + i\eta)|$$

(13)

Then

$$N(a) \leq N^{1-a}(0) N^a(1).$$

(13')

**Proof.** Set $c = \log N(0)/N(1)$; by (13), the function $\phi(\zeta)e^{c\zeta}$ is in absolute value $\leq N(0)$ for $\text{Re} \zeta = 0$ and $\text{Re} \zeta = 1$. So, by the maximum principle applied in the strip $0 \leq \text{Re} \zeta \leq 1$,

$$|\phi(a + i\eta)| e^{ca} \leq N(0);$$

from this and the definition of $c$, (13') follows.

We turn now to the mapping $M$; by definition of the norm,

$$M(p, q) = \sup_{|f|_p = 1} |Mf|_q.$$  

Furthermore, according to theorem 5 of chapter 5—Hölder's inequality and equality—for any $g \in L^q$, $|g|_q = \sup_{|h|_{L^p} = 1} |(g, h)|_q$, where $q'$ is dual to $q$.

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

Combining the last two, with $g = Mf$, we get

$$M(p, q) = \sup_{|f|_p = 1, |h|_{q'} = 1} |(Mf, h)|. $$

(14)

We take $p = p(a), q = q(a)$ as defined in (12). The complex-valued functions $f$ and $h$ can be factored as $f = |f|e^{i\mu}, \quad h = |h|e^{i\nu}$. For any $\zeta$ in the strip $0 \leq \text{Re} \zeta \leq 1$ we define

$$f(\zeta) = |f|^p(a)/p(\zeta) e^{i\mu}, \quad h(\zeta) = |h|^q(a)/q(\zeta) e^{i\nu},$$

(15)

where $p(\zeta), p(\zeta)$, and so on, are defined by formula (12). Note that $f(a) = f, \quad h(a) = h$. Since $1/p(\zeta)$ and $1/q(\zeta)$ are linear functions of $\zeta$, so is $1/q(\zeta)$. Therefore $f(\zeta)$ and $h(\zeta)$ are analytic functions of $\zeta$, and so is

$$\phi(\zeta) = (Mf(\zeta), h(\zeta)) = \int_M f (\zeta) h(\zeta) dm(s).$$

(15')

**Lemma 5.** Let $f$ and $h$ be functions of unit norm, $|f|_p(a) = 1, |h|_q(a) = 1$, and $\phi(\zeta)$ defined as above. Define $N(\alpha)$ as the supremum of $|\phi(\zeta)|$ on the line Re $\zeta = \alpha$; we claim that

$$N(0) \leq M(p_0, q_0), \quad N(1) \leq M(p_1, q_1).$$

(16)

**Proof.** Let's take Re $\zeta = 0$. Then $\zeta = i\eta$, and so by formula (12),

$$p(\zeta) = p(\zeta) + \text{Im} \zeta, \quad q(\zeta) = q(\zeta) + \text{Im} \zeta,$$

(17)

From (15),

$$|f(i\eta)|^p_L = |f(i\eta)|_p^{p(a)} = |h(i\eta)|^q_L = |h(i\eta)|^{q(a)}.$$  

(18)

Since $f$ and $h$ were chosen so that $|f|_{L^p(a)} = 1, |h|_{L^q(a)} = 1$, it follows from (18) that $|f(i\eta)|_{L^p} = 1, |h(i\eta)|_{L^q} = 1$. Therefore,

$$|Mf(i\eta)|_{L^p} \leq M(p_0, q_0).$$

(19)

Estimate $\phi$ defined in (15') by Hölder's inequality using (19), and $|h(i\eta)|_{L^q} = 1,$ we get

$$|\phi(i\eta)| = |(Mf(i\eta), h(i\eta))| \leq |(Mf(i\eta))|_{L^p} |h(i\eta)|_{L^q} \leq M(p_0, q_0).$$

This proves the first part of (16); the second part follows in exactly the same fashion.

We apply now the three lines theorem, (13'), to $p$ defined by (15'); using (16), we get

$$|\phi(\alpha)| \leq N(\alpha) \leq M^{1-a}(p_0, q_0) M^a(p_1, q_1)$$

(20)

Since $f(a) = f, h(a) = h$, by (15')

$$\phi(a) = (Mf, h)$$

According to (14) the supremum of the right side over all $f, h$ of unit norm is the norm of $M$. 

---

**THE CONVEXITY THEOREM OF MARCEL RIESZ**

---
16.3 EXAMPLES OF BOUNDED INTEGRAL OPERATORS

Theorems 2 and 3 both furnish criteria for an integral operator to be bounded in $L^2 \to L^2$. These criteria are very far from being necessary for boundedness, and are insufficient for proving the $L^2 \to L^2$ boundedness of the most important and most beloved mappings. We illustrate this on a number of examples.

16.3.1 The Fourier Transform

The Fourier transform is defined by

$$\hat{f}(\omega) = \int_{\mathbb{R}} e^{-i\omega t} f(t) \frac{dt}{\sqrt{2\pi}},$$

whose kernel is

$$K(t,\omega) = \frac{1}{\sqrt{2\pi}} e^{-it\omega}. \quad (21')$$

$T = S = \mathbb{R}$, $m$ and $n$ Lebesgue measure.

Clearly, for this $K$ the right side of both (7) and (8) is $\infty$, so neither theorem 2 nor theorem 3 can be used to show the $L^2 \to L^2$ boundedness of the Fourier transform. Yet it is well known that it is bounded; see theorem 21 of Appendix B. On the other hand, we can use part (i) of theorem 1 to conclude that $F : L^1 \to L^\infty$ is bounded by $1/\sqrt{2\pi}$. We can now appeal to the M. Riesz convexity theorem, theorem 4, with $(p_0, q_0) = (2, 2), (p_1, q_1) = (1, \infty)$ to conclude, after a brief calculation.

Theorem 6. For $1 \leq p \leq 2$, $F$ is a bounded map of $L^p \to L^{p/(p-1)}$ and

$$\left| F \right| \leq \left( \frac{1}{\sqrt{2\pi}} \right)^{(2-p)/p}. \quad (22)$$

This inequality is called the Hausdorff-Young inequality after its discoverers.

16.3.2 The Hilbert Transform

Let $h(t)$ be a real-valued function on $\mathbb{R}$, fairly smooth—$C^1$ will do—and tending to zero as $|t| \to \infty$ at a reasonable rate, say $O(t^{-2})$.

The Cauchy integral

$$\frac{1}{\pi i} \int_{\mathbb{R}} \frac{h(t)}{t-\xi} dt = f(\xi)$$

defines a function $f(\xi)$, or rather two functions, one analytic in the upper half-plane, the other in the lower half-plane. We will restrict $\xi$ to the upper half-plane.

Writing $\xi = \xi + i\eta$, we can express the real and imaginary parts of $f$ as follows:

$$f(\xi) = \frac{1}{\pi i} \int_{\mathbb{R}} \frac{h(t)(t-\xi)}{|t-\xi|^2 + \eta^2} dt = \frac{1}{\pi} \int_{\mathbb{R}} \frac{n}{(\xi - t)^2 + \eta^2} h(t) dt$$

$$+ \frac{i}{\pi} \int_{\mathbb{R}} \frac{(\xi - t)}{(\xi - t)^2 + \eta^2} h(t) dt. \quad (23')$$

Using the properties imposed on $h$, it is not hard to show that

(i) As $|\xi| \to \infty$,

$$|f(\xi)| = o(|\xi|^{-1}). \quad (24)$$

(ii) $f(\xi)$ is continuous up to the real axis, and its real part there equals $h$:

$$f(\xi) = h(\xi) + i k(\xi), \quad (25)$$

where $k$ is expressed in terms of $h$ as the principal value integral

$$k(\xi) = \frac{1}{\pi} PV \int_{\mathbb{R}} \frac{h(t)}{\xi - t} dt = (\mathcal{H}h)(\xi). \quad (25')$$

The map $\mathcal{H}$ defined in (25') is called the Hilbert transform; it relates the real to the imaginary part of the boundary value of analytic functions in the upper half-plane satisfying (24).

Theorem 7. The Hilbert transform is an isometry of $L^2(\mathbb{R}) \to L^2(\mathbb{R})$.

Proof. Since $f^2$ is analytic in $\operatorname{Im} \xi > 0$, by Cauchy's theorem,

$$\oint f^2 d\xi = 0 \quad (26)$$

over every closed contour there. We take now the contour to consist of a line segment $\xi + i\epsilon, -R \leq \xi \leq R$, and a semicircle $\xi = R \cos \theta, \eta = R \sin \theta + \epsilon$. Now let $\epsilon \to 0$ and $R \to \infty$. It follows from (24) that the integral over the semicircle in (26) tends to zero as $R \to \infty$, while it follows from (24) and (25) that the integral over the segment tends to

$$\int_{\mathbb{R}} (h + i k)^2 d\xi = 0. \quad (26')$$

Taking the real part of (26') gives
EXAMPLES OF BOUNDED LINEAR MAPS

\[ \int h^2 d\xi = \int k^2 d\xi, \]

as asserted in theorem 7. \(\Box\)

**Exercise 1.** Show that

\[ H^2 = -I, \quad \text{where} \quad I = \text{identity}. \]  

(Hint: Consider the relation of the real part of \( -if(\xi) \) to its imaginary part.)

Note that the kernel of \( H \),

\[ K(s, t) = \frac{1}{s-t}, \]

fails miserably the tests for boundedness given in theorems 2 and 3.

The argument used above to show that \( H \) is an isometry of \( L^2 \to L^2 \) can be used to prove:

**Theorem 8.** The Hilbert transform \( H \) is a bounded map of \( L^p \to L^p \) for all \( p \), \( 1 < p < \infty \).

*Proof.* Take \( p = 4 \), and consider the analytic function \( f^4 \). By Cauchy's theorem,

\[ \oint f^4 d\xi = 0. \]

We choose the same contour as in (26) and let \( \epsilon \to 0 \), \( R \to \infty \), to obtain

\[ \int_R (h(\xi) + ik(\xi))^4 d\xi = 0. \]

The real part of this relation is

\[ \int_R (h^4 - 6h^2k^2 + k^4) d\xi = 0. \]  

(29')

According to a well-known inequality, for \( a, b, c \) positive \( ab \leq ca^2/2 + b^2/2c \); applying this to \( a = h^2, b = k^2, c = 6 \), gives

\[ 6h^2k^2 \leq 18h^4 + \frac{1}{2}k^4. \]

Setting this into (29'), we get

\[ \frac{1}{2} \int k^4 d\xi \leq 17 \int h^4 d\xi. \]

This shows that \( H : L^4 \to L^4 \) is bounded, and that \( |H| \leq 34 \).

EXAMPLES OF BOUNDED INTEGRAL OPERATORS

The same argument works for \( p \) any even integer. Then, using M. Riesz's convexity theorem, theorem 4, we deduce the boundedness of \( H \) as a map of \( L^p \to L^p \) for any \( p, 2 \leq p < \infty \).

To complete the proof, we turn to theorem 5 of chapter 15, according to which the transpose \( H' \) of \( H \) has the same norm as \( H \). According to formulas (1), (1'), the kernel of \( H' \) is obtained from the kernel of \( H \) by interchanging the roles of the variables. According to (28), interchanging the variables in the kernel of \( H \) merely changes the sign of the kernel. So

\[ H' = -H. \]  

(30)

If \( H \) maps \( L^p \to L^p \), \( H' \) maps \( (L^p)' \to (L^p)' \). According to theorem 11 of chapter 8, the dual of \( L^p \) is \( L^{p'} \), where

\[ \frac{1}{p'} + \frac{1}{p} = 1. \]  

(31)

Note that if \( p > 2, p' < 2 \). Combining (30) and (31) with theorem 5 of chapter 8, we conclude that the norm of \( H : L^p \to L^{p'} \) equals the norm of \( H : L^p \to L^p \). Since the latter were shown to be finite for \( 2 < p < \infty \), it follows that they are bounded for \( 1 < p' < 2 \) as well. \(\Box\)

Theorem 8, and the astonishing proof above, are due to M. Riesz.

**Exercise 2.** Show that \( H \) is not bounded as a map: \( L^\infty \to L^\infty \). Deduce from this that \( H \) is not bounded as a map: \( L^1 \to L^1 \).

16.3.3 The Laplace Transform

Let \( f(t) \) be a complex-valued function on \( \mathbb{R}_+; t \geq 0 \). Its Laplace transform \( LF \) is the function on \( \mathbb{R}_+; s > 0 \) defined by

\[ g(s) = (LF)(s) = \int_0^\infty f(t) e^{-st} dt. \]  

(32)

**Theorem 9.** The Laplace transform \( L \) is a bounded map of \( L^2(\mathbb{R}_+) \to L^2(\mathbb{R}_+) \), and

\[ \|L\| = \sqrt{\pi}. \]  

(33)

*Proof.* We estimate \( g(s) \) by the Schwarz inequality:

\[ |g(s)|^2 = \left( \int_0^\infty f(t)e^{-st} dt \right)^2 \leq \int_0^\infty (f(t)e^{-st_1/2})(e^{-s't_1/2}) dt \int_0^\infty e^{-s't_1/2} dt. \]  

(34)
EXAMPLES OF BOUNDED LINEAR MAPS

By a change of variable we can write the second integral as

\[ \int_0^\infty e^{-s} t^{-1/2} dt = \int_0^\infty e^{-u} u^{-1/2} du \, s^{-1/2} = Cs^{-1/2}, \]  

where

\[ C = \int_0^\infty e^{-u} u^{-1/2} du = \int_0^\infty e^{-x^2} x^{-1} dx = 2 \int_0^\infty e^{-x^2} dx = \sqrt{\pi}. \]  

Setting (35) into (34) gives

\[ |g(s)|^2 \leq Cs^{-1/2} \int_0^\infty |f(t)|^2 e^{-st} t^{1/2} dt. \]  

(37)

Integrating (37) gives

\[ \|g\|^2 = \int_0^\infty |g(s)|^2 ds \leq C \int_0^\infty \int_0^\infty |f(t)|^2 e^{-st} t^{1/2} s^{-1/2} dt ds. \]  

(38)

Interchange the order of integration, and change variables in the s-integral:

\[ \int_0^\infty e^{-st} t^{1/2} s^{-1/2} ds = \int_0^\infty e^{-ut} u^{-1/2} du = C. \]

So we get from (38) that

\[ \|g\|^2 \leq C^2 \|f\|^2. \]

Using the value of C given by (36), we conclude that \( \|L\| \leq \sqrt{\pi} \). To show that equality holds, take \( f(t) = 1/\sqrt{t} \) for \( a < t < b \), zero outside this interval; for this choice \( \|f\|^2 = \log b/a \). Set \( g = L f \); it is not hard to show that as \( a \) tends to zero and \( b \) to \( \infty \), \( \|g\|^2 \geq \pi (1 - e) \log b/a \). Combined with \( \|L\| \leq \sqrt{\pi} \) this proves (33). \( \Box \)

Again note that the kernel of \( L_0 \), \( e^{-st} \), utterly fails the criteria for \( L^2 \) boundedness contained in either theorem 2 or theorem 3.

Exercise 3. Prove that the Laplace transform \( L \) is not bounded as a map of \( L^p(\mathbb{R}_+) \to L^p(\mathbb{R}_+) \), except for \( p = 2 \). (Hint: Try \( f(t) = e^{-at} \).)

As remarked in chapter 15, theorem 8, if \( L \) is bounded, so is \( L_0 \); therefore it follows by submultiplicity from (33) that

\[ \|L^2\| \leq \|L\|^2 = \pi. \]  

(39)

We claim that in (39) the sign of equality holds. To see this, we note that the kernel of the integral operator \( L \), \( e^{-st} \), is a real symmetric function of \( s \) and \( t \). It is easily verified (see formulas (6), (6')) that an integral operator \( L \) with a symmetric kernel satisfies

\[ (Lu, v) = (u, Lv). \]  

(40)

That is, such an operator is its own adjoint; such operators are called symmetric.

Theorem 10. Let \( L \) be a bounded, symmetric mapping of a real Hilbert space into itself. Then

\[ \|L^2\| = \|L\|^2. \]

Proof. By submultiplicity, \( \|L^2\| \leq \|L\|^2 \) is valid for all mappings, symmetric or not. To show the opposite inequality, we set \( v = Lu \) into (40); we get

\[ (Lu, Lu) = (u, L^2 u). \]

The left side equals \( \|Lu\|^2 \); estimating the right side by the Schwarz inequality gives

\[ \|Lu\|^2 \leq \|u\| \|L^2 u\| \leq \|u\|^2 \|L^2\|. \]

Since this holds for all vectors \( u \) in \( H \), \( \|L\|^2 \leq \|L^2\| \) follows. \( \Box \)

Clearly, it follows from theorem 10 that the sign of equality holds in relation (39).

The mapping \( L^2 \) is easily computed:

\[ (L^2 f)(r) = \int_0^\infty (Lf)(s) e^{-rs} ds = \int_0^\infty \int_0^\infty f(t) e^{-st} dt \, e^{-rs} ds \]

\[ = \int_0^\infty f(t) \int_t^\infty e^{-r(s+t)} ds \, dt = \int_0^\infty \frac{f(t)}{t + r} dt. \]

So we have proved

Theorem 11. The integral operator \( f \to g \):

\[ g(r) = \int_0^\infty \frac{f(t)}{t + r} dt \]

(41)

is bounded as a map of \( L^2(\mathbb{R}_+) \to L^2(\mathbb{R}_+) \), and its norm is equal to \( \pi \).

The map (41) is called the Hilbert-Hankel operator. Note that its kernel, \( 1/\sqrt{s + r} \), utterly fails the test for \( L^2 \) boundedness contained in either theorem 2 or 3.

Exercise 4. Prove that the Hilbert-Hankel operator is a bounded map of \( L^p \to L^p \) for \( 1 < p < \infty \).

For further information about integral operators, see Halmos and Sunder.
16.4 SOLUTION OPERATORS FOR HYPERBOLIC EQUATIONS

We recall from section 5 of chapter 11 the class of symmetric hyperbolic operators of first-order. These are first-order partial differential operators of the form

$$ L = \sum_{j=1}^{m} A_j \partial_j + B, \quad \partial_j = \frac{\partial}{\partial x_j}. \quad (42) $$

The $A_j, B$ are $n \times n$ matrices with real-valued entries that are reasonably smooth functions of $s$. We take them to be periodic in $s$. $L$ acts on vector-valued functions $u(s)$, whose components are real valued, and assumed reasonable smooth, periodic functions of $s$. As scalar product for such functions we take the $L^2$ scalar product over a period parallelogram $F$:

$$ (u, v) = \frac{1}{T} \int_{F} u \cdot v \, ds, \quad (43) $$

where the dot is the standard inner product for vectors. We assume the coefficient matrices $A_j$ to be symmetric:

$$ A^T_j = A_j. \quad (44) $$

In this case the formal adjoint $L^*$ of $L$ takes the form

$$ L^* = -L + K, \quad (44') $$

where

$$ K = B + B^T - \sum_{j=1}^{m} A_{j,j}, \quad (45) $$

Adjointness means that for smooth functions $u$ and $v$,

$$ (u, Lu) = \frac{1}{2} \int_{F} u \cdot (L^* v + v) \, ds, $$

from this and $44'$ we deduce, upon setting $v = u$, that

$$ 2(u, Lu) = 2(u, Ku), \quad (45') $$

see equation (20'), chapter 11.

**Theorem 12.** Let $u(s, t)$ be a solution of

$$ u_t + Lu = 0, \quad (46) $$

and suppose that $u$ is periodic in $s$. Then

$$ \|u(T)\| \leq c\|u(0)\|, \quad (47) $$

with $c$ a constant that may depend on $T$. Here the norm is the $L^2$ norm (43) over a period parallelogram $F$.

**Remark 1.** It follows from (47) that the solution $u$ is uniquely determined by its initial value $u(s, 0)$. Thus $u(T)$ is related to $u(0)$, and since equation (46) is linear, this relation is linear; denote by $S(T)$ the map relating $u(0)$ to $u(T)$:

$$ S(T) : u(0) = u(T). \quad (48) $$

$S(T)$ is called the solution operator. Theorem 12 states that for each $T$ the solution operator is bounded in the norm $L^2(F) \to L^2(F)$.

**Proof.** Assume first that the matrix $K$ in (45) is $0$ for all $s$. Take the scalar product of (46) with $2u$, and integrate over $F$. Using (43), we can write

$$ 2(u, u_t) + 2(u, Lu) = 0, $$

so by (45')

$$ 2(u, u_t) + (u, Ku) = 0. \quad (49) $$

The first term can be written as $d(u, u_t)/dt$. Therefore, if the symmetric matrix $K$ is $0$, it follows from (49) that $\|u(t)\|$ is decreasing as function of $T$; from this (47) follows for all $T > 0$, $c = 1$.

If $K$ is not positive, introduce $v$ by $u = e^{kt}v$ as new dependent variable; set this into (46) to obtain $v_t + (k + L)v = 0$. For $k$ large enough, $k + K \geq 0$, so $v$ satisfies (47) with $c = 1$. Therefore $u$ satisfies (47) with $c = e^{kt}$, $T > 0$.

Obviously the proof works also if $u$ is not periodic in $s$ but tends to zero as $|s| \to \infty$ so fast that $u \in L^2(R^n)$.

We take now a special example of an equation of form (42):

$$ A_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad B = 0. $$

Denoting $u = (u, u')$, we can write (42) componentwise as

$$ u_t + u_x + u_y = 0, \quad \nabla u_t - \nabla u_x + u_y = 0. $$

We can eliminate one of the two components, obtaining after a brief calculation

$$ u_t - u_{xx} - u_{yy} = 0, \quad u_{tt} - u_{xx} - u_{yy} = 0, $$

the classical wave equation. There is an explicit solution formula for the wave equation that puts the solution operator $S(T)$ in the form of an integral operator. The kernel of $S(T)$ is in this case just a function but a distribution; it utterly fails to satisfy the $L^2$ boundedness criteria stated in theorems 2 and 3.
16.5 SOLUTION OPERATOR FOR THE HEAT EQUATION

We consider solutions \(u(x, t)\) of the heat equation

\[ u_t = u_{xx} \quad (50) \]

that are defined for all \(x \) and all \(t \geq 0\), and which \(\to 0\) sufficiently rapidly as \(|x| \to \infty\).

**Theorem 13.** Let \(u(x, t)\) be a solution of the heat equation, as above. Then for all 
\( T > 0 \)

(i) \(|u(T)|_{max} \leq |u(0)|_{max}. \)
(ii) \(|u(T)|_{L^2} \leq |u(0)|_{L^2}. \)
(iii) \(|u(T)|_{L^2} \leq |u(0)|_{L^2}. \)

**Remark 2.** Since (50) is linear, these estimates show that \(u\) is uniquely determined by its initial value and that the dependence of \(u\) on its initial data is linear. Therefore the solution operator

\[ S(t) : u(0) \to u(t) \quad (51) \]

is well defined. In terms of it, theorem 13 can be formulated so: \(|S(t)| \leq 1\) as an operator mapping \(L^p \to L^p, p = \infty, 1, 2.\)

**Proof.** Let \(k\) be an arbitrary positive number. Define \(v(x, t)\) to be

\[ v = u e^{-kt}. \quad (52) \]

Then \(v\) satisfies the equation

\[ v_t + kv = v_{xx}. \quad (50') \]

Since \(u(x, t)\) was assumed to \(\to 0\) as \(|x| \to \infty\), the same is true of \(v(x, t)\). It follows that in the strip \(0 \leq t \leq T, -\infty < x < \infty\) the function \(|v(x, t)|\) takes on its maximum. We claim that this is at a point where \(t = 0\). Say that the maximum occurs where \(t = T\). If \(v(x, T) > 0\) at this point, then the first term on the left in (50') is \(> 0\), and the second term on the left is \(> 0\), while the term \(v_{xx}\) on the right is \(\leq 0\). At a negative minimum we find an analogous contradiction. So it follows that

\[ \max_{0 \leq x \leq T, x} |v(x, t)| = \max_{x} |v(x, 0)|. \]

This shows that \(v\) satisfies property (i) of theorem 13. Letting \(k \to 0\) in the definition (52) of \(v\) shows that also \(u\) satisfies (i):

\[ |S(T)| \leq 1, \quad S : L^\infty \to L^\infty. \]

**SOLUTION OPERATOR FOR THE HEAT EQUATION**

(ii) Consider the space of all solutions \(w(x, t)\) of the backward heat equation

\[ w_t = -w_{xx}, \quad (53) \]

defined for \(0 \leq t \leq T,\) and all \(x\), which tends to zero sufficiently rapidly as \(|x| \to \infty\). Multiply equation (50) by \(u\), (53) by \(u\), and add; the result can be written as

\[ (uw)_t = uw_{xx} - u_{xx}. \]

Integrate this with respect to \(x\) on \(\mathbb{R}\); integrate by parts. The fact that \(u, w\) tend to zero as \(|x| \to \infty\) shows that the integral of the right is zero. So we get

\[ 0 = \int (uw)_t dx = \frac{d}{dt} \int u w dx; \]

that is, \(\int u w dx = (u(t), w(t))\) is independent of \(t\), in particular,

\[ (u(0), w(0)) = (u(T), w(T)). \quad (54) \]

Denote the initial value \(u(0)\) by \(f\); in the notation (51), \(u(T) = S(T) f\). Similarly denote the final value \(w(T)\) by \(g\). Analogously to what we have shown about solutions of (50), for \(t < T, w(t)\) is completely determined by \(w(T)\), and there is a linear relation between \(w(T)\) and \(w(0)\) that we denote by \(S'\):

\[ w(0) = S'(T) g. \]

We rewrite (54) in this new notation as

\[ (f, S'(T) g) = (S(T) f, g). \quad (55) \]

The bracket \((u, w)\) is a bilinear function: for fixed \(w\), it is a linear functional of \(u\) and for fixed \(u\), a linear functional of \(w\). Thus (55) says that \(S\) and \(S'\) are *transposes* of each other with respect to this bilinear pairing.

It is easy to verify that

\[ |u|_{L^1} = \sup_{|u|_{max} = 1} |(u, w)|. \]

According to part (i), \(|S'(T) g|_{max} \leq |g|_{max}, so we deduce from (55) that \(|S(T) f|_{L^1} \leq |f|_{L^1}, as asserted in part (ii). Part (iii), the boundedness of \(S : L^2 \to L^2,\) follows from the Marcel Riesz convexity theorem, theorem 4 above.

**Remark 3.** Here is another, direct proof for part (iii). Multiply equation (50) by \(2u\) and integrate with respect to \(x\) over \(\mathbb{R}\). Integrate by parts on the right. By the fact that \(u(x, t) \to 0\) as \(|x| \to \infty\), we get

\[ \frac{d}{dt} \int u^2 dx = -\int u_x^2 dx. \]
EXAMPLES OF BOUNDED LINEAR MAPS

This shows that \( \int u^2(x, t) \, dx \) is a decreasing function of \( t \), from which part (iii) follows.

A similar direct proof can be given for part (ii). Let \( x_j(t) \) be the points where \( u(x, t) \) changes sign:

\[
    u(x, t) \begin{cases} 
        > 0 & \text{for } x_j < x < x_{j+1}, \ j \text{ even} \\
        < 0 & \text{for } x_j < x < x_{j+1}, \ j \text{ odd}.
    \end{cases}
\]  

(56)

Then

\[
    |u(t)|_{L^1} = \sum (-1)^j \int_{x_j(t)}^{x_{j+1}(t)} u(x, t) \, dx.
\]  

(57)

Differentiate this with respect to \( t \). Using calculus and equation (50), we get

\[
    \frac{d}{dt} |u(t)|_{L^1} = \sum (-1)^j \int_{x_j}^{x_{j+1}} u_t \, dx = \sum (-1)^j \int_{x_j}^{x_{j+1}} u_{xx} \, dx
\]  

\[
    = \sum (-1)^j (u_x(x_{j+1}) - u_x(x_j)).
\]  

(57')

It follows from (56) that the first \( x \)-derivative of \( u \) alternates in sign at the points \( x_j \):

\[
    u_x(x_j, t) \begin{cases} 
        \geq 0 & \text{for } j \text{ even} \\
        \leq 0 & \text{for } j \text{ odd};
    \end{cases}
\]

therefore the right side of (57') is \( \leq 0 \). This shows that \( |u(t)|_{L^1} \) is a decreasing function of \( t \), as asserted in part (ii).

We give now yet another proof of theorem 13; the initial value problem for (50) can be solved explicitly:

\[
    u(x, t) = \frac{1}{2\sqrt{\pi t}} \int f(y) e^{-(x-y)^2/4t} \, dy.
\]

This shows that \( S \) is an integral operator whose kernel is \( \exp((x-y)^2/4t)/2\sqrt{\pi t} \).

We appeal to parts (iii) and (iv) of theorem 1 to prove parts (i) and (ii) of theorem 13, and to theorem 3 to prove part (iii).

(57')

Theorem 13 holds for second-order parabolic equations in any number of space variables; the proofs sketched above apply to the general case, except of course the last one based on the explicit formula for the solution.

16.6 SINGULAR INTEGRAL OPERATORS, PSEUDODIFFERENTIAL OPERATORS AND FOURIER INTEGRAL OPERATORS

The above-named classes of operators play a dominant role in modern analysis, in particular, the modern theory of partial differential equations. They extend enormously the class of traditional integral operators; they unify integral and differential

BIBLIOGRAPHY


NORMED ALGEBRAS

The theorems contained in this section are valid for all Banach algebras \( \mathcal{L} \) with a unit, not just for \( \mathcal{L}(X, X) \).

**Definition.** An element \( M \) of a Banach algebra \( \mathcal{L} \) with a unit is called **invertible** if it has an inverse \( N = M^{-1} \) in \( \mathcal{L} \):

\[
NM = MN = I.
\]

(3)

\( M \) is said to have a **left inverse** \( A \), respectively a **right inverse** \( B \), if

\[
AM = I, \quad I = MB.
\]

(4)

It is an elementary fact of algebra that if \( M \) has both a left inverse \( A \) and a right inverse \( B \), then these are equal. For multiply the first relation in (4) by \( B \) on the right:

\[
AMB = B.
\]

(5)

Using associativity and the second relation (4) we get \( A = B \).

**Theorem 1.**

(i) If \( M \) and \( K \) in \( \mathcal{L} \) are invertible, so is their product \( MK \), and

\[
(MK)^{-1} = K^{-1}M^{-1}.
\]

(6)

(ii) If \( M \) and \( K \) commute,

\[
MK = KM,
\]

and if their product is invertible, so are \( M \) and \( K \) separately.

Proof. (i) is an obvious consequence of associativity. To show (ii), denote the inverse of \( MK \) by \( N \):

\[
(MK)N = I = N(MK).
\]

By \( N \) and \( K \)'s associativity, we conclude that \( KN \) is a right inverse of \( M \). By commutativity of \( M \) and \( K \) and associativity, we get

\[
I = N(MK) = N(KM) = (NK)M,
\]

from which we conclude that \( NK \) is a left inverse of \( M \). So \( M \) is invertible.

\[ \square \]

Theorem 1 is purely algebraic; not so

**Theorem 2.** Suppose that \( K \) in \( \mathcal{L} \) is invertible; then so are all elements of \( \mathcal{L} \) close enough to \( K \). Specifically all elements of form \( I = K - A \) are invertible, provided that

\[
|A| < \frac{1}{|K^{-1}|}.
\]

(8)
FUNCTIONAL CALCULUS

As path of integration we may choose any contour $C$ in the resolvent set of $M$ that winds once around $\sigma(M)$. It follows from the definition (12) that $|\zeta| = |\sigma(M)| + \delta$ is such a contour. We may then estimate from (17):

$$|M^n| \leq c \left(|\sigma(M)| + \delta\right)^{n+1}, \quad c = \max_{|\zeta| = |\sigma(M)| + \delta} |(2) - M^{-1}|.$$

Take the $n$th root,

$$|M^n|^{1/n} \leq c^{1/n} \left(|\sigma(M)| + \delta\right)^{(n+1)/n}$$

and then form the lim sup,

$$\limsup_{n \to \infty} |M^n|^{1/n} \leq |\sigma(M)| + \delta.$$

Since this is true for any $\delta > 0$, it is true for $\delta = 0$;

$$\limsup_{n \to \infty} |M^n|^{1/n} \leq |\sigma(M)|.$$

Comparing (16) and (18), we conclude that the lim inf and lim sup are equal, and we obtain Gelfand's formula (12') for the spectral radius. \(\square\)

17.2 FUNCTIONAL CALCULUS

Since $\mathcal{L}$ is an algebra, we can form any polynomial $p$ of an element $M$ of $\mathcal{L}$ by setting

$$p(M) = \sum_{j=0}^N a_j M^j.$$  (19)

(19) defines a mapping from the algebra of polynomials into the algebra $\mathcal{L}$ that is, clearly, a homomorphism. This homomorphism can be extended to a larger class of functions than polynomials; for instance, we can define

$$e^M = \sum_{n=0}^\infty \frac{M^n}{n!}.$$  

More generally, we can define

$$f(M) = \sum_{n=0}^\infty a_n M^n$$  (20)

for any entire function

$$f(\zeta) = \sum_{n=0}^\infty a_n \zeta^n.$$  (20')

Still more generally it follows from (12') that we can define (20) for any function $f(\zeta)$ whose power series converges in a circle whose radius exceeds $|\sigma(M)|$. We propose now a still further extension:
values lie in a Banach algebra. As the reader may immediately verify, the product of two such analytic functions is analytic. All the standard paraphernalia of the theory of analytic functions—the Cauchy integral theorem, the Cauchy integral formula, power series, Laurent series, and so on—are meaningful and valid for functions that take their values in a Banach algebra.

**Theorem 3.**

(i) The resolvent set $\rho(M)$ is an open subset of $\mathbb{C}$.

(ii) The resolvent of $M$, defined on $\rho(M)$ as $(\zeta - M)^{-1}$, abbreviated as $(\zeta - M)^{-1}$, is an analytic function of $\zeta$ on $\rho(M)$.

**Proof.** Suppose that $\lambda$ is in $\rho(M)$; then by theorem 2 applied to $K = \lambda I - M$ and $A = hI$,

$$(\lambda - h)I - M = (\lambda I - M - hI)$$

is invertible for $h$ small enough. This proves part (i).

By formula (10'),

$$(\lambda - h) - M)^{-1} = \sum_{0}^{\infty} (\lambda - M)^{-1} h^n;$$

(11)

this shows that the resolvent can be expanded in a power series around each point $\lambda$ of $\rho(M)$, convergent for $|h| < |(\lambda - M)^{-1}|^{-1}$, this proves analyticity, as asserted in part (ii).

The series (11) converges when $|h|$ is less than $|(\lambda - M)^{-1}|^{-1}$. From this we deduce

**Corollary 3'.** For any $\lambda$ in $\rho(M)$, denote by $d(\lambda)$ the distance of $\lambda$ to the spectrum of $M$. Then

$$|(\lambda - M)^{-1}| \geq d^{-1};$$

(11')

**Theorem 4 (Gelfand).**

(i) The spectrum $\sigma(M)$ is a closed, bounded, nonempty set in $\mathbb{C}$.

(ii) The spectral radius of $M$, denoted as $|\sigma(M)|$, is defined as

$$|\sigma(M)| = \max_{\lambda \in \sigma(M)} |\lambda|.$$  

(12)

We claim that

$$|\sigma(M)| = \lim_{k \to \infty} |M^k|^{1/k}.$$  

(12')
Definition. Let $M$ be an element of $L$, $f(\xi)$ a function analytic in a domain $G$ containing $\sigma(M)$. Let $C$ be a contour in $G \cap \rho(M)$ that winds once around every point in $\sigma(M)$ but winds zero time around any point of the complement of $G$. We define

$$f(M) = \oint (\xi - M)^{-1} f(\xi) \, d\xi.$$  \hspace{1cm} (21)

By the Cauchy integral theorem, (21) is independent of the choice of the contour.

Theorem 5.

(i) For $f$ a polynomial, definitions (21) and (19) are the same.

(ii) The mapping (21) from the algebra of functions analytic on an open set containing $\sigma(M)$ into $L$ is a homomorphism.

(iii) $\sigma(f(M)) = f(\sigma(M)).$ \hspace{1cm} (22)

(iv) Let $f$ be analytic on an open set containing $\sigma(M)$, and $g$ analytic on an open set containing $f(\sigma(M))$. Denote their composite by $h$,

$$h(\xi) = g(f(\xi)),$$ \hspace{1cm} (23)

then

$$h(M) = g(f(M)).$$ \hspace{1cm} (23')

Proof. (i) Replacing $f(\xi)$ by a polynomial in (21) and using formula (17) shows that (21) and (19) are the same. The same argument shows that (21) is the same as (20) for $f$ analytic in a disk of radius $> \sigma(M)$.

(ii) For any pair of complex numbers $\xi$ and $\omega$,

$$(\xi - I - M) - (\omega I - M) = (\xi - \omega)I.$$

Suppose that both $\xi$ and $\omega$ belong to $\rho(M)$. Multiply the above identity by

$$(\xi - M)^{-1}(\omega - M)^{-1}(\xi - \omega)^{-1}.$$  \hspace{1cm} (24)

Relation (24) is called the resolvent identity.

The mapping $f \mapsto f(M)$ given by (22) is obviously linear. We show now that it is multiplicative. Let $f, g$ both be functions analytic in an open set $G \supset \sigma(M)$. We choose two contours $C$ and $D$, both in $G \cap \rho(M)$ so that they have no point in common, and so that $D$ lies inside $C$. That is, $C$ winds once around every point $\omega$ of $D$, while $D$ winds zero times around every point $\xi$ of $C$. Using definition (21) for $f$ and $g$ with contours $C$ and $D$, we write $f(M)g(M)$ as a product of two integrals, which we express as a double integral. Then we use the resolvent identity (24):

$$f(M)g(M) = \oint \oint (\xi - M)^{-1}(\omega - M)^{-1} f(\xi)g(\omega) \, d\xi \, d\omega$$

$$= \oint \oint (\xi - \omega)^{-1}(\omega - M)^{-1}(\xi - M)^{-1} f(\xi)g(\omega) \, d\xi \, d\omega$$

$$= \oint \left[ \oint (\xi - \omega)^{-1} f(\xi) \, d\xi \right] (\omega - M)^{-1} g(\omega) \, d\omega$$

$$- \oint \left[ \oint (\xi - \omega)^{-1} g(\omega) \, d\omega \right] (\xi - M)^{-1} f(\xi) \, d\xi.$$  \hspace{1cm} (25)

Since $C$ winds once around every point $\omega$ of $D$, the integral with respect to $\xi$ above in the first term is, by the Cauchy integral formula, $f(\omega)$. Since $D$ does not wind around any point $\xi$ on $C$, the $\omega$ integration in the second term above is zero; so we conclude from (25) that

$$f(M)g(M) = \oint (\omega - M)^{-1} f(\omega)g(\omega) \, d\omega,$$

which by (21) is $h(M)$, where $h(\omega) = f(\omega)g(\omega)$. This proves that the mapping (21) is multiplicative.

(iii) We have to show that $\mu$ belongs to the spectrum of $f(M)$ if and only if $\mu$ is of the form

$$\mu = f(\lambda), \quad \lambda \in \sigma(M).$$  \hspace{1cm} (26)

If $\mu$ is not of form (26), then $f(\xi) - \mu$ does not vanish on $\sigma(M)$. Therefore $(f(\xi) - \mu)^{-1} = g(\xi)$ is analytic in an open set containing $\sigma(M)$, thus we may define $g(M)$ by formula (21). According to part (ii), $[f(M) - \mu I] g(M) = h(M)$, where $h(\xi) = f(\xi) - \mu I$. Thus $h(M) = I$, and $g(M)$ is the inverse of $f(M) - \mu I$. This proves that $\mu$ does not lie in $\sigma(f(M))$.

On the other hand, suppose that $\mu$ is of form (26). Define the function $k(\xi)$ by

$$k(\xi) = \frac{f(\xi) - f(\lambda)}{\xi - \lambda}.$$  \hspace{1cm} (27)

Clearly, $k$ is analytic in an open set containing $\sigma(M)$, so $k(M)$ can be defined by (21). Since $(\xi - \lambda) k(\xi) = f(\xi) - f(\lambda)$, it follows from part (ii) that

$$(M - \lambda I) k(M) = f(M) - f(\lambda) I.$$  \hspace{1cm} (28)

Since $\lambda$ belongs to $\sigma(M)$, the first factor is not invertible. We appeal now to part (i) of theorem 1, according to which the product $f(M) - f(\lambda) I$ is not invertible either.

(iv) By assumption, $g(\omega)$ is analytic on $f(\sigma(M))$. Since by part (iii) the spectrum of $f(M)$ is $f(\sigma(M))$, it follows that formula (21) can be applied to $g$ in place of $f$, etc.
and \( f(M) \) in place of \( M \) and \( D \) in place of \( C \):

\[
g(f(M)) = \oint (\omega - f(M))^{-1} g(\omega) \, d\omega,
\]

(28)

For \( \omega \) on \( D \), \((\omega - f(\zeta))^{-1} \) is an analytic function on \( \sigma(M) \), therefore applying formula (21) once more, we get

\[
(\omega I - f(M))^{-1} = \oint (\xi - M)^{-1} (\omega - f(\xi))^{-1} d\xi.
\]

(29)

provided that the contour \( C \) does not wind around any of the points \( \omega \) on \( D \). Now set (29) into (28):

\[
g(f(M)) = \oint \oint (\xi - M)^{-1} (\omega - f(\xi))^{-1} g(\omega) d\xi \, d\omega.
\]

(30)

We reverse the order of integration; since \( C \) does not wind around points of \( D \), it follows that \( D \) winds around every point \( \xi \) of \( C \). By the Cauchy integral formula,

\[
\oint (\omega - f(\xi))^{-1} g(\omega) \, d\omega = g(f(\xi)) = h(\xi),
\]

where we have used (23). Setting this in (30) on the right we get, by (21), \( h(M) \), as asserted in (23). \qed

Definition (21) and properties listed in theorem 5 are called the functional calculus for operators. Relation (22) is called the spectral mapping theorem.

Suppose that the spectrum of \( M \) can be decomposed as the union of \( \text{pairwise disjoint closed components:} \)

\[
\sigma(M) = \sigma_1 \cup \cdots \cup \sigma_N, \quad \sigma_j \cap \sigma_k = \varnothing.
\]

(31)

For each \( j \), denote by \( C_j \) a contour in the resolvent of \( M \) that winds once around each point of \( \sigma_j \) but not \( \sigma_k, k \neq j \). We define

\[
P_j = \oint (\xi - M)^{-1} d\xi_j.
\]

(32)

**Theorem 6.**

(i) The \( P_j \) are disjoint projections, that is,

\[
P_j^2 = P_j \quad \text{and} \quad P_j P_k = 0 \quad \text{for} \quad j \neq k.
\]

(33)

(ii) \[
\sum_j P_j = I.
\]

(34)

(iii) \( P_n \neq 0 \) if \( \sigma_n \) is not empty.

BIBLIOGRAPHY

*Proof.* Relations (33) are corollaries of part (ii) of theorem 5. Since \( C = \sum C_j \) winds once around every point of \( \sigma(M) \), (34) follows by summing (32) over all \( j \), and using (14). We leave the proof of part (iii) to the diligent reader. \( \square \)

**Exercise 1.** Show that if \( P \) is a nonzero projection, that is, satisfies \( P^2 = P \neq 0 \), then

\[
|P| \geq 1.
\]

(35)

**Exercise 2.** Show that the spectral radius \( |\sigma(M)| \) depends upper semicontinuously on \( M \) in the norm topology, namely, that if \( \lim M_n = M \), then

\[
\lim \sup |\sigma(M_n)| \leq |\sigma(M)|.
\]

**Exercise 3.** Show that \( |\exp M| \leq \exp |M| \).

**Exercise 4.** Show that if \( 0 \) does not belong to \( \sigma(M) \), and if \( 0 \) can be connected to \( \infty \) by curve that lies in \( \rho(M) \), then \( \log(M) \) can be defined so that \( \exp \log(M) = M \).

**Exercise 5.** Define \( \mathcal{L}_M \) to be the closure of the algebra generated by \( M \) and \((\xi - M)^{-1}, \xi \in \rho(M) \). Show that \( \mathcal{L}_M \) is a commutative subalgebra of \( \mathcal{L} \).

**Note.** For the history of the spectral theory of operators in a Banach space, see pp. 607–609 of Dunford and Schwartz.

The term "spectrum" is due to Hilbert, a remarkable anticipation of its meaning in quantum mechanics.

**BIBLIOGRAPHY**


