1. Foundations

The sets of functions which form the subject matter of this book are to be considered as sequences in *metric spaces*. Actually we shall be almost exclusively concerned with various L^p spaces, particularly the case $p = 2$, and with subspaces of such spaces. Although the notes which follow in § 1.1 and § 1.2 contain sufficient metric space theory for an understanding of the rest of the book, the reader who is new to metric spaces may wish to fill in from a good text such as Copson (1967). For general background reading Simmons (1963) is also highly recommended.

Throughout the book an effort has been made to present theorems which are sufficiently general to be 'useful', but in a small introductory book of this kind a great deal of detail has to be left out; adequate references are given for those who want to consult more advanced sources.

1.1 Notes on metric spaces

1.1.1 Vector space It is assumed that the reader is familiar with elementary set theory. The word 'space' is used in mathematics to mean a set with some internal 'structure'. Let V be a set whose elements are to be called *vectors*, and let F be a field (the *field of scalars*; we will usually take it to be the field of complex numbers). The basic structure that we shall require for V is that it be closed under an operation of addition of two vectors u and v, denoted by $u+v$, and an operation of multiplication of a vector u by a scalar f of F, denoted by fu. If V is to become a useful mathematical system we shall require more structure within it than the presence of these two operations. If in addition the two operations satisfy the following list of axioms, then V is called a *vector space*.

(1) $u+v \in V$ for every u and v in V (V is closed under the operation of addition).

(2) $u+(v+w) = (u+v)+w$ for every u, v, w in V (addition is associative).

(3) $u+v = v+u$ for every u and v in V (addition is commutative).

(4) V contains a vector θ such that $u+\theta = u$ for every u in V (θ is the 'null vector' of V).

(5) For every $u \in V$ there is a vector $-u$ in V such that $u+(-u) \equiv u-u = \theta$ (each vector has an additive inverse).

(6) fu is in V for every f in F and every u in V (V is closed under multiplication by scalars).

(7) Let 1 denote the multiplicative unit of F. Then $1u = u$ for every u in V.

(8) For every u and v in V and f and g in F we have

$$f(u+v) = fu+fv$$

and $$(f+g)\,u = fu+gu \text{ (distributive laws).}$$

(9) $(fg)\,u = f(gu)$ for every f and g in F and u in V.

The first five axioms express the Abelian group character of V.

From now on, V will denote a vector space; the definitions to follow give further structure to V.

1.1.2 Metric and norm A *metric* on V is a real valued 'distance' function ρ, defined on pairs (u, v) of vectors in V, such that for every u, v, w in V we have

(1) $\rho(u,v) \geqslant 0$, and $\rho(u,v) = 0$ if and only if $u = v$;

(2) $\rho(u,v)+\rho(v,w) \geqslant \rho(u,w)$ (triangle inequality);

(3) $\rho(u,v) = \rho(v,u)$.

Note that a space need not be a vector space in order to define a metric on it.

A *norm* on V is a real valued function defined on V and denoted by $\| \;\|$, such that

(1) $\|u\| \geqslant 0$, $\|u\| = 0$ if and only if $u = \theta$;

(2) $\|u+v\| \leqslant \|u\| + \|v\|$ (triangle inequality);
(3) $\|fu\| = |f|\,\|u\|$ for every complex number f.

The norm generalises the notion of absolute value of complex numbers. Evidently the choice $\rho(u,v) = \|u-v\|$ provides a metric on V, and this special metric is called the *metric induced by the norm*. A space possessing a metric is called a *metric space* and one possessing a norm is called a *normed space*. From now on unless otherwise stated V will denote a normed vector space with metric induced by the norm.

1.1.3 Convergence Let (u_n) be a sequence (see §1.1.10) of elements of V. Then (u_n) is said to be *convergent* if there exists u in V, called the *limit* of the sequence, such that $\|u_n - u\| \to 0$ as $n \to \infty$. A convergent sequence has a unique limit. The reader must notice carefully that convergence means convergence in the norm of V to an element of V. For example, one can construct a sequence of rationals which 'converges' to $\sqrt{2}$, but here convergence must be understood to take place within the normed vector space of real numbers, even though each member of the sequence is a rational number. If one were to speak only of the rationals, the sequence would not be convergent; it would be a 'Cauchy sequence', however, and these ideas lead to a most important and desirable property of metric spaces, that of 'completeness' (see § 1.1.5).

1.1.4 Closed sets Let $S \subset V$ and $u \in V$. Then u is called a *point of closure* of S if, given $\epsilon > 0$, there exists an $s \in S$ such that $\|u-s\| < \epsilon$. The *closure* $\bar{S}$ of S is the collection of all points of closure of S. S is called *closed* if $S = \bar{S}$. For every S we have $S \subset \bar{S} = \bar{\bar{S}}$.

Let (u_n) be a sequence in V. Any expression formed from vectors of this sequence by use of the two basic operations of addition and of multiplication by a scalar is called a *linear combination* of those vectors. The collection U of all such finite linear combinations is called the *linear span* of (u_n) and $\bar{U}$ is called the *closed linear span* of (u_n) frequently denoted by $[u_n]$. One sometimes says that (u_n) *spans* U.

1.1.5 Cauchy sequences and completeness The sequence (u_n) in V is said to be a *Cauchy sequence* if, given $\epsilon > 0$, there exists N such that $\|u_n - u_m\| < \epsilon$ for every n and m greater than N. V is said to be *complete* if every Cauchy sequence in it converges.

1.1.6 Dense subsets A subset $S \subset V$ is said to be *dense* in V if for every $u \in V$ and $\epsilon > 0$, there exists a vector $s \in S$ such that $\|u - s\| < \epsilon$. The reader may verify that if $S_1 \subset S_2 \subset V$ with S_1 dense in S_2 and S_2 dense in V, then S_1 is dense in V. We shall refer to this as the *chain of dense subsets* principle. If V contains a countable dense subset then it is called *separable*.

1.1.7 Banach space In his well-known book, *Théorie des opérations linéaires*, Stefan Banach (1932) referred to certain spaces as '*les espaces du type* (B)' and such spaces have carried his name ever since; we are now in a position to write down the definition: a normed vector space which is complete in the metric induced by the norm is called a *Banach space*.

1.1.8 Hilbert space Let V be a vector space. A complex valued function defined on pairs of vectors of V is called an *inner product*, and written (u, v) for u and v in V, if it satisfies

(1) $(u_1 + u_2, v) = (u_1, v) + (u_2, v) \quad (u_1, u_2, v \in V)$;

(2) $(cu, v) = c(u, v)$ for every $c \in F$; ((1) and (2) express 'linearity' in the first argument);

(3) $(u, v) = \overline{(v, u)}$ (here a bar denotes complex conjugate; this is the 'Hermitian' symmetry property of the inner product);

(4) $(u, u) \geqslant 0$, and $(u, u) = 0$ if and only if $u = \theta$.

A space V in which each pair of vectors has an inner product is called an *inner product space*. Note that by (3) (u, u) is real; in order to construct a norm for V we may put $(u, u)^{\frac{1}{2}} = \|u\|$, and this choice does indeed provide a norm (see problem 1.4).

THEOREM (Schwarz' inequality) *Let V be an inner product* *space with norm given by* $\|u\| = (u,u)^{\frac{1}{2}}$*. Then* $|(u,v)| \leqslant \|u\|\,\|v\|$ *for every* u *and* v *in* V*.*

Proof Assume that neither u nor v is null, since in this case the theorem is obvious. For any scalar c we have $\|u+cv\|^2 \geqslant 0$. As the proof proceeds we shall see how to choose an appropriate c. We have

$$\begin{aligned} 0 &\leqslant (u+cv, u+cv) \\ &= \|u\|^2 + |c|^2\|v\|^2 + (cv,u) + (u,cv) \\ &= \|u\|^2 + |c|^2\|v\|^2 + 2\,\mathrm{Re}\, c(u,v) \text{ by (1), (2) and (3) above.} \end{aligned}$$

Choose arg c such that $c(u,v)$ is real and negative; i.e. choose $\arg c = \pi - \arg(u,v)$. Then

$$\|u\|^2 + |c|^2\|v\|^2 \geqslant -2c(u,v) = 2|c|\,|(u,v)|.$$

Now choose $|c| = \|u\|/\|v\|$, which yields the required result.

Let V be an inner product space. If V is a Banach space with respect to the norm defined by $\|u\| = (u,u)^{\frac{1}{2}}$, then V is said to be a *Hilbert space*. Thus the Hilbert spaces form a subclass of the Banach spaces.

The reader cannot fail to have noticed that metric spaces appear to have 'geometrical' properties analogous to properties of ordinary finite dimensional Euclidean space. For example, the metric itself provides the notion of distance in V, the norm gives the distance from the 'origin' and is thus some measurement of the size of an element, and in Hilbert space the inner product generalises the dot product of ordinary vectors. Indeed, the bases which are the subject of this book are nothing but generalisations of the bases of unit vectors of finite dimensional vector spaces. The reader will find many more such geometrical facts in the following pages, and is enjoined to develop a geometrical way of thinking about Hilbert and Banach space.

Problems

1.1 Show that in a metric space the triangle inequality

$$\|u+v\| \leqslant \|u\| + \|v\|$$

is equivalent to $\quad \|u-v\| \geqslant \big|\|u\| - \|v\|\big|.$

1.2 Show that one has equality in Schwarz' inequality if and only if u and v are linearly dependent.

1.3 Formulate and prove a Pythagoras theorem in Hilbert space.

1.4 Show that the choice $\|u\| = (u,u)^{\frac{1}{2}}$ provides a norm for the inner product space V.

1.5 Prove the 'parallelogram identity'

$$\|u+v\|^2 + \|u-v\|^2 = 2\|u\|^2 + 2\|v\|^2$$

in Hilbert space, and justify the name 'parallelogram'.

1.6 Show that if the norm of a Banach space satisfies the parallelogram identity then it is a Hilbert space. Hint: introduce an inner product by use of the 'polarisation identity'

$$(u, v) = \tfrac{1}{4}(\|u+v\|^2 - \|u-v\|^2 + i\|u+iv\|^2 - i\|u-iv\|^2).$$

1.7 Show that the norm is a continuous function on a Banach space to $\mathbb{R}$.

1.1.9 The projection theorem Let H be a Hilbert space. A *linear manifold* or *subspace* of H is a subset which is algebraically closed under the operation of taking linear combinations of its elements. The reader should check that this definition does lead to what he would expect from the word 'subspace'. A linear manifold which is closed as a subset of H is called a *closed linear manifold.*

Two vectors of H are said to be *orthogonal* if their inner product is zero; a vector is said to be orthogonal to a subspace if it is orthogonal to every vector of that subspace. The *orthogonal complement* $S^\perp$ of a subset S of H is the collection of all vectors orthogonal to S.

Let S and T be subspaces of H such that $S \cap T = \{\theta\}$; then the set of all vectors of the form $u+v$ with u in S and v in T is called the *direct sum* of S and T and is written $S \oplus T$. The following theorem is a most satisfying result, and appeals to our geometrical way of thinking about Hilbert space.

THEOREM (Projection theorem) *Let S be a subspace of Hilbert space H. Then $S \oplus S^\perp = H$.*

For the proof, see Yosida (1965) p. 82.

1.1.10 Mappings The words mapping, function, functional, operator, transformation, etc. occur repeatedly in mathematical writing and, whilst their meanings have become fairly standard, no completely satisfactory standardisation has so far evolved. We shall adopt the following definitions.

A *function* on, *mapping* or *transformation* of, a set V into a set W is a rule f which assigns a unique $w \in W$ to each $v \in V$. This association will be denoted by the usual functional notation $w = f(v)$. V is called the *domain* of f and $W_1 = \{f(v) : v \in V\}$ is called the *range* of f. The notation $f\colon V \to W$ is used to mean 'f maps V into W'. Sometimes we call w the *image of v by f*, and this notation carries over to sets; thus the set

$$\{f(v) : v \in A \subset V\} = f(A)$$

is called the *image of A by f*.

If $W_1 = W$ then f is called a mapping *onto* W.

A mapping f is called *one-to-one* if $v_1 \neq v_2$ implies $f(v_1) \neq f(v_2)$.

If V is a vector space, a mapping f of V is called *linear* if it preserves the two basic operations of addition and multiplication by scalars, that is, if $f(c_1 v_1 + c_2 v_2) = c_1 f(v_1) + c_2 f(v_2)$ for every c_1, c_2 in F and v_1, v_2 in V.

A *functional* is a mapping of V to $\mathbb{R}$. If V is a normed vector space a functional f on V is called *bounded* if there exists a real number c such that, for every $v \in V$, $|f(v)| \leqslant c\|v\|$; the infimum of all such cs is called the *norm* of f and written $\|f\|$.

If $f\colon V \to V$ then f is called an *operator* on V.

A *sequence* in W is a mapping of a subset J of $\mathbb{R}$ into W. Usually, but not always (see, e.g. § 2.2), J, the *indexing set*, is such that $J \subseteq \mathbb{N}$. If a sequence maps $n \in J$ to $\phi_n \in W$ then it is denoted by $(\phi_n)_{n \in J}$.

A mapping f between two normed spaces is called an *isometry* if it is one-to-one and norm preserving, that is, whenever $w = f(v)$ then $\|w\| = \|v\|$.

A mapping which is one-to-one, linear and onto is called an *isomorphism*. A mapping which is both isometric and isomorphic is called an *isometric isomorphism*. An operator which is also an isometry on a Hilbert space is called a *unitary operator*.

1.1.11 The dual of a Banach space, strong and weak convergence Let B be a Banach space. The class B^* of all bounded linear functionals on B is also a Banach space whose norm is the linear functional norm defined in § 1.1.10. B^* is called the *dual space* of B. One can define a linear functional on the dual space by the process of fixing v in B and forming

$$F_v(f) = f(v)$$

where f varies over B^*. Then F_v is obviously a bounded linear functional on B^* for every $v \in B$. It may turn out that B^* admits no other bounded linear functionals, in which case there is a natural one-to-one association of points $v \in B$ with points F_v in $(B^*)^* = B^{**}$, the *second dual* space of B. Banach spaces with this property are called *reflexive*. The L^p spaces (see § 1.2) are reflexive Banach spaces for $p > 1$, but not for $p = 1$.

Convergence in B is often called strong convergence, that is, the sequence (v_n) *converges strongly* to v if $\|v_n - v\| \to 0$ as $n \to \infty$. There is a companion mode of convergence in B called weak convergence, associated with the linear functionals on B. The sequence (v_n) *converges weakly* to v if $|f(v_n) - f(v)| \to 0$ as $n \to \infty$ for every $f \in B^*$. *Strong convergence implies weak convergence to the same limit*, for

$$|f(v_n) - f(v)| = |f(v_n - v)| \leqslant \|f\| \, \|v_n - v\|.$$

The converse is not true (find an example to show this !).

1.2 Notes on the L^p spaces

In this section we present some of the basic facts about the L^p spaces, again in note form.

Let X denote any measurable subset of $\mathbb{R}$, of finite or infinite Lebesgue measure. For any real number p such that $1 \leqslant p < \infty$, we are going to consider the class $L^p(X)$ of complex valued functions whose pth powers are Lebesgue measurable and integrable over X. These classes can be generalised in various ways, e.g. by defining the functions on a σ-finite measure space.

Let p be given, and f lie in the class described above. Put

$$\|f\|_p = \left\{\int_X |f|^p\right\}^{1/p}.$$

We have deliberately chosen the notation of the norm here. We shall also need $L^\infty(X) = \{f : f$ bounded except possibly on a set of measure zero, measurable on a set X of finite measure, and ess sup $|f| = \lim_{p\to\infty} \|f\|_p < \infty\}$.

HÖLDER'S INEQUALITY *Let $f \in L^p(X)$, $g \in L^q(X)$, $1/p + 1/q = 1, p \geqslant 1$. Then fg is integrable over X, that is, $fg \in L^1(X)$, and*

$$\|fg\|_1 \leqslant \|f\|_p \|g\|_q.$$

Equality holds if and only if f and g are linearly dependent.

MINKOWSKI'S INEQUALITY *Let both f and g belong to $L^p(X)$, $p \geqslant 1$. Then $f+g \in L^p(X)$, and*

$$\|f+g\|_p \leqslant \|f\|_p + \|g\|_p$$

Equality holds if and only if f and g are linearly dependent.

There are various extensions of these inequalities, for example to the case of more than two functions, and to other values of p (when $p < 1$ the inequalities are reversed; see Hardy, Littlewood and Polya (1952)).

For a given p, the class of functions and the formula for $\|f\|_p$ described above yield a Banach space $L^p(X)$ and its norm, provided that one more stipulation is made. This is that two functions f and g whose values differ only on a set of measure zero must be considered as the same Banach space element, since otherwise we should have more than one null element and this would contradict the definition of the norm. Thus $L^p(X)$ consists of equivalence classes, two functions being in the same equivalence class if they differ only on a set of measure zero.

Minkowski's inequality is the triangle inequality for the norm.

When $p = 1$ we frequently omit the '1' from notation. When $1 < p < \infty$ $L^p(X)$ is a separable, reflexive Banach space whose dual space is isometrically isomorphic to $L^q(X)$, $1/p + 1/q = 1$; the completeness is a well-known theorem of Riesz and Fischer.

From Hölder's inequality we have $L^p(X) \subset L^r(X)$ if $r < p$ and X is of finite measure.

In the special case $p = 2$ we have

$$\|f\|_2^2 = \int_X |f|^2 = \int_X f\bar{f}.$$

We can define an inner product on $L^2(X)$ by putting

$$(f, g) = \int_X f\bar{g} < \infty;$$

then $L^2(X)$ becomes a Hilbert space. Note that Schwarz' inequality for this Hilbert space is a special case of Hölder's inequality. We shall also need the Hilbert space $L^2(X, w)$ of (equivalence classes of) functions which are square integrable with respect to the 'weight function' w, i.e. those f for which

$$\int_X |f|^2 w < \infty \text{ (see pp. 28–33).}$$

The weak convergence of the sequence (v_n) to v in $L^2(X)$ is equivalent to the condition

$$(w, v_n) \to (w, v) \quad (w \in L^2(X)).$$

The continuous functions are dense in $L^p(a, b)$, $1 \leqslant p < \infty$, for a finite interval (a, b) (so are they in $L^p(\mathbb{R}^n)$). The class $C(a, b)$ of all continuous functions on a closed finite interval $[a, b]$ is a Banach space with norm $\|f\| = \sup\{|f(x)| : x \in [a, b]\}$. Then the Weierstrass approximation theorem (Appendix 1,3(a)) says that the polynomials are dense in C. Now if $f \in C$, then $f \in L^p(a, b)$ and a simple calculation shows that the polynomials are also dense in C in the L^p norm. From the chain of dense subsets principle (§ 1.1.6) it follows that the polynomials are dense in $L^p(a, b)$, $1 \leqslant p < \infty$. Note that $[a, b]$ must be a *finite* interval. We can now state an important result: *The set* $\{\sum a_n x^n\}$ *of all linear combinations of powers* $\{x^n : n = 0, 1, 2, \ldots\}$ *is dense in* $L^p(a, b)$, $1 \leqslant p < \infty$. We shall return to this result and generalise it in § 2.1.