

Contents

Foreword	vii
Preface	xvii
Chapter 1. L^p Spaces and Banach Spaces	1
1 L^p spaces	2
1.1 The Hölder and Minkowski inequalities	3
1.2 Completeness of L^p	5
1.3 Further remarks	7
2 The case $p = \infty$	7
3 Banach spaces	9
3.1 Examples	9
3.2 Linear functionals and the dual of a Banach space	11
4 The dual space of L^p when $1 \leq p < \infty$	13
5 More about linear functionals	16
5.1 Separation of convex sets	16
5.2 The Hahn-Banach Theorem	20
5.3 Some consequences	21
5.4 The problem of measure	23
6 Complex L^p and Banach spaces	27
7 Appendix: The dual of $C(X)$	28
7.1 The case of positive linear functionals	29
7.2 The main result	32
7.3 An extension	33
8 Exercises	34
9 Problems	43
Chapter 2. L^p Spaces in Harmonic Analysis	47
1 Early Motivations	48
2 The Riesz interpolation theorem	52
2.1 Some examples	57
3 The L^p theory of the Hilbert transform	61
3.1 The L^2 formalism	61
3.2 The L^p theorem	64
3.3 Proof of Theorem 3.2	66
4 The maximal function and weak-type estimates	70
4.1 The L^p inequality	71

5	The Hardy space H_r^1	73
5.1	Atomic decomposition of H_r^1	74
5.2	An alternative definition of H_r^1	81
5.3	Application to the Hilbert transform	82
6	The space H_r^1 and maximal functions	84
6.1	The space BMO	86
7	Exercises	90
8	Problems	94
Chapter 3. Distributions: Generalized Functions		98
1	Elementary properties	99
1.1	Definitions	100
1.2	Operations on distributions	102
1.3	Supports of distributions	104
1.4	Tempered distributions	105
1.5	Fourier transform	107
1.6	Distributions with point supports	110
2	Important examples of distributions	111
2.1	The Hilbert transform and $\text{pv}(\frac{1}{x})$	111
2.2	Homogeneous distributions	115
2.3	Fundamental solutions	125
2.4	Fundamental solution to general partial differential equations with constant coefficients	129
2.5	Parametrices and regularity for elliptic equations	131
3	Calderón-Zygmund distributions and L^p estimates	134
3.1	Defining properties	134
3.2	The L^p theory	138
4	Exercises	145
5	Problems	153
Chapter 4. Applications of the Baire Category Theorem		157
1	The Baire category theorem	158
1.1	Continuity of the limit of a sequence of continuous functions	160
1.2	Continuous functions that are nowhere differentiable	163
2	The uniform boundedness principle	166
2.1	Divergence of Fourier series	167
3	The open mapping theorem	170
3.1	Decay of Fourier coefficients of L^1 -functions	173
4	The closed graph theorem	174
4.1	Grothendieck's theorem on closed subspaces of L^p	174

CONTENTS	xiii
5 Besicovitch sets	176
6 Exercises	181
7 Problems	185
Chapter 5. Rudiments of Probability Theory	188
1 Bernoulli trials	189
1.1 Coin flips	189
1.2 The case $N = \infty$	191
1.3 Behavior of S_N as $N \rightarrow \infty$, first results	194
1.4 Central limit theorem	195
1.5 Statement and proof of the theorem	197
1.6 Random series	199
1.7 Random Fourier series	202
1.8 Bernoulli trials	204
2 Sums of independent random variables	205
2.1 Law of large numbers and ergodic theorem	205
2.2 The role of martingales	208
2.3 The zero-one law	215
2.4 The central limit theorem	215
2.5 Random variables with values in \mathbb{R}^d	220
2.6 Random walks	222
3 Exercises	227
4 Problems	235
Chapter 6. An Introduction to Brownian Motion	238
1 The Framework	239
2 Technical Preliminaries	241
3 Construction of Brownian motion	246
4 Some further properties of Brownian motion	251
5 Stopping times and the strong Markov property	253
5.1 Stopping times and the Blumenthal zero-one law	254
5.2 The strong Markov property	258
5.3 Other forms of the strong Markov Property	260
6 Solution of the Dirichlet problem	264
7 Exercises	268
8 Problems	273
Chapter 7. A Glimpse into Several Complex Variables	276
1 Elementary properties	276
2 Hartogs' phenomenon: an example	280

3	Hartogs' theorem: the inhomogeneous Cauchy-Riemann equations	283
4	A boundary version: the tangential Cauchy-Riemann equations	288
5	The Levi form	293
6	A maximum principle	296
7	Approximation and extension theorems	299
8	Appendix: The upper half-space	307
	8.1 Hardy space	308
	8.2 Cauchy integral	311
	8.3 Non-solvability	313
9	Exercises	314
10	Problems	319
Chapter 8. Oscillatory Integrals in Fourier Analysis		321
1	An illustration	322
2	Oscillatory integrals	325
3	Fourier transform of surface-carried measures	332
4	Return to the averaging operator	337
5	Restriction theorems	343
	5.1 Radial functions	343
	5.2 The problem	345
	5.3 The theorem	345
6	Application to some dispersion equations	348
	6.1 The Schrödinger equation	348
	6.2 Another dispersion equation	352
	6.3 The non-homogeneous Schrödinger equation	355
	6.4 A critical non-linear dispersion equation	359
7	A look back at the Radon transform	363
	7.1 A variant of the Radon transform	363
	7.2 Rotational curvature	365
	7.3 Oscillatory integrals	367
	7.4 Dyadic decomposition	370
	7.5 Almost-orthogonal sums	373
	7.6 Proof of Theorem 7.1	374
8	Counting lattice points	376
	8.1 Averages of arithmetic functions	377
	8.2 Poisson summation formula	379
	8.3 Hyperbolic measure	384
	8.4 Fourier transforms	389
	8.5 A summation formula	392

CONTENTS	xv
9 Exercises	398
10 Problems	405
Notes and References	409
Bibliography	413
Symbol Glossary	417
Index	419

1 L^p Spaces and Banach Spaces

In this work the assumption of quadratic integrability will be replaced by the integrability of $|f(x)|^p$. The analysis of these function classes will shed a particular light on the real and apparent advantages of the exponent 2; one can also expect that it will provide essential material for an axiomatic study of function spaces.

F. Riesz, 1910

At present I propose above all to gather results about linear operators defined in certain general spaces, notably those that will here be called *spaces of type (B)*...

S. Banach, 1932

Function spaces, in particular L^p spaces, play a central role in many questions in analysis. The special importance of L^p spaces may be said to derive from the fact that they offer a partial but useful generalization of the fundamental L^2 space of square integrable functions.

In order of logical simplicity, the space L^1 comes first since it occurs already in the description of functions integrable in the Lebesgue sense. Connected to it via duality is the L^∞ space of bounded functions, whose supremum norm carries over from the more familiar space of continuous functions. Of independent interest is the L^2 space, whose origins are tied up with basic issues in Fourier analysis. The intermediate L^p spaces are in this sense an artifice, although of a most inspired and fortuitous kind. That this is the case will be illustrated by results in the next and succeeding chapters.

In this chapter we will concentrate on the basic structural facts about the L^p spaces. Here part of the theory, in particular the study of their linear functionals, is best formulated in the more general context of Banach spaces. An incidental benefit of this more abstract view-point is that it leads us to the surprising discovery of a finitely additive measure on *all* subsets, consistent with Lebesgue measure.

1 L^p spaces

Throughout this chapter (X, \mathcal{F}, μ) denotes a σ -finite measure space: X denotes the underlying space, \mathcal{F} the σ -algebra of measurable sets, and μ the measure. If $1 \leq p < \infty$, the space $L^p(X, \mathcal{F}, \mu)$ consists of all complex-valued measurable functions on X that satisfy

$$(1) \quad \int_X |f(x)|^p d\mu(x) < \infty.$$

To simplify the notation, we write $L^p(X, \mu)$, or $L^p(X)$, or simply L^p when the underlying measure space has been specified. Then, if $f \in L^p(X, \mathcal{F}, \mu)$ we define the L^p **norm** of f by

$$\|f\|_{L^p(X, \mathcal{F}, \mu)} = \left(\int_X |f(x)|^p d\mu(x) \right)^{1/p}.$$

We also abbreviate this to $\|f\|_{L^p(X)}$, $\|f\|_{L^p}$, or $\|f\|_p$.

When $p = 1$ the space $L^1(X, \mathcal{F}, \mu)$ consists of all integrable functions on X , and we have shown in Chapter 6 of Book III, that L^1 together with $\|\cdot\|_{L^1}$ is a complete normed vector space. Also, the case $p = 2$ warrants special attention: it is a Hilbert space.

We note here that we encounter the same technical point that we already discussed in Book III. The problem is that $\|f\|_{L^p} = 0$ does not imply that $f = 0$, but merely $f = 0$ almost everywhere (for the measure μ). Therefore, the precise definition of L^p requires introducing the equivalence relation, in which f and g are equivalent if $f = g$ a.e. Then, L^p consists of all equivalence classes of functions which satisfy (1). However, in practice there is little risk of error by thinking of elements in L^p as functions rather than equivalence classes of functions.

The following are some common examples of L^p spaces.

- (a) The case $X = \mathbb{R}^d$ and μ equals Lebesgue measure is often used in practice. There, we have

$$\|f\|_{L^p} = \left(\int_{\mathbb{R}^d} |f(x)|^p dx \right)^{1/p}.$$

- (b) Also, one can take $X = \mathbb{Z}$, and μ equal to the counting measure. Then, we get the “discrete” version of the L^p spaces. Measurable functions are simply sequences $f = \{a_n\}_{n \in \mathbb{Z}}$ of complex numbers,

and

$$\|f\|_{L^p} = \left(\sum_{n=-\infty}^{\infty} |a_n|^p \right)^{1/p}.$$

When $p = 2$, we recover the familiar sequence space $\ell^2(\mathbb{Z})$.

The spaces L^p are examples of normed vector spaces. The basic property satisfied by the norm is the triangle inequality, which we shall prove shortly.

The range of p which is of interest in most applications is $1 \leq p < \infty$, and later also $p = \infty$. There are at least two reasons why we restrict our attention to these values of p : when $0 < p < 1$, the function $\|\cdot\|_{L^p}$ does not satisfy the triangle inequality, and moreover, for such p , the space L^p has no non-trivial bounded linear functionals.¹ (See Exercise 2.)

When $p = 1$ the norm $\|\cdot\|_{L^1}$ satisfies the triangle inequality, and L^1 is a complete normed vector space. When $p = 2$, this result continues to hold, although one needs the Cauchy-Schwarz inequality to prove it. In the same way, for $1 \leq p < \infty$ the proof of the triangle inequality relies on a generalized version of the Cauchy-Schwarz inequality. This is Hölder's inequality, which is also the key in the duality of the L^p spaces, as we will see in Section 4.

1.1 The Hölder and Minkowski inequalities

If the two exponents p and q satisfy $1 \leq p, q \leq \infty$, and the relation

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

holds, we say that p and q are **conjugate** or **dual** exponents. Here, we use the convention $1/\infty = 0$. Later, we shall sometimes use p' to denote the conjugate exponent of p . Note that $p = 2$ is self-dual, that is, $p = q = 2$; also $p = 1, \infty$ corresponds to $q = \infty, 1$ respectively.

Theorem 1.1 (Hölder) *Suppose $1 < p < \infty$ and $1 < q < \infty$ are conjugate exponents. If $f \in L^p$ and $g \in L^q$, then $fg \in L^1$ and*

$$\|fg\|_{L^1} \leq \|f\|_{L^p} \|g\|_{L^q}.$$

Note. Once we have defined L^∞ (see Section 2) the corresponding inequality for the exponents 1 and ∞ will be seen to be essentially trivial.

¹We will define what we mean by a bounded linear functional later in the chapter.

The proof of the theorem relies on a simple generalized form of the arithmetic-geometric mean inequality: if $A, B \geq 0$, and $0 \leq \theta \leq 1$, then

$$(2) \quad A^\theta B^{1-\theta} \leq \theta A + (1 - \theta)B.$$

Note that when $\theta = 1/2$, the inequality (2) states the familiar fact that the geometric mean of two numbers is majorized by their arithmetic mean.

To establish (2), we observe first that we may assume $B \neq 0$, and replacing A by AB , we see that it suffices to prove that $A^\theta \leq \theta A + (1 - \theta)$. If we let $f(x) = x^\theta - \theta x - (1 - \theta)$, then $f'(x) = \theta(x^{\theta-1} - 1)$. Thus $f(x)$ increases when $0 \leq x \leq 1$ and decreases when $1 \leq x$, and we see that the continuous function f attains a maximum at $x = 1$, where $f(1) = 0$. Therefore $f(A) \leq 0$, as desired.

To prove Hölder's inequality we argue as follows. If either $\|f\|_{L^p} = 0$ or $\|g\|_{L^q} = 0$, then $fg = 0$ a.e. and the inequality is obviously verified. Therefore, we may assume that neither of these norms vanish, and after replacing f by $f/\|f\|_{L^p}$ and g by $g/\|g\|_{L^q}$, we may further assume that $\|f\|_{L^p} = \|g\|_{L^q} = 1$. We now need to prove that $\|fg\|_{L^1} \leq 1$.

If we set $A = |f(x)|^p$, $B = |g(x)|^q$, and $\theta = 1/p$ so that $1 - \theta = 1/q$, then (2) gives

$$|f(x)g(x)| \leq \frac{1}{p}|f(x)|^p + \frac{1}{q}|g(x)|^q.$$

Integrating this inequality yields $\|fg\|_{L^1} \leq 1$, and the proof of the Hölder inequality is complete.

For the case when the equality $\|fg\|_{L^1} = \|f\|_{L^p}\|g\|_{L^q}$ holds, see Exercise 3.

We are now ready to prove the triangle inequality for the L^p norm.

Theorem 1.2 (Minkowski) *If $1 \leq p < \infty$ and $f, g \in L^p$, then $f + g \in L^p$ and $\|f + g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p}$.*

Proof. The case $p = 1$ is obtained by integrating $|f(x) + g(x)| \leq |f(x)| + |g(x)|$. When $p > 1$, we may begin by verifying that $f + g \in L^p$, when both f and g belong to L^p . Indeed,

$$|f(x) + g(x)|^p \leq 2^p(|f(x)|^p + |g(x)|^p),$$

as can be seen by considering separately the cases $|f(x)| \leq |g(x)|$ and $|g(x)| \leq |f(x)|$. Next we note that

$$|f(x) + g(x)|^p \leq |f(x)| |f(x) + g(x)|^{p-1} + |g(x)| |f(x) + g(x)|^{p-1}.$$

If q denotes the conjugate exponent of p , then $(p-1)q = p$, so we see that $(f+g)^{p-1}$ belongs to L^q , and therefore Hölder's inequality applied to the two terms on the right-hand side of the above inequality gives

$$(3) \quad \|f+g\|_{L^p}^p \leq \|f\|_{L^p} \|(f+g)^{p-1}\|_{L^q} + \|g\|_{L^p} \|(f+g)^{p-1}\|_{L^q}.$$

However, using once again $(p-1)q = p$, we get

$$\|(f+g)^{p-1}\|_{L^q} = \|f+g\|_{L^p}^{p/q}.$$

From (3), since $p - p/q = 1$, and because we may suppose that $\|f+g\|_{L^p} > 0$, we find

$$\|f+g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p},$$

so the proof is finished.

1.2 Completeness of L^p

The triangle inequality makes L^p into a metric space with distance $d(f, g) = \|f - g\|_{L^p}$. The basic analytic fact is that L^p is **complete** in the sense that every Cauchy sequence in the norm $\|\cdot\|_{L^p}$ converges to an element *in* L^p .

Taking limits is a necessity in many problems, and the L^p spaces would be of little use if they were not complete. Fortunately, like L^1 and L^2 , the general L^p space does satisfy this desirable property.

Theorem 1.3 *The space $L^p(X, \mathcal{F}, \mu)$ is complete in the norm $\|\cdot\|_{L^p}$.*

Proof. The argument is essentially the same as for L^1 (or L^2); see Section 2, Chapter 2 and Section 1, Chapter 4 in Book III. Let $\{f_n\}_{n=1}^\infty$ be a Cauchy sequence in L^p , and consider a subsequence $\{f_{n_k}\}_{k=1}^\infty$ of $\{f_n\}$ with the following property $\|f_{n_{k+1}} - f_{n_k}\|_{L^p} \leq 2^{-k}$ for all $k \geq 1$. We now consider the series whose convergence will be seen below

$$f(x) = f_{n_1}(x) + \sum_{k=1}^{\infty} (f_{n_{k+1}}(x) - f_{n_k}(x))$$

and

$$g(x) = |f_{n_1}(x)| + \sum_{k=1}^{\infty} |f_{n_{k+1}}(x) - f_{n_k}(x)|,$$

and the corresponding partial sums

$$S_K(f)(x) = f_{n_1}(x) + \sum_{k=1}^K (f_{n_{k+1}}(x) - f_{n_k}(x))$$

and

$$S_K(g)(x) = |f_{n_1}(x)| + \sum_{k=1}^K |f_{n_{k+1}}(x) - f_{n_k}(x)|.$$

The triangle inequality for L^p implies

$$\begin{aligned} \|S_K(g)\|_{L^p} &\leq \|f_{n_1}\|_{L^p} + \sum_{k=1}^K \|f_{n_{k+1}} - f_{n_k}\|_{L^p} \\ &\leq \|f_{n_1}\|_{L^p} + \sum_{k=1}^K 2^{-k}. \end{aligned}$$

Letting K tend to infinity, and applying the monotone convergence theorem proves that $\int g^p < \infty$, and therefore the series defining g , and hence the series defining f converges almost everywhere, and $f \in L^p$.

We now show that f is the desired limit of the sequence $\{f_n\}$. Since (by construction of the telescopic series) the $(K-1)^{\text{th}}$ partial sum of this series is precisely f_{n_K} , we find that

$$f_{n_K}(x) \rightarrow f(x) \quad \text{a.e. } x.$$

To prove that $f_{n_K} \rightarrow f$ in L^p as well, we first observe that

$$\begin{aligned} |f(x) - S_K(f)(x)|^p &\leq [2 \max(|f(x)|, |S_K(f)(x)|)]^p \\ &\leq 2^p |f(x)|^p + 2^p |S_K(f)(x)|^p \\ &\leq 2^{p+1} |g(x)|^p, \end{aligned}$$

for all K . Then, we may apply the dominated convergence theorem to get $\|f_{n_K} - f\|_{L^p} \rightarrow 0$ as K tends to infinity.

Finally, the last step of the proof consists of recalling that $\{f_n\}$ is Cauchy. Given $\epsilon > 0$, there exists N so that for all $n, m > N$ we have $\|f_n - f_m\|_{L^p} < \epsilon/2$. If n_K is chosen so that $n_K > N$, and $\|f_{n_K} - f\|_{L^p} < \epsilon/2$, then the triangle inequality implies

$$\|f_n - f\|_{L^p} \leq \|f_n - f_{n_K}\|_{L^p} + \|f_{n_K} - f\|_{L^p} < \epsilon$$

whenever $n > N$. This concludes the proof of the theorem.

1.3 Further remarks

We begin by looking at some possible inclusion relations between the various L^p spaces. The matter is simple if the underlying space has finite measure.

Proposition 1.4 *If X has finite positive measure, and $p_0 \leq p_1$, then $L^{p_1}(X) \subset L^{p_0}(X)$ and*

$$\frac{1}{\mu(X)^{1/p_0}} \|f\|_{L^{p_0}} \leq \frac{1}{\mu(X)^{1/p_1}} \|f\|_{L^{p_1}}.$$

We may assume that $p_1 > p_0$. Suppose $f \in L^{p_1}$, and set $F = |f|^{p_0}$, $G = 1$, $p = p_1/p_0 > 1$, and $1/p + 1/q = 1$, in Hölder's inequality applied to F and G . This yields

$$\|f\|_{L^{p_0}}^{p_0} \leq \left(\int |f|^{p_1} \right)^{p_0/p_1} \cdot \mu(X)^{1-p_0/p_1}.$$

In particular, we find that $\|f\|_{L^{p_0}} < \infty$. Moreover, by taking the p_0^{th} root of both sides of the above equation, we find that the inequality in the proposition holds.

However, as is easily seen, such inclusion does not hold when X has infinite measure. (See Exercise 1). Yet, in an interesting special case the opposite inclusion does hold.

Proposition 1.5 *If $X = \mathbb{Z}$ is equipped with counting measure, then the reverse inclusion holds, namely $L^{p_0}(\mathbb{Z}) \subset L^{p_1}(\mathbb{Z})$ if $p_0 \leq p_1$. Moreover, $\|f\|_{L^{p_1}} \leq \|f\|_{L^{p_0}}$.*

Indeed, if $f = \{f(n)\}_{n \in \mathbb{Z}}$, then $\sum |f(n)|^{p_0} = \|f\|_{L^{p_0}}^{p_0}$, and $\sup_n |f(n)| \leq \|f\|_{L^{p_0}}$. However

$$\begin{aligned} \sum |f(n)|^{p_1} &= \sum |f(n)|^{p_0} |f(n)|^{p_1-p_0} \\ &\leq \left(\sup_n |f(n)| \right)^{p_1-p_0} \|f\|_{L^{p_0}}^{p_0} \\ &\leq \|f\|_{L^{p_0}}^{p_1}. \end{aligned}$$

Thus $\|f\|_{L^{p_1}} \leq \|f\|_{L^{p_0}}$.

2 The case $p = \infty$

Finally, we also consider the limiting case $p = \infty$. The space L^∞ will be defined as all functions that are “essentially bounded” in the following sense. We take the space $L^\infty(X, \mathcal{F}, \mu)$ to consist of all (equivalence

classes of) measurable functions on X , so that there exists a positive number $0 < M < \infty$, with

$$|f(x)| \leq M \quad \text{a.e. } x.$$

Then, we define $\|f\|_{L^\infty(X, \mathcal{F}, \mu)}$ to be the infimum of all possible values M satisfying the above inequality. The quantity $\|f\|_{L^\infty}$ is sometimes called the **essential-supremum** of f .

We note that with this definition, we have $|f(x)| \leq \|f\|_{L^\infty}$ for a.e. x . Indeed, if $E = \{x : |f(x)| > \|f\|_{L^\infty}\}$, and $E_n = \{x : |f(x)| > \|f\|_{L^\infty} + 1/n\}$, then we have $\mu(E_n) = 0$, and $E = \bigcup E_n$, hence $\mu(E) = 0$.

Theorem 2.1 *The vector space L^∞ equipped with $\|\cdot\|_{L^\infty}$ is a complete vector space.*

This assertion is easy to verify and is left to the reader. Moreover, Hölder's inequality continues to hold for values of p and q in the larger range $1 \leq p, q \leq \infty$, once we take $p = 1$ and $q = \infty$ as conjugate exponents, as we mentioned before.

The fact that L^∞ is a limiting case of L^p when p tends to ∞ can be understood as follows.

Proposition 2.2 *Suppose $f \in L^\infty$ is supported on a set of finite measure. Then $f \in L^p$ for all $p < \infty$, and*

$$\|f\|_{L^p} \rightarrow \|f\|_{L^\infty} \quad \text{as } p \rightarrow \infty.$$

Proof. Let E be a measurable subset of X with $\mu(E) < \infty$, and so that f vanishes in the complement of E . If $\mu(E) = 0$, then $\|f\|_{L^\infty} = \|f\|_{L^p} = 0$ and there is nothing to prove. Otherwise

$$\|f\|_{L^p} = \left(\int_E |f(x)|^p d\mu \right)^{1/p} \leq \left(\int_E \|f\|_{L^\infty}^p d\mu \right)^{1/p} \leq \|f\|_{L^\infty} \mu(E)^{1/p}.$$

Since $\mu(E)^{1/p} \rightarrow 1$ as $p \rightarrow \infty$, we find that $\limsup_{p \rightarrow \infty} \|f\|_{L^p} \leq \|f\|_{L^\infty}$.

On the other hand, given $\epsilon > 0$, we have

$$\mu(\{x : |f(x)| \geq \|f\|_{L^\infty} - \epsilon\}) \geq \delta \quad \text{for some } \delta > 0,$$

hence

$$\int_X |f|^p d\mu \geq \delta (\|f\|_{L^\infty} - \epsilon)^p.$$

Therefore $\liminf_{p \rightarrow \infty} \|f\|_{L^p} \geq \|f\|_{L^\infty} - \epsilon$, and since ϵ is arbitrary, we have $\liminf_{p \rightarrow \infty} \|f\|_{L^p} \geq \|f\|_{L^\infty}$. Hence the limit $\lim_{p \rightarrow \infty} \|f\|_{L^p}$ exists, and equals $\|f\|_{L^\infty}$.

3 Banach spaces

We introduce here a general notion which encompasses the L^p spaces as specific examples.

First, a **normed vector space** consists of an underlying vector space V over a field of scalars (the real or complex numbers), together with a **norm** $\| \cdot \| : V \rightarrow \mathbb{R}^+$ that satisfies:

- $\|v\| = 0$ if and only if $v = 0$.
- $\|\alpha v\| = |\alpha| \|v\|$, whenever α is a scalar and $v \in V$.
- $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in V$.

The space V is said to be **complete** if whenever $\{v_n\}$ is a Cauchy sequence in V , that is, $\|v_n - v_m\| \rightarrow 0$ as $n, m \rightarrow \infty$, then there exists a $v \in V$ such that $\|v_n - v\| \rightarrow 0$ as $n \rightarrow \infty$.

A complete normed vector space is called a **Banach space**. Here again, we stress the importance of the fact that Cauchy sequences converge to a limit in the space itself, hence the space is “closed” under limiting operations.

3.1 Examples

The real numbers \mathbb{R} with the usual absolute value form an initial example of a Banach space. Other easy examples are \mathbb{R}^d , with the Euclidean norm, and more generally a Hilbert space with its norm given in terms of its inner product.

Several further relevant examples are as follows:

EXAMPLE 1. The family of L^p spaces with $1 \leq p \leq \infty$ which we have just introduced are also important examples of Banach spaces (Theorem 1.3 and Theorem 2.1). Incidentally, L^2 is the only Hilbert space in the family L^p , where $1 \leq p \leq \infty$ (Exercise 25) and this in part accounts for the special flavor of the analysis carried out in L^2 as opposed to L^1 or more generally L^p for $p \neq 2$.

Finally, observe that since the triangle inequality fails in general when $0 < p < 1$, $\| \cdot \|_{L^p}$ is not a norm on L^p for this range of p , hence it is not a Banach space.

EXAMPLE 2. Another example of a Banach space is $C([0, 1])$, or more generally $C(X)$ with X a compact set in a metric space, as will be defined in Section 7. By definition, $C(X)$ is the vector space of continuous

functions on X equipped with the sup-norm $\|f\| = \sup_{x \in X} |f(x)|$. Completeness is guaranteed by the fact that the uniform limit of a sequence of continuous functions is also continuous.

EXAMPLE 3. Two further examples are important in various applications. The first is the space $\Lambda^\alpha(\mathbb{R})$ of all bounded functions on \mathbb{R} which satisfy a **Hölder (or Lipschitz) condition of exponent α** with $0 < \alpha \leq 1$, that is,

$$\sup_{t_1 \neq t_2} \frac{|f(t_1) - f(t_2)|}{|t_1 - t_2|^\alpha} < \infty.$$

Observe that f is then necessarily continuous; also the only interesting case is when $\alpha \leq 1$, since a function which satisfies a Hölder condition of exponent α with $\alpha > 1$ is constant.²

More generally, this space can be defined on \mathbb{R}^d ; it consists of continuous functions f equipped with the norm

$$\|f\|_{\Lambda^\alpha(\mathbb{R}^d)} = \sup_{x \in \mathbb{R}^d} |f(x)| + \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|^\alpha}.$$

With this norm, $\Lambda^\alpha(\mathbb{R}^d)$ is a Banach space (see also Exercise 29).

EXAMPLE 4. A function $f \in L^p(\mathbb{R}^d)$ is said to have **weak derivatives** in L^p up to order k , if for every multi-index $\alpha = (\alpha_1, \dots, \alpha_d)$ with $|\alpha| = \alpha_1 + \dots + \alpha_d \leq k$, there is a $g_\alpha \in L^p$ with

$$(4) \quad \int_{\mathbb{R}^d} g_\alpha(x) \varphi(x) dx = (-1)^{|\alpha|} \int_{\mathbb{R}^d} f(x) \partial_x^\alpha \varphi(x) dx$$

for all smooth functions φ that have compact support in \mathbb{R}^d . Here, we use the multi-index notation

$$\partial_x^\alpha = \left(\frac{\partial}{\partial x} \right)^\alpha = \left(\frac{\partial}{\partial x_1} \right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_d} \right)^{\alpha_d}.$$

Clearly, the functions g_α (when they exist) are unique, and we also write $\partial_x^\alpha f = g_\alpha$. This definition arises from the relationship (4) which holds whenever f is itself smooth, and g equals the usual derivative $\partial_x^\alpha f$, as follows from an integration by parts (see also Section 3.1, Chapter 5 in Book III).

²We have already encountered this space in Book I, Chapter 2 and Book III, Chapter 7.

The space $L_k^p(\mathbb{R}^d)$ is the subspace of $L^p(\mathbb{R}^d)$ of all functions that have weak derivatives up to order k . (The concept of weak derivatives will reappear in Chapter 3 in the setting of derivatives in the sense of distributions.) This space is usually referred to as a **Sobolev space**. A norm that turns $L_k^p(\mathbb{R}^d)$ into a Banach space is

$$\|f\|_{L_k^p(\mathbb{R}^d)} = \sum_{|\alpha| \leq k} \|\partial_x^\alpha f\|_{L^p(\mathbb{R}^d)}.$$

EXAMPLE 5. In the case $p = 2$, we note in the above example that an L^2 function f belongs to $L_k^2(\mathbb{R}^d)$ if and only if $(1 + |\xi|^2)^{k/2} \hat{f}(\xi)$ belongs to L^2 , and that $\|(1 + |\xi|^2)^{k/2} \hat{f}(\xi)\|_{L^2}$ is a Hilbert space norm equivalent to $\|f\|_{L_k^2(\mathbb{R}^d)}$.

Therefore, if k is any positive number, it is natural to define L_k^2 as those functions f in L^2 for which $(1 + |\xi|^2)^{k/2} \hat{f}(\xi)$ belongs to L^2 , and we can equip L_k^2 with the norm $\|f\|_{L_k^2(\mathbb{R}^d)} = \|(1 + |\xi|^2)^{k/2} \hat{f}(\xi)\|_{L^2}$.

3.2 Linear functionals and the dual of a Banach space

For the sake of simplicity, we restrict ourselves in this and the following two sections to Banach spaces over \mathbb{R} ; the reader will find in Section 6 the slight modifications necessary to extend the results to Banach spaces over \mathbb{C} .

Suppose that \mathcal{B} is a Banach space over \mathbb{R} equipped with a norm $\|\cdot\|$. A **linear functional** is a linear mapping ℓ from \mathcal{B} to \mathbb{R} , that is, $\ell : \mathcal{B} \rightarrow \mathbb{R}$, which satisfies

$$\ell(\alpha f + \beta g) = \alpha \ell(f) + \beta \ell(g), \quad \text{for all } \alpha, \beta \in \mathbb{R}, \text{ and } f, g \in \mathcal{B}.$$

A linear functional ℓ is **continuous** if given $\epsilon > 0$ there exists $\delta > 0$ so that $|\ell(f) - \ell(g)| \leq \epsilon$ whenever $\|f - g\| \leq \delta$. Also we say that a linear functional is **bounded** if there is $M > 0$ with $|\ell(f)| \leq M\|f\|$ for all $f \in \mathcal{B}$. The linearity of ℓ shows that these two notions are in fact equivalent.

Proposition 3.1 *A linear functional on a Banach space is continuous, if and only if it is bounded.*

Proof. The key is to observe that ℓ is continuous if and only if ℓ is continuous at the origin.

Indeed, if ℓ is continuous, we choose $\epsilon = 1$ and $g = 0$ in the above definition so that $|\ell(f)| \leq 1$ whenever $\|f\| \leq \delta$, for some $\delta > 0$. Hence,

given any non-zero h , an element of \mathcal{B} , we see that $\delta h/\|h\|$ has norm equal to δ , and hence $|\ell(\delta h/\|h\|)| \leq 1$. Thus $|\ell(h)| \leq M\|h\|$ with $M = 1/\delta$.

Conversely, if ℓ is bounded it is clearly continuous at the origin, hence continuous.

The significance of continuous linear functionals in terms of closed hyperplanes in \mathcal{B} is a noteworthy geometric point to which we return later on. Now we take up analytic aspects of linear functionals.

The set of all continuous linear functionals over \mathcal{B} is a vector space since we may add linear functionals and multiply them by scalars:

$$(\ell_1 + \ell_2)(f) = \ell_1(f) + \ell_2(f) \quad \text{and} \quad (\alpha\ell)(f) = \alpha\ell(f).$$

This vector space may be equipped with a norm as follows. The **norm** $\|\ell\|$ of a continuous linear functional ℓ is the infimum of all values M for which $|\ell(f)| \leq M\|f\|$ for all $f \in \mathcal{B}$. From this definition and the linearity of ℓ it is clear that

$$\|\ell\| = \sup_{\|f\| \leq 1} |\ell(f)| = \sup_{\|f\|=1} |\ell(f)| = \sup_{f \neq 0} \frac{|\ell(f)|}{\|f\|}.$$

The vector space of all continuous linear functionals on \mathcal{B} equipped with $\|\cdot\|$ is called the **dual space** of \mathcal{B} , and is denoted by \mathcal{B}^* .

Theorem 3.2 *The vector space \mathcal{B}^* is a Banach space.*

Proof. It is clear that $\|\cdot\|$ defines a norm, so we only check that \mathcal{B}^* is complete. Suppose that $\{\ell_n\}$ is a Cauchy sequence in \mathcal{B}^* . Then, for each $f \in \mathcal{B}$, the sequence $\{\ell_n(f)\}$ is Cauchy, hence converges to a limit, which we denote by $\ell(f)$. Clearly, the mapping $\ell : f \mapsto \ell(f)$ is linear. If M is so that $\|\ell_n\| \leq M$ for all n , we see that

$$|\ell(f)| \leq |(\ell - \ell_n)(f)| + |\ell_n(f)| \leq |(\ell - \ell_n)(f)| + M\|f\|,$$

so that in the limit as $n \rightarrow \infty$, we find $|\ell(f)| \leq M\|f\|$ for all $f \in \mathcal{B}$. Thus ℓ is bounded. Finally, we must show that ℓ_n converges to ℓ in \mathcal{B}^* . Given $\epsilon > 0$ choose N so that $\|\ell_n - \ell_m\| < \epsilon/2$ for all $n, m > N$. Then, if $n > N$, we see that for all $m > N$ and any f

$$|(\ell - \ell_n)(f)| \leq |(\ell - \ell_m)(f)| + |(\ell_m - \ell_n)(f)| \leq |(\ell - \ell_m)(f)| + \frac{\epsilon}{2}\|f\|.$$

We can also choose m so large (and dependent on f) so that we also have $|(\ell - \ell_m)(f)| \leq \epsilon\|f\|/2$. In the end, we find that for $n > N$,

$$|(\ell - \ell_n)(f)| \leq \epsilon\|f\|.$$

This proves that $\|\ell - \ell_n\| \rightarrow 0$, as desired.

In general, given a Banach space \mathcal{B} , it is interesting and very useful to be able to describe its dual \mathcal{B}^* . This problem has an essentially complete answer in the case of the L^p spaces introduced before.

4 The dual space of L^p when $1 \leq p < \infty$

Suppose that $1 \leq p \leq \infty$ and q is the conjugate exponent of p , that is, $1/p + 1/q = 1$. The key observation to make is the following: Hölder's inequality shows that every function $g \in L^q$ gives rise to a bounded linear functional on L^p by

$$(5) \quad \ell(f) = \int_X f(x)g(x) d\mu(x),$$

and that $\|\ell\| \leq \|g\|_{L^q}$. Therefore, if we associate g to ℓ above, then we find that $L^q \subset (L^p)^*$ when $1 \leq p \leq \infty$. The main result in this section is to prove that when $1 \leq p < \infty$, every linear functional on L^p is of the form (5) for some $g \in L^q$. This implies that $(L^p)^* = L^q$ whenever $1 \leq p < \infty$. We remark that this result is in general not true when $p = \infty$; the dual of L^∞ contains L^1 , but it is larger. (See the end of Section 5.3 below.)

Theorem 4.1 *Suppose $1 \leq p < \infty$, and $1/p + 1/q = 1$. Then, with $\mathcal{B} = L^p$ we have*

$$\mathcal{B}^* = L^q,$$

in the following sense: For every bounded linear functional ℓ on L^p there is a unique $g \in L^q$ so that

$$\ell(f) = \int_X f(x)g(x) d\mu(x), \quad \text{for all } f \in L^p.$$

Moreover, $\|\ell\|_{\mathcal{B}^} = \|g\|_{L^q}$.*

This theorem justifies the terminology whereby q is usually called the dual exponent of p .

The proof of the theorem is based on two ideas. The first, as already seen, is Hölder's inequality; to which a converse is also needed. The second is the fact that a linear functional ℓ on L^p , $1 \leq p < \infty$, leads naturally to a (signed) measure ν . Because of the continuity of ℓ the measure ν is absolutely continuous with respect to the underlying measure μ , and our desired function g is then the density function of ν in terms of μ .

We begin with:

Lemma 4.2 *Suppose $1 \leq p, q \leq \infty$, are conjugate exponents.*

(i) *If $g \in L^q$, then $\|g\|_{L^q} = \sup_{\|f\|_{L^p} \leq 1} \left| \int fg \right|$.*

(ii) *Suppose g is integrable on all sets of finite measure, and*

$$\sup_{\substack{\|f\|_{L^p} \leq 1 \\ f \text{ simple}}} \left| \int fg \right| = M < \infty.$$

Then $g \in L^q$, and $\|g\|_{L^q} = M$.

For the proof of the lemma, we recall the **signum** of a real number defined by

$$\text{sign}(x) = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0. \end{cases}$$

Proof. We start with (i). If $g = 0$, there is nothing to prove, so we may assume that g is not 0 a.e., and hence $\|g\|_{L^q} \neq 0$. By Hölder's inequality, we have that

$$\|g\|_{L^q} \geq \sup_{\|f\|_{L^p} \leq 1} \left| \int fg \right|.$$

To prove the reverse inequality we consider several cases.

- First, if $q = 1$ and $p = \infty$, we may take $f(x) = \text{sign } g(x)$. Then, we have $\|f\|_{L^\infty} = 1$, and clearly, $\int fg = \|g\|_{L^1}$.
- If $1 < p, q < \infty$, then we set $f(x) = |g(x)|^{q-1} \text{sign } g(x) / \|g\|_{L^q}^{q-1}$. We observe that $\|f\|_{L^p}^p = \int |g(x)|^{p(q-1)} d\mu / \|g\|_{L^q}^{p(q-1)} = 1$ since $p(q-1) = q$, and that $\int fg = \|g\|_{L^q}$.
- Finally, if $q = \infty$ and $p = 1$, let $\epsilon > 0$, and E a set of finite positive measure, where $|g(x)| \geq \|g\|_{L^\infty} - \epsilon$. (Such a set exists by the definition of $\|g\|_{L^\infty}$ and the fact that the measure μ is σ -finite.) Then, if we take $f(x) = \chi_E(x) \text{sign } g(x) / \mu(E)$, where χ_E denotes the characteristic function of the set E , we see that $\|f\|_{L^1} = 1$, and also

$$\left| \int fg \right| = \frac{1}{\mu(E)} \int_E |g| \geq \|g\|_{L^\infty} - \epsilon.$$

This completes the proof of part (i).

To prove (ii) we recall³ that we can find a sequence $\{g_n\}$ of simple functions so that $|g_n(x)| \leq |g(x)|$ while $g_n(x) \rightarrow g(x)$ for each x . When $p > 1$ (so $q < \infty$), we take $f_n(x) = |g_n(x)|^{q-1} \text{sign } g(x) / \|g_n\|_{L^q}^{q-1}$. As before, $\|f_n\|_{L^p} = 1$. However

$$\int f_n g = \frac{\int |g_n(x)|^q}{\|g_n\|_{L^q}^{q-1}} = \|g_n\|_{L^q},$$

and this does not exceed M . By Fatou's lemma it follows that $\int |g|^q \leq M^q$, so $g \in L^q$ with $\|g\|_{L^q} \leq M$. The direction $\|g\|_{L^q} \geq M$ is of course implied by Hölder's inequality.

When $p = 1$ the argument is parallel with the above but simpler. Here we take $f_n(x) = (\text{sign } g(x)) \chi_{E_n}(x)$, where E_n is an increasing sequence of sets of finite measure whose union is X . The details may be left to the reader.

With the lemma established we turn to the proof of the theorem. It is simpler to consider first the case when the underlying space has finite measure. In this case, with ℓ the given functional on L^p , we can then define a set function ν by

$$\nu(E) = \ell(\chi_E),$$

where E is any measurable set. This definition makes sense because χ_E is now automatically in L^p since the space has finite measure. We observe that

$$(6) \quad |\nu(E)| \leq c(\mu(E))^{1/p},$$

where c is the norm of the linear functional, taking into account the fact that $\|\chi_E\|_{L^p} = (\mu(E))^{1/p}$.

Now the linearity of ℓ clearly implies that ν is finitely-additive. Moreover, if $\{E_n\}$ is a countable collection of disjoint measurable sets, and we put $E = \bigcup_{n=1}^{\infty} E_n$, $E_N^* = \bigcup_{n=N+1}^{\infty} E_n$, then obviously

$$\chi_E = \chi_{E_N^*} + \sum_{n=1}^N \chi_{E_n}.$$

Thus $\nu(E) = \nu(E_N^*) + \sum_{n=1}^N \nu(E_n)$. However $\nu(E_N^*) \rightarrow 0$, as $N \rightarrow \infty$, because of (6) and the assumption $p < \infty$. This shows that ν is countably

³See for instance Section 2 in Chapter 6 of Book III.

additive and, moreover, (6) also shows us that ν is absolutely continuous with respect to μ .

We can now invoke the key result about absolutely continuous measures, the Lebesgue-Radon-Nykodim theorem. (See for example Theorem 4.3, Chapter 6 in Book III.) It guarantees the existence of an integrable function g so that $\nu(E) = \int_E g d\mu$ for every measurable set E . Thus we have $\ell(\chi_E) = \int \chi_E g d\mu$. The representation $\ell(f) = \int fg d\mu$ then extends immediately to simple functions f , and by a passage to the limit, to all $f \in L^p$ since the simple functions are dense in L^p , $1 \leq p < \infty$. (See Exercise 6.) Also by Lemma 4.2, we see that $\|g\|_{L^q} = \|\ell\|$.

To pass from the situation where the measure of X is finite to the general case, we use an increasing sequence $\{E_n\}$ of sets of finite measure that exhaust X , that is, $X = \bigcup_{n=1}^{\infty} E_n$. According to what we have just proved, for each n there is an integrable function g_n on E_n (which we can set to be zero in E_n^c) so that

$$(7) \quad \ell(f) = \int fg_n d\mu$$

whenever f is supported in E_n and $f \in L^p$. Moreover by conclusion (ii) of the lemma $\|g_n\|_{L^q} \leq \|\ell\|$.

Now it is easy to see because of (7) that $g_n = g_m$ a.e. on E_m , whenever $n \geq m$. Thus $\lim_{n \rightarrow \infty} g_n(x) = g(x)$ exists for almost every x , and by Fatou's lemma, $\|g\|_{L^q} \leq \|\ell\|$. As a result we have that $\ell(f) = \int fg d\mu$ for each $f \in L^p$ supported in E_n , and then by a simple limiting argument, for all $f \in L^p$. The fact that $\|\ell\| \leq \|g\|_{L^q}$, is already contained in Hölder's inequality, and therefore the proof of the theorem is complete.

5 More about linear functionals

First we turn to the study of certain geometric aspects of linear functionals in terms of the hyperplanes that they define. This will also involve understanding some elementary ideas about convexity.

5.1 Separation of convex sets

Although our ultimate focus will be on Banach spaces, we begin by considering an arbitrary vector space V over the reals. In this general setting we can define the following notions.

First, a **proper hyperplane** is a linear subspace of V that arises as the zero set of a (non-zero) linear functional on V . Alternatively, it is a linear subspace of V so that it, together with any vector not in V ,

spans V . Related to this notion is that of an **affine hyperplane** (which for brevity we will always refer to as a **hyperplane**) defined to be a translate of a proper hyperplane by a vector in V . To put it another way: H is a hyperplane if there is a non-zero linear functional ℓ , and a real number a , so that

$$H = \{v \in V : \ell(v) = a\}.$$

Another relevant notion is that of a convex set. The subset $K \subset V$ is said to be **convex** if whenever v_0 and v_1 are both in K then the straight-line segment joining them

$$(8) \quad v(t) = (1 - t)v_0 + tv_1, \quad 0 \leq t \leq 1$$

also lies entirely in K .

A key heuristic idea underlying our considerations can be enunciated as the following general principle:

If K is a convex set and $v_0 \notin K$, then K and v_0 can be separated by a hyperplane.

This principle is illustrated in Figure 1.

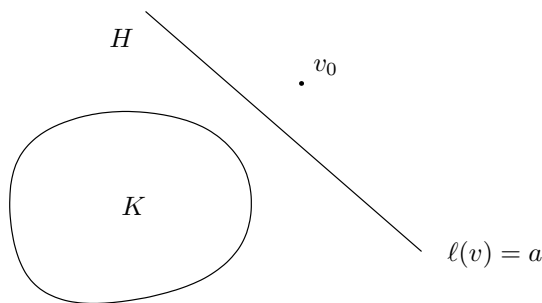


Figure 1. Separation of a convex set and a point by a hyperplane

The sense in which this is meant is that there is a non-zero linear functional ℓ and a real number a , so that

$$\ell(v_0) \geq a, \quad \text{while} \quad \ell(v) < a \text{ if } v \in K.$$

To give an idea of what is behind this principle we show why it holds in a nice special case. (See also Section 5.2.)

Proposition 5.1 *The assertion above is valid if $V = \mathbb{R}^d$ and K is convex and open.*

Proof. Since we may assume that K is non-empty, we can also suppose that (after a possible translation of K and v_0) we have $0 \in K$. The key construct used will be that of the Minkowski **gauge function** p associated to K , which measures (the inverse of) how far we need to go, starting from 0 in the direction of a vector v , to reach the exterior of K . The precise definition of p is as follows:

$$p(v) = \inf_{r>0} \{r : v/r \in K\}.$$

Observe that since we have assumed that the origin is an interior point of K , for each $v \in \mathbb{R}^d$ there is an $r > 0$, so that $v/r \in K$. Hence $p(v)$ is well-defined.

Figure 2 below gives an example of a gauge function in the special case where $V = \mathbb{R}$ and $K = (a, b)$, an open interval that contains the origin.

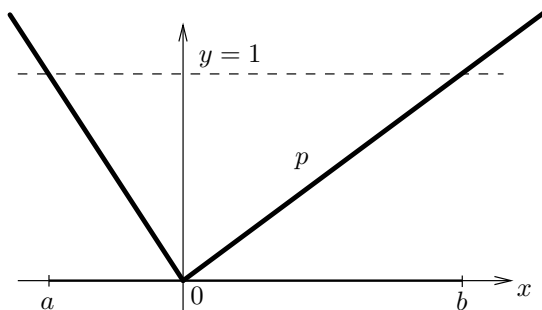


Figure 2. The gauge function of the interval (a, b) in \mathbb{R}

We note, for example, that if V is normed and K is the unit ball $\{\|v\| < 1\}$, then $p(v) = \|v\|$.

In general, the non-negative function p completely characterizes K in that

$$(9) \quad p(v) < 1 \quad \text{if and only if } v \in K.$$

Moreover p has an important sub-linear property:

$$(10) \quad \begin{cases} p(av) = ap(v), & \text{if } a \geq 0, \text{ and } v \in V. \\ p(v_1 + v_2) \leq p(v_1) + p(v_2), & \text{if } v_1 \text{ and } v_2 \in V. \end{cases}$$

In fact, if $v \in K$ then $v/(1 - \epsilon) \in K$ for some $\epsilon > 0$, since K is open, which gives that $p(v) < 1$. Conversely if $p(v) < 1$, then $v = (1 - \epsilon)v'$, for some $0 < \epsilon < 1$, and $v' \in K$. Then since $v = (1 - \epsilon)v' + \epsilon \cdot 0$ this shows $v \in K$, because $0 \in K$ and K is convex.

To verify (10) we merely note that $(v_1 + v_2)/(r_1 + r_2)$ belongs to K , if both v_1/r_1 and v_2/r_2 belong to K , in view of property (8) defining the convexity of K with $t = r_2/(r_1 + r_2)$ and $1 - t = r_1/(r_1 + r_2)$.

Now our proposition will be proved once we find a linear functional ℓ , so that

$$(11) \quad \ell(v_0) = 1, \quad \text{and} \quad \ell(v) \leq p(v), \quad v \in \mathbb{R}^d.$$

This is because $\ell(v) < 1$, for all $v \in K$ by (9). We shall construct ℓ in a step-by-step manner.

First, such an ℓ is already determined in the one-dimensional subspace V_0 spanned by v_0 , $V_0 = \{\mathbb{R}v_0\}$, since $\ell(bv_0) = b\ell(v_0) = b$, when $b \in \mathbb{R}$, and this is consistent with (11). Indeed, if $b \geq 0$ then $p(bv_0) = bp(v_0) \geq b\ell(v_0) = \ell(bv_0)$ by (10) and (9), while (11) is immediate when $b < 0$.

The next step is to choose any vector v_1 linearly independent from v_0 and extend ℓ to the subspace V_1 spanned by v_0 and v_1 . Thus we can make a choice for the value of ℓ on v_1 , $\ell(v_1)$, so as to satisfy (11) if

$$a\ell(v_1) + b = \ell(av_1 + bv_0) \leq p(av_1 + bv_0), \quad \text{for all } a, b \in \mathbb{R}.$$

Setting $a = 1$ and $bv_0 = w$ yields

$$\ell(v_1) + \ell(w) \leq p(v_1 + w) \quad \text{for all } w \in V_0,$$

while setting $a = -1$ implies

$$-\ell(v_1) + \ell(w') \leq p(-v_1 + w'), \quad \text{for all } w' \in V_0.$$

Altogether then it is required that for all $w, w' \in V_0$

$$(12) \quad -p(-v_1 + w') + \ell(w') \leq \ell(v_1) \leq p(v_1 + w) - \ell(w).$$

Notice that there is a number that lies between the two extremes of the above inequality. This is a consequence of the fact that $-p(-v_1 + w') + \ell(w')$ never exceeds $p(v_1 + w) - \ell(w)$, which itself follows from the fact that $\ell(w) + \ell(w') \leq p(w + w') \leq p(-v_1 + w') + p(v_1 + w)$, by (11) on V_0 and the sub-linearity of p . So a choice of $\ell(v_1)$ can be made that is

consistent with (12) and this allows one to extend ℓ to V_1 . In the same way we can proceed inductively to extend ℓ to all of \mathbb{R}^d .

The argument just given here in this special context will now be carried over in a general setting to give us an important theorem about constructing linear functionals.

5.2 The Hahn-Banach Theorem

We return to the general situation where we deal with an arbitrary vector space V over the reals. We assume that with V we are given a real-valued function p on V that satisfies the sub-linear property (10). However, as opposed to the example of the gauge function considered above, which by its nature is non-negative, here we do not assume that p has this property. In fact, certain p 's which may take on negative values are needed in some of our applications later.

Theorem 5.2 *Suppose V_0 is a linear subspace of V , and that we are given a linear functional ℓ_0 on V_0 that satisfies*

$$\ell_0(v) \leq p(v), \quad \text{for all } v \in V_0.$$

Then ℓ_0 can be extended to a linear functional ℓ on V that satisfies

$$\ell(v) \leq p(v), \quad \text{for all } v \in V.$$

Proof. Suppose $V_0 \neq V$, and pick v_1 a vector not in V_0 . We will first extend ℓ_0 to the subspace V_1 spanned by V_0 and v_1 , as we did before. We can do this by defining a putative extension ℓ_1 of ℓ_0 , defined on V_1 by $\ell_1(\alpha v_1 + w) = \alpha \ell_1(v_1) + \ell_0(w)$, whenever $w \in V_0$ and $\alpha \in \mathbb{R}$, if $\ell_1(v_1)$ is chosen so that

$$\ell_1(v) \leq p(v), \quad \text{for all } v \in V_1.$$

However, exactly as above, this happens when

$$-p(-v_1 + w') + \ell_0(w') \leq \ell_1(v_1) \leq p(v_1 + w) - \ell_0(w)$$

for all $w, w' \in V_0$.

The right-hand side exceeds the left-hand side because of $\ell_0(w') + \ell_0(w) \leq p(w' + w)$ and the sub-linearity of p . Thus an appropriate choice of $\ell_1(v_1)$ is possible, giving the desired extension of ℓ_0 from V_0 to V_1 .

We can think of the extension we have constructed as the key step in an inductive procedure. This induction, which in general is necessarily

trans-finite, proceeds as follows. We well-order all vectors in V that do not belong to V_0 , and denote this ordering by $<$. Among these vectors we call a vector v “extendable” if the linear functional ℓ_0 has an extension of the kind desired to the subspace spanned by V_0 , v , and all vectors $< v$. What we want to prove is in effect that all vectors not in V_0 are extendable. Assume the contrary, then because of the well-ordering we can find the smallest v_1 that is not extendable. Now if V'_0 is the space spanned by V_0 and all the vectors $< v_1$, then by assumption ℓ_0 extends to V'_0 . The previous step, with V'_0 in place of V_0 allows us then to extend ℓ_0 to the subspace spanned by V'_0 and v_1 , reaching a contradiction. This proves the theorem.

5.3 Some consequences

The Hahn-Banach theorem has several direct consequences for Banach spaces. Here \mathcal{B}^* denotes the dual of the Banach space \mathcal{B} as defined in Section 3.2, that is, the space of continuous linear functionals on \mathcal{B} .

Proposition 5.3 *Suppose f_0 is a given element of \mathcal{B} with $\|f_0\| = M$. Then there exists a continuous linear functional ℓ on \mathcal{B} so that $\ell(f_0) = M$ and $\|\ell\|_{\mathcal{B}^*} = 1$.*

Proof. Define ℓ_0 on the one-dimensional subspace $\{\alpha f_0\}_{\alpha \in \mathbb{R}}$ by $\ell_0(\alpha f_0) = \alpha M$, for each $\alpha \in \mathbb{R}$. Note that if we set $p(f) = \|f\|$ for every $f \in \mathcal{B}$, the function p satisfies the basic sub-linear property (10). We also observe that

$$|\ell_0(\alpha f_0)| = |\alpha| M = |\alpha| \|f_0\| = p(\alpha f_0),$$

so $\ell_0(f) \leq p(f)$ on this subspace. By the extension theorem ℓ_0 extends to an ℓ defined on \mathcal{B} with $\ell(f) \leq p(f) = \|f\|$, for all $f \in \mathcal{B}$. Since this inequality also holds for $-f$ in place of f we get $|\ell(f)| \leq \|f\|$, and thus $\|\ell\|_{\mathcal{B}^*} \leq 1$. The fact that $\|\ell\|_{\mathcal{B}^*} \geq 1$ is implied by the defining property $\ell(f_0) = \|f_0\|$, thereby proving the proposition.

Another application is to the duality of linear transformations. Suppose \mathcal{B}_1 and \mathcal{B}_2 are a pair of Banach spaces, and T is a bounded linear transformation from \mathcal{B}_1 to \mathcal{B}_2 . By this we mean that T maps \mathcal{B}_1 to \mathcal{B}_2 ; it satisfies $T(\alpha f_1 + \beta f_2) = \alpha T(f_1) + \beta T(f_2)$ whenever $f_1, f_2 \in \mathcal{B}_1$ and α and β are real numbers; and that it has a bound M so that $\|T(f)\|_{\mathcal{B}_2} \leq M \|f\|_{\mathcal{B}_1}$ for all $f \in \mathcal{B}_1$. The least M for which this inequality holds is called the **norm** of T and is denoted by $\|T\|$.

Often a linear transformation is initially given on a dense subspace. In this connection, the following proposition is very useful.

Proposition 5.4 *Let $\mathcal{B}_1, \mathcal{B}_2$ be a pair of Banach spaces and $\mathcal{S} \subset \mathcal{B}_1$ a dense linear subspace of \mathcal{B}_1 . Suppose T_0 is a linear transformation from \mathcal{S} to \mathcal{B}_2 that satisfies $\|T_0(f)\|_{\mathcal{B}_2} \leq M\|f\|_{\mathcal{B}_1}$ for all $f \in \mathcal{S}$. Then T_0 has a unique extension T to all of \mathcal{B}_1 so that $\|T(f)\|_{\mathcal{B}_2} \leq M\|f\|_{\mathcal{B}_1}$ for all $f \in \mathcal{B}_1$.*

Proof. If $f \in \mathcal{B}_1$, let $\{f_n\}$ be a sequence in \mathcal{S} which converges to f . Then since $\|T_0(f_n) - T_0(f_m)\|_{\mathcal{B}_2} \leq M\|f_n - f_m\|_{\mathcal{B}_1}$ it follows that $\{T_0(f_n)\}$ is a Cauchy sequence in \mathcal{B}_2 , and hence converges to a limit, which we define to be $T(f)$. Note that the definition of $T(f)$ is independent of the chosen sequence $\{f_n\}$, and that the resulting transformation T has all the required properties.

We now discuss duality of linear transformations. Whenever we have a linear transformation T from a Banach space \mathcal{B}_1 to another Banach space \mathcal{B}_2 , it induces a **dual transformation**, T^* of \mathcal{B}_2^* to \mathcal{B}_1^* , that can be defined as follows.

Suppose $\ell_2 \in \mathcal{B}_2^*$, (a continuous linear functional on \mathcal{B}_2), then $\ell_1 = T^*(\ell_2) \in \mathcal{B}_1^*$, is defined by $\ell_1(f_1) = \ell_2(T(f_1))$, whenever $f_1 \in \mathcal{B}_1$. More succinctly

$$(13) \quad T^*(\ell_2)(f_1) = \ell_2(T(f_1)).$$

Theorem 5.5 *The operator T^* defined by (13) is a bounded linear transformation from \mathcal{B}_2^* to \mathcal{B}_1^* . Its norm $\|T^*\|$ satisfies $\|T\| = \|T^*\|$.*

Proof. First, if $\|f_1\|_{\mathcal{B}_1} \leq 1$, we have that

$$|\ell_1(f_1)| = |\ell_2(T(f_1))| \leq \|\ell_2\| \|T(f_1)\|_{\mathcal{B}_2} \leq \|\ell_2\| \|T\|.$$

Thus taking the supremum over all $f_1 \in \mathcal{B}_1$ with $\|f_1\|_{\mathcal{B}_1} \leq 1$, we see that the mapping $\ell_2 \mapsto T^*(\ell_2) = \ell_1$ has norm $\leq \|T\|$.

To prove the reverse inequality we can find for any $\epsilon > 0$ an $f_1 \in \mathcal{B}_1$ with $\|f_1\|_{\mathcal{B}_1} = 1$ and $\|T(f_1)\|_{\mathcal{B}_2} \geq \|T\| - \epsilon$. Next, with $f_2 = T(f_1) \in \mathcal{B}_2$, by Proposition 5.3 (with $\mathcal{B} = \mathcal{B}_2$) there is an ℓ_2 in \mathcal{B}_2^* so that $\|\ell_2\|_{\mathcal{B}_2^*} = 1$ but $\ell_2(f_2) \geq \|T\| - \epsilon$. Thus by (13) one has $T^*(\ell_2)(f_1) \geq \|T\| - \epsilon$, and since $\|f_1\|_{\mathcal{B}_1} = 1$, we conclude $\|T^*(\ell_2)\|_{\mathcal{B}_1^*} \geq \|T\| - \epsilon$. This gives $\|T^*\| \geq \|T\| - \epsilon$ for any $\epsilon > 0$, which proves the theorem.

A further quick application of the Hahn-Banach theorem is the observation that in general L^1 is *not* the dual of L^∞ (as opposed to the case $1 \leq p < \infty$ considered in Theorem 4.1).

Let us first recall that whenever $g \in L^1$, the linear functional $f \mapsto \ell(f)$ given by

$$(14) \quad \ell(f) = \int fg \, d\mu$$

is bounded on L^∞ , and its norm $\|\ell\|_{(L^\infty)^*}$ is $\|g\|_{L^1}$. In this way L^1 can be viewed as a subspace of $(L^\infty)^*$, with the L^1 norm of g being identical with its norm as a linear functional. One can, however, produce a continuous linear functional of L^∞ not of this form. For simplicity we do this when the underlying space is \mathbb{R} with Lebesgue measure.

We let \mathcal{C} denote the subspace of $L^\infty(\mathbb{R})$ consisting of continuous bounded functions on \mathbb{R} . Define the linear function ℓ_0 on \mathcal{C} (the “Dirac delta”) by

$$\ell_0(f) = f(0), \quad f \in \mathcal{C}.$$

Clearly $|\ell_0(f)| \leq \|f\|_{L^\infty}$, $f \in \mathcal{C}$. Thus by the extension theorem, with $p(f) = \|f\|_{L^\infty}$, we see that there is a linear functional ℓ on L^∞ , extending ℓ_0 , that satisfies $|\ell(f)| \leq \|f\|_{L^\infty}$, for all $f \in L^\infty$.

Suppose for a moment that ℓ were of the form (14) for some $g \in L^1$. Since $\ell(f) = \ell_0(f) = 0$ whenever f is a continuous trapezoidal function that excludes the origin, we would have $\int fg \, dx = 0$ for such functions f ; by a simple limiting argument this gives $\int_I g \, dx = 0$ for all intervals excluding the origin, and from there for all intervals I . Hence the indefinite integrals $G(y) = \int_0^y g(x) \, dx$ vanish, and therefore $G' = g = 0$ by the differentiation theorem.⁴ This gives a contradiction, hence the linear functional ℓ is not representable as (14).

5.4 The problem of measure

We now consider an application of the Hahn-Banach theorem of a different kind. We present a rather stunning assertion, answering a basic question of the “problem of measure.” The result states that there is a finitely-additive⁵ measure defined on *all* subsets of \mathbb{R}^d that agrees with Lebesgue measure on the measurable sets, and is translation invariant. We formulate the theorem in one dimension.

Theorem 5.6 *There is an extended-valued non-negative function \hat{m} , defined on all subsets of \mathbb{R} with the following properties:*

- (i) $\hat{m}(E_1 \cup E_2) = \hat{m}(E_1) + \hat{m}(E_2)$ whenever E_1 and E_2 are disjoint subsets of \mathbb{R} .

⁴See for instance Theorem 3.11, in Chapter 3 of Book III.

⁵The qualifier “finitely-additive” is crucial.

(ii) $\hat{m}(E) = m(E)$ if E is a measurable set and m denotes the Lebesgue measure.

(iii) $\hat{m}(E + h) = \hat{m}(E)$ for every set E and real number h .

From (i) we see that \hat{m} is finitely additive; however it cannot be countably additive as the proof of the existence of non-measurable sets shows. (See Section 3, Chapter 1 in Book III.)

This theorem is a consequence of another result of this kind, dealing with an extension of the Lebesgue integral. Here the setting is the circle \mathbb{R}/\mathbb{Z} , instead of \mathbb{R} , with the former realized as $(0, 1]$. Thus functions on \mathbb{R}/\mathbb{Z} can be thought of as functions on $(0, 1]$, extended to \mathbb{R} by periodicity with period 1. In the same way, translations on \mathbb{R} induce corresponding translations on \mathbb{R}/\mathbb{Z} . The assertion now is the existence of a generalized integral (the “Banach integral”) defined on all bounded functions on the circle.

Theorem 5.7 *There is a linear functional $f \mapsto I(f)$ defined on all bounded functions f on \mathbb{R}/\mathbb{Z} so that:*

- (a) $I(f) \geq 0$, if $f(x) \geq 0$ for all x .
- (b) $I(\alpha f_1 + \beta f_2) = \alpha I(f_1) + \beta I(f_2)$ for all α and β real.
- (c) $I(f) = \int_0^1 f(x) dx$, whenever f is measurable.
- (d) $I(f_h) = I(f)$, for all $h \in \mathbb{R}$ where $f_h(x) = f(x - h)$.

The right-hand side of (c) denotes the usual Lebesgue integral.

Proof. The idea is to consider the vector space V of all (real-valued) bounded functions on \mathbb{R}/\mathbb{Z} , with V_0 the subspace of those functions that are measurable. We let I_0 denote the linear functional given by the Lebesgue integral, $I_0(f) = \int_0^1 f(x) dx$ for $f \in V_0$. The key is to find the appropriate sub-linear p defined on V so that

$$I_0(f) \leq p(f), \quad \text{for all } f \in V_0.$$

Banach’s ingenious definition of p is as follows: We let $A = \{a_1, \dots, a_N\}$ denote an arbitrary collection of N real numbers, with $\#(A) = N$ denoting its cardinality. Given A , we define $M_A(f)$ to be the real number

$$M_A(f) = \sup_{x \in \mathbb{R}} \left(\frac{1}{N} \sum_{j=1}^N f(x + a_j) \right),$$

and set

$$p(f) = \inf_A \{M_A(f)\},$$

where the infimum is taken over all finite collections A .

It is clear that $p(f)$ is well-defined, since f is assumed to be bounded; also $p(cf) = cp(f)$ if $c \geq 0$. To prove $p(f_1 + f_2) \leq p(f_1) + p(f_2)$, we find for each ϵ , finite collections A and B so that

$$M_A(f_1) \leq p(f_1) + \epsilon \quad \text{and} \quad M_B(f_2) \leq p(f_2) + \epsilon.$$

Let C be the collection $\{a_i + b_j\}_{1 \leq i \leq N_1, 1 \leq j \leq N_2}$ where $N_1 = \#(A)$, and $N_2 = \#(B)$. Now it is easy to see that

$$M_C(f_1 + f_2) \leq M_C(f_1) + M_C(f_2).$$

Next, we note as a general matter that $M_A(f)$ is the same as $M_{A'}(f')$ where $f' = f_h$ is a translate of f and $A' = A - h$. Also the averages corresponding to C arise as averages of translates of the averages corresponding to A and B , so it is easy to verify that

$$M_C(f_1) \leq M_A(f_1) \quad \text{and also} \quad M_C(f_2) \leq M_B(f_2).$$

Thus

$$p(f_1 + f_2) \leq M_C(f_1 + f_2) \leq M_A(f_1) + M_B(f_2) \leq p(f_1) + p(f_2) + 2\epsilon.$$

Letting $\epsilon \rightarrow 0$ proves the sub-linearity of p .

Next if f is Lebesgue measurable (and hence integrable since it is bounded), then for each A

$$I_0(f) = \frac{1}{N} \int_0^1 \left(\sum_{j=1}^N f(x + a_j) \right) dx \leq \int_0^1 M_A(f) dx = M_A(f),$$

and hence $I_0(f) \leq p(f)$. Let therefore I be the linear functional extending I_0 from V_0 to V , whose existence is guaranteed by Theorem 5.2. It is obvious from its definition that $p(f) \leq 0$ if $f \leq 0$. From this it follows that $I(f) \leq 0$ when $f \leq 0$, and replacing f by $-f$ we see that conclusion (a) holds.

Next we observe that for each real h

$$(15) \quad p(f - f_h) \leq 0.$$

In fact, for h fixed and N given, define the set A_N to be $\{h, 2h, 3h, \dots, Nh\}$. Then the sum that enters in the definition of $M_{A_N}(f - f_h)$ is

$$\frac{1}{N} \sum_{j=1}^N (f(x + jh) - f(x + (j-1)h)),$$

and thus $|M_{A_N}(f - f_h)| \leq 2M/N$, where M is an upper bound for $|f|$. Since $p(f - f_h) \leq M_{A_N}(f - f_h) \rightarrow 0$, as $N \rightarrow \infty$, we see that (15) is proved. This shows that $I(f - f_h) \leq 0$, for all f and h . However, replacing f by f_h and then h by $-h$, we see that $I(f_h - f) \leq 0$ and thus (d) is also established, finishing the proof of Theorem 5.7.

As a direct consequence we have the following.

Corollary 5.8 *There is a non-negative function \hat{m} defined on all subsets of \mathbb{R}/\mathbb{Z} so that:*

- (i) $\hat{m}(E_1 \cup E_2) = \hat{m}(E_1) + \hat{m}(E_2)$ for all disjoint subsets E_1 and E_2 .
- (ii) $\hat{m}(E) = m(E)$ if E is measurable.
- (iii) $\hat{m}(E + h) = \hat{m}(E)$ for every h in \mathbb{R} .

We need only take $\hat{m}(E) = I(\chi_E)$, with I as in Theorem 5.7, where χ_E denotes the characteristic function of E .

We now turn to the proof of Theorem 5.6. Let \mathcal{I}_j denote the interval $(j, j + 1]$, where $j \in \mathbb{Z}$. Then we have a partition $\bigcup_{j=-\infty}^{\infty} \mathcal{I}_j$ of \mathbb{R} into disjoint sets.

For clarity of exposition, we temporarily relabel the measure \hat{m} on $(0, 1] = \mathcal{I}_0$ given by the corollary and call it \hat{m}_0 . So whenever $E \subset \mathcal{I}_0$ we defined $\hat{m}(E)$ to be $\hat{m}_0(E)$. More generally, if $E \subset \mathcal{I}_j$ we set $\hat{m}(E) = \hat{m}_0(E - j)$.

With these things said, for any set E define $\hat{m}(E)$ by

$$(16) \quad \hat{m}(E) = \sum_{j=-\infty}^{\infty} \hat{m}(E \cap \mathcal{I}_j) = \sum_{j=-\infty}^{\infty} \hat{m}_0((E \cap \mathcal{I}_j) - j).$$

Thus $\hat{m}(E)$ is given as an extended non-negative number. Note that if E_1 and E_2 are disjoint so are $(E_1 \cap \mathcal{I}_j) - j$ and $(E_2 \cap \mathcal{I}_j) - j$. It follows that $\hat{m}(E_1 \cup E_2) = \hat{m}(E_1) + \hat{m}(E_2)$. Moreover if E is measurable then $\hat{m}(E \cap \mathcal{I}_j) = m(E \cap \mathcal{I}_j)$ and so $\hat{m}(E) = m(E)$.

To prove $\hat{m}(E + h) = \hat{m}(E)$, consider first the case $h = k \in \mathbb{Z}$. This is an immediate consequence of the definition (16) once one observes that $((E + k) \cap \mathcal{I}_{j+k}) - (j + k) = (E \cap \mathcal{I}_j) - j$, for all $j, k \in \mathbb{Z}$.

Next suppose $0 < h < 1$. We then decompose $E \cap \mathcal{I}_j$ as $E'_j \cup E''_j$, with $E'_j = E \cap (j, j + 1 - h]$ and $E''_j = E \cap (j + 1 - h, j + 1]$. The point of this decomposition is that $E'_j + h$ remains in \mathcal{I}_j but $E''_j + h$ is placed in \mathcal{I}_{j+1} . In any case, $E = \bigcup_j E'_j \cup \bigcup_j E''_j$, and the union is disjoint.

Thus using the first additivity property proved above and then (16) we see that

$$\hat{m}(E) = \sum_{j=-\infty}^{\infty} (\hat{m}(E'_j) + \hat{m}(E''_j)).$$

Similarly

$$\hat{m}(E + h) = \sum_{j=-\infty}^{\infty} (\hat{m}(E'_j + h) + \hat{m}(E''_j + h)).$$

Now both E'_j and $E'_j + h$ are in \mathcal{I}_j , hence $\hat{m}(E'_j) = \hat{m}(E'_j + h)$ by the translation invariance of \hat{m}_0 and the definition of \hat{m} on subsets of \mathcal{I}_j . Also E''_j is in \mathcal{I}_j and $E''_j + h$ is in \mathcal{I}_{j+1} , and their measures agree for the same reasons. This establishes that $\hat{m}(E) = \hat{m}(E + h)$, for $0 < h < 1$. Now combining this with the translation invariance with respect to \mathbb{Z} already proved, we obtain conclusion (iii) of Theorem 5.6 for all h , and hence the theorem is completely proved.

For the corresponding extension of Lebesgue measure in \mathbb{R}^d and other related results, see Exercise 36 and Problems 8* and 9*.

6 Complex L^p and Banach spaces

We have supposed in Section 3.2 onwards that our L^p and Banach spaces are taken over the reals. However, the statements and the proofs of the corresponding theorems for those spaces taken with respect to the complex scalars are for the most part routine adaptations of the real case. There are nevertheless several instances that require further comment. First, in the argument concerning the converse of Hölder's inequality (Lemma 4.2), the definition of f should read

$$f(x) = |g(x)|^{q-1} \frac{\overline{\text{sign } g(x)}}{\|g\|_{L^q}^{q-1}},$$

where now “sign” denotes the complex version of the signum function, defined by $\text{sign } z = z/|z|$ if $z \neq 0$, and $\text{sign } 0 = 0$. There are similar occurrences with g replaced by g_n .

Second, while the Hahn-Banach theorem is valid as stated only for real vector spaces, a version of the complex case (sufficient for the applications in Section 5.3 where $p(f) = \|f\|$) can be found in Exercise 33 below.

7 Appendix: The dual of $C(X)$

In this appendix, we describe the bounded linear functionals of the space $C(X)$ of continuous real-valued functions on X . To begin with, we assume that X is a compact metric space. Our main result then states that if $\ell \in C(X)^*$, then there exists a finite signed Borel measure μ (this measure is sometimes referred to as a Radon measure) so that

$$\ell(f) = \int_X f(x) d\mu(x) \quad \text{for all } f \in C(X).$$

Before proceeding with the argument leading to this result, we collect some basic facts and definitions.

Let X be a metric space with metric d , and assume that X is compact; that is, every covering of X by open sets contains a finite sub-covering. The vector space $C(X)$ of real-valued continuous functions on X equipped with the sup-norm

$$\|f\| = \sup_{x \in X} |f(x)|, \quad f \in C(X)$$

is a Banach space over \mathbb{R} . Given a continuous function f on X we define the **support** of f , denoted $\text{supp}(f)$, as the closure of the set $\{x \in X : f(x) \neq 0\}$.⁶

We recall some simple facts about continuous functions and open and closed sets in X that we shall use below.

(i) **Separation.** If A and B are two disjoint closed subsets of X , then there exists a continuous function f with $f = 1$ on A , $f = 0$ on B , and $0 < f < 1$ in the complements of A and B .

Indeed, one can take for instance

$$f(x) = \frac{d(x, B)}{d(x, A) + d(x, B)},$$

where $d(x, B) = \inf_{y \in B} d(x, y)$, with a similar definition for $d(x, A)$.

(ii) **Partition of unity.** If K is a compact set which is covered by finitely many open sets $\{\mathcal{O}_k\}_{k=1}^N$, then there exist continuous functions η_k for $1 \leq k \leq N$ so that $0 \leq \eta_k \leq 1$, $\text{supp}(\eta_k) \subset \mathcal{O}_k$, and $\sum_{k=1}^N \eta_k(x) = 1$ whenever $x \in K$. Moreover, $0 \leq \sum_{k=1}^N \eta_k(x) \leq 1$ for all $x \in X$.

One can argue as follows. For each $x \in K$, there exists a ball $B(x)$ centered at x and of positive radius such that $\overline{B(x)} \subset \mathcal{O}_i$ for some i . Since $\bigcup_{x \in K} B(x)$ covers K , we can select a finite subcovering, say $\bigcup_{j=1}^M B(x_j)$. For each $1 \leq k \leq N$, let U_k be the union of all open balls $B(x_j)$ so that $B(x_j) \subset \mathcal{O}_k$; clearly $K \subset \bigcup_{k=1}^N U_k$. By (i) above, there exists a continuous function $0 \leq \varphi_k \leq 1$ so that $\varphi_k = 1$ on $\overline{U_k}$ and $\text{supp}(\varphi_k) \subset \mathcal{O}_k$. If we define

$$\eta_1 = \varphi_1, \quad \eta_2 = \varphi_2(1 - \varphi_1), \quad \dots, \quad \eta_N = \varphi_N(1 - \varphi_1) \cdots (1 - \varphi_{N-1})$$

⁶This is the common usage of the terminology “support.” In Book III, Chapter 2, we used “support of f ” to indicate the set where $f(x) \neq 0$, which is convenient when dealing with measurable functions.

then $\text{supp}(\eta_k) \subset \mathcal{O}_k$ and

$$\eta_1 + \cdots + \eta_N = 1 - (1 - \varphi_1) \cdots (1 - \varphi_N),$$

thus guaranteeing the desired properties.

Recall⁷ that the **Borel σ -algebra** of X , which is denoted by \mathcal{B}_X , is the smallest σ -algebra of X that contains the open sets. Elements of \mathcal{B}_X are called **Borel sets**, and a measure defined on \mathcal{B}_X is called a **Borel measure**. If a Borel measure is finite, that is $\mu(X) < \infty$, then it satisfies the following “regularity property”: for any Borel set E and any $\epsilon > 0$, there are an open set \mathcal{O} and a closed set F such that $E \subset \mathcal{O}$ and $\mu(\mathcal{O} - E) < \epsilon$, while $F \subset E$ and $\mu(E - F) < \epsilon$.

In general we shall be interested in finite *signed* Borel measures on X , that is, measures which can take on negative values. If μ is such a measure, and μ^+ and μ^- denote the positive and negative variations of μ , then $\mu = \mu^+ - \mu^-$, and integration with respect to μ is defined by $\int f d\mu = \int f d\mu^+ - \int f d\mu^-$. Conversely, if μ_1 and μ_2 are two finite Borel measures, then $\mu = \mu_1 - \mu_2$ is a finite signed Borel measure, and $\int f d\mu = \int f d\mu_1 - \int f d\mu_2$.

We denote by $M(X)$ the space of finite signed Borel measures on X . Clearly, $M(X)$ is a vector space which can be equipped with the following norm

$$\|\mu\| = |\mu|(X),$$

where $|\mu|$ denotes the total variation of μ . It is a simple fact that $M(X)$ with this norm is a Banach space.

7.1 The case of positive linear functionals

We begin by considering only linear functionals $\ell : C(X) \rightarrow \mathbb{R}$ which are **positive**, that is, $\ell(f) \geq 0$ whenever $f(x) \geq 0$ for all $x \in X$. Observe that positive linear functionals are automatically bounded and that $\|\ell\| = \ell(1)$. Indeed, note that $|f(x)| \leq \|f\|$, hence $\|f\| \pm f \geq 0$, and therefore $|\ell(f)| \leq \ell(1)\|f\|$.

Our main result goes as follows.

Theorem 7.1 *Suppose X is a compact metric space and ℓ a positive linear functional on $C(X)$. Then there exists a unique finite (positive) Borel measure μ so that*

$$(17) \quad \ell(f) = \int_X f(x) d\mu(x) \quad \text{for all } f \in C(X).$$

Proof. The existence of the measure μ is proved as follows. Consider the function ρ on the open subsets of X defined by

$$\rho(\mathcal{O}) = \sup \{ \ell(f), \text{ where } \text{supp}(f) \subset \mathcal{O}, \text{ and } 0 \leq f \leq 1 \},$$

⁷The definitions and results on measure theory needed in this section, in particular the extension of a premeasure used in the proof of Theorem 7.1, can be found in Chapter 6 of Book III.

and let the function μ_* be defined on all subsets of X by

$$\mu_*(E) = \inf\{\rho(\mathcal{O}), \text{ where } E \subset \mathcal{O} \text{ and } \mathcal{O} \text{ is open}\}.$$

We contend that μ_* is a metric exterior measure on X .

Indeed, we clearly must have $\mu_*(E_1) \leq \mu_*(E_2)$ whenever $E_1 \subset E_2$. Also, if \mathcal{O} is open, then $\mu_*(\mathcal{O}) = \rho(\mathcal{O})$. To show that μ_* is countably sub-additive on subsets of X , we begin by proving that μ_* is in fact sub-additive on open sets $\{\mathcal{O}_k\}$, that is,

$$(18) \quad \mu_* \left(\bigcup_{k=1}^{\infty} \mathcal{O}_k \right) \leq \sum_{k=1}^{\infty} \mu_*(\mathcal{O}_k).$$

To do so, suppose $\{\mathcal{O}_k\}_{k=1}^{\infty}$ is a collection of open sets in X , and let $\mathcal{O} = \bigcup_{k=1}^{\infty} \mathcal{O}_k$. If f is any continuous function that satisfies $\text{supp}(f) \subset \mathcal{O}$ and $0 \leq f \leq 1$, then by compactness of $K = \text{supp}(f)$ we can pick a sub-cover so that (after relabeling the sets \mathcal{O}_k , if necessary) $K \subset \bigcup_{k=1}^N \mathcal{O}_k$. Let $\{\eta_k\}_{k=1}^N$ be a partition of unity of $\{\mathcal{O}_1, \dots, \mathcal{O}_N\}$ (as discussed above in (ii)); this means that each η_k is continuous with $0 \leq \eta_k \leq 1$, $\text{supp}(\eta_k) \subset \mathcal{O}_k$ and $\sum_{k=1}^N \eta_k(x) = 1$ for all $x \in K$. Hence recalling that $\mu_* = \rho$ on open sets, we get

$$\ell(f) = \sum_{k=1}^N \ell(f\eta_k) \leq \sum_{k=1}^N \mu_*(\mathcal{O}_k) \leq \sum_{k=1}^{\infty} \mu_*(\mathcal{O}_k),$$

where the first inequality follows because $\text{supp}(f\eta_k) \subset \mathcal{O}_k$ and $0 \leq f\eta_k \leq 1$. Taking the supremum over f we find that $\mu_* \left(\bigcup_{k=1}^{\infty} \mathcal{O}_k \right) \leq \sum_{k=1}^{\infty} \mu_*(\mathcal{O}_k)$.

We now turn to the proof of the sub-additivity of μ_* on all sets. Suppose $\{E_k\}$ is a collection of subsets of X and let $\epsilon > 0$. For each k , pick an open set \mathcal{O}_k so that $E_k \subset \mathcal{O}_k$ and $\mu_*(\mathcal{O}_k) \leq \mu_*(E_k) + \epsilon 2^{-k}$. Since $\mathcal{O} = \bigcup \mathcal{O}_k$ covers $\bigcup E_k$, we must have by (18) that

$$\mu_* \left(\bigcup E_k \right) \leq \mu_*(\mathcal{O}) \leq \sum_k \mu_*(\mathcal{O}_k) \leq \sum_k \mu_*(E_k) + \epsilon,$$

and consequently $\mu_* \left(\bigcup E_k \right) \leq \sum_k \mu_*(E_k)$ as desired.

The last property we must verify is that μ_* is metric, in the sense that if $d(E_1, E_2) > 0$, then $\mu_*(E_1 \cup E_2) = \mu_*(E_1) + \mu_*(E_2)$. Indeed, the separation condition implies that there exist disjoint open sets \mathcal{O}_1 and \mathcal{O}_2 so that $E_1 \subset \mathcal{O}_1$ and $E_2 \subset \mathcal{O}_2$. Therefore, if \mathcal{O} is any open subset which contains $E_1 \cup E_2$, then $\mathcal{O} \supset (\mathcal{O} \cap \mathcal{O}_1) \cup (\mathcal{O} \cap \mathcal{O}_2)$, where this union is disjoint. Hence the additivity of μ_* on disjoint open sets, and its monotonicity give

$$\mu_*(\mathcal{O}) \geq \mu_*(\mathcal{O} \cap \mathcal{O}_1) + \mu_*(\mathcal{O} \cap \mathcal{O}_2) \geq \mu_*(E_1) + \mu_*(E_2),$$

since $E_1 \subset (\mathcal{O} \cap \mathcal{O}_1)$ and $E_2 \subset (\mathcal{O} \cap \mathcal{O}_2)$. So $\mu_*(E_1 \cup E_2) \geq \mu_*(E_1) + \mu_*(E_2)$, and since the reverse inequality has already been shown above, this concludes the proof that μ_* is a metric exterior measure.

(continued...)

Index

Relevant items that also arose in Book I, Book II or Book III are listed in this index, preceded by the numerals I, II or III, respectively.

- $C^{(n)}$ -normalized bump function, 135
- L^p norm, 2
- δ -neighborhood, 177
- O notation, 64; (III)12

- affine hyperplane, 17
- algebra, 209
 - tail, 215
- allied series, 50
- almost surely, 192
- almost-orthogonality, 372
- amplitude, 325; (I)3; (II)323
- analytic family of operators, 338
- analytic identity, 279
- approximation to the identity, 64;
 - (I)49; (III)109
- atomic decomposition, 74
- atoms, 74
 - 1-atom, 138
 - p -atoms, 81
 - “faux”, 93
- averaging operator, 322, 323, 366

- Banach integral, 24
- Banach space, 9
 - equivalent, 46
- Banach-Tarski paradox, 46
- Baouendi-Treves approximation theorem, 300
- Bernoulli trials, 205
- Besicovitch set, 176; (III)360, 362, 374
- bijjective mapping, 171
- Blumenthal’s zero-one law, 257
- BMO, 86
- Bochner’s theorem, 292
- Bochner-Martinelli integral, 319
- Borel
 - σ -algebra, 29; (III)23, 267
 - measure, 29, 242; (III)269
 - sets, 29, 242; (III)23, 267
- Borel-Cantelli lemma, 231; (III)42, 63
- bounded mean oscillation, 86
- Brownian motion, 227, 240
 - recurrent, 274
 - transient, 274

- Calderón-Zygmund
 - decomposition, 76
 - distributions, 135
- cancelation condition, 135
- category
 - first, 158
 - second, 158
- Cauchy integral
 - representation, 277
 - upper half-space, 312
- Cauchy-Riemann
 - equations, 277; (II)12
 - tangential weak sense, 300
 - operator, 148
 - vector field, 290
 - tangential, 291
- Cauchy-Szegö integral, 311
- central limit theorem, 195, 220
- characteristic
 - function, 216, 221; (III)27
 - polynomial, 126; (III)221, 258
- Clarkson inequalities, 45
- class C^k ; (I)44
 - function, 101, 290
 - hypersurface, 288
- closed linear map, 174
- closure of a set, 158
- complete normed vector space, 5
- conditional expectation, 209
- cone, 267

- backward, 155
- forward, 155
- outside condition, 268
- conjugate
 - exponents, 3
 - function, 50
 - Poisson kernel, 63; (III)255
- convergence in probability, 195
- convex set, 17, 382; (III)35
- convolution; (I)44, 139, 239; (III)74, 94, 253
 - distributions, 102
 - functions, 38, 60
- covariance matrix, 221
- critical point, 327; (II)326
- curvature
 - form, 333
 - Gauss, 333
 - principal, 333
 - rotational, 366
 - total, 333
- cylinder set, 191; (III)316
- cylindrical set, 242

- defining function, 288
- dense set, 158
- differential form, 291
- Dirac delta function, 23, 101; (III)110, 285
- Dirichlet
 - kernel, 90; (I)37; (III)179
 - problem, 264; (I)20, 28, 64, 170; (II)212, 215, 216; (III)230
- dispersion equations, 348
 - non-linear, 359
- distance
 - Hausdorff, 177; (III)345
- distribution, 99, 100
 - convolution, 102
 - derivative, 101
 - finite order, 150
 - function, 72
 - fundamental solution, 125
 - Gaussian, 196
 - homogeneous, 115
 - joint, 206
 - measure, 195, 220
 - normal, 196
 - periodic, 153
 - positive, 150
 - principal value, 111
 - regular, 117, 132
 - support, 104
 - tempered, 106
 - weak sense convergence, 103
- domain of holomorphy, 320
- Donsker invariance principle, 250
- dual
 - exponents, 3
 - space, 12
 - transformation, 22
- dyadic intervals, 199

- elliptic differential operator, 132
- equivalence, 41
- equivalent Banach spaces, 46
- ergodic, 208; (I)111; (III)294
- error function, 229
- essential-supremum, 8
- event, 192
- expectation conditional, 209
- exponential type, 151; (II)112

- Fourier coefficients, 48; (I)16, 34; (III)170
- Fourier series; (I)34; (II)101; (III)171
 - conjugate function, 50
 - decay of coefficients, 173
 - diverging at a point, 167
 - periodic distributions, 153
 - random, 202
- Fourier transform; (I)134, 136, 181; (II)111
 - surface-carried measure, 334
 - tempered distribution, 108
- fractional derivative, 375
- function
 - analytic in C^n , 276
 - class C^k , 101
 - convolution, 38, 60
 - Dirac delta, 101; (III)110
 - expectation, 196
 - gauge, 18
 - holomorphic in C^n , 276
 - homogeneous, 115
 - mean, 196
 - measurable, 209; (III)28

- mutual independence, 193
- nowhere differentiable, 163, 253; (III)154, 383
- Rademacher, 192
- slowly increasing, 107
- support, 28, 104, 146; (III)53
- variance, 196
- Walsh-Paley, 230
- zig-zag, 165
- fundamental solution, 125

- gauge function, 18
- Gauss
 - curvature, 333
 - map, 405
- Gaussian; (I)135, 181; (III)88
 - distribution, 196
 - subspace, 228
- generalized function, 99
- generic set, 158
- graph of a linear map, 174

- Hahn-Banach Theorem, 20, 43
- Hamel basis, 183
- Hardy space, 73, 308; (III)174, 203, 213
- harmonic measure, 254
- Hartog's phenomenon, 280
- Hausdorff distance, 177; (III)345
- Hausdorff-Young inequality, 49, 57, 90
- heat
 - kernel, 128; (I)120, 146, 209; (III)111
 - operator, 127, 133
- Heaviside function, 101; (III)285
- Heisenberg group, 318
- Hessian matrix, 329
- Hilbert transform, 62; (III)220, 255
- Hölder
 - condition, 10; (I)43
 - inequality, 3, 35, 38, 39
 - converse, 14
- holomorphic coordinates, 294
- Huygens' principle, 156; (I)193
- hyperbolic measure, 385
- hyperplane, 17
 - affine, 17
 - proper, 16
- hypersurface, 288
 - class C^k , 288
 - hypo-elliptic, 133
- identically distributed functions, 205
- injective mapping, 171
- interior of a set, 158; (III)3
- invariance principle (Donsker), 250
- invariant set, 207; (III)302
- iterated logarithm, 237, 275

- Jensen's inequality, 40
- John-Nirenberg inequalities, 95
- joint distribution, 206

- Khinchin's inequality, 203

- Laplacian, 119, 126; (I)20, 149, 185; (II)27; (III)230
- lattice points, 377, 379
- law of large numbers, 213
- law of the iterated logarithm, 237, 275
- Lebesgue's thorn, 275
- Levi form, 295
- Lewy
 - example, 313
 - extension theorem, 306
- linear functional, 11; (III)181
 - bounded, 11
 - continuous, 11
- linear transformation
 - bounded, 21
- Liouville numbers, 185
- Lipschitz
 - boundary, 272
 - condition, 10, 146; (I)82; (III)90, 147, 151, 330, 362
- martingale sequence, 211
 - complete, 211
- maximal function, 70, 76, 85; (III)100, 261
 - spherical, 406
- maximum principle, 296; (I)92; (III)235
- meager set, 158
- mean, 196
- measurable, 209

- measure
 - Borel, 29, 242
 - continuous, 218
 - harmonic, 254
 - hyperbolic, 385
 - Radon, 28, 100
- Minkowski inequality, 4
 - for integrals, 37
- mixed norm, 38
- mixing, 207; (III)305
- multiplier, 134; (III)220
- mutual independence, 193
 - function, 193
 - sub-algebras, 211
- non-linear dispersion equation, 359
- norm, 9, 21
 - of a continuous linear functional, 12
- normal
 - distribution, 196
 - number, 231; (III)318
- normed vector space, 3, 9
- nowhere dense set, 158
- nowhere differentiable function, 163, 253; (I)113, 126; (III)154, 383
- open mapping, 171
- open mapping theorem, 171; (II)92
- Orlicz space, 41, 45
- oscillation of a function, 161; (I)288
- outside cone condition, 268
- parallelogram law, 41, 45; (III)176
- parametrix, 131
- partition of unity, 28
- path, 223
- periodization operator, 153
- phase, 325; (I)3; (II)323
- Poisson
 - kernel, 63; (I)37, 55, 149, 210; (II)67, 78, 109, 113, 216; (III)111, 171, 217
 - conjugate, 63; (I)149; (II)78, 113; (III)255
- Poisson summation formula, 379; (I)154–156, 165, 174; (II)118
- polydisc, 277
- principal
 - curvatures, 333
 - value, 111
- probability
 - convergence, 195
 - measure, 192, 195
 - weak convergence, 219
 - space, 192
- process
 - stationary, 232
 - stochastic, 239
 - stopped, 261
- Prokhorov's lemma, 243
- proper hyperplane, 16
- pseudo-convex, 296
 - strongly, 296
- Rademacher functions, 192
- Radon
 - measure, 28, 100
 - transform, 363; (I)200, 203; (III)363
- random
 - flight, 237
 - Fourier series, 202
 - variable, 190
 - walk, 222
 - recurrent, 223
- recurrent
 - Brownian motion
 - neighborhood, 274
 - pointwise, 274
 - random walk, 223
- reflection, 63
- regular
 - distribution, 117, 132
 - point, 257
- restriction (L^p, L^q), 345
- Riemann-Lebesgue lemma, 93; (I)80; (III)94
- Riesz
 - convexity theorem, 57
 - diagram, 57
 - interpolation theorem, 52
 - product, 235
- rotational
 - curvature, 366
 - matrix, 365
- Schrödinger equation, 348

- Schwartz space, 105; (I)134, 180
- second fundamental form, 333
- section, 243
- separable
 - L^p space, 36
 - Banach space, 43
 - measure space, 36
- set
 - Borel, 29, 242
 - closure, 158; (II)6
 - convex, 17, 382; (II)107
 - cylinder, 191; (III)316
 - cylindrical, 242
 - dense, 158
 - first category, 158
 - generic, 158
 - interior, 158; (II)6
 - invariant, 207
 - meager, 158
 - nowhere dense, 158
 - second category, 158
 - strongly convex, 383
- signature, 294
- signum, 14
- singular integral, 62, 134
- Sobolev
 - embedding, 151; (III)257
 - space, 11, 151
- spherical maximal function, 406
- stationary
 - process, 232
- stationary phase, 325, 398; (II)326
- stochastic process, 239
- stopped process, 261
- stopping time, 254, 255
- Strichartz estimates, 351
- strong Markov property, 258
- strong solution, 360
- strongly convex set, 383
- strongly pseudo-convex, 296
- sub-algebra, 209
- support
 - distribution, 104
 - function, 104; (III)53
 - of a function, 28, 146
- surface-carried measure, 334
- smooth density, 334
- surjective mapping, 171
- tail algebra, 215
- tangential
 - Cauchy-Riemann vector field, 291
 - vector field, 290
- Tchebychev inequality, 73; (III)91
- tempered distribution, 106
- test functions, 100, 105
- three-lines lemma, 53, 339; (II)133
- Tietze extension principle, 269
- tight, 33, 243
- total curvature, 333
- type (of an operator), 56
- uniformly convex, 45
- universal element, 184
- upper
 - half-plane, 61
 - half-space, 307
- van der Corput inequality, 328
- variance, 196; (I)160
- vector field, 290
- Walsh-Paley functions, 230
- wave operator, 155
- weak
 - boundedness, 184
 - compactness of L^p , 37
 - convergence, 37, 221, 243; (III)198
- weak sense
 - continuity, 108
 - convergence, 103
 - derivative, 101
 - derivative in L^p , 10
 - tangential Cauchy-Riemann equations, 300
- weak* convergence, 44
- weak-type, 92
- weak-type inequality, 71; (III)101
- Weierstrass approximation theorem, 299; (I)54, 63, 144, 163
- Weierstrass preparation theorem, 282, 319
- Wiener measure, 240, 241
- Young's inequality, 39, 40, 60
- Yukawa potential, 149
- zero-one law, 199, 215
- zig-zag function, 165