Functional Analysis II
Spring Term 2017

James C. Robinson

Contents

1 Banach spaces: norms and separability *page* 1

1.1 Norms 1

1.2 Examples of Banach spaces 2

1.3 Density and separability 5

1.4 Linear maps 8

2 Dual spaces of sequence spaces and Lebesgue spaces 10

3 The Hahn–Banach Theorem and some applications 15

3.1 Statement of the Hahn–Banach Theorem 15

3.2 Some applications of the Hahn–Banach Theorem 16

4 Proof of the Hahn–Banach Theorem 18

4.1 Zorn's Lemma 18

4.2 The Hahn–Banach Theorem: real case 19

4.3 The Hahn–Banach Theorem: complex case 21

5 Separation theorems 25

6 The second dual and reflexivity 30

6.1 Reflexive spaces 31

6.2 Some examples of reflexive spaces 32

7 Linear maps between Banach spaces 35

7.1 The Baire Category Theorem 35

7.2 The Principle of Uniform Boundedness 36

7.2.1 Fourier series of continuous functions 38

8 Weak and weak-∗ convergence and the Banach–Alaoglu Theorem 40

8.1 Weak convergence 40

8.2 Weak closures 43

8.3 Weak-∗ convergence 44

8.3.1 Two weak compactness theorems 45

9 Open mapping, inverse mapping, and closed graph theorems 48

9.1 Open mapping and inverse mapping theorems 48

9.2 Bases in Banach spaces 50

9.3 The Closed Graph Theorem 53

10 Continuous functions 54

10.1 "Baire one" functions 54

10.2 The Arzelà–Ascoli Theorem 55

10.3 The Stone–Weierstrass Theorem 57

10.4 Mollification 60

11 An introduction to the theory of distributions 62

11.1 Test functions and distributions 62

11.1.1 Multi-index notation 62

11.1.2 Test functions and distributions 64

11.2 Differentiation, multiplication, and other operations 67

11.2.1 The distribution derivative 67

11.2.2 Products of smooth functions and distributions 69

11.2.3 Other operations on distributions in $\mathscr{D}'(\mathbb{R}^n)$ 70

11.3 The support of a distribution 72

11.4 Convolution and mollification of distributions 76

11.4.1 Convolution operators and fundamental solutions of PDEs 80

12 Zorn's Lemma (non-examinable) 84

1
Banach spaces: norms and separability

We will concentrate on Banach spaces and linear operators between Banach spaces.

1.1 Norms

A norm on a vector space X is a map $\|\cdot\|\colon X \to \mathbb{R}$ such that

(i) $\|x\| \geq 0$ for every $x \in X$ and $\|x\| = 0$ if and only if $x = 0$;
(ii) $\|\lambda x\| = |\lambda|\|x\|$ for every $\lambda \in \mathbb{K}$, $x \in X$; and
(iii) $\|x + y\| \leq \|x\| + \|y\|$ for every $x, y \in X$.

Any norm defines a distance $d(x, y) = \|x - y\|$.

The (closed) unit ball is

$$B_X(0,1) = \{x : \ \|x\| \leq 1\}.$$

This set is convex, since if $x, y \in B_X$ and $\lambda \in (0,1)$ then

$$\|\lambda x + (1-\lambda)y\| \leq \|\lambda x\| + \|(1-\lambda)y\| \leq \lambda\|x\| + (1-\lambda)\|y\| \leq 1,$$

and symmetric: if $x \in B_X$ then $-x \in B_X$.

Lemma 1.1 *Suppose that $N\colon X \to \mathbb{R}$ satisfies* (i) *and* (ii) *of the definition of a norm and in addition the set $B := \{x : \ N(x) \leq 1\}$ is convex. Then N is a norm on X.*

Proof We only need to prove the triangle inequality, i.e. show that

$$N(x+y) \leq N(x) + N(y).$$

If $N(x) = 0$ then $x = 0$ and $N(x+y) = N(y) = N(x) + N(y)$, so we can assume that $N(x) > 0$ and $N(y) > 0$.

Then $x/N(x) \in B$ and $y/N(y) \in B$, so using the convexity of B we have

$$\frac{N(x)}{N(x)+N(y)}\left(\frac{x}{N(x)}\right) + \frac{N(y)}{N(x)+N(y)}\left(\frac{y}{N(y)}\right) \in B.$$

So

$$\frac{x+y}{N(x)+N(y)} \in B,$$

which means that

$$N\left(\frac{x+y}{N(x)+N(y)}\right) = \frac{N(x+y)}{N(x)+N(y)} \leq 1 \quad \Rightarrow \quad N(x+y) \leq N(x)+N(y)$$

as required. □

For a related result see Examples 1.

1.2 Examples of Banach spaces

A Banach space is a complete normed space.

The finite-dimensional spaces $\mathbb{R}^n_p$ and $\mathbb{C}^n_p$ with elements

$$\underline{x} = (x_1, \ldots, x_n), \qquad x_j \in \mathbb{R} \text{ or } \mathbb{C}$$

with norms

$$\|\underline{x}\|_p := \left(\sum_{j=1}^n |x_j|^p\right)^{1/p} \qquad 1 \leq p < \infty$$

and

$$\|\underline{x}\|_\infty := \max_{j=1}^n |x_j|.$$

The infinite-dimensional sequence spaces $\ell^p(\mathbb{K})$ of sequences $\underline{x} = (x_j)_{j=1}^{\infty}$ ($x_j \in \mathbb{K}$) such that

$$\|\underline{x}\|_p := \left(\sum_j |x_j|^p\right)^{1/p} < \infty$$

and ℓ^∞ of sequences for which

$$\|\underline{x}\|_\infty := \sup_j |x_j| < \infty.$$

Particular useful element of ℓ^p are the sequences $\underline{e}^{(j)}$,

$$\underline{e}^{(j)} = (0, 0, \dots, 0, 0, 1, 0, 0, \dots),$$

the sequence consisting of entirely zeros apart from a single 1 in the jth place (its components are $e_i^{(j)} = \delta_{ij}$).

Definition 1.2 *The space c_0 is the subspace of ℓ^∞ consisting of null sequences ($x_j \to 0$ as $j \to \infty$); we equip this space with the ℓ^∞ norm.*

The ℓ^1 and ℓ^∞ norm clearly satisfy the triangle inequality, and the argument for ℓ^2 is familiar. For $1 < p < \infty$ we use Lemma 1.1.

Lemma 1.3 (Minkowski's inequality in ℓ^p spaces) *If $\underline{x}, \underline{y} \in \ell^p$ then $\underline{x} + \underline{y} \in \ell^p$ and*

$$\|\underline{x} + \underline{y}\|_{\ell^p} \le \|\underline{x}\|_{\ell^p} + \|\underline{y}\|_{\ell^p}.$$

Proof To show that $\|\cdot\|_p$ is a norm we can use Lemma 1.1 and show that the set

$$B := \{x \in \ell^p : \ \|x\|_p \le 1\} = \{x \in \ell^p : \ \|x\|_p^p \le 1\}$$

is convex, since the only issue is the triangle inequality. We use the fact that the function $t \mapsto |t|^p$ is convex[1] for all $1 \le p < \infty$: if $x, y \in B$ then

$$\begin{aligned}\|\lambda x + (1-\lambda)y\|_p^p &= \sum_p |\lambda x_j + (1-\lambda)y_j|^p \\ &\le \sum_p \lambda |x_j|^p + (1-\lambda)|y_j|^p \le 1.\end{aligned}$$ □

[1] A twice differentiable function on an interval (a, b) is convex on (a, b) iff its second derivative is non-negative, see examples.

We say that two indices $1 \le p, q \le \infty$ are *conjugate* if

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{1.1}$$

The following simple inequality is fundamental.

Lemma 1.4 (Young's inequality) *Let $a, b > 0$ and let (p, q) be conjugate indices with $1 < p, q < \infty$. Then*

$$ab \le \frac{a^p}{p} + \frac{b^q}{q}. \tag{1.2}$$

Proof The function e^x is convex and so

$$\begin{aligned} ab = \exp(\log a + \log b) &= \exp\left(\frac{1}{p}\log a^p + \frac{1}{q}\log b^q\right) \\ &\le \frac{1}{p}\mathrm{e}^{\log(a^p)} + \frac{1}{q}\mathrm{e}^{\log(b^q)} = \frac{a^p}{p} + \frac{b^q}{q}. \end{aligned}$$ □

Lemma 1.5 (Hölder's inequality in ℓ^p spaces) *Let $\underline{x} \in \ell^p$ and $\underline{y} \in \ell^q$ with p, q conjugate, $1 \le p, q \le \infty$. Then if $\underline{z} = (x_1y_1, x_2y_2, \ldots)$, $\underline{z} \in \ell^1$ with*

$$\|\underline{z}\|_{\ell^1} = \sum_{j=1}^{\infty} |x_j y_j| \le \|\underline{x}\|_{\ell^p} \|\underline{y}\|_{\ell^q}. \tag{1.3}$$

Proof For $1 < p < \infty$, consider

$$\sum_{j=1}^{n} \frac{|x_j|}{\|\underline{x}\|_{\ell^p}} \frac{|y_j|}{\|\underline{y}\|_{\ell^q}} \le \sum_{j=1}^{n} \frac{1}{p} \frac{|x_j|^p}{\|\underline{x}\|_{\ell^p}^p} + \frac{1}{q} \frac{|y_j|^q}{\|\underline{y}\|_{\ell^q}^q} \le 1.$$

So for each $n \in \mathbb{N}$

$$\sum_{j=1}^{n} |x_j y_j| \le \|\underline{x}\|_{\ell^p} \|\underline{y}\|_{\ell^q}$$

and (1.3) follows. For $p = 1$, $q = \infty$,

$$\sum_{j=1}^{n} |x_j y_j| \le \max_{j=1,\ldots,n} |y_j| \left(\sum_{j=1}^{n} |x_j|\right) \le \|\underline{x}\|_{\ell^1} \|\underline{y}\|_{\ell^\infty}.$$ □

Proposition 1.6 *For each $1 \le p \le \infty$, the sequence space ℓ^p (equipped with its standard norm) is complete, and so is c_0 (with the ℓ^∞ norm).*

Proof For the completeness of ℓ^p see Functional Analysis I (Theorems 4.8–4.10). For the completeness of c_0 see Examples 1. □

More important, perhaps, are spaces of functions.

The space $C^0([a,b])$ of continuous (real or complex-valued) functions on $[a,b]\subset\mathbb{R}$ equipped with the supremum norm is a Banach space (Theorem 4.5 in FA1).

If (Ω,Σ,μ) is a measure space then the space $L^p(\Omega;\mu)$, $1\le p<\infty$, consists of 'functions' on Ω such that

$$\int_\Omega |f|^p\,\mathrm{d}\mu<\infty$$

is a Banach space with norm

$$\|f\|_{L^p}=\left(\int_\Omega |f|^p\,\mathrm{d}\mu\right)^{1/p}. \tag{1.4}$$

The proof of the triangle inequality for this norm follows as for ℓ^p.

The space $L^\infty(\Omega)$ consists of 'functions' such that

$$\|f\|_{L^\infty}=\operatorname{ess\,sup} f=\inf\{M:\ |f(x)|\le M\ \mu\text{-almost everywhere}\}.$$

Since there are non-zero functions for which $\|f\|_{L^p}=0$ (e.g. taking the value 1 at all rationals) to make $\|\cdot\|_{L^p}$ into a norm we have to identify all elements that agree almost everywhere, so strictly $L^p(\Omega)$ is an equivalence class of functions. We will usually (always?) take $\Omega\subset\mathbb{R}^n$ for some n and let μ be Lebesgue measure. (Translation (if you have not done the measure theory course): the integral in (1.4) is a generalisation of the integral you know that allows you to understand what it means for a much larger class of functions than the 'regulated functions' of Analysis III).

1.3 Density and separability

A subset A of a metric space (X,d) is *dense* if every $x\in X$ can be approximated arbitrarily closely by an element of A: given $x\in X$ and $\epsilon>0$ there exists an $a\in A$ such that

$$d(x,a)<\epsilon.$$

A metric space is *separable* if it has a countable dense subset.

Lemma 1.7 *If (X,d) is separable and $Y \subset X$ then (Y,d) is also separable.*

Proof Given $\{x_n\}$ that are dense in X, for each $n,k \in \mathbb{N}$, if Y contains a point with $d(x_n,y) < 1/k$ call this point $y_{n,k}$ and add it to A. In this way A is (at most) a countable set.

To show that A is dense, take $y \in Y$ and $\epsilon > 0$. Taking k such that $1/k < \epsilon/2$ and $x_n \in X$ with $d(x_n,y) < 1/k$, it follows (since $d(x_n,y) < 1/k$) that there exists $y_{n,k} \in A$ such that $d(x_n,y_{n,k}) < 1/k$ and hence

$$d(y_{n,k},y) \leq d(y_{n,k},x_n) + d(x_n,y) < 2/k < \epsilon. \quad \square$$

Lemma 1.8 *Let X be a normed space. The following three statements are equivalent:*

(i) *X is separable;*
(ii) *$S_X = \{x \in X : \|x\| = 1\}$ is separable; and*
(iii) *X contains a sequence $\{x_1, x_2, x_3, \ldots\}$ whose linear span is dense.*

Note that the *linear span* of $\{x_1, x_2, x_3, \ldots\}$ consists of *finite* linear combinations of the x_i.

Proof Lemma 1.7 shows that (i) ⇒ (ii). For (ii) ⇒ (iii) choose a countable dense subset $\{x_1, x_2, x_3, \ldots\}$ of S_X: then for any $x \in X$ we have $x/\|x\| \in S_X$, and so for any $\epsilon > 0$ there exists an x_k such that

$$\left\| x_k - \frac{x}{\|x\|} \right\| < \frac{\epsilon}{\|x\|}.$$

It follows that

$$\big\| x - \|x\| x_k \big\| < \epsilon,$$

and clearly $\|x\|x_k$ is contained in the linear span of the $\{x_j\}$.

To show that (iii) implies (i) note that the collection of finite linear combinations of the $\{x_j\}$ with rational coefficients is countable. This countable collection is dense: given $x \in X$ and $\epsilon > 0$, choose an element of the linear span of $\{x_1, x_2, \ldots\}$ such that

$$\left\| x - \sum_{j=1}^{n} \alpha_j x_j \right\| < \frac{\epsilon}{2},$$

and then choose $q_j \in \mathbb{Q}$ for $j = 1, \ldots, n$ such that

$$|q_j - \alpha_j| < \frac{\epsilon}{2\|x_j\|}.$$

It then follows from the triangle inequality that

$$\left\| x - \sum_{j=1}^{n} q_j x_j \right\| < \epsilon. \qquad \square$$

$\mathbb{R}^n$ and $\mathbb{C}^n$ are separable. $C^0([a,b])$ is separable (this is a consequence of the Weierstrass Approximation Theorem – Theorem 3.1 in FA1).

Corollary 1.9 *ℓ^p is separable if $1 \leq p < \infty$, and c_0 is separable.*

In these spaces the linear span of the countable collection $\{\underline{e}^{(j)}\}$ is dense.

Proposition 1.10 *ℓ^∞ is not separable.*

Proof Consider the elements of ℓ^∞ that consist of 0 or 1 in each coordinate; these elements form an uncountable set X (Cantor diagonal argument). Furthermore any two distinct elements of X are a distance 1 apart. Now suppose that S is a dense subset of ℓ^∞; it follows that a different element of S is needed to approximate each element of X, and so S must be uncountable. It follows that ℓ^∞ is not separable. $\square$

Proposition 1.11 *For any $E \subset \mathbb{R}^n$ the space $L^p(E)$ is separable for $1 \leq p < \infty$, but $L^\infty(E)$ is not separable.*

Proof If E is a compact subset of $\mathbb{R}^n$ then this can be obtained as a corollary of the Weierstrass Approximation Theorem (Theorem 3.1 in FA1). The general proof for $L^p(E)$ shows that the set of characteristic functions of dyadic cubes satisfies (iii) of Lemma 1.8. For L^∞ see Examples 1. $\square$

1.4 Linear maps

Definition 1.12 *An operator $T : X \to Y$ is linear if*

$$T(\lambda x + \mu y) = \lambda T(x) + \mu T(y), \qquad x, y \in X, \ \lambda, \mu \in \mathbb{K}.$$

Recall that T is *bounded* if there exists $M \geq 0$ such that

$$\|T(x)\|_Y \leq M\|x\|_X \qquad \text{for all } x \in X$$

and that a linear operator is continuous if and only if it is bounded (Theorem 7.3 in FA1).

We denote by $B(X, Y)$ the collection of all bounded operators from X into Y and define

$$\|T\|_{B(X,Y)} := \sup_{\|x\|=1} \|T(x)\|_Y = \sup_{x \neq 0} \frac{\|T(x)\|_Y}{\|x\|_X}.$$

This is a norm: we need only check the triangle inequality, which we can do by taking $x \in X$ with $\|x\| = 1$ and considering

$$\|(T + S)x\|_Y = \|Tx + Sx\|_Y \leq \|Tx\|_Y + \|Sx\|_Y \leq \|T\| + \|S\|;$$

it follows that $\|T + S\| \leq \|T\| + \|S\|$.

Recall the following (Proposition 7.7 in FA1).

Theorem 1.13 *If Y is a Banach space then $B(X, Y)$ is a Banach space.*

Two spaces X and Y are isomorphic if there is a linear map $T \colon X \to Y$ such that T and T^{-1} are bounded. They are isometrically isomorphic if there is a such a T with $\|T\| = \|T^{-1}\| = 1$; in this case we write $X \simeq Y$.

Note that any linear isometry is automatically injective; if $Tx = Ty$ then

$$0 = \|Tx - Ty\| = \|T(x - y)\| = \|x - y\|$$

and so $x = y$.

Lemma 1.14 *If X and Y are Banach spaces and $X \simeq Y$ then X is separable if and only if Y is separable.*

Proof The linear isometric $T\colon X \to Y$ will map a dense subspace of X to a dense subspace of Y, similarly with $T^{-1}\colon Y \to X$. □

Given a linear operator $T\colon X \to Y$, we define its kernel

$$\operatorname{Ker} T = \{x \in X : \ Tx = 0\}$$

and its range

$$\operatorname{Range} T = \{y \in Y : \ y = Tx \text{ for some } x \in X\}.$$

Lemma 1.15 *If $T \in B(X,Y)$ then $\operatorname{Ker} T$ is a closed linear subspace of X.*

Proof It is easy to show that $\operatorname{Ker}(T)$ is a linear subspace: if $\alpha, \beta \in \mathbb{K}$ and $x, y \in \operatorname{Ker}(T)$ then

$$T(\alpha x + \beta y) = \alpha T x + \beta T y = 0.$$

Furthermore if $x_n \to x$ and $Tx_n = 0$ then since T is continuous it follows that $Tx = \lim_{n\to\infty} Tx_n = 0$, so $\operatorname{Ker}(T)$ is closed. □

Note, however, that the range of a map in $B(X,Y)$ is not necessarily closed. Indeed, consider the map from ℓ^2 into itself given by

$$T\underline{x} = (x_1, \frac{x_2}{2}, \frac{x_3}{3}, \frac{x_4}{4}, \ldots).$$

Then clearly $\|T\|_{\mathrm{op}} \le 1$, so T is bounded. Now consider

$$\underline{y}^{(n)} = T(\underbrace{1, 1, \ldots, 1}_{n \text{ times}}, 0 \ldots) = \left(1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, 0, \ldots\right).$$

Then clearly $\underline{y}^{(n)} \to \underline{y}$ with $y_j = j^{-1}$, and since

$$\sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{\pi^2}{6} < \infty$$

it follows that $\underline{y} \in \ell^2$. However, there is no $\underline{x} \in \ell^2$ such that $T(\underline{x}) = \underline{y}$: the only candidate is $\underline{x} = (1, 1, 1, \ldots)$, but this is not in ℓ^2 since its ℓ^2 norm is not finite.

2

Dual spaces of sequence spaces and Lebesgue spaces

If X is a normed space then its dual space $X^* = B(X, \mathbb{K})$ is always a Banach space with norm

$$\|\phi\|_{X^*} = \sup_{\|x\| \le 1} |\phi(x)| \tag{2.1}$$

(Proposition 6.11 in FA1).

Lemma 2.1 *If H is a Hilbert space then $H^* \simeq H$; the map $x \mapsto f_x$, where $f_x(y) = (y, x)$ is an antilinear isometry from H onto H^*. We denote by $R\colon H \to H^*$ the map $R(x) = f_x$.*

We say that a map is antilinear (conjugate linear) if

$$L(\alpha x + \beta z) = \bar{\alpha} L(x) + \bar{\beta} L(z).$$

So for example (and this more or less what we have here), $y \mapsto (x, y)$ is antilinear in y. The map in the above lemma is a linear isometric isomorphism if H is a real Hilbert space.

Proof The map $x \mapsto f_x$ is anti-linear, since

$$f_{\alpha x + \beta z} = (y, \alpha x + \beta z) = \bar{\alpha}(y, x) + \bar{\beta}(y, z).$$

Given $x \in H$, $f_x \in H^*$ since

$$|f_x(y)| = |(y, x)| \le \|y\| \|x\|,$$

so $\|f_x\|_{H^*} \le \|x\|$. However, $|f_x(x)| = \|x\|^2$, and so $\|f_x\|_{H^*} = \|x\|$, i.e. $x \mapsto f_x$ is an isometry.

We now have to show that the map is onto, i.e. every $f \in H^*$ can be expressed as $(\cdot, x)$ for some $x \in H$. This is precisely the content of the Riesz Representation Theorem (Theorem 6.14 in FA1). □

In the case of $\ell^2(\mathbb{R})$ this shows that $(\ell^2)^* \simeq \ell^2$ via the map $x \mapsto f_x$, where

$$f_x(y) = (y, x) = \sum_{j=1}^{\infty} \bar{x}_j y_j.$$

We now use a very similar map to investigate the dual spaces of ℓ^p, $1 \leq p < \infty$, and c_0. We drop the underlinings on elements of the sequence spaces to declutter the notation.

Theorem 2.2 *For $1 < p, q < \infty$ with (p, q) conjugate, $(\ell^p) \simeq (\ell^q)^*$, via the mapping $x \mapsto L_x$, where*

$$L_x(y) = \sum_{j=1}^{\infty} x_j y_j. \tag{2.2}$$

We denote this mapping as $T_q \colon \ell^p \to (\ell^q)^*$.

The case $p = q = 2$ here follows from the Riesz Representation Theorem when H is real.

Proof Given $x \in \ell^p$ define L_x as in (2.2) above. Then from Hölder's inequality

$$|L_x(y)| = \left| \sum_j x_j y_j \right| \leq \sum_j |x_j y_j| \leq \|x\|_{\ell^p} \|y\|_{\ell^q}, \tag{2.3}$$

so we do indeed have $L_x \in (\ell^q)^*$. To show that $\|L_x\|_{(\ell^q)^*} = \|x\|_{\ell^p}$ consider the element $y \in \ell^q$ given by

$$y_j = \begin{cases} |x_j|^p / x_j & x_j \neq 0 \\ 0 & x_j = 0; \end{cases}$$

this is in ℓ^q since

$$\|y\|_{\ell^q}^q = \sum_j |y_j|^q = \sum_j |x_j|^{q(p-1)} = \sum_j |x_j|^p < \infty,$$

as $q(p-1)=p$ and we have

$$|L_x(y)| = \left|\sum_j x_j y_j\right| = \sum_j |x_j|^p = \left(\sum_j |x_j|^p\right)^{1-\frac{1}{p}} \|x\|_{\ell^p} = \|y\|_{\ell^q}\|x\|_{\ell^p},$$

since $1-(1/p)=1/q$.

Since $x \mapsto L_x$ is a linear isometry it is injective: if $x \neq y$ then

$$\|L_x - L_y\|_{\ell^{q*}} = \|L_{x-y}\|_{\ell^{q*}} = \|x-y\|_{\ell^p}.$$

We now show that this is surjective; i.e. that any $L \in (\ell^q)^*$ can be written as L_x for some $x \in \ell^p$.

If we do have $L = L_x$ then for each e_j (the element of ℓ^p with $(e_j)_i = \delta_{ij}$) we must have

$$L(e_j) = L_x(e_j) = x_j.$$

If x defined component-wise in this way is an element of ℓ^p then for any $y \in \ell^q$ we have $y = \sum_j y_j e_j$, and since L is continuous

$$\begin{aligned} L\left(\sum_{j=1}^{\infty} y_j e_j\right) &= L\left(\lim_{n\to\infty}\sum_{j=1}^{\infty} y_j e_j\right) = \lim_{n\to\infty} L\left(\sum_{j=1}^{\infty} y_j e_j\right) \\ &= \lim_{n\to\infty}\sum_{j=1}^{n} y_j L(e_j) = \lim_{n\to\infty}\sum_{j=1}^{n} y_j x_j = L_x(y). \end{aligned}$$

So we need only show that x defined by $x_j = L(e_j)$ is an element of ℓ^p. To do this consider the sequence $\phi^{(n)}$ of elements of ℓ^q defined by

$$\phi_j^{(n)} = \begin{cases} |x_j|^p/x_j & j \leq n \text{ and } x_j \neq 0 \\ 0 & j > n \text{ or } x_j = 0; \end{cases}$$

then

$$L(\phi^{(n)}) = L\left(\sum_{j=1}^{n} \phi_j^{(n)} e_j\right) = \sum_{j=1}^{n} \phi_j^{(n)} L(e_j) = \sum_{j=1}^{n} \phi_j^{(n)} x_j = \sum_{j=1}^{n} |x_j|^p,$$

and so

$$\sum_{j=1}^{n} |x_j|^p = |L(\phi^{(n)}| \le \|L\|_{(\ell^q)^*} \|\phi^{(n)}\|_{\ell^q}$$

$$= \|L\|_{(\ell^q)^*} \left(\sum_{j=1}^{n} |x_j|^{q(p-1)} \right)^{1/q}$$

$$= \|L\|_{(\ell^q)^*} \left(\sum_{j=1}^{n} |x_j|^p \right)^{1/q},$$

which shows that

$$\left(\sum_{j=1}^{n} |x_j|^p \right)^{1/p} \le \|L\|_{(\ell^q)^*}$$

and so $x \in \ell^p$ as required. □

The same is true for one of the endpoints, but not the other.

Theorem 2.3 *We have $\ell^\infty \simeq (\ell^1)^*$ via the the mapping (2.2), which we denote by $T_1 \colon \ell^\infty \to (\ell^1)^*$.*

Proof Given $x \in \ell^\infty$, Hölder's inequality as in (2.3) shows that L_x defined as in (2.2) is an element of $(\ell^1)^*$ with

$$\|L_x\|_{(\ell^1)^*} \le \|x\|_{\ell^\infty}.$$

The equality of norms follows by choosing, for each $\varepsilon > 0$, a $j \in \mathbb{N}$ such that $|x_j| > \|x\|_{\ell^\infty} - \varepsilon$, and then considering $y \in \ell^1$ with $y_i = \delta_{ij}\mathrm{e}^{-\mathrm{i}\theta}$, where $x_j = |x_j|\mathrm{e}^{\mathrm{i}\theta}$ [note that in the real case this is $y_i = \delta_{ij}\mathrm{sgn}(x_j)$]. Then

$$|L_x(y)| = |x_j| \ge (\|x\|_{\ell^\infty} - \varepsilon)\, \|y\|_{\ell^1},$$

since $\|y\|_{\ell^1} = 1$. Since this is valid for any $\varepsilon > 0$, it follows that

$$\|L_x\|_{(\ell^1)^*} = \|x\|_{\ell^\infty}.$$

To show that the map $x \mapsto L_x$ is onto we use the same argument as before and consider x defined by $x_j = L(e_j)$. It is easy to see that this is an element of ℓ^∞, since $e_j \in \ell^1$ and

$$|x_j| = |\bar{x}_j| = |L(e_j)| \le \|L\|_{(\ell^1)^*} \|e_j\|_{\ell^1} = \|L\|_{(\ell^1)^*}.$$ □

However, $(\ell^\infty)^* \not\simeq \ell^1$ (for a proof of this see later); instead we have the following.

Theorem 2.4 $(c^0)^* = \ell^1$.

For a proof see Examples 2.

Similar duality results hold in the Lebesgue spaces L^p, and are proved in the Measure Theory module.

Theorem 2.5 *Suppose that $(\Omega, \mathscr{F}, \mu)$ is a σ-finite measure space. Then for $1 \leq p < \infty$ the space $L^q(\Omega)$ is isometrically isomorphic to $(L^p(\Omega))^*$, where (p, q) are conjugate, via the mapping $g \mapsto \phi_g$, where*

$$\phi_g(f) := \int fg \, \mathrm{d}\mu.$$

Note that coupled with (2.1) the result of the previous theorem gives one way (which is sometimes useful) to find the L^p norm of f for any $1 < q \leq \infty$:

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

3
The Hahn–Banach Theorem and some applications

The Hahn–Banach Theorem guarantees that a linear functional defined on a subspace of X can be extended to a linear functional defined on the whole of X without increasing its norm.

3.1 Statement of the Hahn–Banach Theorem

But first, here is an easy version in a Hilbert space.

Lemma 3.1 *Let H be a Hilbert space and U a closed linear subspace. If $f \in U^*$ then f has an extension to an element $F \in H^*$ such that $F(x) = f(x)$ for every $x \in U$ and $\|F\| = \|f\|$.*

Proof U is a Hilbert space with the same norm/inner product as H. By the Riesz Representation Theorem (FA1 Theorem 6.14) there exists $v \in U$ with $\|v\| = \|f\|_{U^*}$ such that $f(u) = (u, v)$ for every $u \in U$. Define $F(u) = (u, v)$ for every $u \in H$; then $\|F\|_{H^*} = \|v\| = \|f\|_{U^*}$. □

In the rest of this chapter we will explore consequences of the Hahn–Banach Theorem in a Banach space.

Theorem 3.2 (Hahn–Banach) *Let X be a Banach space and U a subspace of X. If $f \in U^*$ then f has an extension to an element $F \in X^*$, i.e. $F(x) = f(x)$ for every $x \in U$, such that $\|F\|_{X^*} = \|f\|_{U^*}$.*

3.2 Some applications of the Hahn–Banach Theorem

As an immediate application, we prove the existence of a particularly useful class of linear functionals, and show therefore that understanding linear functionals is in some way enough to understand elements of X.

Corollary 3.3 (Support functional) *Let $x \in X$. Then there exists an $f \in X^*$ such that $\|f\|_{X^*} = 1$ and $f(x) = \|x\|$.*

Proof Define $\hat{f}$ on the linear space U spanned by x as

$$\hat{f}(\alpha x) = \alpha \|x\|.$$

Then $\hat{f}(x) = \|x\|$ and $|\hat{f}(z)| \leq \|z\|$ for all $z \in U$. Extend $\hat{f}$ to an $f \in X^*$; then $\|f\|_{X^*} = 1$ and $f(x) = \hat{f}(x) = \|x\|$. □

The following simple corollary shows that X^* is rich enough to distinguish between elements of X.

Corollary 3.4 (X^* separates points) *Let $x, y \in X$. If $f(x) = f(y)$ for every $f \in X^*$ then $x = y$.*

Proof If $x \neq y$ then by the previous lemma there exists an f with $\|f\|_{X^*} = 1$ such that $f(x) - f(y) = f(x - y) = \|x - y\| \neq 0$. □

The next result is a key ingredient in many subsequent proofs.

Corollary 3.5 (Closest point witness) *Let Y be a proper closed subspace of a Banach space X and let $x \in X \setminus Y$. Set*

$$d = \inf\{\|x - y\| : \ y \in Y\}.$$

Then there is a $\phi \in X^*$ such that $\|\phi\| = 1$, $\phi(y) = 0$ for every $y \in Y$, and $\phi(x) = d$.

The functional ϕ shows that x is at least a distance d from Y, since

$$d = \phi(x) = \phi(x - y) \leq \|\phi\|\|x - y\| = \|x - y\|.$$

Proof Let $Z = \text{span}\{Y \cup \{x\}\}$ and define $\phi : Z \to \mathbb{R}$ by $\phi(y + \lambda x) = \lambda d$ for $y \in Y$ and $\lambda \in \mathbb{R}$. To see that ϕ is bounded on Z, observe that

$$|\phi(y + \lambda x)| = |\lambda| d \leq |\lambda| \|x - (-y/\lambda)\| = \|\lambda x + y\|$$

since $(-y/\lambda) \in Y$ and d is the distance between x and Y. So $\|\phi\| \leq 1$. To see that $\|\phi\| \geq 1$ take $y_n \in Y$ such that

$$\|x - y_n\| \leq d\left(1 + \frac{1}{n}\right);$$

then

$$d = \phi(-y_n + x) \geq \frac{n}{n+1}\|x - y_n\|$$

and so $\|\phi\| \geq n/n+1$ for every n, i.e. $\|\phi\| \geq 1$.

We now extend ϕ to X using the Hahn–Banach Theorem. □

We have already seen that $(\ell^1)^* \simeq \ell^\infty$; we know that ℓ^1 is separable but that ℓ^∞ is not separable. So in general separability of X does not imply that X^* is separable. However, the converse is true.

Lemma 3.6 *If X^* is separable then X is separable.*

Proof Since X^* is separable, the unit sphere in X^* is separable (by (i) ⇒ (ii) in Lemma 1.8). Let (f_n) be a countable dense subset of S_{X^*}. For each n there exists an $x_n \in X$ with $\|x_n\| = 1$ such that $|f_n(x_n)| \geq 1/2$, by the definition of the norm in X^*.

We now show that M, the closed linear span of the (x_n), is all of X, and hence (by (iii) ⇒ (i) of Lemma 1.8) X is separable. Suppose that the closed linear span is not all of X. Then M is a proper closed subspace of X, and so Corollary 3.5 provides an $f \in X^*$ with $\|f\| = 1$ and $f(x) = 0$ for every $x \in M$. But then $f(x_n) = 0$ for every n and so

$$\frac{1}{2} \leq |f_n(x_n)| = |f_n(x_n) - f(x_n)| \leq \|f_n - f\|\|x_n\| = \|f_n - f\|$$

for every n, which contradicts the fact that $\{f_n\}$ is dense in S_{X^*}. □

Note that this shows that $(\ell^\infty)^* \not\simeq \ell^1$, because ℓ^1 is separable and ℓ^∞ is not, and separability is preserved under isometric isomorphisms.

4
Proof of the Hahn–Banach Theorem

One can prove the Hahn–Banach Theorem in a separable Banach space without using Zorn's Lemma, see Examples 3, but we will give the proof for an arbitrary Banach space.

4.1 Zorn's Lemma

In order to state Zorn's Lemma (which in fact is more of an axiom, since it is equivalent to the axiom of choice) we need to introduce some auxiliary concepts.

Definition 4.1 *A partial order on a set P is a binary relation $\preceq$ on P such that*

(i) $a \preceq a$ for all $a \in P$,

(ii) $a \preceq b$ and $b \preceq a$ implies that $a = b$, and

(iii) $a \preceq b$ and $b \preceq c$ implies that $a \preceq c$.

Definition 4.2 *A subset C of P in which all elements can be ordered is called a* chain*: i.e. for all $a, b \in C$, either $a \preceq b$ or $b \preceq a$ (or both, in which case $a = b$).*

An element $b \in P$ is an upper bound *for a subset S of P if $s \preceq b$ for all $s \in P$. An element m of P is* maximal *if $m \preceq a$ for some $a \in P$ implies that $a = m$.*

Lemma 4.3 (Zorn's Lemma) *Let P be a non-empty partially ordered set. If every chain in P has an upper bound then P has at least one maximal element.*

4.2 The Hahn–Banach Theorem: real case

In fact we will prove a result that allows for upper bounds on $f\colon U \to \mathbb{R}$ in terms of a sublinear functional.

If V is a vector space over $\mathbb{R}$ then a function $p\colon V \to \mathbb{R}$ is *sublinear* if

$$p(x+y) \le p(x) + p(y) \qquad \text{and} \qquad p(\lambda x) = \lambda x,\ \lambda \ge 0.$$

We say that p is a *seminorm* if additionally $p(\lambda x) = |\lambda| p(x)$ for every $\lambda \in \mathbb{K}$. Note that if p is a seminorm then $p(x) \ge 0$ for every $x \in X$ ($p(0) \le p(-x) + p(x) = 2p(x)$); and that if $\|\cdot\|$ is a norm on X then for any $M > 0$, $M\|\cdot\|$ defines a seminorm on X (actually, it defines another norm, and any norm is also a seminorm).

Theorem 4.4 (Sublinear Hahn–Banach Theorem) *Let X be a real vector space, and U a subspace of X. Suppose that $f\colon U \to \mathbb{R}$ is linear and satisfies*

$$f(x) \le p(x) \qquad \textit{for all} \quad x \in U.$$

Then there exists a linear map $F\colon X \to \mathbb{R}$ such that $F(x) = f(x)$ for all $x \in U$ and

$$F(x) \le p(x) \qquad \textit{for all} \quad x \in X.$$

Furthermore, if p is a seminorm then

$$|F(x)| \le p(x) \qquad \textit{for all} \quad x \in X.$$

In particular any $f \in U^*$ has an extension $F \in X^*$ with $\|F\|_{X^*} = \|f\|_{U^*}$.

The final statement will be the most useful version of the theorem for us in what follows.

Proof We will apply Zorn's Lemma to all possible extensions g of f satisfying the bound $g(x) \le p(x)$. More precisely, consider the collection P of all possible extensions of g of f to subspaces G, i.e. the collection of linear functionals $g\colon G \to \mathbb{R}$ for some subspace G of X such that $g = f$ on U and

$$g(x) \le p(x) \qquad \text{for every } x \in G.$$

Then P is non-empty, since $(U, f) \in P$.

We define an order on P by defining $(G,g) \preceq (H,h)$ if h is an extension of g, i.e. $H \supseteq G$ and $h = g$ on G.

Now, any chain $C = \{(g_i, G_i)\}$ has an upper bound, namely the pair (G_∞, g_∞), where

$$G_\infty = \bigcup_i G_i$$

and

$$g_\infty(x) = g_i(x) \qquad x \in G_i.$$

Note that g_∞ is well defined, since any two elements in C are ordered: if $x \in G_j \cap G_i$ then either $(G_j, g_j) \preceq (G_i, g_i)$ (or vice versa), and we know that $g_i = g_j$ on G_j, since g_i extends g_j.

Similarly, g_∞ satisfies $g_\infty(x) \le p(x)$ for every $x \in G_\infty$, since this bound holds for (G_i, g_i) when $x \in G_i$. Also, g_∞ is linear: if $x, y \in G_\infty$, then there exists G_i such that $x, y \in G_i$ ($x \in G_i$, $y \in G_j$, and $G_i \supset G_j$ or $G_j \supset G_i$).

Since any chain has an upper bound, P must have a maximal element (Y, g). We want to show that $Y = X$. If not, then there exists $z \in X \setminus Y$; we want to show that in this case we can extend g to the linear span of $Y \cup \{z\}$, which is a space strictly larger than Y. This contradicts the maximality of (Y, g), and so $Y = X$.

We want to set $F(u + \alpha z) = g(u) + \alpha c$ for some choice of $c \in \mathbb{R}$; the only issue is how to choose c so that

$$g(u) + \alpha c \le p(u + \alpha z) \tag{4.1}$$

for every choice of $\alpha \in \mathbb{R}$ and every $u \in Y$. We know (by assumption) that this holds for $\alpha = 0$, so we have to guarantee that we can find c such that (i) for $\alpha > 0$ we have (dividing by α)

$$c \le p\left(\frac{u}{\alpha} + z\right) - \frac{g(u)}{\alpha} \qquad u \in Y$$

and (ii) for every $\alpha < 0$ we have (dividing by $-\alpha$)

$$c \ge \frac{g(u)}{-\alpha} - p\left(\frac{u}{-\alpha} - z\right) \qquad u \in Y.$$

Since Y is a linear subspace and g is linear this is the same as requiring

$$g(v) - p(v - z) \le c \le p(v + z) - g(v) \qquad \text{for every } v \in Y. \tag{4.2}$$

We show that the right-hand side is bounded below using the triangle property for p: take $v_1, v_2 \in Y$, and then

$$\begin{aligned} g(v_1) + g(v_2) &= g(v_1 + v_2) \\ &\leq p(v_1 + v_2) = p(v_1 - z + v_2 + z) \\ &\leq p(v_1 - z) + p(v_2 + z) \end{aligned}$$

and so

$$g(v_1) - p(v_1 - z) \leq -g(v_2) + p(v_2 + z) \qquad v_1, v_2 \in Y.$$

Hence we an find a c to ensure that (4.2) holds.

We can therefore extend (Y, g) to a linear functional F defined on H, the span of Y and z, that still satisfies $F(x) \leq p(x)$ for all $x \in H$. This contradicts the maximality of (g^*, G^*), so $G^* = X$.

This proves the result for sublinear functions.

If p is a seminorm then we have

$$F(x) \leq p(x) \qquad \text{and} \qquad -F(x) = F(-x) \leq p(-x) = p(x),$$

and so $|F(x)| \leq p(x)$. To prove the result for norms we only need to show that

$$p(x) = M\|x\|$$

defines a seminorm; but this is trivial. □

4.3 The Hahn–Banach Theorem: complex case

To extend the Hahn–Banach Theorem to the complex case (when elements of V^* map elements of V to complex numbers), first observe that any complex vector space V can be viewed as a real vector space by only allowing scalar multiplication by real numbers, we call this space $V_{\mathbb{R}}$. This does not effect the elements of the space – for us, the main significance is that an element $\phi \in V_{\mathbb{R}}^*$ is a map from V into $\mathbb{R}$ that is linear in the sense that

$$\phi(\alpha + \beta y) = \alpha\phi(x) + \beta\phi(y)$$

for all real, but not complex, α and β.

Lemma 4.5 *Let V be a complex vector space. Given any linear functional $f\colon V \to \mathbb{C}$ there exists a unique linear $\phi\colon V_{\mathbb{R}} \to \mathbb{R}$ such that*

$$f(v) = \phi(v) - \mathrm{i}\phi(\mathrm{i}v) \qquad \textit{for all} \qquad v \in V. \tag{4.3}$$

If V is a normed space and $f \in V^*$ then $\|\phi\|_{V_{\mathbb{R}}^*} = \|f\|_{V^*}$. Conversely, if $\phi \in (V_{\mathbb{R}})^*$ then f defined by (4.3) is an element of V^* with $\|f\|_{V^*} = \|\phi\|_{V_{\mathbb{R}}^*}$.

Note that it is also immediate that if $|f(v)| \le p(v)$ then this is inherited by ϕ, since

$$|f(v)|^2 = |\phi(v)|^2 + |\phi(\mathrm{i}v)|^2 \qquad \Rightarrow \qquad |\phi(v)| \le |f(v)| \le p(v).$$

The equality in the case of a normed space requires some proof.

Proof If $v \in V$ then

$$f(v) = \phi(v) + \mathrm{i}\psi(v),$$

where $\phi, \psi \in (V_{\mathbb{R}})^*$ [these maps are linear because we allow only real scalar multiples in the definition of 'linearity' in $V_{\mathbb{R}}^*$]. Since

$$\phi(\mathrm{i}v) + \mathrm{i}\psi(\mathrm{i}v) = f(\mathrm{i}v) = \mathrm{i}f(v) = \mathrm{i}\phi(v) - \psi(\mathrm{i}v)$$

it follows that $\phi(\mathrm{i}v) = -\psi(v)$ which yields (4.3).

Now for any $x \in X$ we have from (4.3)

$$|f(x)|^2 = |\phi(x)|^2 + |\phi(\mathrm{i}x)|^2 \ge |\phi(x)|^2 \tag{4.4}$$

and so $|f(x)| \ge |\phi(x)|$ for every $x \in X$ which implies that $\|f\| \ge \|\phi\|$.

For the reverse inequality, observe that for any x we can write

$$|f(x)| = \mathrm{e}^{\mathrm{i}\theta} f(x)$$

for some $\theta \in \mathbb{R}$. So

$$|f(x)| = \mathrm{e}^{\mathrm{i}\theta} f(x) = f(\mathrm{e}^{\mathrm{i}\theta} x) = \phi(\mathrm{e}^{\mathrm{i}\theta} x) - \mathrm{i}\phi(\mathrm{i}\mathrm{e}^{\mathrm{i}\theta x}).$$

Since $|f(x)|$ is real we must have, for all $x \in V$,

$$|f(x)| = \phi(\mathrm{e}^{\mathrm{i}\theta} x) \le |\phi(\mathrm{e}^{\mathrm{i}\theta} x)| \le \|\phi\|\|\mathrm{e}^{\mathrm{i}\theta} x\| = \|\phi\|\|x\|,$$

and so $\|f\| \le \|\phi\|$. □

For the complex version of the Hahn–Banach Theorem we require our bounding functional p to be at least a seminorm, i.e. $p\colon X \to [0,\infty)$; $p(\lambda x) = |\lambda| p(x)|$ for all $x \in X$, $\lambda \in \mathbb{K}$; $p(x+y) \le p(x) + p(y)$.

Theorem 4.6 (Seminorm Hahn–Banach Theorem) *Let X be a real or complex vector space, U a subspace of X, and p a seminorm on X. Suppose that $f\colon U \to \mathbb{K}$ is linear and satisfies*

$$|f(x)| \le p(x) \qquad \text{for all} \quad x \in U.$$

Then there exists a linear map $F\colon X \to \mathbb{K}$ such that $F(x) = f(x)$ for all $x \in U$ and

$$|F(x)| \le p(x) \qquad \text{for all} \quad x \in X.$$

In particular, if X is a normed space then any $f \in U^*$ can be extended to some $F \in X^*$ with $\|F\|_{X^*} = \|f\|_{U^*}$.

Proof There exists $\phi \in (U_{\mathbb{R}})^*$ such that $f(v) = \phi(v) - \mathrm{i}\phi(\mathrm{i}v)$ as in Lemma 4.5, and then

$$|\phi(w)| \le p(w)$$

for all $w \in U_{\mathbb{R}}$. We can now use the real Hahn–Banach Theorem to extend ϕ from $U_{\mathbb{R}}$ to $\Phi\colon X_{\mathbb{R}} \to \mathbb{R}$, and then define

$$F(u) = \Phi(u) - \mathrm{i}\Phi(\mathrm{i}u),$$

which provides an extension of f.

To show that F is linear, it is clear that $F(u+v) = F(u) + F(v)$, since Φ has this property. We therefore only need to show that $F(\lambda u) = \lambda F(u)$ for $\lambda \in \mathbb{C}$; taking $\lambda = \alpha + \mathrm{i}\beta$ we have

$$\begin{aligned}
F(\lambda u) &= F(\alpha u + \mathrm{i}\beta u) \\
&= \Phi(\alpha u + \mathrm{i}\beta u) - \mathrm{i}\Phi(\mathrm{i}\alpha u - \beta u) \\
&= \alpha\Phi(u) + \beta\Phi(\mathrm{i}u) - \alpha\mathrm{i}\Phi(\mathrm{i}u) + \beta\mathrm{i}\Phi(u) \\
&= \alpha[\Phi(u) - \mathrm{i}\Phi(\mathrm{i}u)] + \mathrm{i}\beta[\Phi(u) - \mathrm{i}\Phi(\mathrm{i}u)] \\
&= [\alpha + \mathrm{i}\beta]F(u)
\end{aligned}$$

as required.

To show that $|F(x)| \le p(x)$ we use the same trick as in the proof of Lemma 4.5. Suppose that

$$|F(x)| = \mathrm{e}^{\mathrm{i}\theta} F(x);$$

then

$$|F(x)| = F(e^{i\theta}x) = \Phi(e^{i\theta}x) - i\Phi(ie^{i\theta}x),$$

and since $|F(x)|$ is real we have

$$|F(x)| = \Phi(e^{i\theta}x) \leq |\Phi(e^{i\theta}x)| \leq p(e^{i\theta}x) = |e^{i\theta}|p(x) = p(x). \qquad \square$$

5
Separation theorems

We now want to apply the Hahn–Banach Theorem to obtain a geometric 'separation theorem'.

We will need to following lemma for the proof.

Lemma 5.1 *If C is an open convex subset of a Banach space X with $0 \in C$ we define the* Minkowski functional of C *by setting*

$$p_C(x) = \inf\{\lambda > 0 : \ \lambda^{-1}x \in C\} \qquad \textit{for each } x \in X.$$

Then p_C is a sublinear functional on X,

$$C = \{x : \ p_C(x) < 1\}$$

and there exists a constant $c > 0$ such that

$$0 \leq p_C(x) \leq c\|x\| \qquad \textit{for every } x \in X. \tag{5.1}$$

Proof To see that p_C is sublinear when C is convex, take $\alpha > p_C(x)$ and $\beta > p_C(y)$; then $\alpha^{-1}x, \beta^{-1}y \in C$, and since C is convex

$$\frac{\alpha}{\alpha+\beta}\,\alpha^{-1}x + \frac{\beta}{\alpha+\beta}\,\beta^{-1}y = \frac{x+y}{\alpha+\beta} \in C.$$

It follows that $p_C(x+y) \leq \alpha + \beta$, and since α, β were arbitrary, we obtain

$$p_C(x+y) \leq p_C(x) + p_C(y)$$

as required.

Since C is open and $0 \in C$, C contains an open ball $B(0, \delta)$ for some $\delta > 0$, and so

$$\|z\| < \delta \qquad \Rightarrow \qquad z \in C \qquad \Rightarrow \qquad |p_C(z)| \leq 1$$

and then (5.1) follows since for any $x \in C$ we can consider

$$z = \frac{\delta}{2} \frac{x}{\|x\|}.$$

Finally, if $x \in C$ then since C is open we have $\lambda^{-1}x \in C$ for some $\lambda < 1$, and so $p_C(x) \leq \lambda < 1$; while if $p_C(x) < 1$ then $\lambda^{-1}x \in C$ for some $\lambda < 1$, and since $0 \in C$ and C is convex, $x = \lambda(\lambda^{-1}x) + (1-\lambda)0 \in C$. □

Theorem 5.2 (Functional separation theorem) *Suppose that X is a real Banach space and $A, B \subset X$ are non-empty, disjoint, convex sets.*

(i) *If A is open then there exist $f \in X^*$ and $\gamma \in \mathbb{R}$ such that*

$$f(a) < \gamma \leq f(b) \qquad a \in A,\ b \in\ B.$$

(ii) *If A is compact and B is closed then there exist $f \in X^*$, $\gamma \in \mathbb{R}$, and $\delta > 0$ such that*

$$f(a) \leq \gamma - \delta < \gamma + \delta \leq f(b), \qquad a \in A,\ b \in B.$$

A simple example of case (ii) is when $A = \{a\}$ is a point and B is closed.

Proof (i) Choose $a_0 \in A$ and $b_0 \in B$, and let $w_0 = b_0 - a_0$. Now consider

$$C = w_0 + A - B.$$

Then C is an open convex set that contains 0. Since $A \cap B = \varnothing$, $w_0 \notin C$, and so $p_C(w_0) \geq 1$.

Let $W = \mathrm{Span}(w_0)$, and define a linear functional f_W on W by setting

$$f_W(\alpha w_0) = \alpha, \qquad \alpha \in \mathbb{R}.$$

If $\alpha \geq 0$ then

$$f_W(\alpha w_0) = \alpha \leq \alpha p_C(w_0) = p_C(\alpha w_0)$$

while if $\alpha < 0$ then

$$f_W(\alpha w_0) < 0 \leq p_C(\alpha w_0),$$

and so $f_W(w) \le p(w)$ for every $w \in W$.

We can therefore use the general Hahn–Banach Theorem to find a linear extension $f\colon X \to \mathbb{R}$ such that

$$f(x) \le p_C(x) \qquad \text{for every } x \in X.$$

Since we have (5.1) this f satisfies

$$f(x) \le p_C(x) \le c\|x\|.$$

Since f is linear and p_C is sublinear,

$$-f(x) = f(-x) \le p_C(-x) = p_C(x) \le c\|x\|,$$

and so $|f(x)| \le c\|x\|$, i.e. f is an element of X^*.

By definition for any $a \in A$ and $b \in B$ we have $w_0 + a - b \in C$, and so

$$1 + f(a) - f(b) \quad = \quad f(w_0 + a - b) \le p_C(w_0 + a - b) \quad < \quad 1.$$

This shows that $f(a) < f(b)$, and so if we define $\gamma = \inf_{b \in B} f(b)$ we obtain

$$f(a) \le \gamma \le f(b) \qquad a \in A,\ b \in B. \tag{5.2}$$

To guarantee that the left-hand inequality is in fact strict, suppose not, i.e. that there exists an $a \in A$ such that $f(a) = \gamma$. Since A is open we must have $a + \delta w_0 \in A$ for some $\delta > 0$, and then we would have

$$f(a + \delta w_0) = f(a) + \delta f_W(w_0) = \gamma + \delta > \gamma,$$

which contradicts (5.2).

Note that if A and B are both open then the same argument shows that

$$f(a) < \gamma < f(b), \qquad a \in A,\ b \in B.$$

To prove (ii) we set

$$\epsilon = \frac{1}{4} \inf\{\|a - b\| :\ a \in A,\ b \in B\} > 0$$

and consider the two open convex sets

$$A_\epsilon := A + B(0, \epsilon) \qquad \text{and} \qquad B_\epsilon := B + B(0, \epsilon).$$

We can now apply part (i) to find $f \in X^*$ and $\gamma \in \mathbb{R}$ such that

$$f(a) < \gamma \leq f(b), \qquad a \in A_\epsilon,\ b \in B_\epsilon.$$

If we let $\delta = \epsilon/2\|w_0\|$ then for any $a \in A$ we have $a + \delta w_0 \in A_\epsilon$, and so

$$f(a) = f(a + \delta w_0) - \delta f_W(w_0) \leq \gamma - \delta$$

and similarly $\gamma + \delta \leq f(b)$ for any $b \in B$. □

We can use this to give a characterisation of convex subsets of X that will be useful later.

Corollary 5.3 *Suppose that C is a closed convex subset of X. Then*

$$C = \{x : \ f(x) \geq \inf_{y \in C} f(y) \text{ for every } f \in X^*\}.$$

Proof That C is contained in the right-hand side is immediate.

Suppose that $x_0 \notin C$. Then $\{x_0\}$ is compact and convex, so we can find $f \in X^*$ such that

$$f(x_0) \leq \gamma - \delta < \gamma + \delta \leq f(y), \qquad \text{for every } y \in C.$$

In particular $f(x_0) < \inf_{y \in C} f(y)$. □

We can recast these results in a more geometric form using the following simple lemma.

Definition 5.4 *A* hyperplane *U in X is a codimension 1 subspace of X, i.e. a maximal proper subspace: $U \neq X$ and if Z is a subspace with $U \subset Z \subset X$ then $Z = U$ or $Z = X$.*

Lemma 5.5 *The following are equivalent:*

(i) *U is a hyperplane in X;*
(ii) *U is a subspace of X with $U \neq X$ but for any $x \in X \setminus U$, the span of $(U, \{x\})$ is X; and*
(iii) *$U = \mathrm{Ker}(\phi)$ for some non-zero linear functional ϕ on X (which may not be bounded).*

For the proof see Examples 3.

Lemma 5.6 *If $Y = \mathrm{Ker}(\phi)$ is a hyperplane in X then Y is closed if and only if ϕ is bounded. Y is dense in X if and only if ϕ is unbounded.*

Proof First, note that since $Y \neq X$ then it cannot be both closed and dense. So it is enough to show that bounded implies closed, and unbounded implies dense.

If ϕ is bounded then it is continuous, so $Y = \mathrm{Ker}(\phi) = \phi^{-1}\{0\}$ is closed.

If ϕ is unbounded then we can find $(x_n) \in X$ such that $\|x_n\| = 1$ but $\phi(x_n) \geq n$. Now given $x \in X$, consider the sequence

$$y_n = x - \frac{\phi(x)}{\phi(x_n)} x_n.$$

Then $\phi(y_n) = 0$, so $y_n \in Y$, and

$$\|x - y_n\| = \left\| \frac{\phi(x)}{\phi(x_n)} x_n \right\| = \frac{|\phi(x)| \|x_n\|}{|\phi(x_n)|} \leq \frac{|\phi(x)|}{n},$$

and so $y_n \to x$ and $n \to \infty$ and Y is dense. □

Corollary 5.7 *Suppose that A, B are non-empty convex subsets of X with A closed and B compact. Then there exists a closed hyperplane that can be translated so that it separates A and B.*

6

The second dual and reflexivity

We have seen that $(\ell^q)^* \simeq \ell^p$ for $1 \le p < \infty$ and (p,q) conjugate. It follows that if we take the 'second dual' (i.e. the dual of the dual) then $[(\ell^q)^*]^* \simeq (\ell^p)^* \simeq \ell^q$ - we get back to where we started by taking the dual twice. We now investigate in what sense this is always true. [We will see that things are a little more subtle than this; for X to be 'reflexive' we will require that $X^{**} \simeq X$ using a particular linear isometry.]

We write X^{**} for the space $(X^*)^*$, i.e. $X^{**} = B(X^*; \mathbb{K})$.

Lemma 6.1 *For any normed space X we can isometrically map X onto a subspace of X^{**} via the canonical linear mapping $x \mapsto x^{**}$, where x^{**} is the element of X^{**} defined by setting*

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

We denote this mapping by $\Lambda_X \colon X \to X^{**}$.

Proof We have to show that for any $x \in X$, x^{**} defines a linear functional on X^* (i.e. an element of X^{**}) with the same norm as x. Given $x \in X$ we set

$$x^{**}(f) := f(x) \qquad \text{for every } f \in X^*.$$

Then since

$$|x^{**}(f)| = |f(x)| \le \|f\|_{X^*}\|x\|_X$$

it certainly follows that $x^{**} \in X^{**}$ and that $\|x^{**}\|_{X^{**}} \le \|x\|_X$.

If we take the element f from Corollary 3.3, for which $\|f\| = 1$ and

$f(x) = \|x\|$ then we have

$$|x^{**}(f)| = |f(x)| = \|x\|_X = \|x\|_X \|f\|_{X^*}$$

(since $\|f\|_{X^*} = 1$) and it follows that $\|x^{**}\| \geq \|x\|_X$ which yields the required equality of norms. □

In general Λ_X does not map X *onto* X^{**}.

6.1 Reflexive spaces

If $x \mapsto x^{**}$ maps X onto X^{**} then X is called reflexive, and in this case it is easy to see that $X \simeq X^{**}$. We will see shortly that there are some key results that are only true in reflexive spaces.

Definition 6.2 *A Banach space X is* reflexive *if $\Lambda_X : X \to X^{**}$ is onto, i.e. if every $F \in X^{**}$ can be written as x^{**} for some $x \in X$.*

Note that the definition of reflexivity is *not* that $X^{**} \simeq X$. There are spaces whose second dual are isometrically isomorphic to X for whose Λ_X is not an isometric isomorphism and *these spaces or not reflexive.*

The following result is very useful; its proof is a good exercise in using the definition of reflexivity.

Theorem 6.3 *Let X be a Banach space. Then X is reflexive if and only if X^* is reflexive.*

Proof Suppose that X is reflexive; we want to show that X^* is reflexive, i.e. that for any $\Phi \in (X^*)^{**} = (X^{**})^*$ we can find an $f \in X^*$ such that "$f^{**} = \Phi$", i.e. such that

$$\Phi(F) = F(f) \qquad \text{for every } F \in X^{**}.$$

Choose as a candidate $f \in X^*$ the functional defined by

$$f(x) = \Phi(x^{**}),$$

where $x^{**} = \Lambda_X(x)$. Clearly f is linear; it is bounded since

$$|f(x)| = |\Phi(x^{**})| \leq \|\Phi\|_{X^{***}} \|x^{**}\|_{X^{**}} = \|\Phi\|_{X^{***}} \|x\|_X.$$

We need to check that $f^{**} = \Phi$.

Since X is reflexive, for any $F \in X^{**}$ we have $F = x^{**}$ for some $x \in X$, and so

$$f^{**}(F) = f^{**}(x^{**}) = x^{**}(f) = f(x) = \Phi(x^{**}) = \Phi(F).$$

For the converse, suppose that X^* is reflexive but X is not, i.e. there is an element $F \in X^{**}$ such that $F \neq x^{**}$ for any $x \in X$. Then the set

$$D = \{x^{**} : \ x \in X\}$$

is a proper closed linear subspace of X^{**}, and hence by by Corollary 3.5 there is a non-zero $\Phi \in X^{***} = (X^{**})^*$ such that

$$\Phi(x^{**}) = 0 \qquad \text{for all } x \in X.$$

Since X^* is reflexive, we know that $\Phi = f^{**}$ for some $f \in X^*$, and so if $x \in X$ we have

$$f(x) = x^{**}(f) = f^{**}(x^{**}) = \Phi(x^{**}) = 0.$$

But this means that $f = 0$, which implies that $\Phi = 0$, a contradiction. $\square$

Since any reflexive space satisfies $X^{**} \simeq X$ (but not vice versa, as commented above), and we know that

$$(c_0)^* \simeq \ell^1 \qquad \text{and} \qquad (\ell^1)^* \simeq (\ell^\infty)$$

the space c_0 is not reflexive. [We know that $c_0 \not\simeq \ell^\infty$ because c_0 is separable and ℓ^∞ is not.] Therefore ℓ^1 is not reflexive, hence ℓ^∞ is not reflexive. L^1 and L^∞ are not reflexive either.

6.2 Some examples of reflexive spaces

We now show that all Hilbert spaces are reflexive, and so are ℓ^p spaces for $1 < p < \infty$. We start with Hilbert spaces.

Proposition 6.4 *All Hilbert spaces are reflexive.*

Proof Take $F \in H^{**}$ we need to find $x \in H$ such that, for every $f \in H^*$ we have

$$f(x) = F(f).$$

We know that the map $R : H \to H^*$ given by $x \mapsto L_x$ where

$$L_x(y) = (y, x) \qquad y \in H$$

is a linear (anti-linear in the complex case) isometric isomorphism. Given any $f \in H^*$ we can write

$$f(y) = (y, R^{-1}f).$$

Now, $F \circ R\colon H \to \mathbb{K}$ is a bounded antilinear map, so $\overline{F \circ R}\colon H \to \mathbb{K}$, defined by setting

$$\overline{F \circ R}(x) = \overline{F \circ R(x)}$$

is a bounded linear map, i.e. another element of H^*. So we can find an element $y \in H$ such that

$$\overline{(F \circ R)(x)} = (x, y)$$

for every $x \in H$. Since R is a bijection, for any $g \in H^*$ we can choose $x = R^{-1}g$ and then

$$\overline{F(g)} = (R^{-1}g, y) = \overline{(y, R^{-1}g)} = \overline{g(y)},$$

so $F(g) = g(y)$ as required. □

Proposition 6.5 *The sequence space ℓ^p is reflexive if $1 < p < \infty$.*

Proof Take $F \in (\ell^p)^{**}$. We need to find $y \in \ell^p$ such that, for every $f \in (\ell^p)^*$ we have

$$F(f) = f(y).$$

Now, we know that

$$T_q\colon \ell^q \to (\ell^p)^*,$$

defined by setting

$$[T_q(x)](y) = x \cdot y$$

for $y \in \ell^p$, is a linear isometric isomorphism. So given any $f \in (\ell^p)^*$ we can write

$$f(y) = T_q^{-1}(f) \cdot y.$$

Now, $F \circ T_q \colon \ell^q \to \mathbb{K}$ is both linear and bounded; so $F \circ T_q \in (\ell^q)^*$. We can therefore find $y \in \ell^p$ such that

$$(F \circ T_q)(x) = y \cdot x$$

for all $x \in \ell^q$. Since $T_p \colon \ell^p \to (\ell^q)^*$ is a bijection, for any $g \in (\ell^p)^*$ we can choose $x = T_p^{-1} g$, and then

$$F(g) = g(y).$$

This shows that $F = y^{**}$, and so ℓ^p is reflexive. □

7
Linear maps between Banach spaces

7.1 The Baire Category Theorem

We now recall the Baire Catgeory Theorem (from Metric Spaces); we will give a number of interesting applications in functional analysis.

1st version: the countable intersection of large sets is still large.

Theorem 7.1 (Baire Category Thereom) *If $\{G_i\}_{i=1}^{\infty}$ is a countable family of dense open subsets of a complete metric space (X,d) then*

$$G = \bigcap_{i=1}^{\infty} G_i$$

is dense in X.

A countable union of open dense sets (G in the theorem) is called *residual*. Note that the union of a countable collection of residual sets is still residual

An alternative formulation is perhaps a little less intuitive. We say that a subset W and (X,d) is *nowhere dense* if the closure of W contains no open sets. Observe that if W is nowhere dense then $X \setminus \bar{W}$ is open and dense: that this set is open is clear; if it were not dense there would be a point $x \in X \setminus \bar{W}$ such that $B(x,r) \cap X \setminus \bar{W} = \varnothing$ for all r sufficiently small, which would imply that $\bar{W} \supset B(x,r)$, a contradiction.

2nd version: the countable union of small sets cannot be everything.

Corollary 7.2 *Let $\{F_j\}_{j=1}^{\infty}$ be a countable collection of nowhere dense subsets of a complete metric space (X,d). Then*

$$\bigcup_{j=1}^{\infty} F_j \neq X.$$

A countable union of closed nowhere dense subsets is called *meagre*.

Proof The sets $X \setminus \bar{F}_j$ are a countable collection of open dense sets. It follows that

$$\bigcap_{j=1}^{\infty} X \setminus \bar{F}_j = X \setminus \left\{ \bigcup_{j=1}^{\infty} \bar{F}_j \right\}$$

is dense, and in particular non-empty. □

7.2 The Principle of Uniform Boundedness

Theorem 7.3 *Let X be a Banach space and Y a normed space. Let $S \subset B(X,Y)$ be a collection of linear operators such that*

$$\sup_{T \in S} \|Tx\|_Y < \infty \qquad \text{for each} \quad x \in X.$$

Then

$$\sup_{T \in S} \|T\| < \infty.$$

Be careful applying this theorem. You need to know that each element of S is a bounded linear map from X to Y (this may sound obvious but it is easy to slip up).

Proof Consider the sets

$$F_j = \{x \in X : \ \|Tx\|_Y \leq j \text{ for all } T \in S\}.$$

Then by assumption

$$X = \bigcup_j F_j.$$

Corollary 7.2 (the BCT, essentially) implies that at least one of the F_j

is not nowhere dense; since all the F_j are closed, this means that at least one F_j must contain an open set; so there must exist $y \in X$ and $r > 0$ such that $B_r(y) \subset F_n$ for some n.

Then for any x with $\|x\| < r$ – for which $y + x \in B_r(y) \subset F_n$ – for every $T \in S$ we must have

$$\|Tx\| = \|T(y + x) + T(-y)\| \leq n + \|Ty\| \leq 2n$$

for some $R > 0$, since $y \in F_n$. So for any x with $\|x\| = r/2$ we have

$$\|Tx\| \leq 2n \qquad \text{for every } T \in S.$$

Since T is linear we can write any $y \in X$ as $y = (2\|y\|/r)(ry/2\|y\|)$, and then

$$\|Ty\| = \frac{2\|y\|}{r} \left\| T \frac{ry}{2\|y\|} \right\| \leq \frac{R}{r}\|y\|.$$

and the conclusion follows. □

Corollary 7.4 (PUB reformulated) *Suppose that $T_n \in B(X, Y)$ and $\|T_n\|$ is unbounded. Then there exists $x \in X$ such that $\|T_n x\|$ is unbounded.*

Proof If not then the PUB shows that $\|T_n\|$ is bounded. □

Corollary 7.5 *Suppose that X and Y are Banach spaces and that $T_n \in B(X, Y)$. Suppose that*

$$Tx := \lim_{n\to\infty} T_n x$$

exists for every $x \in X$. Then $T \in B(X, Y)$.

Proof It is easy to check that T is linear. The only point is to show that it is bounded. Since

$$\lim_{n\to\infty} \|T_n x\|$$

exists it follows that for every $x \in X$ the set $\{Tx_n\}$ is bounded. The PUB now show that $\|T_n\| \leq M$ for every $n \in \mathbb{N}$. It follows that

$$\|Tx\| = \lim_{n\to\infty} \|T_n x\| \leq M\|u\|$$

and so T is bounded. □

7.2.1 Fourier series of continuous functions

We prove that there is a 2π-periodic continuous function $f\colon[-\pi,\pi]\to\mathbb{R}$ such that the Fourier series of f at 0 does not converge, i.e. the partial sums are unbounded.

The nth partial sum is

$$f_n(x)=\frac{1}{2\pi}\sum_{k=-n}^{n}\left(\int_{-\pi}^{\pi}f(t)\mathrm{e}^{\mathrm{i}kt}\,\mathrm{d}t\right)\mathrm{e}^{-\mathrm{i}kx}.$$

At $x=0$ this gives

$$\begin{aligned}f(0)&=\frac{1}{2\pi}\sum_{k=-n}^{n}\left(\int_{-\pi}^{\pi}f(t)\mathrm{e}^{\mathrm{i}kt}\,\mathrm{d}t\right)\\&=\frac{1}{2\pi}\int_{-\pi}^{\pi}f(t)\left(\sum_{k=-n}^{n}\mathrm{e}^{\mathrm{i}kt}\,\mathrm{d}t\right)\,\mathrm{d}t.\end{aligned}$$

The kernel is given by

$$\begin{aligned}K_n(t)&:=\sum_{k=-n}^{n}\mathrm{e}^{\mathrm{i}kt}\\&=\mathrm{e}^{-\mathrm{i}nt}(1+\cdots+\mathrm{e}^{2\mathrm{i}nt})\\&=\mathrm{e}^{-\mathrm{i}nt}\,\frac{\mathrm{e}^{\mathrm{i}(2n+1)t}-1}{\mathrm{e}^{\mathrm{i}t}-1}\\&=\frac{\mathrm{e}^{\mathrm{i}(n+\frac{1}{2})t}-\mathrm{e}^{-\mathrm{i}(n+\frac{1}{2})t}}{\mathrm{e}^{\frac{1}{2}\mathrm{i}t}-\mathrm{e}^{-\frac{1}{2}\mathrm{i}t}}\\&=\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t},\end{aligned}$$

and so

$$f_n(0)=\frac{1}{2\pi}\int_{-\pi}^{\pi}f(t)\,\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}\,\mathrm{d}t.$$

Let

$$P=\{f\in C^0([-\pi,\pi]):\ f(-\pi)=f(\pi)\}$$

and consider the map $S_n\colon P\to\mathbb{R}$ given by $f\mapsto f_n(0)$.

Using Corollary 7.4 it is enough to show that $\|S_n\|$ is unbounded:

if it is then there must exist an $f \in P$ such that $|S_n f| = |f_n(0)|$ is unbounded.

We do this by showing that

$$\|S_n\| = I_n := \int_{-\pi}^{\pi} \left| \frac{\sin(n+\frac{1}{2})t}{\sin \frac{1}{2}t} \right| \mathrm{d}t$$

and that the integral is unbounded (in n).

Clearly $\|S_n\| \le I_n$. To get equality we would like to take $f = \mathrm{sign} K_n(t)$, but this is not an element of P. This can be overcome by approximating $\mathrm{sign} K_n(t)$ by a sequence of elements of P, see Examples 5.

To estimate I_n from below observe that $|\sin(t/2)| \le |t/2|$, and so

$$\left| \frac{\sin(n+\frac{1}{2})t}{\sin \frac{1}{2}t} \right| \ge \left| \frac{\sin(n+\frac{1}{2})t}{\frac{1}{2}t} \right|.$$

We have

$$\left| \frac{\sin(n+\frac{1}{2})t}{\frac{1}{2}t} \right| \ge \left| \frac{\sin(n+\frac{1}{2})t}{\frac{1}{2}k\pi/(n+(1/2))} \right| \qquad t \in \left[\frac{(k-1)\pi}{n+(1/2)}, \frac{k\pi}{n+(1/2)} \right].$$

So

$$\begin{aligned}
\int_{-\pi}^{\pi} |K_n(t)| &\ge \sum_{k=1}^{n} \int_{(k-1)\pi/(n+(1/2))}^{k\pi/(n+(1/2))} \left| \frac{\sin(n+\frac{1}{2})t}{\frac{1}{2}k\pi/(n+(1/2))} \right| \\
&= \sum_{k=1}^{n} \frac{2n+1}{k\pi} \int_{0}^{\pi/(n+1/2)} \sin(n+(1/2))t \,\mathrm{d}t \\
&= \sum_{k=1}^{n} \frac{1}{k\pi} \int_{0}^{\pi} \sin t \,\mathrm{d}t = c \sum_{k=1}^{n} \frac{1}{k} \sim c \log n.
\end{aligned}$$

8
Weak and weak-$*$ convergence and the Banach–Alaoglu Theorem

8.1 Weak convergence

Definition 8.1 *We say that* $\{x_n\} \in X$ *converges weakly to* $x \in X$*, and write* $x_n \rightharpoonup x$*, if*

$$f(x_n) \to f(x) \qquad \textit{for all} \qquad f \in X^*.$$

Note that in a Hilbert space, where every linear functional is of the form $x \mapsto (x, y)$ for some $y \in H$, $x_n \rightharpoonup x$ if

$$(x_n, y) \to (x, y) \qquad \text{for all} \qquad y \in H.$$

This provides an easy example of a sequence that converges weakly but does not converge; pick a countable orthonormal set $\{e_j\}_{j=1}^{\infty}$. Then for any $y \in H$ Bessel's inequality

$$\sum_{j=1}^{\infty} |(y, e_j)|^2 \leq \|y\|^2$$

shows that the sum converges; it follows that $(y, e_j) \to 0$ as $j \to \infty$, and hence that $e_j \rightharpoonup 0$. But the sequence $\{e_j\}$ does not converge (any two elements are a distance $\sqrt{2}$ apart).

Lemma 8.2

(i) Strong convergence implies weak convergence;
(ii) weak limits are unique;
(iii) weakly convergent sequences are bounded; and

(iv) if $x_n \rightharpoonup x$ *then*

$$\|x\| \leq \liminf_{n\to\infty} \|x\|_n.$$

Proof (i) If $x_n \to x$ then for any $f \in X^*$

$$|f(x_n) - f(x)| \leq \|f\|_{X^*}\|x_n - x\|_X \to 0 \qquad \text{as} \qquad n \to \infty,$$

so $f(x_n) \to f(x)$ and hence $x_n \rightharpoonup x$.

(ii) Suppose that $x_n \rightharpoonup x$ and $x_n \rightharpoonup y$. Then for any $f \in X^*$, $f(x) = \lim_{n\to\infty} f(x_n) = f(y)$. So by Lemma 3.4, $x = y$.

(iii) We consider the sequence $x_n^{**} \in X^{**}$. Then for every $f \in X^*$ we know that

$$f(x_n) = x_n^{**}(f)$$

converges, and so is bounded. It follows from the PUB that (x_n^{**}) is bounded in X^{**}. Since $\|x^{**}\|_{X^{**}} = \|x\|_X$ it follows that (x_n) is bounded in X.

(iv) Choose $f \in X^*$ with $\|f\|_{X^*} = 1$ such that $f(x) = \|x\|$. Then

$$\|x\| = f(x) = \lim_{n\to\infty} f(x_n),$$

so

$$\|x\| \leq \liminf_{n\to\infty} |f(x_n)| \leq \liminf_{n\to\infty} \|f\|_{X^*}\|x_n\|_X;$$

the result follows since $\|f\|_{X^*} = 1$. □

How can we convert weak convergence to strong convergence? In a Hilbert space weak convergence plus norm convergence implies strong convergence. This is true in a larger class of Banach spaces, but the proof in a Hilbert space is simple.

Lemma 8.3 *Let* H *be a Hilbert space. If* $x_n \rightharpoonup x$ *and* $\|x_n\| \to \|x\|$ *then* $x_n \to x$.

Proof Observe that

$$\|x - x_n\|^2 = (x - x_n, x - x_n) = \|x\|^2 - 2(x, x_n) + \|x_n\|^2.$$

Since $x_n \rightharpoonup x$ we have $\|x\| \leq \liminf_{n\to\infty} \|x_n\|$ and $(x, x_n) \to \|x\|^2$. □

The following result is often useful.

Lemma 8.4 *Suppose that $T : X \to Y$ is a compact linear operator. Then if $x_n \rightharpoonup$ in X, $Tx_n \to Tx$ in Y.*

Proof First observe that $Tx_n \rightharpoonup Tx$ in Y; indeed if $f \in Y^*$ then observe that $f \circ T$ is is an element of X^*, so that $x_n \rightharpoonup x$ implies that

$$f(Tx_n) \to f(Tx).$$

Now, if $Tx_n \not\to Tx$ then there is an $\epsilon > 0$ and a subsequence (which we relabel) such that $\|Tx_n - Tx\| > \epsilon$ for every j. Since x_n is a bounded sequence in X and T is compact, $\{Tx_n\}$ has a subsequence (which we relabel again) that converges to some $z \in Y$. If $Tx_n \to z$ then certainly $Tx_n \rightharpoonup z$; but weak limits are unique, and so $z = Tx$, a contradiction. □

Corollary 8.5 *Suppose that X, Y are Banach spaces, with Y compactly embedded in X. Then if $y_n \rightharpoonup y$ in Y, $y_n \to y$ in X.*

Proof The identity map $T \colon Y \to X$ is compact. □

Weak convergence in ℓ^p, $1 \le p < \infty$:

$$x^{(n)} \rightharpoonup x \qquad \Leftrightarrow \qquad (x^{(n)}, y) \to (x, y) \quad \text{for every } y \in \ell^q,$$

(p, q) conjugate. For $1 < p < \infty$ there is a nice characterisation. For an example showing that the following result is not true in ℓ^1 see Examples 5.

Lemma 8.6 *Let (x_n) be a sequence in ℓ^p for $1 < p < \infty$. Then*

$$x^{(n)} \rightharpoonup x \qquad \Leftrightarrow \qquad x_k^{(n)} \to x_k \ \forall\ k \quad \textit{and} \quad \|x^{(n)}\| \textit{ is bounded.}$$

Proof $\Rightarrow$ This taking $y = e_k$, and then using the fact that any weakly convergent sequence is bounded.

$\Leftarrow$ Suppose that $\|x^{(n)}\| \le M$. Take any $y \in \ell^q$. Then

$$y = \lim_{k\to\infty} \sum_{j=1}^{k} y_j e_j.$$

Given any $\varepsilon > 0$ there exists k such that

$$\left\| y - \sum_{j=1}^{k} y_j e_j \right\|_{\ell^p} < \frac{\varepsilon}{4M};$$

then

$$\begin{aligned}|(x^{(n)} - x, y)| &= |(x^{(n)} - x, \sum_{j=1}^{k} y_j e_j) + (x^{(n)} - x, y - \sum_{j=1}^{k} y_j e_j)| \\ &\leq |(x^{(n)} - x, \sum_{j=1}^{k} y_j e_j)| + \|x^{(n)} - x\| \|y - \sum_{j=1}^{k} y_j e_j\| \\ &\leq \sum_j |y_j| |(x^{(n)} - x, e_j)| + \frac{\varepsilon}{2}.\end{aligned}$$

Since $(x^{(n)}, e_j) \to (x, e_j)$ for each j it follows that $x^{(n)} \rightharpoonup x$. □

Weak convergence in $X = C^0([0,1])$*:* the maps $f \mapsto f(x)$ for any fixed $x \in [0,1]$ is an element of X^*. If $f_n \rightharpoonup f$ then $f_n(x) \to f(x)$ for every $x \in [0,1]$ (pointwise convergence). Take any subinterval $[a,b] \subset [0,1]$ then

$$f \mapsto \int_a^b f(x)\,\mathrm{d}x$$

is an element of X^*. So

$$\int_a^b f_n(x)\,\mathrm{d}x \to \int_a^b f(x)\,\mathrm{d}x$$

for any weakly convergent sequence. (This can be used to show that pointwise convergence is weaker than weak convergence.)

8.2 Weak closures

We set that a subset A of X is *weakly closed* if whenever $x_n \in A$ and $x_n \rightharpoonup x$, we have $x \in A$. In general this is a stronger property than being closed: weakly closed implies closed but not vice versa. For example, the unit sphere in a separable Hilbert space H is closed but not weakly closed, since we can take a countable orthonormal set (e_j); then $e_j \in S_H$ but $e_j \rightharpoonup 0$ and $0 \notin S_H$.

However, for convex subsets being closed and being weakly closed are the same.

Proposition 8.7 *Closed convex subsets of a Banach space are also weakly closed.*

Proof Using Corollary 5.3 we know that

$$C = \{x : \; f(x) \geq \inf_{y \in C} f(y) \text{ for every } f \in X^*\}.$$

If $x_n \in C$ and $x_n \rightharpoonup x$ then since $x_n \in C$ we have, for any $f \in X^*$ we have $f(x_n) \geq \inf_{y \in C} f(y)$, and so

$$f(x) = \lim_{n \to \infty} f(x_n) \geq \inf_{y \in C} f(y),$$

i.e. $x \in C$. □

8.3 Weak-∗ convergence

There is another notion of weak convergence, weak-∗ convergence, which deals with sequences of elements of X^*.

Definition 8.8 *If $\{f_n\} \in X^*$ then f_n converges weakly-∗ to f, $f_n \overset{*}{\rightharpoonup} f$ if*

$$f_n(x) \to f(x) \qquad \textit{for all} \qquad x \in X.$$

We have a result along similar lines to Lemma 8.2.

Lemma 8.9

(i) Weak-∗ limits are unique;
(ii) weakly-∗ convergent sequences are bounded;
(iii) weak convergence in X^* implies weak-∗ convergence in X^*;
(iv) if X is reflexive then weak-∗ convergence in X^* implies weak convergence in X^*.

Proof (i) Follows from the definition.

(ii) Follows immediately from the PUB.

(iii) $f_n \rightharpoonup f$ in X^* means that for every $\Phi \in X^{**}$ we have

$$\Phi(f_n) \to \Phi(f).$$

Given any element $x \in X$ we can consider the corresponding $x^{**} \in X^{**}$. Since $f_n \rightharpoonup f$ in X^* we have

$$f_n(x) = x^{**}(f_n) \to x^* * f = f(x),$$

and so $f_n \overset{*}{\rightharpoonup} f$.

(iv) When X is reflexive any $\Phi \in X^{**}$ is of the form x^{**} for some $x \in X$. So if $f_n \overset{*}{\rightharpoonup} f$ in X^* we have

$$\Phi(f_n) = x^{**}(f_n) = f_n(x) \to f(x) = x^{**}(f) = \Phi(f),$$

using the weak-∗ convergence of f_n to f to take the limit. So $f_n \rightharpoonup f$ in X^*. □

8.3.1 Two weak compactness theorems

We now prove two key compactness theorems. We begin with a preparatory lemma.

Lemma 8.10 *Let (f_n) be a bounded sequence in X^* with $\|f_n\|_{X^*} \le M$, and suppose that $f_n(x_k)$ converges for every x_k in a dense subset of X. Then $lim_{n\to\infty} f_n(x)$ exists for every $x \in X$, and the map $f\colon X \to \mathbb{R}$ defined by setting*

$$f(x) = \lim_{n\to\infty} f(x), \qquad \textit{for each} \qquad x \in X$$

is an element of X^* with $\|f\|_{X^*} \le M$.

Proof First notice that if $f_n(x_k)$ converges for every k then $f_n(x)$ converges for every $x \in X$. Indeed, if $\|f_n\|_{X^*} \le M$ then given $\varepsilon > 0$ and $x \in X$ first choose k such that

$$\|x - x_k\|_X \le \varepsilon/3M.$$

Now, using the fact that $f_n(x_k)$ converges as $n \to \infty$, choose n_0 sufficiently large that $\|f_n(x_k) - f_n(x_k)\| < \varepsilon/3$ for all $n, m \ge n_0$. Then for all $n, m \geq n_0$ we have

$$\begin{aligned}
&|f_n(x) - f_m(x)| \\
&\qquad \leq |f_n(x) - f_n(x_k)| + |f_n(x_k) - f_m(x_k)| + |f_m(x_k) - f_m(x)| \\
&\qquad \leq \|f_n\|_{X^*}\|x - x_k\| + \frac{\varepsilon}{3} + \|f_m\|_{X^*}\|x_k - x\| \\
&\qquad \leq \varepsilon.
\end{aligned}$$

It follows that $(f_n(x))$ is Cauchy and hence converges.

We now define $f\colon X \to \mathbb{R}$ by setting

$$f(x) := \lim_{n\to\infty} f_n(x).$$

Then f is linear since

$$f(x + \lambda y) = \lim_{n\to\infty} f_n(x + \lambda y) = \lim_{n\to\infty} f_n(x) + \lambda f_n(y) = f(x) + \lambda f(y)$$

and f is bounded since

$$|f(x)| = \lim_{n\to\infty} |f_n(x)| \leq M\|x\|. \quad \square$$

Using this we can prove a weak-∗ compactness result when X is separability.

Theorem 8.11 *Suppose that X is separable. Then a bounded sequence in X^* has a weakly-∗ convergent subsequence.*

Proof Let $\{x_k\}$ be a countable dense subset of X, and $\{f_j\}$ a sequence in X^* such that $\|f_j\|_{X^*} \leq M$. A standard diagonal argument yields a subsequence of the $\{f_j\}$ (which we relabel) such that $f_j(x_k)$ converges for every k: Since $|f_n(x_1)| \leq M\|x_1\|$, we can use the Bolzano–Weierstrass Theorem to find a subsequence $f_{n_{1,i}}$ such that $f_{n_{1,i}}(x_1)$ converges. Now, since $|f_{n_{1,i}}(x_2)| \leq M\|x_2\|$ we can find a subsequence $f_{n_{2,i}}$ of $f_{n_{1,i}}$ such that $f_{n_{2,i}}(x_1)$ converges and so does $f_{n_{2,i}}(x_2)$. We continue in this way to find successive subsequences $f_{n_{m,i}}$ such that

$$f_{n_{m,i}}(x_k) \qquad \text{converges as } i \to \infty \text{ for every } k = 1, \ldots, m.$$

By taking the diagonal subsequence $f_m^* := f_{n_{m,m}}$ we can ensure that $f_m^*(x_k)$ converges for every $k \in \mathbb{N}$.

The proof concludes using Lemma 8.10. $\square$

A consequence of this is the following extremely useful weak compactness result. In fact the unit ball in X is weakly compact if and only if X is reflexive.

Theorem 8.12 *Let X be a reflexive Banach space. Then any bounded sequence in X has a weakly convergent subsequence.*

Proof Take a bounded sequence $(x_n) \in X$ and let $Y = \overline{\mathrm{Sp}}\{x_1, x_2, \ldots\}$. Then, using Lemma 1.8, Y is separable. Since $Y \subset X$ and X is reflexive, so is Y (see Examples 4). Therefore Y^{**} is separable, and so Y^* is separable (using Lemma 3.6).

Now, x_n^{**} is a bounded sequence in Y^{**}, so there is a subsequence x_{n_k} such that $x_{n_k}^{**}$ is weakly-$*$ convergent in Y^{**} to some limit $\Phi \in Y^{**}$. Since Y is reflexive, $\Phi = x^{**}$ for some $x \in Y \subset X$.

Now for any $f \in X^*$ we have $f_Y := f|_Y \in Y^*$, so

$$\lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} f_Y(x_n) = \lim_{n\to\infty} x_n^{**}(f_Y) = x^{**}(f_Y) = f_Y(x) = f(x)$$

and $x_n \rightharpoonup x$. □

Here is an example of the use of weak compactness and 'approximation' to prove the existence of a fixed point.

Lemma 8.13 *Let X be a reflexive Banach space, Y a Banach space, and $T\colon X \to Y$ a compact linear operator. Suppose that there is a bounded sequence $(x_n) \in X$ such that*

$$\|Tx_n - x_n\| \to 0$$

as $n \to \infty$. Then there exists $x \in X$ such that $Tx = x$.

Proof Since (x_n) is a bounded sequence in a reflexive Banach space it has a weakly convergent subsequence, $x_{n_j} \to x$. Since T is compact, it follows that $Tx_{n_j} \to Tx$ strongly in Y. Since $\lim_{j\to\infty} Tx_{n_j} - x_{n_j}$ it follows that $x_{n_j} \to Tx$. Since strong convergence implies weak convergence we have $x_{n_j} \rightharpoonup Tx$, and since weak limits are unique it follows that $x = Tx$. □

9

Open mapping, inverse mapping, and closed graph theorems

9.1 Open mapping and inverse mapping theorems

We start by proving the Open Mapping Theorem; however, its corollary, the Inverse Mapping Theorem, is perhaps more useful.

Theorem 9.1 (Open mapping theorem) *If $T\colon X \to Y$ is a bounded surjective linear map from a Banach space X into a Banach space Y, then T maps open sets in X to open sets in Y.*

We use B_X for the closed unit ball in X, and $B(x,r)$ for the open ball of radius r around x.

Proof It suffices to show that $T(B_X)$ includes an open ball around 0 in Y, say $B(0,r)$ for some $r > 0$: if U is an open set in X then for any point $x \in U$ there exists an $\delta > 0$ such that $x + \delta B_X \subset U$. Then $T(U) \supset T(x + \delta B_X) = Tx + \delta T(B_X) \supset Tx + \delta B(0,r)$.

First we show that $\overline{T(B_X)}$ contains a ball around 0.

Notice that the closed sets

$$\overline{T(nB_X)} = n\overline{T(B_X)}$$

cover Y (since $T(X) = Y$ as T is surjective). It follows use the BCT that at least one of them contains a ball of positive radius (otherwise they would all be nowhere dense, and then they could not cover Y).

All these sets are scaled copies of $\overline{T(B_X)}$, so $\overline{T(B_X)}$ contains $B_Y(z,r)$

for some $z \in Y$ and some $r > 0$. However, $\overline{T(B_X)}$ is convex and symmetric, and so contains $B_Y(0, r)$.

Now we show that $T(2B_X)$ must include $B_Y(0, r)$. Note that for any $y \in B_Y(0, \alpha r)$, since $\overline{T(\alpha B_X)} \ni y$, for any $\varepsilon > 0$ there exists $x \in \alpha B_X$ such that

$$\|y - Tx\|_Y < \varepsilon.$$

We use this argument repeatedly. Given $u \in B_Y(0, r)$, find $x_1 \in B_X$ with $\|u - Tx_1\|_Y < r/2$. Then, since

$$u - Tx_1 \in B(0, r/2)$$

we can find $x_2 \in \frac{1}{2}B_X$ such that

$$\|(u - Tx_1) - Tx_2\| < r/4.$$

Now find $x_3 \in \frac{1}{4}B_X$ such that

$$\|(u - Tx_1 - Tx_2) - Tx_3\| < r/8,$$

and so on, yielding a sequence (x_n) with $x_n \in 2^{-n}B_X$ such that

$$\left\| u - T\left(\sum_{j=1}^{n} x_j\right) \right\| < r2^{-n}.$$

The sequence $\sum_{j=1}^{n} x_j$ is Cauchy; since X is complete it converges to some $x \in 2B_X$ with $Tx = u$.

Since $T(2B_X) \supset B_Y(0, r)$ it follows that $T(B_X) \supset B_Y(0, r/2)$. □

Corollary 9.2 (Inverse Mapping Theorem) *If $T\colon X \to Y$ is a bounded bijective linear map from a Banach space X onto a Banach space Y, then T^{-1} is bounded.*

Proof T has an inverse since it is bijective. The inverse is linear: the inverse is uniquely defined, and

$$T(\alpha T^{-1}(y_1) + \beta T^{-1}y_2) = \alpha y_1 + \beta y_2 = T(T^{-1}(\alpha y_1 + \beta y_2)),$$

i.e. $T^{-1}(\alpha y_1 + \beta y_2) = \alpha T^{-1}y_1 + \beta T^{-1}y_2$.

By the open mapping theorem $T(B_X)$ includes θB_Y for some $\theta > 0$, so

$$T^{-1}(\theta B_Y) \subset B_X,$$

and so $\|T^{-1}\| \le \theta^{-1}$. □

This has an immediate application in spectral theory. For $T\colon \mathbb{R}^n \to \mathbb{R}^n$ we know that $\lambda \in \sigma(T)$ if and only if λ is an eigenvalue, i.e. $T-\lambda I$ is not one-to-one. In infinite dimensions, the spectrum of $T\colon X \to X$ consists of all those λ for which $T-\lambda I$ does not have a bounded inverse. The inverse mapping theorem tells us that if $T \in B(X)$ and $T-\lambda I$ is one-to-one and onto then necessarily $(T-\lambda I)^{-1}$ is bounded, so $\lambda \notin \sigma(T)$.

The following corollary, concerning equivalence of norms on Banach spaces, follows by considering the identity map from $\mathrm{id}\colon (X, \|\cdot\|_1) \to (X, \|\cdot\|_2)$.

Corollary 9.3 *If X is a Banach space that is complete wrt two different norms $\|\cdot\|_1$ and $\|\cdot\|_2$ and $\|x\|_2 \le C\|x\|_1$, then the two norms are equivalent.*

A quick example before a much longer one: you cannot put the ℓ^1 norm on ℓ^2 and make a complete space: we know that $\|x\|_{\ell^2} \le \|x\|_{\ell^1}$. But $(1, 1/2, 1/3, \dots) \in \ell^2$ and not in ℓ^1, so the norms are not equivalent.

9.2 Bases in Banach spaces

We now use Corollary 9.3 to investigate bases in Banach spaces. Suppose that a Banach space X has a countable basis (e_n) so that any element $x \in X$ can be written uniquely in the form

$$x = \sum_{j=1}^{\infty} x_j e_j,$$

where the sum converges in X. Suppose that we consider the 'truncated' expansions

$$P_n x = \sum_{j=1}^{n} x_j e_j.$$

Can we find a constant C such that $\|P_n x\| \le C\|x\|$ for every n and every $x \in X$?

If we knew that P_n was in $B(X, X)$ for every n (i.e. that for each n

we had $\|P_n x\| \le C_n \|x\|$ for every $x \in X$) this would follow from the Principle of Uniform Boundedness. But this is not so easy to show.

In order to do this, we prove the following.

Proposition 9.4 *Suppose that (e_n) is a countable sequence in a Banach space X with $\|e_n\| = 1$ whose closed linear span is all of X, and which is a basis for X in the sense described above. Then*

$$|||x||| = \sup_n \left\| \sum_{j=1}^n a_j e_j \right\|_X \qquad \textit{when} \quad x = \sum_{j=1}^\infty a_j e_j$$

defines a norm $|||\cdot|||$ on X, and X is complete with respect to this norm.

Proof It is straightforward to show that $|||\cdot|||$ is a norm. The only issue is the triangle inequality, but note that by the triangle inequality for $\|\cdot\|$

$$\| \sum_{i=1}^n (x_i + y_i) e_i \| \le \| \sum_{i=1}^n x_i e_i \| + \| \sum_{i=1}^n y_i e_i \|,$$

so

$$\sup_n \| \sum_{i=1}^n (x_i + y_i) e_i \| \le \sup_n \| \sum_{i=1}^n x_i e_i \| + \sup_n \| \sum_{i=1}^n y_i e_i \|,$$

i.e.

$$|||x + y||| \le |||x||| + |||y|||.$$

To show that $(X, |||\cdot|||)$ is complete is more involved. A key observation, however, is that if $x = \sum_{j=1}^\infty a_j e_j$ then

$$\begin{aligned} |a_j| = \|a_j e_j\| &= \left\| \sum_{j=1}^m a_j e_j - \sum_{j=1}^{m-1} a_j e_j \right\| \\ &\le \left\| \sum_{j=1}^m a_j e_j \right\| + \left\| \sum_{j=1}^{m-1} a_j e_j \right\| \le 2|||x|||. \end{aligned}$$

Now suppose that $x^{(n)}$ is a Cauchy sequence in $(X, |||\cdot|||)$, with

$$x^{(n)} = \sum_{j=1}^\infty a_j^{(n)} e_j :$$

given any $\varepsilon > 0$ there exists an $N = N(\varepsilon)$ such that

$$\left|\left|\left| x^{(n)} - x^{(m)} \right|\right|\right| = \sup_k \left\| \sum_{j=1}^{k} [a_j^{(n)} - a_j^{(m)}] e_j \right\| < \varepsilon$$

for all $n, m \geq N(\varepsilon)$.

It follows that for each j the sequence $(a_j^{(n)})$ is Cauchy, so $a_j^{(n)} \to \alpha_j$ as $n \to \infty$. We now show that $\sum_j \alpha_j e_j$ is convergent to some element $x \in X$ and that $\left|\left|\left| x^{(n)} - x \right|\right|\right| \to 0$ as $n \to \infty$.

Take $n \geq N(\varepsilon)$ and any $k \geq 1$. Then

$$\begin{aligned} \left\| \sum_{j=1}^{k} (a_j^{(n)} - \alpha_j) e_j \right\| &= \left\| \sum_{j=1}^{k} (a_j^{(n)} - \lim_{m \to \infty} a_j^{(m)}) \right\| \\ &= \lim_{m \to \infty} \left\| \sum_{j=1}^{k} (x_j^{(n)} - x_j^{(m)}) e_j \right\| < \varepsilon. \end{aligned} \tag{9.1}$$

We use this to show that the partial sums $\sum_{j=1}^{n} \alpha_j e_j$ are Cauchy in $\|\cdot\|$ and hence converge to some $x = \sum_j \alpha_j e_j$. Given this, note that (9.1) shows that $x^{(n)}$ to x in the norm $|||\cdot|||$.

For each $n \geq N(\varepsilon)$ we know that $\sum_j a_j^{(n)} e_j$ converges in X, so there exists $M(\varepsilon)$ such that if $r > s \geq M(\varepsilon)$ we have

$$\left\| \sum_{i=s}^{r} a_i^{(n)} e_i \right\| < \varepsilon. \tag{9.2}$$

Therefore

$$\begin{aligned} \left\| \sum_{i=s}^{r} \alpha_i e_i \right\| &= \left\| \sum_{i=s}^{r} (a_i^(n) - \alpha_i) e_i \right\| + \left\| \sum_{i=s}^{r} a_i^{(n)} e_i \right\| \\ &\leq \left\| \sum_{i=1}^{r} (a_i^(n) - \alpha_i) e_i - \sum_{i=1}^{s-1} (a_i^(n) - \alpha_i) e_i \right\| + \varepsilon \\ &\leq 3\varepsilon, \end{aligned}$$

using (9.1) and (9.2). □

Corollary 9.5 *If (e_n) satisfies the conditions of Proposition 9.4, then there exists a constant $C > 0$ such that*

$$\left\| \sum_{j=1}^{n} x_j e_j \right\| \leq C\|x\|$$

for every $n \in \mathbb{N}$ and $x \in X$, where $x = \sum_j x_j e_j$.

9.3 The Closed Graph Theorem

Theorem 9.6 (Closed Graph Theorem) *Suppose that $T\colon X \to Y$ is a linear map between Banach spaces and that the graph of T,*

$$G := \{(x, Tx) \in X \times Y : \ x \in X\}$$

is a closed subset of $X \times Y$ (with norm $\|(x,y)\|_{X\times Y} = \|x\|_X + \|y\|_Y$). Then T is bounded.

If the graph G is closed then this means that if $x_n \to x$ and $Tx_n \to y$ then $Tx = y$. Continuity is stronger, since it does not require $Tx_n \to y$.

Proof Note that G is closed linear subspace of $X \times Y$ (since T is linear). It is therefore a Banach space when equipped with the norm of $X \times Y$.

Now consider the projection map $\Pi_X\colon G \to X$, defined by

$$\Pi_X(x, y) = x,$$

which is both linear and bounded. This map is also surjective and one-to-one, since

$$\Pi_X(x, Tx) = \Pi_X(y, Ty) \qquad \Rightarrow \qquad x = y \qquad \Rightarrow \qquad (x, Tx) = (y, Ty).$$

By the IMT the map Π_X^{-1} is bounded (this step requires the fact that G is closed $\Rightarrow$ G is a Banach space). It follows that

$$\|\Pi_X^{-1} x\|_{X\times Y} = \|(x, Tx)\|_{X\times Y} = \|x\|_X + \|Tx\|_Y \leq M\|x\|,$$

and so $\|Tx\|_Y \leq M\|x\|$ as required. □

10
Continuous functions

10.1 "Baire one" functions

One nice application of the Baire Category Theorem is the following.

Theorem 10.1 *Let $f_n : \mathbb{R} \to \mathbb{R}$ be a sequence of continuous functions that converge pointwise to some function $f : \mathbb{R} \to \mathbb{R}$, i.e. $f(x) = \lim_{n\to\infty} f_n(x)$ exists for every x. Then f is continuous at a residual set.*

Before the proof we make the following observation: if U and V are open subsets of $\mathbb{R}$ with $V \cap \bar{U} \neq \emptyset$ then $U \cap V \neq \emptyset$. For any $v \in V \cap \bar{U}$ there exists $\varepsilon > 0$ such that $B(v, \varepsilon) \subset V$, and since $v \in \bar{U}$ there exist $u_n \in U$ such that $u_n \to v$. So $u_n \in V$ for all n sufficiently large.

Proof We show that for any $\delta > 0$ the closed set

$$F_\delta = \{x_0 \in \mathbb{R} : \lim_{\epsilon\to 0} \sup_{x:\ |x-x_0|\leq\epsilon} |f(x) - f(x_0)| \geq 3\delta\}$$

is nowhere dense. From this it follows that

$$\cup_{n\in\mathbb{N}} F_{1/n} = \{\text{discontinuity points of } f\}$$

is the countable union of closed nowhere dense sets, and so its complement – the set of continuity points – is residual.

To show that F_δ is nowhere dense, i.e. that its closure contains no open set, let

$$E_n(\delta) = \{x \in \mathbb{R} : \sup_{i,j\geq n} |f_i(x) - f_j(x)| \leq \delta\}.$$

Note that E_n is closed, $E_{n+1} \supset E_n$, and

$$\mathbb{R} = \bigcup_{n=0}^{\infty} E_n.$$

Choose any open set $U \subset \mathbb{R}$, and consider

$$\overline{U} = \bigcup_{n=0}^{\infty} \overline{U} \cap E_n.$$

Since $\overline{U}$ is a complete metric space, it follows from the Baire Category Theorem that there exists an n such that $\overline{U} \cap E_n$ contains an open set V'. From the remark before the proof, $V := V' \cap U$ is an open subset of $\overline{U} \cap E_n$ that is in addition a subset of U.

Since $V \subset E_n$, it follows that $|f_i(x) - f_j(x)| \leq \delta$ for all $x \in V$ and $i, j \geq n$. Fixing $i = n$ and letting $j \to \infty$ it follows that

$$|f_n(x) - f(x)| \leq \delta \qquad \text{for all} \quad x \in V.$$

Now, since f_n is continuous, for any $x_0 \in V$ there is a neighbourhood $N(x_0) \subset V$ such that

$$|f_n(x) - f_n(x_0)| < \delta \qquad \text{for all} \quad x \in N(x_0).$$

Thus by the triangle inequality

$$|f(x_0) - f(x)| < 3\delta \qquad \text{for all} \quad x \in N(x_0).$$

It follows that no element of $N(x_0)$ belongs to F_δ.

This implies, since $N(x_0) \subset V \subset U$ that U contains an open set that contains no element of F_δ. This shows that F_δ is nowhere dense, which concludes the proof. □

10.2 The Arzelà–Ascoli Theorem

Theorem 10.2 *Let X be a compact metric space. A subset of $C(X;\mathbb{R})$ is compact if and only if it is closed, bounded, and equicontinuous.*

A subset A of $C(X;\mathbb{R})$ is equicontinuous if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$d(x,y) < \varepsilon \qquad \Rightarrow \qquad |f(x) - f(y)| < \varepsilon \quad \text{for every} \quad f \in A.$$

Proof We begin by constructing a countable dense subset of X of a particular form. For each n, we have

$$X \subset \bigcup_{x \in X} B(x, 2^{-n}).$$

Since X is compact, there is a finite subcover, so there are a finite number of points $(x_j^{(n)})_{j=1}^{N_n}$, such that for every $x \in X$ we have $d(x, x_j^{(n)})$ for some j. Put these together to form a sequence (x_k) such that given n, we can guarantee that there is an $N(n)$ such that for every $x \in X$, $d(x, x_k) < 2^{-n}$ for some $k \leq N(n)$.

Now, since (f_j) is bounded we can use a 'standard diagonal argument' to find a subsequence (which we relabel) such that $f_j(x_k)$ converges for every k. We now show that (f_j) must be Cauchy in the sup norm.

Given $\varepsilon > 0$, since (f_j) is equicontinuous there exists a $\delta > 0$ such that

$$|x - y| < \delta \qquad \Rightarrow \qquad |f_j(x) - f_j(y)| < \varepsilon/3$$

for every j.

By our construction of the (x_j) there exists an M such that for every $x \in X$ there exists an x_j with $1 \leq j \leq M$ such that $|x - x_j| < \delta$.

Now for N sufficiently large we can guarantee that

$$|f_n(x_i) - f_m(x_i)| < \varepsilon/3 \qquad n, m \geq N,\ 1 \leq i \leq M.$$

Now we can use the triangle inequality: for $n, m \geq N$, for any $x \in X$ we choose $1 \leq i \leq M$ such that $d(x, x_i) < \delta$, and then

$$\begin{aligned} |f_n(x) - f_m(x)| &= |f_n(x) - f_n(x_j) + f_n(x_j) - f_m(x_j) + f_m(x_j) - f_m(x)| \\ &\leq |f_n(x) - f_n(x_j)| + |f_n(x_j) - f_m(x_j)| + |f_m(x_j) - f_m(x)| \\ &< \varepsilon, \end{aligned}$$

which shows that (f_n) is Cauchy in the sup norm, and hence converges to some $f \in C^0(X;\mathbb{R})$ as required.

Now note that if A is compact then it is bounded. To show that A must consist of equicontinuous functions, suppose that it does not. Then there exists $\varepsilon > 0$ such that for each $n \in \mathbb{N}$ there are points $x_n, y_n \in X$ and $f_n \in A$ such that

$$d(x_n, y_n) < 1/n \qquad \text{but} \qquad |f_n(x) - f_n(y)| \geq \varepsilon.$$

But any uniformly convergent sequence is equicontinuous (exercise). □

10.3 The Stone–Weierstrass Theorem

Let X be a compact metric space. Given $f, g \in C(X)$ we can define

$$fg \in C(X) \qquad \text{by} \qquad (fg)(x) = f(x)g(x).$$

A linear subspace A of $C(X)$ is an *algebra* if $1 \in A$ and $f, g \in A$ implies that $fg \in A$.

Lemma 10.3 *Suppose that X is compact and that A is a closed subalgebra of $C(X)$. Then if $f \in A$ with $f \geq 0$, $\sqrt{f} \in A$.*

Proof Suppose that $0 \leq f \leq 1$. Set $g = 1 - f$; so $0 \leq g \leq 1$ and $f = 1 - g$.

Now, the Taylor expansion of $\sqrt{1-x}$ gives

$$\sqrt{1-x} = 1 - \sum_{n=1}^{\infty} a_n x^n,$$

which converges uniformly for $0 \leq x \leq 1$. So

$$\sqrt{f(x)} = 1 - \sum_{n=1}^{\infty} a_n (g(x)^n) \in A.$$

For a general $f \geq 0$ consider λf with $\lambda \geq 0$ such that $0 \leq \lambda f \leq 1$. □

Theorem 10.4 (Real Stone–Weierstrass Theorem) *Suppose that X is compact and A is a closed subalgebra of $C(X; \mathbb{R})$ that 'separates points', i.e. for every $x, y \in X$ there exists an $f \in A$ such that $f(x) \neq f(y)$. Then $A = C(X; \mathbb{R})$.*

Proof If $f \in A$ then $|f| \in A$, since $|f| = \sqrt{f^2}$.

If $f, g \in A$ then

$$\min(f, g) = \frac{1}{2}(f + g - |f - g|) \in A$$

and

$$\max(f, g) = \frac{1}{2}(f + g + |f - g|) \in A.$$

Now, given $x, y \in X$, there exists $h \in A$ such that $h(x) \neq h(y)$.

For any choice of $\lambda, \mu \in \mathbb{R}$ the map g defined by

$$t \mapsto \mu + (\lambda - \mu)\frac{h(t) - h(y)}{h(x) - h(y)}$$

is an element of A, with $g(x) = \lambda$ and $g(y) = \mu$.

Now, fix $f \in C(X)$. Take $\varepsilon > 0$, $x, y \in X$; then there exists $f_{xy} \in A$ such that

$$f_{xy}(x) = f(x) \qquad f_{xy}(y) = f(y).$$

For $x \in X$ the set

$$U_y = \{\xi \in X : \ f_{xy}(\xi) < f(\xi) + \varepsilon\}$$

is an open neighbourhood of y (since every $h \in A$ is continuous).

Since X is compact, there exist $y_1, \ldots, y_n$ such that

$$X = \bigcup_{j=1}^{n} U_{y_j}.$$

Define

$$h_x = \min(f_{x,y_j}).$$

We know that $h_x \in A$, that $h_x(x) = f(x)$, and $h_x < f + \varepsilon$.

Now consider similarly

$$V_x = \{\xi \in X : \ h_x(\xi) > f(\xi) - \varepsilon\}.$$

Then there exist a finite collection (x_j) such that

$$X = \bigcup_{j=1}^{m} V_{x_j}.$$

Define $F = \max_j h_{x_j}$. Then

$$f - \varepsilon < F < f + \varepsilon,$$

i.e. $d(f, F) < \varepsilon$. It follows that $f \in \overline{A} = A$. □

Consequences of this result were discussed in FA1 (Section 3.2): polynomials are dense in $C(X;\mathbb{R})$; periodic functions can be approximated by Fourier series.

Theorem 10.5 (Complex Stone–Weierstrass Theorem) *Suppose that X is compact and A is a closed subalgebra of $C(X;\mathbb{C})$ that is closed under conjugation ($f \in A$ implies that $\bar{f} \in A$) and that 'separates points', i.e. for every $x, y \in X$ there exists an $f \in A$ such that $f(x) \neq f(y)$. Then $A = C(X;\mathbb{C})$.*

Proof We want to show that $A_{\mathbb{R}}$, the elements of A that are real-valued, satisfy the requirements of the real Stone–Weierstrass Theorem. The algebra property is inherited from A itself; we need to show that $A_{\mathbb{R}}$ still separates points. So suppose that $x, y \in X$ and $f \in A$ separates points. Then either $\mathrm{Re} f(x) \neq \mathrm{Re} f(y)$ or $\mathrm{Im} f(x) \neq \mathrm{Im} f(y)$. Since f is closed under conjugation we have

$$\mathrm{Re}(f) = \frac{1}{2}(f + \bar{f}) \in A \qquad \text{and} \qquad \mathrm{Im}(f) = \frac{1}{2\mathrm{i}}(f - \bar{f}) \in A;$$

these are both elements of $C(X;\mathbb{R})$, and hence of $A_{\mathbb{R}}$. So $A_{\mathbb{R}}$ separates points.

Now since any element $f \in C(X;\mathbb{C})$ can be written as $f_1 + \mathrm{i} f_2$, for appropriate $f_1, f_2 \in C(X;\mathbb{R})$, it follows that we can approximate any element of $C(X;\mathbb{C})$ by elements of the form $\phi_+ \mathrm{i}\psi$ with $\phi, \psi \in A_{\mathbb{R}}$; and thus by $\phi + \mathrm{i}\psi \in A$. □

10.4 Mollification

Finally, we show how to approximate continuous functions by infinitely differentiable functions. First, note that the function

$$\psi(x) = \begin{cases} e^{-1/(1-|x|^2)} & |x| \leq 1 \\ 0 & |x| > 1. \end{cases}$$

is C^∞.

The support of f is given by

$$\mathrm{supp}(f) := \text{closure of } \{x : \ f(x) \neq 0\}.$$

In the statement we write $\partial_j = \partial/\partial x_j$.

Lemma 10.6 *Let $\rho \in C^\infty(\mathbb{R}^n)$ such that*

$$\rho \geq 0, \qquad \mathrm{supp}(\rho) \subset \{|x| \leq 1\}, \qquad \textit{and} \qquad \int \rho = 1.$$

Take $f \in C^0(\mathbb{R}^n)$ with compact support; write $\rho_\varepsilon = \varepsilon^{-n}\rho(x/\varepsilon)$. Then

$$f_\varepsilon(x) = (\rho_\varepsilon * f)(x) = \varepsilon^{-n} \int f(y)\rho\left(\frac{x-y}{\varepsilon}\right) \mathrm{d}y$$

is an element of $C^\infty(\mathbb{R}^n)$ with the support of f_ε contained in an ε-neigbourhood of the support of f. Furthermore, if in addition $f \in C^1(\mathbb{R}^n)$ then $\partial_j f_\varepsilon$ converges uniformly to $\partial_j f$ for each $j = 1, \ldots, n$.

If $f \in C^k(\mathbb{R}^n)$ the argument of the proof shows that all derivatives up to order k converge uniformly; it also shows that

$$\partial_j f_\varepsilon = (\partial_j f)_\varepsilon, \tag{10.1}$$

i.e. mollification and derivatives commute.

Proof That $f_\varepsilon \in C^\infty$ follows by differentiating under the integral sign to give

$$\partial^\alpha f_\varepsilon(x) = \varepsilon^{-n} \int f(y)\partial_x^\alpha \left[\rho\left(\frac{x-y}{\varepsilon}\right)\right] \mathrm{d}y.$$

That its support is contained in an ε-neighbourhood of the support of f follows since $f_\varepsilon = 0$ if the distance of x from the support of f is greater than ε.

To prove convergence of f_ε to f, note that we can change variables to write

$$f_\varepsilon(x) = \int f(x - \varepsilon z)\rho(z)\,\mathrm{d}z, \tag{10.2}$$

and then

$$\begin{aligned}|f_\varepsilon(x) - f(x)| &= \left|\int [f(x - \varepsilon z) - f(x)]\rho(z)\,\mathrm{d}z\right| \\ &\le \int |f(x - \varepsilon z) - f(x)|\rho(z)\,\mathrm{d}z \\ &\le \sup_{|y|\le\varepsilon} |f(x - y) - f(x)|.\end{aligned}$$

Since f is uniformly continuous, this expression tends to zero uniformly as $\varepsilon \to 0$.

For the convergence of derivatives, it follows from (10.2) that

$$\partial_j f_\varepsilon(x) = \int (\partial_j f)(x - \varepsilon z)\rho(z)\,\mathrm{d}z$$

and we can use the same argument as above. Note that this equality also gives (10.1). □

11
An introduction to the theory of distributions

11.1 Test functions and distributions

11.1.1 Multi-index notation

A multi-index α is a collection of non-negative integers

$$\alpha = (\alpha_1, \dots, \alpha_n);$$

we write $|\alpha| = \sum_{j=1}^n \alpha_j$.

For any vector $v = (v_1, \dots, v_n)$ we write

$$v^\alpha = v_1^{\alpha_1} \cdots v_n^{\alpha_n},$$

and in particular

$$\partial^\alpha = \frac{\partial^{|\alpha|}}{\partial_1^{\alpha_1} \cdots \partial_n^{\alpha_n}}.$$

By $\alpha!$ we mean

$$\alpha! = \alpha_1! \cdots \alpha_n!,$$

and we write

$$\binom{\alpha}{\beta} = \frac{\alpha!}{(\alpha-\beta)!\beta!} = \binom{\alpha_1}{\beta_1} \cdots \binom{\alpha_n}{\beta_n},$$

so that the Leibniz Rule for differentiation of products can be written

$$\partial^\alpha (fg) = \sum_{0 \le \beta \le \alpha} \binom{\alpha}{\beta} \partial^\beta f \partial^{\alpha-\beta} g,$$

where we write $\beta \le \alpha$ if $\beta_i \le \alpha_i$ for all $i = 1, \ldots, n$.

The following version of Taylor's Theorem will be useful later.

Theorem 11.1 (Multi-dimensional Taylor's Theorem) *Suppose that $f \in C^\infty(\mathbb{R}^n)$. Then for each $x \in \mathbb{R}^n$ and $m \ge 0$ there exists $c \in (0,1)$ such that*

$$f(x) = \sum_{|\alpha| \le m} \frac{\partial^\alpha f(0)}{\alpha!} x^\alpha + \sum_{|\alpha| = m+1} \frac{\partial^\alpha (cx)}{\alpha!} x^\alpha. \tag{11.1}$$

Proof If $g \in C^\infty(\mathbb{R})$ then Taylor's Theorem guarantees that there exists $c \in (0,1)$ such that

$$g(1) = \sum_{0 \le k \le m} \frac{g^{(k)}(0)}{k!} + \frac{g^{(k+1)}(0)}{(k+1)!} c^{k+1}. \tag{11.2}$$

Apply this to the function $g(t) := f(xt)$. Then one can show by induction that

$$\frac{1}{k!} \partial_t^k f(xt) = \frac{1}{k!} \left(\sum_{j=1}^n x_i \partial_i \right)^k f|_{y=xt} = \sum_{|\alpha|=k} \frac{x^\alpha}{\alpha!} f^{(\alpha)}(xt),$$

from which (11.1) follows. Indeed, if this holds for k, then

$$\begin{aligned} \frac{1}{(k+1)!} \partial_t^{k+1} f(xt) &= \frac{1}{k+1} \partial_t \left(\sum_{|\alpha|=k} \frac{x^\alpha}{\alpha!} f^{(\alpha)}(xt) \right) \\ &= \frac{1}{k+1} \sum_{|\alpha|=k} \sum_{j=1}^n \frac{x_i x^\alpha}{\alpha!} f^{(\alpha+e_i)}(xt), \end{aligned}$$

where e_i is the multiindex that is zero apart from a single 1 in the ith place. Now simply note that $|\alpha + e_i| = k+1$, and that every β with $|\beta| = k+1$ can be obtained in $k+1$ different ways as $\alpha + e_i$ ($|\alpha| = k$), whence

$$\frac{1}{(k+1)!} \partial_t^{k+1} f(xt) = \sum_{|\alpha|=k+1} \frac{x^\alpha}{\alpha!} f^{(\alpha)}(xt).$$

Using this form of the derivative in (11.2) yields (11.1). □

11.1.2 Test functions and distributions

We will take Ω to be an open subset of $\mathbb{R}^n$.

With any continuous function $f \in C^0(\Omega)$ we can associate a linear map u_f given by

$$\phi \mapsto \langle u_f, \phi\rangle := \int_\Omega f(x)\phi(x)\,\mathrm{d}x, \qquad (11.3)$$

for ϕ in some "appropriate" space of functions, and this linear map will have certain continuity properties. A generalised function, or *distribution*, will be a linear map with the same continuity properties.

Since we want the integral in (11.3) to make sense for as many possible f as possible, we choose ϕ to have compact support. Since we will later define generalised derivatives using an integration by parts, we ensure that ϕ is infinitely differentiable.

Definition 11.2 *The space $\mathscr{D}(\Omega)$ of* test functions *on Ω are infinitely differentiable functions f on Ω with compact support, i.e. there is a compact set K with $f \equiv 0$ on $\Omega \setminus K$. We say* $\mathrm{supp}(f) \subset K$.

Other notations for $\mathscr{D}(\Omega)$ are $C_c^\infty(\Omega)$ and $C_0^\infty(\Omega)$.

As distributions are to be defined by continuity properties, we need a notion of convergence in $\mathscr{D}(\Omega)$. In the same way that we were free to choose as "nice" as possible a space of functions, we choose a very strong notion of convergence.

Definition 11.3 *A sequence $\{\phi_n\}$ in $\mathscr{D}(\Omega)$ is said to* converge to ϕ in $\mathscr{D}(\Omega)$ *($\phi_n \xrightarrow{\mathscr{D}} \phi$) if there is a compact set K with* $\mathrm{supp}(\phi_n) \subset K$ *for all n, and ϕ_n and all its derivatives converge uniformly to ϕ on Ω.*

If f is continuous then if $\phi_n \xrightarrow{\mathscr{D}} \phi$ it is easy to see that

$$\langle u_f, \phi_n\rangle = \int_\Omega f\phi_n \to \int_\Omega f\phi = \langle u_f, \phi\rangle,$$

since

$$\left|\int_\Omega f(\phi_n - \phi)\right| \le \int_K |f||\phi_n - \phi| \le \|f\|_K \|\phi_n - \phi\|_K, \qquad (11.4)$$

where $\|f\|_K$ is the supremum of $|f(x)|$ over K. The right hand side of (11.4) tends to zero with n.

We now define the space $\mathscr{D}'(\Omega)$ of distributions on Ω, as the space of all sequentially continuous linear functionals on $\mathscr{D}(\Omega)$. (This is not quite the dual space of $\mathscr{D}(\Omega)$, since $\mathscr{D}(\Omega)$ is not a normed space.)

Definition 11.4 *A distribution f is a sequentially continuous linear functional $\phi \mapsto \langle f, \phi\rangle$ from $\mathscr{D}(\Omega)$ into $\mathbb{R}$; i.e. if $\phi_n \to \phi$ in $\mathscr{D}(\Omega)$ then $\langle f, \phi_n\rangle \to \langle f, \phi\rangle$.*

The u_f defined above for any $f \in C^0(\Omega)$ is therefore a distribution. It is convenient to write just $\langle f, \phi\rangle$ for $\langle u_f, \phi\rangle$, although this is a slight abuse of notation. If $f, g \in C^0$ and $u_f = u_g$ then $f = g$, see Examples 7.

Similarly, given any $f \in L^1_{\mathrm{loc}}(\Omega)$ ($L^1(K)$ for every compact subset K of Ω) we can define u_f by the same formula as (11.3). In this case if $u_f = u_g$ then $f = g$ almost everywhere. (This result is known as the 'fundamental lemma of the calculus of variations'.)

An important example is the Dirac delta distribution: for each $y \in \Omega$ the distribution $\delta \in \mathscr{D}'(\Omega)$ is defined by

$$\langle \delta, \phi\rangle = \phi(y).$$

Lemma 11.5 (Seminorm estimates) *A linear map $u\colon \mathscr{D}(\Omega) \to \mathbb{R}$ is an element of $\mathscr{D}'(\Omega)$ if and only if for every compact set $K \subset \Omega$ there exist constants C, k such that*

$$|\langle u, \phi\rangle| \le C \sum_{|\alpha|\le k} \sup |\partial^\alpha \phi(x)| \tag{11.5}$$

for every $\phi \in \mathscr{D}(\Omega)$ with $\mathrm{supp}(\phi) \subset K$.

A distribution is said to have finite order if you can take the same k in (11.5) for every $K \subset \Omega$; the order of a distribution is the minimum such k.

Proof If u is a distribution then suppose that there is a compact K for which this inequality fails. Then for each $k \ge 1$ there exists $\phi_k \in \mathscr{D}(\Omega)$

with $\operatorname{supp}(\phi_k) \subset K$ and

$$|\langle u, \phi_k \rangle| \geq k \sum_{|\alpha| \leq k} \sup |\partial^\alpha \phi_k|.$$

Now consider

$$\psi_k = \frac{\phi_k}{k \sum_{|\alpha| \leq k} \sup |\partial^\alpha \phi_k|};$$

then $\psi_k \to 0$ in $\mathscr{D}$ but $|\langle u, \psi_k \rangle| \geq 1$, which is impossible.

Now suppose that the inequality holds, and $\phi_n \in \phi$ in $\mathscr{D}$. Let K be a compact set that contains the support of ϕ_n and ϕ. Then there exist C, k, such that

$$|\langle u, \phi_n \rangle - \langle u, \phi \rangle| = |\langle u, \phi_n - \phi \rangle| \leq C \sum_{|\alpha| \leq k} \sup |\partial^\alpha (\phi_n - \phi)|$$

so the right-hand side tends to zero as $n \to \infty$, which implies that $u \in \mathscr{D}'$. □

Examples: δ is a distribution of order zero; $\langle u, \phi \rangle = \partial^\alpha \phi(0)$ a distribution of order $|\alpha|$;

$$\langle u, \phi \rangle = \sum_{k=0}^{\infty} \phi^{(k)}(k)$$

(for $\phi \in \mathscr{D}(\mathbb{R})$) a distribution of infinite order.

Just as we had a notion of convergence in $\mathscr{D}(\Omega)$, there is a corresponding notion of convergence of distributions (cf. weak-$*$ convergence).

Definition 11.6 *A sequence (u_n) in $\mathscr{D}'(\Omega)$ converges to $u \in \mathscr{D}'(\Omega)$ ($u_n \xrightarrow{\mathscr{D}'} u$) if $\langle u_n, \phi \rangle \to \langle u, \phi \rangle$ for all $\phi \in \mathscr{D}(\Omega)$.*

Example: take u_n to be the distribution corresponding to $\rho_{1/n}$ from Lemma 10.6. Then $u_n \xrightarrow{\mathscr{D}'} \delta$ (we will revisit this example).

A version of the Principle of Uniform Boundedness implies that if $(u_j) \in \mathscr{D}'(\Omega)$ and $\langle u_j, \phi \rangle$ converges for every $\phi \in \mathscr{D}(\Omega)$ then the linear map $\phi \mapsto \langle u, \phi \rangle$ defined by setting

$$\langle u, \phi \rangle := \lim_{j \to \infty} \langle u_j, \phi \rangle \qquad \text{for each } \phi \in \mathscr{D}(\Omega)$$

is an element $u \in \mathscr{D}'(\Omega)$.

11.2 Differentiation, multiplication, and other operations

11.2.1 The distribution derivative

We now introduce the generalised derivative, by analogy with integrating by parts

$$\int_\Omega \frac{\partial f}{\partial x_j} g \,\mathrm{d}x = -\int_\Omega f \frac{\partial g}{\partial x_j} \,\mathrm{d}x$$

when f and g are both differentiable.

Definition 11.7 *The distribution derivative of $f \in \mathscr{D}'(\Omega)$ with with respect to x_j, written $\partial_j f$ is given by the linear map*

$$\langle \partial_j f, \phi \rangle := -\langle f, \partial_j \phi \rangle. \tag{11.6}$$

It follows directly from the differentiability and continuity properties of the test functions that $\partial_j f$ is also a distribution.

Proposition 11.8 *As defined above, $\partial_j f \in \mathscr{D}'(\Omega)$. Furthermore,*

(i) if $f \in C^1(\Omega)$ then the definition agrees with the classical one, in that

$$\partial_j u_f = u_{\partial_j f};$$

and

(ii) if $u_k \xrightarrow{\mathscr{D}'} u$ then $\partial_j u_k \xrightarrow{\mathscr{D}'} \partial_j u$.

Proof Suppose that $\phi_n \xrightarrow{\mathscr{D}} \phi$. Then $\partial_j \phi_n \xrightarrow{\mathscr{D}} \partial_j \phi$, and so

$$\langle \partial_j u, \phi_n \rangle = -\langle u, \partial_j \phi_n \rangle \to -\langle u, \partial_j \phi \rangle = (\partial_j u, \phi \rangle,$$

the convergence in the previously line holding since $u \in \mathscr{D}'$; so $\partial_j u$ is a distribution.

For (i) we have

$$\begin{aligned}\langle \partial_j u_f, \phi \rangle = -\langle u_f, \partial_j \phi \rangle &= -\int_\Omega f(x) \partial_j \phi(x) \,\mathrm{d}x \\ &= \int_\Omega \partial_j f(x) \phi(x) \,\mathrm{d}x = \langle u_{\partial_j f}, \phi \rangle.\end{aligned}$$

For (ii),

$$\langle \partial_j u_k, \phi \rangle = -\langle u_k, \partial_j \phi \rangle;$$

since $\partial_j \phi \in \mathscr{D}$, the definition of convergence in $\mathscr{D}'$ gives the limit as

$$-\langle u, \partial_j \phi \rangle = \langle \partial_j u, \phi \rangle. \qquad \square$$

A distribution therefore has derivatives of all orders: iterating (11.6) yields

$$\langle \partial^\alpha f, \phi \rangle := (-1)^{|\alpha|} \langle f, \partial^\alpha \phi \rangle.$$

Example 11.9 *The Heaviside step function*

$$H(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 1 \end{cases}$$

corresponds to the Heaviside distribution $H \in \mathscr{D}'(\mathbb{R})$ defined by setting

$$\langle H, \phi \rangle = \int_0^\infty \phi(x)\,\mathrm{d}x \qquad \text{for every } \phi \in \mathscr{D}(\mathbb{R}).$$

Then

$$\begin{aligned} \langle \partial H, \phi \rangle &= -\langle H, \partial \phi \rangle \\ &= -\int_0^\infty \phi'(x)\,\mathrm{d}x \\ &= \phi(0) = \langle \delta_0, \phi \rangle, \end{aligned}$$

so $\partial H = \delta_0$.

Example 11.10 *The derivatives of the δ distribution, $\delta^{(\alpha)}$, are given by*

$$\langle \delta^{(\alpha)}, \phi \rangle := \langle \partial^\alpha \delta, \phi \rangle = (-1)^{|\alpha|} \langle \delta, \partial^\alpha \phi \rangle = (-1)^{|\alpha|} \partial^\alpha \phi(0).$$

Example 11.11 *For a longer example, let us consider the distribution in $\mathscr{D}'(\mathbb{R})$ derived from the function $\log|x| \in L^1_{\mathrm{loc}}(\mathbb{R})$:*

$$\langle \log|x|, \phi \rangle = \int_{-\infty}^{\infty} \log|x| \phi(x)\,\mathrm{d}x.$$

This is a distribution because $\log|x|$ is integrable on any compact subset of $\mathbb{R}$.

Now we calculate its derivative. We have

$$\begin{aligned}\langle \partial \log|x|, \phi\rangle &= -\langle \log|x|, \phi'\rangle \\ &= -\int_{-\infty}^{\infty} \log|x|\phi'(x)\,\mathrm{d}x \\ &= -\lim_{\varepsilon\to 0}\left(\int_{-\infty}^{-\varepsilon} + \int_{\varepsilon}^{\infty}\right)\log|x|\phi'(x)\,\mathrm{d}x,\end{aligned}$$

since $\log|x|\phi'(x)$ *is integrable. Now we integrate by parts to obtain*

$$\begin{aligned}&\lim_{\varepsilon\to 0}\left[\left(\int_{-\infty}^{-\varepsilon} + \int_{\varepsilon}^{\infty}\right)\frac{\phi(x)}{x}\,\mathrm{d}x\right] + \lim_{\varepsilon\to 0}\log(\varepsilon)(\phi(\varepsilon)-\phi(-\varepsilon)) \\ &\quad= \lim_{\varepsilon\to 0}\left[\int_{-\infty}^{-\varepsilon}\frac{\phi(x)}{x}\,\mathrm{d}x + \int_{\varepsilon}^{\infty}\frac{\phi(x)}{x}\,\mathrm{d}x\right] \\ &\quad=: \mathrm{PV}\int_{-\infty}^{\infty}\frac{\phi(x)}{x}\,\mathrm{d}x,\end{aligned}$$

where the final term on the first line tends to zero as $\varepsilon \to 0$ *using the Mean Value Theorem (*$\phi(\varepsilon) - \phi(-\varepsilon) = 2\varepsilon\phi'(c)$ *for some* $c \in (-\varepsilon, \varepsilon)$*). So* $\partial \log|x| = PV\frac{1}{x}$ *in the sense of distributions.*

11.2.2 Products of smooth functions and distributions

If $u \in C(\Omega)$ and $f \in C^\infty(\Omega)$ then

$$\int (fu)\phi = \int u(f\phi).$$

We define multiplication of distributions by smooth functions analagously.

Lemma 11.12 *Suppose that* $u \in \mathscr{D}'(\Omega)$ *and* $\psi \in C^\infty(\Omega)$*; if we define*

$$\langle \psi u, \phi\rangle := \langle u, \psi\phi\rangle \qquad \text{for every } \phi \in \mathscr{D}(\Omega)$$

then $\psi u \in \mathscr{D}'(\Omega)$*.*

Note that $\psi\phi \in \mathscr{D}(\Omega)$, so this definition makes sense. This does define a distribution, since if $\phi_k \xrightarrow{\mathscr{D}} \phi$ we have $\psi\phi_k \xrightarrow{\mathscr{D}} \psi\phi$ thanks to the Leibniz rule, and hence $\langle \psi u, \phi_k\rangle \to \langle \psi u, \phi\rangle$ as required for ψu to be a distribution.

We now extend the Leibniz rule to such products.

Proposition 11.13 *Take $u \in \mathscr{D}'(\Omega)$ and $\psi \in C^\infty(\Omega)$. Then*

$$\partial^\alpha(u\psi) = \sum_{\beta \le \alpha} \binom{\alpha}{\beta} \partial^\beta \psi \partial^{\alpha-\beta} u.$$

Proof This equality is required to hold in the sense of distributions, which means that

$$\Big\langle u, \underbrace{(-1)^{|\alpha|}\psi\partial^\alpha\phi - \sum_{\beta \le \alpha} \binom{\alpha}{\beta} \partial^{\alpha-\beta}[\phi\partial^\beta\psi]}_{\Phi} \Big\rangle = 0.$$

Note that $\Phi \in \mathscr{D}(\Omega)$. We know that the Leibniz rule holds if $u, \psi \in C^\infty$; so if we consider $u = \Phi$ (the distribution arising from Φ) the same integrations by parts yield $\int |\Phi|^2 = 0$. Since Φ is continuous, it follows that $\Phi \equiv 0$. Then $\langle u, \Phi \rangle = 0$ for every $u \in \mathscr{D}'(\Omega)$ as required. □

Example 11.14 *For any $\phi \in \mathscr{D}(\Omega)$, $\psi \in C^\infty(\Omega)$,*

$$\langle \psi\delta, \phi \rangle = \langle \delta, \psi\phi \rangle = \psi(0)\phi(0) = \langle \psi(0)\delta, \phi \rangle$$

so $\psi\delta = \psi(0)\delta$.

Example 11.15 *For any $f \in C^\infty(\Omega)$*

$$\begin{aligned} f\delta' &= \partial(f\delta) - f'\delta \\ &= \partial(f(0)\delta) - f'(0)\delta \\ &= f(0)\delta' - f'(0)\delta. \end{aligned}$$

11.2.3 Other operations on distributions in $\mathscr{D}'(\mathbb{R}^n)$

We can define the following by analogy with change of variables in integrals when $f \in C(\mathbb{R}^n)$ and $\phi \in \mathscr{D}(\mathbb{R}^n)$.

Reflection. Let $\check{f}(x) := f(-x)$; then

$$\begin{aligned} \int_{\mathbb{R}^n} \check{f}(x)\phi(x)\,\mathrm{d}x &= \int_{\mathbb{R}^n} f(-x)\phi(x)\,\mathrm{d}x = \int_{\mathbb{R}^n} f(y)\phi(-y)\,\mathrm{d}y \\ &= \int_{\mathbb{R}^n} f(y)\check{\phi}(y)\,\mathrm{d}y. \end{aligned}$$

For $u \in \mathscr{D}'(\mathbb{R}^n)$ we define $\check{u}$ by setting

$$\langle \check{u}, \phi \rangle := \langle u, \check{\phi} \rangle \qquad \text{for every } \phi \in \mathscr{D}(\mathbb{R}^n).$$

Translation. Write $\tau_h f(x) := f(x-h)$. Then

$$\begin{aligned}\int_{\mathbb{R}^n} \tau_h f(x)\phi(x)\,\mathrm{d}x &= \int_{\mathbb{R}^n} f(x-h)\phi(x)\,\mathrm{d}x = \int_{\mathbb{R}^n} f(y)\phi(y+h)\,\mathrm{d}y \\ &= \int_{\mathbb{R}^n} f(y)\tau_{-h}\phi(y)\,\mathrm{d}y.\end{aligned}$$

For $u \in \mathscr{D}'(\mathbb{R}^n)$ we define $\tau_h u$ by setting

$$\langle \tau_h u, \phi\rangle := \langle u, \tau_{-h}\phi\rangle \qquad \text{for every } \phi \in \mathscr{D}(\mathbb{R}^n).$$

Dilation. Write $f_t(x) = f(tx)$. Then

$$\begin{aligned}\int_{\mathbb{R}^n} f_t(x)\phi(x)\,\mathrm{d}x &= \int_{\mathbb{R}^n} f(tx)\phi(x)\,\mathrm{d}x = t^{-n}\int_{\mathbb{R}^n} f(y)\phi(y/t)\,\mathrm{d}y \\ &= t^{-n}\int_{\mathbb{R}^n} f(y)\phi_{1/t}(y)\,\mathrm{d}y.\end{aligned}$$

For $u \in \mathscr{D}'(\mathbb{R}^n)$ we define u_t by setting

$$\langle u_t, \phi\rangle = \langle u, t^{-n}\phi_{1/t}\rangle \qquad \text{for every } \phi \in \mathscr{D}(\mathbb{R}^n)..$$

Definition 11.16 *We say that $u \in \mathscr{D}'(\mathbb{R}^n)$ is homogeneous of degree λ if*

$$u_t = t^\lambda u \qquad \textit{for all } t > 0.$$

E.g. δ is homogeneous of degree $-n$:

$$\langle \delta_t, \phi\rangle = \langle \delta, t^{-n}\phi_{1/t}\rangle = t^{-n}\phi(0) = \langle t^{-n}\delta, \phi\rangle.$$

Note that if u is homogeneous of degree λ then $\partial_j u$ is homogeneous of degree $\lambda - 1$:

$$\begin{aligned}\langle (\partial_j u)_t, \phi\rangle &= \langle \partial_j u, t^{-n}\phi_{1/t}\rangle = -\langle u, t^{-n}\partial_j\phi_{1/t}\rangle = -t^{-1}\langle u, t^{-n}(\partial_j\phi)_{1/t}\rangle \\ &= -t^{-1}\langle u_t, \partial_j\phi\rangle = -t^{\lambda-1}\langle u_t, \partial_j\phi\rangle = t^{\lambda-1}\langle \partial_j u_t, \phi\rangle.\end{aligned}$$

It follows that $\partial^\alpha u$ is then homogeneous of degree $\lambda - |\alpha|$; so $\delta^{(\alpha)}$ is homogeneous of degree $-n - |\alpha|$.

11.3 The support of a distribution

Definition 11.17 *Take $u, v \in \mathscr{D}'(\Omega)$, and let U be an open subset of Ω. We say that $u = v$ on U if*

$$\langle u, \phi \rangle = \langle v, \phi \rangle \qquad \textit{for all} \quad \phi \in \mathscr{D}(U).$$

Proposition 11.18 *Let $u, v \in \mathscr{D}'(\Omega)$ and let $\{Y_i\}_{i \in I}$ be a collection of open subsets of Ω. If $u = v$ on Y_i for every i then $u = v$ on $\cup_i Y_i$.*

Proof Take $\phi \in \mathscr{D}(\cup_i Y_i)$. Then $\operatorname{supp}(\phi)$ is a compact subset of $\cup_i Y_i$, so there exist $i_1, \ldots, i_k$ such that $\operatorname{supp}(\phi) \subset \cup_{j=1}^k Y_{i_j}$.

Choose functions $\psi_j \in \mathscr{D}(\Omega)$ such that $\operatorname{supp}(\psi_j) \subset Y_{i_j}$ and $\sum \psi_j \equiv 1$ on $\operatorname{supp}(\phi)$. [This is a partition of unity – for a construction see Friedlander & Joshi.] Then we have

$$\begin{aligned} \langle u, \phi \rangle = \langle u, \sum_{j=1}^k \psi_j \phi \rangle &= \sum_{j=1}^k \langle u, \psi_j \phi \rangle \\ &= \sum_{j=1}^k \langle v, \psi_h \phi \rangle = \langle v, \sum_{j=1}^k \psi_j \phi \rangle = \langle v, \phi \rangle, \end{aligned}$$

where we use the fact that $u = v$ on Y_{i_j} to move from the first to the second line. □

Definition 11.19 *The support of $u \in \mathscr{D}'(\Omega)$, $\operatorname{supp}(u)$, is the set defined by*

$$x \in \Omega \setminus \operatorname{supp}(u) \quad \Leftrightarrow \quad u = 0 \textit{ on some open neighbourhood of } x.$$

Example: $\operatorname{supp}(\delta^{(\alpha)}) = \{0\}$ for every α. We will see that these are the only distributions with point support.

Corollary 11.20 *If $u \in \mathscr{D}'(\Omega)$ then $u = 0$ on $\Omega \setminus \operatorname{supp}(u)$.*

Proof Take $x \in \Omega \setminus \operatorname{supp}(u)$. Then by definition $u = 0$ on an open neighbourhood U_x of x, and $\cup_{x \in \Omega \setminus \operatorname{supp}(u)} U_x$ covers $\Omega \setminus \operatorname{supp}(u)$. □

This means that if $\operatorname{supp}(\phi) \cap \operatorname{supp}(u) = \varnothing$ then $\langle u, \phi \rangle = 0$. So if $\phi = 0$ on an open neighbourhood of $\operatorname{supp}(u)$ then $\langle u, \phi \rangle = 0$.

Note that in general the fact that $\phi = 0$ on $\text{supp}(u)$ alone is not sufficient to imply that $\langle u, \phi \rangle = 0$. Consider $u = \delta'$ on $\mathbb{R}$ and $\phi(x) = x$: then $\phi = 0$ on $\text{supp}(\delta')$, but $\langle \delta', \phi \rangle = -\phi'(0) = -1$.

Theorem 11.21 (Distributions with compact support) *Suppose that $u \in \mathscr{D}'(\Omega)$ has compact support. Then u is a distribution of finite order, i.e. for any compact $K \subset \Omega$ such that* $\text{supp}(u) \subset \text{int}(K)$*, there exist constants C, k, such that*

$$|\langle u, \phi \rangle| \leq C \sum_{|\alpha| \leq k} \sup_{x \in K} |\partial^\alpha \phi(x)| \qquad \textit{for all} \quad \phi \in \mathscr{D}(\Omega).$$

Proof Fix $\psi \in C^\infty$ such that $\text{supp}(\psi) \subset K$ and $\psi \equiv 1$ on a neighbourhood of $\text{supp}(u)$. By Lemma 11.5 there exist constants C, k, such that

$$|\langle u, \phi \rangle| \leq C \sum_{|\alpha| \leq k} \sup_{x \in K} |\partial^\alpha \phi(x)|$$

for all $\phi \in \mathscr{D}(\Omega)$ such that $\text{supp}(\phi) \subset K$.

For a general $\phi \in \mathscr{D}(\Omega)$ we can write

$$\begin{aligned} |\langle u, \phi \rangle| &= |\langle u, \psi\phi \rangle| \\ &\leq C \sum_{|\alpha| \leq k} \sup_{x \in K} |\partial^\alpha (\psi\phi)(x)| \\ &\leq C' \sum_{|\alpha| \leq k} \sup_{x \in K} |\partial^\alpha \phi(x)|, \end{aligned}$$

using the Leibniz rule. □

We write $\mathscr{E}'(\Omega)$ for the subset of $\mathscr{D}'(\Omega)$ consisting of all distributions with compact support in Ω. Note that if $u \in \mathscr{E}'(\Omega)$ then we can define the action of u on any test function $\phi \in C^\infty(\Omega)$; taking $\rho \in \mathscr{D}(\Omega)$ such that $\rho \equiv 1$ on a neighbourhood of the support of u we can define

$$\langle u, \phi \rangle := \langle u, \rho\phi \rangle. \tag{11.7}$$

This observation will be useful later.

The following result is extremely important in applications to fundamental solutions of PDEs.

Theorem 11.22 (Distributions with point support) *If* $u \in \mathscr{D}'(\mathbb{R}^n)$ *with* $\mathrm{supp}(u) = \{0\}$ *then*

$$u = \sum_{|\alpha| \le k} c_\alpha \delta^{(\alpha)}$$

for some $k \ge 0$, $c_\alpha \in \mathbb{R}$.

Proof Using Lemma 11.5 there exist constants C, k such that

$$|\langle u, \phi \rangle| \le C \sum_{|\alpha| \le k} \sup_{|x| \le 1} |\partial^\alpha \phi(x)|.$$

for all $\phi \in \mathscr{D}(\Omega)$ with $\mathrm{supp}(\phi) \subset \{|x| \le 1\}$.

First we show that if $\phi \in \mathscr{D}(\mathbb{R}^n)$ and $\partial^\alpha \phi(0) = 0$ for all $|\alpha| \le k$ then $\langle u, \phi \rangle = 0$.

Fix $\psi \in C^\infty(\mathbb{R}^n)$ with $\psi \equiv 1$ on $\{|x| < 1/2\}$ and $\mathrm{supp}(\psi) \subset \{|x| < 1\}$. Then for $\varepsilon > 0$

$$\begin{aligned} |\langle u, \phi \rangle| &= |\langle u, \psi(x/\varepsilon)\phi(x) \rangle| \\ &\le C \sum_{|\alpha| \le k|} \sup_{|x| \le \varepsilon} |\partial^\alpha \{\psi(x/\varepsilon)\phi(x)\}| \\ &\le C' \sum_{\beta,\gamma:\ |\beta|+|\gamma| \le k} \varepsilon^{-|\beta|} \sup_{|x| \le \varepsilon} |\partial^\gamma \phi|. \end{aligned}$$

By Taylor's Theorem

$$\sup_{|x| \le \varepsilon} |\partial^\gamma \phi| = O(\varepsilon^{k+1-|\gamma|}),$$

so $\langle u, \phi \rangle = O(\varepsilon)$ as $\varepsilon \to 0$, i.e. $\langle u, \phi \rangle = 0$.

Take the same ψ as in part (i). Given $\phi \in \mathscr{D}(\mathbb{R}^n)$, set

$$\tilde{\phi}(x) = \phi(x) - \psi(x) \sum_{|\alpha| \le k} \frac{\partial^\alpha \phi(0) x^\alpha}{\alpha!}.$$

Then $\tilde{\phi} \in \mathscr{D}(\mathbb{R}^n)$ with $\partial^\alpha \tilde{\phi}(0) = 0$ for all $|\alpha| \le k$. By (i) it follows that $\langle u, \tilde{\phi} \rangle = 0$ and so

$$\langle u, \phi \rangle = \sum_{|\alpha| \le k} \langle u, \psi(x) x^\alpha / \alpha! \rangle \partial^\alpha \phi(0).$$

So $u = \sum_{|\alpha| \leq k} c_\alpha \delta^{(\alpha)}$ where

$$c_\alpha = \frac{1}{\alpha!} \langle u, \psi(x) x^\alpha \rangle. \qquad \square$$

An important application is to the fundamental solutions of PDEs. We first treat the Laplace operator ($\Delta = \sum_j \partial_j^2$).

Theorem 11.23 *We have*

$$\Delta |x|^{2-n} = (2-n)\omega_{n-1}\delta,$$

where $\omega_n = 2\pi^{n/2}/\Gamma(n/2)$ *is the area of the unit* n*-sphere.*

Note that $G(x) = |x|^{2-n}$ defines a distribution, since $\int G\phi$ converges for all ϕ with compact support.

Proof For a radial function $f(x) = f(r)$, where $r = |x|$, we have

$$\Delta f = \frac{\mathrm{d}^2 f}{\mathrm{d}r^2} + \frac{n-1}{r}\frac{\mathrm{d}f}{\mathrm{d}r}.$$

So $\Delta |x|^{2-n} = 0$ on $\mathbb{R}^n \setminus \{0\}$. It follows that ΔG is a distribution whose support is $\{0\}$.

By the previous theorem it follows that

$$\Delta G = \sum_{|\alpha| \leq k} c_\alpha \delta^{(\alpha)}$$

for some $k \geq 0$, $c_\alpha \in \mathbb{R}$.

Note that $|x|^{2-n}$ is homogeneous of degree $2-n$, and so ΔG is homogeneous of degree $-n$.

We showed earlier that $\delta^{(\alpha)}$ is homogeneous of degree $-n - |\alpha|$ in $\mathbb{R}^n$. So we must have

$$t^{-n} \sum_{|\alpha| \leq k} c_\alpha \delta^{(\alpha)} = \sum_{|\alpha| \leq k} c_\alpha t^{-n-|\alpha|} \delta^{(\alpha)}.$$

Applying both sides to some $\phi \in \mathscr{D}(\mathbb{R}^n)$ such that $\phi = x^\alpha$ near zero we obtain

$$t^{-n} c_\alpha \alpha! = t^{-n-|\alpha|} c_\alpha \alpha! \qquad \text{for all } t > 0,$$

from which it follows that $c_\alpha = 0$ unless $\alpha = 0$.

Therefore

$$\Delta G = c_0 \delta,$$

and it only remains to find the constant c_0. To do this, apply both sides to a radial $\phi \in \mathscr{D}(\mathbb{R}^n)$ that is one in a neighbourhood of zero to obtain

$$\begin{aligned} c_0 &= \langle \Delta G, \phi \rangle = \langle G, \Delta\phi \rangle = \langle |x|^{2-n}, \Delta\phi \rangle \\ &= \omega_{n-1} \int_0^\infty r^{2-n} \left[\frac{\mathrm{d}^2\phi}{\mathrm{d}r^2} + \frac{n-1}{r}\frac{\mathrm{d}\phi}{\mathrm{d}r} \right] r^{n-1}\,\mathrm{d}r \\ &= \omega_{n-1} \int_0^\infty r\frac{\mathrm{d}^2\phi}{\mathrm{d}r^2} + (n-1)\frac{\mathrm{d}\phi}{\mathrm{d}r}\,\mathrm{d}r \\ &= \omega_{n-1} \int_0^\infty (n-2)\phi'\,\mathrm{d}r = -(n-2)\omega_{n-1} \end{aligned}$$

after an integration by parts. □

For the fundamental solution of the heat equation $\partial_t u - \Delta u = \delta$ see Examples 7.

11.4 Convolution and mollification of distributions

While we cannot take the product of two distributions, we will see that we can take their convolution. First we return to our mollification result from the previous chapter and rewrite it.

Recall that for $u, v \in \mathscr{D}(\mathbb{R}^n)$ we define the convolution of u and v, $u * v$, as

$$(u * v)(x) = \int_{\mathbb{R}^n} u(y)v(x-y)\,\mathrm{d}y.$$

Lemma 11.24 *Take $\rho \in \mathscr{D}(\mathbb{R}^n)$ with $\rho \geq 0$ and $\int_{\mathbb{R}^n} \rho = 1$. Let $\psi_k(x) = k^n \rho(kx)$. Then for every $\phi \in \mathscr{D}(\mathbb{R}^n)$ we have*

$$\psi_k * \phi \xrightarrow{\mathscr{D}} \phi$$

as $k \to \infty$.

Proof The support of $\psi_k * \phi$ is contained within a 1-neighbourhood of the support of ϕ, and $\partial^\alpha(\psi_k * \phi)$ converges uniformly to $\partial^\alpha\phi$ for every $\alpha \geq 0$. This is precisely convergence in $\mathscr{D}(\mathbb{R}^n)$. □

Using the fact that

$$(\tau_x \check{u})(y) = \check{u}(y - x) = u(x - y)$$

we can rewrite the definition of $u * v$ above as

$$(u * v)(x) = \int u(y)(\tau_x \check{v})(y)\, \mathrm{d}y,$$

or more concisely

$$(u * v)(x) = \langle u, \tau_x \check{v} \rangle.$$

For $u \in \mathscr{D}'(\mathbb{R}^n)$, $\phi \in \mathscr{D}(\mathbb{R}^n)$ we can therefore define

$$(u * \phi)(x) = \langle u, \tau_x \check{\phi} \rangle. \tag{11.8}$$

Note that this defines a function of x, not (at least immediately) another distribution.

A simple but useful example:

$$(\delta * \phi)(x) = \langle \delta, \tau_x \check{\phi} \rangle = \tau_x \check{\phi}(0) = \check{\phi}(-x) = \phi(x);$$

so $\delta * \phi = \phi$. Similarly

$$(\delta^{(\alpha)} * \phi)(x) = \langle \delta^{(\alpha)}, \tau_x \check{\phi} \rangle = (-1)^{|\alpha|}[\partial^\alpha \tau_x \check{\phi}](0) = \partial^\alpha \phi(x);$$

so

$$\delta^{(\alpha)} * \phi = \partial^\alpha \phi \qquad \alpha \geq 0,\ \phi \in \mathscr{D}(\mathbb{R}^n). \tag{11.9}$$

We first prove some useful identities.

Lemma 11.25 *For $u \in \mathscr{D}'(\mathbb{R}^n)$ and $\phi \in \mathscr{D}(\mathbb{R}^n)$ we have*

$$\tau_x(u * \phi) = \tau_x u * \phi = u * \tau_x \phi. \tag{11.10}$$

Proof These follows from the definitions:

$$(\tau_x(u * \phi))(y) = (u * \phi)(y - x) = \langle u, \tau_{y-x} \check{\phi} \rangle,$$

$$((\tau_x u) * \phi)(y) = \langle \tau_x u, \tau_y \check{\phi} \rangle = \langle u, \tau_{y-x} \check{\phi} \rangle,$$

and also

$$(u * (\tau_x \phi))(y) = \langle u, \tau_y (\tau_x \phi)\check{} \rangle = \langle u, \tau_{y-x} \check{\phi} \rangle,$$

so that all three expressions are in fact equal. □

We now investigate properties of $u * \phi$ itself.

Lemma 11.26 *For $u \in \mathscr{D}'$ and $\psi \in \mathscr{D}$, $u * \phi \in C(\mathbb{R}^n)$.*

Proof We know that $\tau_x\phi \xrightarrow{\mathscr{D}} \phi$ as $x \to 0$ (see Examples 7). It follows that $\tau_x\phi \xrightarrow{\mathscr{D}} \tau_c\phi$ as $x \to c$. Therefore $\tau_x\check{\phi} \to \tau_c\check{\phi}$ as $x \to c$, and therefore, since u is a distribution,

$$(u * \phi)(x) = \langle u, \tau_x\check{\phi}\rangle \to \langle u, \tau_c\check{\phi}\rangle = (u * \phi)(c)$$

as $x \to c$, i.e. $u * \phi \in C(\mathbb{R}^n)$. □

In fact we can do much better than this, and show that such convolutions are smooth.

Proposition 11.27 *Let $u \in \mathscr{D}'$, $\phi \in \mathscr{D}$; then*

(i) $u * \phi \in C^\infty$;
(ii) $\mathrm{supp}(u * \phi) \subset \mathrm{supp}(u) + \mathrm{supp}(\phi)$; *and*
(ii) $\partial^\alpha(u * \phi) = (\partial^\alpha u) * \phi = u * (\partial^\alpha \phi)$.

Note that if u is a distribution with compact support then it follows that $u * \phi \in \mathscr{D}(\mathbb{R}^n)$.

Proof To prove (ii), since $u * \phi$ is continuous (from Lemma 11.26) it is enough to show that $u * \phi(x) = 0$ for $x \notin \mathrm{supp}(u) + \mathrm{supp}(\phi)$. For such an x there is no $y \in \mathrm{supp}(u)$ with $x - y \in \mathrm{supp}(\phi)$; so there is no $y \in \mathrm{supp}(u)$ with $y \in \mathrm{supp}(\tau_x\check{\phi})$, so $u * \phi(x) = 0$.

For (i) and (iii) we start from the identity

$$\tau_x(\partial^\alpha\phi)\check{} = (-1)^{|\alpha|}\partial^\alpha(\tau_x\check{\phi}).$$

Applying u to both sides and using (11.8) yields

$$(u * (\partial^\alpha\phi))(x) = ((\partial^\alpha u) * \phi)(x).$$

Now consider a unit vector e_j in the jth direction, and set

$$\eta_r = \frac{1}{r}(\tau_0 - \tau_{re_j});$$

then

$$\eta_r\phi(x) = \frac{\phi(x) - \phi(x - re_j)}{r},$$

and by the definition of the derivative

$$\eta_r\phi \xrightarrow{\mathscr{D}} \partial_j\phi \qquad \text{as} \qquad r \to 0,$$

where the convergence is uniform due to Taylor's Theorem. Since τ_x and $\check{}$ are continuous operations in $\mathscr{D}$, we have

$$\tau_x((\eta_r\phi)\check{}) \xrightarrow{\mathscr{D}} \tau_x(\partial_j\phi)\check{}.$$

Now, noting – using Lemma 11.25 – that

$$\eta_r(u * \phi) = u * (\eta_r\phi) = \langle u, \tau_x((\eta_r\phi)\check{})\rangle,$$

it follows on taking $\lim_{r\to 0}$ that

$$\partial_j(u * \phi) = u * (\partial_j\phi).$$

It follows, since we can arbitrarily many derivatives, that $u * \phi \in C^\infty$ and that $\partial^\alpha(u * \phi) = u * (\partial^\alpha\phi)$ for any $\alpha \geq 0$. Once we know that $u * \phi \in C^\infty$ we can show the other identity in (iii) simply using the definitions; and it suffice to show that $\partial_j(u * \phi) = \partial_j u * \phi$:

$$\partial_j u * \phi(x) = \langle \partial_j u, \tau_x\check{\phi}\rangle = \langle u, \tau_x\check{(\partial_j\phi)}\rangle. \qquad \square$$

The following proposition is key to defining the convolution of two distributions.

Proposition 11.28 *If $u \in \mathscr{D}'(\mathbb{R}^n)$, $\phi, \psi \in \mathscr{D}(\mathbb{R}^n)$ then*

$$u * (\phi * \psi) = (u * \phi) * \psi.$$

The same identity holds if $u \in \mathscr{E}'(\mathbb{R}^n)$, $\phi \in \mathscr{D}(\mathbb{R}^n)$, and only one of ϕ, ψ has compact support.

The proof relies on the fact that

$$\left\langle u, \int_K \Phi(y)\,\mathrm{d}y \right\rangle = \int_K \langle u, \Phi(y)\rangle\,\mathrm{d}y,$$

when $\Phi\colon \mathbb{R}^n \to \mathscr{D}(\mathbb{R}^n)$ is continuous. You can prove this by constructing the integral from step functions.

Proof We start with the identity

$$(\phi * \psi)\check{}(t) = \int \check{\psi}(s)(\tau_s\check{\phi})(t)\,\mathrm{d}s.$$

If the supports of $\check{\phi}$ and $\check{\psi}$ are K_1 and K_2 respectively, and we set $K = K_1 + K_2$ then the map $s \mapsto \check{\psi}(s)\tau_s\check{\phi}$ takes $\mathbb{R}^n$ continuously into $\mathscr{D}(K)$, and is zero outside K_2. Thus

$$(\phi * \psi)\check{} = \int_{K_2} \check{\psi}(s)\tau_s\check{\phi}\,\mathrm{d}s \in \mathscr{D}(\mathbb{R}^n).$$

Now,

$$\begin{aligned}(u * (\phi * \psi))(0) &= \langle u, (\phi * \psi)\check{}\rangle \\ &= \int_{K_2} \check{\psi}(s)\langle u, \tau_s\check{\phi}\rangle\,\mathrm{d}s \\ &= \int_{\mathbb{R}^n} \psi(-s)(u * \phi)(s)\,\mathrm{d}s \\ &= ((u * \phi) * \psi(0).\end{aligned}$$

Taking $\psi \mapsto \tau_{-x}\psi$ gives the result for general x. □

We now show that we can approximate distributions by test functions.

Theorem 11.29 *Suppose that* $(\psi_j) \in \mathscr{D}$ *with the property that*

$$\psi_j * \phi \xrightarrow{\mathscr{D}} \phi \qquad \phi \in \mathscr{D}.$$

Then

$$u * \psi_j \xrightarrow{\mathscr{D}'} u \qquad u \in \mathscr{D}'.$$

Proof For any $\phi \in \mathscr{D}$ we have

$$\begin{aligned}\langle (u * \psi_j), \phi\rangle = [(u * \psi_j) * \check{\phi}](0) &= [u * (\psi_j * \check{\phi})](0) \\ &\to [u * \check{\phi}](0) = \langle u, \phi\rangle,\end{aligned}$$

i.e. $u * \psi_j \xrightarrow{\mathscr{D}'} u$. □

11.4.1 Convolution operators and fundamental solutions of PDEs

Given $u \in \mathscr{D}'(\mathbb{R}^n)$ we can define an operator $\mathscr{L}_u\colon \mathscr{D}(\mathbb{R}^n) \to C^\infty(\mathbb{R}^n)$ by setting

$$\mathscr{L}_u(\phi) = u * \phi;$$

by definition this means that

$$\mathscr{L}_u(\phi)(x) = \langle u, \tau_x \check{\phi} \rangle \qquad \phi \in \mathscr{D}(\mathbb{R}^n),\ x \in \mathbb{R}^n.$$

Note that

(i) $\tau_h \circ \mathscr{L}_u = \mathscr{L}_u \circ \tau_h$ for every $h \in \mathbb{R}^n$; and
(ii) if $\phi_k \xrightarrow{\mathscr{D}} \phi$ then $\mathscr{L}_u \phi_k(0) \to \mathscr{L}_u \phi(0)$.

These two properties characterise convolution.

Theorem 11.30 (Characterisation of convolution) *For any operator $\mathscr{L} : \mathscr{D}(\mathbb{R}^n) \to C^\infty(\mathbb{R}^n)$ that satisfies (i) and (ii) there exists a unique $u \in \mathscr{D}'(\mathbb{R}^n)$ such that $\mathscr{L} = \mathscr{L}_u$.*

Proof If such a u exists, then for every $\phi \in \mathscr{D}(\mathbb{R}^n)$ we have

$$\langle u, \check{\phi} \rangle = \langle u, \tau_0 \check{\phi} \rangle = \mathscr{L}_u \phi(0) = \mathscr{L}\phi(0),$$

hence u is unique.

Conversely, if we define u by setting

$$\langle u, \phi \rangle = \mathscr{L}\check{\phi}(0)$$

then condition (ii) implies that $u \in \mathscr{D}'(\mathbb{R}^n)$. For any $\phi \in \mathscr{D}(\mathbb{R}^n)$ and $x \in \mathbb{R}^n$ we have

$$\begin{aligned}(\mathscr{L}\phi)(x) &= \tau_{-x}(\mathscr{L}\phi)(0) \\ &= \mathscr{L}(\tau_{-x}\phi)(0) \qquad \text{by condition (i)} \\ &= \langle u, (\tau_{-x}\phi)\check{}\ \rangle \\ &= \langle u, \tau_x \check{\phi} \rangle \\ &= \mathscr{L}_u \phi(x),\end{aligned}$$

and so $\mathscr{L} = \mathscr{L}_u$ as required. □

Now we have a way of defining $u * v$ for $u \in \mathscr{E}'(\mathbb{R}^n)$ and $v \in \mathscr{D}'(\mathbb{R}^n)$. Recall that we showed in Proposition 11.28 that when $u \in \mathscr{E}'(\mathbb{R}^n)$, $\psi \in C^\infty$, and $\phi \in \mathscr{D}(\mathbb{R}^n)$ then

$$(u * \psi) * \phi = u * (\psi * \phi).$$

If $u \in \mathscr{E}'(\mathbb{R}^n)$ and $v \in \mathscr{D}'(\mathbb{R}^n)$ then $v * \phi \in C^\infty(\mathbb{R}^n)$, and the expression $u * (v * \phi)$ makes sense. Since the resulting map

$$\phi \mapsto u * (v * \phi)$$

satisfies properties (i) and (ii) above, Theorem 11.30 guarantees that there exists a unique distribution, which we write as $u * v$, such that

$$(u * v) * \phi = u * (v * \phi) \qquad \text{for all } \phi \in \mathscr{D}(\mathbb{R}^n).$$

Definition 11.31 *If $u \in \mathscr{E}'(\mathbb{R}^n)$ and $v \in \mathscr{D}'(\mathbb{R}^n)$ then $u * v$ is the unique distribution such that*

$$(u * v) * \phi = u * (v * \phi) \qquad \textit{for all } \phi \in \mathscr{D}(\mathbb{R}^n).$$

Some simple but extremely important examples: take $w = \delta$, then $\delta * v$ is the unique distribution such that

$$(\delta * v) * \phi = \delta * (v * \phi) = v * \phi,$$

so $\delta * v = v$, extending our previous result valid for $v \in \mathscr{D}(\mathbb{R}^n)$ to $v \in \mathscr{E}'(\mathbb{R}^n)$.

If $w \in \mathscr{E}'(\mathbb{R}^n)$ then we can extend $\mathscr{L}_w$ to $\mathscr{L}_w : \mathscr{D}'(\mathbb{R}^n) \to \mathscr{D}'(\mathbb{R}^n)$ if we set

$$\mathscr{L}_w(u) = w * u.$$

We call $E \in \mathscr{D}'(\mathbb{R}^n)$ a fundamental solution of $\mathscr{L}_w$ if

$$\mathscr{L}_w(E) = \delta.$$

Theorem 11.32 *Let $w \in \mathscr{E}'(\mathbb{R}^n)$ and $v \in \mathscr{D}(\mathbb{R}^n)$, and let $E \in \mathscr{D}'(\mathbb{R}^n)$ be a fundamental solution of $\mathscr{L}_w$. Then*

(i) if $u = E * v$ then $\mathscr{L}_w(u) = v$;
(ii) conversely, if $\mathscr{L}_w(u) = v$ and $\operatorname{supp}(u)$ is compact, then $u = E * v$.

Proof (i) We have

$$\mathscr{L}_w(E * v) = w * (E * v) = (w * E) * v = \delta * v = v.$$

(ii) if $\mathscr{L}_w(u) = v$ then $w * u = v$

$$E * v = E * (w * u) = (E * w) * u = \delta * u = u. \qquad \square$$

When $\operatorname{supp}(w) = \{0\}$ then by Theorem 11.22 we have $w = \sum_{|\alpha|\le k} c_\alpha \delta^{(\alpha)}$ and so

$$\mathscr{L}_w = \sum_{|\alpha|\le k} c_\alpha \partial^\alpha. \tag{11.11}$$

We can rewrite Theorem 11.32 as a result for linear PDEs.

Corollary 11.33 *Let $\mathscr{L}$ be a partial differential operator of the form (11.11) and let $E \in \mathscr{D}'(\mathbb{R}^n)$ be such that*

$$\mathscr{L}E = \delta.$$

Then for every $v \in \mathscr{D}(\mathbb{R}^n)$ the equation $\mathscr{L}u = v$ has a unique solution given by $u = E * v$.

Example 11.34 *If $f \in \mathscr{D}(\mathbb{R}^3)$ then the solution of $-\Delta u = f$ on $\mathbb{R}^3$ is given by*

$$u(x) = \frac{1}{4\pi}\int_{\mathbb{R}^3} \frac{f(y)}{|x-y|}\,\mathrm{d}y.$$

12
Zorn's Lemma (non-examinable)

We will use Zorn's Lemma to prove one of the results at the heart of functional analysis, the Hahn–Banach Theorem. We will show that Zorn's Lemma is a consequence Axiom of Choice (in fact the two are equivalent). We follow the lecture notes of Bergman, which can be found online. For more details see Bergman, G.M. (1998) *An Invitation to General Algebra and Universal Constructions*, pub. Henry Helson, Berkeley, CA.

We begin with a formal statement of the Axiom of Choice.

Axiom of Choice *If* $(X_i)_{i\in I}$ *is any family of sets, there exists a function* φ *on* I *such that* $\varphi(i) \in X_i$.

The statement of Zorn's Lemma requires some more terminology.

A set P is *partially ordered* with respect to the relation $\preceq$ provided that

(i) $x \preceq x$ for all $x \in P$;
(ii) if $x, y, z \in P$ with $x \preceq y$ and $y \preceq z$ then $x \preceq z$;
(iii) if $x, y \in P$, $x \preceq y$, and $y \preceq x$ then $x = y$.

Two elements $x, y \in P$ are *comparable* if $x \preceq y$ or $y \preceq x$. A subset $C \subset P$ is called a *chain* if every two elements of C are comparable, and P is *totally ordered* if every two elements of P are comparable.

Note that in a partial order two arbitrary elements of P need not be ordered: consider for example, the case when P consists of all subsets of $\mathbb{R}$ and $X \preceq Y$ if $X \subseteq Y$; one cannot order $[0, 1]$ and $[1, 2]$.

An element $b \in P$ in an *upper bound* for a subset $T \subset P$ if $x \preceq b$ for every $x \in T$, and $m \in T$ is a *maximal element* for T if $x \in T$ and $m \preceq x$ implies that $x = m$.

Zorn's Lemma *If P is a non-empty partially ordered set in which every chain has an upper bound then P contains at least one maximal element.*

We will need some other terminology and minor results for the proof.

An *initial segment* of a chain S is a subset $T \subset S$ such that if $u, v \in S$ with $u \preceq v$ and $v \in T$ then $u \in T$.

A *well-ordered* set is a totally ordered set in which every non-empty subset has a least element (i.e. every non-empty subset A contains an element s such that $s \preceq a$ for every $a \in A$).

Fact 1. If S is a well-ordered subset of a partially ordered set P and $t \notin S$ is an upper bound for S in P, then $S \cup \{t\}$ is well ordered.

Proof If $a, b \in S \cup \{t\}$ then (i) $a, b \in S$ so are comparable; (ii) $a \in S$, $b = t$ so $a \preceq t$; (iii) $a = b = t$ so $a \preceq b$ and $b \preceq a$. Any subset of $S \cup \{t\}$ contains a least element: $\{t\}$ has least element t; a subset of S contains a least element; and for $A \subset S$ the set $A \cup \{t\}$ has the same least element as A. □

Fact 2. If C is a set of well-ordered subsets of a partially ordered set P, such that for all $X, Y \in C$, either X is an initial segment of Y, or Y is an initial segment of X, then $\cup_{X \in C} X$ is well ordered.

Proof First we show that $U = \cup X$ is totally ordered. Given any two elements of U, $a \in X$ and $b \in Y$, where $X, Y \in C$; but one of X and Y is an initial segment of the other, so a, b are comparable. Now take any non-empty subset of U; it has a non-empty intersection with some $X \in C$, and this intersection has a least element s. Now suppose that we also have $U \cap Y \neq \varnothing$. Then (i) Y is an initial segment of X, in which case $Y \subseteq X$ so $U \cap Y \subset U \cap X$, and we already know that s is the least element of $U \cap X$; (ii) X is an initial segment of Y and $v \in Y$ with $v \preceq s$, since $s \in X$ it follows that $v \in X$; but then s is the least element of $U \cap X$, so $s \preceq v$. □

We follow Bergman's lecture notes.

Theorem 12.1 *Zorn's Lemma is equivalent to the Axiom of Choice.*

Proof First we show that the Axiom of Choice implies Zorn's Lemma.

Let P be a non-empty partially ordered set with the property that every chain in P is bounded.

In particular, for any chain C the set of all upper bounds for C is non empty. Suppose that C does not contain an element that is maximal for P; then C must have upper bounds that do not lie in C. Otherwise, suppose that $b \in C$ is an upper bound for C and $m \in P$ satisfies $b \preceq m$; then m is an upper bound for C and so $m \in S$; therefore $m \preceq b$, whence $m = b$. It follows that b is a maximal element of P.

We denote that set of these upper bounds for C that do not lie in C by $B(C)$, and using the Axiom of Choice for each chain C we choose one element of $B(C)$, and denote it by $\varphi(C)$.

Now we would like to argue as follows: choose some $p_0 \in P$. If this is not maximal then let $p_1 = \varphi(\{p_0\}) \succeq p_0$. If p_1 is not maximal then let $p_2 = \varphi(\{p_0, p_1\}) \succeq p_1$. If this process never terminates then we let

$$p^* = \varphi(\{p_0, p_1, p_2, \ldots\}).$$

If p^* is not maximal in P then we append p^* to the above chain and continue...

Now, fix an element $p \in P$, and let Z denote the set of subsets S of P that have the following properties:

(i) S is a well-ordered chain in P;
(ii) p is the least element of S;
(iii) for every proper non-empty initial segment $T \subset S$ the least element of $S \setminus T$ is $\varphi(T)$.

Note that Z is non-empty, since it contains $\{p\}$.

If S and S' are two members of Z then one is an initial segment of the other. To see this, let R denote the union of all sets that are initial segments of both S and S' - the 'greatest common initial segment' (R is non-empty since $\{p\}$ is an initial segment of every $S \in Z$). If R is a proper subset of S and of S', then by (iii) the element $\varphi(R)$ is the least element of both $S \setminus R$ and $S' \setminus R$; this would mean that $R \cup \varphi(R)$ is an initial segment of both S and S', but this contradicts the maximality of R. Therefore $R = S$ or $R = S'$, i.e. one is an initial segment of the other.

By Fact 2, the set U, the union of all members of Z, is well ordered. All members of Z are initial segments of U [suppose that $u, v \in U$, $u \preceq v$, and $v \in X$ ($X \in Z$); if $u \in Y$ then either Y is an initial segment of X, in which case $u \in X$ immediately; or X is an initial segment of Y and it follows from this that $u \in X$] and the least element of U is $\{p\}$. Furthermore, U also satisfies (iii): if T is a proper nonempty initial segment of U then there exists some $u \in U \setminus T$. By construction of U, $u \in S$ for some $S \in Z$, and so T must be a proper initial segment of S. Hence (iii) ensures that $\varphi(T)$ is the least element of $S \setminus T$, and since S is an initial segment of U, $\varphi(T)$ is also the least element of $U \setminus T$.

Therefore U is a member of Z. If U does not contain a maximal element of P then $U \cup \{\varphi(U)\}$ will be an element of Z that is not a subset of U, which contradicts the definition of U.

Now we show that Zorn's Lemma implies the Axiom of Choice.

Let P be the collection of all subsets $\Phi \subset I \times \cup_{i \in I} X_i$ with the property that (i) for each $i \in I$ there is at most one element of the form $(i, \varphi) \in \Phi$ and (ii) if $(i, \varphi) \in \Phi$ then φ is a single element of X_i.

We partially order P by inclusion. P is non-empty because the empty set is a member of P. Now suppose that C is a chain in P; then $U = \cup_{S \in C} S$ is an upper bound for C, since $S \subseteq U$ for every $S \in C$. It follows that P has a maximal element Φ^*; if there exists an $i \in I$ such that Φ^* contains no element of the form (i, ξ) with $\xi \in X_i$ we can consider $\Psi := \Phi^* \cup (i, \xi)$ for some $\xi \in X_i$, and then $\Psi \in P$, any $\Phi \in P$ satisfies $\Phi \preceq \Psi$, but $\Psi \neq \Phi^*$ which contradicts the maximality of Φ^*. $\square$