Chapter 2

Banach and Hilbert space analysis

2.1 Hilbert Spaces

2.1.1 Review of Hilbert space theory

We will briefly review the theory of Hilbert spaces in order to fix conventions. A (complex) inner product space is a complex vector space H, equipped with an inner product $(\cdot,\cdot) : H \times H \to \mathbb{C}$ such that

i) $(\cdot,\cdot)$ is sesquilinear[1]:

$$(x+y,z) = (x,z) + (y,z) \qquad (x,y+z) = (x,y) + (x,z)$$
$$(\alpha x, y) = \overline{\alpha}(x,y) \qquad (x,\alpha,y) = \alpha(x,y).$$

for all $x, y, z \in H, \alpha \in \mathbb{C}$

ii) $(\cdot,\cdot)$ is hermitian and positive:

$$(x,y) = \overline{(y,x)}, \qquad (x,x) \geqslant 0$$

for all $x, y \in H$, with equality in the second expression iff $x = 0$.

We define $\|\cdot\| : x \mapsto \sqrt{(x,x)}$.

Lemma 2.1. *The map $\|\cdot\| : H \to [0,\infty)$ is a norm on H, which satisfies the Cauchy-Schwarz inequality:*

$$|(x,y)| \leqslant \|x\| \, \|y\| \,, \qquad \text{for all } x,y \in H, \text{ with `$=$' iff } x = \alpha y, \text{ for some } \alpha \in \mathbb{C},$$

and the parallelogram identity:

$$2\|x\|^2 + 2\|y\|^2 = \|x+y\|^2 + \|x-y\|^2$$

[1] There is a choice of convention over whether the first or second entry is anti-linear. Our choice here is that typically used in quantum mechanics, in pure maths the opposite convention is often used. What matters, however, is to be consistent!

Proof. Positivity and homogeneity of $\|\cdot\|$ are immediate from the definition. Fix $x, y \in H$. For any $t \in \mathbb{R}$ we have that $\|x+ty\|^2 \geqslant 0$, so expanding using we see

$$\|x\|^2 + 2t\Re(x,y) + t^2\|y\|^2 \geqslant 0.$$

A non-negative definite quadratic must have non-positive discriminant, hence we must have $(\Re(x,y))^2 \leqslant \|x\|^2\|y\|^2$. Replacing $x \to e^{i\theta}x$ for suitable θ, we deduce the Cauchy-Schwarz inequality. Now, we compute:

$$\|x+y\|^2 = \|x\|^2 + 2\Re(x,y) + \|y\|^2 \leqslant \|x\|^2 + 2\|x\|^2\|y\|^2 + \|y\|^2 = (\|x\| + \|y\|)^2$$

so taking square roots the triangle inequality follows. The parallelogram law can be easily verified by expanding the right-hand side. □

Definition 2.1. *A* Hilbert space *is a complex inner product space* $(H, (\cdot,\cdot))$*, such that the associated metric* $\|\cdot\|$ *is complete.*

Thus a Hilbert space is a special case of a Banach space, however the presence of the inner product gives the space a more geometric character. In particular, we have a natural notion of 'orthogonality'. If $K \subset H$, we write

$$K^\perp = \{x \in H : (z,x) = 0, \text{ for all } z \in K\}.$$

Note that $K \cap K^\perp \subset \{0\}$ and $K \subset M \implies M^\perp \subset K^\perp$.

Lemma 2.2. *For any* $K \subset H$*,* $K^\perp$ *is a closed subspace.*

Proof. For any $z \in H$, the map $\Lambda_z : H \to \mathbb{C}$ given by $x \mapsto (z,x)$ is a bounded linear map. The linearity follows from properties of the inner product, and boundedness follows as:

$$|\Lambda_z x| = |(z,x)| \leqslant \|z\|\,\|x\|$$

by Cauchy-Schwarz. Thus Ker Λ_z is a closed subspace of H for any $z \in H$. By definition

$$K^\perp = \bigcap_{z\in K} \text{Ker } \Lambda_z,$$

so $K^\perp$ is a closed subspace. □

An important result concerns closed, convex sets in a Hilbert space (see Definition A.3)

Theorem 2.3. *Let* K *be a nonempty, closed, convex set in a Hilbert space* $(H, (\cdot,\cdot))$*. Then* K *contains a unique element of least norm.*

Proof. Let $d = \inf\{\|x\| : x \in K\}$. For any $x, y \in K$ the parallelogram identity applied to $x/2$, $y/2$ gives:

$$\frac{1}{4}\|x-y\|^2 = \frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2 - \left\|\frac{x+y}{2}\right\|^2$$

The convexity of K implies $(x+y)/2 \in K$, so

$$\|x-y\|^2 \leqslant 2\|x\|^2 + 2\|y\|^2 - 4d^2. \tag{2.1}$$

We immediately deduce that if $\|x\| = \|y\| = d$, then $x = y$, so the infimum of the distance can be attained by at most one point. Now suppose $(x_n)_{n=1}^{\infty}$ is a sequence with $x_n \in K$ and $\|x_n\| \to d$. By (2.1) we have:

$$\|x_n - x_m\|^2 \leqslant 2\|x_n\|^2 + 2\|x_m\|^2 - 4d^2$$

which implies (x_n) is Cauchy, hence $x_n \to x$ for some $x \in H$. Since K is closed, $x \in K$, and since the norm is continuous $\|x\| = d$ and we're done. □

Noting that K is closed and convex if and only if $K + y$ is closed and convex, we immediately deduce

Corollary 2.4. *Let K be a nonempty, closed, convex set and let $x \in H$. There is a unique $y \in K$ which minimises $\|x-y\|$.*

A more important corollary of this result concerns orthogonal projection onto closed linear subspaces.

Lemma 2.5. *Let L be a closed linear subspace of H. There exists a bounded linear operator $P : H \to L$ such that*

i) $Px = x$ *if* $x \in L$,

ii) $x - Px \in L^{\perp}$,

iii) $Px = 0$ *if* $x \in L^{\perp}$,

iv) $\|Px\| \leqslant \|x\|$.

Proof. For any $x \in H$, let Px be the unique point in L which minimises $\|x - Px\|$. It is immediate that $Px = x$ if $x \in L$. Now, suppose $y \in L$ and consider $\|x - Px - t\alpha y\|$ for $t \in \mathbb{R}$, where α is chosen to satisfy $\Re(\alpha y, x - Px) = |(y, x - Px)|$, $|\alpha| = 1$. We find

$$\|x - Px - t\alpha y\|^2 = \|x - Px\|^2 + 2t\,|(y, x-Px)| + t^2\|y\|^2 \geqslant \|x - Px\|^2$$

so $|(y, x - Px)| = 0$, and hence $x - Px \in L^{\perp}$. This immediately implies that if $x \in L^{\perp}$, then $Px \in L \cap L^{\perp} = \{0\}$. Next, we note that

$$\|x\|^2 = \|x - Px + Px\|^2 = \|x - Px\|^2 + \|Px\|^2,$$

from which the bound on $\|Px\|$ follows. Finally, to see that L is linear, for $\lambda \in \mathbb{C}$, $x, y \in H$ write $P(x + \lambda y) = P(x) + \lambda P(y) + z$ for some $z \in L$. We observe:

$$\begin{aligned}\|(x+\lambda y) - P(x+\lambda y)\|^2 &= \|x - P(x) + \lambda(y - P(y)) - z\|^2\\ &= \|x - P(x) + \lambda(y - P(y))\| + \|z\|^2\end{aligned}$$

which is clearly minimised for $z = 0$, hence $P(x + \lambda y) = P(x) + \lambda P(y)$. □

An immediate corollary of the result is

Corollary 2.6. *If L is a closed linear subspace of H, then*

$$H = L \oplus L^\perp.$$

Proof. Given $x \in H$ we can write $x = Px + (x - Px)$ with $Px \in L$, $(x - Px) \in L^\perp$, and certainly $x \in L \cap L^\perp$ if and only if $x = 0$. □

We also have the following useful characterisation of $K^{\perp\perp} := (K^\perp)^\perp$

Corollary 2.7. *If L is a closed linear subspace of H then $L^{\perp\perp} = L$. If $K \subset H$ is any set, we have $K^{\perp\perp} = \overline{\text{span } K}$. In particular, K is dense in H if and only if $K^\perp = \{0\}$.*

Proof. Clearly $L \subset L^{\perp\perp}$. Suppose $z \in L^{\perp\perp}$, then $z = x + y$ with $x \in L$, $y \in L^\perp$ by the previous Corollary. Thus $0 = (z, y) = (x + y, y) = \|y\|^2$, so $y = 0$ and $z = x \in L$, thus $L^{\perp\perp} \subset L$.

Now consider an arbitrary $K \subset H$. It is clear from the definition of the orthogonal complement that $\overline{\text{span } K} \subset K^{\perp\perp}$. Since $K \subset \overline{\text{span } K}$, we deduce $\overline{\text{span } K}^\perp \subset K^\perp$ and $K^{\perp\perp} \subset \overline{\text{span } K}^{\perp\perp} = \overline{\text{span } K}$, making use of the fact that $\overline{\text{span } K}$ is a closed subspace. □

The final result of this section shows that for a Hilbert space H, we can describe the dual space H' in a straightforward fashion. Recall that for a topological vector space X, the dual space X' is defined to be the set of continuous linear maps $\Lambda : X \to \mathbb{C}$, sometimes known as the continuous linear *functionals* on X. We saw in the proof of Lemma 2.2 that to any $z \in H$ we can associate a continuous linear map $\Lambda_z : x \mapsto (z, x)$. The following result, known as Riesz representation theorem shows that in fact *any* element of H' must be of this form.

Theorem 2.8 (Riesz representation theorem for Hilbert spaces)**.** *If $\Lambda : H \to \mathbb{C}$ is a continuous linear functional on H, then there is a unique $z \in H$ such that $\Lambda = \Lambda_z$.*

Proof. If $\Lambda x = 0$ for all x, take $z = 0$, otherwise let

$$L = \{x : \Lambda x = 0\}.$$

L is a subspace by the linearity of Λ, and it is closed by the continuity of Λ. Moreover, since $\Lambda x \neq 0$ for some x, $L^\perp$ cannot be trivial, as $H = L \oplus L^\perp$ and $L \neq H$.

Pick $y \in L^\perp$ such that $\|y\| = 1$ and for any $x \in H$ let

$$w = (\Lambda x)y - (\Lambda y)x.$$

Clearly $\Lambda w = 0$, so $w \in L$, and as a result $(y, w) = 0$, which implies

$$\Lambda x = (\Lambda y)(y, x) = \left((\overline{\Lambda y})y, x\right)$$

so taking $z = (\overline{\Lambda y})y$ we have established $\Lambda = \Lambda_z$. To see that z is unique, suppose $\Lambda_z = \Lambda_{z'}$, then for any $x \in H$ we have:

$$(z - z', x) = 0,$$

in particular this holds for $x = z - z'$, so that $z = z'$. □

2.1.2 The Hilbert space L^2

Given any measure space $(E, \mathcal{E}, \mu)$, the space $L^2(E, \mu)$ is naturally a Hilbert space, with inner product given by:

$$(f, g)_{L^2} := \int_E \overline{f} g d\mu.$$

We will mostly focus on the cases where E is $\mathbb{R}^n$ (or a subset) equipped with Lebesgue measure, although for some purposes it is convenient to keep the discussion more general.

Orthogonal systems of functions and their completeness

Suppose $S = \{u_j\}_{j \in J}$ is a subset of a Hilbert space H indexed by some (not necessarily countable) set J. We say S is *orthogonal* if $(u_j, u_k) = 0$ for all $j, k \in J$ with $j \neq k$. We say S is *orthonormal* if additionally $\|u_j\| = 1$ for all $j \in J$. We say that S is *complete* if $\overline{\text{Span } S} = H$, where Span S is the set of finite linear combinations of elements of S. An orthonormal set which is complete, we refer to as an *orthonormal basis*. For many purposes, we can take the set J to be $\mathbb{N}$, thanks to the following result

Theorem 2.9. *A Hilbert space H is separable if and only if it admits a countable set S which is orthonormal and complete.*

Proof. If S is countable and complete, then the set

$$\{(\alpha_1 + i\beta_1)s_1 + \cdots + (\alpha_n + i\beta_n)s_n | s_j \in S, \alpha_j, \beta_j \in \mathbb{Q}\}$$

is countable and can be seen to be dense by the completeness of S, hence H is separable. Conversely, if H is separable, then it has a countable dense subset D. By applying the Gram-Schmidt process to this set we can find a countable orthonormal set S such that Span S = Span D, and thus $\overline{\text{Span } S} = H$. □

A useful result concerning orthonormal sets is the following

Lemma 2.10. *Suppose $\{u_j\}_{j=1}^{\infty}$ is an orthonormal set. Then for any $x \in H$:*

$$\sum_{j=1}^{\infty} |(u_j, x)|^2 \leqslant \|x\|^2$$

Proof. For each j pick θ_j such that $(e^{i\theta_j} u_j, x) = |(u_j, x)|$. Consider

$$0 \leqslant \left\| x - \sum_{j=1}^{n} (e^{i\theta_j} u_j, x) e^{i\theta_j} u_j \right\|^2 = \|x\|^2 - 2\sum_{j=1}^{n} |(u_j, x)|^2 + \sum_{j=1}^{n} |(u_j, x)|^2$$

so

$$\sum_{j=1}^{n} |(x, u_j)|^2 \leqslant \|x\|^2$$

and taking the supremum over n we're done. □

Since we know that $L^2(U)$ is separable for U any subset of $\mathbb{R}^n$, we deduce that $L^2(U)$ admits a complete countable orthonormal basis. We give some examples of orthogonal sets of functions.

Example 1. *Consider $L^2([0,1])$ equipped with Lebesgue measure. The set $S = \{e^{-2\pi inx}\}_{n\in\mathbb{Z}}$ is an orthonormal set. By the Stone–Weiserstrass theorem, any function $f \in C^0([0,1])$ can be approximated uniformly by a finite linear combination of elements of S. Since we can approximate any element of $L^2([0,1])$ by a continuous function by Theorem 1.13, we deduce that S is complete. We will see another proof of the completeness of the set S later in the course.*

Example 2. *Let ψ be the function given by:*

$$\psi(x) := \begin{cases} 1 & 0 \leqslant x < \frac{1}{2} \\ -1 & \frac{1}{2} \leqslant x < 1 \\ 0 & \textit{otherwise} \end{cases}$$

and define $\psi_{n,k}$ by

$$\psi_{n,k}(x) := 2^{\frac{n}{2}}\psi(2^n x - k)$$

Then $\{\psi_{n,k}\}_{n,k\in\mathbb{Z}}$ is a complete orthonormal basis for $L^2(\mathbb{R})$ (see Exercise 1.11). This is known as the Haar system. *It is the simplest example of a* wavelet *basis, which give an (imperfect) localisation of a function in both space and frequency. Such bases are widely used in signal processing.*

Example 3. *Consider the space $L^2(\mathbb{R}, e^{-x^2}dx)$, which is a Hilbert space equipped with the Gaussian-weighted inner product*

$$(f,g) = \int_{\mathbb{R}} \overline{f(x)}g(x)e^{-x^2}dx.$$

Applying the Gram-Schmidt process to the linearly independent set $\{1, x, x^2, x^3, \ldots\}$ we can construct a sequence of polynomials $H_k(x)$ of degree k such that

$$\int_{\mathbb{R}} \overline{H_k(x)}H_l(x)e^{-x^2}dx = 0$$

for all $l < k$. For historical reasons, the normalisation is usually chosen such that the coefficient of x^k in $H_k(x)$ is 2^k, but this is purely a convention. The set $\{H_k\}_{k=0}^{\infty}$ is a complete orthogonal set for $L^2(\mathbb{R}, e^{-x^2}dx)$, known as the Hermite *polynomials. We will justify this assertion later in the course.*

Exercise 1.11. Let $S = \{\psi_{n,k}\}_{n,k\in\mathbb{Z}}$ be the Haar system, as defined in lectures.

a) Show that

$$\int_{\mathbb{R}} \psi_{n_1,k_1}(x)\psi_{n_2,k_2}(x)dx = \delta_{n_1n_2}\delta_{k_1k_2}.$$

b) Show that $\mathbb{1}_I \in \overline{\text{Span } S}$ for any finite interval I, where the closure is understood with respect to the L^2 norm.

c) Deduce that S is an orthonormal basis for $L^2(\mathbb{R})$.

2.1.3 The Radon–Nikodym Theorem

An important application of the Riesz representation theorem for Hilbert spaces in the context of measure theory is the Radon–Nikodym theorem, which is important in its own right and will moreover will be valuable when we come to study the dual spaces to the L^p spaces. We first introduce some nomenclature.

Definition 2.2. *Suppose $(E, \mathcal{E})$ is a measurable space, equipped with measures μ, ν. We say ν is* absolutely continuous *with respect to μ, written $\nu \ll \mu$, if for any measurable set A*

$$\mu(A) = 0 \qquad \Longrightarrow \nu(A) = 0.$$

We say μ, ν are mutually singular*, written $\mu \perp \nu$ if there exists a measurable set A such that*

$$\mu(A) = 0 = \nu(A^c)$$

With this definition in hand, we can state

Theorem 2.11 (Radon–Nikodym Theorem)**.** *Suppose $(E, \mathcal{E})$ is a measurable space, with* finite *measures μ, ν such that $\nu \ll \mu$. Then there exists a non-negative $w \in L^1(E, \mu)$ such that*

$$\nu(A) = \int_A w d\mu$$

for any $A \in \mathcal{E}$. In particular this implies

$$\int_E F(x) d\nu(x) = \int_E F(x) w(x) d\mu(x)$$

for any non-negative measurable F.

Proof. Let

$$\alpha = \mu + 2\nu, \qquad \beta = 2\mu + \nu,$$

then α, β are finite measures on $(E, \mathcal{E})$ in an obvious way. On the Hilbert space $H = L^2(E, \alpha) = L^2(E, \mu) \cap L^2(E, \nu)$, we consider the map $\Lambda : H \to \mathbb{C}$ given by:

$$\Lambda(f) = \int_E f d\beta.$$

Noting that $|\Lambda(f)| \leqslant \int_E |f| \, d\beta \leqslant 2 \int_E |f| \, d\alpha \leqslant 2\sqrt{\alpha(E)} \, \|f\|_{L^2(E,\alpha)}$, we see that Λ is bounded on H, and it is manifestly linear, so by Riesz representation theorem (Thm 2.8) there exists $g \in H$ such that for any $f \in H$ we have

$$\int_E f(x) d\beta(x) = \int_E f(x) g(x) d\alpha(x)$$

Rearranging, we deduce

$$\int_E f(2g - 1) d\nu = \int_E f(2 - g) d\mu. \tag{2.2}$$

for all $f \in H$. Taking $f = \mathbb{1}_{A_j}$ for $A_j = \{x \in E : g(x) < \frac{1}{2} - \frac{1}{j}\}$ we deduce $\mu(A_j) = \nu(A_j) = 0$, and so $g \geqslant \frac{1}{2}$ μ-ae and ν-ae. Similarly by considering $A_j = \{x \in E : g(x) > 2 + \frac{1}{j}\}$ we see that $g \leqslant 2$ μ-ae and ν-ae. Thus by redefining g on a set which is null with respect to all measures in the problem, we may assume $\frac{1}{2} \leqslant g(x) \leqslant 2$ for all $x \in E$. By the monotone convergence theorem, we can deduce that (2.2) holds for any non-negative measurable f.

Let $Z = \{g(x) = \frac{1}{2}\}$. Setting $f = \mathbb{1}_Z$ in (2.2) we see that $\mu(Z) = 0$. Since $\nu \ll \mu$ we deduce that $\nu(Z) = 0$, so given a non-negative measurable function F, we can define

$$f(x) = \frac{F(x)}{2g(x) - 1}, \qquad w(x) = \frac{2 - g(x)}{2g(x) - 1}$$

for all $x \in Z^c$ and set $f(x) = w(x) = 0$ otherwise. Applying (2.2) and using $\mu(Z) = \nu(Z) = 0$, we deduce

$$\begin{aligned}\int_E F(x) d\nu(x) &= \int_{E \setminus Z} F(x) d\nu(x) \\ &= \int_E f(2g - 1) d\nu = \int_E f(2 - g) d\mu \\ &= \int_{E \setminus Z} F(x) w(x) d\mu(x) = \int_E F(x) w(x) d\mu(x)\end{aligned}$$

Setting $F(x) = 1$ shows $w \in L^1(E, \mu)$. □

Exercise 1.12. (*) Suppose $(E, \mathcal{E})$ is a measurable space, with finite measures μ, ν. Show that ν may be uniquely written as $\nu = \nu_a + \nu_s$, for measures ν_a, ν_s such that $\nu_s \perp \mu$ and $\nu_a \ll \mu$.
[Hint: Return to the proof of the Radon–Nikodym theorem, but drop the assumption that $\nu \ll \mu$]

2.2 Dual spaces

Given a topological vector space[2] X, we define the dual space X' to be the set of continuous linear functions $\Lambda : X \to \mathbb{C}$. This is a vector space with the obvious operations:

$$(\Lambda_1 + \alpha\Lambda_2)(x) := \Lambda_1(x) + \alpha\Lambda_2(x), \qquad \text{for all } x \in X, \Lambda_1, \Lambda_2 \in X', \alpha \in \mathbb{C}.$$

If X is a normed space, we can equip X' with a norm by setting

$$\|\Lambda\|_{X'} = \sup_{x \in X, \|x\| = 1} |\Lambda(x)|.$$

Often (though not always), we will take X to be a Banach space.

Exercise 2.1. Let X be a normed space. Show that X' equipped with its norm forms a Banach space. If $\overline{X}$ is the completion of X with respect to the metric induced by its norm, show that $X' = \overline{X}'$.

[2] See Definition A.7

Exercise 2.2. Suppose X is a Banach space. Show that if $\Lambda \in X'$ with $\Lambda \neq 0$ then Λ is an open mapping (i.e. $\Lambda(U)$ is open whenever $U \subset X$ is open).

An important fact concerning the dual of a Banach space is that it *separates points*. That is

Lemma 2.12. *Let X be a Banach space. Suppose $x, y \in X$ with $x \neq y$. Then there exists $\Lambda \in X'$ such that $\Lambda(x) \neq \Lambda(y)$.*

We shall not prove this result at this stage, it will follow as a corollary of the Hahn–Banach theorem which we shall prove later.

Suppose X is a Banach space. Given $x \in X$, there is a natural map:

$$\begin{aligned} f_x \quad : \quad & X' \to \mathbb{C} \\ & \Lambda \mapsto \Lambda(x). \end{aligned}$$

This is a bounded linear map, thus belongs to X''. Furthermore, if $f_x(\Lambda) = f_y(\Lambda)$ for all Λ then $x = y$ by Lemma 2.12, thus we have a natural injection of X into X'' given by $x \mapsto f_x$. If this map is surjective, then we say that X is *reflexive*, and write $X = X''$ (by a slight abuse of notation).

2.2.1 The dual of $L^p(\mathbb{R}^n)$

Suppose $f \in L^p(\mathbb{R}^n)$ for some $1 \leqslant p \leqslant \infty$, and let q be such that $p^{-1} + q^{-1} = 1$. Then by Hölder's inequality, we know that if $g \in L^q(\mathbb{R}^n)$ we have:

$$\left| \int_{\mathbb{R}^n} g(x) f(x) dx \right| \leqslant \|g\|_{L^q} \|f\|_{L^p}$$

This tells us that the map $\Lambda_g : L^p(\mathbb{R}^n) \to \mathbb{C}$ given by:

$$\Lambda_g(f) := \int_{\mathbb{R}^n} g(x) f(x) dx$$

is a bounded linear map from $L^p(\mathbb{R}^n)$ to $\mathbb{C}$, thus $\Lambda_g \in L^p(\mathbb{R}^n)'$. Furthermore, it can be shown (Exercise 1.4) that $\|\Lambda_g\|_{L^{p\prime}} = \|g\|_{L^q}$. Thus the map

$$\begin{aligned} \kappa : \quad L^q(\mathbb{R}^n) \quad & \to \quad L^p(\mathbb{R}^n)' \\ g \quad & \mapsto \quad \Lambda_g \end{aligned}$$

is linear, isometric and injective. We see that $L^q(\mathbb{R}^n) \subset L^p(\mathbb{R}^n)'$ in a natural way.

For the case $p = q = 2$, we know by Riesz representation theorem that in fact[3] $L^2(\mathbb{R}^n) = L^2(\mathbb{R}^n)'$. It is a very natural question to ask whether this happens for other values of p. In fact, it is true for all values of p except one.

Theorem 2.13. *[The dual of $L^p(\mathbb{R}^n)$] Let $1 \leqslant p < \infty$, and let q satisfy $p^{-1} + q^{-1} = 1$. Then $L^q(\mathbb{R}^n) = L^p(\mathbb{R}^n)'$, where we understand elements of $L^q(\mathbb{R}^n)$ as linear maps on $L^p(\mathbb{R}^n)$ according to the map κ described above.*

[3] Being pedantic, one should say that Riesz representation theorem gives an isometric bijection between $L^2(\mathbb{R}^n)$ and $L^2(\mathbb{R}^n)'$, but we will leave this as understood.

Note that this result asserts that $L^1(\mathbb{R}^n)' = L^\infty(\mathbb{R}^n)$, however it makes no statement about $L^\infty(\mathbb{R}^n)'$. In fact, $L^\infty(\mathbb{R}^n)' \neq L^1(\mathbb{R}^n)$. We also comment that the result holds more generally for the spaces $L^p(E,\mu)$, where μ is a σ-finite measure on E, but for simplicity we stick to the case of $\mathbb{R}^n$.

The main work of the proof of Theorem 2.13 has already been done in the proof of the Radon–Nikodym theorem, which is the key result we shall require. We first simplify the problem by reducing to the case of positive real linear functionals. Let $L^p(\mathbb{R}^n;\mathbb{R})$ denote the subset of $L^p(\mathbb{R}^n)$ consisting of functions taking values in $\mathbb{R}$ almost everywhere. Clearly $L^p(\mathbb{R}^n;\mathbb{R})$ is a vector space over $\mathbb{R}$, and any element $f \in L^p(\mathbb{R}^n)$ can be written uniquely as $f_r + if_i$ with $f_r, f_i \in L^p(\mathbb{R}^n;\mathbb{R})$. Given a bounded (complex-)linear map $\Lambda : L^p(\mathbb{R}^n) \to \mathbb{C}$, we can define two bounded (real-)linear maps $\Lambda_r, \Lambda_i : L^p(\mathbb{R}^n;\mathbb{R}) \to \mathbb{R}$ by:

$$\Lambda_r(f) := \Re(\Lambda(f)), \quad \Lambda_i(f) := \Im(\Lambda(f))$$

and we can recover Λ from Λ_r, Λ_i by:

$$\Lambda(f_r + if_i) = \Lambda_r(f_r) - \Lambda_i(f_i) + i\left(\Lambda_r(f_i) + \Lambda_i(f_r)\right).$$

We say that a real-linear map $u : L^p(\mathbb{R}^n;\mathbb{R}) \to \mathbb{R}$ is *positive* if $u(f) \geqslant 0$ for all $f \geqslant 0$. We claim that any bounded real-linear map $u : L^p(\mathbb{R}^n;\mathbb{R}) \to \mathbb{R}$ may be written as $u = u_+ - u_-$, where $u_\pm : L^p(\mathbb{R}^n;\mathbb{R}) \to \mathbb{R}$ are bounded, positive, real-linear maps (see Exercise 2.3). In view of these facts, in order to prove Theorem 2.13 it will suffice to establish:

Lemma 2.14. *Let $1 \leqslant p < \infty$, $p^{-1} + q^{-1} = 1$. Suppose $u : L^p(\mathbb{R}^n;\mathbb{R}) \to \mathbb{R}$ is a bounded, positive (real-)linear map. Then there exists a non-negative $g \in L^q(\mathbb{R}^n;\mathbb{R})$ with $\|g\|_{L^q} = \|u\|_{(L^p)'}$ such that:*

$$u(f) = \int_{\mathbb{R}^n} f(x)g(x)dx$$

for all $f \in L^p(\mathbb{R}^n;\mathbb{R})$.

Proof. Let $\mu = e^{-|x|^2}dx$ be the Gaussian measure on the Lebesgue sets of $\mathbb{R}^n$, which has the property that $\mu(\mathbb{R}^n) < \infty$. Define, for a measurable set A:

$$\nu(A) := u(e^{-\frac{|x|^2}{p}}\mathbb{1}_A).$$

Clearly $\nu(A) \in [0,\infty]$ and $\nu(\emptyset) = 0$. Further, if $B = \cup_{n=1}^\infty A_n$ for disjoint measurable A_n, setting $B_k = \cup_{n=1}^k A_n$ we have

$$\left\| e^{-\frac{|x|^2}{p}}\mathbb{1}_B - e^{-\frac{|x|^2}{p}}\mathbb{1}_{B_k} \right\|_{L^p} = [\mu(B \setminus B_k)]^{\frac{1}{p}} \to 0$$

so by the continuity and linearity of u, we have $\nu(B_k) \to \nu(B)$. Thus ν defines a measure on the Lebesgue sets of $\mathbb{R}^n$. Further $\nu(\mathbb{R}^n) < \infty$ and moreover $\nu \ll \mu$ since if $\mu(A) = 0$ then

$$\left\| e^{-\frac{|x|^2}{p}}\mathbb{1}_A \right\|_{L^p} = [\mu(A)]^{\frac{1}{p}} = 0.$$

By the Radon–Nikodym theorem, we deduce that there exists a non-negative $G \in L^1(\mathbb{R}^n, \mu)$ such that

$$\nu(A) = \int_A G(x) d\mu = \int_A G(x) e^{-|x|^2} dx.$$

Now by linearity, we deduce that if $f = e^{-\frac{|x|^2}{p}} F$ for some simple function F then we have

$$u(f) = \int_{\mathbb{R}^n} f(x) g(x) dx.$$

where $g(x) = e^{-\frac{|x|^2}{q}} G(x)$. Now, functions of the form $e^{-\frac{|x|^2}{p}} F$, with F simple, are dense in $L^p(\mathbb{R}^n; \mathbb{R})$ and moreover we know from the boundedness of u that

$$\int_{\mathbb{R}^n} |f(x)g(x)|\, dx = \int_{\mathbb{R}^n} |f(x)|\, g(x) dx = u(|f|) \leqslant \|u\|_{L^{p'}} \|f\|_{L^p}.$$

This implies that

$$\sup\left\{ \int_{\mathbb{R}^n} |f(x)g(x)|\, dx : f \in L^p(\mathbb{R}^n; \mathbb{R}), \|f\|_{L^p} \leqslant 1 \right\} \leqslant \|u\|_{L^{p'}}$$

By Exercise 1.4 we deduce that $g \in L^q(\mathbb{R}^n; \mathbb{R})$ with $\|g\|_{L^q} \leqslant \|u\|_{L^{p'}}$. On the other hand $\|g\|_{L^q} \geqslant \|u\|_{L^{p'}}$ follows from Hölder and we're done. □

Exercise 2.3. Let $u : L^p(\mathbb{R}^n; \mathbb{R}) \to \mathbb{R}$ be a bounded, linear functional.

a) For $f \in L^p(\mathbb{R}^n; \mathbb{R}), f \geqslant 0$, define

$$\tilde{u}(f) = \sup\{u(g) : g \in L^p(\mathbb{R}^n; \mathbb{R}), 0 \leqslant g \leqslant f\}.$$

Show that $0 \leqslant \tilde{u}(f)$ and $u(f) \leqslant \tilde{u}(f) \leqslant \|u\|_{L^{p'}} \|f\|_{L^p}$, and establish

$$\tilde{u}(f + ag) = \tilde{u}(f) + a\tilde{u}(g)$$

for all $f, g \in L^p(\mathbb{R}^n; \mathbb{R})$ with $f, g \geqslant 0$ and $a \in \mathbb{R}$, $a > 0$.

b) For $f \in L^p(\mathbb{R}^n; \mathbb{R})$, define $w(f) = \tilde{u}(f^+) - \tilde{u}(f^-)$, where $f^+(x) = \max\{0, f(x)\}$, $f^-(x) = \max\{0, -f(x)\}$. Show that w is linear and bounded, and that w and $w - u$ are positive.

c) Deduce that $u = u_+ - u_-$, where $u_\pm$ are bounded, positive, linear functionals.

2.2.2 The Riesz Representation Theorem for spaces of continuous functions

Another space whose dual space can be conveniently described is $C_c^0(\mathbb{R}^n)$, the space of continuous, compactly supported, functions on $\mathbb{R}^n$ equipped with the supremum norm. By a similar reduction to §2.2.1, we can reduce to the problem of understanding positive bounded functionals on $C_c^0(\mathbb{R}^n; \mathbb{R})$.

A classical result known, somewhat confusingly, as the Riesz Representation Theorem shows that any positive functional on $C_c^0(\mathbb{R}^n; \mathbb{R})$ can be represented as integration against a suitable measure. To motivate this result, we first suppose that we are given a σ-algebra $\mathcal{M}$ on $\mathbb{R}^n$ and a measure μ such that $\mu(\mathbb{R}^n) < \infty$. We will also require that the measure space is *regular*:

Definition 2.3. *Suppose that E is a topological space, and $\mathcal{E}$ is a σ-algebra on E which contains the Borel algebra. Then a measure μ defined on $(E, \mathcal{E})$ is regular if for any $A \in \mathcal{E}$, and any $\epsilon > 0$ we can find a closed set C and an open set O such that $C \subset A \subset O$ and:*

$$\mu(O \setminus C) < \epsilon.$$

Since $\mathcal{B}(\mathbb{R}^n) \subset \mathcal{M}$, we know that any $f \in C_c^0(\mathbb{R}^n; \mathbb{R})$ is measurable. The map $\Lambda : C_c^0(\mathbb{R}^n; \mathbb{R}) \to \mathbb{R}$ given by:

$$\Lambda(f) = \int_{\mathbb{R}^n} f(x) d\mu(x) \tag{2.3}$$

is then a positive, bounded linear map. Now, suppose we are given the map Λ, can we recover the measure μ? We note that, if we could set $f = \mathbb{1}_A$ for $A \in \mathcal{M}$, then

$$\text{‘}\Lambda(\mathbb{1}_A) = \int_{\mathbb{R}^n} \mathbb{1}_A(x) d\mu(x) = \mu(A)\text{’},$$

however $\mathbb{1}_A \notin C_c^0(\mathbb{R}^n; \mathbb{R})$. At least for certain sets, however, we can approximate $\mathbb{1}_A$ from below by elements of $C_c^0(\mathbb{R}^n; \mathbb{R})$. Suppose $O \subset \mathbb{R}^n$ is open, and for $k \in \mathbb{N}$ let $O_k = O \cap \{|x| < k\}$. Define:

$$\chi_k(x) := \begin{cases} 1, & x \in O_k,\ d(x, O_k^c) \geqslant k^{-1} \\ kd(x, O_k^c) & x \in O_k,\ d(x, O_k^c) < k^{-1} \\ 0 & x \in O_k^c \end{cases}$$

Then $\chi_k \in C_c^0(\mathbb{R}^n; \mathbb{R})$ and $\chi_k(x)$ increases monotonically to $\mathbb{1}_O$. Thus, by the monotone convergence theorem,

$$\mu(O) = \lim_{k \to \infty} \int_{\mathbb{R}^n} \chi_k(x) d\mu(x) = \lim_{k \to \infty} \Lambda(\chi_k).$$

This shows in particular that

$$\mu(O) = \sup\{\Lambda(g) : g \in C_c^0(\mathbb{R}^n; \mathbb{R}),\ 0 \leqslant g \leqslant \mathbb{1}_O\}. \tag{2.4}$$

Now, since μ is regular, it suffices to know how to compute $\mu(O)$ for open sets in order to find $\mu(A)$ for any $A \in \mathcal{M}$. We have shown:

Lemma 2.15. *Suppose we are given a σ-algebra $\mathcal{M}$ containing $\mathcal{B}(\mathbb{R}^n)$ on $\mathbb{R}^n$ and a regular measure μ such that $\mu(\mathbb{R}^n) < \infty$. Then $\Lambda : C_c^0(\mathbb{R}^n; \mathbb{R}) \to \mathbb{R}$ given by (2.3) defines a bounded, positive, linear operator. Furthermore, μ is uniquely determined by Λ.*

Riesz Representation Theorem makes the stronger statement that *all* positive bounded linear operators take the form (2.3) for some regular measure.

Theorem 2.16. *Given a positive bounded linear operator $\Lambda : C_c^0(\mathbb{R}^n; \mathbb{R}) \to \mathbb{R}$, there exists a σ-algebra $\mathcal{M}$ on $\mathbb{R}^n$, containing $\mathcal{B}(\mathbb{R}^n)$, and a unique regular measure μ such that $\mu(\mathbb{R}^n) < \infty$ and:*

$$\Lambda(f) = \int_{\mathbb{R}^n} f(x) d\mu(x).$$

We shall not include the proof of this result here, as it is fairly long and technical, and not especially enlightening. The main idea is to use (2.4) to define μ on open sets, and then use Carathéodory's Theorem (or some equivalent approach) to complete this measure. Those interested in the proof will find it in Chapter 2 of Rudin's Real and Complex Analysis.

2.2.3 The strong, weak and weak-$*$ topologies

If X is a Banach space, then X', with the dual norm, is also a Banach space. Thus both spaces are naturally equipped with a topology which makes them topological vector spaces. For certain purposes, however, we may wish to introduce an alternative topology on X or X'. For example, a hugely useful result in the analysis of $\mathbb{R}^n$ is the Bolzano–Weierstrass theorem:

Theorem 2.17. *Let $(x_k)_{k=1}^{\infty}$ with $x_k \in \mathbb{R}^n$ be a bounded sequence. Then $(x_k)_{k=1}^{\infty}$ has a convergent subsequence.*

This result is *not* true when $\mathbb{R}^n$ is replaced by an infinite dimensional Banach space, so the closed unit ball in such a space is not compact. This is quite inconvenient for many problems: for example in the calculus of variations one often wishes to minimise some continuous function defined on a Banach or Hilbert space. Without compactness as a tool, this can be difficult to achieve.

> **Exercise 2.4.** Suppose X is a normed space, and $V \subset X$ is a closed proper subspace of X and let $0 < \alpha < 1$. Show that there exists $x \in X$ with $\|x\| = 1$ such that $\|x - y\| \geqslant \alpha$ for all $y \in V$. Deduce that the Bolzano–Weiserstrass theorem does not hold if X is an infinite dimensional Banach space.
> *[The first result above is known as Riesz' Lemma]*

One way to restore (a version of) compactness for the closed unit ball in X is to consider a different topology defined on X. In order to describe new topologies on X we will make use of seminorms. For full details of this discussion, see §A.2

Definition 2.4. *A seminorm on a vector space X over a field $\Phi = \mathbb{C}$ or $\mathbb{R}$ is a map $p : X \to \mathbb{R}$ satisfying:*

i) $p(x + y) \leqslant p(x) + p(y)$, for all $x, y \in X$,

ii) $p(\lambda x) = |\lambda|\, p(x)$ for all $x \in X$, $\lambda \in \Phi$,

iii) $p(x) \geqslant 0$ for all $x \in X$.

A family $\mathscr{P}$ of seminorms is said to be separating *if for every $x \in X$ with $x \neq 0$, there exists $p \in \mathscr{P}$ with $p(x) \neq 0$.*

Strictly speaking, condition *iii*) follows from *i*) and *ii*) (check this!), but we include it in the definition for convenience. Given a separating family of seminorms, we can construct a topology $\tau_{\mathscr{P}}$ as follows. First, for $p \in \mathscr{P}$, $n \in \mathbb{N}$, we define the set $V(p,n) \subset X$ by

$$V(p,n) := \left\{ x \in X : p(x) < \frac{1}{n} \right\}$$

We let $\dot{\beta}$ be the collection of finite intersections of $V(p,n)$'s, and $\beta = \{x + B : B \in \dot{\beta}\}$.

Theorem 2.18. *Let $\mathscr{P}$ be a separating family of seminorms. The collection of sets β, as described above, is a base for a Hausdorff topology $\tau_{\mathscr{P}}$ on X such that the vector space operations are continuous, and each $p \in \mathscr{P}$ is continuous.*

A topological space $(X, \tau_{\mathscr{P}})$ constructed in the manner above is known as a *locally convex topological vector space.* If $\mathscr{P} = \{p_i\}_{i=1}^{\infty}$ is countable, then the topology is a metric topology, with metric given by:

$$d(x,y) = \sum_{i=1}^{\infty} 2^{-i} \frac{p_i(x-y)}{1 + p_i(x-y)}$$

If this metric is complete, we say that $(X, \tau_{\mathscr{P}})$ is a *Fréchet* space.

> **Exercise 2.5.** Let $\mathscr{P}$ be a separating family of seminorms on a vector space X. Show that a sequence $(x_k)_{k=1}^{\infty}$ with $x_k \in X$ converges to $x \in X$ in the topology $\tau_{\mathscr{P}}$ if and only if $p(x_k - x) \to 0$ for all $p \in \mathscr{P}$.

For a Banach space X, a trivial family of separating seminorms is given by $\mathscr{P} = \{\|\cdot\|\}$. The topology $\tau_s := \tau_{\mathscr{P}}$ induced by this family is simply the usual norm topology. In this context, we sometimes refer to this as the *strong* topology on X. A sequence $(x_k)_{k=1}^{\infty}$ with $x_k \in X$ converges to x in the strong topology if

$$\|x_k - x\| \to 0.$$

An alternative topology on X is given by making use of X' to construct a family of seminorms. It is straightforward to verify that if $\Lambda \in X'$ then $p_\Lambda : x \mapsto |\Lambda(x)|$ is a seminorm. Setting

$$\mathscr{P} := \{p_\Lambda : \Lambda \in X'\}$$

we have a family of seminorms. Moreover, it is separating, since X' separates points of X. Thus $\tau_w := \tau_{\mathscr{P}}$ makes X into a locally convex topological space. This topology is known as the *weak* topology. With respect to the weak topology, the elements of X' are still continuous, however convergence of sequences in the weak topology differs from the strong topology. A sequence $(x_k)_{k=1}^{\infty}$ with $x_k \in X$ converges to x in the weak topology if

$$|\Lambda(x_k - x)| \to 0, \qquad \text{for all } \Lambda \in X'.$$

When $(x_k)_{k=1}^{\infty}$ converges to x in the weak topology, we write $x_k \rightharpoonup x$.

Now, X' is a Banach space itself in a natural fashion, and so has its own associated strong and weak topologies. It also has a further topology, known as the *weak-$*$* topology (pronounced 'weak star'). To define this topology, we note that for each $x \in X$, we can define a seminorm $p_x : X' \to \mathbb{R}$ by $p_x(\Lambda) = |\Lambda(x)|$. Setting

$$\mathscr{P} := \{p_x : x \in X\}$$

we have a family of seminorms, which is separating since if $\Lambda \in X'$, $\Lambda \neq 0$, then there exists $x \in X$ such that $\Lambda(x) \neq 0$. The associated topology we denote τ_{w*}. A sequence $(\Lambda_k)_{k=1}^{\infty}$ with $\Lambda_k \in X'$ converges to Λ in the weak-$*$ topology if:

$$|\Lambda_k(x) - \Lambda(x))| \to 0, \qquad \text{for all } x \in X.$$

When $(\Lambda_k)_{k=1}^{\infty}$ converges to Λ in the weak-$*$ topology, we write $\Lambda_k \overset{*}{\rightharpoonup} \Lambda$. Note that if X is reflexive, then $X'' = X$, and the weak and weak-$*$ topologies coincide.

> **Exercise 2.6.** Suppose that X is a Banach space, and let $(\Lambda_k)_{k=1}^{\infty}$ be a sequence with $\Lambda_k \in X'$. Show that:
>
> $$\Lambda_k \to \Lambda \quad \Longrightarrow \quad \Lambda_k \rightharpoonup \Lambda \quad \Longrightarrow \quad \Lambda_k \overset{*}{\rightharpoonup} \Lambda.$$
>
> (*) Show the stronger statement that $\tau_{w*} \subset \tau_w \subset \tau_s$, where $\tau_{w*}, \tau_w, \tau_s$ are the weak-$*$, weak and strong topologies on X' respectively.

As an example, suppose that $1 \leqslant p < \infty$. Then we know that $L^p(\mathbb{R}^n)' = L^q(\mathbb{R}^n)$ with $p^{-1} + q^{-1} = 1$. If $(f_i)_{i=1}^{\infty}$ is a sequence of functions $f_i \in L^p(\mathbb{R}^n)$, then of course $f_i \to f$ in L^p if:

$$\|f_i - f\|_{L^p} \to 0.$$

On the other hand, $f_i \rightharpoonup f$ in L^p if

$$\int_{\mathbb{R}^n} g(x) f_i(x) dx \to \int_{\mathbb{R}^n} g(x) f(x) dx, \qquad \text{for all } g \in L^q(\mathbb{R}^n).$$

If $1 < p < \infty$, $L^p(\mathbb{R}^n)$ is reflexive, so the weak-$*$ topology that arises from viewing $L^p(\mathbb{R}^n)$ as the dual of $L^q(\mathbb{R}^n)$ agrees with the weak topology. We have not identified any space X such that $X' = L^1(\mathbb{R}^n)$, so no weak-$*$ topology on L^1 is available to us. Since $L^1(\mathbb{R}^n)' = L^\infty(\mathbb{R}^n)$, we can consider the weak-$*$ topology on $L^\infty(\mathbb{R}^n)$. We have $f_i \overset{*}{\rightharpoonup} f$ in L^∞ if:

$$\int_{\mathbb{R}^n} g(x) f_i(x) dx \to \int_{\mathbb{R}^n} g(x) f(x) dx, \qquad \text{for all } g \in L^1(\mathbb{R}^n).$$

Since we don't have a concrete realisation of $L^\infty(\mathbb{R}^n)'$, we don't have a simple description of weak convergence in this space (other than the abstract condition $\Lambda(f_i) \to \Lambda(f)$ for all $\Lambda \in L^\infty(\mathbb{R}^n)')$.

> **Exercise 2.7.** For a bounded measurable set $E \subset \mathbb{R}^n$ of positive measure, and any $f \in L^1_{loc.}(\mathbb{R}^n)$, define the mean of f on E to be:
>
> $$⨍_E f(x) dx = \frac{1}{|E|} \int_E f(x) dx.$$

Suppose $1 < p < \infty$ and let $(f_j)_{j=1}^{\infty}$ be a bounded sequence in $L^p(\mathbb{R}^n)$. Show that $f_j \rightharpoonup f$ for some $f \in L^p(\mathbb{R}^n)$ if and only if

$$\fint_E f_j(x)dx \to \fint_E f(x)dx$$

for all bounded measurable sets $E \subset \mathbb{R}^n$ of positive measure.

Exercise 2.8. Suppose $(H, (\cdot,\cdot))$ is an infinite dimensional Hilbert space and let $(x_i)_{i=1}^{\infty}$ be a sequence with $x_i \in H$.

i) Show that $x_i \rightharpoonup x$ if and only if $(y, x_i) \to (y, x)$ for all $y \in H$.

ii) Show there exists a sequence such that $x_i \rightharpoonup 0$, but $x_i \not\to 0$.

iii) Suppose $x_i \rightharpoonup x$. Show that

$$\|x\| \leqslant \liminf_{i\to\infty} \|x_i\|,$$

and $\|x_i\| \to \|x\|$ iff $x_i \to x$.

2.2.4 Compactness, Banach–Alaoglu

As we have discussed above, if X is an infinite dimensional Banach space, then the closed unit ball $\overline{B} = \{x : \|x\| \leqslant 1\}$ is not compact. This is unfortunate, and we would like to try and restore compactness in some way. There are essentially two (related) approaches: we can either restrict our attention to a subset of $\overline{B}$ for which we have compactness, or else we can weaken the topology on $\overline{B}$.

To explain this, let us recall the Arzelà–Ascoli theorem. We set $I = [0, 1]$

Theorem 2.19. *Suppose $(f_k)_{k=1}^{\infty}$ is a sequence of continuous functions $f_k : I \to \mathbb{C}$ which is bounded, i.e. for all k:*

$$\sup_{x\in I} |f_k(x)| \leqslant M$$

and equicontinuous: for all $\epsilon > 0$, there exists $\delta > 0$ such that for all k and all $|x - y| < \delta$ we have

$$|f_k(x) - f_k(y)| < \epsilon.$$

Then $(f_k)_{k=1}^{\infty}$ admits a uniformly convergent subsequence.

To put this into the language we have been discussing, recall that for $0 < \gamma \leq 1$, we say a continuous function $f : I \to \mathbb{C}$ is γ-Hölder continuous if

$$\|f\|_{C^{0,\gamma}} := \sup_{x\in I} |f(x)| + \sup_{x,y\in I, x\neq y} \frac{|f(x) - f(y)|}{|x - y|^{\gamma}} < \infty.$$

The set $C^{0,\gamma}(I)$, of γ-Hölder continuous functions, is a Banach space with this norm, and $C^{0,\gamma}(I) \subset C^0(I)$. A consequence of the Arzelà–Ascoli theorem is the following:

Corollary 2.20. *The closed unit ball in $C^{0,\gamma}(I)$ is compact in the C^0-topology.*

Proof. Since the C^0-topology is a metric topology, compactness is equivalent to sequential compactness. If $(f_k)_{k=1}^{\infty}$ is a sequence with $f_k \in C^{0,\gamma}(I)$ satisfying

$$\|f_k\|_{C^{0,\gamma}} \leqslant 1,$$

then $(f_k)_{k=1}^{\infty}$ is a bounded equicontinuous sequence of functions, so admits a uniformly convergent subsequence, i.e. a subsequence which converges in the C^0-topology to some f. It is a short exercise to check that $f \in C^{0,\gamma}$ with $\|f\|_{C^{0,\gamma}} \leqslant 1$. □

This gives us a paradigmatic example of a compactness result for Banach spaces: the closed unit ball is compact, but only in a weaker topology than the strong topology. The central result is the Banach–Alaoglu theorem:

Theorem 2.21. *Let X be a normed space, and let $\overline{B}' = \{\Lambda \in X' : \|\Lambda\|_{X'} \leqslant 1\}$ be the closed unit ball in X'. Then $\overline{B}'$ is compact in the weak-$*$ topology on X'.*

This result in its full generality is typically proven using Tychonoff's theorem, which relies on (a version of) the axiom of choice. We will content ourselves with the proof in the case where X is a separable Banach space for which it is possible to give a constructive proof. The majority of the applications of Banach–Alaoglu that arise in (for example) the calculus of variations or PDE are covered by this special case.

We first note that if X is a separable Banach space, then the weak-$*$ topology on $\overline{B}'$ is in fact a metric topology.

Lemma 2.22. *Let X be a separable Banach space, with a countable dense subset $D = \{x_k\}_{k=1}^{\infty}$. Let $\tilde{\mathscr{P}} = \{p_k\}_{k=1}^{\infty}$ be the family of seminorms on X' defined by:*

$$p_k : \Lambda \mapsto |\Lambda(x_k)|,$$

and let $\tau_{\tilde{\mathscr{P}}}$ be the associated topology. Then $\tau_{\tilde{\mathscr{P}}}|_{\overline{B}'} = \tau_{w*}|_{\overline{B}'}$. In particular, the weak-$*$ topology on $\overline{B}'$ is a metric topology.

Proof. From the definition of the topologies, it is immediate that every open set in $\tau_{\tilde{\mathscr{P}}}$ is open in τ_{w*}, thus $\tau_{\tilde{\mathscr{P}}}|_{\overline{B}'} \subset \tau_{w*}|_{\overline{B}'}$. For any $x \in X$, $n \in \mathbb{N}$ let

$$V(x,n) := \left\{\Lambda \in \overline{B}' : |\Lambda(x)| < \frac{1}{n}\right\}.$$

In order to show $\tau_{\tilde{\mathscr{P}}}|_{\overline{B}'} \supset \tau_{w*}|_{\overline{B}'}$ it suffices to show that for any $x \in X$, $n \in \mathbb{N}$, we can find $x_k \in D$, $m \in \mathbb{N}$ such that

$$V(x_k, m) \subset V(x,n).$$

Fix $x \in X$. For any $\epsilon > 0$, there exists $x_i \in D$ such that $\|x - x_i\| < \epsilon$. If $\Lambda \in V(x_i, m)$, then:

$$|\Lambda(x)| = |\Lambda(x - x_i) + \Lambda(x_i)| \leqslant \|\Lambda\| \, \|x - x_i\| + |\Lambda(x_i)| < \epsilon + \frac{1}{m}.$$

Taking $\epsilon < 1/(2n)$, $m > 2n$ we have $\Lambda \in V(x,n)$ and we're done. □

This result is useful in two ways. Firstly, we see that a sequence $(\Lambda_j)_{j=1}^{\infty}$ with $\Lambda_j \in \overline{B}'$ converges to Λ in the weak-$*$ topology if and only if:

$$\Lambda_j(x_k) \to \Lambda(x_k) \text{ as } j \to \infty, \text{ for all } k.$$

Secondly, since the weak-$*$ topology on $\overline{B}'$ is metric, compactness is equivalent to sequential compactness. We can establish sequential compactness with a very similar method to the proof of the Arzelà–Ascoli theorem.

Theorem 2.23. *Let X be a separable Banach space. Let $(\Lambda_j)_{j=1}^{\infty}$ be a sequence with $\Lambda_j \in \overline{B}'$. Then there exists a subsequence $(\Lambda_{j_k})_{k=1}^{\infty}$ and $\Lambda \in \overline{B}'$ such that $\Lambda_j \overset{*}{\rightharpoonup} \Lambda$.*

Proof. Let $D = \{x_k\}_{k=1}^{\infty}$ be a countable dense subset. Consider the sequence $(\Lambda_j(x_1))_{j=1}^{\infty}$. This is a uniformly bounded sequence of complex numbers, since:

$$|\Lambda_j(x_1)| \leqslant \|\Lambda_j\| \, \|x_1\| \leqslant \|x_1\| \, .$$

Thus, by Bolzano–Weierstrass, there exists a subsequence $(\Lambda_{j_k}(x_1))_{k=1}^{\infty}$ and a number $\Lambda(x_1) \in \mathbb{C}$ with $|\Lambda(x_1)| \leqslant \|x_1\|$ such that:

$$\Lambda_{j_k}(x_1) \to \Lambda(x_1).$$

We write $\Lambda_{1,k} := \Lambda_{j_k}$, then $(\Lambda_{1,j})_{j=1}^{\infty}$ is a subsequence of $(\Lambda_j)_{j=1}^{\infty}$. By a similar argument, we can find a subsequence $(\Lambda_{1,j_k})_{j=1}^{\infty}$ of $(\Lambda_{1,j})_{j=1}^{\infty}$ such that $\Lambda_{1,j_k}(x_2) \to \Lambda(x_2)$. We write $\Lambda_{2,k} := \Lambda_{1,j_k}$. Continuing in this fashion, we construct for each $l \geqslant 1$ a sequence $(\Lambda_{l,j})_{j=1}^{\infty}$, and a complex number $\Lambda(x_l)$ with $|\Lambda(x_l)| \leqslant \|x_l\|$ with the property that $(\Lambda_{l,j})_{j=1}^{\infty}$ is a subsequence of $(\Lambda_{l-1,j})_{j=1}^{\infty}$, and $\Lambda_{l,j}(x_k) \to \Lambda(x_k)$ as $j \to \infty$ for all $l \leqslant k$.

Now, consider $(\Lambda_{j,j})_{j=1}^{\infty}$. This is a subsequence of $(\Lambda_j)_{j=1}^{\infty}$ with the property that for each $x \in D$ we have:

$$\Lambda_{j,j}(x) \to \Lambda(x).$$

If we can show that there exists $\tilde{\Lambda} \in \overline{B}'$ with $\tilde{\Lambda}(x) = \Lambda(x)$ for all $x \in D$, then we are done. We first claim that $\Lambda : D \to \mathbb{C}$ is uniformly continuous. Fix $\epsilon > 0$, and suppose $x, y \in D$ with $\|x - y\| < \frac{\epsilon}{3}$. Since $\Lambda_{j,j}(x) \to \Lambda(x)$ and $\Lambda_{j,j}(y) \to y$, there exists k such that for all $j \geqslant k$ we have $|\Lambda_{j,j}(x) - \Lambda(x)| < \frac{\epsilon}{3}$ and $|\Lambda_{j,j}(y) - \Lambda(y)| < \frac{\epsilon}{3}$. For such a j we estimate:

$$|\Lambda(x) - \Lambda(y)| \leqslant |\Lambda(x) - \Lambda_{j,j}(x)| + |\Lambda_{j,j}(y) - \Lambda(y)| + |\Lambda_{j,j}(x - y)| < \epsilon.$$

we conclude that $\Lambda : D \to \mathbb{C}$ is continuous. Thus Λ extends to a continuous function $\tilde{\Lambda} : X \to \mathbb{C}$. We abuse notation and drop the tilde at this point. Next, we claim Λ is linear. Suppose $x, y \in X$ and $a \in \mathbb{C}$ and for $z = x + ay$ estimate:

$$\begin{aligned} |\Lambda(z) - \Lambda(x) - a\Lambda(y)| \leqslant & \left|\Lambda(z) - \Lambda(z')\right| + \left|\Lambda(x) - \Lambda(x')\right| + |a| \left|\Lambda(y) - \Lambda(y')\right| \\ & + \left|\Lambda(z') - \Lambda_{j,j}(z')\right| + \left|\Lambda(x') - \Lambda_{j,j}(x')\right| + |a| \left|\Lambda(y') - \Lambda_{j,j}(y')\right| \\ & + \left|\Lambda_{j,j}(z' - x' - ay')\right| . \end{aligned}$$

By choosing $x', y', z' \in D$ sufficiently close to x, y, z respectively we may arrange that the first and final line are arbitrarily small. Taking j sufficiently large we see that the middle line can also be made arbitrarily small. We conclude that $\Lambda(x + ay) = \Lambda(x) + a\Lambda(y)$. Thus $\Lambda : X \to \mathbb{C}$ is a continuous linear map. Finally, since D is dense in X we have:

$$\|\Lambda\| = \sup_{x \in X, \|x\| \leqslant 1} |\Lambda(x)| = \sup_{x \in D, \|x\| \leqslant 1} |\Lambda(x)| \leqslant 1.$$

Hence $\Lambda \in \overline{B}'$. □

As a corollary, we find the following compactness result for the Lebesgue spaces:

Corollary 2.24. *Suppose $1 < p \leqslant \infty$, and let $(f_j)_{j=1}^{\infty}$ be a sequence of functions $f_j \in L^p(\mathbb{R}^n)$ satisfying*

$$\|f_j\|_{L^p} \leqslant K.$$

Then there exists $f \in L^p(\mathbb{R}^n)$ and a subsequence $(f_{j_k})_{k=1}^{\infty}$ such that $\|f\|_{L^p} \leqslant K$ and

$$\int_{\mathbb{R}^n} g(x) f_{j_k}(x) dx \to \int_{\mathbb{R}^n} g(x) f(x) dx$$

for all $g \in L^q(\mathbb{R}^n)$, where $p^{-1} + q^{-1} = 1$.

Proof. Apply the previous result to f_j/K. □

Note that we do not have a corresponding compactness result for $L^1(\mathbb{R}^n)$. It is possible to gain some compactness by considering $L^1(\mathbb{R}^n)$ as a subspace of the dual space of $C^0_c(\mathbb{R}^n)$, however the limiting objects constructed in this way are typically measures, not elements of $L^1(\mathbb{R}^n)$.

> **Exercise 2.9.** Construct a bounded sequence $(f_i)_{i=1}^{\infty}$ of functions $f_i \in L^1(\mathbb{R})$ such that no subsequence is weakly convergent.

2.3 Hahn–Banach

The next result we shall cover is the Hahn–Banach theorem. This modest-seeming result permits us to extend a bounded linear functional defined on a subspace, M, of a vector space X into a bounded linear functional defined on the whole space. While it seems like this should be straightforward, in full generality it requires the axiom of choice (or at least some method of transfinite induction). We will proceed modestly by first showing that we can extend a linear functional in one direction.

We first show that we can reduce to the case of a real vector space. Suppose X is a complex vector space. We note that X is also a real vector space in a natural fashion. A real-linear map $\ell : X \to \mathbb{R}$ is a map satisfying:

$$\ell(x + ay) = \ell(x) + a\ell(y), \qquad \text{for all } x, y \in X, a \in \mathbb{R}.$$

Suppose $\Lambda : X \to \mathbb{C}$ is a complex-linear map, then $\ell(x) = \Re(\Lambda(x))$ defines a real-linear map on X. Conversely, given a real-linear map $\ell : X \to \mathbb{R}$, we have that:

$$\Lambda(x) = \ell(x) - i\ell(ix)$$

defines a complex-linear map $\Lambda : X \to \mathbb{C}$, with $\Re(\Lambda(x)) = \ell(x)$. Thus provided we can establish a result which allows us to extend linear functionals in a bounded fashion on a real vector space, we immediately have a corresponding result for the complex setting.

At this point we should address our use of the word 'bounded' above. When X is a Banach space, then we have already discussed what it means for a functional to be bounded. It turns out to be useful to consider a slightly more general notion of boundedness at this stage, however. For this we introduce

Definition 2.5. *A sublinear functional on a real vector space X is a map $p : X \to \mathbb{R}$ satisfying*

$$p(x+y) \leqslant p(x) + p(y), \qquad p(tx) = tp(x),$$

for any $x, y \in X$ and $t \geq 0$.

For example if $\Lambda : X \to \mathbb{R}$ linear, then $p(x) = |\Lambda(x)|$ is sublinear. Any semi-norm (hence any norm) is sublinear, but the converse doesn't hold, so be careful!

We first show that we can extend a linear functional ℓ defined on a subspace in one direction, maintaining a bound by a sublinear functional p. We will work with one-sided bounds of the form $\ell(x) \leqslant p(x)$, but we note that this implies the two-sided bound $-p(-x) \leqslant \ell(x) \leqslant p(x)$.

Lemma 2.25. *Let X be a real vector space, $p : X \to \mathbb{R}$ sublinear and $M \subset X$ a subspace. Suppose $\ell : M \to \mathbb{R}$ is linear and satisfies $\ell(y) \leqslant p(y)$ for all $y \in M$. Fix $x \in X \setminus M$, then setting $\tilde{M} = span\,\{M, x\}$, there exists a linear operator $\tilde{\ell} : \tilde{M} \to \mathbb{R}$ such that*

$$\tilde{\ell}(z) \leqslant p(z) \quad \text{for all } z \in \tilde{M}.$$

and

$$\ell(y) = \tilde{\ell}(y), \quad \text{for all } y \in M.$$

Proof. Any $z \in \tilde{M}$ can be uniquely written as $z = \lambda x + y$ for $y \in M$, so to define the extension $\tilde{\ell}$, by linearity it suffices to specify $\tilde{\ell}(x) = a$ as then $\tilde{\ell}(\lambda x + y) = \lambda a + \ell(y)$. Suppose $y, z \in M$, then

$$\ell(y) + \ell(z) = \ell(y+z) \leqslant p(y+z) \leqslant p(y-x) + p(z+x)$$

and hence

$$\ell(y) - p(y-x) \leqslant p(z+x) - \ell(z). \tag{2.5}$$

Let

$$a = \sup_{y \in M} \left(\ell(y) - p(y-x) \right).$$

This is well defined by (2.5) and further we deduce

$$\ell(y) - a \leqslant p(y-x), \qquad \ell(z) + a \leqslant p(z+x)$$

for all $y, z \in M$. If $\lambda > 0$, replace z with $\lambda^{-1}y$ and multiply by λ. If $\lambda < 0$ replace y with $-\lambda^{-1}y$ and multiply by $-\lambda$ to deduce

$$\ell(y) + a\lambda \leqslant p(y + \lambda x),$$

holds for all $y \in M, \lambda \in \mathbb{R}$. □

Now, if dim X/M is finite, this result allows us to iteratively extend $\ell : M \to \mathbb{R}$ to the whole space in a finite number of steps. If X is infinite, but *separable*, then it's possible to construct an extension inductively (try it!). If, however, X is not infinite, then (speaking loosely) it's not possible to exhaust X with a countable number of finite extensions. We require some way to make an inductive type argument in a non-countable setting. There are several approaches to this, all of which require the axiom of choice. We shall use Zorn's Lemma[4]. For this we require some background.

> **Exercise($*$).** Suppose X is a *separable* real Banach space. Prove the Hahn–Banach theorem on X *without* invoking the axiom of choice through Zorn's Lemma (or equivalent).

2.3.1 Zorn's Lemma

Zorn's Lemma is a statement concerning *partial orderings* of a set S.

Definition 2.6. *Let S be a set. Then a* partial order *on S is a binary relation $\leq$ satisfying, for any $a, b, c \in S$:*

i) $a \leq a$ for all $a \in S$. (Reflexivity)

ii) If $a \leq b$ and $b \leq a$, then $a = b$. (Antisymmetry)

iii) If $a \leq b$ and $b \leq c$, then $a \leq c$. (Transitivity)

A set with a partial order is called a partially ordered set, or *poset*. Note that we do not assert that for any a, b either $a \leq b$ or $b \leq a$. If this does hold, we say $\leq$ is a *total order*.

A subset T of a partially ordered set which is totally ordered is called a *chain*. An element $u \in S$ is an *upper bound* for $T \subset S$ if $a \leq u$ for all $a \in T$. A *maximal element* of S is an element $m \in S$ such that $m \leq x$ implies $x = m$.

Example 4. *a) If S is any set, then the power set 2^S is a poset, with $\leq$ given by inclusion, i.e. $A \leq B$ iff $A \subset B$. S is a maximal element.*

b) The real numbers with their usual order is a totally ordered set, with no maximal element.

c) The collection S of open balls in $\mathbb{R}^n$ is a poset with order given by inclusion. The subset

$$T = \{B_r(0) \subset \mathbb{R}^n : 0 < r \leqslant 1\}$$

is a chain. $B_1(0)$ is a maximal element, and $B_2(0)$ is an upper bound.

Zorn's Lemma can now be stated as:

---

[4]For a good discussion of how and why Zorn's Lemma is useful, see Prof. Gower's blog: https://gowers.wordpress.com/2008/08/12/how-to-use-zorns-lemma/

Proposition 1. *Let $(S, \leq)$ be a partially ordered set in which every chain has an upper bound. Then $(S, \leq)$ contains at least one maximal element.*

We shall not prove this claim. In fact, it is equivalent (within the Zermelo-Fraenkel framework) to the axiom of choice, so we can reasonably treat Zorn's Lemma as an axiom itself. The proof of the Hahn–Banach theorem we give below is a typical application of Zorn's Lemma.

Theorem 2.26 (Hahn–Banach)**.** *Let X be a real vector space, $p : X \to \mathbb{R}$ sublinear, and $M \subset X$ a subspace. Suppose $\ell : M \to \mathbb{R}$ is linear and satisfies $\ell(y) \leqslant p(y)$ for all $y \in M$. Then there exists a linear operator $\tilde{\ell} : X \to \mathbb{R}$ such that*

$$\tilde{\ell}(z) \leqslant p(z) \quad \text{for all } z \in X.$$

and

$$\ell(y) = \tilde{\ell}(y), \quad \text{for all } y \in M.$$

Proof. We consider the set S of extensions of ℓ to a linear subspace of X. That is a pair $(N, \ell^*) \in S$ if:

i) N is a linear subspace of X containing M.

ii) $\ell^* : N \to \mathbb{R}$ is a linear map.

iii) $\ell^*(x) \leqslant p(x)$ for all $x \in N$

iv) $\ell^*(y) = \ell(y)$ for all $y \in M$

S is a poset, with the partial ordering given by $(N_1, \ell_1) \leq (N_2, \ell_2)$ if N_1 is a subspace of N_2 and $\ell_1(x) = \ell_2(x)$ for all $x \in N_1$. Suppose that T is a totally ordered subset of S. We define $(\mathcal{N}, L) \in S$ by:

$$\mathcal{N} = \bigcup_{(N,\ell^*)\in T} N,$$

and for any $x \in \mathcal{N}$, we define $L(x) = \ell^*(x)$, where $(N, \ell^*) \in T$ with $x \in N$. This is well defined since T is totally ordered, and moreover we have $(N, \ell^*) \leq (\mathcal{N}, L)$ for all $(N, \ell^*) \in T$, thus T has an upper bound. By Zorn's Lemma, S has a maximal element, $(\mathscr{N}, \tilde{\ell})$. We claim that $\mathscr{N} = X$. Suppose not, then there exists $x \in X \setminus \mathscr{N}$ and we can extend $\tilde{\ell}$ to a functional $\tilde{\ell}^*$ on $\mathscr{N}^* = \text{span}\,\{\mathscr{N}, x\}$ by Lemma 2.25. Then $(\mathscr{N}, \tilde{\ell}) \leq (\mathscr{N}^*, \tilde{\ell}^*)$, but $(\mathscr{N}, \tilde{\ell}) \neq (\mathscr{N}^*, \tilde{\ell}^*)$, contradicting the maximality of $(\mathscr{N}, \tilde{\ell})$. Thus $\tilde{\ell} : X \to \mathbb{R}$ is the extension we seek. □

Notice that this proof of the Hahn–Banach theorem is *non-constructive*: while we assert the existence of at least one extension, the proof provides no mechanism to construct a particular example. This is typical of proofs which invoke the axiom of choice through Zorn's Lemma (or otherwise).

Corollary 2.27. *Let X be a Banach space over Φ, where $\Phi = \mathbb{R}$, or $\mathbb{C}$, and $M \subset X$ be a subspace. Let $\Lambda : M \to \Phi$ be a bounded linear operator. Then there exists a bounded linear operator $\tilde{\Lambda} : X \to \Phi$ with $\|\Lambda\|_{M'} = \|\tilde{\Lambda}\|_{X'}$ such that $\Lambda(y) = \tilde{\Lambda}(y)$ for all $y \in M$.*

Proof. If $\Phi = \mathbb{R}$, then we may apply the previous result with $p(x) = \|\Lambda\|\|x\|$.

If $\Phi = \mathbb{C}$, then recall that we may write $\Lambda(x) = \ell(x) - i\ell(ix)$ for a real-linear map $\ell(x) = \Re(\Lambda(x))$. Further, by noting that $|\Lambda(x)| = \ell(e^{i\theta}x)$ for suitable θ, we can see that for any subspace $N \subset X$

$$\sup_{x\in N, \|x\|\leqslant 1} |\Lambda(x)| = \sup_{x\in N, \|x\|\leqslant 1} |\ell(x)|\,,$$

and we may apply the $\Phi = \mathbb{R}$ result to ℓ. □

We will now establish some more geometric consequences of the Hahn–Banach theorem that go by the name of separation theorems. The first of these is related to the hyperplane separation theorem, which states that given two disjoint convex sets in $\mathbb{R}^n$ we may find a co-dimension one plane such that the sets are on opposite sides of the plane. The theorem (as with many of our results on Banach spaces) can be generalised to other topological vector spaces. Those interested in more general statements may wish to consult Rudin's "Functional Analysis".

Theorem 2.28. *Suppose A and B are disjoint, nonempty, convex sets in a real or complex Banach space X.*

a) If A is open, there exist $\Lambda \in X'$ and $\gamma \in \mathbb{R}$ such that

$$\Re(\Lambda x) < \gamma \leqslant \Re(\Lambda y) \tag{2.6}$$

for all $x \in A$, $y \in B$. If B is further assumed to be open the second inequality may be taken to be strict.

b) If A is compact, B is closed then there exist $\Lambda \in X'$ and $\gamma_1, \gamma_2 \in \mathbb{R}$ such that

$$\Re(\Lambda x) < \gamma_1 < \gamma_2 < \Re(\Lambda y).$$

Proof. We first observe that it suffices to establish the result for real scalars. If we have done so then for X a complex Banach space we may find a real-linear $\ell : X \to \mathbb{R}$ which separates A and B as required, and we may then set $\Lambda(x) = \ell(x) - i\ell(ix)$.

a) Pick $a_0 \in A, b_0 \in B$ and let $x_0 = b_0 - a_0$. Let $C = A - B + x_0$. This is a convex neighbourhood of 0 in X, and since A, B are disjoint $x_0 \notin C$. Let $p(x) = \inf\{t > 0 : t^{-1}x \in C\}$. By Exercise 2.10 this is a sublinear function satisfying $p(x) \leqslant k\|x\|$ for some $k > 0$ and $p(y) < 1$ for $y \in C$. Since $x_0 \notin C$ we have $p(x_0) \geqslant 1$.

Let M be the subspace generated by x_0 and on this space define the linear functional $f(tx_0) = t$. If $t \geqslant 0$ then $f(tx_0) = t \leqslant tp(x_0) = p(tx_0)$, while if $t < 0$ then $f(tx_0) = t < 0 \leqslant p(tx_0)$, so $f \leqslant p$ on M. By the Hahn–Banach theorem we can extend f to a linear functional Λ satisfying $-k\|x\| \leqslant -p(-x) \leqslant \Lambda(x) \leqslant p(x) \leqslant k\|x\|$, so $\Lambda \in X'$.

Now suppose $a \in A, b \in B$. Then

$$\Lambda a - \Lambda b + 1 = \Lambda(a - b + x_0) \leqslant p(a - b + x_0) < 1$$

since $a - b + x_0 \in C$. Thus $\Lambda a < \Lambda b$. This implies that $\Lambda(A)$ and $\Lambda(B)$ are disjoint convex subsets of $\mathbb{R}$. $\Lambda(A)$ is open since every non-constant linear functional is an open mapping by Exercise 2.2. We may take γ to be the right end-point of $\Lambda(A)$ and (2.6) follows. If B is open, then so is $\Lambda(B)$ and we can replace the $\leqslant$ in (2.6) with $<$.

b) Since now A is compact and B is closed, we have $\inf\{\|a - b\| : a \in A, b \in B\} = d > 0$. Let $V = B_{\frac{d}{2}}(0)$ and consider $A + V$. This is open, convex and disjoint from B. Applying part *a*) with $A + V$ in place of A shows there exists $\Lambda \in X'$ such that $\Lambda(A + V)$ and $\Lambda(B)$ are disjoint convex subsets of $\mathbb{R}$ with $\Lambda(A + V)$ to the left of $\Lambda(B)$. Since $\Lambda(A + V)$ is open and $\Lambda(A)$ is a compact subset of $\Lambda(A + V)$, the result follows. □

This result immediately gives the proof of Lemma 2.12, which states that X' separates points in X: we set $A = \{x\}$, $B = \{y\}$ and apply part *b*). Another consequence is

Corollary 2.29. *Suppose M is a subspace of a Banach space X and $x_0 \in X$. If x_0 is not in the closure of M then there exists $\Lambda \in X'$ such that $\Lambda x_0 = 1$ and $\Lambda x = 0$ for every $x \in M$.*

Proof. Applying part *b*) of the previous Theorem with $A = \{x_0\}$ and $B = \overline{M}$, there exists $\Lambda \in X'$ such that Λx_0 and $\Lambda(M)$ are disjoint. But $\Lambda(M)$ must be a proper subspace of the scalar field, so must be $\{0\}$. The desired functional can be obtained by dividing Λ by Λx_0. □

Exercise 2.10. Let X be a Banach space and suppose $A \subset X$ is a convex neighbourhood of 0. For $x \in X$ define $\mu_A(x) = \inf\{t > 0 : t^{-1}x \in A\}$. Show that μ_A is sublinear and satisfies $\mu_A(x) \leqslant k\|x\|$ for some $k > 0$. Show further that $\mu_A(y) < 1$ for $y \in A$.
μ_A is called the Minkowski functional of A

Exercise 2.11. Let $\{x_1, \dots x_n\}$ be a set of linearly independent elements of a Banach space X. Let $a_1, \dots, a_n \in \mathbb{C}$. Show that there exists $\Lambda \in X'$ such that $\Lambda(x_i) = a_i$, for $i = 1, \dots, n$.

Exercise 2.12. Let M be a vector subspace of the Banach space X, and suppose that $K \subset X$ is open, convex and disjoint from M. Show that there exists a co-dimension one subspace $N \subset X$ which contains M and is disjoint from K.
This is Mazur's theorem.

Exercise 2.13. let X be a reflexive Banach space, and suppose $Y \subset X$ is a closed subspace. Show that Y is reflexive.