MATHEMATICS 3103 (Functional Analysis)
YEAR 2012–2013, TERM 2

HANDOUT #6: THE HAHN–BANACH THEOREM AND DUALITY OF BANACH SPACES

The Hahn–Banach theorem

Let X be a normed linear space. Three weeks ago we posed the question of whether there are "enough" continuous linear functionals on X to separate the points of X. This week we will prove that the answer is yes (this result is a kind of analogue, for continuous *linear* functionals on a *normed linear* space X, of Urysohn's lemma for *general* continuous functions on an arbitrary *metric* space X). We will actually prove more: namely, we will prove an extension theorem for continuous linear functionals defined on a proper linear subspace of X (this result is a kind of analogue of the Tietze extension theorem for general continuous functions defined on a proper *closed subset* of an arbitrary *metric* space X):

Theorem 6.1 (Hahn–Banach theorem for normed linear spaces)[1] *Let X be a real or complex normed linear space, let $M \subseteq X$ be a linear subspace, and let $\ell \in M^*$ be a bounded linear functional on M. Then there exists a linear functional $\widetilde{\ell} \in X^*$ that extends ℓ (i.e. $\widetilde{\ell} \restriction M = \ell$) and satisfies $\|\widetilde{\ell}\|_{X^*} = \|\ell\|_{M^*}$.*

As in the Tietze extension theorem, the important fact here is not just the existence of a continuous extension, but the existence of a continuous extension *that does not increase the norm*.

Note also that here (*unlike* in the Tietze extension theorem) the linear subspace M need not be closed. That is because a bounded *linear* functional (unlike a general continuous function) can always be automatically extended continuously from M to $\overline{M}$ (see Proposition 3.20); so it makes no difference whether M is closed or not.

For simplicity we will prove the Hahn–Banach theorem only in the *real* case. The complex case is not really much more difficult, but it involves fiddly work that would divert us from more important issues.

The proof of the Hahn–Banach theorem has two parts: First, we show that ℓ can be extended (without increasing its norm) from M to a subspace *one dimension larger*: that is, to any subspace $M_1 = \text{span}\{M, x_1\} = M + \mathbb{R}x_1$ spanned by M and a vector $x_1 \in X \setminus M$. Secondly, we show that these one-dimensional extensions can be combined to provide an extension from M to all of X.

Here is the first step:

[1]The Hahn–Banach theorem was first proven in 1912 by the Austrian mathematician Eduard Helly (1884–1943). It was rediscovered independently in the 1920s by the Austrian mathematician Hans Hahn (1879–1934) and the Polish mathematician Stefan Banach (1892–1945).

Lemma 6.2 (one-dimensional extension, real case) *Let X be a real normed linear space, let $M \subseteq X$ be a linear subspace, and let $\ell \in M^*$ be a bounded linear functional on M. Then, for any vector $x_1 \in X \setminus M$, there exists a linear functional ℓ_1 on $M_1 = \mathrm{span}\{M, x_1\}$ that extends ℓ (i.e. $\ell_1 \restriction M = \ell$) and satisfies $\|\ell_1\|_{M_1^*} = \|\ell\|_{M^*}$.*

Proof. If $\ell = 0$ the result is trivial, so we can assume without loss of generality that $\|\ell\| = 1$ (why?) (this assumption is made only to simplify the formulae). Now every $x \in M_1$ can be uniquely represented in the form $x = \lambda x_1 + y$ with $\lambda \in \mathbb{R}$ and $y \in M$. To define ℓ_1 as an extension of ℓ, it suffices to choose the value of $\ell_1(x_1)$, call it c_1: we then have

$$\ell_1(\lambda x_1 + y) \;=\; \lambda c_1 \,+\, \ell(y)\,. \tag{6.1}$$

We want to choose c_1 so that $|\ell_1(x)| \le \|x\|$ for all $x \in M_1$, i.e.

$$-\|\lambda x_1 + y\| \;\le\; \lambda c_1 \,+\, \ell(y) \;\le\; \|\lambda x_1 + y\| \tag{6.2}$$

for all $\lambda \in \mathbb{R}$ and $y \in M$. This holds for $\lambda = 0$ by hypothesis on ℓ, and for $\lambda \neq 0$ it can be rewritten as

$$-\left\|x_1 + \frac{y}{\lambda}\right\| \,-\, \ell(y/\lambda) \;\le\; c_1 \;\le\; \left\|x_1 + \frac{y}{\lambda}\right\| \,-\, \ell(y/\lambda) \tag{6.3}$$

for all $\lambda \in \mathbb{R}$ and $y \in M$ (you should check that this is correct both for $\lambda > 0$ and for $\lambda < 0$), or equivalently

$$-\|x_1 + z\| \,-\, \ell(z) \;\le\; c_1 \;\le\; \|x_1 + z\| \,-\, \ell(z) \tag{6.4}$$

for all $z \in M$. But for $z_1, z_2 \in M$ we have

$$\ell(z_2) - \ell(z_1) \;=\; \ell(z_2 - z_1) \;\le\; \|z_2 - z_1\| \;\le\; \|x_1 + z_1\| \,+\, \|x_1 + z_2\| \tag{6.5}$$

by $\|\ell\| = 1$ and the triangle inequality, so that

$$-\|x_1 + z_1\| \,-\, \ell(z_1) \;\le\; \|x_1 + z_2\| \,-\, \ell(z_2) \tag{6.6}$$

for all $z_1, z_2 \in M$. It follows that

$$c_- \;\equiv\; \sup_{z_1 \in M} \Big[-\|x_1 + z_1\| \,-\, \ell(z_1) \Big] \tag{6.7a}$$

$$c_+ \;\equiv\; \inf_{z_2 \in M} \Big[\|x_1 + z_2\| \,-\, \ell(z_2) \Big] \tag{6.7b}$$

are finite and satisfy $c_- \le c_+$; so we can choose any $c_1 \in [c_-, c_+]$. $\square$

OK, what next? If X is finite-dimensional — or more generally if M has finite codimension in X, i.e. the quotient space X/M is finite-dimensional — then we simply need to repeat the one-dimensional extension step a finite number of times, and we are done.

If X is separable — or more generally if the quotient space $X/\overline{M}$ is separable — then a slight refinement of this argument works: We first extend ℓ from M to $\overline{M}$ using Proposition 3.20. Then we choose a total sequence of linearly independent vectors $[x_1], [x_2], [x_3], \ldots$ in $X/\overline{M}$ (see Problem 8(b) of Problem Set #3), and we then successively extend ℓ to a linear functional ℓ_n defined on the space $M_n = \text{span}\{\overline{M}, x_1, \ldots, x_n\}$ for each n. Since these are *successive* extensions, we have $\ell_n \restriction M_{n'} = \ell_{n'}$ whenever $n' < n$. It follows that the union of the ℓ_n defines a linear functional ℓ_∞ on the linear subspace $M_\infty = \bigcup\limits_{n=1}^{\infty} M_n$. But by construction M_∞ is dense in X, so by Proposition 3.20, ℓ_∞ can be extended (uniquely) to a bounded linear functional $\widetilde{\ell}$ on X (without changing its norm).

Alas, if X is nonseparable, no such simple inductive construction can work, and we need to appeal to more powerful set-theoretic tools to show (nonconstructively) that the one-dimensional extensions can be pieced together to reach the whole space X. The tool we need is **Zorn's lemma**.[2] (The technique of using Zorn's lemma to make nonconstructive existence proofs is sometimes called *Zornication*.) Here are the needed concepts:

Definition 6.3 *Let S be a set. Then a* **partial order** *on S is a binary relation $\preceq$ on S that satisfies*

*(a) $a \preceq a$ (***reflexivity***);*

*(b) $a \preceq b$ and $b \preceq a$ imply $a = b$ (***antisymmetry***); and*

*(c) $a \preceq b$ and $b \preceq c$ imply $a \preceq c$ (***transitivity***)*

for all $a, b, c \in S$. The pair $(S, \preceq)$ is called a **partially ordered set** *(or* **poset***). We sometimes also refer to S alone as a partially ordered set if the relation $\preceq$ is understood from the context.*

Now let $(S, \preceq)$ be a partially ordered set. A subset $T \subseteq S$ is called **totally ordered** (with respect to $\preceq$) if for every pair $a, b \in T$ we have either $a \preceq b$ or $b \preceq a$. A totally ordered subset is also called a **chain**. An element $u \in S$ is said to be an **upper bound** for a subset $T \subseteq S$ if $a \preceq u$ for all $a \in T$. (Note that the upper bound u need not belong to T itself.) Finally, a **maximal element** of S is an element $m \in S$ such that $m \preceq x$ implies $m = x$. (A maximal element need not exist; and if one exists, it need not be unique.)

Examples. 1. The usual order $\le$ on $\mathbb{R}$ is a total order. There is no maximal element.

2. The usual order $\le$ on $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ is also a total order. Now there is a unique maximal element $+\infty$.

3. The usual partial order $\le$ on $\mathbb{R}^n$ is defined by $x \le y$ if and only if $x_i \le y_i$ for $1 \le i \le n$. For $n \ge 2$ it is *not* a total order. There is no maximal element.

4. Consider the usual partial order $\le$ on $\mathbb{R}^2$ restricted to the three-element subset $S = \{(0,0),\ (0,1),\ (1,0)\}$. Then $(0,1)$ and $(1,0)$ are maximal elements.

5. The **lexicographic order** on $\mathbb{R}^2$ is defined by $x \preceq y$ if and only if either $x_1 < y_1$ or else $x_1 = y_1$ and $x_2 \le y_2$. (Think of the ordering of words in a dictionary!) This is a total order (why?). There is no maximal element.

6. Let A be an arbitrary set, and let $\mathcal{P}(A)$ be the set of all subsets of A. Then the relation $\subseteq$ of set inclusion is a partial order on $\mathcal{P}(A)$. (It is not a total order except in two degenerate cases — can you see what they are?) There is a unique maximal element A.

[2] Zorn's lemma was first proved by the Polish mathematician Kazimierz Kuratowski (1896–1980) in 1922. It was rediscovered and applied by the German/American mathematician Max Zorn (1906–1993) in 1935.

7. Let V be a vector space, and let $\mathcal{L}(V)$ be the set of all linear subspaces of V. Then the relation $\subseteq$ of set inclusion is a partial order on $\mathcal{L}(V)$. (It is not a total order except in two degenerate cases — can you see what they are?) There is a unique maximal element V. $\square$

We then have:

Proposition 6.4 (Zorn's lemma) *Let $(S, \preceq)$ be a partially ordered set in which every totally ordered subset has an upper bound. Then $(S, \preceq)$ contains at least one maximal element.*

Zorn's lemma is a result of set theory that can be proven using the axiom of choice. More precisely, Zorn's lemma is *equivalent* to the axiom of choice in Zermelo–Fraenkel (ZF) set theory. Other important statements of set theory that are equivalent to the axiom of choice in ZF set theory are the well-ordering theorem and the Hausdorff maximal principle. We shall not enter into the details of these statements or the proof of their equivalence, which belong to a course in set theory or mathematical logic; rather, we shall simply take Zorn's lemma as a set-theoretic result that we can use without worry.[3]

We are now ready to prove the Hahn–Banach theorem:

Proof of the Hahn–Banach theorem (real case). Let $\mathcal{E}$ denote the set of all extensions of ℓ to linear subspaces of X (not necessarily to all of X) that satisfy the properties claimed in the Hahn–Banach theorem. More precisely, $\mathcal{E}$ consists of all pairs (N, f) such that

(a) N is a linear subspace of X that contains M;

(b) f is a bounded linear functional on N;

(c) $f \restriction M = \ell$; and

(d) $\|f\|_{N^*} = \|\ell\|_{M^*}$.

Now equip $\mathcal{E}$ with a partial order $\preceq$ by declaring that

$$(N, f) \preceq (N', f') \iff N \subseteq N' \text{ and } f' \restriction N = f \; . \tag{6.8}$$

In other words, $(N, f) \preceq (N', f')$ iff f' is an extension of f. (It is easy to check that $\preceq$ is indeed a partial order; you should do this.)

[3]A nice introduction to all these issues can be found in `http://en.wikipedia.org/wiki/Axiom_of_Choice` I can't resist the following quote from American mathematician Jerry Bona (1945–):

> "The Axiom of Choice is obviously true, the well-ordering principle obviously false, and who can tell about Zorn's lemma?"

As the Wikipedia article comments,

> This is a joke: although the three are all mathematically equivalent, many mathematicians find the axiom of choice to be intuitive, the well-ordering principle to be counterintuitive, and Zorn's lemma to be too complex for any intuition.

Now suppose that $\mathcal{F}$ is a totally ordered subset of $\mathcal{E}$. I claim that $\mathcal{F}$ has an upper bound in $\mathcal{E}$ (in fact a least upper bound, though we do not need this fact), defined as follows: First let

$$Y \;=\; \bigcup_{(N,f)\in\mathcal{F}} N \,. \tag{6.9}$$

You should verify, using the fact that $\mathcal{F}$ is totally ordered, that Y is a *linear subspace* of X; it is, in fact, the smallest linear subspace containing all the subspaces N where $(N, f) \in \mathcal{F}$. Next define on Y a linear functional g as the union of all the linear functionals f with $(N, f) \in \mathcal{F}$, i.e.

$$g(y) \;=\; f(y) \text{ whenever } (N, f) \in \mathcal{F} \text{ with } y \in N \,. \tag{6.10}$$

You should verify, using again the total ordering of $\mathcal{F}$, that g is well-defined in the sense that $f(y) = f'(y)$ whenever $(N, f) \in \mathcal{F}$ and $(N', f') \in \mathcal{F}$ with $y \in N$ and $y \in N'$; and you should verify, using once again the total ordering of $\mathcal{F}$, that g is indeed linear. Finally, you should check that $\|g\|_{Y^*} = \|\ell\|_{M^*}$. It follows that $(Y, g) \in \mathcal{E}$ and that $(N, f) \preceq (Y, g)$ for all $(N, f) \in \mathcal{F}$. Hence (Y, g) is an upper bound for $\mathcal{F}$ (in fact the least upper bound, though we do not need this fact).

So all the hypotheses of Zorn's lemma are satisfied. We can therefore conclude that $\mathcal{E}$ has a maximal element (N_*, f_*). I now claim that $N_* = X$; for if this were not the case, i.e. if we had $N_* \subsetneq X$, then Lemma 6.2 would provide an extension $(N_{**}, f_{**}) \in \mathcal{E}$ with $N_* \subsetneq N_{**}$ and $(N_*, f_*) \preceq (N_{**}, f_{**})$, contradicting the maximality of (N_*, f_*). The linear functional f_* is then the desired extension $\widetilde{\ell}$. $\square$

There is actually much more to the Hahn–Banach theorem than the result quoted in Theorem 6.1. Firstly, the norm $\|\cdot\|_X$ used in Theorem 6.1 to bound $\ell(x)$ can be replaced by an arbitrary **sublinear functional**: see Problem 2 of Problem Set #6. Secondly, this more general Hahn–Banach theorem implies important results on the separation of convex sets by bounded linear functionals: see Problem 3 of Problem Set #6. The term "Hahn–Banach theorems" is sometimes used to refer to this whole circle of results — which have applications in statistical physics, mathematical economics, numerical analysis (convex optimization) and many other fields.

Some corollaries of the Hahn–Banach theorem

We now note some fairly easy corollaries of the Hahn–Banach theorem. Let us begin with a result that I already announced a few weeks ago as Proposition 3.31:

Proposition 6.5 *Let X be a normed linear space. Then for each nonzero $x_0 \in X$, there exists $\ell_0 \in X^*$ with $\|\ell_0\| = 1$ such that $\ell_0(x_0) = \|x_0\|$.*

PROOF. On the one-dimensional subspace $M = \mathrm{span}(x_0) = \mathbb{R}x_0$, define a linear functional ℓ by $\ell(\alpha x_0) = \alpha\|x_0\|$. Clearly we have $\|\ell\|_{M^*} = 1$ (why?). Now extend ℓ to all of X by the Hahn–Banach theorem. $\square$

It follows in particular that for every nonzero $x_0 \in X$ there exists $\ell_0 \in X^*$ such that $\ell_0(x_0) \neq 0$. And it follows from this that X^* separates points of X.[4]

As already discussed a few weeks ago (Proposition 3.32), this result has the important consequence that the natural embedding of X into X^{**} is an *isometry* (and hence in particular injective).

Here is a useful generalization of Proposition 6.5:

Proposition 6.6 *Let X be a normed linear space, let $M \subset X$ be a proper closed linear subspace, and let $x_0 \in X \setminus M$ [so that in particular $d(x_0, M) > 0$]. Then there exists $\ell \in X^*$ with $\|\ell\| = 1$ such that $\ell \restriction M = 0$ (i.e. $M \subseteq \ker \ell$) and $\ell(x_0) = d(x_0, M)$.*

PROOF. Consider the linear subspace $M_1 = \mathrm{span}\{M, x_0\} = M + \mathbb{R}x_0$. Every vector $x \in M_1$ can be uniquely represented in the form $x = \lambda x_0 + y$ with $\lambda \in \mathbb{R}$ and $y \in M$. Define a linear functional ℓ_1 on M_1 by

$$\ell_1(\lambda x_0 + y) \;=\; \lambda \,. \tag{6.11}$$

Then clearly $\ker \ell_1 = M$ and $\ell_1(x_0) = 1$. Moreover, $\ell_1^{-1}[1] = x_0 + M$ (why?). It follows from Proposition 3.30 that

$$\|\ell_1\|_{M_1^*} \;=\; \frac{1}{d(0, x_0 + M)} \;=\; \frac{1}{d(x_0, M)} \,. \tag{6.12}$$

We can now invoke the Hahn–Banach theorem to extend ℓ_1 to a linear functional ℓ_* on all of X, with norm $\|\ell_*\|_{X^*} = 1/d(x_0, M)$. Then $\ell = d(x_0, M)\, \ell_*$ is the required linear functional. $\square$

Do you see why Proposition 6.6 includes Proposition 6.5 as a special case? (What should M be taken to be?)

Corollary 6.7 *Let X be a normed linear space. Every proper closed linear subspace $M \subset X$ is the intersection of the closed hyperplanes containing it.*

Why does this follow immediately from Proposition 6.6?

More on the duality of Banach spaces

We can use the Hahn–Banach theorem to deduce some interesting results concerning the relations between a normed linear space X and its duals X^*, X^{**}, X^{***}, etc.

Here is one result concerning the separability of X and its dual space X^*. We know that ℓ^1 is separable but $\ell^\infty \simeq (\ell^1)^*$ is not, so the separability of a Banach space X does *not* imply the separability of its dual. However, the converse *is* true:

Theorem 6.8 *Let X be a normed linear space. If X^* is separable, then so is X.*

[4] Consider any pair $x, y \in X$ with $x \neq y$, apply Proposition 6.5 to $x_0 = x - y$, and use the linearity of ℓ_0.

PROOF. The unit sphere of X^* is separable (why?), so let $\{\ell_n\}$ be a countable dense set in the unit sphere of X^*. For each n there exists $x_n \in X$ with $\|x_n\| = 1$ such that $|\ell_n(x_n)| \ge \frac{1}{2}$ (why?). Now let M be the closed linear span of $\{x_n\}$. We claim that $M = X$; if we can prove this, it follows that $\{x_n\}$ is total in X and hence that X is separable. To prove that $M = X$, suppose otherwise; then M is a proper closed linear subspace of X, so by Proposition 6.6 there exists $\ell \in X^*$ with $\|\ell\| = 1$ and $\ell \restriction M = 0$. (Note that we are really using much less than Proposition 6.6 asserts.) We then have $\ell(x_n) = 0$ for all n and hence

$$\frac{1}{2} \;\le\; |\ell_n(x_n)| \;=\; |\ell_n(x_n) - \ell(x_n)| \;\le\; \|\ell_n - \ell\| \, \|x_n\| \;=\; \|\ell_n - \ell\| \tag{6.13}$$

for all n. But this contradicts the hypothesis that $\{\ell_n\}$ is dense in the unit sphere of X^*. $\square$

Using the Hahn–Banach theorem we can also determine the duals of subspaces and quotient spaces. We need the following concept:

Definition 6.9 *Let X be a normed linear space, and let $M \subseteq X$ be a subset. Then the* **annihilator** *of M is the subset of X^* defined by*

$$M^\perp \;=\; \{\ell \in X^* \colon \ell(x) = 0 \textit{ for all } x \in M\} \,. \tag{6.14}$$

Note that $M^\perp$ is always a closed linear subspace of X^* (why?). Note also that

$$M^\perp \;=\; (\overline{M})^\perp \;=\; (\operatorname{span} M)^\perp \;=\; (\overline{\operatorname{span} M})^\perp \tag{6.15}$$

(why?).

Remark. The notation $M^\perp$ for the annihilator apparently conflicts with the notation $M^\perp$ that we used previously to denote the *orthogonal complement* of a set in a *Hilbert* space $\mathcal{H}$, since in our previous notation $M^\perp$ is a subset of $\mathcal{H}$, not $\mathcal{H}^*$. But the conflict is harmless, because we also showed that $\mathcal{H}^*$ can be canonically identified with $\mathcal{H}$, and under this identification the two meanings of $M^\perp$ *do* coincide. $\square$

The following important result shows that subspaces and quotient spaces are in a very specific sense "dual" to each other:

Theorem 6.10 *Let X be a normed linear space, and let $M \subseteq X$ be a linear subspace. Then:*

(a) M^* is isometrically isomorphic to $X^*/M^\perp$.

(b) If M is closed, then $(X/M)^*$ is isometrically isomorphic to $M^\perp$.

Furthermore, the isometric isomorphisms here are "natural", as will be seen in the course of the proof.

PROOF. (a) Given $\ell \in M^*$, the Hahn–Banach theorem tells us that there exists an extension $\widetilde{\ell} \in X^*$ satisfying $\|\widetilde{\ell}\|_{X^*} = \|\ell\|_{M^*}$. Now consider the mapping $\iota \colon M^* \to X^*/M^\perp$ defined by

$$\iota(\ell) \;=\; \widetilde{\ell} + M^\perp \,. \tag{6.16}$$

This mapping is well-defined, because if $\widetilde{\ell}' \in X^*$ is any other extension of ℓ, then we have $\widetilde{\ell} - \widetilde{\ell}' \in M^\perp$ (why?). Clearly ι is linear. Furthermore, ι is a bijection of M^* onto $X^*/M^\perp$, because the inverse of ι is given by restriction to M, namely

$$\iota^{-1}(h + M^\perp) \;=\; h \restriction M \tag{6.17}$$

(why is this independent of the choice of h in the coset? why is this ι^{-1}?).

Since each element in $\widetilde{\ell} + M^\perp$ is an extension of ℓ (why?), we have

$$\|\ell\|_{M^*} \;\le\; \inf_{g \in M^\perp} \|\widetilde{\ell} + g\|_{X^*} \;=\; \|\widetilde{\ell} + M^\perp\|_{X^*/M^\perp} \,. \tag{6.18}$$

On the other hand, since $\|\widetilde{\ell}\|_{X^*} = \|\ell\|_{M^*}$ we have

$$\|\ell\|_{M^*} \;=\; \|\widetilde{\ell}\|_{X^*} \;\ge\; \inf_{g \in M^\perp} \|\widetilde{\ell} + g\|_{X^*} \;=\; \|\widetilde{\ell} + M^\perp\|_{X^*/M^\perp} \,. \tag{6.19}$$

These two inequalities, together with the fact that ι is a bijection, prove that ι is an isometric isomorphism of M^* onto $X^*/M^\perp$.

(b) Recall that the quotient map $\pi\colon X \to X/M$ is defined by $\pi(x) = x + M$. Now define the mapping $\iota\colon (X/M)^* \to X^*$ by

$$\iota(\ell) \;=\; \ell \circ \pi \tag{6.20}$$

for $\ell \in (X/M)^*$. Clearly ι is linear. Moreover, for $x \in M$ we have

$$\iota(\ell)(x) \;=\; \ell(\pi(x)) \;=\; \ell(x+M) \;=\; \ell(M) \;=\; 0 \tag{6.21}$$

since the coset M is the zero element of X/M. It follows that ι actually maps $(X/M)^*$ into $M^\perp$. Furthermore, for any $h \in M^\perp$ we can express $\iota^{-1}(h) \in (X/M)^*$ by

$$\iota^{-1}(h)(x+M) \;=\; h(x) \tag{6.22}$$

for $x \in X$ (why is this well-defined? why is it ι^{-1}?). So ι is a bijection of $(X/M)^*$ onto $M^\perp$.

Now for every $\ell \in (X/M)^*$ and $x \in X$ we have

$$|\iota(\ell)(x)| \;=\; |\ell(x+M)| \;\le\; \|\ell\|_{(X/M)^*} \|x+M\|_{X/M} \;\le\; \|\ell\|_{(X/M)^*} \|x\|_X \tag{6.23}$$

and hence

$$\|\iota(\ell)\|_{X^*} \;\le\; \|\ell\|_{(X/M)^*} \,. \tag{6.24}$$

On the other hand, for $h \in M^\perp$ and $x \in X$ we have

$$|\iota^{-1}(h)(x+M)| \;=\; |h(x)| \;\le\; \|h\|_{X^*} \|x\|_X \tag{6.25}$$

and hence

$$|\iota^{-1}(h)(x+M)| \;\le\; \|h\|_{X^*} \|x+M\|_{X/M} \tag{6.26}$$

(why?) and hence

$$\|\iota^{-1}(h)\|_{(X/M)^*} \;\le\; \|h\|_{X^*} \,. \tag{6.27}$$

These two inequalities, together with the fact that ι is a bijection, prove that ι is an isometric isomorphism of $(X/M)^*$ onto $M^\perp$. □

Finally, let us prove:

Theorem 6.11 *Let X be a Banach space. Then X is reflexive if and only if X^* is reflexive.*

It follows from this that X and its duals and preduals are either all reflexive or all nonreflexive.

In preparation for the proof, let us recall that the natural embedding of X into X^{**} is defined by $x \mapsto \widehat{x}$ where

$$\widehat{x}(\ell) \;=\; \ell(x) \tag{6.28}$$

for $\ell \in X^*$. We denote the image of this natural embedding by $\widehat{X}$; it is a linear subspace of X^{**}. The natural embedding is an isometric isomorphism of X onto $\widehat{X}$. The space X is called reflexive in case $\widehat{X}$ is all of X^{**}.

We will similarly denote the natural embedding of X^* into X^{***} by $\ell \mapsto \widehat{\ell}$ where

$$\widehat{\ell}(y) \;=\; y(\ell) \tag{6.29}$$

for $y \in X^{**}$. We denote the image of this natural embedding by $\widehat{X^*}$; the space X^* is reflexive in case $\widehat{X^*}$ is all of X^{***}.

Proof of Theorem 6.11. Consider any $L \in X^{***}$. Since L is a bounded linear functional on X^{**}, we can restrict it to the linear subspace $\widehat{X} \subseteq X^{**}$, which is isometric to X via the natural embedding; this defines a linear functional $\ell \in X^*$ by

$$\ell(x) \;=\; L(\widehat{x}) \tag{6.30}$$

for $x \in X$. But $\ell(x) = \widehat{x}(\ell)$ by definition, so we have

$$L(\widehat{x}) \;=\; \widehat{x}(\ell) \tag{6.31}$$

for $x \in X$. Suppose now that X is reflexive, so that $\widehat{X} = X^{**}$; then every $y \in X^{**}$ is of the form $\widehat{x}$ for some $x \in X$, so we have

$$L(y) \;=\; y(\ell) \;=\; \widehat{\ell}(y) \tag{6.32}$$

for all $y \in X^{**}$. But this shows that $L = \widehat{\ell}$, so we can conclude that $\widehat{X^*}$ is all of X^{***}, i.e. that X^* is reflexive.

Conversely, suppose that X is not reflexive. Then $\widehat{X}$ is a proper linear subspace of X^{**}, which is closed since by hypothesis X is complete. It then follows from Proposition 6.6 that there exists a nonzero $L \in X^{***}$ such that $L \restriction \widehat{X} = 0$. (Note that we are using much less than Proposition 6.6 asserts.) If X^* were reflexive, then there would exist $\ell \in X^*$ such that $\widehat{\ell} = L$. But then $L \restriction \widehat{X} = 0$ says that for all $x \in X$ we have

$$0 \;=\; L(\widehat{x}) \;=\; \widehat{\ell}(\widehat{x}) \;=\; \widehat{x}(\ell) \;=\; \ell(x)\;, \tag{6.33}$$

or in other words $\ell = 0$, hence $L = 0$, which is a contradiction. □