We recall that for a normed space $X$, we introduced it’s dual space $X^*$ which consists of bounded linear functionals $f : X \to F$. In this lecture we would in interested in how rich is the space $X^*$. In particular, we address the questions:

- **point separation**: for $x \neq y$ in $X$, can we find $f \in X^*$ such that $f(x) \neq f(y)$?
- **extension**: Suppose that $Z \subset X$ is a subspace of $X$ and $f \in Z^*$. Can we construct a linear functional $\bar{f} \in X^*$ such that $\bar{f} = f$ on $Z$?

The Hahn–Banach Theorem gives an affirmative answer to these questions. It provides a powerful tool for studying properties of normed spaces using linear functionals.

The proof of the Hahn-Banach theorem is using an inductive argument. However, since we are dealing with infinite objects, we need a new tool – the Zorn Lemma – to make this induction rigorous.

### Zorn Lemma

Let us begin with some standard definitions.

**Definition 6.1** (partially ordered set). A partially ordered set is a set with a partial ordering, meaning a binary relation $\leq$ which satisfies:

1. $x \leq x$,
2. if $x \leq y$ and $y \leq x$, then $x = y$,
3. if $x \leq y$ and $y \leq z$, then $x \leq z$.

It is “partial” because there may be incomparable elements.

- A chain is a totally ordered set, meaning it has no incomparable elements, that is, for all $x$ and $y$ either $x \leq y$ or $y \leq x$.

- Suppose $S$ is a partially ordered set, and $A$ is a subset of $S$. An upper bound for $A$ is an element $s \in S$ satisfying $a \leq s$ for every $a \in A$.

- A maximal element of $A \subset S$ is an element $m \in A$ with the property that if $m \leq a$ for $a \in A$, then $a = m$.

**Example.** 1. The set $\mathbb{R}$ of real numbers constitutes a totally ordered set, without a maximal element. Every bounded subset of $\mathbb{R}$ has an upper bound.
2. the set \( \mathbb{R}^n \) a partially ordered set with the relation \( \mathbf{x} \leq \mathbf{y} \) (where \( \mathbf{x} = (x_1, \ldots, x_n) \)) if \( x_i \leq y_i \).

3. The power set \( \mathcal{P}(S) \) of a set \( S \) is the set of all subsets of \( S \), and is itself a partially ordered set with the binary relation \( \subseteq \) with the usual meaning. That is, \( A \subseteq B \) for \( A, B \in \mathcal{P}(S) \) if \( A \) is a subset of \( B \). Of course, \( \mathcal{P}(S) \) is not totally ordered if \( S \) has more than one element. The whole set \( S \) is of course a member of its own power set \( \mathcal{P}(S) \), and is its only maximal element.

4. On the other hand, suppose we take \( \mathcal{P}_f(S) \) to be the collection of all finite subsets of a set \( S \). If \( S \) is infinite, then \( \mathcal{P}_f(S) \) has no maximal element.

**Zorn Lemma.** *If \( S \) is a nonempty partially ordered set in which every chain has an upper bound, then \( S \) has a maximal element.*

We regard the Zorn Lemma as an axiom. One can show it equivalent to other axioms in set theory. In particular, it is equivalent to the axiom of choice.

**Axiom of Choice.** *If \( I \) is any nonempty (indexing) set and \( A_i \) is a nonempty set for all \( i \in I \), then there exists a function \( c : I \to \bigcup_{i \in I} A_i \) with the property that \( c(i) \in A_i \) for every \( i \in I \).*

Using the Zorn Lemma we can, in particular, prove existence of complete orthonormal sets:

**Theorem 6.2.** *Every non-trivial Hilbert space has a complete orthonormal set.*

**Proof.** Let \( S \) be the set of all orthonormal subsets of our non-trivial Hilbert space \( H \). The set \( S \) is nonempty: any nonzero element of \( H \) can be normalized to produce an orthonormal set containing one vector. \( S \) is partially ordered, with \( \subseteq \). Furthermore, every chain \( \mathcal{C} \subseteq S \) has an upper bound: take the union of all the orthonormal sets constituting \( \mathcal{C} \). Now Zorn’s Lemma tells us that there is a maximal element \( M \in S \). Now we only have to prove that \( M \) is complete. Suppose there is a nonzero \( z \in H \) with \( z \perp M \). Then \( M_1 = M \cup \{z/\|z\|\} \) is orthonormal, and contains \( M \). By maximality of \( M \), we must have \( M_1 = M \), which contradicts \( z \perp M \). \( \square \)

**Hahn-Banach theorems**

We would like to extend linear functionals from subspaces to whole spaces. Moreover, we would like to do it in a way that respects the ‘boundedness’ properties of the given functional. The Hahn–Banach Theorem articulates this ‘boundedness’ via *sublinear functionals*. We note norms give the most standard example of sublinear functionals, but for many applications it is convenient to consider more general objects.
Definition 6.3. A sublinear functional is a real-valued function $p$ on a vector space $X$ which satisfies

$$p(x + y) \leq p(x) + p(y) \quad \text{for all } x, y \in X,$$  
(subadditivity)

and

$$p(ax) = ap(x) \quad \text{for all } a \geq 0, x \in X.$$  
(positive homogeneity)

Example 6.4. For a linear map $f : X \to \mathbb{R}$, the map $x \mapsto |f(x)|$ is a sublinear functional.

Theorem 6.5 (Hahn–Banach Theorem for real vector spaces). Let $X$ be a real vector space and $p$ a sublinear functional on $X$. Let $f$ be a linear functional defined on a subspace $Z \subset X$, and satisfying $f(z) \leq p(x)$ for all $z \in Z$. Then there exists a linear functional $\bar{f}$ on $X$ satisfying

- $\bar{f}(z) = f(z)$ for all $z \in Z$; and,
- $\bar{f}(x) \leq p(x)$ for all $x \in X$.

Proof. We want to use Zorn’s Lemma, so we must find a partially ordered set where a maximal element would be relevant to us. Let $S$ be the set of all linear extensions of $f$ that are dominated by $p$. Those are linear maps $\phi : D_\phi \to \mathbb{R}$, where $D_\phi$ is a subspace containing $Z$, such that $\phi = f$ on $Z$ and $\phi \leq p$ on $D_\phi$. We consider the set

$$S = \{ \phi : \phi : D_\phi \to \mathbb{R} \text{ is linear}, D_\phi \supset Z, \phi = f \text{ on } Z, \phi \leq p \text{ on } D_\phi \}.$$  

The binary relation $\prec$ is ‘extension’: We say $\phi_1 \prec \phi_2$ if $\phi_2$ is an extension of $\phi_1$, that is, $D_{\phi_1} \subset D_{\phi_2}$ and $\phi_1 = \phi_2$ on $D_{\phi_1}$. One easily sees that it is a partial ordering.

We claim that every chain $C \subset S$ has an upper bound. Define $\tilde{\phi}$ by setting $\tilde{\phi}(x) = \phi(x)$ whenever $x \in D_\phi$ for some $\phi \in C$. This definition is consistent because $C$ is a chain. Then $\tilde{\phi}$ is a linear functional on the domain

$$D_{\tilde{\phi}} = \bigcup_{\phi \in C} D_\phi,$$  

which is a vector space.\(^1\) It is clear that $f \prec \tilde{\phi}$, hence $\tilde{\phi} \in S$, and that $\tilde{\phi}$ is an upper bound for the chain $C$. Therefore, by Zorn’s Lemma, $S$ has a maximal element, $\bar{f} \in S$.

We now have a linear extension $\bar{f}$ of $f$ satisfying $\bar{f}(x) \leq p(x)$ for all $x \in D_{\bar{f}}$ and which is not extendible by any other linear functional also

\(^1\)This is because $C$ is a chain, so that whenever $x, y \in D_\phi$, we know that there is some $\phi \in C$ for which $x, y$ are in the vector space $D_\phi$, hence $ax + \beta y$ is also. The linearity of $\tilde{\phi}$ follows from linearity of $\phi$’s.
dominated by \( p \) (by its maximality in \( S \)). It is only left to show that \( D_f = X \). Suppose that \( D_f \neq X \) and choose some \( y_1 \notin D_f \). We note that \( y_1 \) together with \( D_f \) span a subspace \( Y_1 \) containing \( D_f \) properly. Any element \( x \in Y_1 \) can be written \( x = y + t y_1 \) where \( y \in D_f \) and \( t \in \mathbb{R} \), in exactly one way.\(^2\) Now define \( F : Y_1 \to \mathbb{R} \) by \( F(x) = F(y + t y_1) = \bar{f}(y) + a t \) where \( a \in \mathbb{R} \). The parameter \( a \) will be chosen later. It is clear that \( F \) is a linear extension of \( \bar{f} \) and that \( F \neq \bar{f} \). So, if we are able to show that \( F(x) \leq p(x) \) for all \( x \in D_F \), then we will have contradicted \( \bar{f} \)'s maximality in \( S \), implying that \( D_f = X \) and proving the theorem.

To arrange that \( F \) is dominated by \( p \), we choose the parameter \( a \) in a suitable way. More precisely we have to choose \( a \) such that for all \( u \in D_f \) and all \( t \in \mathbb{R} \),

\[
F(u + ty_1) = \bar{f}(u) + ta \leq p(u + ty_1). 
\]

In particular, these inequalities imply that for all \( u_1, u_2 \in D_f \),

\[
\bar{f}(u_1) + a \leq p(u_1 + y_1) \quad \text{and} \quad \bar{f}(u_2) - a \leq p(u_2 - y_1).
\]

Hence, \( a \) must satisfy

\[
\bar{f}(u_2) - p(u_2 - y_1) \leq a \leq p(u_1 + y_1) - \bar{f}(u_1). \tag{2}
\]

for all \( u_1, u_2 \in D_f \). We observe that by subadditivity of \( p \),

\[
\bar{f}(u_1) + \bar{f}(u_2) = \bar{f}(u_1 + u_2) \leq p(u_1 + u_2) \leq p(u_1 + y_1) + p(u_2 - y_1),
\]

and re-arranging terms we find that

\[
\bar{f}(u_2) - p(u_2 - y_1) \leq p(u_1 + y_1) - \bar{f}(u_1).
\]

for all \( u_1, u_2 \in D_f \). Therefore, taking a supremum of the left-hand side over \( u_2 \) to obtain \( m \), and an infimum over the right-hand side over \( u_1 \) to obtain \( M \), we find that \( m \leq M \). Taking \( a \in [m, M] \), we deduce that (2) holds. It follows from (2) that

\[
F(u + y_1) = \bar{f}(u) + a \leq p(u + y_1) \quad \text{and} \quad F(u - y_1) = \bar{f}(u) - a \leq p(u - y_1).
\]

Multiplying these inequalities by \( t > 0 \) and using positive homogeneity of \( p \), we deduce that (1) holds. That is \( F \leq p \) on \( D_F \). As it was noted above this gives the contradiction, and implies that \( D_f = X \). Hence, \( \bar{f} \) gives the required extension.

Now we proved several other versions of the Hahn-Banach theorems which are consequences of Theorem (6.5).

\(^2\)For if there were another way, say \( y + t y_1 = \bar{g} + s y_1 \), then we would have that \( y - \bar{g} = (s - t) y_1 \). The left-hand side is in \( D_f \) and the right-hand side is a multiple of \( y_1 \notin D_f \), meaning that they both must be 0.
Definition 6.6. A **seminorm** on a vector space $X$ is simply a map $p : X \to \mathbb{R}$ that satisfies all the defining properties of a norm except

$$p(x) = 0 \iff x = 0.$$  

Theorem 6.7 (Hahn–Banach Theorem for seminorms). Let $f$ be a linear functional on a subspace $Z$ of a normed linear space $X$. Suppose $p : X \to \mathbb{R}$ is a seminorm on $X$ and that $|f(z)| \leq p(z)$ for all $z \in Z$. Then there is a linear functional $\tilde{f}$ on $X$ satisfying $\tilde{f}(z) = f(z)$ for all $z \in Z$ and $|\tilde{f}(x)| \leq p(x)$ for all $x \in X$.

**Proof.** If $X$ is a real vector space, then this follows easily from Theorem 6.5. The assumption that $|f(z)| \leq p(z)$ implies that $f(z) \leq p(z)$, so Theorem 6.5 implies that there is a functional $\bar{f}$ on $X$ that extends $f$ and satisfies $\bar{f}(x) \leq p(x)$ for all $x \in X$. In particular, we have $\bar{f}(-x) \leq p(-x)$, which implies by linearity of $\bar{f}$ and homogeneity of $p$ that $-\bar{f}(x) \leq |-1|p(x)$, for all $x$. Therefore, $|\bar{f}(x)| \leq p(x)$ for all $x$.

If $X$ is a complex vector space, then $f$ is a complex-valued functional on the subspace $Z \subset X$, hence is expressible as $f(x) = f_1(x) + if_2(x)$, where $f_1$ and $f_2$ are real-valued. The remaining steps are:

1. Show that $f_1$ and $f_2$ are linear functionals on $Z_{\mathbb{R}}$, where $Z_{\mathbb{R}}$ is just $Z$, thought of as a real vector space, and show that $f_1(z) \leq p(z)$ for all $z \in Z_{\mathbb{R}}$. Deduce from Theorem 6.5 that there is a linear extension $\bar{f}_1$ of $f_1$ from $Z_{\mathbb{R}}$ to $X_{\mathbb{R}}$.

2. Show that $f_2(z) = -f_1(iz)$ for all $z \in Z$, and that if we set

$$\tilde{f}(x) = \bar{f}_1(x) - i\bar{f}_1(ix),$$

then $\tilde{f}(z) = f(z)$ for all $z \in Z$, hence is an extension.

3. Show that $\tilde{f}$ as defined above is complex-linear.

4. Show that $|\tilde{f}(x)| \leq p(x)$ for all $x \in X$.

These steps are set as exercises.

Theorem 6.8 (Hahn–Banach Theorem for normed spaces). Let $f$ be a bounded linear functional on a non-trivial subspace $Z$ of a normed space $X$. Then there is a bounded linear functional $\tilde{f}$ on $X$ which is an extension of $f$ to $X$ and has the same norm: $\|\tilde{f}\|_X = \|f\|_Z$.

**Proof.** The idea is to use Theorem 6.7, so we need to find our seminorm $p$. But this is easy, since we have started with a bounded linear functional $f$ on a normed space $Z$, meaning that for all $z \in Z$, we have that

$$|f(z)| \leq \|f\|_Z \|z\|.$$
Let us define for \( x \in X \), the map \( p(x) = \|f\|_Z \|x\| \). It is routine to check that \( p \) is a seminorm on \( X \). We can therefore use Theorem 6.7 to assert that there is a linear extension \( \bar{f} \) of \( f \) to all of \( X \) that satisfies

\[
|\bar{f}(x)| \leq p(x) = \|f\|_Z \|x\|
\]

for all \( x \in X \). This implies that \( \|\bar{f}\|_X \leq \|f\|_Z \). On the other hand, it is clear that an extension cannot have smaller norm, so that we also have \( \|\bar{f}\|_X \geq \|f\|_Z \), implying the equality \( \|\bar{f}\|_X = \|f\|_Z \).

Using Theorem 6.8 we construct \( f \in X^* \) with prescribed values.

**Theorem 6.9.** Let \( X \) be a normed space and \( x_0 \neq 0 \) an element of \( X \). Then there exists a bounded linear functional \( f \) on \( X \) such that \( \|f\| = 1 \) and \( f(x_0) = \|x_0\| \).

**Proof.** We will use Theorem 6.8. Notice that we only need to find a subspace \( Z \subset X \) containing the element \( x_0 \), and a linear functional \( f \in Z^* \) with \( f(x_0) = \|x_0\| \) and \( \|f\|_Z = 1 \). This way, Theorem 6.8 will imply the existence of \( \bar{f} \in X^* \) with \( \bar{f}(x_0) = f(x_0) = \|x_0\| \) and \( \|\bar{f}\|_X = \|f\|_Z = 1 \), as desired. The most natural choices turn out to be suitable. Let \( Z = \{a x_0 \mid a \in \mathbb{F}\} \) be the one-dimensional subspace of \( X \) spanned by \( x_0 \), and let \( f : Z \to \mathbb{F} \) be the functional defined by \( f(a x_0) = a \|x_0\| \).

**Corollary 6.10.** Let \( X \) and \( Y \) be normed spaces and \( A : X \to Y \) a linear map. Then \( \|A^*\| = \|A\| \).

**Proof.** The inequality \( \|A^*\| \leq \|A\| \) has already been proven in a previous lecture. To prove the opposite inequality, we use the above result. Let \( x \in X \). Take \( f \in Y^* \) such that \( \|f\| = 1 \) and \( f(Ax) = \|Ax\| \). Then

\[
\|Ax\| = |f(Ax)| = |(A^* f)(x)| \leq \|A^* f\| \|x\| \leq \|A^*\| \|f\| \|x\| = \|A^*\| \|x\|.
\]

This implies that \( \|A\| \leq \|A^*\| \).

Many features of a normed space can be obtained from looking on its dual space. For example, we can compute norms of elements.

**Corollary 6.11.** For every \( x \) in a normed space \( X \) we have

\[
\|x\| = \sup \left\{ \frac{|f(x)|}{\|f\|} \mid f \in X^*, f \neq 0 \right\}.
\]

**Proof.** Theorem 6.9 implies that there is some functional \( f \in X^* \) with norm 1 and taking \( x \) to \( \|x\| \), which implies that \( \sup \frac{|f(x)|}{\|f\|} \geq \|x\| \). The other inequality follows from \( |f(x)| \leq \|f\| \|x\| \).

Often, the Hahn–Banach Theorem is phrased as “there are enough linear functionals to separate points of a normed space.” Indeed, if \( f(x) = f(y) \) for all bounded linear functionals \( f \), this implies that \( f(x - y) = 0 \) for every \( f \in X^* \). Corollary 6.11 then implies that \( x - y = 0 \).
Bounded linear functional on $C([a,b])$

To give another application of the Hahn-Banach theorem, we describe the dual space $C([a,b])^*$. For this we need to introduce the notion of the Riemann–Stieltjes integral which generalises the Riemann integral.

We define the total variation of a function $w$ on $[a,b]$ as

$$\text{Var}(w) = \sup \sum_{j=1}^{n} |w(t_j) - w(t_{j-1})|,$$

where the supremum is taken over partitions $a = t_0 < t_1 < \cdots < t_n = b$ of the interval $[a,b]$. We denote by $BV([a,b])$ the set of functions with $\text{Var}(w) < \infty$. Clearly, this is a vector space, which is called the space of functions with bounded variation. We equip this space with the norm $\|w\| = |w(a)| + \text{Var}(w)$.

Given a function $w \in BV([a,b])$, we now define the Riemann–Stieltjes integral $\int_a^b \phi(t)dw(t)$. For a partition $P = \{t_0 < t_1 < \cdots < t_n\}$ of $[a,b]$, we denote by $|P|$ the length of its largest interval and set

$$S(\phi,P) = \sum_{j=1}^{n} \phi(t_j)(w(t_j) - w(t_{j-1}))$$

Suppose that there exists a number $I$ with the property: for every $\epsilon > 0$ there exists $\delta > 0$ such that for all partitions $P_n$ satisfying $|P| < \delta$,

$$|I - S(\phi,P)| < \epsilon.$$

Then we call $I$ the Riemann–Stieltjes integral which is denoted by $\int_a^b \phi(t)dw(t)$.

It follows from the uniform continuity property that the Riemann–Stieltjes integral exists for continuous $\phi$. It is clear from the definition that it is a linear map on $C([a,b])$ and $\left| \int_a^b \phi(t)dw(t) \right| \leq \|\phi\| \text{Var}(w)$. So that it defines a linear functional on $C([a,b])$ with norm $\leq \text{Var}(w)$. Remarkably, it turns out that every element of $C([a,b])^*$ is of this form.

**Theorem 6.12 (Riesz).** Every bounded linear functional $f$ on $C([a,b])$ can be represented as

$$f(\phi) = \int_a^b \phi(t)dw(t)$$

for some $w \in BV([a,b])$. Moreover, $\|f\| = \text{Var}(w)$.

**Proof.** Let $B([a,b])$ be the space of all bounded functions on $[a,b]$ equipped with the maximum norm. By the Hahn-Banach theorem, $f$ could be extended to a bounded linear functional on $B([a,b])$ such that $\|F\| = \|f\|$.
We define the function $w$ as follows. Let $\rho_t$ be the function on $[a, b]$ such that $\rho_t = 1$ on $[a, t]$ and $\rho_t = 0$ on $(t, b]$. We set

$$w(a) = 0 \quad \text{and} \quad w(t) = F(\rho_t) \text{ for } t \in (a, b].$$

We claim that $w$ has bounded variation. For a partition $P = \{t_0 < t_1 < \cdots < t_n\}$ of $[a, b]$,

$$\sum_{j=1}^n |w(t_j) - w(t_{j-1})| = |F(\rho_{t_1})| + \sum_{j=2}^n |F(\rho_{t_j}) - F(\rho_{t_{j-1}})|$$

$$= a_1 F(\rho_{t_1}) + \sum_{j=2}^n a_j (F(\rho_{t_j}) - F(\rho_{t_{j-1}}))$$

for some constants $a_j$ such that $|a_j| = 1$. Then

$$\sum_{j=1}^n |w(t_j) - w(t_{j-1})| = F \left( a_1 \rho_{t_1} + \sum_{j=2}^n a_j (\rho_{t_j} - \rho_{t_{j-1}}) \right)$$

$$\leq \|F\left\|_{\infty} \left( a_1 \rho_{t_1} + \sum_{j=2}^n a_j (\rho_{t_j} - \rho_{t_{j-1}}) \right) \leq \|F\|$$

since $\rho_{t_j} - \rho_{t_{j-1}} = 1$ only on $(t_{j-1}, t_j]$ and is zero otherwise. This proves that $\text{Var}(\rho) \leq \|F\| = \|f\|$.

Now we show that

$$f(\phi) = \int_a^b \phi(t) dw(t) \quad \text{for } \phi \in C([a, b]). \quad (3)$$

Given a partition $P = \{t_0 < t_1 < \cdots < t_n\}$ of $[a, b]$, we define a piecewise constant approximation to $\phi$ by

$$\psi_P = \phi(a) \rho_{t_1} + \sum_{j=2}^n \phi(t_{j-1}) (\rho_{t_j} - \rho_{t_{j-1}}).$$

It follows from uniform continuity of $\phi$ that for every $\epsilon > 0$ there exists $\delta > 0$ such that if $|s - t| < \delta$, then $|\phi(s) - \phi(t)| < \epsilon$. This implies that for every partition $P$ such that $|P| < \delta$, we have $\|\phi - \psi_P\|_{\infty} < \epsilon$. Then

$$|F(\phi) - F(\psi_P)| \leq \|F\| \epsilon.$$
On the other hand,

\[
F(\psi P) = \phi(a)F(\rho_1) + \sum_{j=2}^{n} \phi(t_{j-1})(F(\rho_j) - F(\rho_{j-1})) \\
= \phi(a)w(t_1) + \sum_{j=2}^{n} \phi(t_{j-1})(w(t_j) - w(t_{j-1})) \\
= \sum_{j=1}^{n} \phi(t_{j-1})(w(t_j) - w(t_{j-1})).
\]

So by the definition of the Riemann-Stieltjes integral,

\[
F(\psi P) \rightarrow \int_{a}^{b} \phi(t)dw(t) \quad \text{as} \quad |P| \rightarrow 0.
\]

This implies (3).

Finally, for every \( \phi \in C([a, b]) \),

\[
|f(\phi)| = \left| \int_{a}^{b} \phi(t)dw(t) \right| \leq \|\phi\|_{\infty} \text{Var}(w).
\]

So \( \|f\| \leq \text{Var}(w) \), and we conclude that \( \|f\| = \text{Var}(w) \). \( \square \)