1 Zorn's Lemma

Definition 1.1. *A **partially ordered set** $(M, \leq)$ is a set M equipped with a binary relation $\leq$, called a partial ordering, satisfying*

(a) $x \leq x$ *for every* $x \in M$. *(Reflexivity)*

(b) If $x \leq y$ *and* $y \leq x$*, then* $x = y$. *(Antisymmetry)*

(c) If $x \leq y$ *and* $y \leq z$*, then* $x = z$. *(Transitivity)*

- The word "partially" is chosen because M may contain elements x, y such that neither $x \leq y$ nor $y \leq x$ holds. Such pair of elements is said to be **incomparable**. One example is $M = \Big\{\{1\}, \{2\}, \{1, 2\}\Big\}$, with set inclusion $\subseteq$ as a partial ordering. It is clear that $\{1\}$ and $\{2\}$ are not comparable.
- Analogously, $x, y \in M$ are **comparable** if $x \leq y$ or $y \leq x$ or both.

Definition 1.2.

*(a) A **chain** C is a partially ordered set where every pair of elements is comparable.*

*(b) An **upper bound** (if it exists) of a subset W of a partially ordered set M is an element $u \in M$ such that*

$$x \leq u \text{ for all } x \in W. \tag{1}$$

*(c) A **maximal element** (if it exists) of a partially ordered set M is an element $m \in M$ such that*

$$\text{if } m \leq x \text{ for some } x \in M, \text{ then } x = m. \tag{2}$$

In other words, there is no $x \in M$ such that $m \leq x$ but $x \neq m$.

If a maximal element exists for a totally ordered set/chain, then it must be unique; this follows immediately since every pair of elements is comparable in a chain. Now, it seems intuitive that a maximal element is also an upper bound. Unfortunately, this is not always the case, as illustrated in the following example.

Example 1.3. *Consider the set $W = \Big\{\varnothing, \{1\}, \{2\}, \{3\}, \{1, 2\}\Big\}$ with set inclusion $\subseteq$ as a partial ordering.*

- The maximal elements are $\{1, 2\}$ and $\{3\}$.
- On the other hand, if we view W as a subset of the power set of $\{1, 2, 3\}$, then the upper bound of W is the element $\{1, 2, 3\}$.
- Here is what went wrong: (1) requires an upper bound u to be comparable with all elements in W, while (2) does not. Another simple explanation is that an upper bound might not be element of W.

We are now ready to formulate Zorn's lemma and provide two applications of it.

Theorem 1.4 (Zorn's lemma)**.** *Let $(M, \leq)$ be a non-empty partially ordered set. If every chain $C \subset M$ has an upper bound, then M has at least one maximal element.*

- Note the upper bound need not be an element of C, but it must be an element of M.

Theorem 1.5 (Hamel basis)**.** *Every non-empty vector space X has a Hamel basis.*

Proof. Let M be the set of all linearly independent subsets of X ordered by set inclusion $\subseteq$. M is non-empty since

$$X \neq \{\mathbf{0}\} \implies \text{there exists a nonzero } x \in X \implies x \in M.$$

Every chain $C \subset M$ has an upper bound, given by the union of all elements of C; by Zorn's lemma, M contains a maximal element B. We claim that B is a Hamel basis for X. Suppose not, then there exists an $z \in X \setminus B$. The set $Y = B \cup \{z\}$ form a linearly independent subset of X, containing B as a proper subset. This contradicts the maximality of B.

■

Theorem 1.6 (Total orthonormal set)**.** *Every non-empty Hilbert space H has a total orthonormal set.*

Proof. Let M be the set of all orthonormal subsets of H ordered by set inclusion $\subseteq$. M is non-empty since

$$H \neq \{\mathbf{0}\} \implies \text{there exists a nonzero } x \in X \implies \frac{x}{\|x\|} \in M.$$

Every chain $C \subset M$ has an upper bound, given by the union of all elements of C; by Zorn's lemma, M contains a maximal element F. We claim that F is total in H. Suppose not, then there exists a nonzero $z \in H$ such that $z \perp F$. The set $Y = F \cup \{z/\|z\|\}$ form an orthonormal subset of H, containing F as a proper subset. This contradicts the maximality of F. Note that completeness is required to appeal to the sufficient condition for totality.

■

Remark 1.7. *Observe that the proofs of Theorem 1.5, 1.6 shares the same steps. We first construct a partially ordered set $(M, \leq)$ that possess certain properties depending on the context, such that Zorn's lemma is applicable. The maximal element, say, B, will be a candidate of something we want to show; this is shown by constructing an element in M that contradicts the maximality of B. The last step is usually the most difficult part. We will see in the next section that the proof of the Hahn-Banach theorem follows exactly this same approach.*

2 Hahn-Banach Theorem.

The Hahn-Banach theorem is one of the most fundamental result in linear functional analysis. A simple but powerful consequence of the theorem is there are sufficiently many bounded linear functionals in a given normed space X. Essentially, one construct functionals with certain properties on a lower-dimensional subspace of X, then extend it to the entire space X. By now, you should have seen several applications of the Hahn-Banach theorem.

Definition 2.1. *Given a vector space X, a **sublinear functional** is a real-valued functional $p\colon X \longrightarrow R$ satisfying*

(a) $p(x+y) \le p(x) + p(y)$ for all $x, y \in X$. *(Subadditive).*

(b) $p(\alpha x) = \alpha p(x)$ for all $\alpha \in \mathbb{R}_{\ge 0}$ and $x \in X$. *(Positive-homogeneous).*

Theorem 2.2 (Hahn-Banach Theorem: Extension of linear functionals)**.**
Let X be a real vector space and p a sublinear functional on X. Let f be a linear functional defined on a subspace $Y \subset X$, satisfying

$$f(y) \le p(y) \qquad \text{for all } y \in Y.$$

Then f has a linear extension $\tilde{f}$ from Y to X satisfying

(a) $\tilde{f}$ is a linear functional on X,

(b) $\tilde{f}|_Y = f$, i.e. the restriction of $\tilde{f}$ to Y agrees with f,

(c) $\tilde{f}(x) \le p(x)$ for all $x \in X$.

Proof. The proof is different from the textbook, in the sense that in step (A) we define the partially ordered set M as an ordered pair consists of a subspace of X and a linear extension, whereas in step (C) we show how to choose δ by a "backward argument", which is more intuitive instead of starting on some random equations and claim the choice of δ will get the job done. Following Remark 1.7, we split the proof into three parts:

(A) Let M be the partially ordered set of pairs (Z, f_Z), where

i. Z is a subspace of X containing Y,

ii. $f_Z\colon Z \longrightarrow \mathbb{R}$ is a linear functional extending f, satisfying

$$f_Z(z) \le p(z) \qquad \text{for all } z \in Z.$$

with the partial ordering $(Z_1, f_{Z_1}) \le (Z_2, f_{Z_2})$ if $Z_1 \subset Z_2$ and $(f_{Z_2})|_{Z_1} = f_{Z_1}$. It is clear that M is non-empty since $(Y, f) \in M$. Choose an arbitrary chain $C = \Big\{ (Z_\alpha, f_{Z_\alpha}) \Big\}_{\alpha \in \Lambda}$ in M, Λ some indexing set. We will show that C has an upper bound in M. Let $W = \bigcup_{\alpha \in \Lambda} Z_\alpha$ and construct a functional $f_W\colon W \longrightarrow \mathbb{R}$ defined as follows: If $w \in W$, then $w \in Z_\alpha$ for some $\alpha \in \Lambda$ and we set $f_W(w) = f_{Z_\alpha}(w)$ for that particular α.

- This definition is well-defined. Indeed, suppose $w \in Z_\alpha$ and $w \in Z_\beta$. If $Z_\alpha \subset Z_\beta$, say, then $f_{Z_\beta}|_{Z_\alpha} = f_{Z_\alpha}$, since we are in a chain.
- W clearly contains Y, and we show that W is a subspace of X and f_W is a linear functional on W. Choose any $w_1, w_2 \in W$, then $w_1 \in Z_{\alpha_1}, w_2 \in Z_{\alpha_2}$ for some $\alpha_1, \alpha_2 \in \Lambda$. If $Z_{\alpha_1} \subset Z_{\alpha_2}$, say, then for any scalars $\beta, \gamma \in \mathbb{R}$ we have

$$w_1, w_2 \in Z_{\alpha_2} \implies \beta w_1 + \gamma w_2 \in Z_{\alpha_2} \subset W.$$

Also, with $f_W(u) = f_{Z_{\alpha_1}}(u)$ and $f_W(v) = f_{Z_{\alpha_2}}(v)$,

$$\begin{aligned} f_W(\beta u + \gamma v) &= f_{Z_{\alpha_2}}(\beta u + \gamma v) \\ &= \beta f_{Z_{\alpha_2}}(u) + \gamma f_{Z_{\alpha_2}}(v) \quad \left[\text{linearity of } f_{Z_{\alpha_2}}\right] \\ &= \beta f_{Z_{\alpha_1}}(u) + \gamma f_{Z_{\alpha_2}}(v) \quad \left[\text{since we are in a chain}\right] \\ &= \beta f_W(u) + \gamma f_W(v). \end{aligned}$$

The case $Z_{\alpha_2} \subset Z_{\alpha_1}$ follows from a symmetric argument.
- Choose any $w \in W$, then $w \in Z_\alpha$ for some $\alpha \in \Lambda$ and

$$f_W(w) = f_{Z_\alpha}(w) \le p(w) \qquad \text{since } (w, Z_\alpha) \in M.$$

Hence, (W, f_W) is an element of M and an upper bound of C since $(Z_\alpha, f_{Z_\alpha}) \le (W, f_W)$ for all $\alpha \in \Lambda$. Since C was an arbitrary chain in M, by Zorn's lemma, M has a maximal element $(Z, f_Z) \in M$, and f_Z is (by definition) a linear extension of f satisfying $f_Z(z) \le p(z)$ for all $z \in Z$.

(B) The proof is complete if we can show that $Z = X$. Suppose not, then there exists an $\theta \in X \setminus Z$; note $\theta \neq \mathbf{0}$ since Z is a subspace of X. Consider the subspace $Z_\theta = \text{span}\{Z, \{\theta\}\}$. Any $x \in Z_\theta$ has a unique representation $x = z + \alpha\theta, z \in Z, \alpha \in \mathbb{R}$. Indeed, if

$$x = z_1 + \alpha_1\theta = z_2 + \alpha_2\theta, z_1, z_2 \in Z, \alpha_1, \alpha_2 \in \mathbb{R},$$

then $z_1 - z_2 = (\alpha_2 - \alpha_1)\theta \in Z$ since Z is a subspace of X. Since $\theta \notin Z$, we must have $\alpha_2 - \alpha_1 = 0$ and $z_1 - z_2 = \mathbf{0}$. Next, we construct a functional $f_{Z_\theta} \colon Z_\theta \longrightarrow \mathbb{R}$ defined by

$$f_{Z_\theta}(x) = f_{Z_\theta}(z + \alpha\theta) = f_Z(z) + \alpha\delta, \tag{3}$$

where δ is any real number. It can be shown that f_{Z_θ} is linear and f_{Z_θ} is a proper linear extension of f_Z; indeed, we have, for $\alpha = 0$, $f_{Z_\theta}(x) = f_{Z_\theta}(z) = f_Z(x)$. Consequently, if we can show that

$$f_{Z_\theta}(x) \le p(x) \qquad \text{for all } x \in Z_\theta, \tag{4}$$

then $(Z_\theta, f_{Z_\theta}) \in M$ satisfying $(Z, f_Z) \le (Z_\theta, f_{Z_\theta})$, thus contradicting the maximality of (Z, f_Z).

(C) From (3), observe that (4) is trivial if $\alpha = 0$, so suppose $\alpha \neq 0$. We do have a single degree of freedom, which is the parameter δ in (3), thus the problem reduces to showing the existence of a suitable $\delta \in \mathbb{R}$ such that (4) holds. Consider any $x = z + \alpha\theta \in Z_\theta, z \in Z, \alpha \in \mathbb{R}$. Assuming $\alpha > 0$, (4) is equivalent to

$$\begin{aligned} f_Z(z) + \alpha\delta &\leq p(z + \alpha\theta) = \alpha p(z/\alpha + \theta) \\ f_Z(z/\alpha) + \delta &\leq p(z/\alpha + \theta) \\ \delta &\leq p(z/\alpha + \theta) - f_Z(z/\alpha). \end{aligned}$$

Since the above must holds for all $z \in Z, \alpha \in \mathbb{R}$, we need to choose δ such that

$$\delta \leq \inf_{z_1 \in Z} \Big[p(z_1 + \theta) - f_Z(z_1) \Big] = m_1. \tag{5}$$

Assuming $\alpha < 0$, (4) is equivalent to

$$\begin{aligned} f_Z(z) + \alpha\delta &\leq p(z + \alpha\theta) = -\alpha p(-z/\alpha - \theta) \\ -f_Z(z/\alpha) - \delta &\leq p(-z/\alpha - \theta) \\ \delta &\geq -p(-z/\alpha - \theta) - f_Z(z/\alpha). \end{aligned}$$

Since the above must holds for all $z \in Z, \alpha \in \mathbb{R}$, we need to choose δ such that

$$\delta \geq \sup_{z_2 \in Z} \Big[-p(-z_2 - \theta) - f_Z(z_2) \Big] = m_0. \tag{6}$$

We are left with showing condition (5), (6) are compatible, i.e.

$$-p(-z_2 - \theta) - f_Z(z_2) \leq p(z_1 + \theta) - f_Z(z_1) \quad \text{for all } z_1, z_2 \in Z.$$

The inequality above is trivial if $z_1 = z_2$, so suppose not. We have that

$$\begin{aligned} p(z_1 + \theta) - f_Z(z_1) + p(-z_2 - \theta) + f_Z(z_2) &= p(z_1 + \theta) + p(-z_2 - \theta) + f_Z(z_2 - z_1) \\ &\geq f_Z(z_2 - z_1) + p(z_1 + \theta - z_2 - \theta) \\ &= f_Z(z_2 - z_1) + p(z_1 - z_2) \\ &= -f_Z(z_1 - z_2) + p(z_1 - z_2) \geq 0. \end{aligned}$$

where linearity of f_Z and subadditivity of p are used. Hence, the required condition on δ is $m_0 \leq \delta \leq m_1$.

■

3 Complex Hahn-Banach Theorem.

We begin by proving a lemma that gives us the relation between real and imaginary parts of a complex linear functional.

Lemma 3.1. *Let X be a complex vector space and $f\colon X \longrightarrow \mathbb{C}$ a linear functional. There exists a real linear functional $f_1\colon X \longrightarrow \mathbb{R}$ such that*

$$f(x) = f_1(x) - if(ix) \qquad \text{for all } x \in X.$$

Proof. Write $f(x) = f_1(x) + if_2(x)$, where f_1, f_2 are real-valued functionals. f_1, f_2 are real-linear since for any $x, y \in X$ and scalars $\alpha, \beta \in \mathbb{R}$ we have

$$\begin{aligned} f_1(\alpha x + \beta y) + if_2(\alpha x + \beta y) &= f(\alpha x + \beta y) \\ &= \alpha f(x) + \beta f(y) \\ &= \Big[\alpha f_1(x) + \beta f_1(y)\Big] + i\Big[\alpha f_2(x) + \beta f_2(y)\Big]. \end{aligned}$$

Moreover, we also have

$$\begin{aligned} f_1(ix) + if_2(ix) = f(ix) = if(x) &= if_1(x) - f_2(x) \\ \implies f_1(x) = f_2(ix) &\text{ and } f_2(x) = -f_1(ix). \\ \implies f(x) = f_1(x) &- if_1(ix). \end{aligned}$$

■

Theorem 3.2 (Complex Hahn-Banach Theorem)**.** *Let X be a real or complex vector space and p a real-valued functional on X satisfying*

(a) $p(x + y) \leq p(x) + p(y)$ *for all* $x, y \in X$*,*

(b) $p(\alpha x) = |\alpha| p(x)$ *for all* $\alpha \in \mathbb{C}$*.*

Let f be a linear functional defined on a subspace $Y \subset X$, satisfying

$$|f(y)| \leq p(y) \qquad \text{for all } y \in Y.$$

Then f has a linear extension $\tilde{f}$ from Y to X satisfying

(a) $\tilde{f}$ *is a linear functional on* X*,*

(b) $\tilde{f}|_Y = f$*, i.e. the restriction of* $\tilde{f}$ *to* Y *agrees with* f*,*

(c) $|\tilde{f}(x)| \leq p(x)$ *for all* $x \in X$

Proof. Observe that the result follows from Theorem 2.2 if X is real, so suppose X is complex; this means Y is complex. Lemma 3.1 states that $f(x) = f_1(x) - if(ix)$, with f_1 real-linear functional on Y. By the real Hahn-Banach Theorem 2.2, there exists a linear extension $\tilde{f}_1$ from Y to X. Setting $\tilde{f}(x) = \tilde{f}_1(x) - i\tilde{f}_1(ix)$, one can show that $\tilde{f}$ is indeed a complex linear extension of f from Y to X. We are left to show that $\tilde{f}$ satisfies (c).

Observe that the inequality is trivial if $\tilde{f}(x) = 0$, since $p(x) \geq 0$. Indeed,

$$0 = p(0) = p(x - x) \leq p(x) + p(-x) = p(x) + |-1|p(x) = 2p(x).$$

Suppose $\tilde{f}(x) \neq 0$. Writing $\tilde{f}(x) = |\tilde{f}(x)|e^{i\theta}$, we have

$$|\tilde{f}(x)| = e^{-i\theta}\tilde{f}(x) = \tilde{f}(e^{-i\theta}x) = \tilde{f}_1(e^{-i\theta}x) \leq p(e^{-i\theta}x) = |e^{-i\theta}|p(x) = p(x).$$

■

Remark 3.3. *One could make a crude estimate on $|\tilde{f}(x)|$ without appealing to polar form, but it fails to deliver what we want since*

$$|\tilde{f}(x)|^2 = \tilde{f}_1(x)^2 + \tilde{f}_1(ix)^2 \leq p(x)^2 + p(ix)^2 = 2p(x)^2.$$