321 2005–09

Chapter 2: Fundamentals

The objective in this chapter is to convey a very naive idea of Mathematical Logic (not even page 1 of course 371), and then to explain what the 'Axiom of Choice' is and why it is useful (by giving some example applications). Later we will use it for the Hahn-Banach theorem.

In this course, as in most others at this level, we need to pay close attention to details. It helps to keep in mind the structure of logical arguments. There are certain rules (called Axioms) which we follow, or take to be self-evident truths, and we often prove theorem which say

$$\text{hypothesis } A \Rightarrow \text{ conclusion } B.$$

The $\Rightarrow$ means 'implies' and the meaning of the theorem (the assertion we intend to prove) is that

> Under our usual axiom system, any time we are in a situation where A is valid, we can be sure that B will also hold.

We have to be sure we have an incontrovertible proof of any theorem, or we run the risk of ending up with nonsensical conclusions.

The codification of what a proof is belongs to the Logic course, but there are genuine concerns. These include

(i) Russels paradox: the set of all sets A with the property that $A \notin A$, or

$$S = \{A : A \notin A\}$$

causes a problem. Is $S \in S$? or not?

The solution to this is to make restrictions on the things one is allowed to write down. First order logic is one system (a very restrictive one) that avoids this paradox.

(ii) Nevertheless, we can never prove that a set of rules is consistent (will not lead to contradictions).

(iii) There are 'theorems' (or assertions) which can neither be proved not disproved in any given system of axioms.

One motivation to be ultra careful about all these rules would be to build an automated theorem-prover or proof-checker. Though there is a lot of work in this direction, it seems to have limited success and the kind of proof that can be mechanically validated usually cannot be easily read by a human. For example, conventions that we often have such as that $x \in \mathbb{R}$, $z \in \mathbb{C}$ or $n \in \mathbb{N}$, can make things easier to follow. But for a mechanical proof, we have to spell out all such things in every line.

The axiom of choice is known not to be provable, but also assuming it to be true does not introduce any inconsistency. The axiom seems harmless enough when we see it, and it is usual to take it to be true. However, it does lead to proofs of remarkable results and so it is perhaps not so harmless. Many of the remarkable results proved using the axiom of choice cannot be proved without it (because knowing them one could prove the axiom of choice).

Before stating the axiom, some further reinforcement to make sure you remember the rules. The symbol $\forall$ is translated as "for all", the symbols $\exists$ as "there exists" and the symbol $\iff$ as "if and only if". This last one is really important to digest. If we claim a theorem

$$\text{statement } A \iff \text{statement } B$$

then we *always* have two **independent** things to establish. One is $A \Rightarrow B$ (so starting from the knowledge that A holds in some situation, we have to show B). The other part is $B \Rightarrow A$ (starting from a maybe quite different scenario where we know B is valid, we have to establish A).

So a proof of $A \iff B$ should always be divided into a proof of $A \Rightarrow B$ and a second proof that $B \Rightarrow A$. Without this structure, the proof cannot even start to be right. Another thing to keep in mind is that the steps in the proof of the $A \Rightarrow B$ will be (usually) valid because we are supposing A to be true. We cannot re-use those steps in the $B \Rightarrow A$ argument because the assumption is now that B holds (and A is what we have to prove, not what we know).

The relationship between the $A \iff B$ notation and the wording "A if and only if B" is that $A \Rightarrow B$ (A implies B) can be phrased "A only if B". The other direction $A \Leftarrow B$ (alternatively written $B \Rightarrow A$) can be phrased "A if B".

Definition 2.1. If I is any set (which we take as an index set) and $\{A_i : i \in I\}$ is a family of sets indexed by I, then a *choice function* for the family is a function

$$f\colon I \to \bigcup_{i \in I} A_i$$

with the property that $f(i) \in A_i$ holds for all $i \in I$.

Axiom 2.2 (Axiom of Choice)**.** *If $\{A_i : i \in I\}$ is a family of sets with the property that $A_i \neq \emptyset$ holds for all $i \in I$, then there is a choice function for the family.*

Remarks 2.3. (i) It is clear that if we have a family of sets $\{A_i : i \in I\}$ where there is just one $i_0 \in I$ so that the corresponding set A_{i_0} is empty, then we cannot have any choice function. Clearly the requirement $f(i_0) \in A_{i_0}$ cannot be satisfied if $A_{i_0} = \emptyset$.

(ii) If $I = \{i_0\}$ has just one element and we have a family of one nonempty set $\{A_{i_0}\}$, then we can prove there is a choice function. Simply use the fact that $A_{i_0} \neq \emptyset$ to deduce that there must be at least one element $x_0 \in A_{i_0}$. Pick one such element and define $f(i_0) = x_0$. Then we have a choice function

$$f\colon I \to \bigcup_{i \in I} A_i = A_{i_0}$$

(iii) If $I = \{1, 2, \ldots, n\}$ or if I is any finite set, we can also prove that the conclusion of the axiom of choice is valid. The reason is basically that we have only to make a finite number of choices $x_j \in A_j$ for $j = 1, 2, \ldots, n$ and to define the choice function f by $f(j) = x_j$.

To organise this proof better, we could make it into a proof by induction that the assertion about n is true:

> If $I = \{1, 2, \ldots, n\}$ and $\{A_i : i \in I\}$ is a family of nonempty sets indexed by I, then there is a choice function for the family.

We proved the initial induction step ($n = 1$) already. The idea to go from assuming true for n to establishing it for $n + 1$ is to make a choice $x_{n+1} \in A_{n+1}$. Then use the induction hypothesis to get a choice function when A_{n+1} is omitted from the family, that is $f\colon \{1, 2, \ldots, n\} \to \bigcup_{i=1}^{n} A_i$ so that $f(i) \in A_i$ ($1 \leq i \leq n$). And finally define the desired choice function to be the same as this f on $\{1, 2, \ldots, n\}$ and to take the value x_{n+1} at $n + 1$.

(iv) We could even start this induction at $n = 0$ if we wanted. If I has 0 elements, it is the empty set, and there are no sets in the family $\{A_i : i \in I\}$. Since there are no sets A_i it is vacuously true they are all nonempty. But it is also true that there is a choice function!

A function with domain $I = \emptyset$ can't have any values. If we think of a function $f\colon X \to Y$ between two sets as a set of ordered pairs in $X \times Y$ satisfying certain rules, then the empty function will satisfy the rules when $X = \emptyset$. In the case $I = \emptyset$, the empty function works as a choice function. So technically (and rather bizarrely) this case is ok.

(v) The difficulty with proving the axiom of choice is in trying to write down the fact that there are (if I is a big set) very many choices to be made so as to pick $f(i) \in A_i$ for each $i \in I$. For any one $i \in I$, we have the assumption $A_i \neq \emptyset$ and so there is some element $x_i \in A_i$. We could say "pick such an x_i and we will put $f(i) = x_i$". The problem is that there are so many choices to be made that we would never be done making the choices.

The rules of logic (which we have not gone into) prevent you from formally writing down a proof that this should be possible. That is where the axiom of choice comes in.

Definition 2.4. If I is an index set, and $\{A_i : i \in I\}$ is a family of sets, then their *Cartesian product*, which we denote by

$$\prod_{i \in I} A_i,$$

is the set of all choice functions for the family.

Remarks 2.5. (i) The Axiom of Choice (2.2) can be restated

$$\prod_{i \in I} A_i \neq \emptyset \text{ if } A_i \neq \emptyset \forall i \in I.$$

(ii) When we have just two sets, say $I = \{1, 2\}$ and sets A_1, A_2, there are only two values $f(1)$ and $f(2)$ needed to specify any choice function.

So choice functions f can be specified by listing the two values $(f(1), f(2))$ as an ordered pair. We get then the more familiar view of the Cartesian product of two sets as a set of ordered pairs

$$A_1 \times A_2 = \{(x_1, x_2) : x_1 \in A_1, x_2 \in A_2\}.$$

(iii) From this perspective, the Axiom of Choice seems innocuous. Our next goal is to state Zorn's lemma, something equivalent that is quite often used as part of an argument. To do that we need some terminology about partially ordered sets.

Definition 2.6. A *partial order* $\leq$ on a set S is a relation on S satisfying the following properties

(i) $x \leq x$ is true for all $x \in S$

(ii) $x, y \in S$, $x \leq y$ and $y \leq x \Rightarrow x = y$

(iii) (transitivity) $x, y, z \in S, x \leq y, y \leq z \Rightarrow x \leq z$

The set S together with some chosen partial order is called a *partially ordered set* $(S, \leq)$.

Examples 2.7. (i) The most familiar example of $\leq$ is the one for numbers (real numbers $S = \mathbb{R}$, rational numbers $S = \mathbb{Q}$, integers $S = \mathbb{Z}$, natural numbers $S = \mathbb{N}$). In fact we can take $S \subseteq \mathbb{R}$ to be any subset and keep the usual interpretation of $\leq$ to get a partially ordered set $(S, \leq)$.

These examples are rather too special to get a general feel for what a partial order might be like.

(ii) A fairly general kind of example comes by taking $\leq$ to be set inclusion $\subseteq$.

Start with any set T and take $S = \mathcal{P}(T) =$ the power set of T, by which we mean the set of all subsets of T,

$$\mathcal{P}(T) = \{R : R \subseteq T\}.$$

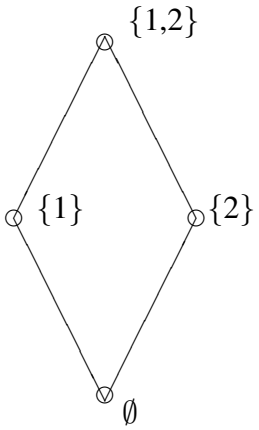
Declare $R_1 \leq R_2$ to mean $R_1 \subseteq R_2$.

It is not hard to see that the rules for a partial order are satisfied.

(iii) If we take small sets for T, we can picture what is going on.

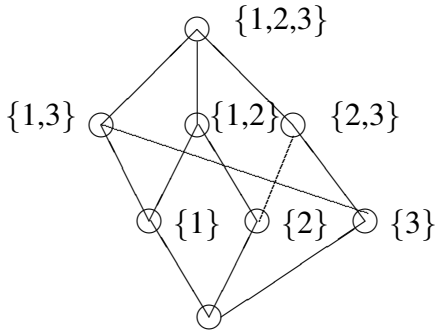
If $T = \{1, 2\}$ has 2 elements, then $\mathcal{P}(T)$ has $2^2 = 4$ elements

$$\mathcal{P}(\{1, 2\}) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}.$$



What the picture represents is the inclusion ordering between the sets. $A \subseteq B$ means we can get from A to B traveling upwards always. The picture shows that $A = \{1\}$ and $B = \{2\}$ are not related by inclusion. Neither is contained in the other. This is why the word 'partial' is used.

If we go to a set with 3 elements and $2^3 = 8$ subsets, the picture will become a little trickier to draw. Some paths cross.



You can see now that the single element subsets $\{1\}$, $\{2\}$, $\{3\}$ are not comparable one to another, and neither are the 2 element subsets. There are sets like $\{1\}$ and $\{2, 3\}$ with different numbers of elements which are also not comparable.

(iv) One way to get new partially ordered sets from old ones is to take subsets. If $(S, \leq)$ is any partially ordered set nd $S_0 \leq S$ is any subset we can use the same $\leq$ on S_0, or restricted to S_0, to get a partial order on S_0.

If we want to make it more formal, we define the relation $\leq_0$ on S_0 by the rule

$$x \leq_0 y \iff x \leq y.$$

So $x \leq_0 y$ for $x, y \in S_0$ means that $x \leq y$ holds in S. The difference between $\leq_0$ and $\leq$ is that $\leq_0$ forgets whatever relations there were involving elements of $S \setminus S_0$.

It is not hard to verify that $\leq_0$ is a partial order on S_0.

We call it the induced order on S_0. Looking at the picture above where $S = \mathcal{P}(\{1, 2, 3\})$ we could take

$$S_0 = \{\emptyset, \{1\}, \{2\}, \{1, 2, 3\}\}$$

and we would get the picture for $(S_0, \leq_0)$ by erasing all nodes in the picture corresponding to the other 4 subsets. In fact, if we do this, we will end up with a picture that is really similar to the picture for a power set of a set with 2 elements.

We can end up with other pictures by taking different S_0.

Definition 2.8. A *linear ordering* $\leq$ on a set S is a partial order with the additional property

$$x, y \in S \Rightarrow x \leq y \text{ or } y \leq x.$$

(In other words, every pair of elements are comparable.)

A *linearly ordered set* is a set S with a linear ordering $\leq$ specified.

Definition 2.9. If $(S, \leq)$ is a partially ordered set, then a *chain* in S is a subset $C \subseteq S$ that becomes linearly ordered in the induced order.

Example 2.10. In the case $S = \mathcal{P}(\{1, 2, 3\})$, one possible chain is

$$C = \{\emptyset, \{1, 3\}, \{1, 2, 3\}\}.$$

Definition 2.11. If $(S, \leq)$ is a partially ordered set, and $R \subseteq S$ is any subset, then an element $u \in S$ is called an *upper bound* for R if it is true that

$$x \leq u \forall x \in R.$$

Example 2.12. In $\mathbb{R}$, with the usual order, there is no upper bound for the subset $\mathbb{N}$.

In $S = \mathcal{P}(\{1, 2, 3\})$ (with set inclusion $\subseteq$ as our $\leq$), if we take

$$R = \{\emptyset, \{1, 3\}, \{2, 3\}\}$$

then we do have an upper bound $u = \{1, 2, 3\}$ in S for R.

Notice however, in the same setting, that $\{1,3\}$ is *not* an upper bound for R. For example $\{2,3\} \not\subseteq \{1,3\}$ So $\{1,3\}$ fails to be an upper bound because it is not larger than everything in R.

But, there is a little difference that comes because we have only a partial order and not a linear order. In the case of $\mathbb{R}$, if something is not an upper bound for a subset, then there is something larger in the subset. In our example, there is nothing in R strictly bigger than $\{1,3\}$. There are elements in R that are not comparable to $\{1,3\}$.

Definition 2.13. If $(S, \leq)$ is a partially ordered set, then $m \in S$ is called a *maximal element* of S if

$$x \in S, m \leq x \Rightarrow m = x.$$

(There is nothing in S strictly larger than m.)

Example 2.14. This is **not** the same as an element of S that is an upper bound for S. If we take a subset of the example drawn above where $S_0 = \{\emptyset, \{1\}, \{2\}, \{3\}\} \subseteq \mathcal{P}(\{1,2,3\})$, we can see that each of $\{1\}$, $\{2\}$ and $\{3\}$ is maximal inside S_0. (There is nothing in S_0 strictly larger.)

Theorem 2.15 (Zorn's lemma)**.** *If $(S, \leq)$ is a (nonempty) partially ordered set, with the property that every chain $C \subset S$ has an upper bound in S, then S has a maximal element.*

Proof. We postpone the proof (which uses the Axiom of Choice) and in fact we will leave the hard parts to an appendix. It is interesting that the proof uses the axiom of choice, but the details are somewhat tedious.

We will show some examples of using Zorn's lemma before saying anything about the proof. □

Proposition 2.16. *A (nonzero) unital ring contains a maximal proper ideal.*

Recall that a *ring* R is an abelian group under an operation $+$ and has a multiplication satisfying reasonable properties (distributivity $x(y+z) = xy + xz$ and $(x+y)z = xz+yz$). Associativity ($(xy)z = x(yz)$) is usually assumed but it is not really necessary for us. A unital ring is one where there is a multiplicative identity $1_R \in R$ (satisfying $1_R x = x = x1_R$, $x \in R$). It is common to assume $1_R \neq 0$ but if $1_R = 0$ then for each $x \in R$ we have $x = 1_R x = 0x = 0$, and so $R = \{0\}$. An ideal $I \subseteq R$ is a subgroup under addition that satisfies $rx, xr \in I \forall x \in I, r \in R$.

By a maximal (proper) ideal $M \subseteq R$ we mean an ideal with $M \neq R$ and the maximality property that if $M \subseteq I \subsetneq R$ for any (proper) ideal I, then $M = I$.

Proof. The idea is to apply Corns lemma 2.15 where the the partially ordered set S is the set of all proper ideals $I \subsetneq R$ and the relation $\leq$ is $\subseteq$. The proposition states exactly that $(S, \leq)$ must have a maximal element and this is also the conclusion of Corns lemma. So what we need to do is establish the hypotheses of Corns lemma (that every chain in S has an upper bound).

Let $C \subseteq S$ be a chain. If $C = \emptyset$ then we can take $\mathcal{I} = \{0\}$ as the upper bound. Since $1 \neq 0$, $\mathcal{I} \in S$ (that is it is a proper ideal). $\mathcal{I}$ is vacuously an upper bound for the empty chain.

If C is nonempty, we put $\mathcal{I} = \bigcup_{I \in C} I$. Clearly $I \subseteq \mathcal{I} \forall I \in C$ and the issue in proving that $\mathcal{I}$ is an upper bound for C (in S) is to show $\mathcal{I} \in S$. Now $0 \in \mathcal{I}$ since $C \neq \emptyset$, so $\exists I \in C$ and thus $0 \in I \Rightarrow 0 \in \mathcal{I}$. Next to establish that I is an additive subgroup of R we need to show that $x, y \in \mathcal{I} \Rightarrow x - y \in \mathcal{I}$.

Now $x, y \in \mathcal{I}$ implies $\exists I_1, I_2 \in C$ with $x \in I_1$ and $y \in I_2$. As C is a chain we have either $I_1 \subseteq I_2$ or $I_2 \subseteq I_1$. Say $I_1 \subseteq I_2$ as the other case is similar. Then we have $x, y \in I_2$ and so $x - y \in I_2 \subseteq \mathcal{I}$ (because I_2 is an ideal).

Next to show $\mathcal{I}$ has the ideal property, pick $x \in \mathcal{I}$ and $r \in R$. Then $x \in I$ for some $I \in C$ and so $rx, xr \in I \subseteq \mathcal{I}$. Thus $rx, xr \in \mathcal{I}$.

Finally, we must show $\mathcal{I} \neq R$ (that is that $\mathcal{I}$ is a proper ideal and so $\mathcal{I} \in S$). To do this we note that $1_R \notin \mathcal{I}$. If $1_R \in \mathcal{I}$, then $1_R \in I$ for some $I \in C$ and so $r = r1_R \in I \forall r \in R$ (by the ideal property of I) and this would mean $R = I$, $I \notin S$. □

The above is a fairly typical application of Corns lemma, but other applications can require more technicalities.

Definition 2.17. Let V be an arbitrary vector space over any field F.

A subset $E \subseteq V$ is called *linearly independent* if whenever $n \geq 1$, $e_1, e_2, \ldots, e_n \in E$ distinct elements, $\lambda_1, \lambda_2, \ldots, \lambda_n \in F$ and $\lambda_1 e_1 + \lambda_2 e_2 + \cdots + \lambda_n e_n = 0$, then it must be that $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0$.

A subset $E \subseteq V$ is said to *span* V if $v \in V$ implies there is $n \geq 0$, $e_1, e_2, \ldots, e_n \in E$ and $\lambda_1, \lambda_2, \ldots, \lambda_n \in F$ so that $v = \lambda_1 e_1 + \lambda_2 e_2 + \cdots + \lambda_n e_n$. Here we interpret the empty sum (the case of $n = 0$) to be the zero vector $0 \in V$.

A subset $E \subseteq V$ is called a *basis* for V if it is both linearly independent and a spans V.

Observe that the empty set is linearly independent and forms a basis for the zero vector space.

Lemma 2.18. *Let V be a vector space over a field F. Then a subset $E \subseteq V$ is a basis for $V \iff$ it is a maximal linearly independent subset of V.*

Here by maximal we mean maximal with respect to $\subseteq$. So taking S to be the set (or collection) of all linearly independent subsets of V and the relation $\subseteq$ on S we are considering a maximal element of $(S, \subseteq)$.

Proof. $\Rightarrow$: We start with $E \subseteq V$ a basis.

Let $E \subseteq E_1 \subseteq V$ with E_1 linearly independent. If $E_1 \neq E$, then there exists $v \in E_1 \setminus E$. As E spans V, there are $e_1, e_2, \ldots, e_n \in E$ and scalars $\lambda_1, \lambda_2, \ldots, \lambda_n$ with $v = \lambda_1 e_1 + \lambda_2 e_2 + \cdots + \lambda_n e_n$. We can suppose that $e_1, e_2, \ldots, e_n$ are distinct elements of E because if (say) $e_1 = e_2$ we can rewrite $v = \lambda_1 e_1 + \lambda_2 e_2 + \cdots + \lambda_n e_n = (\lambda_1 + \lambda_2)e_2 + \lambda_3 e_3 + \cdots + \lambda_n e_n$ as shorter linear combination. We can continue to do this to eliminate repetitions among the $e_j \in E$. Suppose then that this has been done, that the value of n has been suitably adjusted and the elements of E renumbered to get $e_1, e_2, \ldots, e_n \in E$ distinct.

Now we can rearrange $v = \lambda_1 e_1 + \lambda_2 e_2 + \cdots + \lambda_n e_n$ as

$$(-1)v + \lambda_1 e_1 + \lambda_2 e_2 + \cdots + \lambda_n e_n = 0.$$

As $v, e_1, e_2, \ldots, e_n \in E_1$ are distinct and the scalar coefficients are not all 0 (because $-1 \neq 0$ in all fields), we see that E_1 cannot be linearly independent. this contradiction forces $E = E_1$ and shows that E is a maximal linearly independent subset of V.

$\Leftarrow$: We start with $E \subseteq V$ a maximal linearly independent subset of V.

To show it spans V, pick $v \in V$. If $v \in E$, then we can write $v = 1.v$ as a linear combination of elements of E. If $v \notin E$, then $E \subsetneq E \cup \{v\} \subseteq V$ means that $E \cup \{v\}$ cannot be linearly independent by maximality of E. So there are distinct vectors $e_1, e_2, \ldots, e_n \in E \cup \{v\}$ and scalars $\lambda_1, \lambda_2, \ldots, \lambda_n$ not all zero with $\lambda_1 e_1 + \lambda_2 e_2 + \cdots + \lambda_n e_n = 0$. We must have $v \in \{e_1, e_2, \ldots, e_n\}$ as otherwise we would contradict linear independence of E. So we may as well assume $e_1 = v$ (by renumbering if necessary). Also then $\lambda_1 \neq 0$ (as otherwise we would again contradict linear independence of E). Thus

$$v = (-\lambda_2/\lambda_1)e_2 + (-\lambda_3/\lambda_1)e_3 + \cdots + (-\lambda_n/\lambda_1)e_n$$

is a linear combination of elements of E. (If $n = 1$ it is the empty sum and $v = 0$ but that is ok.) This shows that E spans V. Hence E is a basis for V. □

Theorem 2.19. *Every vector space has a basis.*

Proof. Let V be a vector space.

The idea is to apply Zorn's lemma 2.15 to $S = \{E : E \subseteq V \text{ linearly independent}\}$ ordered by $\subseteq$.

S is not empty since the empty set $E = \emptyset \in S$. Thus the empty chain in S has an upper bound in S.

For any (empty or not) chain $C \in S$, we can find an upper bound for C by taking $E_0 = \bigcup\{E : E \in C\}$. What we have to show is that $E_0 \in S$ (that is that E_0 is linearly independent) as $E \subseteq E_0$ is clearly true $\forall E \in C$.

Let $n \geq 1$ and $e_1, e_2, \ldots, e_n \in E_0$ distinct elements. Suppose $\lambda_1, \lambda_2, \ldots, \lambda_n$ are scalars (elements of the base field) with $\lambda_1 e_1 + \lambda_2 e_2 + \cdots + \lambda_n e_n = 0$. For each $1 \leq j \leq n$ there is $E_j \in C$ with $e_j \in E_j$.

By induction on n we can find $E'_n \in \{E_1, E_2, \ldots, E_n\}$ with $e_1, e_2, \ldots, e_n \in E'_n$. To show this consider E_1 and E_2. Both are in the chain C and so $E_1 \subseteq E_2$ or $E_2 \subseteq E_1$. In the first case let $E'_2 = E_2$ and in the second let $E'_2 = E_1$. Then $e_1, e_2 \in E'_2$.

Next $E'_2 \subseteq E_3$ or $E_3 \subseteq E'_2$. Taking E'_3 to be the larger of the two sets, we get $e_1, e_2, e_3 \in E'_3$. Continuing in this way (or, more formally, using induction on n) we can show the existence of E'_n.

By linear independence of E'_n we can conclude $\lambda_1 = \lambda_2 = \cdots = \lambda_n$.

Thus E_0 is linearly independent. $E_0 \in S$ is an upper bound for the chain C.

We have shown that every chain in S has an upper bound. By Zorn's lemma, S has a maximal element. By the previous lemma, such a maximal element of S is a basis for V. □

Proposition 2.20. *Let V be a vector space (over any field) and $I \subset V$ a linearly independent set. Then there is a basis E of V with $I \subset E$.*

Proof. This proposition is a generalisation of Theorem 2.19. The reason is that we can take $I =$ set as a linearly independent subset of V and then the conclusion that there is a basis E with $\emptyset \subset E$ just amounts to the statement that there is a basis E for V.

The proof is a modification of the proof of Theorem 2.19.

The idea is to apply Zorn's lemma 2.15 to $S = \{E : E \subseteq V \text{ linearly independent and } I \subset E\}$ ordered by $\subseteq$. The proof is no different except the fact that $S \neq \emptyset$ now follows from $I \in S$.

We then have an upper bounded $E = I$ for the empty chain in S. The exact same proof as above shows that a non-empty chain C in S has a linearly independent union $E_0 = \bigcup_{E \in C} E$. To show $E_0 \in S$, note that $I \subset E_0$ because C is nonempty. □

Examples 2.21. (i) $\mathbb{R}$ is a vector space over $\mathbb{Q}$. For instance, $\sqrt{2}$ is irrational and it follows that $I = \{1, \sqrt{2}\}$ is linearly independent over the rationals. From the proposition, it follows that $\mathbb{R}$ has a basis B over $\mathbb{Q}$ with $\{1, \sqrt{2}\} \subset B$.

Because $\mathbb{R}$ is uncountable, any basis B has to be uncountable. The reason is that, if B was countable, the collection of finite subsets of B would also be countable. For each finite subset $\{b_1, b_2, \ldots, b_n\} \subset B$, there are just countably many rational linear combinations

$$\sum_{j=1}^{n} q_j b_j$$

with rational coefficients $(q_1, q_2, \ldots, q_n) \in \mathbb{Q}^n$.

(ii) Every Banach space E over $\mathbb{K}$ (where $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$ as usual) has a basis over $\mathbb{K}$. Here we mean an algebraic basis, which involves only finite sums. This kind of basis is sometimes called a *Hamel basis*, because in Banach spaces one can consider infinite linear combinations. The kind of basis where infinite linear combinations are allowed is called a *Schauder basis*. A special case arses in Hilbert spaces (which we will come to soon) and are called an *orthonormal basis*. Please take note that these concepts are quite different from the algebraic bases we have discussed above.

We have not said anything about the proof of Zorn's lemma (using the axiom of choice). Just one hint. It is quite easy to prove it from something else, which we now state. Unfortunately the proof of this theorem is not easy. All will go in the appendix in case you are curious.

Theorem 2.22 (Hausdorff's maximality principle)**.** *If $(S, \leq)$ is a partially ordered set, then S contains a maximal chain.*

By this we mean that the set (or collection) of all chains in S, ordered by set inclusion $\subseteq$, has a maximal element.

A Proof of Hausdorff's maximality principle

The first ingredient in the proof will be the following rather technical lemma.

Lemma A.1. *Let $\mathcal{F}$ be a nonempty collection of subsets of a set X and $g\colon \mathcal{F} \to \mathcal{F}$ a function. Suppose*

*(i) $\mathcal{F}$ contains the union of every nonempty chain in $\mathcal{F}$ (*i.e. *the union of any nonempty subcollection of $\mathcal{F}$ which is linearly ordered by set inclusion).*

(ii) $\forall A \in \mathcal{F}$, $A \subseteq g(A)$ and $g(A) \setminus A$ has at most one element.

Then there exists some $A \in \mathcal{F}$ with $g(A) = A$.

Proof. Fix $A_0 \in \mathcal{F}$. Call a subcollection $\mathcal{F}' \subset \mathcal{F}$ a *tower* if

(a) $A_0 \in \mathcal{F}'$

(b) $A \in \mathcal{F}' \Rightarrow g(A) \in \mathcal{F}'$, and

(c) $\mathcal{F}'$ satisfies (i) above.

The family of all towers is nonempty. To see this look at

$$\mathcal{F}_1 = \{A \in \mathcal{F} : A_0 \subseteq A\}$$

(or even $\mathcal{F}_2 = \mathcal{F}$). $\mathcal{F}_1$ is a tower (and so is $\mathcal{F}_2$).

Now let $\mathcal{F}_0$ denote the intersection of all towers. Then $\mathcal{F}_0$ is a tower (check).

We next show that $\mathcal{F}_0$ is a chain in $\mathcal{F}$. We achieve this by a rather indirect approach.

Consider

$$\begin{aligned} \Gamma &= \{D \in \mathcal{F}_0 : \text{ every } A \in \mathcal{F}_0 \text{ satisfies either } A \subseteq D \text{ or } D \subseteq A\} \\ &= \{D \in \mathcal{F}_0 : \text{ every } A \in \mathcal{F}_0 \text{ is comparable to } D\}. \end{aligned}$$

(Aside: Maybe Γ is not too mysterious. We want to prove that $\mathcal{F}_0$ is a chain, which is the same as $\mathcal{F}_0 = \Gamma$.)

For each $D \in \Gamma$, let

$$\Phi_D = \{A \in \mathcal{F}_0 : \text{ either } A \subseteq D \text{ or } g(D) \subseteq A\}.$$

Now $A_0 \in \mathcal{F}_0$ because $\mathcal{F}_0$ is a tower. $\mathcal{F}_0 \subseteq \mathcal{F}_1$ and hence $A_0 \in \Gamma$. Also $A_0 \in \Phi_D$ for all $D \in \Gamma$ since $A_0 \subseteq D$. Therefore Γ and each Φ_D satisfies (a).

If $\mathcal{C}$ is a chain in Γ, consider

$$E = \bigcup_{D \in \mathcal{C}} D.$$

For any $A \in \mathcal{F}_0$, we have either $A \subseteq D$ or $D \subseteq A$, for each $D \in \mathcal{C}$. If $A \subseteq D$ for one $D \in \mathcal{C}$, then $A \subseteq E$. Otherwise, $D \subseteq A$ for all $D \in \mathcal{C}$ and so $E \subseteq A$. Thus $E \in \Gamma$. This shows that Γ satisfies (c).

Consider now a chain $\mathcal{C}$ in Φ_D (some D) and let $B = \bigcup_{A \in \mathcal{C}} A$. Since each $A \in \mathcal{C}$ is in Φ_D, we have either $A \subseteq D$ or $g(D) \subseteq A$ for $A \in \mathcal{C}$.

If $A \subseteq D$ for each $A \in \mathcal{C}$, then $B \subseteq D$. Otherwise $g(D) \subseteq A$ for some A and then $g(D) \subseteq B$. Thus we have $B \in \Phi_D$. This shows that Φ_D satisfies (c).

Next we prove that Φ_D satisfies (b) (for fixed $D \in \Gamma$). Let $A \in \Phi_D$ and our aim now is to check that $g(A) \in \Phi_D$.

We know $A \subseteq D$ or $g(D) \subseteq A$. We consider three separate cases:

$$A \subsetneq D, \quad A = D, \quad g(D) \subseteq A.$$

we know $g(A) \in \mathcal{F}_0$ and therefore either $D \subseteq g(A)$ or $g(A) \subseteq D$.

If $A \subsetneq D$, then we cannot have $D \subsetneq g(A)$, because that would mean

$$g(A) \setminus A = (g(A) \setminus D) \cup (D \setminus A)$$

would be a union of two nonempty sets, and so have at least two elements — contradicting assumption (ii). Hence $g(A) \subseteq D$ if $A \subsetneq D$.

If $A = D$ or $g(D) \subseteq A$, we have $g(D) \subseteq g(A)$.

Altogether then, we have $g(A) \subseteq D$ or $g(D) \subseteq g(A)$, so that $g(A) \in \Phi_D$. Thus ϕ_D satisfies (b).

Now we know that ϕ_D is a tower. Since $\Phi_D \subseteq \mathcal{F}_0$ and $\mathcal{F}_0$ is the smallest possible tower, we must have

$$\Phi_D = \mathcal{F}_0$$

and this is true of all $D \in \Gamma$.

Thus, for $D \in \Gamma$,

$$\begin{aligned} A \in \mathcal{F}_0 &\Rightarrow A \subseteq D \text{ or } g(D) \subseteq A \\ &\Rightarrow A \subseteq g(D) \text{ or } g(D) \subseteq A \end{aligned}$$

Thus $g(D) \in \Gamma$.

Therefore Γ satisfies (b). So Γ is a tower. Just as for Φ_D, this means $\Gamma = \mathcal{F}_0$ (because Γ is a tower inside the smallest possible tower). So, finally, we know that $\mathcal{F}_0$ is linearly ordered (every two elements $A, D \in \mathcal{F}_0$ are comparable).

Let $B = \bigcup_{A \in \mathcal{F}_0} A$. By (b), $B \in \mathcal{F}_0$ (because $\mathcal{F}_0$ is a chain in itself). By (c), $g(B) \in \mathcal{F}_0$. Thus

$$B \subseteq g(B) \subseteq \bigcup_{A \in \mathcal{F}_0} A = B.$$

Therefore $B = g(B)$. □

Proof. (of 2.22): Let $(S, \leq)$ be a nonempty partially ordered set. Let $\mathcal{F}$ be the collection of all linearly ordered subsets of S. Since singleton subsets of S are linearly ordered (and $S \neq \emptyset$), $\mathcal{F}$ is not empty.

Notice that the union of any chain of linearly ordered subsets of S is again linearly ordered (check this). So $\mathcal{F}$ satisfies condition (i) of Lemma A.1.

For each $A \in \mathcal{F}$, let A^* denote the set of all $x \in S \setminus A$ such that $A \cup \{x\}$ is linearly ordered (*i.e.* $A \cup \{x\} \in \mathcal{F}$). If $A^* = \emptyset$ for any $A \in \mathcal{F}$, then A is a maximal chain in S and we are done.

Otherwise, let f be a choice function for the family

$$\{A^* : A \in \mathcal{F}\}.$$

(This is where we use the axiom of choice!) This means that $A \cup \{f(A)\}$ is linearly ordered for each $A \in \mathcal{F}$. Put

$$g(A) = A \cup \{f(A)\}$$

for $A \in \mathcal{F}$.

Then Lemma A.1 tells us that there is some $A \in \mathcal{F}$ with $g(A) = A$. This contradicts the construction of g. To avoid the contradiction, we must have some $A \in \mathcal{F}$ with $A^* = \emptyset$, *i.e* a maximal chain in S. This proves Theorem 2.22. □

Proof. (of Zorn's Lemma 2.15 using 2.22) Let $(S, \leq)$ be a partially ordered set with the property that each chain in S has an upper bound in S.

Applying Hausdorff's maximality principle 2.22, we know there is a maximal chain C in S.

By hypothesis C has an upper bound $u \in S$. We claim that any such upper bound must be in C.

To show this we notice first that $C \cup \{u\}$ is a chain in S. This is simple to check, but here are the details.

Let $x, y \in C \cup \{u\}$. We have to show that either $x \leq y$ or $y \leq x$. There are two cases to consider

(i) $x, y \in C$.

In this case we know $x \leq y$ or $y \leq x$ because C is a chain.

(ii) x or y (or both) is u. If (say) $x = u$ then $y \leq u = x$ because u is an upper bound for C (or because $u \leq u$ in case $y = u$ also). Similarly, if $y = u$ we can conclude $x \leq y$.

Now $C \cup \{u\}$ is a chain that contains the (maximal) chain C and so $C \cup \{u\} = C \Rightarrow u \in C$.

Finally we claim that u is a maximal element of S. If $u \leq a \in S$ then a is also an upper bound for C (because $x \in C \Rightarrow x \leq u \Rightarrow x \leq a$ by transitivity of $\leq$). By the previous part of the argument the upper bound $a \in C$. So $a \leq u$. Combining with $u \leq a$ we conclude that $u = a$. Thus u is maximal in S. $\square$