Zorn’s Lemma, ZL, is an equivalent of the Axiom of Choice, AC. However it is often a more useful variant, especially in parts of algebra, geometry, and topology. We will look at various applications later in this document.

The naming of this result is one of those curiosities that happen in Mathematics. Here is the story as I remember it, with some information from the St Andrews web-page on Max Zorn.

Max Zorn, 1906–1993, was a good mathematician who did useful work in several areas. Early in his career, in 1935, he wrote a paper in which he used what he called a ‘maximum principle’ and is now called Zorn’s Lemma. He was working in a branch of Algebra, and at the time the Zermelo Well-Ordering principle, every set can be well-ordered, was the standard technique. It turned out that ZL is a much neater method of producing the required results. At the end of the paper he said that the AC, ZL, and the well-ordering principle are equivalent, but he never wrote up the proof.

Zorn did some useful work in several areas of Mathematics, but he is now remembered mainly for ZL. This didn’t please him too much.

It turned out that several similar principles have been used earlier.

On page 161 of ‘Basic Set Theory’ by A. Levy it is credited to Hausdorff in 1914. This is now known as Hausdorff’s maximizing principle. In ‘Elements of set theory’ by H.B. Enderton it is credited to Kuratowski in 1922, but it does say that Hausdorff result was similar in spirit. In ‘Set Theory’ by T. Jech the result is called the Kuratowski-Zorn Lemma. There is also something called the Teichmüller-Tukey Lemma, but I think this came later. We will look at these variants in this document.

You have to remember that these earlier proofs were put together at the time when set theory, both informal and axiomatic, was being sorted out. This could mean that the precise contents of the proofs may not be too clear.

In this document we first describe ZL, and then we look at various applications. We also look at the Krull Separation Lemma which is a variant of ZL useful in ring theory and certain other areas.

In Section 3 we look at a proof extracted from ‘Naive Set Theory’ by Paul Halmos. There it says that this proof is adapted from one by Zermelo, but doesn’t give a date.

Contents

1 Posets and their properties .................................................. 2
  1.1 Exercises: Posets and their properties ................................ 4
  1.2 Problems: Posets and their properties ............................... 5
2 Zorn’s Lemma and some uses .................................................. 5
  2.1 Exercises: Zorn’s Lemma and some uses ................................. 9
  2.2 Problem: Zorn’s Lemma and some uses ................................. 9
3 AC ⇒ ZL ............................................................................ 10
4 The Krull Separation Lemma ................................................... 14
  4.1 Exercises: The Krull Separation Lemma ............................... 16
  4.2 Problems: The Krull Separation Lemma ............................... 17
1 Posets and their properties

Zorn’s Lemma, ZL, gives us a method of ensuring that a poset has certain kinds of elements. To do that we need to impose certain restrictions on the poset. In this section we gather together all the relevant material needed.

Recall that a poset is a set $S$ which carries a comparison $\leq$, a reflexive, transitive, and anti-symmetric relations. Thus

$$x \leq x \quad x \leq y \leq z \implies x \leq z \quad x \leq y \land y \leq x \implies x = y$$

for all $x, y, z \in S$. Of course, we don’t always use ‘$\leq$’ for the carried comparison, especially if there are two posets around, and certainly if the one set carries two comparisons.

You should remember that the empty set $\emptyset$ is a poset. I will try to point out where this might be important.

We are interested in locating certain kinds of elements in a poset, those which are maximal or a maximum. To explain these notions it is instructive to look at more general versions.

1.1 DEFINITION. Let $S$ be a poset and consider a subset $X$ of $S$.

An element $a \in X$ is maximal in $X$ if

$$a \leq x \implies x = a$$

for all $x \in X$.

An element $a$ of $S$ is maximal if it is maximal in $S$, that is with $X = S$.

An element $a \in X$ is a maximum in $X$ if

$$x \leq a$$

for all $x \in X$.

An element $a$ of $S$ is a maximum if it is a maximum in $S$, that is with $X = S$. ■

Of course, if a set $X$ has a maximum element then this is the only possible maximal element of the set. However, a set can have many maximals but no maximum. You should also note that having maximals is concerned with the nature of the subset, not the whole poset. Be careful with this when you see ‘maximal’ used. Luckily, for the time being we are concerned only with the maximals or the maximum of the parent poset $S$.

A poset need not have any maximal elements. Think of $\mathbb{N}$ as a poset. So how can we show that a particular poset does have maximals? We check that it has certain properties and then apply ZL.

A poset $S$ is linear, a linear ordering, a linearly ordered set, if

$$x \leq y \text{ or } y \leq x$$

for all $x, y \in S$. On the whole we don’t usually consider linear posets, but we do often consider linear subsets of a poset.

1.2 DEFINITION. Let $S$ be a poset.

A chain in $S$ is a subset $X \subseteq S$ which is linearly ordered by the comparison on $S$, that is

$$x \leq y \text{ or } y \leq x$$
for all \( x, y \in X \).

An ascending \( \omega \)-chain in \( S \) is a linear subset
\[
a(0) \leq a(1) \leq \cdots \leq a(i) \leq \cdots \quad (i < \omega)
\]
which is indexed by \( \omega \) and where the comparison matches that on \( \omega \).

A descending \( \omega \)-chain in \( S \) is a linear subset
\[
a(0) \geq a(1) \geq \cdots \geq a(i) \leq \cdots \quad (i < \omega)
\]
which is indexed by \( \omega \) and where the comparison matches that on \( \omega \) in reverse order. ■

Of course, each \( \omega \)-chain, of either kind, is a chain. The reason for defining the two notions explicitly is that some people often think ‘chain’ means \( \omega \)-chain, and even more often they think it means ascending \( \omega \)-chain. It doesn’t.

Chains play an important role in the analysis, so it is useful to have some notation.

1.3 DEFINITION. Let \( S \) be an arbitrary poset. We write \( \mathcal{C}S \) for the collection of chain in \( S \), and we view this as a poset under extension. ■

We sometimes write \( \alpha, \beta, \gamma, \ldots \) for typical chains in \( S \). These are partially ordered by inclusion. For two chains \( \alpha, \beta \) of a parent poset we write
\[
\alpha \subseteq \beta
\]
to indicate that each member of \( \alpha \) is also a member of \( \beta \). The comparison on \( \alpha \) is the restriction of the comparison on \( \beta \), for both are restriction of the comparison on the parent poset \( S \). Note that the members of \( \alpha \) need not occur in any special positions in \( \beta \).

Finally, not that the empty set is a chain, so that \( \mathcal{C}S \) is always non-empty.

We also need a generalization of the idea of a non-empty chain.

1.4 DEFINITION. Let \( S \) be a poset. A subset \( X \) of \( S \) is directed (or more precisely, upwards directed) if \( X \neq \emptyset \), and for each \( x, y \in X \) there is some \( z \in X \) with \( x, y \leq z \).

Trivially, each non-empty chain is directed, but not every directed set is a chain. It doesn’t take to long to draw a picture of a non-linear directed set.

As I said at the start of this section, ZL is concerned with the existence of a certain kind of element of a poset.

1.5 DEFINITION. Let \( S \) be a poset and consider a subset \( X \) of \( S \). An element \( a \in S \) is an upper bound of \( X \) if
\[
x \leq a
\]
for each \( x \in X \).

An element \( a \in S \) is a least upper bound, or a supremum, of \( X \) if it is an upper bound, and if
\[
a \leq b
\]
for every upper bound \( b \) of \( X \). ■
It doesn’t take too long to see that each subset $X$ of a poset can have at most one supremum. We often write $\bigvee X$ for this supremum, when it exists.

Of course, a subset of a poset need not have a supremum, or even an upper bound. To organize the existence of these we need four more properties that a poset may or may not have. In the literature these properties are rarely, if ever, made explicit, so the following notation is not standard.

1.6 DEFINITION. Let $S$ be an arbitrary poset.
We say $S$ has $Z(S)$ if each chain of $S$ has an upper bound.
We say $S$ has $Y(S)$ if each directed subset of $S$ has an upper bound.
We say $S$ has $L(S)$ if $S$ has a maximal element.
We say $S$ has $K(S)$ if each element of $S$ lies below a maximal element.

Trivially we have
$$Y(S) \implies Z(S) \implies Z(S) \neq \emptyset$$

since each chain is directed and the empty set is a chain. We also have
$$K(S) \implies L(S)$$

provided $S$ is non empty.

Zorn’s Lemma and its variants are concerned with other implications between these properties.

1.1 Exercises: Posets and there properties

1.1 (a) Find a poset $S$ and a subset $X$ which has thousands of maximals but no maximum.
(b) Write down the definitions of a minimal and a minimum element of a subset of a poset.
(c) Find a poset and certain subsets which have
   (i) a maximum and no minimum
   (ii) a maximal and thousands of minimals

and so on.

1.2 Let $S$ be an arbitrary poset with a subset $X$.
(a) Show that if $X$ has a maximum then $X$ has a supremum $\bigvee X$.
(b) Show, by example, that $X$ may have two or more maximals and still have a supremum.
(c) Show, by example, that $X$ may have just one maximal but no supremum.
(d) Show, by example, that $X$ may have a supremum but no maximals

1.3 Find an example of two posets $S \subseteq T$, where the comparison on $S$ is the restriction of that on $T$, with a subset $X \subseteq S$, where $X$ has a supremum in both $S$ and $T$, but these are not the same.
  Find such an example where $S$ and $T$ are linear orders.
1.4 Let $X$ be a directed subset of a poset $S$.
    (a) Show that each finite subset of $X$ has an upper bound in $X$.
    (b) Find an example to show that an infinite subset of $X$ need not have an upper bound in $X$, nor in $S$.

1.5 Let $(S, \leq)$ be an arbitrary poset. A chain of $S$ is a subset $C \subseteq S$ which is linearly ordered by the comparison $\leq$ carried by $S$. Let $\mathcal{CS}$ be the set of such chains.
    (a) Show that $\mathcal{CS}$ is a poset under inclusion.
    (b) Show that $\mathcal{CS}$ is closed under directed unions.

1.6 Problems: Posets and their properties

1.6 Let $S$ be an arbitrary set and consider the family $\mathcal{POS}$ of partial orderings on $S$. Compare these by extension.
    (a) Show that $\mathcal{POS}$ is a poset.
    (b) Show that each directed subfamily of $\mathcal{POS}$ has an upper bound in $\mathcal{POS}$ which, in fact, is the supremum of the family.

1.7 (a) Let $(S, \leq)$ be a poset. Suppose $a, b$ are two elements that are not comparable, that is both $a \not\leq b$ and $b \not\leq a$ hold. Consider the relation $\sqsubseteq$ on $S$ given by
    \[
    x \sqsubseteq y \iff (x \leq y) \text{ or } (x \leq a \& y \leq b)
    \]
    for arbitrary $x, y \in S$. Show that $\sqsubseteq$ is a partial order that extends $\leq$.
    (b) Suppose that the comparison $\leq$ is maximal in the poset of all possible comparisons on $S$. Show that $\leq$ is linear.

2 Zorn’s Lemma and some uses

In this section we set up Zorn’s Lemma, ZL, and two of its variants. We also indicate some of its uses. As yet we can not prove ZL, for we don’t have the required background. However, as an exercise we indicate the strength of ZL, namely that $ZL \implies AC$. Later, in Section 3, we show that the converse, so $ZL$ and $AC$ are equivalent.

To discuss Zorn’s Lemma we introduce some abbreviations, one of which is ZL as used already. We make use of the abbreviations from Definition 1.6.

2.1 DEFINITION. Consider the following statements about the universe of sets.

(ZL$^+$) For each non-empty poset $S$ we have $Z(S) \implies K(S)$.

(ZL) For each non-empty poset $S$ we have $Z(S) \implies L(S)$.

(YL) For each non-empty poset $S$ we have $Y(S) \implies K(S)$.

(XL) For each non-empty poset $S$ we have $Y(S) \implies L(S)$.

The central one, ZL, is often referred to as Zorn’s Lemma. ■
At the end of Section 1 we observed that
\[ Y(S) \implies Z(S) \quad K(S) \implies L(S) \]
for each non-empty poset \( S \). Thus the implications
\[
\begin{align*}
ZL^+ & \implies ZL \\
\uparrow & \uparrow \\
YL & \implies XL
\end{align*}
\]
hold. Later in this section we indicate that these four properties are equivalent.

As above we often refer to ZL as Zorn’s Lemma. If necessary we call \( ZL^+ \) the strong Zorn’s Lemma. There isn’t a standard name for \( YL \) and \( XL \), and I will refrain from calling it Yorn’s Lemma.

At this point I should sort out the source of some linguistic confusion. I said that \( ZL \) is Zorn’s Lemma. Technically this is wrong. The actual Zorn’s Lemma is that

\[ ZL \text{ is true} \]

or more precisely

\[ \text{The standard axiom of set theory imply } ZL \]

which is what Zorn, and certain others, actually proved. However, here we don’t need to be so pedantic.

In the remainder of this section we look at some examples of the use of \( ZL \), and then we begin to consider the relationship with \( AC \).

For the first example consider a group \( G \). We think of this written multiplicatively. This group has subgroups. For instance the trivial group is a subgroup. Of course, this may be the only subgroup, and more generally it may be the only proper subgroup. Let \( A \) be the set of abelian (commutative) subgroups of \( G \). Again the trivial subgroup is a member of \( A \), so \( A \) is non-empty. This set \( A \) is a poset under inclusion. We wish to produce maximal members of \( A \). Of course, if \( G \) is abelian then \( G \in A \), and \( G \) is the maximum member of \( G \). We look at a more general case.

2.2 LEMMA. For the situation described above the poset \( A \) is closed under directed unions.

Proof. Let \( D \) be a directed subset of \( A \), so for \( A, B \in D \) there is some \( C \in D \) with \( A, B \subseteq C \). We need to check that \( \bigcup D \in A \). This is a couple of simple observations which we leave as an exercise.

This show that the poset \( A \) has \( Y(A) \), so an application of \( YL \) gives the following.

2.3 COROLLARY. Let \( G \) be an arbitrary group. Then each abelian subgroup is included in a maximal abelian subgroup.

There are many similar results concerning rings and ideals. We look at some of those in Section 4.

For the next example we consider the vectors \( V \) spaces over a given field \( K \). We know that when it has finite dimension the space \( V \) has a basis. We just keep producing larger and larger linearly independent sets, and the finite dimensionality ensures that this process will stop at some finite stage. That doesn’t need any choice principle. When the space has infinite dimension we have to use a maximizing argument.
2.4 **LEMMA.** Let \( V \) be a vector space over a field \( K \). Let \( \mathcal{L} \) be the poset of linearly independent subsets under inclusion. Then \( \mathcal{L} \) is closed under directed unions.

**Proof.** A subset \( X \) of \( V \) is linearly independent precisely when each finite subset of \( X \) is linearly independent. \( \blacksquare \)

Now an application of XL to \( \mathcal{L} \) shows that there is a maximal linearly independent set. A little bit more work, virtually the same as the finite dimensional case, gives the following.

2.5 **COROLLARY.** Each vector space over a field has a basis.

As an example of this result consider the reals \( \mathbb{R} \) as a vector space over the rationals \( \mathbb{Q} \). Then \( \mathbb{R} \) has a basis over \( \mathbb{Q} \). This is known as a Hamel basis. A consequence of this is that some rather simple looking functions on \( \mathbb{R} \) have a rather weird behaviour. This is discussed in Problem 2.7

As mentioned earlier in the Preamble, K. Kuratowski proved a version of ZL in 1922 before the idea became widely known. The following is sometimes known as KL, Kuratowski’s Lemma

2.6 **LEMMA.** For each poset \( S \), each chain of \( S \) is included in a maximal chain.

We can now sort out the relationship between the four variant of Zorn’s Lemma stated in Definition 2.1.

2.7 **THEOREM.** The four properties

\[
\text{ZL}^+ \quad \text{ZL} \quad \text{YL} \quad \text{XL}
\]

of the universe of sets are equivalent.

**Proof.** By the observations at the beginning of this section it suffices to show that

\[ \text{XL} \implies \text{ZL}^+ \]

holds. To this end consider any poset \( S \) which has \( Z(S) \). Consider also any element \( a \in S \). In due course we invoke XL to produce a maximal element of \( S \) above \( a \).

Let \( \mathcal{A} \) be the poset of all chains of \( S \) which have \( a \) as the first element. In particular, \( \{a\} \) is one such chain, so that \( \mathcal{A} \) is non-empty. By Exercise 1.5 we know we have \( Y(\mathcal{A}) \). Thus an application of XL to \( \mathcal{A} \) gives \( L(\mathcal{A}) \), that is there is a maximal chain \( \mu \) of \( S \) with \( a \) as its first element. But \( S \) has \( Z(S) \), so that is some element \( b \in S \) with \( \mu \leq b \). It suffices to show that \( b \) is a maximal element of \( S \).

The chain \( \mu \cup \{b\} \) extends \( \mu \). Since \( \mu \) is a maximal chain this ensures that \( b \) is the top of \( \mu \). Now consider any element \( c \) with \( b \leq c \). The same observation shows that \( b = c \), and hence \( b \) is a maximal element of \( S \). \( \blacksquare \)

Finally, for this section, we can indicate the strength of ZL. We can show that

\[ \text{ZL} \implies \text{AC} \]
holds. Rather, we can obtain enough in formation to prove this implication. The actual proof is your job.

We need to set up a bit of gadgetry.

We work with two non-empty sets $A, B$ together with a subset $R \subseteq B \times A$ such that $(\forall b \in B)(\exists a \in A)[bRa]$
holds. Thus $R$ is a special kind of relation between $A$ and $B$.

2.8 LEMMA. For the situation described above a use of XL provides a function

$g : B \longrightarrow A$ such that $bRg(b)$

for all $b \in B$.

**Proof.** We consider the partial function from $B$ to $A$ that do this kind of job. Thus we consider those functions

$g : Y \longrightarrow A$

where $Y \subseteq B$ and $yRg(y)$ for all $y \in Y$. Let $G$ be the set of all such functions.

The set $G$ is non-empty. In fact, for each $b \in B$ there is at least one such function with $b \in Y$. We may take $Y = \{b\}$ and that let $g(y)$ be any member of $A$ given by the property of $R$.

We partially order $G$ by extension. Thus for two members

$(Y, g) (Z, h)$

we let

$(Y, g) \leq (Z, h)$ mean $Y \subseteq Z$ and $g = h|_Y$

that is $g$ and $h$ agree on the smaller set $Y$.

We show that this poset $G$ is closed under directed suprema.

Consider any directed $D \subseteq G$. Thus for each $g_1, g_2 \in D$ there is some $g_3 \in D$ with $g_1 \leq g_3$ and $g_2 \leq g_3$. We produce a function $\bigvee D \in G$ which is the supremum of $D$ in $G$.

Technically, $D$ is a set of pairs $(Y, g)$ where $g$ is a function (of a certain kind) with source $Y$. Let

$Z = \bigcup \{Y \mid \text{There is some } (Y, g) \in D\}$

that is we take the union of all the sources of the functions in $D$. We show there is a function

$h : Z \longrightarrow A$

in $G$ with $g \leq h$ for each $g \in D$. In fact, the $h$ we produce is not just an upper bound for $D$ it is the least upper bound of $D$.

Consider any element $z \in Z$. We have $z \in Y$ for at least one $(Y, g) \in G$. We want to take $g(z)$ as the output $h(z)$, but we need to check that this doesn’t depend on which $(Y, g)$ we use.

Consider two $(Y_1, g_1), (Y_2, g_2) \in D$ with $z \in Y_1 \cap Y_2$. Since $D$ is directed there is some $(Y_3, g_3) \in D$ which extends both $g_1, g_2$. But now

$g_1(z) = g_3(z) = g_2(z)$
to show the independence of output from \( z \).

This shows there is a function \( h : Z \to A \) with

\[
h(z) = g(z) \text{ for any } (Y, g) \in D \text{ with } z \in Y
\]

so we show that this function is the required supremum of \( D \).

We must first check that \( (Z, h) \in G \), that is

\[
zRh(z)
\]

for each \( z \in Z \). This is more or less immediate since \( h \) is a conglomerate of the \( g \in D \).

Trivially, this \( h \) is an upper bound of \( D \).

This shows that \( G \) has \( Y(G) \), and hence a use of XL provides a maximal member \((Y, g)\) of \( G \). A simple observation now shows that \( Y = b \), and we are done.

There are one or two details missing from this proof, but they are not hard, and are left as an exercise.

2.1 Exercises: Zorn’s Lemma and some uses

2.1 Complete the proof of Lemma 2.2.

2.2 Sort out the details of Lemma 2.4 and Corollary 2.5.
   (a) Write down the definition of a subset of \( V \) being linearly independent.
   (b) Show that the union of a directed family of linearly independent sets is itself linearly independent.
   (c) Show that a maximal linearly independent is a basis of \( V \).

2.3 Show that for an arbitrary poset \( S \), each chain in \( S \) is included in a maximal chain.
   (This is Hausdorff’s Maximality Principle.)

2.4 (a) Using ZL, or a variant, prove Lemma 2.6 
   (b) Show that KL \( \implies \) ZL.

2.5 State and prove (using ZL) an analogue of ZL for the existence of minimal elements.

2.2 Problem: Zorn’s Lemma and some uses

2.6 Show that each partial order on a set can be extended to a linear order of the set.

2.7 Consider an arbitrary function

\[
f : \mathbb{R} \to \mathbb{R}
\]

which satisfies

\[
f(x + y) = f(x) + f(y)
\]

for arbitrary \( x, y \in \mathbb{R} \). Let \( c = f(1) \).
   (a) Show that the behaviour of \( f \) on \( \mathbb{Q} \) is uniquely determined by \( c \).
   (b) Show that if \( f \) is continuous then it is uniquely determined by \( c \).
   (c) Using a Hamel basis for \( \mathbb{R} \) over \( \mathbb{Q} \), show there are many more exotic functions \( f \) where each value of \( f \) is rational.
2.8 Let $U, T$ be a pair of vector spaces over the same field $K$. Recall that a function

$$f : U \longrightarrow T$$

is linear, a linear transformation, if

$$f(rx) = r(f(x)) \quad f(x + y) = f(x) + f(y)$$

for all $x, y \in U$ and $r \in K$.

Suppose $U \subseteq W$ are a pair of vector spaces over $K$. Show that each linear function $f : U \longrightarrow T$ extends to a linear function $h : W \longrightarrow T$ with the same range, that is $f[U] = h[W]$.

3 AC $\Rightarrow$ ZL

By Theorem 2.7 we know that the four properties

$$\text{ZL}^+ \quad \text{ZL} \quad \text{YL} \quad \text{XL}$$

so from now on we refer to all of them as ZL, Zorn’s Lemma. We don’t know about the minor differences between these variants. However, we don’t yet know that ZL is true. By Lemma 2.8 we know that

$$\text{ZL} \implies \text{AC}$$

so that ZL is as least as strong as the Axiom of Choice. In this section we show

$$\text{AC} \implies \text{ZL}$$

and hence ZL is equivalent to AC.

The proof is a bit long but most of the ideas are straight forward, and we don’t need any background information. In particular, we don’t need anything about ordinals. The proof is essentially that given on pages 63-65 of ‘Naive Set Theory’ by Paul Halmos. He says that the proof is a modified version of the one originally given by Zermelo.

The proof

We obtain the implication via a contradiction. Thus suppose $S$ is a poset, and consider the family $C$ of all chains $\alpha \subseteq S$. We reserve $\alpha, \beta, \gamma, \ldots$ to range over $C$. We assume the following hold.

(+) Each $\alpha \in C$ has an upper bound in $S$.

(−) The poset $S$ has no maximal element.

Using AC we derive a contradiction from these, and so verify ZL.

The argument is developed in six phases.
Phase 1

In this phase we use AC to produce a function

\[ f : C \rightarrow C \]

which converts each chain into a slightly longer chain. This is the only part where we use any choice.

We arrange that the function satisfies

1. \( f(\alpha) \in C \)
2. \( \alpha \subset f(\alpha) \)
3. \( f(\alpha) - \alpha \) is a singleton

for each \( \alpha \in C \).

Consider any \( \alpha \in C \). Let \( \alpha^* \subseteq S \) be given by

\[ a \in \alpha^* \iff a \in S - \alpha \text{ and } \alpha \subseteq \downarrow a \]

for each \( a \in S \). We show that \( \alpha^* \) is non-empty.

By (+) we have

\[ \alpha \subseteq \downarrow b \]

for at least one \( b \in S \). By (−) we have \( b < a \) for some \( a \in S \). This produces \( a \in \alpha^* \).

We have attached to each \( \alpha \in C \) a non-empty set \( \alpha^* \). By AC there is a function

\[ g : C \rightarrow S \]

with \( g(\alpha) \in \alpha^* \) for each \( \alpha \in C \). We have

\[ \alpha \subseteq \downarrow g(\alpha) \text{ and } g(\alpha) \notin \alpha \]

for each \( \alpha \in C \). Thus

\[ f(\alpha) = \alpha \cup \{g(\alpha)\} \]

is a strictly longer chain. This gives (1; i, ii, iii).

Phase 2

Next we produce a certain collection of chains. To help with this let us say a collection \( \mathcal{T} \subseteq C \) is a tower if the following hold.

1. \( \emptyset \in \mathcal{T} \)
2. \( \mathcal{T} \) is closed under \( f \), that is \( \alpha \in \mathcal{T} \implies f(\alpha) \in \mathcal{T} \).
3. If \( \mathcal{L} \subseteq \mathcal{T} \) is linearly ordered by inclusion then \( \bigcup \mathcal{L} \in \mathcal{T} \).
Notice that the $\mathcal{L}$ in part (2,iii) could be called a chain, but it is not a chain of elements of $S$, it is a chain of certain kinds of subsets. Thus to avoid confusion we won’t call it a chain.

There is at least one tower, for $\mathcal{C}$ itself is a tower. This is the largest possible tower. We want to go to the other extreme. Let $\mathcal{T}$ be the family of all towers. An easy calculation shows that

$$\bigcap \mathcal{T}$$

is also a tower. We now let $\mathcal{T}$ be this particular tower, so we have the following.

(2,iv) If $\mathcal{U}$ is a tower that $\mathcal{T} \subseteq \mathcal{U}$.

Our next job it to show that this particular tower $\mathcal{T}$ is linearly ordered by inclusion, so we may apply (2,iii) to get $\bigcup \mathcal{T} \in \mathcal{T}$, and this will lead to the contradiction.

Phase 3

Let us say $\gamma \in \mathcal{T}$ is comparable if the following holds.

(3,i) $(\forall \alpha \in \mathcal{T})[\gamma \subseteq \alpha$ or $\alpha \subseteq \gamma]$  

Trivially, $\emptyset$ is comparable. An easy calculation show that the union of any linear family of comparable sets is itself comparable. We want to show that the comparable sets form a tower and hence are the whole of $\mathcal{T}$. To do that we need to show that $f$ converts each comparable set into a comparable set.

This takes a bit of time.  
Fix one particular comparable set $\gamma$. We show that the following holds.

(3,ii) $(\forall \alpha \in \mathcal{T})[\alpha \subsetneq \gamma \implies f(\alpha) \subseteq \gamma]$  

To this end consider any $\alpha \in \mathcal{T}$ with $\alpha \subsetneq \gamma$. By (2,ii) we have $f(\alpha) \in \mathcal{T}$, so that

$$\gamma \subseteq f(\alpha) \text{ or } f(\alpha) \subseteq \gamma$$

by (3,i). The left hand case gives

$$\alpha \subsetneq \gamma \subseteq f(\alpha)$$

so that

$$\gamma = f(\alpha)$$

by (1,iii). Thus in both cases we have $f(\alpha) \subseteq \gamma$, as required.

Phase 4

As in Phase 3 we fix a comparable $\gamma \in \mathcal{T}$. Using $\gamma$ we consider the subfamily $\mathcal{D} \subseteq \mathcal{T}$ defined as follows.

$$\alpha \in \mathcal{D} \iff \alpha \in \mathcal{T} \text{ and } \left( \alpha \subseteq \gamma \text{ or } f(\gamma) \subseteq \alpha \right)$$

We wish to show that $\mathcal{D}$ is a tower, and hence $\mathcal{D} = \mathcal{T}$ by (2,iv), the minimality of $\mathcal{T}$.

Trivially, we have $\emptyset \in \mathcal{D}$.  

12
Next we show that
\[ \alpha \in D \implies f(\alpha) \in D \]
holds. To this end consider any \( \alpha \in D \). By the construction of \( D \) one of
\[ \alpha \subseteq \gamma \quad \alpha = \gamma \quad f(\gamma) \subseteq \alpha \]
holds. In turn these give
\[ f(\alpha) \subseteq \gamma \quad f(\alpha) = f(\gamma) \quad f(\gamma) \subseteq \alpha \subseteq f(\alpha) \]
using (3,ii) for the first case, and (1,ii) for the third case. These combine to give \( f(\alpha) \in D \).

Finally, consider any \( L \subseteq D \) which is linearly ordered by inclusion. We certainly have \( \bigcup L \in T \), since \( T \) is a tower. If \( \alpha \subseteq \gamma \) for each \( \alpha \in L \), then \( \bigcup L \subseteq \gamma \). Otherwise, there is some \( \alpha \in L \) with \( \alpha \not\subseteq \gamma \), and then
\[ f(\gamma) \subseteq \alpha \subseteq \bigcup L \]
to verify the third required condition.

Phase 5

We can now show that
\[ \gamma \text{ is comparable } \implies f(\gamma) \text{ is comparable} \]
for members \( \gamma \in T \).

To this end consider any comparable \( \gamma \). As in Phases 3 and 4 we have an associated subfamily \( D \subseteq T \). By Phase 4 we have \( D = T \). Thus the definition of \( D \) rephrases as follows.
\[
(\forall \alpha \in T)[\alpha \subseteq \gamma \text{ or } f(\gamma) \subset \alpha]
\]
This gives
\[
(\forall \alpha \in T)[f(\gamma) \subset \alpha \text{ or } \alpha \subseteq \gamma \subseteq f(\gamma)]
\]
which, by comparison with (3,i), show that \( f(\gamma) \) is comparable.

Phase 6

Trivially, \( \emptyset \) is comparable. An easy calculation shows that the union of a linear family of comparable sets is comparable. Thus, by the result of Phase 5, we see that the comparable sets form a tower, and hence each member of \( T \) is comparable.

For each pair \( \alpha, \gamma \in T \) we have \( \gamma \subseteq \alpha \) or \( \alpha \subseteq \gamma \), since \( \gamma \) is comparable (and so is \( \alpha \)). Thus \( T \) is linearly ordered by inclusion. This gives
\[ T = \bigcup T \in T \]
by (2,iii). But now (2,ii) gives
\[ f(T) \in T \text{ so that } f(T) \subseteq T \]
and then (2,ii) gives
\[ T \subsetneq f(T) \subseteq T \]
which is the contradiction.
4 The Krull Separation Lemma

In this section we look at a consequence of Zorn’s Lemma which is suitable for applications to rings. We deal mostly with commutative rings, but some of the problems look at non-commutative versions.

Recall that a ring is a structure

\[(R, +, 0, \times, 1)\]

with the usual properties (which you should write down). Usually, for \(x, y \in R\) we write

\[xy\]

for \(x \times y\)

so that ‘\(\times\)’ doesn’t clutter up various expressions. The KSL (Krull Separation Lemma) is a technique for producing prime ideals in \(R\). We need to recall some standard material concerning such ideals.

4.1 DEFINITION. Let \(R\) be a commutative ring.

An ideal of \(R\) is a non-empty subset \(I\) such that

\[x, y \in I \implies x \pm y \in Y \quad x \in I \implies lxr \in I\]

for all \(x, y, l, r \in I\).

Both

\[O = \{0\} \quad R\]

are ideals, the trivial ideal and the improper ideal.

An ideal \(M\) is maximal if it is maximal in the poset of proper ideals under inclusion. An ideal \(P\) is prime if it is proper and

\[ab \in P \implies a \in P\ or b \in P\]

for all \(a, b \in R\).

Observe that because the ring is commutative the definition of an ideal could be slightly simplified. However, often the version given is more useful.

You should also recall that when \(R\) need not be commutative the definition of a prime ideal is not correct.

Each element \(a \in R\) generates a principal ideal

\[\langle a \rangle\]

the smallest ideal that contains \(a\). This is the set of all finite sums

\[\sum l a r\]

for varying \(l, r \in R\). Again, when \(R\) is commutative this can be slightly simplified, but often this version is more useful.

More generally, given an ideal \(I\) and an element \(a \in R\) we may generate the smallest ideal that includes \(I\) and contains \(a\).

\[I + \langle a \rangle\]

This is the set of all finite sums

\[x + \sum l a r\]

for varying \(x \in I\) and \(l, r \in R\). Of course, if \(a \in I\) then \(I + \langle a \rangle = I\).
4.2 LEMMA. For a commutative ring $R$ each maximal ideal $M$ is prime.

Proof. By way of contradiction suppose the maximal ideal $M$ is not prime. Then we have

$$ab \in M \quad a \notin M \quad b \notin M$$

for two elements $a, b$. Consider what happens in the two ideals

$$M + \langle a \rangle \quad M + \langle b \rangle$$

to obtain the contradiction.

The remainder of this proof, and that of many other proofs, is left as an exercise.

This is a useful result but we also need a more general method of producing prime ideals. For that we need another notion.

4.3 DEFINITION. Let $R$ be a commutative ring. A subset $X \subseteq R$ is \times-closed if $1 \in X$ and

$$x, y \in X \implies xy \in X$$

holds for all $x, y \in R$.

There is also a non-commutative version of this notion, and for that the definition is not quite correct.

4.4 EXAMPLE. Let $R$ be a commutative ring. For each element $a \in R$ and prime ideal $P$ of $R$ the sets

$$\{1\} \quad \{a^m \mid m \in \mathbb{N}\} \quad R - P$$

are \times-closed.

By taking $X = \{1\}$ we see that Lemma 4.2 is a particular case of the following.

4.5 LEMMA. Let $R$ be a commutative ring and consider any \times-closed subset $X$. Consider the poset of those ideals which are disjoint from $X$. Each ideal $M$ which is maximal in this poset is prime.

Proof. By way of contradiction suppose the ideal $M$ is not prime. Then we have

$$ab \in M \quad a \notin M \quad b \notin M$$

for two elements $a, b$. Consider what happens in the two ideals

$$M + \langle a \rangle \quad M + \langle b \rangle$$

each of which must meet $X$. This leads to the contradiction.

With this we can state, and prove, the KSL.

4.6 LEMMA. (The Krull Separation Lemma) Let $R$ be a commutative ring and consider any \times-closed subset $X$. Then each ideal which is disjoint from $X$ is included in a prime ideal which is disjoint from $X$.  

Proof. For convenience let \( I \) be the poset of ideals which are disjoint from \( X \). Since \( O \in I \) we know that \( I \) is non-empty. A simple calculation shows that \( I \) is closed under directed unions. Thus, by ZL, there each member of \( I \) is included in a maximal member of \( I \). By Lemma 4.5 each such maximal member is prime.

Of course, it is nice to see an example of this technique in use.

4.7 EXAMPLE. Let \( R \) be a commutative ring, and consider any ideal \( I \) of \( R \), and it does no harm to assume that this is proper. We set

\[
\sqrt{I} = \{ a \in R \mid (\exists m \in \mathbb{N})[a^m \in I]\}
\]

to obtain the radical (the prime radical) of \( I \). It turns out that \( \sqrt{I} \) is an ideal. More generally, we find that

\[
\sqrt{I} = \bigcap \{ P \mid I \subseteq P \text{ where } P \text{ is prime}\}
\]

hence the name.

There are other properties of prime ideal which are not obtained using KSP, but by ZL. One of these is given in the exercises. In the problems it is shown how some of these results can be extended to non-commutative rings.

4.1 Exercises: The Krull Separation Lemma

4.1 (a) Complete the proof of Lemma 4.2, that is for a commutative ring \( R \) show that each maximal ideal is prime.

(b) Complete the proof of Lemma 4.5, that is for a commutative ring \( R \) and a \( \times \)-closed subset \( X \), show that each ideal that is maximally disjoint from \( X \) is prime.

4.2 (a) Complete the proof of Lemma 4.6, that is show that the poset \( I \) of ideals is closed under directed unions.

(b) Show that each proper ideal is included in a maximal ideal.

4.3 Complete the details of Example 4.7. In other words, prove the following for a commutative ring.

(a) For each ideal \( I \) the set \( \sqrt{I} \) is an ideal.

(b) In fact

\[
\sqrt{I} = \bigcap \{ P \mid I \subseteq P \text{ where } P \text{ is prime}\}
\]

holds.

(c) The set of nilpotent elements of \( R \) is an ideal, and is just the intersection of all prime ideals of \( R \). (An element \( a \) is nilpotent if \( a^m = 0 \) for some \( m \in \mathbb{N} \).)

4.4 Let \( R \) be a commutative ring, and let \( Z \) be its set of zeros divisors, the set of all those elements \( z \) with \( zw = 0 \) for some \( w \neq 0 \).

(a) Show that the intersection of any chain of prime ideals is a prime ideal.

(b) Can you see a slight generalization of (a)?

(c) Show that each prime ideal includes a minimal prime ideal.
(d) For an arbitrary prime ideal $P$, show that

$$X = \{ xu \mid x \in R - Z, u \in R - P \}$$

is $\times$-closed.

(e) For an arbitrary prime ideal $P$, show that $Z \cap P$ includes a minimal prime ideal.

(f) Show that $Z$ includes all minimal prime ideals.

(g) What does this show when $R$ is an integral domain?

4.2 Problems: The Krull Separation Lemma

In the following problems a ring $R$ need not be commutative.

An ideal $P$ is prime if it is proper and

$$aRb \subseteq P \implies a \in P \text{ or } b \in P$$

where

$$aRb = \{arb \mid r \in R \}$$

is the set included in $P$.

A set $X \subseteq R$ is an $m$-set if $1 \in X$ and

$$x, y \in X \implies (\exists r \in R)[xry \in X]$$

holds.

4.5 Prove the analogue of Lemma 4.5, namely that if $X$ is $m$-closed and $M$ is an ideal which is maximally disjoint from $X$, then $M$ is prime.

4.6 Prove the more general Krull Separation Lemma, namely that if $X$ is $m$-closed the each ideal which is disjoint from $X$ is included in a prime which is disjoint from $X$.

4.7 Let $R$ be an arbitrary ring. We say an ideal $E$ of $R$ is essential in $R$ if

$$E \cap J = O \implies J = O$$

for each ideal $J$. (This ideal $E$ could be the whole of $R$.)

Show that for each proper ideal $I$ of $R$ there is a prime ideal $P$ such that the ideal $I + P$ is essential in $R$. 

17