Math 7411 1. Logic Spring 2008

I will assume basic knowledge of logic. I will use the following notation, where P and Q are any statements.

$\neg P$	not P	the statement P does not hold,
$P \wedge Q$	P and Q	both P and Q hold,
$P \vee Q$	P or Q	either P or Q holds (or both),
$P \Rightarrow Q$	P implies Q	if P holds then Q holds (true unless P holds but Q does not),
$P \Leftrightarrow Q$	P iff Q	P holds if and only if Q holds.

Note that there are many relationships between these, for example

$$\begin{aligned}
P \Rightarrow Q &\text{ is equivalent to } (\neg P) \vee Q, \quad \text{or} \quad \neg(P \wedge \neg Q), \quad \text{or} \quad (\neg Q) \Rightarrow (\neg P), \\
P \Leftrightarrow Q &\text{ is equivalent to } (P \Rightarrow Q) \wedge (Q \Rightarrow P), \\
\neg(P \vee Q) &\text{ is equivalent to } (\neg P) \wedge (\neg Q), \\
\neg(P \wedge Q) &\text{ is equivalent to } (\neg P) \vee (\neg Q).
\end{aligned}$$

If $P(x)$ is a statement involving a variable x, then I will use the following notation for the quantifiers.

$\forall x\colon P(x)$	the statement $P(x)$ holds for all values of x,
$\exists x\colon P(x)$	there exists an x for which $P(x)$ holds.

Sometimes we shall encounter statements with several quantifiers in a row, e.g.,

$$\forall x\colon \exists y\colon \forall z\colon \exists t\colon zt < x + y$$

(assume variables are real numbers). It sometimes helps to think of such statements as a game between yourself and an opponent. You wish to make the statement true, but your opponent is trying to make it false. You go through each quantifier in turn. When you see an $\exists$, you get to choose the value of the variable, and when you see a $\forall$, your opponent gets to choose. You are both allowed to base your choices on previously chosen values. The truth of the statement is determined by who wins (assuming you play optimally). For example, you can win the above game by choosing $y > -x$ and $t = 0$, so the statement is true. The order of (different types of) quantifiers matters. For example $\forall x\colon \exists y\colon y > x$ is true but $\exists y\colon \forall x\colon y > x$ is false.

When negating a statement involving quantifiers, one interchanges $\exists$s and $\forall$s. For example

$$\neg\forall x\colon \exists y\colon \forall z\colon P(x,y,z) \quad \text{is equivalent to} \quad \exists x\colon \forall y\colon \exists z\colon \neg P(x,y,z),$$

(you and your opponent have changed rôles). I will often use abbreviations such as

$$\begin{aligned}
&\forall x > 0\colon P(x) && \text{for} \quad \forall x\colon (x > 0) \Rightarrow P(x), \\
&\exists x > 0\colon P(x) && \text{for} \quad \exists x\colon (x > 0) \wedge P(x), \\
&\exists x, y, z\colon \text{ or } \forall x, y, z\colon && \text{for} \quad \exists x\colon \exists y\colon \exists z\colon \text{ or } \forall x\colon \forall y\colon \forall z\colon .
\end{aligned}$$

Exercise: Assuming all variables are real numbers. Which of the following are true?

1. $\forall x\colon \forall \varepsilon > 0\colon \exists \delta > 0\colon |(x+\delta)^2 - x^2| < \varepsilon$.
2. $\forall \varepsilon > 0\colon \exists \delta > 0\colon \forall x\colon |(x+\delta)^2 - x^2| < \varepsilon$.

Math 7411 2. Set Theory Spring 2008

In set theory, one postulates the existence of 'sets' or 'collections' which can contain objects (or other sets) as members. We write $x \in A$ if x is a member of the set A. Key to the definition of a set is the

A1. Extensionality axiom: $\forall A, B\colon A = B \Leftrightarrow (\forall x\colon x \in A \Leftrightarrow x \in B)$.
Two sets are equal if and only if they contain the same elements.

For example, the sets $\{1, 1, 2\}$, $\{1, 2\}$, and $\{2, 1\}$ are all the same set. We say A is a *subset* of B if every member of A is a member of B, i.e.,

$$A \subseteq B \qquad \text{iff} \qquad \forall x\colon x \in A \Rightarrow x \in B.$$

The axiom A1 can now be rephrased as

$$A = B \qquad \text{iff} \qquad A \subseteq B \text{ and } B \subseteq A.$$

When trying to prove two sets A and B are the same, one often proves the two statements $A \subseteq B$ and $B \subseteq A$ separately, and then applies A1.

In naïve descriptions of set theory, one describes sets by listing elements (e.g., $\{1, 2\}$), or by specifying a set by a property that the elements must satisfy (e.g., $\{x \mid x \text{ is irrational}\}$). Unfortunately, one can get into trouble with this approach. For example, consider the set

$$S = \{x \mid x \notin x\}.$$

We get a contradiction since $S \in S \Leftrightarrow S \notin S$. (This example is known as as Russell's Paradox, and was discovered in 1901 by Bertrand Russell). To avoid this, we insist that sets are 'built up' using certain allowed constructions, and sets of the form $\{x \mid P(x) \text{ holds}\}$ are only allowed if x is restricted in some way. The most common version of set theory used is called *Zermelo-Fraenkel* or ZFC set theory, and, as far as we know, does not cause the contradictions that plague more simpleminded approaches. (ZFC = ZF + the Axiom of Choice, where ZF is a modification by Abraham Fraenkel of an axiomatic set theory originally developed by Ernst Zermelo in 1908.) In ZFC we can combine existing sets using

A2. Pair set axiom: $\forall x, y\colon \exists A\colon \forall z\colon z \in A \Leftrightarrow (z = x \vee z = y)$.
Given any x and y, we can construct the set $\{x, y\}$.

A3. Union axiom: $\forall \mathcal{C}\colon \exists U\colon \forall z\colon z \in U \Leftrightarrow (\exists A\colon z \in A \wedge A \in \mathcal{C})$.
Given any collection of sets $\mathcal{C}$, we can construct the union $U = \bigcup \mathcal{C} = \bigcup_{A \in \mathcal{C}} A$.

These allow us to construct any explicit finite set. For example, to construct $\{1, 3, 7\}$ we use A2 to construct $\{1, 3\}$, $\{1, 7\}$, and then $\{\{1, 3\}, \{1, 7\}\}$. Now take the union using A3. We can also form both finite and infinite unions. For example, to construct

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\},$$

we first use A2 to get the set $\{A, B\}$ and then use A3 to take the union.

A4. Power set axiom: $\forall A\colon \exists \mathcal{P}\colon \forall B\colon B \in \mathcal{P} \Leftrightarrow B \subseteq A$.
Given any set A, we can construct $\mathcal{P}(A) = \{B \mid B \subseteq A\}$, the set of all subsets of A.

These axioms allow us to build big sets, but for more versatility we need to be able to modify and/or remove elements as well. In the following $P(x)$ is some statement involving x.

A5. Subset axiom: $\forall A\colon \exists B\colon \forall x\colon x \in B \Leftrightarrow (x \in A \wedge P(x))$.
Given any set A and property $P(x)$, we can construct the subset $\{x \in A \mid P(x)\}$.

Note that this is really an infinite collection of axioms, one for every property $P(x)$. From this axiom one can construct intersections, set differences, etc.,

$$A \cap B = \{x \in A \mid x \in B\}, \qquad A \triangle B = \{x \in A \cup B \mid x \notin A \cap B\},$$
$$A \setminus B = \{x \in A \mid x \notin B\}, \qquad \bigcap_{A\in\mathcal{C}} A = \{x \in \bigcup \mathcal{C} \mid \forall A \in \mathcal{C}\colon x \in A\} \quad (\mathcal{C} \neq \emptyset).$$

A6. Replacement: $\forall A\colon \exists B\colon \forall y\colon y \in B \Leftrightarrow (\exists x \in A\colon P(x,y) \wedge (\forall z\colon P(x,z) \Rightarrow z = y))$.
Given any set A and any 'function' $F(x)$ we can construct the set $\{F(x) \mid x \in A\}$.

This one needs a bit of explaining. The 'function' $y = F(x)$ is really just a property $P(x,y)$ that is satisfied for (at most) one value of y when $x \in A$. This axiom, for example, allows us to convert, for example, $\{1,2,3\}$ into $\{\{1,\{1\}\},\{2,\{2\}\},\{3,\{3\}\}\}$ using $F(x) = \{x,\{x\}\}$. By allowing the function to be undefined for some $x \in A$, we also recover A5 as a special case of A6 (let $P(x,y)$ be $(x = y) \wedge P(x)$).

So far, we have only been able to construct sets from existing sets. So how do we get started? We want an axiom of the form 'there exists a set'. The simplest set to use would be the empty set.

A7. Empty set Axiom: $\exists\emptyset\colon \forall x\colon x \notin \emptyset$.
A set $\emptyset$ exists that does not contain any elements.

From the axioms so far, one can construct only finite sets, but we want infinite sets as well.

A8. Axiom of Infinity: $\exists N\colon \emptyset \in N \wedge (\forall x \in N\colon x \cup \{x\} \in N)$.
There exists an infinite set.

There are many equivalent versions of A8. This one gives a nice construction of the natural numbers $\mathbb{N} = \{0,1,2,\dots\}$. To see this, define the non-negative integers by $0 = \emptyset$ and $n = \{0,1,2,\dots,n-1\}$. Then A8 implies N contains $0 = \emptyset$, $1 = \{0\} = \{\emptyset\}$, $2 = 1 \cup \{1\} = \{0,1\} = \{\emptyset,\{\emptyset\}\}$, $3 = 2 \cup \{2\} = \{0,1,2\} = \{\emptyset,\{\emptyset\},\{\emptyset,\{\emptyset\}\}\}$, $\dots$
Note that A8+A5 implies A7 by taking $P(x) =$ False. Also the N of A8 may contain more elements than $\mathbb{N}$. To get $\mathbb{N}$ one needs to take the intersection of all subsets of N satisfying the condition in A8, which means we need A4 and A5 as well as A8.

The next axiom allows induction over sets, and will not be needed in this course.

A9. Axiom of Foundation: $(\forall x\colon (\forall y \in x\colon P(y)) \Rightarrow P(x)) \Rightarrow (\forall x\colon P(x))$.
Equivalent, but cryptic version: $\forall A \neq \emptyset\colon \exists x \in A\colon x \cap A = \emptyset$.
If $P(y)$ for all $y \in x$ implies $P(x)$ then $P(x)$ always holds.

The last axiom is the Axiom of Choice, which I will describe later.

Math 7411 3. Functions Spring 2008

Ordered pairs

An *ordered pair* is an object (a, b) consisting of two elements a and b, in such a way that order matters, i.e., $(a, b) = (c, d)$ iff $a = c$ and $b = d$, so that in general $(a, b) \neq (b, a)$. There are a number of possible constructions of ordered pairs in terms of sets. For example, one could define $(a, b) = \{\{a\}, \{a, b\}\}$ (check that this works!). Ordered triples (a, b, c) or n-tuples $(a_1, \ldots, a_n)$ can also be defined, (for example, in terms of ordered pairs by $(a_1, \ldots, a_n) = \{(1, a_1), \ldots, (n, a_n)\}$).

Functions

A *function* (or *map*) $f\colon X \to Y$ assigns for each $x \in X$ a value $f(x) \in Y$. A function can be defined more precisely as an ordered triple (X, Y, Γ) consisting of the *domain* X, the *co-domain* Y, and the *graph* Γ of f. The graph is a set of ordered pairs (x, y) with the property that for each $x \in X$ there is precisely one value of $y \in Y$ such that $(x, y) \in \Gamma$. This y is then written as $f(x)$. Note that we consider functions with distinct co-domains as different functions, even if they have the same values, e.g., $f\colon \mathbb{R} \to \mathbb{R}$; $f(x) = x^2$, is not the same as $g\colon \mathbb{R} \to [0, \infty)$; $g(x) = x^2$. The *image* of a map $f\colon X \to Y$ is the set of values taken, $f[X] = \{f(x) \mid x \in X\} = \{y \in Y \mid \exists x \in X\colon f(x) = y\} \subseteq Y$.

In order to define a function it is vitally important to check that

- The function is *defined*, i.e., produces a value *in* Y for *every* $x \in X$.
- The function is *well defined*, i.e., each $x \in X$ gives *only one value* $y \in Y$.

A *sequence* is a function a from $\mathbb{N}$ (or some subset of $\mathbb{N}$) to a set X, and we often write a_n instead of $a(n)$, and $(a_n)_{n\in\mathbb{N}}$ or $(a_n)_{n=0}^{\infty}$ instead of a. More generally, it is quite common to to think of a function $a\colon I \to X$ as a collection of elements $a_i = a(i)$ *indexed* by $i \in I$. If we have a collection of sets $A_i = A(i)$ indexed by $i \in I$, expressions such as $\bigcup_{i\in I} A_i$ are really just an alternative notation for $\bigcup\{A_i \mid i \in I\} = \bigcup A[I]$.

If $f\colon X \to Y$ and $g\colon Y \to Z$ are two maps then the *composition* $g \circ f$ (or gf) is the map $g \circ f\colon X \to Z$ defined by $g \circ f(x) = g(f(x))$.

For example, if $n\colon \mathbb{N} \to \mathbb{N}$ is a strictly increasing function ($i < j$ implies $n_i < n_j$), and $a\colon X \to \mathbb{N}$ is a sequence, the composition $a \circ n = (a_{n_i})_{i\in\mathbb{N}}$ is called a *subsequence* of $(a_i)_{i\in\mathbb{N}}$.

A map $f\colon X \to Y$ is *injective* (or *1-to-1*) if $f(x) = f(x')$ implies $x = x'$, i.e., each value $y \in Y$ occurs *at most* once as a value of f.
A map $f\colon X \to Y$ is *surjective* (or *onto*) if for every $y \in Y$ there is some x with $f(x) = y$, i.e., each value $y \in Y$ occurs *at least* once as a value of f.
A map $f\colon X \to Y$ is *bijective* if it is both injective and surjective, i.e., each value occurs *exactly* once, and so the function pairs up elements of X with elements of Y.

The map $f\colon X \to Y$ is a bijection if and only if it has a *two-sided inverse* $f^{-1}\colon Y \to X$, i.e., $f^{-1} \circ f = 1_X$, $f \circ f^{-1} = 1_Y$, where $1_X(x) = x$ and $1_Y(y) = y$ are the *identity* maps on

X and Y. If $A \subseteq Y$ then the *inclusion map* $i\colon A \to Y$ is defined by $i(x) = x$ for all $x \in A$. The *identity map* 1_X is just the inclusion $X \to X$. If $f\colon Y \to Z$ is any function then the *restriction* of f to A is $f_{|A} = f \circ i\colon A \to Z$, so $f_{|A}(x) = f(x)$ for all $x \in A$.

Given a function $f\colon X \to Y$, one can define the image $f[A]$ of $A \subseteq X$ and inverse image $f^{-1}[B]$ of $B \subseteq Y$ by

$$f[A] = \{f(x) \mid x \in A\} \subseteq Y, \qquad f^{-1}[B] = \{x \in X \mid f(x) \in B\} \subseteq X.$$

If f is bijective, $f^{-1}[B]$ could either be interpreted as an inverse image of f, or an image of the inverse f^{-1}. Fortunately, these sets are the same. The inverse image 'commutes' with all the set operations, but the image only commutes with unions in general:

$$\begin{array}{rcl@{\qquad}rcl}
f[\bigcup_i A_i] & = & \bigcup_i f[A_i], & f^{-1}[\bigcup_i B_i] & = & \bigcup_i f^{-1}[B_i], \\
f[\bigcap_i A_i] & \subseteq & \bigcap_i f[A_i], & f^{-1}[\bigcap_i B_i] & = & \bigcap_i f^{-1}[B_i], \\
f[X \setminus A] & ?? & Y \setminus f[A], & f^{-1}[Y \setminus B] & = & X \setminus f^{-1}[B], \\
f^{-1}[f[A]] & \supseteq & A & f[f^{-1}[B]] & \subseteq & B.
\end{array}$$

Cartesian Products

The *Cartesian Product* $X \times Y$ is the set of all ordered pairs (x, y) with $x \in X$ and $y \in Y$.

More generally if X_i are sets for all $i \in I$, then the Cartesian product $\prod_{i\in I} X_i$ can be defined as the set of functions $x\colon I \to \bigcup_{i\in I} X_i$, with $x_i \in X_i$ for all i. This definition works for any set I including infinite or even uncountable sets. If $I = \{1, \ldots, n\}$ we can identify this with the set of all ordered n-tuples $(x_1, \ldots, x_n)$ with $x_i \in X_i$.

A function of two (or more) variables can be represented as a function from a Cartesian product. E.g., if a function $f(x, y)$ takes values $x \in X$ and $y \in Y$ and produces $f(x, y) \in Z$, then f can be considered as a map $f\colon X \times Y \to Z$ with $f(x, y)$ a shorthand for $f((x, y))$.

The **Axiom of Choice** states that if $X_i \neq \emptyset$ for all $i \in I$ then $\prod_{i\in I} X_i \neq \emptyset$. In other words, there is a function $(x_i)_{i\in I}$ with $x_i \in X_i$ for all $i \in I$. This function simultaneously 'chooses' an element x_i from each X_i.

Exercises

1. Show that the following are equivalent
   (a) $f\colon X \to Y$ is injective,
   (b) For all Z and all maps $g, h\colon Z \to X$, $f \circ g = f \circ h$ implies $g = h$.
2. Show that the following are equivalent
   (a) $f\colon X \to Y$ is surjective,
   (b) For all Z and all maps $g, h\colon Y \to Z$, $g \circ f = h \circ f$ implies $g = h$.
3. Show that if $f\colon X \to Y$ is injective and $X \neq \emptyset$ then f has a *left inverse* g, $g \circ f = 1_X$.
4. Show that if $f\colon X \to Y$ is surjective then f has a *right inverse* g, $f \circ g = 1_Y$.
   [You will need the Axiom of Choice.]

Math 7411 4. Zorn's Lemma Spring 2008

A *binary relation* on X is a subset $R \subseteq X \times X$. We write xRy (or say that xRy *holds*) iff $(x, y) \in R$. For example we can define '$<$' on $X = \{1, 2, 3\}$ as the set $\{(1, 2), (1, 3), (2, 3)\}$.

A relation $\leq$ is called a *partial ordering* if for all x, y, z,

O1. $x \leq x$ (reflexive),

O2. if $x \leq y$ and $y \leq x$ then $x = y$ (antisymmetric),

O3. if $x \leq y$ and $y \leq z$ then $x \leq z$ (transitive).

For any partial ordering, we can define $\geq$, $<$, $>$, by e.g., $x < y$ iff $x \leq y$ and $x \neq y$.
Note that it is possible for two elements to be *incomparable*, i.e., $x \not\leq y$ and $y \not\leq x$.
A relation $\leq$ is called a *total* or *linear ordering* if it also satisfies

O4. either $x \leq y$ or $y \leq x$ (trichotomy).

For example, $\subseteq$ is a partial order on $\mathcal{P}(X)$, the collection of subsets of X, but is not in general a total order. On the other hand, the usual $\leq$ on the real numbers is a total order.

If $(\mathcal{X}, \leq)$ is a partially ordered set, a *chain* in $\mathcal{X}$ is a non-empty subset $\mathcal{T} \subseteq \mathcal{X}$ that is totally ordered by $\leq$.

If $\mathcal{T} \subseteq \mathcal{X}$, and $x \in \mathcal{X}$, we say x is an *upper bound* for $\mathcal{T}$ if $y \leq x$ for all $y \in \mathcal{T}$.
Note: We do **not** require x to be an element of $\mathcal{T}$.

A *maximal element* of $\mathcal{X}$ is an element x such that for any $y \in \mathcal{X}$, $x \leq y$ implies $x = y$.
Note: This does not imply that $y \leq x$ for all $y \in \mathcal{X}$ since $\leq$ is only a partial order. In particular there may be many maximal elements.

Theorem (Zorn's Lemma) *If $(\mathcal{X}, \leq)$ is a non-empty partially ordered set for which every chain has an upper bound, then $\mathcal{X}$ has a maximal element.*

Note: If we had defined things so that $\emptyset$ were a chain, we would not need the condition that $\mathcal{X} \neq \emptyset$ in Zorn's Lemma, since the existence of an upper bound for $\emptyset$ is just the condition that an element of $\mathcal{X}$ exists. However, in practice it is usually easier to check that each *non-empty* totally ordered subset has an upper bound and then check $\mathcal{X} \neq \emptyset$. If $\mathcal{X}$ is finite, then $\mathcal{X}$ always has a maximal element by induction on the size of $\mathcal{X}$, so Zorn's lemma is mainly of interest when $\mathcal{X}$ is infinite.

We now give an application of Zorn's Lemma to linear algebra.

Let V be a vector space (possibly infinite dimensional). Recall that $S \subseteq V$ is called *linearly independent* if there are no non-trivial *finite* linear combinations that give 0, i.e., if $\sum_{i=1}^{n} \lambda_i s_i = 0$ and the $s_i \in S$ are distinct, then $\lambda_i = 0$ for each i. A set $S \subseteq V$ is called *spanning* if every $v \in V$ can be written as a *finite* linear combinations of elements of S, $v = \sum_{i=1}^{n} \lambda_i s_i$. A set $S \subseteq V$ is called a *basis* if it is a linearly independent spanning set. Note that every element $v \in V$ can be written as a linear combination of elements of a basis in a unique way. (Spanning implies existence, linear independence implies uniqueness.)

Theorem *Every vector space has a basis.*

Proof. Let $\mathcal{X}$ be the set of all linearly independent sets in V partially ordered by $\subseteq$. Since $\emptyset$ is linearly independent, $\mathcal{X} \neq \emptyset$. Let $\mathcal{T}$ be a chain in $\mathcal{X}$ and let $S = \bigcup_{S_\alpha \in \mathcal{T}} S_\alpha$. We shall show that S is linearly independent.

Suppose $\sum_{i=1}^n \lambda_i s_i = 0$ and $s_i \in S_{\alpha_i} \in \mathcal{T}$ (the s_i are distinct but the α_i need not be). Then by total ordering of the S_{α_i}, there must be one S_{α_j} that contains all the others (by induction on n there is a maximal S_{α_j} in $\{S_{\alpha_1}, \ldots, S_{\alpha_n}\}$). But then $\sum_{i=1}^n \lambda_i s_i = 0$ is a linear relation in S_{α_j} which is linearly independent. Thus $\lambda_i = 0$ for all i. Hence S is linearly independent, so $S \in \mathcal{X}$ and is an upper bound for $\mathcal{T}$.

Now apply Zorn's Lemma to give a maximal linearly independent set M. We shall show that M spans V and so is a basis. Clearly any element of M is a linear combination of elements of M, so pick any $v \notin M$ and consider $M \cup \{v\}$. By maximality of M this cannot be linearly independent. Hence there is a linear combination $\lambda v + \sum_{i=1}^n \lambda_i s_i = 0$, $s_i \in M$, with not all the λ's zero. If $\lambda = 0$ this gives a linear relation in M, contradicting linear independence of M. Hence $\lambda \neq 0$ and $v = \sum_{i=1}^n (-\lambda_i/\lambda) s_i$ is a linear combination of elements of M. □

Another variant of Zorn's Lemma is

Theorem (Hausdorff's Maximal Principle) *If $(\mathcal{X}, \leq)$ is a partially ordered set, then $\mathcal{X}$ contains a maximal totally ordered subset $\mathcal{T}$, i.e., if $\mathcal{T}'$ is a totally ordered subset of $\mathcal{X}$ and $\mathcal{T} \subseteq \mathcal{T}'$ then $\mathcal{T} = \mathcal{T}'$.*

Both results are equivalent to the Axiom of Choice, which states that if X_i are non-empty sets then $\prod_{i \in I} X_i$ is non-empty. As examples of their use, I will prove HMP⇒ZL⇒AC.

HMP⇒ZL: Assume $(\mathcal{X}, \leq)$ is a non-empty partially ordered set for which every chain has an upper bound. Pick a maximal totally ordered subset $\mathcal{T}$. Since $\mathcal{X} \neq \emptyset$ and singleton subsets of $\mathcal{X}$ are totally ordered, $\mathcal{T} \neq \emptyset$. Thus $\mathcal{T}$ has an upper bound $x \in \mathcal{X}$. If $x < y$ for some $y \in \mathcal{X}$ then $y \notin \mathcal{T}$, so $\mathcal{T} \cup \{y\} \supsetneq \mathcal{T}$. Also $z \in \mathcal{T}$ implies $z \leq x < y$, so $\mathcal{T} \cup \{y\}$ is totally ordered, a contradiction. Hence x is a maximal element. □

ZL⇒AC: Assume $X_i \neq \emptyset$, $i \in I$, and let $\mathcal{X}$ be the set of partial functions $(x_i)_{i \in J}$, where $x_i \in X_i$ for all $i \in J$ and $J \subseteq I$. We order them by declaring $(x_i)_{i \in J} \leq (x'_i)_{i \in J'}$ provided $J \subseteq J'$ and $x_i = x'_i$ for all $i \in J$. (Check this is a partial order.) If we have a chain $\mathcal{T} = \{(x_i^{(k)})_{i \in J^{(k)}}\}$, we can construct an upper bound by setting $J = \bigcup J^{(k)}$ and $x_i = x_i^{(k)}$ whenever $i \in J^{(k)}$. This is well defined, since if $i \in J^{(k)}$ for several k, then the definitions for x_i agree by the total ordering of $\mathcal{T}$. It is clear that $(x_i)_{i \in J}$ is an upper bound for $\mathcal{T}$. The empty function (with $J = \emptyset$) lies in $\mathcal{X}$, so $\mathcal{X} \neq \emptyset$. Hence by Zorn we can find a maximal $(x_i)_{i \in J}$. If $J \neq I$, pick an $i_0 \in I \setminus J$ and an $x_0 \in X_{i_0}$ and define a new function to be x_i when $i \in J$ and x_0 when $i = i_0$. This lies in $\mathcal{X}$ and is larger than $(x_i)_{i \in J}$, contradicting the maximality of $(x_i)_{i \in J}$. Hence $J = I$ and $(x_i)_{i \in J} \in \prod_i X_i$. □

Exercise: Show ZL ⇒ HMP. [Hint: partially order the chains by inclusion.]

Math 7411 5. Well ordering Spring 2008

A total ordering $\leq$ on X is called a *well ordering* (or X is *well ordered* by $\leq$) if

WO. Every non-empty subset $S \subseteq X$ contains a smallest element,
i.e, $\exists x_0 \in S \colon \forall x \in S \colon x_0 \leq x$.

We shall write this smallest element as $\min(S)$. In fact, it is enough to assume $\leq$ is a partial ordering since WO implies trichotomy (take $S = \{x, y\}$).

Examples

1. The usual $\leq$ on $\mathbb{N}$ is a well ordering, but $\leq$ on the reals is not (take $S = \{x \mid x > 0\}$).
2. The set $\mathbb{N} \cup \{\omega\} = \{0, 1, \ldots, \omega\}$ with $n < \omega$ for all $n \in \mathbb{N}$ is a well ordered set.
3. Define the *lexicographical* (*dictionary*) ordering on $\mathbb{N} \times \mathbb{N}$ by $(a, b) \leq (c, d)$ iff either $a < c$, or ($a = c$ and $b \leq d$). Then this is a well ordering of $\mathbb{N} \times \mathbb{N}$.

Induction

Well ordering allows us to use *induction*. Recall that there are two forms of induction on $\mathbb{N}$: *weak* induction, where we use $P(0)$ and $P(n) \Rightarrow P(n+1)$ to deduce $\forall n \colon P(n)$, and *strong* induction, where we use $(\forall m < n \colon P(m)) \Rightarrow P(n)$ to deduce $\forall n \colon P(n)$. Strong induction is in fact more generally applicable than weak induction since weak induction requires every element (except 0) to be of the form $n + 1$. Strong induction works in any well ordered set, without requiring the idea of a 'previous' element.

Theorem (Transfinite induction) *Assume $\leq$ is a well-ordering on X, and $P(x)$ is a property. If from $\forall y < x \colon P(y)$ one can deduce $P(x)$, then $P(x)$ holds for all $x \in X$.*

Proof. Consider $\min(\{x \in X \mid P(x) \text{ fails}\})$. □

It appears that we don't have a 'base' for the induction, but this is an illusion. If $x_0 = \min(X)$ then the statement $\forall y < x_0 \colon P(y)$ is vacuously true, and so $P(x_0)$ needs to be proved without any assumptions.

Note that Example 2 above shows that that weak induction does not in general hold for well-ordered sets (take $P(x)$ to be '$x < \omega$'). Example 3 often appears in 'nested' induction arguments on two (or more) variables.

The following result is useful since it allows induction in any set. It is equivalent to the Axiom of Choice.

Theorem (Well Ordering Principle) *Every set can be well ordered.*

Recursion

One can also use well ordering to produce *recursively defined functions*. To illustrate this, lets start with recursively defined functions on $\mathbb{N}$.

Theorem *Assume $g\colon Y \to Y$ is any function, and $c \in Y$ is specified. Then there exists a unique sequence $(a_n)_{n=0}^{\infty}$ of elements of Y such that $a_0 = c$ and $a_{n+1} = g(a_n)$.*

Proof. First show by induction that for every N, there is a unique finite sequence $(a_i^{(N)})_{i=0}^{N}$ such that $a_0^{(N)} = c$ and for all $n < N$, $a_{n+1}^{(N)} = g(a_n^{(N)})$. Clearly this holds for $N = 0$, and if true for N, then we can define $a_n^{(N+1)} = a_n^{(N)}$ for $n \leq N$ and $a_{N+1}^{(N+1)} = g(a_N^{(N)})$. This sequence is unique: clear for $N = 0$, and for $N+1$ the restriction to $\{0, \ldots, N\}$ is uniquely determined by induction, and then so is the value at $N + 1$. Now any sequence $(a_i)_{i=0}^{\infty}$ satisfying the conditions gives $(a_i^{(N)})_{i=0}^{N}$ when restricted to $\{0, \ldots, N\}$, so $a_n = a_n^{(n)}$ for all n. Conversely, defining $(a_n)_{n=0}^{\infty}$ by $a_n = a_n^{(n)}$ works. □

Theorem (Transfinite recursion) *Suppose $\leq$ is a well-ordering of X, and assume $g\colon \mathcal{F} \to Y$, where $\mathcal{F}$ is the set of functions of the form $h\colon \{z \in X \mid z < x\} \to Y$ for some $x \in X$. Then there is a unique function $f\colon X \to Y$ such that for all $x \in X$, $f(x) = g(f_{|\{z|z<x\}})$.*

The proof is very similar to the proof for $\mathbb{N}$, only using strong induction wherever necessary.

Exercises

It is intuitively obvious that $\leq$ on $\mathbb{N}$ is a well ordering, but it is perhaps instructive to prove it using the construction of Section 2. For any set x, define the *successor* of x to be the set $x^+ = x \cup \{x\}$. Recall that the axiom of infinity gives us a set N such that (a) $\emptyset \in N$ and (b) $x \in N \Rightarrow x^+ \in N$. Define the natural numbers, $\mathbb{N}$, as the intersection of all subsets $S \subseteq N$ such that (a) $\emptyset \in S$ and (b) $x \in S \Rightarrow x^+ \in S$. We define *zero* to be $\emptyset$.

1. Prove the *Peano Axioms* of arithmetic:
   - P1. Zero is a natural number.
   - P2. Every natural number has a successor, which is also a natural number.
   - P3. Zero is not the successor of any natural number.
   - P4. Any two distinct natural numbers have distinct successors.
   - P5. If a set S contains zero, and for every $n \in S$, S contains the successor of n, then S contains every natural number.

   [Hints: For P5, show S satisfies (a) and (b), so is one of the sets in the intersection defining $\mathbb{N}$. For P4, apply P5 to $S = \{n \in \mathbb{N} \mid \forall m \in n^+\colon m \subseteq n\}$ and use this to show that if $m^+ = n^+$ then $m = n$.]
2. Show that if $m \in n$ then either $m^+ \in n$ or $m^+ = n$. [Hint: induction (i.e., P5) on n.]
3. Show that either $m \in n$ or $n \subseteq m$. [Hint: induction on m, using question 2.]
4. Show that $\subseteq$ is a well ordering of $\mathbb{N}$.
   [Hint: Using question 3, show the following property holds for all n:

   $$P(n) = \text{'For any } S \subseteq \mathbb{N}, \text{ if } S \cap n \neq \emptyset \text{ then } S \text{ has a smallest element'}.$$

   Now note that if $n \in S$ then $S \cap n^+ \neq \emptyset$.]

Math 7411 6. WOP⇔HMP⇔ZL⇔AC Spring 2008

Recall that we have proved HMP⇒ZL⇒AC. We shall now prove AC⇒WOP⇒HMP.

WOP⇒HMP: We are given an arbitrary partial ordering $\leq$ on $\mathcal{X}$, and, by the WOP, a well ordering $\leq_W$ on $\mathcal{X}$. Define a function $f\colon \mathcal{X} \to \mathcal{X} \cup \{\star\}$ recursively (using $\leq_W$):

$$f(x) = \begin{cases} x & \text{if } (f[\{z \mid z <_W x\}] \setminus \{\star\}) \cup \{x\} \text{ is totally ordered by } \leq; \\ \star & \text{otherwise.} \end{cases}$$

We now show that $\mathcal{T} = f[\mathcal{X}] \setminus \{\star\}$ is a maximal totally ordered subset of $\mathcal{X}$. If $\mathcal{T}$ is not totally ordered, then $\exists x, y \in \mathcal{T}$ with $x \not\leq y$ and $x \not\leq y$. W.l.o.g., $x <_W y$. Note that $x \in \mathcal{T}$ implies $f(x) = x$. But then $f(y) = \star$ and so $y \notin \mathcal{T}$. It is clearly maximal, since if any element x could be added to $\mathcal{T}$ then $f(x) = x$ would already be in $\mathcal{T}$. □

Note that WOP⇒AC is even easier: Let $\leq$ be a well ordering on $\bigcup X_i$. If $X_i \neq \emptyset$ for all $i \in I$, set $x_i = \min(X_i)$. Then $(x_i)_{i\in I} \in \prod_{i\in I} X_i$.

Definition If $\leq$ is a well ordering of X, an *initial segment* of X is a subset S of X such that if $x \in S$ and $y \leq x$ then $y \in S$. Any such set is either of the form $\{z \in X \mid z < x\}$, or is the set X itself. (Consider $\min(\{x \in X \mid x \notin S\})$.)

Definition An *order isomorphism* between two partially ordered sets $(A, \leq)$ and $(B, \leq')$ is a bijective function $f\colon A \to B$ such that $x \leq y$ iff $f(x) \leq' f(y)$.

Lemma 1 *Suppose $(A, \leq)$ and $(B, \leq')$ are two well ordered sets. Then there is an order isomorphism from one to an initial segment of the other.*

Proof. (sketch) Define a function $f\colon A \to B \cup \{\star\}$, $\leq$-recursively, so that

$$f(x) = \begin{cases} \min_{\leq'}(B \setminus f[\{z \mid z < x\}]) & B \subsetneq f[\{z \mid z < x\}]; \\ \star & \text{otherwise.} \end{cases}$$

If $\star \notin f[A]$, then check that f defines an order isomorphism from A to $f[A]$, which is an initial segment of B. If $\star \in f[A]$, then check that f defines an order isomorphism from B to $f^{-1}[B]$, which is an initial segment of A. □

AC⇒WOP: Assume X is a set. By AC there exists a function $f\colon (\mathcal{P}(X) \setminus \{X\}) \to X$ so that $f(A) \notin A$ for all $A \subsetneq X$. We wish to define $\leq$ so that the 'next' element x in the order is $f(\{z \mid z < x\})$.

Consider the set $\mathcal{C}$ of all pairs $(A, \leq)$ where $\leq$ is a well ordering on $A \subseteq X$ with the property that for all $x \in A$, $x = f(\{z \mid z < x\})$. Suppose $(A, \leq), (B, \leq') \in \mathcal{C}$. By Lemma 1, there exists an order preserving bijection h from, say, A to an initial segment of B. Now assume h is not an inclusion and pick the $\leq$-smallest x such that $h(x) \neq x$. But $h(x) = f(\{z \in B \mid z <' h(x)\}) = f(h[\{z \in A \mid z < x\}]) = f(\{z \in A \mid z < x\}) = x$, a contradiction. Hence $(A, \leq)$ is an initial segment of $(B, \leq')$.

Now consider the relation $\leq$ on $A = \bigcup A_i$ obtained by taking the union of all the well orderings $\leq_i$, where $(A_i, \leq_i) \in \mathcal{C}$. We shall show that $A = X$ and $\leq$ is a well ordering.
O1: If $x \in A$ then $x \in A_i$ for some i. Thus $x \leq_i x$, so $x \leq x$.
O2: Assume $x \leq y$ and $y \leq x$. Suppose $x \leq_i y$ and $y \leq_j x$. We may assume $(A_i, \leq_i)$ is an initial segment of $(A_j, \leq_j)$. Now $x \leq_j y$, so $x = y$.
O3: Assume $x \leq y \leq z$, and suppose $x \leq_i y \leq_j z$. If $(A_i, \leq_i)$ is an initial segment of $(A_j, \leq_j)$ then $x \leq_j y \leq_j z$, so $x \leq_j z$, and so $x \leq z$. Similarly $x \leq z$ if $(A_j, \leq_j)$ is an initial segment of $(A_i, \leq_i)$.
WO: Assume $\emptyset \neq S \subseteq A$ and pick $x \in S$ and $(A_i, \leq_i) \in \mathcal{C}$ with $x \in A_i$. Let $x_0 = \min_{\leq_i}(S \cap A_i)$. If $y \in S \cap A_i$ then $x_0 \leq_i y$, so $x_0 \leq y$. If $y \in S \setminus A_i$, then $y \in A_j$ where $(A_i, \leq_i)$ is an initial segment of $(A_j, \leq_j)$. But $x_0 \in A_i$, $y \notin A_i$, so $x_0 \leq_j y$. Thus $x_0 \leq y$.
Finally, we need to show $A = X$. Assume otherwise, and set $x = f(A)$. Defining $A' = A \cup \{x\}$ and $\leq'$ by $\leq = \leq'$ on A and $y \leq' x$ for all $y \in A$, we get $(A', \leq') \in \mathcal{C}$. But this implies $A' \subseteq \bigcup_i A_i = A$, a contradiction. Thus $X = A$ and $\leq$ is a well ordering on X. $\square$

Ordinal Numbers

One can define an *ordinal number* to be a set α such that the elements of α are (strictly) totally ordered by $\in$ and every element of α is a subset of α. From this definition (due to John von Neumann), one can prove the following:

1. The elements of an ordinal are well ordered by $\subseteq$. (Needs Axiom of Foundation.)
2. For any two ordinals α, β, either $\alpha \subseteq \beta$ or $\beta \subseteq \alpha$, (Use Lemma 1.)
3. Every ordinal is equal to the set of all ordinals strictly less ($\subsetneq$) than itself.
4. $\emptyset$ is an ordinal. If α is an ordinal, then so is $\alpha^+ = \alpha \cup \{\alpha\}$. If $\mathcal{C}$ is a set of ordinals, then $\bigcup_{\alpha \in \mathcal{C}} \alpha$ is an ordinal.
5. Any well ordered set is order isomorphic to a unique ordinal.
   (Needs Axiom of Replacement.)
6. The ordinals themselves are 'well ordered' by $\subseteq$.
   (There is no 'set of all ordinals', hence the quotes around 'well ordered'.)

The natural numbers under the construction of Section 2 are all ordinals, as is the set $\omega = \mathbb{N}$. One can define addition and multiplication on ordinals by

1. $\alpha + \beta$ is the ordinal which is order isomorphic to the lexicographic (well) ordering on $(\{0\} \times \alpha) \cup (\{1\} \times \beta)$ (i.e., write down the elements of β after all the elements of α).
2. $\alpha\beta$ is the ordinal which is order isomorphic to the lexicographic (well) ordering on $\beta \times \alpha$ (note reversed order — we have β copies of α).

For example $\omega + 1$ has order type $\{0 < 1 < 2 < \cdots < 0'\}$ so is just $\omega^+ = \omega \cup \{\omega\}$. However $1 + \omega$ has order type $\{0 < 0' < 1' < 2' < \cdots\}$ so is just ω. Note addition is not commutative! Similarly $\omega 2 = \{(0,0) < (0,1) < (0,2) < \cdots < (1,0) < (1,1) < (1,2) < \cdots\} = \omega + \omega$ but $2\omega = \{(0,0) < (0,1) < (1,0) < (1,1) < (2,0) < \cdots\} = \omega$.

Math 7411 7. Cardinality Spring 2008

It is possible to define a notion of the 'size' of a set. If A and B are two sets, we say the *cardinality* of A is less than or equal to the cardinality of B, $|A| \leq |B|$, iff there exists an injective function $f\colon A \to B$. Since the composition of two injective functions is injective, if $|A| \leq |B|$ and $|B| \leq |C|$ then $|A| \leq |C|$. We say A and B have the same cardinality, $|A| = |B|$, if there exists a bijection from A to B. The following theorem shows that these definitions define a 'partial order' on cardinalities.

Theorem (Cantor-Schröder-Bernstein theorem) *If there exist injective functions $f\colon A \to B$ and $g\colon B \to A$ between the sets A and B, then there exists a bijective function $h\colon A \to B$. In terms of cardinalities, if $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$.*

Proof. Set $C_0 = A \setminus g[B]$, and inductively $C_{n+1} = g[f[C_n]]$ for $n \geq 0$. Let $C = \bigcup_{n=0}^{\infty} C_n$. Then $C = C_0 \cup g[f[C]]$, so $g^{-1}[C] = g^{-1}[C_0] \cup f[C] = f[C]$ (since g is injective). Define

$$h(x) = \begin{cases} f(x) & \text{if } x \in C; \\ g^{-1}(x) & \text{if } x \notin C. \end{cases}$$

If $x \notin C$, then $x \notin C_0$ and $x \in g[B]$, so there is a unique element $g^{-1}(x) \in B$, and $h(x)$ is well-defined. Now f is a bijection from C to $f[C]$ and g^{-1} is a bijection from $A \setminus C$ to $g^{-1}[A \setminus C] = B \setminus g^{-1}[C] = B \setminus f[C]$. Thus $h\colon A \to B$ is the desired bijection. □

Without AC one cannot show total ordering in general, i.e., that either $|A| \leq |B|$ or $|B| \leq |A|$ holds. Also, one cannot show that the existence of a surjective map $g\colon A \to B$ implies $|B| \leq |A|$. However AC does implies both of these. For the first, put a well-ordering on both sets and apply Lemma 1 of Section 6. The second follows from Exercise 4 of Section 3.

We say a set A is *countable* if $|A| \leq |\mathbb{N}|$, i.e., if there is an injective function $f\colon A \to \mathbb{N}$. This is equivalent to either $A = \emptyset$ or the existence of a surjective function $\mathbb{N} \to A$ (i.e., A is the image of a sequence). Note that we do not need AC for this since $\mathbb{N}$ is well ordered.

Theorem *A countable union of countable sets is countable.*

Proof. Let the sets be A_i, $i \in I$, and choose injections $g\colon I \to \mathbb{N}$ and $f_i\colon A_i \to \mathbb{N}$ (this requires AC). Define $h\colon \bigcup A_i \to \mathbb{N}$ by $h(x) = 2^{g(i)}(2f_i(x) + 1)$ where $g(i)$ is the smallest $g(i)$ such that $x \in A_i$. If $2^i m = 2^j n$ and n, m are odd, then $i = j$ and $m = n$. From this one can deduce that h is injective. □

Note that this implies lots of sets are countable, e.g., $\mathbb{Z} = \mathbb{N} \cup -\mathbb{N}$, $\mathbb{Q} = \bigcup \frac{1}{n}\mathbb{Z}$, $\mathbb{N} \times \mathbb{N}$.

Theorem *The set of all finite sequences $(a_i)_{i=1}^n$ in a countable set X is countable.*

Proof. (Without AC). Let $f\colon X \to \mathbb{N}$ be injective. Set $b_i = \sum_{j=1}^{i}(f(a_j) + 1)$ and map $(a_i)_{i=1}^n$ to $x = \sum_{i=1}^{n} 2^{b_i} \in \mathbb{N}$. This is an injection since one can read off the a_i from the gaps between the 1's in the binary expansion of x, (and n from the number of 1's). □

Theorem (Cantor) *If X is any set then $|X| < |\mathcal{P}(X)|$.*

Proof. Clearly there is an injection $X \to \mathcal{P}(X)$ taking x to $\{x\}$. We shall show there is no bijection, by showing there is no surjective map $g\colon X \to \mathcal{P}(X)$. Consider $A = \{x \in X \mid x \notin g(x)\}$. Then since g is surjective, we can write $A = g(x)$. But then $x \in g(x) \Leftrightarrow x \notin g(x)$, a contradiction. Hence no such g can exist. [cf. Russell's Paradox.] □

This shows that $\mathcal{P}(\mathbb{N})$, and hence $\mathbb{R}$ is uncountable (there exists an injection from $\mathcal{P}(\mathbb{N})$ to $\mathbb{R}$ by sending $A \subseteq \mathbb{N}$ to $\sum_{i\in A} 10^{-i}$). However, the following direct proof is of interest.

Theorem (Cantor's Diagonal Argument) *$\mathbb{R}$ is uncountable.*

Proof. By Exercise 3 of Section 3, if there is an injection $\mathbb{R} \to \mathbb{N}$ then there is a surjection $\mathbb{N} \to \mathbb{R}$, so assume there is a sequence $(x_i)_{i=0}^{\infty}$ whose image is the whole of $\mathbb{R}$. Write each x_i in decimal notation, $x_i = \sum_{j\in\mathbb{Z}} a_{ij}10^{-j}$, $a_{ij} \in \{0, \ldots, 9\}$ (note that this is not unique in general, $0.999\cdots = 1.000\cdots$). Define a real number $x = \sum_{j=0}^{\infty} a_j 10^{-j}$, where $a_j = 2$ if $a_{jj} \geq 5$ and $a_j = 7$ if $a_{jj} < 5$. Check that $x \neq x_i$ for all i, a contradiction. □

One can construct a sequence of larger and larger sets $\mathbb{N}$, $\mathcal{P}(\mathbb{N})$, $\mathcal{P}(\mathcal{P}(\mathbb{N}))$, and it is not hard to show that $|\mathcal{P}(\mathbb{N})| = |\mathbb{R}|$. One can ask: is there a set with cardinality in between these. For example, is there a set A with $|\mathbb{N}| < |A| < |\mathbb{R}|$?

Continuum Hypothesis *There are no sets with $|\mathbb{N}| < |A| < |\mathbb{R}|$, i.e., $|\mathbb{R}|$ is the 'smallest' uncountable cardinality.*

It turns out that the Continuum Hypothesis is independent of the axioms of ZFC, so can neither be proved nor disproved (assuming consistency of ZFC).

Finally, one can construct the smallest uncountable ordinal.

Theorem *There exists a well ordered set Ω such that Ω is uncountable, but every proper initial segment $\{z \in \Omega \mid z < x\}$ is countable.*

Proof. (The result doesn't need AC, but we will use it). Well order $\mathbb{R}$, and let $S = \{x \in \mathbb{R} \mid \{z \mid z < x\} \text{ is uncountable}\}$. If $S = \emptyset$ then $\Omega = \mathbb{R}$ will do. Otherwise, let $\Omega = \{z \mid z < \min(S)\}$. □

Cardinal Numbers

One can define a *cardinal number* to be an ordinal number α such that $|\beta| < |\alpha|$ for all $\beta < \alpha$. The finite ordinals (0, 1, 2, ...) are all cardinal numbers, as is ω, but, e.g., $\omega + 1$ and $\omega 2$ are not since $|\omega + 1| = |\omega 2| = |\omega|$. We can define the infinite cardinals $\aleph_0 = \omega$, $\aleph_1 = \Omega$ as the first ordinal bigger than $\aleph_0$ (i.e., the first uncountable ordinal), and in general, $\aleph_\alpha$ (indexed by ordinals) as the first ordinal with $|\aleph_\beta| < |\aleph_\alpha|$ for all $\beta < \alpha$.

We can define arithmetic as for ordinals, however $+$ and $\times$ are now commutative. Indeed, $\aleph_\alpha + \aleph_\beta = \aleph_\alpha \aleph_\beta = \aleph_{\max\{\alpha,\beta\}}$. If we define $2^{\aleph_\alpha}$ to be the cardinality of $\mathcal{P}(\aleph_\alpha)$, then we know $2^{\aleph_\alpha} > \aleph_\alpha$. The Continuum Hypothesis claims that $2^{\aleph_0} = \aleph_1$.

Math 7411 8. Equivalence Relations Spring 2008

An *equivalence relation* on X is a binary relation $\sim$ such that for all $x, y, z \in X$,

E1. $x \sim x$ (reflexive)

E2. $x \sim y \Rightarrow y \sim x$ (symmetric)

E3. $x \sim y$ and $y \sim z \Rightarrow x \sim z$ (transitive)

The *equivalence class* of x is $\bar{x} = \{y \mid y \sim x\}$.

Example If $f\colon X \to Y$ is any map then the relation defined by $x \sim y$ iff $f(x) = f(y)$ is an equivalence relation. The equivalence classes are all non-empty sets of the form $f^{-1}[\{z\}]$.

The equivalence classes form a *partition* of X, i.e., X is the disjoint union of equivalence classes, each of which is non-empty. Equivalently, $X = \bigcup_{x \in X} \bar{x}$, $\bar{x} \neq \emptyset$, and for all x and y, either $\bar{x} = \bar{y}$ or $\bar{x} \cap \bar{y} = \emptyset$.

Write $X/\!\sim$ for the set of equivalence classes and $\pi\colon X \to X/\!\sim$ for the *projection* map $x \mapsto \bar{x}$. Note that π is always surjective and $x \sim y$ iff $\pi(x) = \pi(y)$ (so our example above is in fact quite general).

Defining functions on equivalence classes

If we have a function $f\colon X \to Y$ and an equivalence relation $\sim$ on X such that if $x \sim y$ then $f(x) = f(y)$, then one can define $\bar{f}\colon X/\!\sim \to Y$ by $\bar{f}(\bar{x}) = f(x)$. The important point to notice is that (a) $\bar{f}$ is defined for every element of $X/\!\sim$ since every element of $X/\!\sim$ can be written in the form $\bar{x}$, and (b) $\bar{f}$ is well-defined, since if $\bar{x} = \bar{y}$ then $x \sim y$ and so $f(x) = f(y)$. Thus $\bar{f}(\bar{x})$ is unambiguous.

Note If you wish to define a function $X/\!\sim \to Y$ it is often easier to construct a function $f\colon X \to Y$ and check that $x \sim y$ implies $f(x) = f(y)$.

Suppose now we have a binary operation $\star\colon X \times X \to X$ with the property that $a \sim a'$ and $b \sim b'$ imply $a \star b \sim a' \star b'$. We can then define a binary operation $\bar{\star}$ on $X/\!\sim$ by $\bar{a} \bar{\star} \bar{b} = \overline{a \star b}$. This is defined since any element of $X/\!\sim$ can be written as $\bar{a}$ or $\bar{b}$. It is well defined, since if $\bar{a} = \bar{a'}$ and $\bar{b} = \bar{b'}$ then $a \sim a'$, $b \sim b'$, and so $a \star b \sim a' \star b'$ and $\overline{a \star b} = \overline{a' \star b'}$.

Example (Construction of $\mathbb{Z}$ from $\mathbb{N}$).
Let $X = \mathbb{N} \times \mathbb{N}$ and define $(a, b) \sim (c, d)$ iff $a + d = b + c$. Check that this is an equivalence relation and write $\mathbb{Z} = X/\!\sim$. [$\overline{(a, b)}$ corresponds to $a - b \in \mathbb{Z}$.] Define $+\colon X \times X \to X$ by $(a, b) + (c, d) = (a + c, b + d)$. Check that if $(a, b) \sim (a', b')$ and $(c, d) \sim (c', d')$ then $(a, b) + (c, d) = (a', b') + (c', d')$. Thus one can define $+\colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ by $\overline{(a, b)} + \overline{(c, d)} = \overline{(a + c, b + d)}$. Similarly, one can define $\times$ by $\overline{(a, b)} \times \overline{(c, d)} = \overline{(ac + bd, ad + bc)}$. One can then check that $+$ and $\times$ have all the usual properties on $\mathbb{Z}$, and that we can regard $\mathbb{N}$ as a subset of $\mathbb{Z}$ with the same $+$ and $\times$ by mapping $a \mapsto \overline{(a, 0)}$.

Example (Construction of $\mathbb{Q}$ from $\mathbb{Z}$).
Let $X = \mathbb{Z}\times(\mathbb{Z}\setminus\{0\})$ and define $(a,b) \sim (c,d)$ iff $ad = bc$. Check that this is an equivalence relation and write $\mathbb{Q} = X/\sim$. [$\overline{(a,b)}$ corresponds to $a/b \in \mathbb{Q}$.] Define $+\colon X \times X \to X$ by $(a,b) + (c,d) = (ad + bc, bd)$. Check that if $(a,b) \sim (a',b')$ and $(c,d) \sim (c',d')$ then $(a,b) + (c,d) = (a',b') + (c',d')$. Thus one can define $+\colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ by $\overline{(a,b)} + \overline{(c,d)} = \overline{(a+c, b+d)}$. Similarly, one can define $\times$ by $\overline{(a,b)} \times \overline{(c,d)} = \overline{(ac,bd)}$. One can then check that $+$ and $\times$ have all the usual properties on $\mathbb{Q}$, and that we can regard $\mathbb{Z}$ as a subset of $\mathbb{Q}$ with the same $+$ and $\times$ by mapping $a \mapsto \overline{(a,1)}$.

Exercises

1. Using the construction of $\mathbb{Z}$ above, prove that $+$ is commutative, $a + b = b + a$; associative, $a + (b + c) = (a + b) + c$; has an identity 0, $a + 0 = a$; and every element a has an inverse $-a$, $a + (-a) = 0$. (So $\mathbb{Z}$ is an *abelian group* under $+$.)
2. Using the construction of $\mathbb{Z}$ above, prove that $\times$ is commutative, $ab = ba$; associative, $a(bc) = (ab)c$; has identity 1, $1a = a$; and distributes over addition, $a(b+c) = ab+ac$. (So $\mathbb{Z}$ is a *commutative ring*.)
3. Repeat questions 1 and 2 with $\mathbb{Q}$. Show also that if $a \neq 0$ then a has a multiplicative inverse a^{-1}, $aa^{-1} = 1$. (So $\mathbb{Q}$ is a *field*.)
4. Let $n > 0$ and define a relation on $X = \mathbb{Z}$ by $a \sim b$ iff $n \mid a - b$. Show that $\sim$ is an equivalence relation with exactly n equivalence classes. Describe these equivalence classes. Show that one can define $+$ and $\times$ on $X/\sim$.
5. If S and T are two relations on a set X, define ST by $xSTy$ iff there exists a z with xSz and zTy. Show that this 'multiplication' is associative: $(ST)U = S(TU)$.
6. If S is a relation on X, let S^{-1} be the relation defined by $xS^{-1}y$ iff ySx and $S^r = SS\ldots S$ (r times, see previous question) with the convention that S^0 is the relation with xS^0y iff $x = y$. Show that $\sim = \bigcup_{r=0}^{\infty}(S \cup S^{-1})^r$ is an equivalence relation and it is the smallest equivalence relation containing S. We call $\sim$ the equivalence relation *generated* by S.
7. Let X be the set of functions $a\colon \mathbb{N} \to \mathbb{Z}$ (i.e., sequences) such that $|a_{i+j} - a_i - a_j|$ is bounded for all $i, j \geq 0$. Define $(a_i)_{i=0}^{\infty} \sim (b_i)_{i=0}^{\infty}$ if $|a_i - b_i|$ is bounded as $i \to \infty$. Show that $X/\sim$ can be identified with the set $\mathbb{R}$ of real numbers.