$$\in$$

B1.2 Set Theory

$$\emptyset$$

Lecture notes

$$\omega$$

HT 2019

$$\aleph$$

Jonathan Pila

$$\forall x \forall y \langle \forall z \langle z \in x \leftrightarrow z \in y \rangle \leftrightarrow x = y \rangle$$

Contents

Administration

LECTURES – M 12:00 and T 14:00 in [L1]
PREREQUISITES – none, but familiarity with some Prelims/A material assumed
CLASSES AND PROBLEM SETS – essential!
SYLLABUS – everything is examinable unless expressly noted otherwise
CONTACT ME – Office N2.17, pila@maths.ox.ac.uk

The axioms

ZF1 - Extensionality **ZF2** - Nullset **ZF3** - Pairs
ZF4 - Unions **ZF5** - Comprehension **ZF6** - Powerset
ZF7 - Infinity **ZF8** - Replacement **ZF9** - Foundation
ZFC10 - Choice

These notes are based on those of Robin Knight [7], though ubstantially revised and rearranged. Please tell me if you find errors or typos. My thanks to Levon Haykazyan, Janpreet Khabra, Derek Khu, Chen-Wei Lin, Kamil Nizinski, Dominik Peters, and Lorin Samija for suggestions and corrections.

1. Introduction

What is a Set?

You are all familiar with the idea of sets of objects, the notions of set membership, union, intersection etc. The curly brackets. This is perhaps the most primitive notion in mathematics.

Cantor wrote (as translated in [10]) "*By an "aggregate" [set] we are to understand any collection into a whole M of definite and separate objects m of our intuition or our thought. These objects we call the "elements" of M. In signs we express this thus: $M = \{m\}$.* "

One of the achievements of Set Theory is to give (arguably) a framework in which *all* mathematics can be formalised: in particular all mathematical objects may be conceived in terms of sets. We will see in part how this is carried out. Therefore, the "set" notion should be conceived as broadly as possible: and collection we can "comprehend" should be a set. However, a very broad interpretation of a set as the "extension" of a property leads to problems.

Paradoxes

Early in the development of Set Theory, examples were found showing that a completely naive formalisation of Set Theory, in which one has unrestricted "comprehension" of sets leads to inconsistencies. Naive comprehension means: given a property (formulated in some suitable language), one can form a set whose elements are *all* the sets with that property.

The property "x does not belong to itself" can be formulated in the standard languages. Then forming $X = \{x : x \notin x\}$ leads to a problem as each of $X \in X$ and $X \notin X$ imply the other. Since we must have either $X \in X$ or $X \notin X$, such a system is inconsistent. This is Russell's paradox (1902). Other, more technical, paradoxes were found by Cantor and Burali-Forti (1897) They all involve objects that are "too big" to be sets.

In view of such problems, one must be **careful** in formulating the basic axioms of Set Theory. In particular, and interestingly, there will be no "set" of all sets. A pertinent question is:

What is an element of set?

You are familiar with integers as elements of $\mathbb{N}, \mathbb{Z}$, and with rational, or real, or complex numbers as elements of $\mathbb{Q}, \mathbb{R}, \mathbb{C}$. But what are individual rational/real/complex numbers? We know that (e.g.) a *group* consists of a *set* of elements with a (suitable) *composition law* defined on it. What are the elements of a group supposed to be? They are permitted to be "anything", and can typically be integers, rational/real/complex numbers, matrices, functions on any of the above.

It is convenient to formulate Set Theory in such a way that we deal *only* with sets, so the elements of sets are also sets. Such sets are called *hereditary sets*. This is not necessary, but very neat.

What is Set Theory?

The mathematical study of sets and membership, formalized in some suitable way by axioms in some (first order) language. There are several ways to do this.

We will work with ZFC, a variant of the first formalisation due to Zermelo (the 'Z') as refined by Fraenkel (the 'F') and with the Axiom of Choice (the 'C'). ZFC is perhaps the most natural and commonly adopted formalisation of Set Theory.

In Set Theory one can study foundational issues, since all mathematics can (arguably) be formulated there, and mathematical issues, mainly around infinite sets.

The origins

Historical comments here and throughout are largely drawn from Jech [6].

Set theory was created by George Cantor (1845-1918) in the years 1874-1897. He was led to his ideas by certain problems in real analysis (Fourier series). In 1874 Cantor discovered that one cannot enumerate the real numbers by the natural numbers, showing that infinite sets come in different "sizes". It was a marvellous period of discovery.

The axioms for set theory (except the Replacement Scheme and Foundation) are due to Zermelo in 1908, following the paradoxes found by Burali-Forti, Cantor, Russell, and Zermelo.

Our objectives

These are set out in more detail in the course synopsis. Essentially we study:

(1) **ZFC**, Zermelo-Fraenkel set theory with the Axiom of Choice.

(2) **Formalization of mathematics within ZFC** (though not in too much detail).

(3) **Infinite cardinal numbers and ordinal numbers and their arithmetic.** The natural numbers are ordinarily used in two distinct ways: as *cardinal numbers* they express the size of a set, or as *ordinal numbers* they assign an ordering to elements of a set. On a cheque, the amount is a cardinal number, the other numbers appearing are ordinals (example from Bedford [26]). Cantor extended the notions of **cardinal** and **ordinal** numbers to infinite sets.

(4) **The Axiom of Choice**. Historically controversial (though see [2, p4]), it is now generally accepted. Key questions (beyond this course) are whether it is (a) consistent with and (b) independent from the other axioms. The answers are: 'yes' (Godel [14]) and 'yes' (Cohen [11]). The first 'yes' is part of the course in Axiomatic Set Theory. In this course we want to get some appreciation of the function of this axiom, so we don't introduce it until late in the course.

What we won't cover/Reading

We will spend little time on the origins and the paradoxes – and generally on the ideas informing and surrounding Set Theory. E.g. What mathematical objects "are", whether infinity "exists", whether the Axiom of Choice is "true". I highly recommend the books by Goldrei [1], Lavine [2], Rucker [3], and Stillwell [24] as excellent places to start gaining further insight on these issues.

Goldrei is also our text and should be read for what we are going to study as well as material we are not going to study. Other books to refer to on course material are listed in the course synopsis. Some further reading is suggested at the end of the notes.

2. The language of set theory and the first axioms

The *Language of Set Theory* is the language of First Order Predicate Calculus (FOPC) having one non-logical symbol "$\in$", a binary relation which is supposed to denote the membership relation

$$x \in X$$

between two sets. So this language $\mathcal{L}$ consists of $\in$, the logical symbols $\wedge, \vee, \neg, \rightarrow, \forall, \exists, =$, and brackets $($, $)$ so we can parse formulae, and an infinite list of variables which we will denote $x, y, z, X, Y, Z, a, b, c, \ldots$, which will range over sets. Here the predicates $\forall, \exists$ ("for all" and "there exists") also range over sets. The meaningful statements in $\mathcal{L}$ are called *formulae.* Note the language does not contain the symbols $\{, \}$, which (like other symbols) we introduce and use informally.

Formally, an *atomic formula* is an expression of the form $x \in y$ or $x = y$ for some variables x, y. The collection of *formulae* is built from the atomic formulae by recursively using the logical connectives (" if ϕ, ψ are formulae then so are $\phi \wedge \psi, \phi \vee \psi, \neg\phi, \phi \rightarrow \psi$") and quantifiers ("if ϕ is a formula and x is a variable then $\forall x\phi$ and $\exists x\phi$ are formulae"). An occurence of a variable in a formula is *bound* if it is in the scope of a quantifier; otherwise it is *free.* We recall that a formula $\phi \rightarrow \psi$ is logically equivalent to $\neg\phi \vee \psi$, i.e. it is true unless ϕ is true **and** ψ is false, and in particular it is true if ϕ is false.

We will see that statements about sets in "ordinary" mathematical English may be expressed by formulae in $\mathcal{L}$. You will hopefully become comfortable with this. However working completely in $\mathcal{L}$ quickly gets cumbersome (as we will see!). So we will express arguments in the usual mathematical way, but those of you who have studied FOPC (as in B1a or equivalent) should see that everything could be done formally in $\mathcal{L}$. See Goldrei [1], §4.2.

A formal language is necessary in order to make our axiom system precise, in particular the comprehension and replacement axioms (and avoid paradoxes). It also means that the whole system can be studied formally, as is done in Part C Axiomatic Set Theory.

We will want to introduce some additional notation, observing that formulae using the "additional" notation can be equivalently expressed without it, though in a less readable form. E.g. we will want to use familiar things like curly brackets, union and intersection symbols in the usual ways.

Example. The $\mathcal{L}$ formula

$$\forall x\big(x \in y \rightarrow x \in z\big)$$

expresses what we would say as "y is a subset of z", and we will write "$y \subseteq z$" bearing in mind that we have defined this by means of the $\mathcal{L}$-formula above.

Note. For me $y \subset x$ denotes "y is a subset of x" allowing equality. Others interpret this as a "proper subset", excluding equality. I will try to avoid the issue by using $\subseteq$.

We now come to the first axioms. These start simply then speed up somewhat.

ZF1 – The axiom of extensionality: *Two sets are equal if and only if they have the same elements.*

In $\mathcal{L}$ this may be expressed by the formula $\forall x \forall y \big((x = y) \leftrightarrow \forall z(z \in x \leftrightarrow z \in y)\big)$. Note: $\phi \leftrightarrow \psi$ means $(\phi \to \psi) \wedge (\psi \to \phi)$. This axiom serves to formalise the fact that we will deal only with sets. Oberve also that equality depends just on what elements are in the set, not on how it was defined.

ZF2 – Null set axiom: *There exists a set having no elements.*

In $\mathcal{L}$ this may be expressed by the formula $\exists x \forall y (\neg(y \in x))$.

Note. We will also use $x \notin y$, as a shorthand for $\neg(x \in y)$, and then we can write the null set axiom equivalently as $\exists x \forall y (y \notin x)$.

With these two axioms we may prove a theorem.

2.1. Theorem. *There is a* **unique** *set having no elements.*

For those with B1.1 (or equivalent): the theorem may be expressed in $\mathcal{L}$ by the formula

$$\big[\forall x \forall y \big((x = y) \leftrightarrow \forall z(z \in x \leftrightarrow z \in y)\big) \wedge \exists x \forall y \big(\neg(y \in x)\big)\big]$$

$$\to \exists x \Big(\forall t \neg(t \in x) \wedge \forall y \big(\forall w(w \notin y) \to y = x\big)\Big).$$

It is not very readable, but is formally provable in FOPC (first-order predicate calculus). However, we will be proving theorems in the "usual" style, e.g. as follows.

Proof. By ZF2 there is an empty set x. Suppose y is also an empty set. Then x and y have the same elements (i.e. none), and so are equal by ZF1. (I.e. for every z, $z \in x \leftrightarrow z \in y$ is true because BOTH $z \in x$ and $z \in y$ are false). □

2.2. Definition. The set with no elements is called the *empty set* and denoted $\emptyset$.

The term and symbol are of course standard. Formally, $\emptyset$ is not in our language. But informally we will add $\emptyset$ to the language, remembering that any formula ϕ using $\emptyset$ can be expressed equivalently without $\emptyset$, as follows. Form a new formula $\psi(z)$ by replacing each occurrence of $\emptyset$ by a new variable z (there might also be other free variables). Now form

$$\Phi = \forall z \Big(\forall w(w \notin z) \to \psi(z)\Big).$$

Of course this restatement uses additional quantifiers, but has not changed what is expressed.

Several subsequent axioms assert that, given some sets $x, y, z, \ldots$, some other sets exist, such as unordered pairs, union etc, and enable us to start building a reasonable "universe" of sets. Otherwise we would only have $\emptyset$.

ZF3 – Unordered pairs: *If x, y are any two sets there is a set whose elements are precisely x and y.*

This may be expressed in $\mathcal{L}$ as: $\forall x \forall y \exists z \Big(x \in z \wedge y \in z \wedge \forall w (w \in z \to [w = x \vee w = y]) \Big)$.

Given x, y the set z is unique (see Problem Set 1, but fairly clear) and will be denoted $\{x, y\}$. That is, we will add $\{,\}$ to our language as well, and allow ourselves to write unordered pairs in this way. As with $\emptyset$, this makes our language more readable without actually changing what can be expressed.

Observe that, if x is a set, then $\{x\} = \{x, x\}$ is a set by ZF3.

In mathematics we very often use *ordered pairs*, e.g. cartesian coordinates, and say e.g. that a function is a set of ordered pairs. However, sets are unordered. So we introduce a somewhat cumbersome device to "imitate" ordered pairs. Once we have seen that it works, we will introduce a notation for ordered pairs and proceed to use it in the usual ways.

2.3. Definition. Let x, y be sets. The set $\{\{x\}, \{x, y\}\}$ is called the *ordered pair* of x and y. The set x is called the *first coordinate*, and y the *second coordinate*, of the ordered pair. We set

$$\langle x, y \rangle = \{\{x\}, \{x, y\}\}.$$

For the definition to "work" means

(1) with every ordered pair (x, y) of sets we associate a set $\langle x, y \rangle$ which will serve as the ordered pair. This we clearly have, using our pairs axiom, twice: $\{x\}$ is a set and $\{x, y\}$ is a set so $\{\{x\}, \{x, y\}\}$ is a set.

(2) given a set z it can be the ordered pair $\langle x, y \rangle$ for only one ordered pair of sets (x, y), i.e we can uniquely recover x as first coordinate and y as second coordinate given only the set $\langle x, y \rangle$. This is established by the following theorem.

2.4. Theorem. *Let x, y, u, v be sets. If $\langle x, y \rangle = \langle u, v \rangle$ then $x = u$ and $y = v$.*

Proof. Suppose that

$$\{\{x\}, \{x, y\}\} = \{\{u\}, \{u, v\}\}.$$

We will heavily use ZF1 (extensionality), referring to $\{\{u\}, \{u, v\}\}$ as "RHS" (easier to write down).

Suppose $x = y$ (which is certainly not excluded). Then $\{x, y\} = \{x\}$ and

$$\{\{x\}, \{x, y\}\} = \{\{x\}, \{x\}\} = \{\{x\}\}$$

is a set with just one element $\{x\}$. By ZF1, each element of RHS is an element of LHS, so

$$\{u\} = \{x\}, \quad \{u, v\} = \{x\}$$

so that $\{u, v\} = \{u\}$, which implies $u = v$, and $u = x$, and so $y = x = u = v$. If $u = v$, a symmetrical argument gives $x = y = u$.

So we can assume that $x \neq y$ and $u \neq v$. Since $\{x\}$ is an element of LHS, it must be an element of the RHS. As RHS has just two element, we are in one of two cases.

Case 1: $\{x\} = \{u, v\}$. This implies $u = x, v = x$, whence $u = v$, a contradiction. So we must be in

Case 2: $\{x\} = \{u\}$, which implies $x = u$.

Next $\{x, y\}$ is an element of RHS, and $\{x, y\} = \{u\}$ leads to a contradiction. So we have $\{x, y\} = \{u, v\}$. Now y must equal one of the elements u, v of (new) RHS. But $y = u$ and $u = x$ leads to $x = y$, a contradiction. So $y = v$. □

Note. Observe the vital importance of using the right number of $\{\}$'s, as I hope I did. Please be extremely careful, as $x, \{x\}, \{\{x\}\}$ etc are different sets. The set x may have many elements, may be infinite, or might be empty, while $\{x\}$ always has precisely one element: x. The set $\{\{x\}\}$ also has precisely one element but it is $\{x\}$, not x. And please let me know if I make slips of this kind (or other kinds) in notes or lectures.

This gives us individual ordered pairs. We will want to form Cartesian products of any two sets as a *set* of ordered pairs. This will come a bit later.

One thing we certainly want is to take unions (and intersections, complements etc). This is accomplished in a somewhat round-about way.

ZF4 – Unions: *Given a set x, there is a set consisting of all the elements of all the elements of x.*

In $\mathcal{L}$ this may be expressed by the formula $\forall x \exists y \forall z \Big(z \in y \leftrightarrow \exists w \big(z \in w \wedge w \in x \big) \Big)$.

The set whose existence is asserted above is denoted $\bigcup x$. It is what we would mean by the "naive" comprehension $\bigcup x = \{z : \exists w \in x : z \in w\}$.

2.5. Proposition. *If x, y are sets, the union of x and y is a set.*

Proof. By pairs, we form $z = \{x, y\}$. The elements of $\bigcup z = \bigcup \{x, y\}$ are then precisely the elements of x and y, so it is the union of x and y. □

The union of x and y will be denoted $x \cup y$ in the usual way. This axiom will also allow us to define infinite unions as unions over infinite sets, while an axiom asserting the existence of $x \cup y$ would extend to finite unions, but not to infinite ones.

ZF5 – Comprehension scheme: *If $\phi(z, w_1, \ldots, w_k)$ is a formula in $\mathcal{L}$, with free variables among $z, w_1, \ldots, w_k$, and x is a set then for any sets $w_1, \ldots, w_k$ there is set y consisting of the elements z of x for which $\phi(z, w_1, \ldots, w_k)$ holds.*

In $\mathcal{L}$ (possibly clearer!): If $\phi(z, w_1, \ldots, w_k)$ is a formula of $\mathcal{L}$ whose free variables are among $z, w_1, \ldots, w_k$ (and so "asserts something" about $z, w_1, \ldots, w_k$), then

$$\forall x \forall w_1 \ldots \forall w_k \exists y \forall z \Big[z \in y \leftrightarrow \big(z \in x \wedge \phi(z, w_1, \ldots, w_k) \big) \Big].$$

The w_i are "parameters". This set y will be denoted

$$\{z \in x : \phi(z, w_1, \ldots, w_k)\}.$$

Observe that this is an **axiom scheme**, one axiom for each formula ϕ, not a single axiom. It is a much restricted version of naive comprehension – which would give (for each choice of $w_1, \ldots, w_k$) a set

$$\{z : \phi(z, w_1, \ldots, w_k)\}.$$

This version proceeds from the idea that a problem with naive comprehension is that it allows one to define a set by quantifying over a domain that includes the set being defined, even though that set does not exist "yet". Here we allow comprehension of subsets of a set which already exists. The Comprehension scheme is also called the **Subset axiom scheme** and the **Separation scheme**.

With this axiom one can do quite a bit.

2.6 Proposition. *If x, y are sets, so is their intersection $x \cap y$ and their difference $x \backslash y$.*

Proof. By comprehension using $\phi(z, w) = z \in w$ and $\psi(z, w) = \neg(z \in w)$ respectively and $w = y$ we have

$$x \cap y = \{z \in x : \phi(z, y)\}, \quad x \backslash y = \{z \in x : \psi(z, y)\}. \square$$

More generally

2.7. Proposition. *Let x be a non-empty set. Then the intersection of all elements of x is a set, denoted $\bigcap x$.*

Proof. The set $\bigcup x$ that we already have includes the set we want as the subset of elements having a certain property, and so

$$\bigcap x = \{z \in \bigcup x : \forall y\Big((y \in x) \to (z \in y)\Big)\}. \square$$

Note that if we define $\bigcap \emptyset$ by saying it is "the set of z such that, for all sets y, if $y \in x$ then $z \in y$" we would get all sets! This is not a valid definition since we are not comprehending z over a set. If we comprehend over $\bigcup \emptyset$, as in the proof, then $\bigcap \emptyset = \emptyset$, since $\bigcup \emptyset = \emptyset$.

With this axiom one also finds that certain sets cannot exist.

2.8. Proposition. *There is no set of all sets.*

Proof. By contradiction. If x is a set we can 'comprehend'

$$y = \{z \in x : z \notin z\}.$$

If there were a set x of all sets then we recover Russell's Paradox, for $y \in x$, so $y \in y \leftrightarrow y \notin y$. □

Similarly, while we can form set differences $x \backslash y$, there is no "absolute" complement (where this idea is used one is always operating inside some fixed large set).

2.9. Proposition. *Let x be a set. There is no set consisting of all sets that are not elements of x.*

Proof. See Problem Set 1. □

Classes

Still, it is useful to be able to refer to the "collection" of all sets, and we introduce the informal notion of a *class*. It is informal in the sense that variables in our language range over sets; classes are not part of the formal setup.

2.10. Definition. A *class* is a collection of the form

$$\mathbf{X} = \{x : \phi(x)\}$$

where ϕ is a formula of $\mathcal{L}$, and x ranges over sets. The formula ϕ is permitted to have other free variables, which we view as parameters. The class $\mathbf{V} = \{x : x = x\}$ is the traditional notation for "the universe" of all sets.

A class (being a collection of sets) may be a set, or it may not. A class which is not a set is called a *proper class*. If a class belongs to another class, then it is a set (since the elements of a class are sets), and since any set belongs to the class

$$\mathbf{V} = \{x : x = x\}$$

of all sets, we see that a proper class is just a class that is not an element of any class, or of any set. We can form Russell's class

$$\mathbf{Y} = \{x : x \notin x\}$$

and the paradox disappears, becoming simply a proof that $\mathbf{Y}$ is not a set.

We will see some other proper classes. If $\mathbf{X}$ (as above) and $\mathbf{Z} = \{x : \psi(x)\}$ are classes we can write statements like

$$\mathbf{X} \subseteq \mathbf{Z}$$

(whose truth or falsity depends on $\mathbf{X}, \mathbf{Z}$) remembering that this can be rephrased without mentioning classes as

$$\forall x\big(\phi(x) \to \psi(x)\big).$$

It is useful for readability to write such statements, but we must be careful not to treat classes as sets (e.g. by "comprehending" over them).

3. The Powerset axiom

The first of a group of very powerful axioms (especially in combination).

ZF6 – Powerset axiom: *Let x be a set. There is a set whose elements are precisely the subsets of x.*

This set is uniquely determined by x (check left to the reader, but similar to Theorem 2.1 and to Problem Set 1, Question 2).

3.1. Definition. The set of subsets of a set x is called the *power set* of x, and is denoted $\mathcal{P}(x)$.

While we have defined individual ordered pairs, the power set axiom enables us to define Cartesian products.

3.2. Proposition. *Let X, Y be sets. There is a set whose elements are all the ordered pairs $\langle x, y\rangle$ where $x \in X$ and $y \in Y$.*

Proof. The key observation is that, if $x \in X$ and $y \in Y$, then

$$\{x\} \in \mathcal{P}(X \cup Y), \quad \{x, y\} \in \mathcal{P}(X \cup Y)$$

and so

$$\langle x, y\rangle = \{\{x\}, \{x, y\}\} \in \mathcal{P}\Big(\mathcal{P}\big(X \cup Y\big)\Big) = P.$$

The set we want can then be separated out using the comprehension axiom. The set

$$Z = \{z \in P : \exists x \exists y\Big(x \in X \wedge y \in Y \wedge z = \langle x, y\rangle\Big)\}$$

has the desired properties: all ordered pairs $\langle x, y\rangle, x \in X, y \in Y$ are elements of Z (as they all belong to P), and every element in it is such an ordered pair. □

Of course this set is uniquely determined by X and Y, so we can refer to it as "the" Cartesian product, and give it its usual name and notation.

3.3. Definition. Let X, Y be sets. The set consisting of all ordered pairs $\langle x, y\rangle$ where $x \in X, y \in Y$ is called the *Cartesian product* of X and Y and denoted $X \times Y$.

Relations and functions

3.4. Definition. Let X, Y be sets. A *relation* between X and Y is a subset $R \subseteq X \times Y$. We sometimes write xRy instead of $\langle x, y\rangle \in R$. A relation between X and X is called a *relation on X*.

3.5. Definition. A relation $R \subseteq X \times Y$ is a *function from X to Y* if, for all $x \in X$, there exists a unique $y \in Y$ such that $\langle x, y\rangle \in R$. If f is a function from X to Y and $\langle x, y\rangle \in f$ we will write $y = f(x)$. A function will also be called a *map* (etc).

This may of course be expressed by an $\mathcal{L}$-formula: a set F is a function from A to B provided

$$\big(F \subseteq A \times B\big) \wedge \forall x\Big(x \in A \to \exists y\big((y \in B) \wedge ((x, y) \in F) \wedge \forall z((x, z) \in F \to z = y)\big)\Big).$$

The collection of all relations between X and Y is just $\mathcal{P}(X\times Y)$, and so is a set. Inside this we may "comprehend" the set of functions (See Question 1 on Problem Set 2).

3.6. Definition. The set of functions from X to Y is denoted Y^X. Note: sometimes the notation ${}^XY$ is used to avoid confusion later with exponentials of cardinal and ordinal numbers.

Explanation of this "exponential" notation: suppose $X = \{a, b\}$ is a two element set. Specifying a function $f : X \to Y$ means choosing an element of Y to be $f(a)$, and an element of Y to be $f(b)$. So it amounts to an element of $Y \times Y$, an ordered pair. If X has 3 elements, one needs 3 choices from Y, i.e. an element of $Y \times Y \times Y$.

In general we choose an element of Y for each element of X, so it is a kind of exponentiation of sets, and the notation works in a familiar way.

If f is a function (i.e. a set of ordered pairs) then the *domain* of f

$$\mathrm{Dom}(f) = \{x \in \bigcup\bigcup f : \exists y \in \bigcup\bigcup f\Big(\langle x, y\rangle \in f\Big)\}$$

is a set (Union, Comprehension) as is the *range* of f

$$\mathrm{Range}(f) = \{y \in \bigcup\bigcup f : \exists x \in \bigcup\bigcup f\Big(\langle x, y\rangle \in f\Big)\}.$$

If we refer to $f : X \to Y$ as a function we mean that X, Y are sets and f is a suitable subset of $X \times Y$. We will make the usual assumption that f is defined on all X, so that $\mathrm{Dom}(f) = X$, but only that $\mathrm{Range}(f) \subseteq Y$. We will also write $f[X]$ for $\mathrm{Range}(f)$.

We consider a function to be a set of ordered pairs. Thus, two functions are equal if they are the same set of ordered pairs (Extensionality), so the "codomain" of a function is not essential: $f : X \to Y$ and $g : X \to Z$ can be equal even if $Y \neq Z$. Of course the *image* of a function f is determined by f, and so if $f = g$ as function then we must have $f[X] = g[X]$. This will be some subset of $Y \cap Z$.

The empty function. According to our definitions, $\emptyset$ is a function from $\emptyset$ to $\emptyset$. This is because the condition $\forall x(x \in \emptyset \to \ldots)$ holds vacuously. Indeed, is is injective and surjective. Thus $\emptyset^\emptyset = \{\emptyset\}$. This may seem odd but will be important.

Equivalence relations and order relations

Apart from functions, two other special kinds of relations get a lot of attention in mathematics: equivalence relations, and order relations.

3.7. Definition. An *equivalence relation* E on a set X is a relation E on X which is

(i) Reflexive: $\langle x, x\rangle \in E$ for all $x \in X$;
(ii) Symmetric: For all $x, y \in X$, if $\langle x, y\rangle \in E$ then $\langle y, x\rangle \in E$
(iii) Transitive: for all $x, y, z \in X$, if $\langle x, y\rangle \in E$ and $\langle y, z\rangle \in E$ then $\langle x, z\rangle \in E$.

There are several kinds of order relations, but they fall into two basic types: the strict ones (like $<$) and the weak ones (like $\leq$).

3.8. Definition. A *weak partial order* on a set X is a relation $R \subseteq X \times X$ which is

(i) Reflexive: for all $x \in X$, $\langle x, x \rangle \in R$;

(ii) Antisymmetric: for all $x, y \in X$, if $\langle x, y \rangle \in R$; and $\langle y, x \rangle \in R$ then $x = y$

(iii) Transitive: for all $x, y, z \in X$, if $\langle x, y \rangle \in R$ and $\langle y, z \rangle \in R$ then $\langle x, z \rangle \in R$.

A *weak total order* on a set X is a weak partial order in which any two elements are comparable: for all $x, y \in X$, either $\langle x, y \rangle \in R$ or $\langle y, x \rangle \in R$.

3.9. Definition. A *strict partial order* on a set X is a relation $S \subseteq X \times X$ which is

(i) Irreflexive: for all $x \in X$, $\langle x, x \rangle \notin S$; and

(ii) Transitive: for all $x, y, z \in X$, if $\langle x, y \rangle \in S$ and $\langle y, z \rangle \in S$ then $\langle x, z \rangle \in S$.

[Observe that (i) and (ii) two **imply** a 'strict antisymmetry'

(iii) For $x, y \in X$, $\langle x, y \rangle$ and $\langle y, x \rangle$ cannot both hold.]

A *strict total order* is a strict partial order such that, for all $x, y \in X$ with $x \neq y$, either xSy, or ySx.

3.10. Proposition. *Let S be a strict partial order on a set X. Then the relation*

$$R = S \cup \{\langle x, x \rangle : x \in X\}$$

("S or equal to") is a weak partial order on X. If S is total then R is total.

Proof. Easy checks. □

Orders are often written "xRy". We will use symbols like $<, \prec$ for strict orders and $\leq, \preceq$ for weak orders. Sometimes (when clear in context) $\preceq$ will denote "$\prec$ or equal to". For these kinds of symbols we will use $y > x$ as an alternative to $x < y$ wherever it is convenient. We will use symbols like $\sim, \approx$ for equivalence relations. The collection of weak partial orders on a set X forms a set by comprehension. Likewise the weak total orders and the strict partial or total orders. (See Question 1, Set 2)

Note that $x \subseteq y$, which looks like a relation isn't (why?).

4. The Axiom of Infinity and the natural numbers

So far we have no axioms that would lead to the existence of an infinite set. The next axiom does this, and enables us to define sets that will serve as the natural numbers. To motivate the axiom, we will indicate how this will be done before stating the axiom.

4.1. Definition. We define (informally as usual, as these symbols are not in our language)

$$0 = \emptyset, \quad 1 = 0^+ = \{0\}, \quad 2 = 1^+ = \{0, 1\}, \quad 3 = 2^+ = \{0, 1, 2\}, \quad \ldots$$

Thus 1 is the unique set satisfying (for x) the formula

$$\exists z \Big(\forall y (\neg y \in z) \wedge z \in x \wedge \forall w (w \in x \to w = z) \Big).$$

The sets 2,3, etc satisfy similar (but longer) formulae.

All these sets already exist as individual sets, but the set consisting of all of them does not. This definition of the natural numbers has the nice feature that the set n has n elements. This is consistent with the idea that n is the abstract property shared by all sets of n elements – we just choose one to be its representative. The set $n+1$ is obtained from the set n in a consistent way:

$$n+1 = \{0, 1, \dots, n\} = \{0, 1, \dots, n-1\} \cup \{n\} = n \cup \{n\}.$$

One thinks of $n+1$ as the successor of n, i.e. the next number.

4.2. Definition. Let x be a set. The *successor* of x is the set $x \cup \{x\}$, which we denote x^+.

By pairs and union, every set has a successor set.

4.3. Definition. A set x is called *inductive* or a *successor set* if $\emptyset \in x$ and if $y \in x$ then $y^+ \in x$ too. I.e. if x is closed under the successor operation.

ZF7 – Axiom of Infinity: *There is an inductive set.*

With this axiom we can define the natural numbers. They are just the elements of the inductive set formed by starting with $\emptyset$ and closing under successor.

4.4. Proposition. *There exists a* **unique** *inductive set which is a subset of every inductive set.*

Proof. The statement "x is a successor set" is expressed by a formula $\phi(x)$ in the language of set theory. Let X be a successor set. Set

$$N = \{x \in X : \forall y \big(\phi(y) \to x \in y\big)\}.$$

The claim is that N is a successor set, that it is a subset of every successor set, and that it is unique with these two properties.

The second claim is immediate from the definition: if $x \in N$ then $x \in y$ for every successor set y.

Suppose $x \in N$. Let Y be a successor set. Then $x \in Y$, and so $x^+ \in Y$. Also $x \in X$ and so $x^+ \in X$. So $x^+ \in N$. Thus N is a successor set, and establishes the first claim.

Finally, suppose N' is (like N) a successor set that is a subset of every successor set. Then $N' \subseteq N$, but also $N \subseteq N'$, and so they are equal. □

4.5. Definition. The unique inductive set which is a subset of every inductive set is called the set of *natural numbers*. It is denoted $\mathbb{N}$ and also ω.

4.6. Theorem "Proof by induction". *Suppose $\phi(x)$ is a property of natural numbers such that $\phi(0)$ holds, and, for all $x \in \mathbb{N}$, if $\phi(x)$ holds then $\phi(x^+)$ holds. Then $\phi(x)$ holds for all $x \in \mathbb{N}$.*

Proof. Let

$$X = \{x \in \mathbb{N} : \phi(x)\}.$$

Then X is an inductive set, and so $\mathbb{N} \subseteq X$. Also $X \subseteq \mathbb{N}$, therefore $X = \mathbb{N}$. □

Now we can (using induction) prove further familiar properties. First we recover the usual ordering $\leq$ on $\mathbb{N}$.

4.7. Definition. Define

$$R_{\leq} = \{\langle n, m\rangle \in \mathbb{N} \times \mathbb{N} : (n \in m) \vee (n = m)\}.$$

We will then write $n \leq m$ for $\langle n, m\rangle \in R_{\leq}$.

4.8. Proposition. *Let $n, m \in \mathbb{N}$. Then $m \leq n$ iff $m \subseteq n$.*

Proof. $\Rightarrow$ By induction on $\mathbb{N}$ we prove the statement $\phi(n)$: "if $m \leq n$ then $m \subseteq n$ ".

Suppose $n = 0$. If $m \leq n$ then $m \in n$ or $m = n$. The first is impossible as $n = \emptyset$ has no elements, so we must have $m = n$, whence $m \subseteq n$, establishing $\phi(0)$.

Next suppose $\phi(n)$ holds for some $n \in \mathbb{N}$ and suppose $m \leq n^+$. Then either $m = n^+$ or $m \in n^+$. Since $m = n^+$ gives immediately $m \subseteq n^+$, we can suppose $m \in n^+ = n \cup \{n\}$. Then either $m = n$ or $m \in n$. Then $m \leq n$ and $m \subseteq n$ by induction. Now $m \subseteq n \subseteq n^+$. So $\phi(n^+)$ holds. So $\phi(n)$ holds for all n.

$\Leftarrow$ By induction on $\mathbb{N}$ we prove $\phi(n)$: "if $m \in \mathbb{N}$ with $m \subseteq n$ then $m \leq n$".

Suppose $n = 0 = \emptyset, m \subseteq n$. Then $m = 0$ too, so $m = n$ and $m \leq n$. So $\phi(0)$ holds.

Suppose $\phi(n)$ holds and that $m \in \mathbb{N}$ with $m \subseteq n^+ = n \cup \{n\}$. First suppose $n \notin m$. Then $m \subseteq n$ giving $m \leq n$ by induction. So $m = n$ or $m \in n$, giving $m \in n^+$. Otherwise $n \in m$. Then $n \leq m$ and this implies (by $\Rightarrow$) $n \subseteq m$. So $n \in m$ and $n \subseteq m \subseteq n \cup \{n\}$. Then $m = n^+$. Either way we have $m \leq n^+$ and $\phi(n^+)$ holds. Therefore $\phi(n)$ holds for all n. □

4.9. Proposition. *The relation $R_{\leq}$ on $\mathbb{N}$ is a (weak) partial order.*

Proof. From the last proposition we know that, for $m, n \in \mathbb{N}$, $m \leq n$ iff $m \subseteq n$. We use this ($\subseteq$ is a weak partial order!):

Reflexivity: $n \subseteq n$, for any n.

Antisymmetry: If $m \leq n$ and $n \leq m$, then $m \subseteq n$ and $n \subseteq m$, whence $m = n$.

Transitivity: Suppose $\ell \leq m, m \leq n$. Then $\ell \subseteq m, m \subseteq n$, so $\ell \subseteq n$, so $\ell \leq n$. □

4.10. Proposition. *The ordering $\leq$ on $\mathbb{N}$ is total.*

Proof. We prove by induction the statement $\phi(n)$: "If $m \in \mathbb{N}$ and $m \not\leq n$ then $n \leq m$".

Suppose $n = 0$. Then $0 = \emptyset \leq m$ for any m (as $\emptyset \subseteq m$ for any m), so $\phi(0)$ holds.

Suppose $\phi(n)$ holds and $m \in \mathbb{N}$ with $m \not\leq n^+$. Then $m \not\leq n$ (by transitivity), and so $n \leq m$ by the induction hypothesis. Since $m \neq n$ (from $m \not\leq n$) we must have $n \in m$ (by Definition 4.7). Then $n^+ \subseteq m$ and so $n^+ \leq m$ (by 4.8). Thus $\phi(n^+)$ holds. □

4.11. Proposition. *Every element of $\mathbb{N}$ is either 0 or the successor x^+ of a unique $x \in \mathbb{N}$.*

Proof. A very easy induction proves $\phi(n)$: "either $n = 0$ or $\exists m \in \mathbb{N} : n = m^+$". This means that every non-zero n is m^+ for some m. To see uniqueness, suppose $n = \ell^+ = m^+$. Then $\ell \cup \{\ell\} = m \cup \{m\}$. Then $\ell \leq m$ and $m \leq \ell$, so that $\ell = m$. □

4.12. Corollary. The set ω is transitive.

Proof. By induction, prove if $m \in n \in \omega$ then $m \in \omega$. Base case is trivial. Use 4.11. □

4.13. Theorem (Strong induction on ω). *Suppose $\phi(x)$ is a property of natural numbers and, for every $n \in \mathbb{N}$, if $\phi(m)$ holds for all $m \in n$ then $\phi(n)$ holds.*

Then $\phi(n)$ holds for all $n \in \mathbb{N}$.

Proof. We set $\psi(n)$: "if $m \in n$ then $\phi(m)$ holds " and prove $\forall n \psi(n)$ by induction. First $\psi(0)$ holds since there are no $m \in 0$. Suppose $\psi(n)$ holds. Then $\phi(m)$ holds for all $m \in n$ and so $\phi(n)$ holds by the hypothesis of the Theorem. But then $\phi(m)$ holds for all $m \in n^+$, whence $\psi(n^+)$ holds. □

The next theorem, which we prove by strong induction, might also be termed "Strong induction for $\mathbb{N}$", as we can recover the previous theorem from it. We will generalize this form later.

4.14. Theorem. (Strong induction on ω) *A non-empty subset of ω has a least element.*

Proof. Let S be a subset of $\mathbb{N}$. Let $\phi(n)$ be the property "$n \notin S$" for a natural number n. Suppose that S has no least element. Clearly, $0 \notin S$, so $\phi(0)$ holds. Suppose that $\phi(m)$ holds for all $m \in n$. So all such $m \notin S$. Then $n \notin S$, or it would be least. So $\phi(n)$ holds. By strong induction, $\phi(n)$ holds for all n, and S is empty. □

Proof of 4.13 from 4.14. Given such ϕ, let $S = \{n \in \mathbb{N} : \neg\phi(n)\}$. Then (by the hypotheses of 4.13), S has no least element. So S must be empty. □

5. Recursion on ω

In the following theorem, "there exists" means that we can prove, by our axioms, that such a set (or function) exists.

5.1. Theorem (Definition by Recursion on ω). *Let X be a set, $x_0 \in X$ and $g : X \to X$ a function. Then there exists a unique function $f : \omega \to X$ with $f(0) = x_0$ and $f(n^+) = g(f(n))$ for all $n \in \omega$.*

Proof. Let $n \in \mathbb{N}$. A function $h : n \to X$ will be called *n-nice* if $n = 0$ or, if $0 \in n$, $h(0) = x_0$ and $h(m^+) = g(h(m))$ for all m with $m^+ \in n$. A function $h : \omega \to X$ will be called *very nice* if, for each n, the restriction of f to n is n-nice. The conclusion of the theorem may be restated as follows.

There is a unique very nice function $f : \omega \to X$.

Claim 1. For each n there exists an n-nice function.

Proof of claim 1. By induction. For $n = 0$ there is nothing to prove. Suppose the assertion is true for n. If $n = 0$, we set $h(0) = x_0$ to define a 1-nice function i.e., formally, the relation $\{\langle 0, x_0\rangle\}$. If $n \neq 0$, so $n = k^+$ and h is n-nice we set $h' = h \cup \{\langle n, g(h(k))\rangle\}$.

Claim 2. Suppose h_1 is n_1-nice, h_2 is n_2-nice, and $m \in n_1 \cap n_2$. Then $h_1(m) = h_2(m)$.

Proof of claim 2. By induction on m. For $m = 0$ the statement holds because $0 \in n_1 \cap n_2$ implies $h_1(0) = x_0 = h_2(0)$. Suppose the assertion is true for m, and suppose that $m^+ \in n_1 \cap n_2$ and h_i are n_i-nice. By the induction hypothesis (in the middle) and the nice-ness (either side),

$$h_1(m^+) = g(h_1(m)) = g(h_2(m)) = h_2(m^+).$$

So the assertion holds for all m.

Claim 3. (Corollary) For each n there is a unique n-nice function.

Claim 4. A very nice function, if it exists, is unique.

Proof of claim 4. Suppose $f_1, f_2 : \omega \to X$ are very nice. For any n, the restrictions of f_1, f_2 to domain n are n-nice, and so agree. So $f_1 = f_2$.

An n-nice function is an element of $\mathcal{P}(\omega \times X)$, and among all elements of $\mathcal{P}(\omega \times X)$ we can "comprehend" the subset N of functions which are n-nice for some n (by a formula $\phi(x)$ expressing "$\exists n \in \omega : x$ is n-nice").

As N is a set of functions, i.e. a set of *sets* of ordered pairs $\langle k, \ell\rangle$, its union

$$f = \bigcup N$$

is a set of ordered pairs – a subset of $\omega \times X$.

Claim 5. f is a very nice function.

Proof of claim 5. We first show f is a function $\omega \to X$. Suppose $n \in \omega$. Then there exists an $(n + 1)$-nice function h, so there is some pair $\langle n, x\rangle \in \bigcup N$, where $x = h(n) \in X$. If $\langle n, x'\rangle \in \bigcup N$ then there is a k-nice function h' with $n \in k$ and $h'(n) = x'$. By Claim 2 we must have $x' = h'(n) = h(n) = x$. So f is a function $\omega \to X$. Finally we must show that f is very nice. Let $n \in \omega$ and let $h : n^{++} \to X$ be n^{++}-nice. We have $h(0) = x_0$, whence $f(0) = x_0$, and $f(n^+) = h(n^+) = g(h(n)) = g(f(n))$. So f is indeed very nice. □

Thus a very nice function exists (by Claim 5) and is unique (by Claim 4). □ □

5.2. Corollary. *Suppose A, X are sets, $h : A \to X$ is a function, and $g : A \times X \to X$ is a function. Then there exists a unique function*

$$f : A \times \omega \to X$$

such that

(i) $f(a, 0) = h(a)$ for all $a \in A$

(ii) $f(a, n^+) = g\big(a, f(a, n)\big)$ for all $a \in A, n \in \omega$.

Proof. For each $a \in A$, apply the Theorem with $x_0 = h(a)$ and $g_a : X \to X$ given by $g_a(x) = g(a,x)$ to get a unique function $f_a : \omega \to X$ with $f_a(0) = h(a)$ and $f_a(n^+) = g_a\big(f_a(n)\big)$ for all $n \in \omega$. We now define $f : A \times \omega \to X$ by $f(a,n) = f_a(n)$. Formally, f exists because it is

$$f = \{x \in \big(A \times \omega\big) \times X : \exists a \in A \exists n \in \omega : x = \langle\langle a,n\rangle, f_a(n)\rangle\}.$$

Clearly f has the required properties. □

6. Arithmetic on ω

6.1. Proposition.
1. There exists a unique function $+ : \omega \times \omega \to \omega$ with the following properties:
(i) $n + 0 = n$ for all $n \in \omega$
(ii) $n + m^+ = (n+m)^+$ for all $n, m \in \omega$.
2. There exists a unique function $. : \omega \times \omega \to \omega$ with the following properties:
(i) $n.0 = 0$ for all $n \in \omega$
(ii) $n.m^+ = n.m + n$ for all $n, m \in \omega$.
3. There exists a unique function $\hat{}\ : \omega \times \omega \to \omega$ with the following properties:
(i) $n\hat{}0 = 1$ for all $n \in \omega$ (we just take $0^0 = 1$)
(ii) $n\hat{}m^+ = (n\hat{}m).n$ for all $n, m \in \omega$.

6.2. Definition. The function $+$ we call *addition* and the function . we call *multiplication*. The function $n\hat{}m$ we call *exponentiation*, and write also n^m.

Proof. We use recursion on ω as in Corollary 5.2 with $A = X = \omega$ and h, g as follows.
1. Take $h(n) = n$ to be the identity function, and $g(n,m) = m^+$.
2. Take $h(n) = 0$ and $g(n,m) = m + n$ (note re-arrangement), which exists by (1).
3. Take $h(n) = 1$ and $g(n,m) = m.n$, which exists by (2). □

We can prove all the familiar properties by induction (which is the natural method for checking properties of sets defined by recursion).

6.3. Theorem. *The following hold for all $n, m, k \in \omega$.*

1. $(n+m)+k = n+(m+k)$ "addition is associative"
2. $n + k^+ = n^+ + k$
3. $k + 0 = 0 + k$
4. $n + 1 = n^+$
5. $n + k = k + n$ "addition is commutative"
6. $n.1 = n$
7. $n.(m+k) = n.m + n.k$ "multiplication distributes over addition"
8. $(n.m).k = n.(m.k)$ "multiplication is associative"
9. $n.k = k.n$ "multiplication is commutative"
10. $m^{n+k} = m^n.m^k$
11. $m^{n.k} = (m^n)^k$

Proof. For each part (except 4 and 6) we prove, by induction on k, that the statement holds for k and all n and/or $m \in \omega$.

1. Base case $k = 0$: $(n + m) + 0 = n + m = n + (m + 0)$. Induction step: $(n+m)+k^+ = ((n+m)+k)^+ = (n+(m+k))^+ = n+(m+k)^+ = n+(m+k^+)$.

2. Base case: $n + 0^+ = (n + 0)^+ = n^+ = n^+ + 0$. Induction step: $n + k^{++} = (n+k^+)^+ = (n^+ + k)^+ = n^+ + k^+$.

3. Base case: $0 + 0 = 0 + 0$. Induction step: $k^+ + 0 = k + 0^+ = (k + 0)^+ = (0+k)^+ = 0 + k^+$.

4. $n+1 = n+0^+ = (n+0)^+ = n^+$.

5. The base case is (3). Induction step (using (2) twice): $n + k^+ = n^+ + k = k + n^+ = k^+ + n$.

6. $n.1 = n.0^+ = n.0 + n = 0 + n = n$.

7. Base case: $n.(m+0) = n.m = n.m+0 = n.m+n.0$. Induction step: $n.(m+k^+) = n.(m^+ + k) = n.m^+ + n.k = (n.m + n) + n.k = n.m + (n + n.k) = n.m + (n.k + n) = n.m + nk^+$.

8. Base case: $(n.m).0 = 0 = n.0 = n.(m.0)$. Induction step: $(n.m).k^+ = (n.m).k + n.m = n.m.k^+$.

9. Problem set 2.

10. Base case: $m^{n+0} = m^n = m^n.1 = m^n.m^0$. Induction step: $m^{n+k^+} = m^{(n+k)^+} = m^{n+k}.m = (m^n.m^k).m = m^n.(m^k.m) = m^n.m^{k^+}$.

11. Base case: $m^{n.0} = m^0 = 1 = (m^n)^0$. Induction step: $m^{n.k^+} = m^{n.(k+1)} = m^{(n.k)+n} = m^{n.k}.m^n = (m^n)^k.m^n = (m^n)^{k^+}$. □

Now we have defined $\mathbb{N}$ and its arithmetic operations, we can proceed to define $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, etc. This is described in Goldrei, but is not part of our syllabus. One constructs $\mathbb{Z}$ as the set of equivalence classes in $\mathbb{N} \times \mathbb{N}$ under $\langle a, b\rangle \sim \langle c, d\rangle$ if $a + d = b + c$. Then $\langle a, b\rangle$ represents the integer $a - b$. One extends the arithmetic operations to $\mathbb{Z}$. Let $\mathbb{Z}^+$ denote the set of positive integers. One constructs the rational numbers $\mathbb{Q}$ as the set of equivalence classes in $\mathbb{Z} \times \mathbb{Z}^+$, with now $\langle a, b\rangle \sim \langle c, d\rangle$ if $ad = bc$, so that $\langle a, b\rangle$ represents a/b. One extends the arithmetic operations to $\mathbb{Q}$. One now defines the real numbers $\mathbb{R}$ as the set of Dedekind cuts on $\mathbb{Q}$, which are subsets of $\mathbb{Q}$ having certain properties. One extends the arithmetic operations to $\mathbb{R}$. One defines $\mathbb{C}$ as pairs (x, y) of real numbers with suitable operations on them considered as $x + iy$, etc.

It is all quite straightforward though there are a lot of details to check. The constructions of $\mathbb{Q}, \mathbb{R}, \ldots$ are not on our syllabus (so not examinable), but please read Goldrei to see how it goes.

7. Replacement and Foundation

These two axioms were not among those proposed originally by Zermelo. (And note they were not needed to construct $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ etc.) Their addition was proposed by Fraenkel, Skolem, von Neumann and others.

7.1. Definition. A formula $\phi(x,y)$ (perhaps with other variables as parameters) is called a *class function* if, for all sets x, **there exists a unique** set y such that $\phi(x,y)$ holds.

Intuitively, a class function defines a "function" $\mathbf{V} \to \mathbf{V}$, but it is not a function as it is not a set. Hence the terminology. Often one is only interested in the "values" of $\phi(x,y)$ on some class $\mathbf{X}$ and one can just assume it takes the value $\emptyset$ elsewhere.

Examples. The formulae $y = \mathcal{P}(x)$, $y = x$, $y = x^+$, $y = \{x\}$, $y = \mathcal{P}(x \cup z)$ etc.

ZF8 – Replacement scheme: *For each formula $\phi(x,y)$ as above, the statement saying that if $\phi(x,y)$ is a class function and X is a set then the collection Y of sets y such that $\phi(x,y)$ holds for some $x \in X$ is a set.*

Informally: the image of a set under a class function is a set. With this axiom, we can refer to Y as the *image* of X under the class function $\phi(x,y)$.

The Axiom of Replacement is a powerful axiom that asserts the existence of many (large) sets. It gives a stronger form of the recursion principle.

7.2. Theorem (Recursion principle for ω, class form). *Let x_0 be a set and $\phi(x,y)$ a class function. There is a set Y and a unique function $f : \omega \to Y$ with $f(0) = x_0$ and $\phi(f(n), f(n^+))$ for all $n \in \omega$.*

Proof. This is almost the same as the proof of the Recursion principle, Theorem 5.1. We just invoke Replacement at a certain point. As previously, a function h defined on $n \in \omega$ will be called n-nice under the same conditions as previously. Claims 1,2,3,4 are proved **exactly** as previously.

Claim 1. For each $n \in \omega$ there exists an n-nice function.
Claim 2. An n_1-nice function and an n_2-nice function agree on $n_1 \cap n_2$.
Claim 3. (Corollary): For each $n \in \omega$ there is a unique n-nice function.
Claim 4. A very nice function (if it exists) is unique.

By Claim 3, the formula $\psi(x,y)$ asserting "y is an x-nice function if $x \in \omega$ and otherwise $y = \emptyset$" is a class function. Applying Replacement, the image of ω under $\psi(x,y)$ is a set

$$N = \{h_0, h_1, \ldots\}$$

where h_i is the unique i-nice function. Then

$$f = \bigcup N$$

is a very nice function as previously. □

We will later use a still more sophisticated version to define arithmetic on ordinals. For now, just an example of its use.

7.3. Proposition. *There is a set*

$$X = \{\omega,\ \mathcal{P}(\omega),\ \mathcal{P}\big(\mathcal{P}(\omega)\big),\ \ldots\}.$$

Proof. Let $\phi(x, y)$ be the class function $y = \mathcal{P}(x)$ and let $x_0 = \omega$, and apply the Recursion principle, class form. This gives a set Y and a function $f : \omega \to Y$ with

$$f(0) = \omega, \quad f(1) = \mathcal{P}(\omega), \quad f(2) = \mathcal{P}\big(\mathcal{P}(\omega)\big), \quad \dots$$

We can then form

$$X = \{y \in Y : \exists n \in \omega\big(f(n) = y\big)\}$$

by comprehension. □

ZF8 is used in an essential way to show that various statements are equivalent to the Axiom of Choice (see 12.13). As such it is an essential part of ZFC. It was introduced by Fraenkel (1922), and independently by Skolem (1923).

ZF9 – Foundation: *Every non-empty set has an $\in$-minimal element.*

That is, an element $y \in x$ which does not have as an element any element of x:

$$\forall x\Big(x \neq \emptyset \to \exists y\big((y \in x) \wedge (x \cap y = \emptyset)\big)\Big).$$

The Axiom of Foundation precludes the existence of certain "strange" sets.

7.4. Theorem. *Let x be a set. Then $x \notin x$.*

One can however develop set theory without Foundation. This leads to interesting questions. E.g. if $x = \{x\}$ and $y = \{y\}$ does $x = y$? See the text by Aczel [5]

Proof. Suppose not. Put $X = \{x\}$. Then X violates the Axiom of Foundation: for x is the only element of X but $x \in x$ and $x \in X$ mean that $x \cap X = \{x\} \neq \emptyset$. □

More generally, ZF9 precludes the existence of infinite descending chains of sets.

7.5. Definition. A *sequence* (in a set X) is a function $\phi : \omega \to X$.

Note that if $\phi(x, y)$ is a formula such that, for every $x \in \omega$ there is a unique set y such that $\phi(x, y)$ holds then, by Replacement, the collection of y as x ranges over ω is a set. So in practice we don't need to worry too much about the ambient set of a sequence.

7.6. Theorem. *There is no sequence of sets $\{a_0, a_1, a_2, \dots\}$ with $a_0 \ni a_1 \ni a_2 \ni \dots$*

Proof. Problem Set 3. □

Most sets one encounters in mathematics are well-founded anyway, so this axiom does not often come up. For example (Kunen [8, 2.14]) on AC, every group is isomorphic to a group in WF (the class of well-founded sets, i.e. having an $\in$-minimal element), and every topological space is homeomorphic to a topological space in WF. So in restricting the universe to WF we aren't missing any groups or topological spaces etc.

The axioms ZF1-9 constitute Zermelo-Fraenkel Set Theory (ZF). There is one further axiom, the **Axiom of Choice**, which we defer for now. The Axiom of Choice was one of the axioms proposed by Zermelo, but nowadays it is the 'C' in ZFC. It was very controversial (see e.g. [22], though Lavine [2] says it wasn't). Though it is now generally accepted as a basic axiom of Set Theory, it is customary to keep track of where it is and is not needed. For this reason, we defer introducing.

8. Cardinality

We use numbers in two basic ways: to assign *sizes* to sets by saying how many elements are in the set, and to *order* elements of some set. These are related, as the usual way to determine how many elements there are in some set is to count them i.e. to enumerate them in some order.

These two ideas are extended to infinite sets by the notion of *cardinal* and *ordinal* numbers, respectively. Both ideas are due to G. Cantor.

It is convenient to begin by introducing the notion of when two sets have the "same number" of elements, hence the "same size". The "number" that represents this "size" will be introduced later (with the method of "enumerating").

8.1. Definition. Let X and Y be sets. We say that X and Y *have the same cardinality*, or are *equinumerous*, or are *equipotent*, if there is a bijection between them. We write $X \sim Y$ (or $|X| = |Y|$ or $\#X = \#Y$). We say that X has *cardinality less than or equal to* Y if there is an injection from X to Y, and we write $X \preceq Y$ (or $|X| \leq |Y|$ or $\#X \leq \#Y$). If $X \preceq Y$ but $X \not\sim Y$ we will say that X *has cardinality less than* Y and write $X \prec Y$ (or $|X| < |Y|$ or $\#X < \#Y$).

Since Y^X is a set, each of these properties is expressed by a suitable $\mathcal{L}$-formula $\phi(X, Y)$, and we can refer to them as "class relations" (on $\mathbf{V} \times \mathbf{V}$).

8.2. Proposition. *Let A, B, C be sets. Then $A \sim A$. If $A \sim B$ then $B \sim A$. If $A \sim B$ and $B \sim C$ then $A \sim C$. Thus the class relation $\sim$ is a class equivalence relation.*

Proof. Trivial using properties of bijections. □

8.3. Proposition. *Let A, B, C, X, Y be sets. If $A \preceq B$ and $B \preceq C$ then $A \preceq C$. If $X \subseteq Y$ then $X \preceq Y$. In particular $X \preceq X$. Thus the class relation $\preceq$ is reflexive and transitive.*

Proof. The transitivity is easy, as the composition of two injections is an injection. The assertion about subsets is immediate as the identity map $X \to Y$ is injective. □

If $\phi(x, y)$ expresses $x \preceq y$ then the proposition is expressed in $\mathcal{L}$ by the sentences $\forall x \forall y \big(\forall z (z \in x) \to (z \in y)\big) \to \phi(x, y)$, which implies $\forall x \phi(x, x)$ (reflexivity) and $\forall x \forall y \forall z \big(\phi(x, y) \wedge \phi(y, z) \to \phi(x, z)\big)$ (transitivity). The previous proposition concerning $\sim$ can likewise be expressed in $\mathcal{L}$.

The equivalence classes of $\mathbf{V}$ under $\sim$ are proper classes, and so we might say that $\preceq$ is a (class) weak partial order on them. We will show that $\preceq$ is antisymmetric on the $\sim$-classes.

8.4. Lemma. (Tarski's fixed point theorem) *Suppose X is a set and $H : \mathcal{P}(X) \to \mathcal{P}(X)$ is a function with the property that $H(A) \subseteq H(B)$ whenever $A \subseteq B$. Then there exists $C \in \mathcal{P}(X)$ such that $H(C) = C$ (i.e H has a fixed point).*

Proof. Call a set $A \subseteq X$ *expansive* if $A \subseteq H(A)$. Note $\emptyset$ is expansive. If $\mathcal{A}$ is a set of expansive sets then $\bigcup \mathcal{A}$ is also expansive: For if $a \in \bigcup \mathcal{A}$ there exists $A \in \mathcal{A}$ with $a \in A$, so $a \in A \subseteq H(A)$. But $A \subseteq \bigcup \mathcal{A}$ so $H(A) \subseteq H(\bigcup \mathcal{A})$, and so $\bigcup \mathcal{A} \subseteq H(\bigcup \mathcal{A})$. Then the union

$$C = \bigcup \{A \in \mathcal{P}(X) : A \subseteq H(A)\}$$

of all expansive sets does it. By the above C is expansive, i.e $C \subseteq H(C)$. Then $H(C)$ is also expansive, so $H(C) \subseteq C$. □

8.5. Theorem. (Cantor-/Schroeder-Bernstein) *If $X \preceq Y$ and $Y \preceq X$ then $X \sim Y$.*

Remark. According to Levy [9], the result is in fact due to Dedekind. Schroeder is sometimes written Schröder. See the article by Ferreiros [17].

Proof. Suppose that $f : X \to Y$ and $g : Y \to X$ are injective maps. We would like to create a bijection by using f on some elements and g^{-1} on others. What we need are subsets $Z \subseteq X, W \subseteq Y$ with the following properties.

(i) $f[Z] = Y - W$, so then f restricts to a bijection of Z with $Y - W$; and

(ii) $g[W] = X - Z$, so that g restricts to a bijection of W with $X - Z$.

Then g^{-1} gives a bijection of $X - Z$ with W. Now the map $h : X \to Y$ with $h(x) = f(x)$ for $x \in Z$ and $h(x) = g^{-1}(x)$ for $x \in X - Z$ is a bijection. I leave the details to you to check. The problem is to find such Z, W. What we need is $Z \subseteq X$ such that

$$Z = X - g[Y - f[Z]]$$

for then Z and $W = Y - f[Z]$ will have the desired properties: $f[Z] = Y - (Y - f[Z]) = Y - W$, and $Z = X - g[W]$ implies $g[W] = X - Z$.

So we consider the function $H : \mathcal{P}(X) \to \mathcal{P}(X)$

$$H(A) = X - g[Y - f[A]].$$

We want a set $Z \in \mathcal{P}(X)$ with $H(Z) = Z$, that is we want a "fixed point" for H.

Suppose $A \subseteq B \subseteq X$. Then $Y - f[B] \subseteq Y - f[A]$, so that $g[Y - f[B]] \subseteq g[Y - f[A]]$ and $H(A) = X - g[Y - f[A]] \subseteq X - g[Y - f[B]] = H(B)$. So $H(A) \subseteq H(B)$. Now the existence of a fixed point for H follows from the Lemma. □

And so $\preceq$ is a "class weak partial order on $\mathbf{V}/\sim$". One might expect intuitively that, for any two sets X, Y, either $X \preceq Y$ or $Y \preceq X$, so that $\preceq$ is a "class total partial order on $\mathbf{V}/\sim$". One feels that sizes should be totally ordered. This does not follow from the axioms we have exhibited so far. However it does follow with AC, as we will see later.

9. Countable sets

Finite sets

9.1. Definition. Suppose X is a set and $n \in \omega$. We say that X *has cardinality* n if $X \sim n$, and we write $|X| = n$ (or $\#X = n$).

We say that X is *finite* if $X \sim n$ for some $n \in \omega$, and we say X is *infinite* if it is not finite.

So the elements of ω are Cardinal numbers, the finite ones. We of course want that different $n, m \in \omega$ have different cardinalities, so that we do not have $|X| = n, |X| = m, n \neq m$.

9.2. Theorem. *Let X be a finite set. If $f : X \to X$ is one-to-one then it is onto.*

Proof. A finite set is (by definition) in bijection with some $n \in \omega$, so it suffices to prove the theorem when $X = n$ for some $n \in \omega$. We prove this by induction on n. For $n = 0$, every function $f : \emptyset \to \emptyset$ is both one-to-one and onto. (For $n = 1$ the statement is also immediate.)

Suppose the assertion holds for n and $f : n^+ \to n^+$ is one-to-one. Suppose $k \in n^+ \backslash \text{Range}(f)$. Define $g : n^+ \to n$ as follows. If $k = n$, set $g = f$. Otherwise, set $g(i) = f(i)$ if $f(i) \neq n$ and $g(i) = k$ if $f(i) = n$. Then g is one-to-one, and the restriction $g|_n$ is a one-to-one function from n to $n - \{g(n)\}$, contradicting the induction hypothesis. □

9.3. Theorem. *If $n, m \in \omega$ and $n \neq m$ then $n \not\sim m$.*

Proof. We may suppose that $m < n$. Suppose $f : n \to m$ is a bijection. Then the map $f : n \to n$ consisting of the same ordered pairs is one-to-one but not onto. Contradiction. □

9.4. Corollary. *Let $n, m \in \omega$. Then $n \preceq m$ if an only if $n \leq m$.*

Proof. If $n \leq m$ then $n \subseteq m$ and so $n \preceq m$ by 8.3. If $m < n$ and $n \preceq m$ we would have $m \preceq n$ too and hence $m \sim n$, a contradiction. □

9.5. Theorem. *A subset of a finite set is finite.*

Proof. Problem Set 3. Since a finite set is in bijection with some $n \in \omega$ it suffices to prove that, for $n \in \omega$, a subset of n is finite. This is a straightforward induction starting with the emptyset. □

Countable sets

9.6. Definition. A set X with $X \sim \omega$ is called *countably infinite*. We write $|X| = \aleph_0$.

9.7. Definition. A set X is called *countable* if it is finite or countably infinite.

Thus $\aleph_0$ is our first example of an infinite cardinal, and we evidently have $|\omega| = \aleph_0$.

9.8. Theorem. *If $X \subseteq \omega$ then X is countable.*

Proof. Suppose $X \subseteq \omega$ is not finite. Let x_0 be the least element of X. Let $g : \omega \to \omega$ be defined by $g(n) = \min\{x \in X : n < x\}$.

Claim. $g(n)$ always exists so the function g exists (by Comprehension).

Pf. Else we have $X \subseteq n$ for some $n \in \omega$ implying X is finite.

Now define $f : \omega \to X$ by recursion: $f(0) = x_0$, and $f(n^+) = g(f(n)) = \min\{x \in X : f(n) < x\}$.

Claim. f maps onto X.

Pf. Otherwise, let $a \in X \backslash \text{Range}(f)$. By induction we find that $f(n) < a$ for all $n \in \omega$ (using that ω is strictly totally ordered by $<$). However another induction shows that $f(n) \geq n$ for all n. Contradiction.

Clearly f is one-to-one. So it is a bijection and X is countably infinite. □

9.9. Corollary. *A set X is countable iff $X \preceq \omega$.*

Proof. $\Rightarrow$ is obvious from the definitions. For $\Leftarrow$, suppose $f : X \to \omega$ is one-to-one. Then f is a bijection with Range(f), which is countable by the previous theorem. □

9.10. Theorem. *A non-empty set X is countable iff there is a function from ω onto X.*

Proof.

$\Rightarrow$ Suppose X is countable. If X is countably infinite there is a bijection from X to ω. The inverse is then a map from ω onto X. If X is finite there is a bijection $f : n \to X$ for some $n \in \omega$. Since X is non-empty, $n \geq 1$. This map is already onto, so defining $g : \omega \to X$ by $g(i) = f(i), i \in n$ and $g(i) = f(0)$ otherwise does it.

$\Leftarrow$ Suppose $f : \omega \to X$ is onto. Define $g : X \to \omega$ by $g(x) = \min\{n : f(n) = x\}$. Then g is one-to-one, so $X \preceq \omega$ whence X is countable. □

10. Basic cardinal arithmetic

10.1. Provisional definition. A *cardinal number* is the cardinality $|X|$ of some set X. For $n \in \omega$ we set $n = |n|$ (so $0 = |\emptyset|, 1 = |\{\emptyset\}|$ etc) and call these the *finite cardinals*.

10.2. Definition. Suppose a and b are cardinal numbers, with $a = |A|$ and $b = |B|$ where A, B are disjoint sets (as can always be arranged). We define

$a + b$ is the cardinality of $A \cup B$
$a.b$ is the cardinality of $A \times B$
a^b is the cardinality of A^B.

We must verify that these definitions are well-defined – that is that they do not depend on the choice of A and B.

10.3. Proposition. *These operations are well-defined.*

Proof. Straightforward playing with the bijections. □

10.4. Theorem. *Suppose κ, λ, μ are cardinal numbers. Then*

1. $\kappa + \lambda = \lambda + \kappa$
2. $\kappa + (\lambda + \mu) = (\kappa + \lambda) + \mu$
3. $\kappa + 0 = \kappa$
4. $\kappa.\lambda = \lambda.\kappa$
5. $\kappa.(\lambda.\mu) = (\kappa.\lambda).\mu$
6. $\kappa.1 = \kappa$
7. $\kappa.(\lambda + \mu) = \kappa.\lambda + \kappa.\mu$ *(usual rules of precedence)*
8. $\kappa^{\lambda+\mu} = \kappa^{\lambda}.\kappa^{\mu}$
9. $\kappa^{\lambda.\mu} = (\kappa^{\lambda})^{\mu}$
10. $(\kappa.\lambda)^{\mu} = \kappa^{\mu}.\lambda^{\mu}$
11. *These operations agree on finite cardinals with the operations defined in 6.1.*

Proof. I don't want to belabour these, some parts are on Problem Set 3. In 1,2,3 the sets involved are the same. In 4 the map $(x, y) \mapsto (y, x)$ gives a bijection of $X \times Y$ with $Y \times X$. Similar "natural" bijections work in 5,6,7,8,9,10. Part 11 is fairly straightforward by induction. □

10.5. Theorem. *For cardinals $\kappa \leq \kappa', \lambda \leq \lambda'$ we have $\kappa + \lambda \leq \kappa' + \lambda'$ and $\kappa.\lambda \leq \kappa'.\lambda'$.*

Proof. Very straightforward to show $\preceq$ for the corresponding sets. □

Also $\kappa^{\lambda} \leq \kappa'^{\lambda'}$ usually (by extension of functions) though there is an exceptional case involving the empty set. See Problem Set 3.

10.6. Proposition. $2^X \sim \mathcal{P}(X)$ *for any set X.*

Proof. Define $\Phi : 2^X \to \mathcal{P}(X)$ as follows. For $f : X \to \{0, 1\}$ then $\Phi(f) = \{x \in X : f(x) = 1\}$. It is straightforward to verify that Φ is a bijection. □

10.7. Corollary. $2^{|X|} = |\mathcal{P}(X)|$. □

10.8. Theorem (Cantor). *Let X be a set. There is no surjective map from X to $\mathcal{P}(X)$. In particular there is no bijective map.*

Proof. True for $X = \emptyset$. Suppose $X \neq \emptyset$. Let $f : X \to \mathcal{P}(X)$ be any function. We will show that Range$(f) \neq \mathcal{P}(X)$. By the Comprehension Scheme the "diagonal"

$$D = \{x \in X : x \notin f(x)\}$$

is a set, and it is a subset of X hence an element of $\mathcal{P}(X)$. However, for $a \in X$, $D \neq f(a)$ since $a \in D$ iff $a \notin f(a)$. So $f(a)$ and D "differ" at a. □

This key discovery implies that infinite sets come in different sizes.

10.9. Corollary. $\kappa < 2^{\kappa}$ *for any cardinal number κ.*

Proof. First we show $\kappa \leq 2^\kappa$ by means of a one-to-one map $X \to \mathcal{P}(X)$ where $\kappa = |X|$. For $\kappa = 0$, the empty map does it. For $\kappa \neq 0$, hence non-empty X, map $x \in X$ to $\{x\} \in \mathcal{P}(X)$. By Cantor's Theorem we cannot have $X \sim 2^X$, hence $\kappa < 2^\kappa$. □

10.10. Corollary of the Corollary. $\aleph_0 < 2^{\aleph_0} < 2^{2^{\aleph_0}} < \ldots$. *In particular, there are infinitely many different infinite cardinal numbers.* □

10.11. Proposition. $\omega \times \omega \sim \omega$.

Proof. This is in Problem Set 0. One solution uses *Godel's pairing function.* $f(m,n) = 2^m.(2n+1) - 1$. I outline three other proofs.

(1) Let p_n denote the n-th prime number ($p_0 = 2, p_1 = 3, \ldots$). The map $(n,m) \mapsto (p_n)^{m+1}$ is injective by uniqueness of prime factorization. (A little thought is required to see that this "map" is truly a function. The prime numbers form a set etc.)

(2) Pictorially, one can obviously enumerate, for each k, those pairs (n,m) with $n + m = k$.

(3) Let θ be a positive irrational number (like $\sqrt{2}$). Then the map $(n,m) \mapsto n + m\theta$ is an injective map from $\omega \times \omega$ to $\mathbb{R}^+$. We can then map k to the k-th pair: the order $(n,m) <<< (n',m')$ iff $n + m\theta < n' + m'\theta$ is indeed a well-order (see §11) because there are only finitely many (n,m) below any given bound.

(4) Use $g(m,n) = 2^m 3^n$. □ □ □

10.12. Corollary. $\aleph_0.\aleph_0 = \aleph_0$. $|\mathbb{Q}| = \aleph_0$. □

Observe that, for any $n \in \omega$, $\aleph_0 + n = \aleph_0$ and (for $n > 0$) $\aleph_0.n = \aleph_0$, and (inductively) $\aleph_0^n = \aleph_0$. However

10.13. Theorem. $|\mathbb{R}| = 2^{\aleph_0}$*, the cardinality (or "power") of the continuum.*

Proof. We first show $2^{\aleph_0} \leq |\mathbb{R}|$ by giving an injective map $\Phi : 2^\omega \to \mathbb{R}$. For $f \in 2^\omega$ we define

$$\Phi(f) = \sum_{n=0}^{\infty} \frac{f(n)}{3^n}.$$

This is injective (check – you will see why a 3 in the denominator is needed rather than a 2, which would lead to non-injectivity).

Next we show $|\mathbb{R}| \leq 2^{\aleph_0}$ by giving an injective map $\Psi : \mathbb{R} \to \mathcal{P}(\mathbb{Q})$. For $x \in \mathbb{R}$ set

$$\Psi(x) = \{q \in \mathbb{Q} : q < x\}.$$

This is injective because $\mathbb{Q}$ is dense in $\mathbb{R}$: between any two distinct real numbers there is a rational number. The image sets are of course of special form, they are Dedekind cuts, i.e. the map is far from being onto. So we have $2^\omega \preceq \mathbb{R} \preceq \mathcal{P}(\mathbb{Q}) \sim \mathcal{P}(\omega) \sim 2^\omega$. Now apply Cantor/Schroeder-Bernstein. □

Algebraic and transcendental numbers

10.14. Definition. A complex number α is called *algebraic* if it is there is a nonzero polynomial $P \in \mathbb{Q}[x]$ with $P(\alpha) = 0$. A complex number which is not algebraic is called *transcendental.*

10.15. Theorem. *There exist transcendental numbers.*

Proof. The set $\mathbb{Q}[x]$ is countably infinite, and therefore the set of algebraic numbers is countably infinite. But the set of complex numbers has cardinality $2^{\aleph_0}$, and so the set of transcendental numbers has cardinality $2^{\aleph_0}$. □

This fact was observed by Cantor, though the existence of transcendental numbers had been established earlier (in 1844) by Liouville who showed e.g. that $\sum_{i=1}^{\infty} 10^{-i!}$ is transcendental.

11. Well-orders

Here we consider the use of numbers to "enumerate" the elements of a set. What should this mean for infinite sets? This is the key new concept in this course.

If we enumerate a set X this will certainly lead to a strict total ordering $\prec$ of X, since for any two distinct elements one of them will occur "before" the other in the enumeration. But having a total order is not enough: the real numbers are totally ordered by the usual ordering $<$ but this does not at all lend itself to "enumerating" them. Our ordering must give us a "first" element, and, at any stage of the enumeration, a "next" element. Here a "stage" means we have already enumerated some subset $Y \subseteq X$, so Y is such that whenever $y \in Y$ and $w \prec y$ then $w \in Y$ too (because it occurs "before" y).

11.1. Definition. A strict total order $\prec$ on a set X is a *well-ordering* of X if every non-empty subset $S \subseteq X$ has a least element. Such an order is also called a *well-order*, and a set X with a well-order on it is called a *well-ordered set.*

A well-order on a set X does give an "enumeration". At any stage of enumeration, if the subset Y has been enumerated already, we take $S = X - Y$. If it is empty, we have finished enumerating X. Otherwise there is a least element of S which serves as the "next" element of X. The definition gives us a bit more than we really need (we need a least element for any $X - Y$ for Y as above) though in fact the weaker property gives this one.

Examples. The set

$$\omega = \{0, 1, 2, \ldots\}$$

is well ordered by $\in$ (this was established as Theorem 4.13). The set

$$Y = \{0, 1/2, 3/4, 7/8, 15/16, \ldots, (2^n - 1)/2^n, \ldots\}$$

is well-ordered by $<$. The set

$$X = \{0, 1/2, 3/4, 7/8, 15/16, \ldots, (2^n - 1)/2^n, \ldots, 1, 3/2, 15/8, \ldots, (2.2^n - 1)/2^n, \ldots\}$$

is well-ordered by $<$. The least element of $X - Y$ is 1, the "next" element is 3/2 etc. However the sets

$$\{\ldots, \frac{1}{n}, \ldots, \frac{1}{3}, \frac{1}{2}, 1\}, \quad \{0, \ldots, \frac{1}{n}, \ldots, \frac{1}{3}, \frac{1}{2}, 1\},$$

are not well-ordered by $<$. In the first set there is no least element, while in the second set there is no least element among the positive elements.

11.2. Proposition. *A strict total ordering of a finite set is a well-order.*

Proof. Let X be a finite set strictly totally ordered by $\prec$. Let S be a non-empty subset. We must show that S has a least element. Since S is again a finite set (Theorem 9.5, Problem Set 3), the proposition is equivalent to the statement that *a non-empty finite strictly totally ordered set S has a least element*. We prove this by induction on the cardinality of S. This is clearly true if $|S| = 0$ or if $|S| = 1$. Suppose the assertion is true for every finite set T with $|T| \leq n$, and suppose the cardinality $|S| = n + 1$. So there is a bijection $f : n + 1 \to S$. Then $T = S - f(n)$ has cardinality n, and has a least element t. Then t and $f(n)$ are distinct, we have either $t \prec f(n)$, in which case t is least in S, or $f(n) \prec t$, in which case $f(n)$ is least in S. □

So nothing new for finite sets; for countably infinite there are. The usual ordering of $\mathbb{Q}$ is not a well-order. Indeed:

11.3. Theorem. *A subset of $\mathbb{R}$ that is well-ordered by the usual order $<$ is countable.*

Proof. Problem Set 4. □

You might think about the question: *should there exist a well-ordering of $\mathbb{R}$?* Strong Induction on ω may be generalized to any well-ordered set.

11.4. Theorem. (Strong induction on a well-order) *Let $\prec$ be a well-ordering of a set X. Suppose $\phi(x)$ is a formula with the property that, for every $x \in X$, if $\phi(y)$ holds for every $y \prec x$ then $\phi(x)$ holds. Then $\phi(x)$ holds for all $x \in X$.*

Proof. Let $S = \{x \in X : \neg\phi(x)\}$. Suppose S is non-empty. Then it has a least element, s. Therefore, $\phi(y)$ holds for every $y \in X$ with $y \prec s$. But then $\phi(s)$ holds, which contradicts $s \in S$. So S cannot have a least element, and must be empty. □

There is also a Recursion principle for well-orders, but we do not state this now as we will have a more definitive form later, along with a "transfinite induction principle".

We start our study of well-orders with some basic properties. One has a notion of order-preserving function between ordered sets X and Y. We could write $\prec_X$ for the order on an ordered set X , and $\prec_Y$ for the order on Y, etc, but often we will write the orders on the various sets using the same symbol $\prec$.

11.5. Definition. Let X, Y be strictly totally ordered sets. An *order embedding* is a function

$$f : X \to Y$$

such that $f(u) \prec f(v)$ (in Y) whenever $u \prec v$ (in X).

Such a map is an injection (check!), but need not be a bijection: e.g. $\mathbb{Q} \subset \mathbb{R}$. An order embedding $f : X \to Y$ is called an *order isomorphism* if f is a bijection.

11.6. Proposition. *Let X, Y, Z be strictly totally ordered sets. If $f : X \to Y$ is an order isomorphism then its inverse is also an order isomorphism. If $f : X \to Y$ and $g : Y \to Z$ are order isomorphisms then the composition $g \circ f : X \to Z$ is an order isomorphism.*

Proof. Easy check. □

On the real (or rational) numbers one has many order isomorphisms such as $x \mapsto ax + b$ for any positive reals (rationals) a, b, or more complicated functions. However order isomorphisms between well orders are very restricted.

If X is a well ordered set, then any subset $Z \subseteq X$ is also well-ordered by the restriction of the ordering on X to Z.

11.7. Definition. Let X be a set well ordered by $\prec$. An *initial segment* of X is a subset $Z \subseteq X$ such that if $z \in Z$ and $w \in X$ with $w \prec z$ then $w \in Z$. For $x \in X$ define

$$\mathrm{Seg}_X x = \{z \in X : z \prec x\}.$$

11.8. Proposition. *Let X be a well-ordered set. Then*

(i) X is an initial segment of X
(ii) $\mathrm{Seg}_X x$ *is an initial segment of X for any $x \in X$*
(iii) If Z is an initial segment of X, either $Z = X$ or $Z =$ $\mathrm{Seg}_X x$ *for some $x \in X$.*

Proof. Statements (i) and (ii) are easy to check. For (iii), if $Z \neq X$ then $X - Z$ (being non-empty) has a least element x and it is easy to check that $Z = \{x \in X : x \prec z\}$. □

11.9. Definition. If X is a well-ordered set and $Z \subseteq X$ we will say that x is an *upper bound* for Z if $z \prec x$ for all $z \in Z$. If Z has an upper bound we will say it is *bounded above in X*. If Z is bounded above in X then the set of (strict) upper bounds for Z in X is non-empty, and has a least element. This element we call the *least upper bound* or *supremum* of Z in X and denote $\mathrm{Sup}_X Z$ or just $\mathrm{Sup} Z$ if X is implicit.

Note: the way we have defined them, upper bounds are **strict**. Unlike what is done in analysis.

11.10. Theorem. *Let X and Y be well-ordered sets. Then either*

1. *There is a unique order embedding f from X onto an initial segment of Y, or*
2. *There is a unique order embedding from Y onto an initial segment of X.*

Proof. (From [11]). Call an order-embedding from an initial segment A of X onto an initial segment B of Y a *good* function on A. We use $\prec$ for the well-order on both X and Y, and we use $\preceq$ for "$\prec$ or equal to". We wrote $y \succ x$ for $x \prec y$.

Claim 1. If f is a good function then, for all x in its domain, $f(x) = \mathrm{Sup}\{f(y) : y \prec x\}$.

Proof of claim 1. Since f is an order embedding we have $f(y) \prec f(x)$ for any $y \prec x$, so $f(x) \succeq t = \mathrm{Sup}\{f(y) : y \prec x\}$. If $f(x) \succ t$ then t is in the range of f (which is an initial segment), so $t = f(y)$ for some $y \prec x$, as f is an order embedding. But this contradicts the definition of t as a strict upper bound, so we must have $f(x) = t$. Note this holds also for the least element $x \in X$, in which case $\{f(y) : y \prec x\}$ is empty. □

Claim 2. For well-ordered X, Y there is at most one good function from X into Y.

Proof of claim 2. Suppose f, g are good functions from X onto initial segments of Y. If $f \neq g$ then the set $S = \{x \in X : f(x) \neq g(x)\}$ is non-empty and has a least element x. But then $f(x) = \text{Sup}\{f(y) : y \prec x\} = \text{Sup}\{g(y) : y \prec x\} = g(x)$, a contradiction. □

Claim 3. The restriction of a good function to an initial segment Z is a good function. Hence, the image of $f|_Z$ is an initial segment of Y.

Proof of claim 3. Such a map is an order embedding, and if $x \in Z, z \prec f(x)$ then $z = f(y)$ for some y since f is good, and $y \in Z$ as Z is an initial segment. □

Now define C to be the set of $x \in X$ such that there exists a good function from $\{y \in X : y \preceq x\}$ into Y. If $x \in C$ and $y \in X$ with $y \prec x$, and f is good function from $\{y \in X : y \preceq x\}$ into Y then the restriction of f to $\{z \in X : z \preceq y\}$ is a good function. So $y \in C$, and so C is an initial segment.

For each $x \in C$ denote by f_x the *unique* (by Claim 2) good function from $\{z \in X : z \preceq x\}$ into Y. Then $f_x(z) = f_y(z)$ whenever $z \preceq x \preceq y$. For $x \in C$ define $f(x) = f_x(x)$. If $x, y \in C$ and $x \prec y$ then $f(x) = f_x(x) = f_y(x) \prec f_y(y) = f(y)$ so f is an order embedding. If $x \in C$ and $t \prec f(x)$, then $t \prec f_x(x)$ and so there exists $y \in X$ with $y \prec x$ and $t = f_x(y) = f_y(y) = f(y)$. Thus f is a good function.

If $C = X$ we have established case (i) of the Theorem. If $C \neq X$ let $t = \text{Sup}C$ and let $D = f[C]$. If $D = Y$ then the inverse of f is a good function from Y to an initial segment of X, and we have established case (ii) of the Theorem. If $D \neq Y$, let $u = \text{Sup}D$. If we put $\overline{f}(t) = u$ we see (easy to check) that we have extended f to a good function $\overline{f}$ on $\{z \in X : z \preceq t\}$, contradicting the definition of C. So we must have (i) or (ii), and this proves the Theorem. □ □

11.11. Theorem. *Let X, Y be well-ordered sets. If there exists a good function f from X into Y and a good function g from Y into X, then both f and g are onto and they are inverse to each other.*

Proof. (ibid.) Suppose the image of f is the initial segment W of Y. Then f^{-1} is a good function from W onto X. The restriction of g to W is a good function of W into X. Therefore (11.10, Claim 2) it coincides with f^{-1}. But f^{-1} is onto X. Since g is injective on Y, we must have $W = Y$ (nowhere else for the images of other points to go), so $g = f^{-1}$ and f, g are onto. □

11.12. Definition. Let X, Y be well-ordered sets. Write $X \ll Y$ if there is a good function from X into Y, and $X \approx Y$ if there is a good function from X onto Y (so that X, Y are order isomorphic).

The first Theorem tells us that any two well-orders are comparable under $\ll$: one of them is order isomorphic to an initial segment of the other. The second Theorem says that if $X \ll Y$ and $Y \ll X$ then $X \approx Y$.

Now order-isomorphism is a (class) equivalence relation on well-ordered sets. Our theorems now tell us that the equivalence classes are (class) weakly totally ordered by $\ll$. In fact they are (class) well-ordered by the relation "X has an order embedding onto a *proper* initial segment of Y".

The theory of well-ordered sets, including transfinite induction, was developed by Cantor. Ordinals, which we now introduce, give us representatives of the equivalence classes. They were introduced by Zermelo and von Neumann. They are the backbone objects of Set Theory.

12. Ordinal numbers

12.1. Definition. A set X is called *transitive* if every element of X is a subset of X. I.e. if $x \in X$ then $x \subseteq X$.

12.2. Proposition.

1. A set X is transitive if and only if $x \in y \in X$ implies $x \in X$ (hence the term)

2. The intersection of any set of transitive sets is transitive

3. The union of any set of transitive sets is transitive.

Proof. Problem Set 1. □

12.3. Definition. (von Neumann) A set α is an *ordinal number* or *ordinal* if it is transitive and well-ordered by $\in$. (The key definition of this course.)

The property of being an ordinal may be expressed by a formula of $\mathcal{L}$. (**Exercise:** write it down). Hence the ordinals form a class, denoted **ON**.

Note. An ordinal is a well-founded set by its definition (it has an $\in$-minimal element) even without using Foundation. In particular, $\alpha \notin \alpha$ for any ordinal α. We sometimes write $<$ for the well-order $\in$, as it is easier to think of the former relation as an order.

12.4. Proposition. *Let α and β be ordinals.*

1. An element of α is an ordinal

2. A transitive proper subset $y \subset \alpha$ is equal to the least element of $\alpha - y$ (and hence is an ordinal)

3. Either $\alpha \subseteq \beta$ or $\beta \subseteq \alpha$

4. We have $\alpha \subseteq \beta$ if and only if $\alpha \in \beta$ or $\alpha = \beta$

5. We have $\alpha \in \beta$ or $\alpha = \beta$ or $\beta \in \alpha$. So $\in$ strictly totally orders ordinals.

Proof. Let α be an ordinal.

1. Suppose $z \in \alpha$. Suppose $x \in y \in z$. Since α is an ordinal, $y \in \alpha$ and so also $x \in \alpha$. The relation $\in$ is transitive on α. Since $x \in y \in z, x, y, z \in \alpha$ we have $x \in z$. So z is transitive. Since $z \subseteq \alpha$ it is well-ordered by $\in$.

2. Let u be the least element of α not in y. Thus if $x \in u$ then $x \in y$. So $u \subseteq y$. Suppose $x \in y$. Then $x \in \alpha$, so we have either $x \in u, x = u, u \in x$. Either $x = u, u \in x$ imply $u \in y$ (by transitivity), a contradiction. So $x \in u$ and so $y \subseteq u$. So $y = u \in \alpha$.

3. Suppose neither holds. Then $y = \alpha \cap \beta$ is a transitive proper subset of both α and β. Let γ be the least element of $\alpha - y$ and δ the least element of $\beta - y$. By 12.4.2 we have $\gamma = y = \delta$, so that $\gamma \in \alpha \cap \beta = y$, a contradiction.

4. First $\Leftarrow$. If $\alpha = \beta$ the implication is obvious, while if $\alpha \in \beta$ then $\alpha \subseteq \beta$ by transitivity. Next $\Rightarrow$. Suppose $\alpha \subseteq \beta$ but $\alpha \neq \beta$. Then α is a transitive proper subset of β. Let $\gamma \in \beta - \alpha$ be least. Then $\alpha = \gamma \in \beta$.

5. By 12.4.3 and 12.4.4. □

From 12.4.6 we see that an ordinal α is equal to the set of all ordinals smaller then (for ϵ) it.

12.5. Proposition.

1. The empty set is the smallest ordinal

2. The successor of an ordinal is an ordinal

3. The union of any set of ordinals is an ordinal.

Proof.

1. The conditions for being an ordinal hold (vacuously) for $\emptyset$. Also $\emptyset$ is a transitive proper subset of any other ordinal. So if $\alpha \neq \emptyset$ then $\emptyset \in \alpha$ by 12.4.6.

2. We must show $\alpha^+ = \alpha \cup \{\alpha\}$ is an ordinal. First we show α^+ is transitive (done in Problem Set 1). Suppose $\beta \in \alpha^+$. Then $\beta \in \alpha$ or $\beta = \alpha$. In the first case $\beta \subseteq \alpha \subseteq \alpha^+$. In the second case $\beta = \alpha \subseteq \alpha^+$. This gives transitivity. Now suppose $S \subseteq \alpha^+$ is non-empty. If $S \cap \alpha$ is non-empty, then it has a least element β as α is an ordinal. Then $\beta \in \alpha$, the only possible element of S outside of α, and so β is least in S. If $S \cap \alpha$ is empty then $S = \{\alpha\}$ has least element α.

3. Let A be a set of ordinals and let $U = \bigcup A$. We must show U is an ordinal. Suppose $x \in y \in U$. Then $y \in a$ for some $a \in A$, so $x \in a$, and so $x \in U$. Thus (as done in Problem Set 1), U is transitive. Let $\alpha, \beta, \gamma \in U$. Then $\alpha \notin \alpha$, since α is an ordinal. And if $\alpha \in \beta$ and $\beta \in \gamma$ then (as γ is an ordinal) certainly $\alpha \in \gamma$. So $\in$ is a strict partial order on U. Next, by 12.4.6, either $\alpha \in \beta$ or $\beta \in \alpha$ or $\alpha = \beta$, so U is strictly totally ordered by $\in$.

Let S be a non-empty subset of U. Suppose $\alpha \in S$. It may be that α is the least element of S. If not, the set $S \cap \alpha$ is non-empty, and has a least element β. Then β is the least element of S (independent of the choice of α). For if $\gamma \in S$ is any element we must have $\alpha < \gamma$ or $\alpha = \gamma$ or $\gamma < \alpha$. In each case it is clear that $\beta \leq \gamma$. □

12.6. Corollaries.

1. Each $n \in \omega$ is an ordinal, and $\omega = \bigcup_{n \in \omega} n$ is an ordinal.

2. The sets $\omega^+ = \omega \cup \{\omega\}, \omega^{++}, \ldots$ are ordinals. The set

$$\alpha = \bigcup \{\omega, \omega^+, \omega^{++}, \omega^{+++}, \ldots\}$$

(which exists by Recursion) is an ordinal. The sets $\alpha^+, \alpha^{++}, \ldots$ are ordinals, etc.

Proof. 1. This follows from 12.5.1 and 12.5.2, but also from earlier results and problem sets. In problem sets you showed: $\emptyset$ is transitive, and the successor of a transitive set is transitive. Then by induction every $n \in \omega$ is transitive. Now ω is the set of natural numbers, each $n \in \omega$ is the set of smaller natural numbers. So if $m \in n \in \omega$ then $m \in \omega$, so ω is transitive. We have proved (4.13) that ω and its elements are well-ordered by

$\in$. We proved it is a total order on ω and that a non-empty subset of ω has a least element (statement equivalent to "Strong induction").

2. These also follow by 12.5.1, 12.5.3. □

The following result tells us that ordinals can be thought of as "numbers" representing the order-types of well-ordered sets. Order-types go back to Cantor; the ordinals numbers as representatives of them were introduced by von Neumann.

12.7. Theorem. *Let X be a well-ordered set. Then there exist unique α and f such that α is an ordinal and f is an order isomorphism from X onto α.*

Proof. (from Cohen) We may assume X is non-empty, or X is uniquely order-isomorphic to $\emptyset$ by the empty function.

Claim 1. The set $\{\emptyset\}$ is the unique ordinal with one element.

Pf. First it is an ordinal, being $\emptyset^+$. Suppose $x = \{a\}$ is an ordinal. Suppose $b \in a$. Then $b \in x$ by transitivity, but $b \neq a$ since $\in$ is a well-order on x. So $a = \emptyset$. □

We have already shown in 11.10 that, for any α, there can be at most one order-isomorphism of X with α. We define A to be the set of all $x \in X$ such that, for all $t \in X, t \leq x$ the initial segment $I_t = \{y \in X : y \leq t\}$ has an order-isomorphism f_t onto a unique ordinal α_t. Then A is non-empty, since if x is the least element of X then $\{x\}$ is uniquely isomorphic to $\{\emptyset\}$, and the latter is the unique such ordinal.

Clearly A is an initial segment. By Replacement the union of the ranges of the $f_x, x \in A$ is a set, which we call α. Being a union of ordinals, α is an ordinal. Then f defined by $f(x) = f_x(x)$ is an order isomorphism from A onto α.

If $g : A \to \beta$ is another order isomorphism with an ordinal β, then for each $x \in A$ the restriction of g to I_x is an order isomorphism of I_x onto an initial segment J of β, so J is an ordinal and $g(x) = f(x)$ by the defining property of A. So α and f are unique.

We must show finally that $A = X$. If it is not, set $y = \mathrm{Sup} A$ and define $\overline{f}$ on $A \cup \{y\}$ by setting $\overline{f}(x) = f(x)$ for $x \in A$ and $\overline{f}(y) = \alpha$. Then $\overline{f}$ is an order isomorphism of $A \cup \{y\}$ onto $\alpha \cup \{\alpha\}$. It is readily checked that such $\overline{f}$ is unique. (This includes the claim that, for any ordinal α, α^+ is the unique ordinal with $\alpha^+ - \alpha$ a singleton.) Then $y \in A$ is a contradiction. □

Thus every well-ordered set is order isomorphic to a unique ordinal (with its $\in$-ordering), its *order type*.

12.8. Corollary. *Ordinals α, β are order-isomorphic iff they are equal.* □

By Proposition 12.4 the class **ON** of ordinals is (class) strictly ordered by $\in$.

12.9. Definition. An ordinal α is called a *successor ordinal* if $\alpha = \beta^+$ for some ordinal β. An ordinal which is neither 0 nor a successor ordinal is called a *limit ordinal.*

Example. $1, 2, 3, \ldots$ are successor ordinals, ω is a limit ordinal. $\omega^+, \omega^{++}, \ldots$ are successor ordinals, α (in 12.6.2 above) is a limit ordinal (takes a short proof, see 12.11.2).

12.10. Theorem. *An ordinal λ is a limit ordinal iff $\lambda \neq 0$ and for all $\alpha \in \lambda, \alpha^+ \in \lambda$.*

Proof. $\Rightarrow$ Suppose λ limit. Then $\lambda \neq 0$. Suppose $\alpha \in \lambda$. Then $\alpha^+ \subseteq \lambda$. But λ is not a successor ordinal, so $\alpha^+ \neq \lambda$. So we must have $\alpha^+ \in \lambda$.

$\Leftarrow$ First, $\lambda \neq 0$. If λ were a successor ordinal, say $\lambda = \beta^+$, then $\beta \in \lambda$, but then $\beta^+ \in \lambda$, contradicting $\lambda = \beta^+$. □

12.11. Theorem. *Let A be a non-empty set of ordinals.*

1. *If A has a greatest element α then $\bigcup A = \alpha$*
2. *Otherwise, $\bigcup A$ is a limit ordinal and $\bigcup A = \mathrm{Sup} A$.*

Proof. Suppose A has a greatest element α. So for all $\beta \in A$ we have $\beta \subseteq \alpha$, and $\bigcup A \subseteq \alpha$. Clearly $\alpha \subseteq \bigcup A$.

Suppose A has no greatest element. Suppose $\beta \in \bigcup A$. Then there exists $\gamma \in A$ with $\beta \in \gamma$. Since A has no biggest element, $\gamma \in \delta$ for some $\delta \in A$. Then $\gamma^+ \subseteq \delta$. So $\beta^+ \subseteq \gamma \in \delta$, whence $\beta^+ \in \delta \subseteq \bigcup A$, and so $\bigcup A$ (being non-empty) is a limit ordinal. By the above, it is a least upper bound for A, and is least as any (strict) upper bound for A is a superset of $\bigcup A$. □

12.12. Corollary. *An ordinal λ is a limit ordinal iff it is non-empty and $\lambda = \bigcup \lambda$, i.e. $\lambda = \bigcup \{\alpha : \alpha \in \lambda\}$. For any ordinal (or transitive set) α^+, $\bigcup \alpha^+ = \alpha$.* □

Big ordinals

12.13. Theorem (Hartogs's Theorem). *Let X be a set. There exists an ordinal α such that $|\alpha| \not\leq |X|$. I.e. the cardinality of α is not less than or equal to that of X.*

Proof. Let Z be the set of well orderings $<$ of subsets Y of X (which exists by Comprehension over $\mathcal{P}(X \times X)$). Any element of Z is then order-isomorphic to some (unique) ordinal $\beta_<$. Let

$$\alpha = \bigcup \{\beta_<^+ : < \in Z\}$$

which is a set by Replacement (crucial use of ZF8). We show that $|\alpha| \not\leq |X|$. If not, there is an injective function $f : \alpha \to X$. Let $Y = \mathrm{Range} f$. Use f to define a well-order on Y: $f(\gamma) < f(\delta)$ iff $\gamma < \delta$. Then $(Y, <)$ is order isomorphic to α. Then $\alpha \in \alpha^+ \subseteq \alpha$ implies $\alpha \in \alpha$, contradiction. □

12.16. Corollary. *There exists an uncountable ordinal.*

Proof. Apply Hartogs's theorem to $X = \omega$. □

The smallest uncountable ordinal is denoted ω_1.

13. Transfinite induction and recursion

We already observed that the class of ordinals is (class) strictly ordered by $\in$. This allows us to do induction not just on individual ordinals but on **ON**.

13.1. Theorem. *Let $\phi(x)$ be a formula such that $\phi(\alpha)$ holds for some ordinal α. Then there is a least ordinal β for which $\phi(\beta)$ holds.*

This might be rephrased as: *A non-empty class of ordinals has a least element.* I.e. **ON** is (class) well-ordered by $\in$. Or: there is no infinite strictly decreasing sequence of ordinals.

Proof. It may be that α is the least ordinal for which $\phi(x)$ holds. Otherwise, $\phi(x)$ holds for some element of α. So the set $\{\gamma \in \alpha : \phi(\gamma)\}$ is non-empty, and has a least element β. Suppose δ is any ordinal for which $\phi(\delta)$ holds. Either $\alpha \leq \delta$, whence $\beta < \delta$, or $\delta \in \alpha$, in which case $\beta \leq \delta$. □

13.2. Theorem (Strong transfinite induction). *Suppose $\phi(x)$ is a formula such that, for every ordinal α, if $\phi(\beta)$ holds for all $\beta < \alpha$, then $\phi(\alpha)$. [So $\phi(0)$ holds.]*

Then $\phi(\alpha)$ holds for all ordinals α.

Proof. If not, $\neg\phi(x)$ holds for some ordinal, and hence for some least ordinal α. But then $\phi(\beta)$ holds for all $\beta < \alpha$, so that $\phi(\alpha)$ holds. Contradiction. □

Recall that a class function $\mathbf{F} : \mathbf{X} \to \mathbf{Y}$ is a formula $\phi(x, y)$ of $\mathcal{L}$ with the property that for each set $x \in \mathbf{X}$ there is a unique set $y \in \mathbf{Y}$ such that $\phi(x, y)$ holds.

13.3. Theorem (Transfinite recursion on ON). *Let x_0 be a set and $\mathbf{G} : \mathbf{V} \to \mathbf{V}$ a class function. Then there is a unique class function*

$$\mathbf{F} : \mathbf{ON} \to \mathbf{V}$$

with

(1) $\mathbf{F}(0) = x_0$ *and*

(2) $\mathbf{F}(\alpha) = \mathbf{G}\big(\mathbf{F}|_\alpha\big)$ *for every ordinal $\alpha > 0$.*

Note. One could set $\mathbf{G}(\emptyset) = x_0$ and just say $\mathbf{F}(\alpha) = \mathbf{G}\big(\mathbf{F}|_\alpha\big)$ for all α; the form above is intended to look more like the previous versions.)

Observe. The Theorem asserts the existence of a **formula** with certain properties. Restricted to any ordinal α, the formula determines a *bona fide* function on α (by Replacement), and the uniqueness means that these functions agree with each other on common domains. And this is how the proof proceeds.

Sketch proof (not examinable). This follows the proof in the case of ω, just "upgrading" everything for the class setting.

Observe that if $\mathbf{H} : \mathbf{ON} \to \mathbf{V}$ is a class function then, for any ordinal α, the restriction $\mathbf{H}|_\alpha$ is a function – its range is a set by Replacement.

A function $H : \alpha \to \mathbf{V}$ will be called α - *nice* if $\alpha = 0$ or if $H(0) = x_0$ and $H(\beta^+) = \mathbf{G}(H|_\beta)$ for every β with $\beta^+ \in \alpha$. A class function $\mathbf{H}$ is called *very nice* if, for every ordinal α, the restriction $\mathbf{H}|_\alpha$ is α - nice. The theorem says

There exists a unique very nice class function.

However, the formula giving the class function won't be unique. Then we prove the following claims simultaneously by induction:

Claim 1: For each ordinal α there exists an α-nice function.
Claim 2: An α-nice function and a β-nice function agree on their common domain.
Claim 3: (Corollary) For each α there is a unique α-nice function.
Claim 4: A very nice class function, if it exists, is unique.

Finally, the formula $\phi(x, y)$ which expresses that (x, y) is a pair belonging for some α to an α-nice function defines our very nice class function. □

As we did with recursion on ω we can give a version with additional "bells and whistles", allowing the set x_0 as a "parameter". We need this version.

13.4. Theorem (Transfinite recursion – improved version). *Let*

$$\mathbf{H} : \mathbf{V} \to \mathbf{V}, \quad \mathbf{G} : \mathbf{V} \times \mathbf{V} \to \mathbf{V}$$

be a class functions. Then there is a unique class function $\mathbf{F} : \mathbf{V} \times \mathbf{ON} \to \mathbf{V}$ *such that*
(1) $\mathbf{F}(x, 0) = \mathbf{H}(x)$
(2) $\mathbf{F}(x, \alpha) = \mathbf{G}\big(x, \mathbf{F}(x, *)|_\alpha\big)$ *for every ordinal* $\alpha > 0$.

Sketch proof (not examinable). This is just an elaboration of the previous theorem, as 5.2 is an elaboration of 5.1. □

14. Ordinal arithmetic

We use transfinite induction to define addition and multiplication on ordinals, extending the definition of these operations on ω. So these operations are class functions on $\mathbf{ON} \times \mathbf{ON}$; recall this means there is a *formula* that gives the operations. E.g. addition is given by a formula $\phi(x, y, z)$ that says "x, y, z are ordinals and $x + y = z$".

Warning. The operations of ordinal and cardinal arithmetic use the same notation $(+, .)$ but they are radically different! They agree however on natural numbers (Set 4)

14.1. Definition (Ordinal addition). For all ordinals α:

1. $\alpha + 0 = \alpha$
2. $\alpha + \beta^+ = (\alpha + \beta)^+$ for all β
3. $\alpha + \lambda = \bigcup\{\alpha + \beta : \beta < \lambda\}$, for limit λ.

Examples.
(a) $1 + \omega = \bigcup_{n\in\omega}(1 + n) = \bigcup_{m\in\omega} m = \omega$
(b) $\omega + 1 = \omega + 0^+ = (\omega + 0)^+ = \omega^+ = \omega \cup \{\omega\} \neq \omega$

14.2. Conclusion. Ordinal addition is not commutative!

14.3. Definition (Ordinal multiplication). For all ordinals α:

1. $\alpha.0 = 0$
2. $\alpha.\beta^+ = \alpha.\beta + \alpha$ for all β
3. $\alpha.\lambda = \bigcup\{\alpha.\beta : \beta < \lambda\}$ for all limit λ

Examples.

(a) $2.\omega = \bigcup_{n\in\omega} 2n = \omega$

(b) $\omega.2 = \omega.1 + \omega = \omega + \omega \neq \omega$

14.4. Conclusion. Ordinal multiplication doesn't commute! Also distributivity from the right fails ($(1 + 1).\omega \neq \omega + \omega$).

14.5. Definition (Ordinal exponentiation). For $\alpha > 0$:

1. $\alpha^0 = 1$
2. $\alpha^{\beta^+} = \alpha^\beta.\alpha$ for all β
3. $\alpha^\lambda = \bigcup\{\alpha^\beta : \beta < \lambda\}$ for limit λ

14.6. Theorem. *If α, β are countable ordinals then α^β is countable.*

Proof. Problem Set 4. □

We won't deal with ordinal exponentiation very much, so no confusion should occur between the ordinal α^β and the set of functions $f : \beta \to \alpha$ with same denotation α^β. These are very different!

Remark. It is not immediately obvious how set up these definitions in terms of transfinite recursion as set out in 13.4. For addition, say, one would define $\mathbf{F}(x, \alpha)$ to be the graph of $y \mapsto y + \alpha$ on $y \in x$, with $\mathbf{G}$ being the function that extends such a graph to x in the appropriate way (depending on whether x is a limit or successor).

Ordinal arithmetic can be demystified as follows. Imagine a queue of order-type ω. Putting a new person at the *front* of the queue does not change the order-type of the queue ($1 + \omega = \omega$), but putting a new person at the *back* of the queue does ($\omega + 1 = \omega^+$). Problem 2, Set 4 develops this idea, which may be useful in Problem 4.

Ordinal addition $\alpha + \beta$ can be defined as the order type of a suitable order on a disjoint union $X \cup Y$ of ordered sets of order types α, β: the elements of X precede all the elements of Y. Inside X and Y elements retain their orders. Call this the *sum* of the orders on X, Y. The sum of well orders is a well-order (Problem Set 4).

14.7. Proposition. *Let X, Y be disjoint sets, α, β ordinals. Suppose $(X, <_X)$ has order type α and $(Y, <_Y)$ has order type β. Then $X \cup Y$ with the sum order has order type $\alpha + \beta$.*

Proof. Fix α and prove by induction on β. If $\beta = 0$ then $Y = \emptyset$ and the assertion holds. Suppose $\beta = \delta^+$. Let $\phi_Y : Y \to \delta \cup \{\delta\}$ be the (unique) order isomorphism and $y = \phi_Y^{-1}(\delta)$ the "last" element of Y. Put $Z = Y - \{y\}$. Then Z with the induced ordering has order type δ (by the restriction of ϕ), and the order type of the sum on $X \cup Z$ is $\alpha + \delta$, by induction, with a map ψ. Now we extend ψ to go from $X \cup Y$ to δ^+ by setting $\psi(y) = \delta$, and it is straightforward to check it is an order isomorphism. So suppose β is a limit. For each proper initial segment Z of Y we get an order isomorphism of $\alpha + \delta$ with some ordinal $\delta \in \beta$, order isomorphic to Z, and by uniqueness these all agree with each other on common domains. The union is an order isomorphism of $X \cup Y$ onto $\alpha + \beta$. □

14.8. Theorem. *Ordinal addition is associative.*

Proof. Suppose we have ordered sets X, Y, Z of order types α, β, γ. We must show that the sum orders on $(X \cup Y) \cup Z$ and $X \cup (Y \cup Z)$ have the same order type. The identity map is easily checked to be an order isomorphism; so this follows by 14.8. □

14.9. Lemma. *Ordinal addition is strictly monotonic in the second argument.*

Proof. By (transfinite) induction. □

14.10. Corollary. $\alpha + \beta \geq \beta$ *for all ordinals* α, β.

Proof. By induction. □

14.11. Theorem. *Suppose* $\alpha \leq \beta$. *Then there exists* γ *such that* $\alpha + \gamma = \beta$.

Proof. Write $A = \alpha, B = \beta$. So $A \subseteq B$. Let $C = B - A$. Then C is well-ordered as a subset of B, and is isomorphic to some ordinal γ. The order on B is just the sum of A and C. Thus $\alpha + \gamma = \beta$. □

Multiplication arises from a (reverse) lexicographic (*product*) order on $X \times Y$. One sets

$$(x, y) \prec (x', y') \text{ iff } y \prec_Y y', \text{ or } y = y' \text{ and } x \prec_X x'.$$

You can check this is a well-order (Problem Set 4). Then $2.\omega$ amounts to ordering $\{0, 1\} \times \omega$ by taking (e.g.) a queue of people of order-type ω in which every person is joined by a partner who slots in behind them. The new queue has order type $2.\omega = \omega$. But $\omega.2 = \omega + \omega$, ordering $\omega \times \{0, 1\}$, is a queue of order-type ω followed by another queue of order-type ω.

14.12. Proposition. *If* X *and* Y *have order types* α, β *then the product order on* $X \times Y$ *has order type* $\alpha.\beta$.

Proof. By induction on β for fixed α. □

14.13. Theorem. *Ordinal multiplication is associative.*

Proof. Let X, Y, Z have order types α, β, γ. We need to show that the product orders on $(X \times Y) \times Z$ and $X \times (Y \times Z)$ have the same order type. But this is straightforward because the natural map given by $((x, y), z) \mapsto (x, (y, z))$ is easily checked to be an order isomorphism. □

In a similar way to 14.8 one can establish the distributive law (from the left).

14.14. Theorem. $\alpha.(\beta + \gamma) = \alpha.\beta + \alpha.\gamma$ *for all ordinals* α, β, γ.

Proof. Let B, C be disjoint sets with respective order types β, γ, and A a set of order-type α. The two orderings of $A \times (B \cup C)$ agree. □

One may also check the following basic monotonicity property.

14.15. Lemma. *Suppose* $\alpha \neq 0$ *and* $\beta < \gamma$. *Then* $\alpha.\beta < \alpha.\gamma$.

Proof. By induction on β. □

Goodstein's theorem

We follow [4]. Given a positive integer N and a base (also a positive integer) b we can write N in *Cantor normal form to base* b, which is explained by the following examples:

In base $b = 5$, $97 = 3.5^2 + 4.5 + 2$

In base $b = 3$, $97 = 3^4 + 3^2 + 2.3 + 1 = 3^{3+1} + 3^2 + 2.3^1 + 1$

In base $b = 2$, $97 = 2^6 + 2^5 + 1 = 2^{2^2+2} + 2^{2^2+1} + 1$.

I.e. we expand in base b, but then do the same for all exponents, until all numbers appearing are at most b.

14.16. Definition. A *Goodstein sequence* is defined as follows. Choose a positive integer N_0.

Write N_0 in Cantor normal form to base 2.

Let N_0' be the number obtained by replacing 2 by 3 everywhere in the expression for N_0.

Let $N_1 = N_0' - 1$.

Replace 3 by 4 everywhere in the Cantor normal form for N_1, calling the resulting number N_1', and let $N_2 = N_1' - 1$. Continue to define $N_3, N_4, \ldots$ with bases $5, 6, \ldots$.

If some $N_k = 0$ then we stop and say that the sequence *terminates*.

Examples.

1. If $N_0 = 3 = 2 + 1$ we get:
$N_0' = 3 + 1 = 4$, $N_1 = 3$, $N_1' = 4$, $N_2 = 3$, $N_3 = 2$, $N_4 = 1$, $N_5 = 0$. Terminates.

2. If $N_0 = 4 = 2^2$ we have $N_0' = 3^3$,
$N_1 = 3^3 - 1 = 2.3^2 + 2.3 + 2$, $N_1' = 2.4^2 + 2.4 + 2$,
$N_2 = 2.4^2 + 2.4 + 1$, $N_2' = 2.5^2 + 2.5 + 1$, $N_3 = 2.5^2 + 2.5$ $N_3' = 2.6^2 + 2.6$,
$N_4 = 2.6^2 + 6 + 5$.
Now for several steps we just increase the base and reduce the 5:
$N_5 = 2.7^2 + 7 + 4, \ldots, N_9 = 2.11^2 + 11$.
Now $N_{10} = 2.12^2 + 11$ and we go on increasing the base and reducing the 11:
$N_{21} = 2.23^2$, $N_{22} = 24^2 + 23.24 + 23$,
Now for a while we reduce the 23 and increase the base for a while:
$N_{381} = 383^2 + 20.383, \ldots$

14.17. Theorem. (Goodstein, 1944) *Every Goodstein sequence eventually terminates.*

Proof. Sketch, non-examinable. By induction! To each N_i in base $b = i + 2$ we associate an ordinal α_i by replacing b_i by ω. Thus in our second example:

$\alpha_0 = \omega^\omega$,

$\alpha_1 = \omega^2.2 + \omega.2 + 2$,

$\alpha_2 = \omega^2.2 + \omega.2 + 1$,

$\vdots$

$\alpha_5 = \omega^2.2 + \omega + 4$,

$\vdots$

$\alpha_{21} = \omega^2.2$, $\alpha_{22} = \omega^2 + \omega.23 + 23$,

$\vdots$

$\alpha_{381} = \omega^2 + \omega.20$, ...

and this is a strictly decreasing sequence of ordinals. So it terminates (at 0). □

Example. In the descending sequence of ordinals, the choice of an ordinal beneath a limit ordinal arising is determined by the base in use at that point. If we started with $N_0 = 2^{2^2}$, with $\alpha_0 = \omega^{\omega^\omega}$, we have $N_0' = 3^{3^3}$ and

$$N_1 = 3^{3^3} - 1 = 3^{3^3-1} + 3^{3^3-2}.2 + \ldots + 2$$

$$= 3^{3^2+3.2+2} + 3^{3^2+3.2+1}.2 + 3^{3^2+3.2}.2 + 3^{3^2+3+2}.2 + 3^{3^2+3+1}.2 + 3^{3^2+3}.2 + 3^{3^2+2}.2 + \ldots + 2$$

Thus

$$\alpha_1 = \omega^{\omega^2+\omega.2+2} + \omega^{\omega^2+\omega.2+1}.2 + \omega^{\omega^2+\omega.2}.2 + \omega^{\omega^2+\omega+2}.2 + \ldots + 2$$

The ordinals involved are all less than ϵ_0, which is the supremum of $\omega, \omega^\omega, \omega^{\omega^\omega}, \ldots$ that is "$\epsilon_0 = \omega^{\omega^{\omega^{\cdots}}}$" which is an important ordinal (the proof-theoretic ordinal of Peano Arithmetic). It looks large, but it is countable! Properly set-up, the proof of 14.17 is by induction on ϵ_0.

15. The Axiom of Choice

In this section we introduce a further axiom, the Axiom of Choice. Everything proved so far has of course been proved without it, but some natural statements turn out to require it – indeed to be equivalent to it.

With AC we get a fairly clear picture of ordinals and cardinals. The axiom was controversial in the first part of the twentieth century, and some mathematicians are still uncomfortable with it, but it is now generally accepted.

We have not proved that the cardinalities of any two sets are comparable. One would naturally like to have the following hold.

Cardinal comparability (CC): *If X, Y are sets then $|X| \leq |Y|$ or $|Y| \leq |X|$.*

(CC is also called: the "dichotomy principle".) Ordinal numbers were introduced with the idea of "enumerating" sets. One would therefore like to have the following.

Well-ordering principle (WO): *Every set X may be well-ordered.*

15.1. Theorem. *Assume WO. Then every set is equinumerous with an ordinal.*

Proof. Let X be a set. By WO there is a well-order $<$ on X. By Theorem 12.7 the well-ordered set X is order-isomorphic to a (unique) ordinal α. The order-isomorphism is a bijection, so $X \sim \alpha$. □

15.2. Corollary. *WO implies CC.*

Proof. Let X, Y be sets. By the Theorem we have $X \sim \alpha, Y \sim \beta$ for ordinals α, β. But ordinals are comparable: either $\alpha \leq \beta$ or $\beta \leq \alpha$, and composing the relevant maps gives $X \preceq Y$ or $Y \preceq X$. □

15.3. Theorem. *CC implies WO.*

Proof. Assume CC. Let X be any set. By Hartogs's theorem 12.15 there is an ordinal α such that $\alpha \not\preceq X$. By CC, $X \preceq \alpha$. Let $f : X \to \alpha$ be an injection. Define $<$ on X by setting $x < y$ iff $f(x) < f(y)$. This is a well-order on X. □

CC seems a reasonable statement. WO perhaps less so: it is by no means intuitive that there exists a well-ordering of $\mathbb{R}$. But (given the other axioms) the two statements are equivalent! One might introduce either one as an axiom. For historical reasons the following (also in various versions) is the statement adopted.

ZFC10 – Axiom of Choice (AC): *Let X be a non-empty set of disjoint non-empty sets. Then there exists a set B such that for all $A \in X$ we have $|A \cap B| = 1$.*

Note that B containing other elements ("random set junk") is not ecluded.

The set B "chooses" one element out of every set $A \in X$, even if there may be no obvious way of doing this. As Russell described it (following Fraenkel): if X is an infinite set of pairs of shoes, one can go through and choose all the left shoes. But if they are pairs of socks, how do we choose? The Axiom of Choice can be neither proved (Cohen) nor disproved (Gödel) from the other axioms.

15.4. Definition. A *choice function* for a set X is a function $f : \mathcal{P}(X) \backslash \{\emptyset\} \to X$ such that $f(A) \in A$ for all $A \in \mathcal{P}(X) \backslash \{\emptyset\}$.

15.5. Theorem. *AC is equivalent to the statement that every set has a choice function.*

Proof. $\Rightarrow$ The elements of $\mathcal{P}(X) \backslash \{\emptyset\}$ are not disjoint. We make them so: For $A \in \mathcal{P}(X) \backslash \{\emptyset\}$ set $A^* = \{A\} \times A$. If $A_1 \neq A_2$ then $A_1^* \cap A_2^* = \emptyset$. Let $Z = \{A^* : A \in \mathcal{P}(X) \backslash \{\emptyset\}\}$. This is a set of disjoint non-empty sets. Let B be (by AC) a set with $|B \cap A^*| = 1$ for each $A^* \in Z$. The formula

$$\phi(x, y) = (x \in \mathcal{P}(X) \backslash \{\emptyset\}) \wedge (y \in x) \wedge (x, y) \in B$$

defines (by replacement) a choice function for X. (One could almost use B itself but it could have random other elements.)

$\Leftarrow$ Let Z be a set of disjoint non-empty sets. Let f be a choice function for $\bigcup Z$. Let $B = \{f(A) : A \in Z\}$. Then $|B \cap A| = 1$ for each $A \in Z$ as these sets are disjoint. □

15.6. Theorem. *AC implies WO.*

Proof. Let X be a set. Let $f : \mathcal{P}(X) \backslash \{\emptyset\} \to X$ be a choice function (by AC). Let A be a set that does not belong to X, e.g. $A = X$.

By transfinite recursion define $\mathbf{F} : \mathbf{ON} \to \mathbf{V}$ as follows. Let α be an ordinal. Write x_β for $\mathbf{F}(\beta)$. If $X \backslash \{x_\beta : \beta < \alpha\}$ is non-empty set $\mathbf{F}(\alpha) = f(X \backslash \{x_\beta : \beta < \alpha\})$. Otherwise set $\mathbf{F}(\alpha) = A$.

Suppose $\mathbf{F}(\beta) \neq A$ for all $\beta \in \alpha$. Then $\{(\beta, x_\beta) : \beta \in \alpha\}$ is a set and it is a one-to-one function from α into X.

By Hartogs's Theorem we certainly have $\mathbf{F}(\alpha) = A$ for some α. (One could choose this ordinal first and define $\mathbf{F}$ only on α.) Let α be the least such ordinal. Then $\{(\beta, x_\beta) : \beta \in \alpha\}$ gives a bijection of α with X, which we may use to define a well-order on X. □

15.7. Theorem. *WO implies AC.*

Proof. Let X be a set and $<$ a well-ordering of X. The map $f : \mathcal{P}(X)\backslash\{\emptyset\} \to X$ defined by $f(A) = \min A$ is a choice function for X. □

15.8. Corollary. *WO, CC, and AC are all equivalent (given the axioms of ZF).* □

A further equivalent statement looks a bit odd but is very useful in mathematics.

15.9. Definition.

1. Let Z be a set (of sets). A *chain* in Z is a non-empty subset C of Z with the property that, for any two elements $c_1, c_2 \in C$, either $c_1 \subseteq c_2$ or $c_2 \subseteq c_1$. I.e., C is totally ordered by $\subseteq$.

2. Let Z be a set. An element $z \in Z$ is *maximal* if is is not a proper subset of any other $z' \in Z$.

3. We say Z *satisfies the chain condition* if, whenever C is a chain in Z, we have $\bigcup C \in Z$. I.e. such Z is closed under unions of chains.

Zorn's Lemma (ZL): *A non-empty set Z which satisfies the chain condition has a maximal element.*

15.10. Theorem. *AC and ZL are equivalent (given the axioms of ZF).*

Proof. WO implies ZL: Suppose Z is a non-empty set with the chain condition. Let $<$ be a well-ordering of Z. Then $(Z, <)$ is order-isomorphic to some (unique) ordinal α, with (unique) order-isomorphism $\pi : \alpha \to Z$.

Define $f : \alpha \to \{0, 1\}$ by recursion as follows. $f(0) = 1$. Suppose $f(\gamma)$ is already defined for $\gamma < \beta$. If $\pi(\gamma) \subseteq \pi(\beta)$ for all $\gamma < \beta$ with $f(\gamma) = 1$ define $f(\beta) = 1$. Otherwise set $f(\beta) = 0$.

I claim that, for all $\beta > 0$, the set $\{\pi(\gamma) : \gamma < \beta, f(\gamma) = 1\}$ is a chain. For if $c_1 = \pi(\gamma), c_2 = \pi(\delta)$ are in this set with (say) $\gamma \leq \delta$ the definition of f requires that $c_1 \subseteq c_2$. In particular $C = \{\pi(\gamma) : \gamma < \alpha, f(\gamma) = 1\}$ is a chain. Let $z = \bigcup C$. So $z \in Z$.

I claim that z is a maximal element of Z. For suppose $z \subseteq z' \in Z$. Suppose $z' = \pi(\epsilon)$. Let $\gamma < \epsilon$. If $f(\gamma) = 1$ then $\pi(\gamma) \in C$, whence $\pi(\gamma) \subseteq z \subseteq z'$, and so $f(\epsilon) = 1$. Then $z' \in C$ and so $z' \subseteq z$. So z is maximal. (The same argument shows $f(\delta) = 1$ where $\pi(\delta) = z$.)

ZL implies AC: Problem Set 4. □

15.11. Corollary. *CC, WO, ZL, and AC are all equivalent (assuming ZF).* □

For the remainder of the course we assume AC. A typical use of Zorn's Lemma is the following (for others see Problem Set 4 and past papers).

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

Proof. Let V be a vector space over a field K. (So V, K are sets with certain distinguished elements $0_K, 1_K, 0_V$ and functions $+ : K \times K \to K$, $. : K \times K \to K$, $+ : V \times V \to V$, $K \times V \to V$ satisfying the required properties.)

Call a subset $B \subseteq V$ *linearly independent* if every finite subset of it is linearly independent in the usual sense, and *spanning* if for every $v \in V$ there exists a **finite** list of elements $v_1, \ldots, k_m \in B$ and $a_1, \ldots, a_m \in K$ such that $v = a_1 v_1 + \ldots + a_m v_m$. Call B a *basis* if it is linear independent spanning set.

Let Z be the set of all linear independent subsets of V. Then Z is non-empty as $\emptyset$ is linearly independent (if $V = \{0_V\}$ it is a basis!). Let C be a chain in Z.

Claim. $\bigcup C \in Z$.

Pf. Suppose $v_1, \ldots, v_m \in \bigcup C$. Then there exists $z_1, \ldots, z_m \in C$ such that $v_i \in z_i$. Since C is a chain, one of the z_i, say z_j, is the largest among them. Then all $v_i \in z_j$. Since z_j is linearly independent by assumption, so are the v_i. So $\bigcup C$ is linearly independent, and belongs to Z.

So ZL may be applied to Z. Let B be a maximal element.

Claim. B spans.

Pf. Suppose $v \in V$. If $v \in B$ then it is certainly in the span of B. Otherwise, since B is maximal, $B \cup \{v\}$ is not linearly independent, and has a finite subset which is linearly dependent. Since B is linearly independent, this linear dependency must involve v, and can be written $v = \sum_{i=1}^{m} a_i v_i$, $a_i \in K, v_i \in V$.

So the maximal linearly independent set B is a basis. □

Remark. You must always be sure that a set Z is non-empty in order to apply ZL!

Another example use of ZL is the following.

15.13. Theorem. *Let X be a set. Every weak partial order T on X is contained in some total order $\overline{T}$ on X.*

15.14. Proposition. *A (weak partial) order S on X is maximal iff it is total.*

Proof. Suppose S is total and $S \subset \overline{S}$. Suppose $(x, y) \in \overline{S}$. Since $x, y \in X$ and S is total we must have $(x, y) \in S$. So $S = \overline{S}$ and is maximal.

Suppose S is not total. If S does not contain all (x, x) it is not maximal, so assume it does. So we have $a, b \in X$ with $(a, b) \notin S, a \neq b, (b, a) \notin S$. Define

$$\overline{S} = S \cup \{(x, y) : (x, a) \in S \text{ and } (b, y) \in S\}.$$

Claim: $\overline{S}$ is an order. This requires checking quite a few cases. But let's observe that it is reflexive, and $(a, b) \in \overline{S}$ so S is not maximal.

First, $\overline{S}$ is transitive: suppose $(x, y), (y, z) \in \overline{S}$.

(i) if both $(x, y), (y, z) \in S$ then also $(x, z) \in S$ as its an order

(ii) If both are not in $\overline{S}$, we must have $(x,a),(b,y),(y,a),(b,z) \in S$, which imply $(x,z) \in \overline{S}$.

(iii) If $(x,y) \in \overline{S} - S$, $(y,z) \in S$ then $(x,a),(b,y),(y,z) \in S$, whence $(b,z) \in S$, whence $(x,z) \in \overline{S}$.

(iv) Exercise!

Next, $\overline{S}$ is antisymmetric. For suppose $(x,y),(y,x) \in \overline{S}$ with $x \neq y$. Then

(i) $x, y \in S$ is impossible.

(ii) $x, y \in \overline{S} - S$ implies $(x,a),(b,y),(y,a),(b,x) \in S$, implying $(b,a) \in S$, a contradiction.

(iii) if $(x,y) \in \overline{S} - S, (y,x) \in S$ then $(x,a),(b,y),(y,x) \in S$, whence $(b,x) \in S$, whence $(b,a) \in S$, contradiction!

(iv) Exercise! □

Proof of Theorem. Let

$$Z = \{S : S \text{ is an order on } X \text{ and } T \subset S\}.$$

Then $T \in Z$ so it is non-empty, and Z satisfies the chain condition (Check!). So Z contains a maximal element, and by the proposition it is total. □

16. Cardinal numbers

Assuming AC, as we now do, the theory of cardinal numbers becomes moderately well-behaved.

16.1. Definition. An infinite ordinal α is called an *initial ordinal* if it is not equinumerous with any strictly smaller ordinal.

16.2. Theorem. *Every infinite set is equinumerous with a unique initial ordinal.*

Proof. Let X be an infinite set. We know (by WO and 15.1) that X is equinumerous with an ordinal γ. Let α be the least such ordinal. Then $X \sim \alpha$. Moreover, α is an initial ordinal, for any smaller ordinal equinumerous with α would be equinumerous with X. Similarly α is the unique initial ordinal equinumerous with X (a larger ordinal equinumerous with X is equinumerous with α and so not initial). □

16.3. Theorem. *Let $\{\beta_\gamma : \gamma \in \alpha\}$ be a set of initial ordinals. Then there exists an initial ordinal exceeding all the β_γ.*

Proof. The union $\lambda = \bigcup\{\beta_\gamma : \gamma \in \alpha\}$ is an ordinal. Then $\mathcal{P}(\lambda)$ is equinumerous with some initial ordinal μ, and since $|\mathcal{P}(\lambda)| > |\lambda| \geq |\beta_\gamma|$ for all $\gamma \in \alpha$ we must have $\beta_\gamma < \mu$ for all $\gamma \in \alpha$. □

With this Theorem we may define initial ordinals ω_α for all ordinals α by transfinite recursion as follows.

16.4. Definition.

1. ω_0 is the smallest initial ordinal (i.e. $\omega_0 = \omega$)
2. ω_{α^+} is the smallest initial ordinal exceeding ω_α
3. $\omega_\lambda = \bigcup_{\alpha \in \lambda} \omega_\alpha$ for a limit ordinal λ (then ω_λ is initial; Pf: exercise).

16.5. Definition. We define $\aleph_\alpha$ to be the cardinal number of ω_α, for all α.

Thus $\aleph_0 = |\omega_0| = |\omega|$. We can in fact identify cardinal numbers with finite or initial ordinals (maintaining the notations $\aleph_\alpha, \omega_\alpha$ to distinguish whether we are considering the set in question as cardinal or ordinal – thus avoiding ambiguity when exponentiating).

16.6. Definition. A set X is called a *cardinal number* if $X \in \omega$ or $X = \omega_\alpha$ for an ordinal α.

The condition "x is a cardinal number" may be expressed by a suitable $\mathcal{L}$-formula and so cardinal numbers form a class **CN**.

16.7. Theorem. *There is no set of all ordinal numbers.*

Proof. Suppose the set O was the set of all ordinals. Then $\alpha = \bigcup O$ is an ordinal, and so is α^+. Then $\alpha \in \alpha^+ \in O$, whence $\alpha \in \bigcup O = \alpha$; contradiction. □

The "paradox" resulting from the existence of a set comprising all the ordinals is known as the Burali-Forti paradox. Like Russell's paradox it is resolved in ZFC by the fact that it is not a set. Rucker's book [3] is recommended for discussing the "size" issues. Thus **ON** is a proper class by 16.7, and (by 16.8) so is **CN**.

16.8. Corollary. *There is no set of all cardinal numbers.*

Proof. If there were such a set C then $O = \{\alpha : \aleph_\alpha \in C\}$ would be a set comprising all ordinals. □

The adoption of AC makes cardinal addition and multiplication exceedingly simple.

16.9. Theorem. *Suppose X is an infinite set. Then $X \times X \sim X$.*

Proof. Let Z be the set of all bijections $f : Y \times Y \to Y$ where $Y \subseteq X$. Then Z is non-empty because, by CC, $\aleph_0 \leq |X|$ and so X has a countable infinite subset Y, and there certainly exists a bijection $f : Y \times Y \to Y$. If C is a chain in Z then $\bigcup C \in Z$. So, by ZL, Z has a maximal element $F : Y \times Y \to Y$, and Y must be infinite.

Suppose that $|X \backslash Y| \geq |Y|$. Let $W \subseteq X \backslash Y$ be equinumerous with Y. Since $|Y| \leq 3|Y| \leq |Y|.|Y| = |Y|$, we find $|Y| = 3|Y|$ and there is a bijection $g : (Y \times W) \cup (W \times W) \cup (W \times Y) \to W$. Then $F \cup g : (Y \cup W) \times (Y \cup W) \to Y \cup W$ is a bijection, contradicting the maximality of Y.

Hence, by CC, we must have $|X \backslash Y| < |Y|$. Then $|X| = |Y| + |X \backslash Y| \leq 2|Y| \leq |Y|.|Y| = |Y| \leq |X|$. So $|X| = |Y|$, and so $X \times X \sim Y \times Y \sim Y \sim X$. □

16.10. Corollary. *Suppose κ, λ are cardinal numbers with $1 \leq \lambda \leq \kappa$ and κ infinite. Then $\kappa + \lambda = \kappa.\lambda = \kappa$.*

Proof. $\kappa \leq \kappa + \lambda \leq \kappa + \kappa = \kappa.2 \leq \kappa.\lambda \leq \kappa.\kappa = \kappa$. □

Cardinal exponentiation, on the other hand, is rather mysterious.

These remaining remarks are not part of our syllabus, but look toward the further developments. The axioms simply do not determine which $\aleph_\gamma$ is equal to $\aleph_\alpha^{\aleph_\beta}$. The simplest non-trivial exponentiation of cardinals already leads beyond ZFC.

16.11. Cantor's Continuum Hypothesis (CH). $2^{\aleph_0} = \aleph_1$.

16.12. Theorem. (Gödel)
CH is consistent with ZFC. That is, one cannot **disprove** *CH in ZFC.* □

16.13 Theorem. (Cohen)
¬CH is consistent with ZFC. That is, one cannot **prove** *CH in ZFC.* □

The same theorems, by the same people, hold with AC instead of CH; the consistency of AC and CH are studied in Part C Axiomatic Set Theory.

We know that $2^{\aleph_0} = \aleph_\alpha$ for some α. But the axioms of ZFC are insufficient to stipulate which α it has to be. There are some mild restrictions on what α can be, and subject to those one can build (starting with some model of ZFC) a model in which $2^{\aleph_0}$ has any value. These restrictions involve the notion of cofinality.

16.14. Definition. Let α be a limit ordinal. The *cofinality* of α is the least ordinal κ such that there exists a function $f : \kappa \to \alpha$ that is not bounded above below α. I.e. for all $\beta < \alpha$ there exists $\lambda < \kappa$ such that $f(\lambda) > \beta$.

It is easy to see that the cofinality of α is always an initial ordinal.

Examples.
ω has cofinality ω (also any countable limit ordinal, by means of a bijection)
ω_1 has cofinality ω_1 (if not it is a countable union of countable sets so countable)
ω_ω has cofinality ω (due to the function $\omega \to \omega_\omega$ given by $n \mapsto \omega_n$)

16.15. Theorem. *Let κ be a cardinal. The cofinality of 2^κ is greater than κ.* □

16.16. Theorem. *Let κ be a cardinal of uncountable cofinality. Then ZFC does not prove $2^{\aleph_0} \neq \kappa$.* □

Therefore the cofinality of $2^{\aleph_0}$ exceeds $\aleph_0$, and so $2^{\aleph_0}$ cannot equal $\aleph_\omega$. But that is the only (kind of) restriction: $2^{\aleph_0}$ could be (i.e. it is consistent with ZFC that it is) $\aleph_2, \aleph_3, \aleph_{100}, \aleph_{\omega+1}, \ldots$.

16.17. The Generalised Continuum Hypothesis (GCH). For all α, $2^{\aleph_\alpha} = \aleph_{\alpha+1}$.

Under GCH one can compute cardinal exponentiation in terms of cofinalities (see e.g. [8, 10.42]). And it is consistent with ZFC (Gödel [14]), i.e. cannot be disproved in ZFC. However ZFC cannot disprove any assignment $2^{\aleph_\alpha} = \beta_\alpha$ on non-limit ordinals α subject to mild restrictions on the β_α (Cohen [11] and subsequent work by others). The value of $2^{\aleph_\lambda}$ for limit ordinals λ is more restricted.

The proofs of these results require more Set Theory and Model Theory. Some mathematicians feel that there is a "true" theory of sets in which CH is either true or false (opinions differ on which alternative holds!), and that perhaps over time some additional axioms will become generally accepted and will "decide" statements such as CH one way or the other. See e.g. Woodin [25]; others feel we shall not find such axioms, see Hamkins [18].

Infinite combinatorics: Ramsey's Theorem

The following theorem (in more general form) was discovered by F. P. Ramsey in 1930. By a *graph* we mean a vertex set V and a set E of "edges", $E \subset V^{(2)}$, the set of 2-element subsets of V. So edges undirected, and between each pair of vertices there is either an edge, or not: we don't allow "multiple" edges, or "loops" on a single vertex.

16.18. Theorem. *Let V be a countably infinite set, and E an arbitrary subset of $V^{(2)}$, the set of two-element subsets of V. (So (V, E) is a graph on a countably infinite vertex set.)*

Then there exists a countably infinite subset $W \subset V$ such that the induced graph $(W, W^{(2)} \cap E)$ either has no edges, or is a complete graph.

That is $W^{(2)} \cap E$ is either empty or all of $W^{(2)}$.

Proof. Let $V_0 = V$ and choose some $v_0 \in V_0$. Now for each $v \in V_0 - \{v_0\}$, the edge $\{v_0, v\}$ either belongs to E or not, and one of these sets (at least) is infinite. Let V_1 be an infinite such set, choose $v_1 \in V_1$, and continue. This produces a sequence

$$v_0, v_1, v_2, \ldots$$

which has the following property: each adjacent pair v_i, v_{i+1} either belongs to E or not, and in either case the same holds for v_i with every subsequent v_j.

Now either "belongs to E" or "doesn't belong to E" holds for infinitely many adjacent pairs v_i, v_{i+1}. Choose one of these for which the number of pairs v_i, v_{i+1} is infinite, and take W to be the set of v_i. □

Thus inside an arbitrary graph (V, E), which has no particular structure, is an induced subgraph with a high degree of structure: "complete disorder is impossible" is the slogan of Ramsey Theory.

The above theorem has (indeed implies) a finitary version. The simplest case is: in any group of 6 people, there either exist 3 people who all know each other, or three people who all don't know each other. Here "know each other" is the relation defining the edges of the graph; we assume it is symmetric.

For every k there exists $R(k, k)$ such that, in any group of R people, there exists either a subset of k who all know each other, or a subset of k who all don't know each other. One can look for different sizes in each case, $R(k_1, k_2)$, or allow "ℓ-coloured" graphs and $R(k_1, \ldots, k_\ell)$. These numbers are very difficult to compute exactly: $R(4, 4) = 18$ but $R(5, 5)$ is unknown (Wikipedia).

References

Text to expand and supplement these notes:

1. D. Goldrei, *Classic set theory,* Chapman and Hall/CRC, Boca Raton, 1998.

Highly recommended general audience books:

2. S. Lavine, *Understanding the infinite,* Harvard, 1994.
3. R. Rucker, *Infinity and the mind,* Princeton University Press, 1995.
4. J. Stillwell, *Roads to Infinity*, CRC Press, AK Peters, 2010.

Other textbooks (more advanced) and notes:

5. P. Aczel, *Non-well founded sets,* CSLI Lecture Notes, 14, Stanford, CA. On web.
6. T. Jech, *Set theory,* Academic Press, 1978.
7. R. Knight, b1 Set Theory Lecture Notes.
8. K. Kunen, *Set theory,* North-Holland, Amsterdam, 1980.
9. A. Levy, *Basic Set Theory,* Springer, Berlin, 1979; reprinted by Dover.

Read the original texts and interpretations:

10. G. Cantor, *Contributions to the founding of the theory of transfinite numbers* (translated by P. Jourdain), Dover, 1955. p85.
11. P. Cohen, *Set theory and the continuum hypothesis,* W. A. Benjamin, 1966.
12. P. Cohen and R. Hersch, *Non-Cantorian set theory, Scientific American* **217** (1967), 104–116.
13. R. Dedekind, *Essays on the theory of numbers,* Dover, 1963.
14. K. Gödel, *The consistency of the axiom of choice and the generalized continuum hypothesis with the axioms of set theory,* Annals of Mathematics Studies 3, Princeton University Press, 1940.
15. K. Gödel, What is Cantor's continuum problem? *American Mathematical Monthly,* **54** (1964), 515–525. Revised version in Benacerraf and Putnam, *Philosophy of mathemaics: selected readings,* Prentice-Hall, 1964.

Modern perspectives, histories, and commentaries:

16. S. Feferman, *In the Light of Logic,* OUP, Oxford, 1998.
17. J. Ferreiros, On the relations between Georg Cantor and Richard Dedekind, *Historia Math.* **20** (1993), 343–363.
18. J. Hamkins, Is the ream solution of the continuum hypothesis attainable? *Notre Dame J. Symbolic Logic* **56** (2015), 135–145.
19. Y. Manin, George Cantor and his heritage, http://arxiv.org/abs/math/0209244 (just ignore the stuff about $P \neq NP$)
20. P. Maddy, *Defending the Axioms,* OUP, 2011.
21. C. McLarty, What does it take to prove Fermat's Last Theorem? Grothendieck and the logic of number theory, *Bull. Symbolic Logic* **16** (2010), 359–377.
22. G. H. Moore, *Zermelo's axiom of choice,* Springer, 1982.
23. S. Shelah, Logical dreams, *Bull. Amer. Math. Soc.* **40** (2003), 203–228.
24. J. Stillwell, The continuum problem, *Amer. Math. Monthly* **109** (2002), 286–297.
25. W. H. Woodin, Strong axioms of infinity and the search for V, *Proc. ICM Hyderabad* (2010). Available online at http://www.mathunion.org/ICM/ICM2010.1/ Also the corresponding lecture can be viewed online.

Other:

26. E. Bedford, On the logical structure of natural numbers, preprint.