1. Axioms of Set Theory

Axioms of Zermelo-Fraenkel

1.1. Axiom of Extensionality. If $X$ and $Y$ have the same elements, then $X = Y$.

1.2. Axiom of Pairing. For any $a$ and $b$ there exists a set $\{a, b\}$ that contains exactly $a$ and $b$.

1.3. Axiom Schema of Separation. If $P$ is a property (with parameter $p$), then for any $X$ and $p$ there exists a set $Y = \{u \in X : P(u, p)\}$ that contains all those $u \in X$ that have property $P$.

1.4. Axiom of Union. For any $X$ there exists a set $Y = \bigcup X$, the union of all elements of $X$.

1.5. Axiom of Power Set. For any $X$ there exists a set $Y = P(X)$, the set of all subsets of $X$.

1.6. Axiom of Infinity. There exists an infinite set.

1.7. Axiom Schema of Replacement. If a class $F$ is a function, then for any $X$ there exists a set $Y = F(X) = \{F(x) : x \in X\}$.

1.8. Axiom of Regularity. Every nonempty set has an $\in$-minimal element.

1.9. Axiom of Choice. Every family of nonempty sets has a choice function.

The theory with axioms 1.1–1.8 is the Zermelo-Fraenkel axiomatic set theory ZF; ZFC denotes the theory ZF with the Axiom of Choice.

Why Axiomatic Set Theory?

Intuitively, a set is a collection of all elements that satisfy a certain given property. In other words, we might be tempted to postulate the following rule of formation for sets.
1.10. **Axiom Schema of Comprehension (false).** If $P$ is a property, then there exists a set $Y = \{x : P(x)\}$.

This principle, however, is false:

1.11. **Russell’s Paradox.** Consider the set $S$ whose elements are all those (and only those) sets that are not members of themselves: $S = \{X : X \notin X\}$. Question: Does $S$ belong to $S$? If $S$ belongs to $S$, then $S$ is not a member of itself, and so $S \notin S$. On the other hand, if $S \notin S$, then $S$ belongs to $S$. In either case, we have a contradiction.

Thus we must conclude that

$$\{X : X \notin X\}$$

is not a set, and we must revise the intuitive notion of a set.

The safe way to eliminate paradoxes of this type is to abandon the Schema of Comprehension and keep its weak version, the **Schema of Separation**:

*If $P$ is a property, then for any $X$ there exists a set $Y = \{x \in X : P(x)\}$.*

Once we give up the full Comprehension Schema, Russell’s Paradox is no longer a threat; moreover, it provides this useful information: The set of all sets does not exist. (Otherwise, apply the Separation Schema to the property $x \notin x$.)

In other words, it is the concept of the set of all sets that is paradoxical, not the idea of comprehension itself.

Replacing full Comprehension by Separation presents us with a new problem. The Separation Axioms are too weak to develop set theory with its usual operations and constructions. Notably, these axioms are not sufficient to prove that, e.g., the union $X \cup Y$ of two sets exists, or to define the notion of a real number.

Thus we have to add further construction principles that postulate the existence of sets obtained from other sets by means of certain operations.

The axioms of ZFC are generally accepted as a correct formalization of those principles that mathematicians apply when dealing with sets.

**Language of Set Theory, Formulas**

The Axiom Schema of Separation as formulated above uses the vague notion of a *property*. To give the axioms a precise form, we develop axiomatic set theory in the framework of the first order predicate calculus. Apart from the equality predicate $=$, the language of set theory consists of the binary predicate $\in$, the *membership relation*. 
The *formulas* of set theory are built up from the *atomic formulas*

\[ x \in y, \quad x = y \]

by means of *connectives*

\[ \varphi \land \psi, \quad \varphi \lor \psi, \quad \neg \varphi, \quad \varphi \rightarrow \psi, \quad \varphi \leftrightarrow \psi \]

(conjunction, disjunction, negation, implication, equivalence), and *quantifiers*

\[ \forall x \varphi, \quad \exists x \varphi. \]

In practice, we shall use in formulas other symbols, namely defined predicates, operations, and constants, and even use formulas informally; but it will be tacitly understood that each such formula can be written in a form that only involves \(\in\) and \(=\) as nonlogical symbols.

Concerning formulas with free variables, we adopt the notational convention that all free variables of a formula

\[ \varphi(u_1, \ldots, u_n) \]

are among \(u_1, \ldots, u_n\) (possibly some \(u_i\) are not free, or even do not occur, in \(\varphi\)). A formula without free variables is called a *sentence*.

**Classes**

Although we work in ZFC which, unlike alternative axiomatic set theories, has only one type of object, namely sets, we introduce the informal notion of a *class*. We do this for practical reasons: It is easier to manipulate classes than formulas.

If \(\varphi(x, p_1, \ldots, p_n)\) is a formula, we call

\[ C = \{ x : \varphi(x, p_1, \ldots, p_n) \} \]

a *class*. Members of the class \(C\) are all those sets \(x\) that satisfy \(\varphi(x, p_1, \ldots, p_n)\):

\[ x \in C \quad \text{if and only if} \quad \varphi(x, p_1, \ldots, p_n). \]

We say that \(C\) is *definable from* \(p_1, \ldots, p_n\); if \(\varphi(x)\) has no parameters \(p_i\) then the class \(C\) is *definable*.

Two classes are considered equal if they have the same elements: If

\[ C = \{ x : \varphi(x, p_1, \ldots, p_n) \}, \quad D = \{ x : \psi(x, q_1, \ldots, q_m) \}, \]

then \(C = D\) if and only if for all \(x\)

\[ \varphi(x, p_1, \ldots, p_n) \leftrightarrow \psi(x, q_1, \ldots, q_m). \]
The *universal class*, or *universe*, is the class of all sets:

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

We define *inclusion* of classes (*C* is a *subclass* of *D*)

$$C \subset D \text{ if and only if for all } x, x \in C \text{ implies } x \in D,$$

and the following operations on classes:

$$C \cap D = \{ x : x \in C \text{ and } x \in D \},$$

$$C \cup D = \{ x : x \in C \text{ or } x \in D \},$$

$$C - D = \{ x : x \in C \text{ and } x \notin D \},$$

$$\bigcup C = \{ x : x \in S \text{ for some } S \in C \} = \bigcup \{ S : S \in C \}.$$  

Every set can be considered a class. If *S* is a set, consider the formula \( x \in S \) and the class

$$\{ x : x \in S \}.$$  

That the set *S* is uniquely determined by its elements follows from the Axiom of Extensionality.

A class that is not a set is a *proper class*.

**Extensionality**

*If X and Y have the same elements, then X = Y:*

$$\forall u (u \in X \leftrightarrow u \in Y) \rightarrow X = Y.$$  

The converse, namely, if \( X = Y \) then \( u \in X \leftrightarrow u \in Y \), is an axiom of predicate calculus. Thus we have

$$X = Y \text{ if and only if } \forall u (u \in X \leftrightarrow u \in Y).$$

The axiom expresses the basic idea of a set: A set is determined by its elements.

**Pairing**

*For any a and b there exists a set \( \{a, b\} \) that contains exactly a and b:*

$$\forall a \forall b \exists c \forall x (x \in c \leftrightarrow x = a \lor x = b).$$
By Extensionality, the set $c$ is unique, and we can define the pair

$$\{a, b\} = \text{the unique } c \text{ such that } \forall x (x \in c \leftrightarrow x = a \lor x = b).$$

The singleton $\{a\}$ is the set

$$\{a\} = \{a, a\}.$$

Since $\{a, b\} = \{b, a\}$, we further define an ordered pair

$$(a, b)$$

so as to satisfy the following condition:

$$(1.1) \quad (a, b) = (c, d) \text{ if and only if } a = c \text{ and } b = d.$$  

For the formal definition of an ordered pair, we take

$$(a, b) = \{\{a\}, \{a, b\}\}.$$  

We leave the verification of (1.1) to the reader (Exercise 1.1).

We further define ordered triples, quadruples, etc., as follows:

$$(a, b, c) = ((a, b), c),\quad (a, b, c, d) = ((a, b, c), d),$$

$$\vdots$$

$$ (a_1, \ldots, a_{n+1}) = ((a_1, \ldots, a_n), a_{n+1}).$$

It follows that two ordered $n$-tuples $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ are equal if and only if $a_1 = b_1$, ..., $a_n = b_n$.

**Separation Schema**

Let $\varphi(u, p)$ be a formula. For any $X$ and $p$, there exists a set $Y = \{u \in X : \varphi(u, p)\}$:

$$(1.2) \quad \forall X \forall p \exists Y \forall u \ (u \in Y \leftrightarrow u \in X \land \varphi(u, p)).$$

For each formula $\varphi(u, p)$, the formula (1.2) is an Axiom (of Separation). The set $Y$ in (1.2) is unique by Extensionality.

Note that a more general version of Separation Axioms can be proved using ordered $n$-tuples: Let $\psi(u, p_1, \ldots, p_n)$ be a formula. Then

$$(1.3) \quad \forall X \forall p_1 \ldots \forall p_n \exists Y \forall u \ (u \in Y \leftrightarrow u \in X \land \psi(u, p_1, \ldots, p_n)).$$
Simply let \( \varphi(u, p) \) be the formula
\[
\exists p_1, \ldots, \exists p_n \ (p = (p_1, \ldots, p_n) \text{ and } \psi(u, (p_1, \ldots, p_n))
\]
and then, given \( X \) and \( p_1, \ldots, p_n \), let
\[
Y = \{ u \in X : \varphi(u, (p_1, \ldots, p_n)) \}\).

We can give the Separation Axioms the following form: Consider the class
\[
C = \{ u : \varphi(u, p_1, \ldots, p_n) \};
\]
then by (1.3)
\[
\forall X \exists Y (C \cap X = Y).
\]
Thus the intersection of a class \( C \) with any set is a set; or, we can say even more informally
\[
a \text{subclass of a set is a set.}
\]

One consequence of the Separation Axioms is that the intersection and the difference of two sets is a set, and so we can define the operations
\[
X \cap Y = \{ u \in X : u \in Y \} \quad \text{and} \quad X - Y = \{ u \in X : u \notin Y \}.
\]
Similarly, it follows that the empty class
\[
\emptyset = \{ u : u \neq u \}
\]
is a set—the empty set; this, of course, only under the assumption that at least one set \( X \) exists (because \( \emptyset \subset X \)):
\[
(1.4) \quad \exists X (X = X).
\]
We have not included (1.4) among the axioms, because it follows from the Axiom of Infinity.

Two sets \( X, Y \) are called disjoint if \( X \cap Y = \emptyset \).

If \( C \) is a nonempty class of sets, we let
\[
\bigcap C = \bigcap \{ X : X \in C \} = \{ u : u \in X \text{ for every } X \in C \}.
\]
Note that \( \bigcap C \) is a set (it is a subset of any \( X \in C \)). Also, \( X \cap Y = \bigcap \{ X, Y \} \).

Another consequence of the Separation Axioms is that the universal class \( V \) is a proper class; otherwise,
\[
S = \{ x \in V : x \notin x \}
\]
would be a set.
Union

For any $X$ there exists a set $Y = \bigcup X$:

\begin{equation}
\forall X \exists Y \forall u (u \in Y \leftrightarrow \exists z (z \in X \land u \in z)).
\end{equation}

Let us introduce the abbreviations

$(\exists z \in X) \varphi$ for $\exists z (z \in X \land \varphi)$,

and

$(\forall z \in X) \varphi$ for $\forall z (z \in X \rightarrow \varphi)$.

By (1.5), for every $X$ there is a unique set

$$Y = \{ u : (\exists z \in X) u \in z \} = \bigcup \{ z : z \in X \} = \bigcup X,$$

the union of $X$.

Now we can define

$$X \cup Y = \bigcup \{ X, Y \}, \quad X \cup Y \cup Z = (X \cup Y) \cup Z, \quad \text{etc.,}$$

and also

$$\{ a, b, c \} = \{ a, b \} \cup \{ c \},$$

and in general

$$\{ a_1, \ldots, a_n \} = \{ a_1 \} \cup \ldots \cup \{ a_n \}.$$  

We also let

$$X \triangle Y = (X - Y) \cup (Y - X),$$

the symmetric difference of $X$ and $Y$.

Power Set

For any $X$ there exists a set $Y = P(X)$:

$$\forall X \exists Y \forall u (u \in Y \leftrightarrow u \subset X).$$

A set $U$ is a subset of $X$, $U \subset X$, if

$$\forall z (z \in U \rightarrow z \in X).$$

If $U \subset X$ and $U \neq X$, then $U$ is a proper subset of $X$.

The set of all subsets of $X$,

$$P(X) = \{ u : u \subset X \},$$

is called the power set of $X$. 
Using the Power Set Axiom we can define other basic notions of set theory. The product of $X$ and $Y$ is the set of all pairs $(x, y)$ such that $x \in X$ and $y \in Y$:

\[(1.6) \quad X \times Y = \{(x, y) : x \in X \text{ and } y \in Y\}.\]

The notation $\{(x, y) : \ldots \}$ in (1.6) is justified because

$$\{(x, y) : \varphi(x, y)\} = \{u : \exists x \exists y (u = (x, y) \text{ and } \varphi(x, y))\}.$$  

The product $X \times Y$ is a set because

$$X \times Y \subset PP(X \cup Y).$$

Further, we define

$$X \times Y \times Z = (X \times Y) \times Z,$$

and in general

$$X_1 \times \ldots \times X_{n+1} = (X_1 \times \ldots \times X_n) \times X_{n+1}.$$  

Thus

$$X_1 \times \ldots \times X_n = \{(x_1, \ldots, x_n) : x_1 \in X_1 \land \ldots \land x_n \in X_n\}.$$  

We also let

$$X^n = X \times \ldots \times X, \quad \text{n times}.$$  

An $n$-ary relation $R$ is a set of $n$-tuples. $R$ is a relation on $X$ if $R \subset X^n$. It is customary to write $R(x_1, \ldots, x_n)$ instead of

$$(x_1, \ldots, x_n) \in R,$$

and in case that $R$ is binary, then we also use

$$x \, R \, y$$

for $(x, y) \in R$.

If $R$ is a binary relation, then the domain of $R$ is the set

$$\text{dom}(R) = \{u : \exists v (u, v) \in R\},$$

and the range of $R$ is the set

$$\text{ran}(R) = \{v : \exists u (u, v) \in R\}.$$  

Note that dom$(R)$ and ran$(R)$ are sets because

$$\text{dom}(R) \subset \bigcup \bigcup R, \quad \text{ran}(R) \subset \bigcup \bigcup R.$$  

The field of a relation $R$ is the set field$(R) = \text{dom}(R) \cup \text{ran}(R)$. 
In general, we call a class \( R \) an \( n \)-ary relation if all its elements are \( n \)-tuples; in other words, if
\[
R \subseteq V^n = \text{the class of all } n\text{-tuples},
\]
where \( C^n \) (and \( C \times D \)) is defined in the obvious way.

A binary relation \( f \) is a function if \((x, y) \in f \) and \((x, z) \in f \) implies \( y = z \). The unique \( y \) such that \((x, y) \in f \) is the value of \( f \) at \( x \); we use the standard notation
\[
y = f(x)
\]
or its variations \( f : x \mapsto y, y = f_x, \) etc. for \((x, y) \in f \).

\( f \) is a function on \( X \) if \( X = \text{dom}(f) \). If \( \text{dom}(f) = X^n \), then \( f \) is an \( n \)-ary function on \( X \).

\( f \) is a function from \( X \) to \( Y \),
\[
f : X \to Y,
\]
if \( \text{dom}(f) = X \) and \( \text{ran}(f) \subseteq Y \). The set of all functions from \( X \) to \( Y \) is denoted by \( Y^X \). Note that \( Y^X \) is a set:
\[
Y^X \subseteq P(X \times Y).
\]
If \( Y = \text{ran}(f) \), then \( f \) is a function onto \( Y \). A function \( f \) is one-to-one if
\[
f(x) = f(y) \text{ implies } x = y.
\]

An \( n \)-ary operation on \( X \) is a function \( f : X^n \to X \).

The restriction of a function \( f \) to a set \( X \) (usually a subset of \( \text{dom}(f) \)) is the function
\[
f|_X = \{(x, y) : x \in X\}.
\]
A function \( g \) is an extension of a function \( f \) if \( g \supseteq f \), i.e., \( \text{dom}(f) \subseteq \text{dom}(g) \) and \( g(x) = f(x) \) for all \( x \in \text{dom}(f) \).

If \( f \) and \( g \) are functions such that \( \text{ran}(g) \subseteq \text{dom}(f) \), then the composition of \( f \) and \( g \) is the function \( f \circ g \) with domain \( \text{dom}(f \circ g) = \text{dom}(g) \) such that \((f \circ g)(x) = f(g(x)) \) for all \( x \in \text{dom}(g) \).

We denote the image of \( X \) by \( f \) either \( f^" X \) or \( f(X) \):
\[
f^" X = f(X) = \{y : (\exists x \in X) y = f(x)\},
\]
and the inverse image by
\[
f^{-1}(X) = \{x : f(x) \in X\}.
\]
If \( f \) is one-to-one, then \( f^{-1} \) denotes the inverse of \( f \):
\[
f^{-1}(x) = y \text{ if and only if } x = f(y).
\]

The previous definitions can also be applied to classes instead of sets. A class \( F \) is a function if it is a relation such that \((x, y) \in F \) and \((x, z) \in F \).
implies \( y = z \). For example, \( F^\mathcal{C} \) or \( F(\mathcal{C}) \) denotes the image of the class \( \mathcal{C} \) by the function \( F \).

It should be noted that a function is often called a mapping or a correspondence (and similarly, a set is called a family or a collection).

An equivalence relation on a set \( X \) is a binary relation \( \equiv \) which is reflexive, symmetric, and transitive: For all \( x, y, z \in X \),

\[
x \equiv x, \\
x \equiv y \text{ implies } y \equiv x, \\
\text{if } x \equiv y \text{ and } y \equiv z \text{ then } x \equiv z.
\]

A family of sets is disjoint if any two of its members are disjoint. A partition of a set \( X \) is a disjoint family \( P \) of nonempty sets such that

\[
X = \bigcup \{Y : Y \in P\}.
\]

Let \( \equiv \) be an equivalence relation on \( X \). For every \( x \in X \), let

\[
[x] = \{y \in X : y \equiv x\}
\]

(the equivalence class of \( x \)). The set

\[
X/\equiv = \{[x] : x \in X\}
\]

is a partition of \( X \) (the quotient of \( X \) by \( \equiv \)). Conversely, each partition \( P \) of \( X \) defines an equivalence relation on \( X \):

\[
x \equiv y \text{ if and only if } (\exists Y \in P)(x \in Y \text{ and } y \in Y).
\]

If an equivalence relation is a class, then its equivalence classes may be proper classes. In Chapter 6 we shall introduce a trick that enables us to handle equivalence classes as if they were sets.

**Infinity**

There exists an infinite set.

To give a precise formulation of the Axiom of Infinity, we have to define first the notion of finiteness. The most obvious definition of finiteness uses the notion of a natural number, which is as yet undefined. We shall define natural numbers (as finite ordinals) in Chapter 2 and give only a quick treatment of natural numbers and finiteness in the exercises below.

In principle, it is possible to give a definition of finiteness that does not mention numbers, but such definitions necessarily look artificial.

We therefore formulate the Axiom of Infinity differently:

\[
\exists S \left( \emptyset \in S \land (\forall x \in S)x \cup \{x\} \in S \right).
\]

We call a set \( S \) with the above property inductive. Thus we have:
Axiom of Infinity. There exists an inductive set.

The axiom provides for the existence of infinite sets. In Chapter 2 we show that an inductive set is infinite (and that an inductive set exists if there exists an infinite set).

We shall introduce natural numbers and finite sets in Chapter 2, as a part of the introduction of ordinal numbers. In Exercises 1.3–1.9 we show an alternative approach.

Replacement Schema

If a class $F$ is a function, then for every set $X$, $F(X)$ is a set.

For each formula $\varphi(x, y, p)$, the formula (1.7) is an Axiom (of Replacement):

\[
(1.7) \quad \forall x \forall y \forall z (\varphi(x, y, p) \land \varphi(x, z, p) \to y = z) \\
\to \forall X \exists Y \forall y (y \in Y \leftrightarrow (\exists x \in X) \varphi(x, y, p)).
\]

As in the case of Separation Axioms, we can prove the version of Replacement Axioms with several parameters: Replace $p$ by $p_1, \ldots, p_n$.

If $F = \{(x, y) : \varphi(x, y, p)\}$, then the premise of (1.7) says that $F$ is a function, and we get the formulation above. We can also formulate the axioms in the following ways:

If a class $F$ is a function and $\text{dom}(F)$ is a set, then $\text{ran}(F)$ is a set.
If a class $F$ is a function, then $\forall X \exists f (F|X = f)$.

The remaining two axioms, Choice and Regularity, will be introduced in Chapters 5 and 6.

Exercises

1.1. Verify (1.1).

1.2. There is no set $X$ such that $P(X) \subset X$.

Let 
\[
\mathcal{N} = \bigcap\{X : X \text{ is inductive}\}.
\]

$\mathcal{N}$ is the smallest inductive set. Let us use the following notation:

\[
0 = \emptyset, \quad 1 = \{0\}, \quad 2 = \{0, 1\}, \quad 3 = \{0, 1, 2\}, \quad \ldots .
\]

If $n \in \mathcal{N}$, let $n + 1 = n \cup \{n\}$. Let us define $<$ (on $\mathcal{N}$) by $n < m$ if and only if $n \in m$.

A set $T$ is transitive if $x \in T$ implies $x \subset T$. 

1.3. If $X$ is inductive, then the set $\{x \in X : x \subset X\}$ is inductive. Hence $N$ is transitive, and for each $n$, $n = \{m \in N : m < n\}$.

1.4. If $X$ is inductive, then the set $\{x \in X : x$ is transitive$\}$ is inductive. Hence every $n \in N$ is transitive.

1.5. If $X$ is inductive, then the set $\{x \in X : x$ is transitive and $x \not\in x\}$ is inductive. Hence $n \not\in n$ and $n \neq n + 1$ for each $n \in N$.

1.6. If $X$ is inductive, then $\{x \in X : x$ is transitive and every nonempty $z \subset x$ has an $\in$-minimal element$\}$ is inductive ($t$ is $\in$-minimal in $z$ if there is no $s \in z$ such that $s \subset t$).

1.7. Every nonempty $X \subset N$ has an $\in$-minimal element.

[Pick $n \in X$ and look at $X \cap n$.]

1.8. If $X$ is inductive then so is $\{x \in X : x = \emptyset$ or $x = y \cup \{y\}$ for some $y\}$. Hence each $n \neq 0$ is $m + 1$ for some $m$.

1.9 (Induction). Let $A$ be a subset of $N$ such that $0 \in A$, and if $n \in A$ then $n + 1 \in A$. Then $A = N$.

A set $X$ has $n$ elements (where $n \in N$) if there is a one-to-one mapping of $n$ onto $X$. A set is finite if it has $n$ elements for some $n \in N$, and infinite if it is not finite.

A set $S$ is $T$-finite if every nonempty $X \subset P(S)$ has a $\subset$-maximal element, i.e., $u \in X$ such that there is no $v \in X$ with $u \subset v$ and $u \neq v$. $S$ is $T$-infinite if it is not $T$-finite. (T is for Tarski.)

1.10. Each $n \in N$ is $T$-finite.

1.11. $N$ is $T$-infinite; the set $N \subset P(N)$ has no $\subset$-maximal element.

1.12. Every finite set is $T$-finite.

1.13. Every infinite set is $T$-infinite.

[If $S$ is infinite, consider $X = \{u \subset S : u$ is finite$\}$.]


[Given $\varphi$, let $F = \{(x, x) : \varphi(x)\}$. Then $\{x \in X : \varphi(x)\} = F(X)$, for every $X$.]

1.15. Instead of Union, Power Set, and Replacement Axioms consider the following weaker versions:

(1.8) $\forall X \exists Y \bigcup X \subset Y$, i.e., $\forall X \exists Y (\forall x \in X)(\forall u \in x) u \in Y$,

(1.9) $\forall X \exists Y \bigcap P(X) \subset Y$, i.e., $\forall X \exists Y \forall u (u \subset X \rightarrow u \in Y)$,

(1.10) If a class $F$ is a function, then $\forall X \exists Y F(X) \subset Y$.

Then axioms 1.4, 1.5, and 1.7 can be proved from (1.8), (1.9), and (1.10), using the Separation Schema (1.3).
Historical Notes

Set theory was invented by Georg Cantor. The first attempt to consider infinite sets is attributed to Bolzano (who introduced the term *Menge*). It was however Cantor who realized the significance of one-to-one functions between sets and introduced the notion of cardinality of a set. Cantor originated the theory of cardinal and ordinal numbers as well as the investigations of the topology of the real line. Much of the development in the first four chapters follows Cantor’s work. The main reference to Cantor’s work is his collected works, Cantor [1932]. Another source of references to the early research in set theory is Hausdorff’s book [1914].

Cantor started his investigations in [1874], where he proved that the set of all real numbers is uncountable, while the set of all algebraic reals is countable. In [1878] he gave the first formulation of the celebrated Continuum Hypothesis.

The axioms for set theory (except Replacement and Regularity) are due to Zermelo [1908]. The Replacement Schema is due to Fraenkel [1922a] and Skolem (see [1970], pp. 137–152).

Exercises 1.12 and 1.13: Tarski [1925a].
2. Ordinal Numbers

In this chapter we introduce ordinal numbers and prove the Transfinite Recursion Theorem.

Linear and Partial Ordering

Definition 2.1. A binary relation $<$ on a set $P$ is a partial ordering of $P$ if:

(i) $p \not< p$ for any $p \in P$;
(ii) if $p < q$ and $q < r$, then $p < r$.

$(P, <)$ is called a partially ordered set. A partial ordering $<$ of $P$ is a linear ordering if moreover

(iii) $p < q$ or $p = q$ or $q < p$ for all $p, q \in P$.

If $<$ is a partial (linear) ordering, then the relation $\leq$ (where $p \leq q$ if either $p < q$ or $p = q$) is also called a partial (linear) ordering (and $<$ is sometimes called a strict ordering).

Definition 2.2. If $(P, <)$ is a partially ordered set, $X$ is a nonempty subset of $P$, and $a \in P$, then:

- $a$ is a maximal element of $X$ if $a \in X$ and $(\forall x \in X) a \not< x$;
- $a$ is a minimal element of $X$ if $a \in X$ and $(\forall x \in X) x \not< a$;
- $a$ is the greatest element of $X$ if $a \in X$ and $(\forall x \in X) x \leq a$;
- $a$ is the least element of $X$ if $a \in X$ and $(\forall x \in X) a \leq x$;
- $a$ is an upper bound of $X$ if $(\forall x \in X) x \leq a$;
- $a$ is a lower bound of $X$ if $(\forall x \in X) a \leq x$;
- $a$ is the supremum of $X$ if $a$ is the least upper bound of $X$;
- $a$ is the infimum of $X$ if $a$ is the greatest lower bound of $X$.

The supremum (infimum) of $X$ (if it exists) is denoted sup $X$ (inf $X$).

Note that if $X$ is linearly ordered by $<$, then a maximal element of $X$ is its greatest element (similarly for a minimal element).

If $(P, <)$ and $(Q, <)$ are partially ordered sets and $f : P \to Q$, then $f$ is order-preserving if $x < y$ implies $f(x) < f(y)$. If $P$ and $Q$ are linearly ordered, then an order-preserving function is also called increasing.
A one-to-one function of $P$ onto $Q$ is an isomorphism of $P$ and $Q$ if both $f$ and $f^{-1}$ are order-preserving; $(P, \prec)$ is then isomorphic to $(Q, \prec)$. An isomorphism of $P$ onto itself is an automorphism of $(P, \prec)$.

Well-Ordering

Definition 2.3. A linear ordering $\prec$ of a set $P$ is a well-ordering if every nonempty subset of $P$ has a least element.

The concept of well-ordering is of fundamental importance. It is shown below that well-ordered sets can be compared by their lengths; ordinal numbers will be introduced as order-types of well-ordered sets.

Lemma 2.4. If $(W, \prec)$ is a well-ordered set and $f : W \to W$ is an increasing function, then $f(x) \geq x$ for each $x \in W$.

Proof. Assume that the set $X = \{x \in W : f(x) < x\}$ is nonempty and let $z$ be the least element of $X$. If $w = f(z)$, then $f(w) < w$, a contradiction. $\Box$

Corollary 2.5. The only automorphism of a well-ordered set is the identity.

Proof. By Lemma 2.4, $f(x) \geq x$ for all $x$, and $f^{-1}(x) \geq x$ for all $x$. $\Box$

Corollary 2.6. If two well-ordered sets $W_1$, $W_2$ are isomorphic, then the isomorphism of $W_1$ onto $W_2$ is unique. $\Box$

If $W$ is a well-ordered set and $u \in W$, then $\{x \in W : x < u\}$ is an initial segment of $W$ (given by $u$).

Lemma 2.7. No well-ordered set is isomorphic to an initial segment of itself.

Proof. If $\text{ran}(f) = \{x : x < u\}$, then $f(u) < u$, contrary to Lemma 2.4. $\Box$

Theorem 2.8. If $W_1$ and $W_2$ are well-ordered sets, then exactly one of the following three cases holds:

(i) $W_1$ is isomorphic to $W_2$;
(ii) $W_1$ is isomorphic to an initial segment of $W_2$;
(iii) $W_2$ is isomorphic to an initial segment of $W_1$.

Proof. For $u \in W_i, (i = 1, 2)$, let $W_i(u)$ denote the initial segment of $W_i$ given by $u$. Let

$$f = \{(x, y) \in W_1 \times W_2 : W_1(x) \text{ is isomorphic to } W_2(y)\}.$$ 

Using Lemma 2.7, it is easy to see that $f$ is a one-to-one function. If $h$ is an isomorphism between $W_1(x)$ and $W_2(y)$, and $x' < x$, then $W_1(x')$ and $W_2(h(x'))$ are isomorphic. It follows that $f$ is order-preserving.
If \( \text{dom}(f) = W_1 \) and \( \text{ran}(f) = W_2 \), then case (i) holds.

If \( y_1 < y_2 \) and \( y_2 \in \text{ran}(f) \), then \( y_1 \in \text{ran}(f) \). Thus if \( \text{ran}(f) \neq W_2 \) and \( y_0 \) is the least element of \( W_2 - \text{ran}(f) \), we have \( (x_0, y_0) \in f \), where \( x_0 = \) the least element of \( W_1 - \text{dom}(f) \). Thus case (ii) holds.

Similarly, if \( \text{dom}(f) \neq W_1 \), then case (iii) holds.

In view of Lemma 2.7, the three cases are mutually exclusive.

If \( W_1 \) and \( W_2 \) are isomorphic, we say that they have the same \textit{order-type}.

Informally, an ordinal number is the order-type of a well-ordered set.

We shall now give a formal definition of ordinal numbers.

\textbf{Ordinal Numbers}

The idea is to define ordinal numbers so that

\[ \alpha < \beta \quad \text{if and only if} \quad \alpha \in \beta, \quad \text{and} \quad \alpha = \{ \beta : \beta < \alpha \}. \]

\textbf{Definition 2.9.} A set \( T \) is \textit{transitive} if every element of \( T \) is a subset of \( T \).

(Equivalently, \( \bigcup T \subset T \), or \( T \subset P(T) \).)

\textbf{Definition 2.10.} A set is an \textit{ordinal number} (an \textit{ordinal}) if it is transitive and well-ordered by \( \in \).

We shall denote ordinals by lowercase Greek letters \( \alpha, \beta, \gamma, \ldots \). The class of all ordinals is denoted by \( \text{Ord} \).

We define

\[ \alpha < \beta \quad \text{if and only if} \quad \alpha \in \beta. \]

\textbf{Lemma 2.11.}

(i) \( 0 = \emptyset \) is an ordinal.

(ii) If \( \alpha \) is an ordinal and \( \beta \in \alpha \), then \( \beta \) is an ordinal.

(iii) If \( \alpha \neq \beta \) are ordinals and \( \alpha \subset \beta \), then \( \alpha \in \beta \).

(iv) If \( \alpha, \beta \) are ordinals, then either \( \alpha \subset \beta \) or \( \beta \subset \alpha \).

\textbf{Proof.} (i), (ii) by definition.

(iii) If \( \alpha \subset \beta \), let \( \gamma \) be the least element of the set \( \beta - \alpha \). Since \( \alpha \) is transitive, it follows that \( \alpha \) is the initial segment of \( \beta \) given by \( \gamma \). Thus \( \alpha = \{ \xi \in \beta : \xi < \gamma \} = \gamma \), and so \( \alpha \in \beta \).

(iv) Clearly, \( \alpha \cap \beta \) is an ordinal, \( \alpha \cap \beta = \gamma \). We have \( \gamma = \alpha \) or \( \gamma = \beta \), for otherwise \( \gamma \in \alpha \), and \( \gamma \in \beta \), by (iii). Then \( \gamma \in \gamma \), which contradicts the definition of an ordinal (namely that \( \in \) is a \textit{strict} ordering of \( \alpha \)). \( \Box \)
Using Lemma 2.11 one gets the following facts about ordinal numbers (the proofs are routine):

(2.1) $<$ is a linear ordering of the class $\text{Ord}$.
(2.2) For each $\alpha$, $\alpha = \{\beta : \beta < \alpha\}$.
(2.3) If $C$ is a nonempty class of ordinals, then $\bigcap C$ is an ordinal, $\bigcap C \in C$ and $\bigcap C = \inf C$.
(2.4) If $X$ is a nonempty set of ordinals, then $\bigcup X$ is an ordinal, and $\bigcup X = \sup X$.
(2.5) For every $\alpha$, $\alpha \cup \{\alpha\}$ is an ordinal and $\alpha \cup \{\alpha\} = \inf\{\beta : \beta > \alpha\}$.

We thus define $\alpha + 1 = \alpha \cup \{\alpha\}$ (the successor of $\alpha$). In view of (2.4), the class $\text{Ord}$ is a proper class; otherwise, consider $\sup \text{Ord} + 1$.

We can now prove that the above definition of ordinals provides us with order-types of well-ordered sets.

**Theorem 2.12.** Every well-ordered set is isomorphic to a unique ordinal number.

**Proof.** The uniqueness follows from Lemma 2.7. Given a well-ordered set $W$, we find an isomorphic ordinal as follows: Define $F(x) = \alpha$ if $\alpha$ is isomorphic to the initial segment of $W$ given by $x$. If such an $\alpha$ exists, then it is unique. By the Replacement Axioms, $F(W)$ is a set. For each $x \in W$, such an $\alpha$ exists (otherwise consider the least $x$ for which such an $\alpha$ does not exist). If $\gamma$ is the least $\gamma \notin F(W)$, then $F(W) = \gamma$ and we have an isomorphism of $W$ onto $\gamma$. \qed

If $\alpha = \beta + 1$, then $\alpha$ is a successor ordinal. If $\alpha$ is not a successor ordinal, then $\alpha = \sup\{\beta : \beta < \alpha\} = \bigcup \alpha$; $\alpha$ is called a limit ordinal. We also consider $0$ a limit ordinal and define $\sup \emptyset = 0$.

The existence of limit ordinals other than $0$ follows from the Axiom of Infinity; see Exercise 2.3.

**Definition 2.13 (Natural Numbers).** We denote the least nonzero limit ordinal $\omega$ (or $\mathbb{N}$). The ordinals less than $\omega$ (elements of $\mathbb{N}$) are called finite ordinals, or natural numbers. Specifically,

$$0 = \emptyset, \quad 1 = 0 + 1, \quad 2 = 1 + 1, \quad 3 = 2 + 1, \quad \text{etc.}$$

A set $X$ is finite if there is a one-to-one mapping of $X$ onto some $n \in \mathbb{N}$. $X$ is infinite if it is not finite.

We use letters $n, m, l, k, j, i$ (most of the time) to denote natural numbers.
Induction and Recursion

Theorem 2.14 (Transfinite Induction). Let $C$ be a class of ordinals and assume that:

(i) $0 \in C$;
(ii) if $\alpha \in C$, then $\alpha + 1 \in C$;
(iii) if $\alpha$ is a nonzero limit ordinal and $\beta \in C$ for all $\beta < \alpha$, then $\alpha \in C$.

Then $C$ is the class of all ordinals.

Proof. Otherwise, let $\alpha$ be the least ordinal $\alpha \notin C$ and apply (i), (ii), or (iii). \qed

A function whose domain is the set $\mathbb{N}$ is called an \textit{(infinite) sequence} (A sequence in $X$ is a function $f : \mathbb{N} \to X$.) The standard notation for a sequence is

$$\langle a_n : n < \omega \rangle$$

or variants thereof. A \textit{finite sequence} is a function $s$ such $\text{dom}(s) = \{i : i < n\}$ for some $n \in \mathbb{N}$; then $s$ is a sequence of length $n$.

A \textit{transfinite sequence} is a function whose domain is an ordinal:

$$\langle a_\xi : \xi < \alpha \rangle.$$  

It is also called an $\alpha$-sequence or a sequence of length $\alpha$. We also say that a sequence $\langle a_\xi : \xi < \alpha \rangle$ is an \textit{enumeration} of its range $\{a_\xi : \xi < \alpha\}$. If $s$ is a sequence of length $\alpha$, then $s \upharpoonright x$ or simply $sx$ denotes the sequence of length $\alpha + 1$ that extends $s$ and whose $\alpha$th term is $x$:

$$s \upharpoonright x = sx = s \cup \{(\alpha, x)\}.$$  

Sometimes we shall call a “sequence”

$$\langle a_\alpha : \alpha \in \text{Ord} \rangle$$

a function (a proper class) on $\text{Ord}$.

“Definition by transfinite recursion” usually takes the following form: Given a function $G$ (on the class of transfinite sequences), then for every $\theta$ there exists a unique $\theta$-sequence

$$\langle a_\alpha : \alpha < \theta \rangle$$

such that

$$a_\alpha = G(\langle a_\xi : \xi < \alpha \rangle)$$

for every $\alpha < \theta$.

We shall give a general version of this theorem, so that we can also construct sequences $\langle a_\alpha : \alpha \in \text{Ord} \rangle$. 

Theorem 2.15 (Transfinite Recursion). Let $G$ be a function (on $V$), then (2.6) below defines a unique function $F$ on $\text{Ord}$ such that

$$F(\alpha) = G(F|\alpha)$$

for each $\alpha$.

In other words, if we let $a_\alpha = F(\alpha)$, then for each $\alpha$,

$$a_\alpha = G(\langle a_\xi : \xi < \alpha \rangle).$$

(Note that we tacitly use Replacement: $F|\alpha$ is a set for each $\alpha$.)

Corollary 2.16. Let $X$ be a set and $\theta$ an ordinal number. For every function $G$ on the set of all transfinite sequences in $X$ of length $< \theta$ such that $\text{ran}(G) \subset X$ there exists a unique $\theta$-sequence $\langle a_\alpha : \alpha < \theta \rangle$ in $X$ such that $a_\alpha = G(\langle a_\xi : \xi < \alpha \rangle)$ for every $\alpha < \theta$.

Proof. Let

$$F(\alpha) = x \leftrightarrow \text{there is a sequence } \langle a_\xi : \xi < \alpha \rangle \text{ such that:}$$

(i) $(\forall \xi < \alpha) a_\xi = G(\langle a_\eta : \eta < \xi \rangle)$;

(ii) $x = G(\langle a_\xi : \xi < \alpha \rangle)$.

For every $\alpha$, if there is an $\alpha$-sequence that satisfies (i), then such a sequence is unique: If $\langle a_\xi : \xi < \alpha \rangle$ and $\langle b_\xi : \xi < \alpha \rangle$ are two $\alpha$-sequences satisfying (i), one shows $a_\xi = b_\xi$ by induction on $\xi$. Thus $F(\alpha)$ is determined uniquely by (ii), and therefore $F$ is a function. It follows, again by induction, that for each $\alpha$ there is an $\alpha$-sequence that satisfies (i) (at limit steps, we use Replacement to get the $\alpha$-sequence as the union of all the $\xi$-sequences, $\xi < \alpha$). Thus $F$ is defined for all $\alpha \in \text{Ord}$. It obviously satisfies

$$F(\alpha) = G(F|\alpha).$$

If $F'$ is any function on $\text{Ord}$ that satisfies

$$F'(\alpha) = G(F'|\alpha)$$

then it follows by induction that $F'(\alpha) = F(\alpha)$ for all $\alpha$.

Definition 2.17. Let $\alpha > 0$ be a limit ordinal and let $\langle \gamma_\xi : \xi < \alpha \rangle$ be a nondecreasing sequence of ordinals (i.e., $\xi < \eta$ implies $\gamma_\xi \leq \gamma_\eta$). We define the limit of the sequence by

$$\lim_{\xi \to \alpha} \gamma_\xi = \text{sup}\{\gamma_\xi : \xi < \alpha\}.$$

A sequence of ordinals $\langle \gamma_\alpha : \alpha \in \text{Ord} \rangle$ is normal if it is increasing and continuous, i.e., for every limit $\alpha$, $\gamma_\alpha = \lim_{\xi \to \alpha} \gamma_\xi$. 
Ordinal Arithmetic

We shall now define addition, multiplication and exponentiation of ordinal numbers, using Transfinite Recursion.

**Definition 2.18 (Addition).** For all ordinal numbers $\alpha$

(i) $\alpha + 0 = \alpha$,
(ii) $\alpha + (\beta + 1) = (\alpha + \beta) + 1$, for all $\beta$,
(iii) $\alpha + \beta = \lim_{\xi \to \beta} (\alpha + \xi)$ for all limit $\beta > 0$.

**Definition 2.19 (Multiplication).** For all ordinal numbers $\alpha$

(i) $\alpha \cdot 0 = 0$,
(ii) $\alpha \cdot (\beta + 1) = \alpha \cdot \beta + \alpha$ for all $\beta$,
(iii) $\alpha \cdot \beta = \lim_{\xi \to \beta} \alpha \cdot \xi$ for all limit $\beta > 0$.

**Definition 2.20 (Exponentiation).** For all ordinal numbers $\alpha$

(i) $\alpha^0 = 1$,
(ii) $\alpha^{\beta+1} = \alpha^\beta \cdot \alpha$ for all $\beta$,
(iii) $\alpha^\beta = \lim_{\xi \to \beta} \alpha^\xi$ for all limit $\beta > 0$.

As defined, the operations $\alpha + \beta$, $\alpha \cdot \beta$ and $\alpha^\beta$ are normal functions in the second variable $\beta$. Their properties can be proved by transfinite induction. For instance, $+$ and $\cdot$ are associative:

**Lemma 2.21.** For all ordinals $\alpha$, $\beta$ and $\gamma$,

(i) $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$,
(ii) $\alpha \cdot (\beta \cdot \gamma) = (\alpha \cdot \beta) \cdot \gamma$.

*Proof.* By induction on $\gamma$. $\square$

Neither $+$ nor $\cdot$ are commutative:

$$1 + \omega = \omega \neq \omega + 1, \quad 2 \cdot \omega = \omega \neq \omega \cdot 2 = \omega + \omega.$$ 

Ordinal sums and products can be also defined geometrically, as can sums and products of arbitrary linear orders:

**Definition 2.22.** Let $(A, <_A)$ and $(B, <_B)$ be disjoint linearly ordered sets. The sum of these linear orders is the set $A \cup B$ with the ordering defined as follows: $x < y$ if and only if

(i) $x, y \in A$ and $x <_A y$, or
(ii) $x, y \in B$ and $x <_B y$, or
(iii) $x \in A$ and $y \in B$. 

**Definition 2.23.** Let \((A, <)\) and \((B, <)\) be linearly ordered sets. The product of these linear orders is the set \(A \times B\) with the ordering defined by

\[(a_1, b_1) < (a_2, b_2)\] if and only if either \(b_1 < b_2\) or \((b_1 = b_2\) and \(a_1 < a_2)\).

**Lemma 2.24.** For all ordinals \(\alpha\) and \(\beta\), \(\alpha + \beta\) and \(\alpha \cdot \beta\) are, respectively, isomorphic to the sum and to the product of \(\alpha\) and \(\beta\).

**Proof.** By induction on \(\beta\).

Ordinal sums and products have some properties of ordinary addition and multiplication of integers. For instance:

**Lemma 2.25.**

(i) If \(\beta < \gamma\) then \(\alpha + \beta < \alpha + \gamma\).
(ii) If \(\alpha < \beta\) then there exists a unique \(\delta\) such that \(\alpha + \delta = \beta\).
(iii) If \(\beta < \gamma\) and \(\alpha > 0\), then \(\alpha \cdot \beta < \alpha \cdot \gamma\).
(iv) If \(\alpha > 0\) and \(\gamma\) is arbitrary, then there exist a unique \(\beta\) and a unique \(\rho < \alpha\) such that \(\gamma = \alpha \cdot \beta + \rho\).
(v) If \(\beta < \gamma\) and \(\alpha > 1\), then \(\alpha^\beta < \alpha^\gamma\).

**Proof.** (i), (iii) and (v) are proved by induction on \(\gamma\).
(ii) Let \(\delta\) be the order-type of the set \(\{\xi : \alpha \leq \xi < \beta\}\); \(\delta\) is unique by (i).
(iv) Let \(\beta\) be the greatest ordinal such that \(\alpha \cdot \beta \leq \gamma\). The uniqueness of the normal form is proved by induction.

For more, see Exercises 2.10 and 2.11.

**Theorem 2.26 (Cantor's Normal Form Theorem).** Every ordinal \(\alpha > 0\) can be represented uniquely in the form

\[\alpha = \omega^{\beta_1} \cdot k_1 + \ldots + \omega^{\beta_n} \cdot k_n,\]

where \(n \geq 1\), \(\alpha \geq \beta_1 > \ldots > \beta_n\), and \(k_1, \ldots, k_n\) are nonzero natural numbers.

**Proof.** By induction on \(\alpha\). For \(\alpha = 1\) we have \(1 = \omega^0 \cdot 1\); for arbitrary \(\alpha > 0\) let \(\beta\) be the greatest ordinal such that \(\omega^\beta \leq \alpha\). By Lemma 2.25(iv) there exists a unique \(\delta\) and a unique \(\rho < \omega^\beta\) such that \(\alpha = \omega^\beta \cdot \delta + \rho\); this \(\delta\) must necessarily be finite. The uniqueness of the normal form is proved by induction.

In the normal form it is possible to have \(\alpha = \omega^\alpha\); see Exercise 2.12. The least ordinal with this property is called \(\varepsilon_0\).
Well-Founded Relations

Now we shall define an important generalization of well-ordered sets.

A binary relation $E$ on a set $P$ is well-founded if every nonempty $X \subset P$ has an $E$-minimal element, that is $a \in X$ such that there is no $x \in X$ with $x \ E \ a$.

Clearly, a well-ordering of $P$ is a well-founded relation.

Given a well-founded relation $E$ on a set $P$, we can define the height of $E$, and assign to each $x \in P$ an ordinal number, the rank of $x$ in $E$.

**Theorem 2.27.** If $E$ is a well-founded relation on $P$, then there exists a unique function $\rho$ from $P$ into the ordinals such that for all $x \in P$,

$$(2.7) \quad \rho(x) = \sup \{\rho(y) + 1 : y \ E \ x\}.$$  

The range of $\rho$ is an initial segment of the ordinals, thus an ordinal number. This ordinal is called the height of $E$.

**Proof.** We shall define a function $\rho$ satisfying (2.7) and then prove its uniqueness. By induction, let

$$P_0 = \emptyset, \quad P_{\alpha + 1} = \{x \in P : \forall y (y \ E \ x \rightarrow y \in P_\alpha)\},$$

$$P_\alpha = \bigcup_{\xi < \alpha} P_\xi \quad \text{if } \alpha \text{ is a limit ordinal.}$$

Let $\theta$ be the least ordinal such that $P_{\theta + 1} = P_\theta$ (such $\theta$ exists by Replacement).

First, it should be easy to see that $P_\alpha \subset P_{\alpha + 1}$ for each $\alpha$ (by induction). Thus $P_0 \subset P_1 \subset \ldots \subset P_\theta$. We claim that $P_\theta = P$. Otherwise, let $a$ be an $E$-minimal element of $P - P_\theta$. It follows that each $x \ E \ a$ is in $P_\theta$, and so $a \in P_{\theta + 1}$, a contradiction. Now we define $\rho(x)$ as the least $\alpha$ such that $x \in P_{\alpha + 1}$. It is obvious that if $x \ E \ y$, then $\rho(x) < \rho(y)$, and (2.7) is easily verified. The ordinal $\theta$ is the height of $E$.

The uniqueness of $\rho$ is established as follows: Let $\rho'$ be another function satisfying (2.7) and consider an $E$-minimal element of the set $\{x \in P : \rho(x) \neq \rho'(x)\}$. $\square$

**Exercises**

2.1. The relation "$(P, <)$ is isomorphic to $(Q, <)$" is an equivalence relation (on the class of all partially ordered sets).

2.2. $\alpha$ is a limit ordinal if and only if $\beta < \alpha$ implies $\beta + 1 < \alpha$, for every $\beta$.

2.3. If a set $X$ is inductive, then $X \cap \text{Ord}$ is inductive. The set $N = \bigcap \{X : X$ is inductive$\}$ is the least limit ordinal $\neq 0$. 
2.4. (Without the Axiom of Infinity). Let $\omega =$ least limit $\alpha \neq 0$ if it exists, $\omega =$ Ord otherwise. Prove that the following statements are equivalent:

(i) There exists an inductive set.
(ii) There exists an infinite set.
(iii) $\omega$ is a set.

[For (ii) $\rightarrow$ (iii), apply Replacement to the set of all finite subsets of $X$.]

2.5. If $W$ is a well-ordered set, then there exists no sequence $\langle a_n : n \in \mathbb{N} \rangle$ in $W$ such that $a_0 > a_1 > a_2 > \ldots$.

2.6. There are arbitrarily large limit ordinals; i.e., $\forall \alpha \exists \beta > \alpha$ ($\beta$ is a limit).

[Consider $\lim_{n \to \omega} \alpha_n$, where $\alpha_{n+1} = \alpha_n + 1$.]

2.7. Every normal sequence $\langle \gamma_\alpha : \alpha \in \text{Ord} \rangle$ has arbitrarily large fixed points, i.e., $\alpha$ such that $\gamma_\alpha = \alpha$.

[Let $\alpha_{n+1} = \gamma_{\alpha_n}$, and $\alpha = \lim_{n \to \omega} \alpha_n$.]

2.8. For all $\alpha$, $\beta$ and $\gamma$,

(i) $\alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma$,
(ii) $\alpha^{\beta + \gamma} = \alpha^\beta \cdot \alpha^\gamma$,
(iii) $(\alpha^\beta)^\gamma = \alpha^{\beta \cdot \gamma}$.

2.9. (i) Show that $(\omega + 1) \cdot 2 \neq \omega \cdot 2 + 1 \cdot 2$.
(ii) Show that $(\omega \cdot 2)^2 \neq \omega^2 \cdot 2^2$.

2.10. If $\alpha < \beta$ then $\alpha + \gamma \leq \beta + \gamma$, $\alpha \cdot \gamma \leq \beta \cdot \gamma$, and $\alpha^\gamma \leq \beta^\gamma$.

2.11. Find $\alpha$, $\beta$, $\gamma$ such that

(i) $\alpha \leq \beta$ and $\alpha + \gamma = \beta + \gamma$,
(ii) $\alpha < \beta$ and $\alpha \cdot \gamma = \beta \cdot \gamma$,
(iii) $\alpha < \beta$ and $\alpha^\gamma = \beta^\gamma$.

2.12. Let $\varepsilon_0 = \lim_{n \to \omega} \alpha_n$ where $\alpha_0 = \omega$ and $\alpha_{n+1} = \omega^{\alpha_n}$ for all $n$. Show that $\varepsilon_0$ is the least ordinal $\varepsilon$ such that $\omega^\varepsilon = \varepsilon$.

A limit ordinal $\gamma > 0$ is called indecomposable if there exist no $\alpha < \gamma$ and $\beta < \gamma$ such that $\alpha + \beta = \gamma$.

2.13. A limit ordinal $\gamma > 0$ is indecomposable if and only if $\alpha + \gamma = \gamma$ for all $\alpha < \gamma$ if and only if $\gamma = \omega^\alpha$ for some $\alpha$.

2.14. If $E$ is a well-founded relation on $P$, then there is no sequence $\langle a_n : n \in \mathbb{N} \rangle$ in $P$ such that $a_1 E a_0$, $a_2 E a_1$, $a_3 E a_2$, $\ldots$.

2.15 (Well-Founded Recursion). Let $E$ be a well-founded relation on a set $P$, and let $G$ be a function. Then there exists a function $F$ such that for all $x \in P$, $F(x) = G(x, F|\{y \in P : y E x\})$.

Historical Notes

The theory of well-ordered sets was developed by Cantor, who also introduced transfinite induction. The idea of identifying an ordinal number with the set of smaller ordinals is due to Zermelo and von Neumann.