0

Prologue

The purpose of this introductory chapter is to establish the notation and terminology that will be used throughout the book and to present a few diverse results from set theory and analysis that will be needed later. The style here is deliberately terse, since this chapter is intended as a reference rather than a systematic exposition.

0.1 THE LANGUAGE OF SET THEORY

It is assumed that the reader is familiar with the basic concepts of set theory; the following discussion is meant mainly to fix our terminology.

Number Systems. Our notation for the fundamental number systems is as follows:

$\mathbb{N} =$ the set of positive integers (not including zero)

$\mathbb{Z} =$ the set of integers

$\mathbb{Q} =$ the set of rational numbers

$\mathbb{R} =$ the set of real numbers

$\mathbb{C} =$ the set of complex numbers

Logic. We shall avoid the use of special symbols from mathematical logic, preferring to remain reasonably close to standard English. We shall, however, use the abbreviation **iff** for "if and only if."

One point of elementary logic that is often insufficiently appreciated by students is the following: If A and B are mathematical assertions and $-A$, $-B$ are their negations, the statement "A implies B" is logically equivalent to the contrapositive statement "$-B$ implies $-A$." Thus one may prove that A implies B by assuming $-B$ and deducing $-A$, and we shall frequently do so. This is not the same as *reductio ad absurdum,* which consists of assuming both A and $-B$ and deriving a contradiction.

Sets. The words "family" and "collection" will be used synonymously with "set," usually to avoid phrases like "set of sets." The empty set is denoted by $\varnothing$, and the family of all subsets of a set X is denoted by $\mathcal{P}(X)$:

$$\mathcal{P}(X) = \{E : E \subset X\}.$$

Here and elsewhere, the inclusion sign $\subset$ is interpreted in the weak sense; that is, the assertion "$E \subset X$" includes the possibility that $E = X$.

If $\mathcal{E}$ is a family of sets, we can form the union and intersection of its members:

$$\bigcup_{E\in\mathcal{E}} E = \{x : x \in E \text{ for some } E \in \mathcal{E}\},$$

$$\bigcap_{E\in\mathcal{E}} E = \{x : x \in E \text{ for all } E \in \mathcal{E}\}.$$

Usually it is more convenient to consider indexed families of sets:

$$\mathcal{E} = \{E_\alpha : \alpha \in A\} = \{E_\alpha\}_{\alpha\in A},$$

in which case the union and intersection are denoted by

$$\bigcup_{\alpha\in A} E_\alpha, \qquad \bigcap_{\alpha\in A} E_\alpha.$$

If $E_\alpha \cap E_\beta = \varnothing$ whenever $\alpha \neq \beta$, the sets E_α are called **disjoint**. The terms "disjoint collection of sets" and "collection of disjoint sets" are used interchangeably, as are "disjoint union of sets" and "union of disjoint sets."

When considering families of sets indexed by $\mathbb{N}$, our usual notation will be

$$\{E_n\}_{n=1}^\infty \quad \text{or} \quad \{E_n\}_1^\infty,$$

and likewise for unions and intersections. In this situation, the notions of **limit superior** and **limit inferior** are sometimes useful:

$$\limsup E_n = \bigcap_{k=1}^\infty \bigcup_{n=k}^\infty E_n, \qquad \liminf E_n = \bigcup_{k=1}^\infty \bigcap_{n=k}^\infty E_n.$$

The reader may verify that

$$\limsup E_n = \{x : x \in E_n \text{ for infinitely many } n\},$$

$$\liminf E_n = \{x : x \in E_n \text{ for all but finitely many } n\}.$$

If E and F are sets, we denote their **difference** by $E \setminus F$:

$$E \setminus F = \{x : x \in E \text{ and } x \notin F\},$$

and their **symmetric difference** by $E \triangle F$:

$$E \triangle F = (E \setminus F) \cup (F \setminus E).$$

When it is clearly understood that all sets in question are subsets of a fixed set X, we define the **complement** E^c of a set E (in X):

$$E^c = X \setminus E.$$

In this situation we have **deMorgan's laws**:

$$\Big(\bigcup_{\alpha \in A} E_\alpha\Big)^c = \bigcap_{\alpha \in A} E_\alpha^c, \qquad \Big(\bigcap_{\alpha \in A} E_\alpha\Big)^c = \bigcup_{\alpha \in A} E_\alpha^c.$$

If X and Y are sets, their **Cartesian product** $X \times Y$ is the set of all ordered pairs (x, y) such that $x \in X$ and $y \in Y$. A **relation** from X to Y is a subset of $X \times Y$. (If $Y = X$, we speak of a relation **on** X.) If R is a relation from X to Y, we shall sometimes write xRy to mean that $(x, y) \in R$. The most important types of relations are the following:

- Equivalence relations. An **equivalence relation** on X is a relation R on X such that

  $$xRx \text{ for all } x \in X,$$
  $$xRy \text{ iff } yRx,$$
  $$xRz \text{ whenever } xRy \text{ and } yRz \text{ for some } y.$$

  The **equivalence class** of an element x is $\{y \in X : xRy\}$. X is the disjoint union of these equivalence classes.

- Orderings. See §0.2.

- Mappings. A **mapping** $f : X \to Y$ is a relation R from X to Y with the property that for every $x \in X$ there is a unique $y \in Y$ such that xRy, in which case we write $y = f(x)$. Mappings are sometimes called **maps** or **functions**; we shall generally reserve the latter name for the case when Y is $\mathbb{C}$ or some subset thereof.

If $f : X \to Y$ and $g : Y \to Z$ are mappings, we denote by $g \circ f$ their **composition**:

$$g \circ f : X \to Z, \qquad g \circ f(x) = g(f(x)).$$

If $D \subset X$ and $E \subset Y$, we define the **image** of D and the **inverse image** of E under a mapping $f : X \to Y$ by

$$f(D) = \{f(x) : x \in D\}, \qquad f^{-1}(E) = \{x : f(x) \in E\}.$$

It is easily verified that the map $f^{-1} : \mathcal{P}(Y) \to \mathcal{P}(X)$ defined by the second formula commutes with union, intersections, and complements:

$$f^{-1}\Big(\bigcup_{\alpha\in A} E_\alpha\Big) = \bigcup_{\alpha\in A} f^{-1}(E_\alpha), \qquad f^{-1}\Big(\bigcap_{\alpha\in A} E_\alpha\Big) = \bigcap_{\alpha\in A} f^{-1}(E_\alpha),$$
$$f^{-1}(E^c) = \big(f^{-1}(E)\big)^c.$$

(The direct image mapping $f : \mathcal{P}(X) \to \mathcal{P}(Y)$ commutes with unions, but in general not with intersections or complements.)

If $f : X \to Y$ is a mapping, X is called the **domain** of f and $f(X)$ is called the **range** of f. f is said to be **injective** if $f(x_1) = f(x_2)$ only when $x_1 = x_2$, **surjective** if $f(X) = Y$, and **bijective** if it is both injective and surjective. If f is bijective, it has an **inverse** $f^{-1} : Y \to X$ such that $f^{-1} \circ f$ and $f \circ f^{-1}$ are the identity mappings on X and Y, respectively. If $A \subset X$, we denote by $f|A$ the restriction of f to A:

$$(f|A) : A \to Y, \qquad (f|A)(x) = f(x) \text{ for } x \in A.$$

A **sequence** in a set X is a mapping from $\mathbb{N}$ into X. (We also use the term **finite sequence** to mean a map from $\{1, \ldots, n\}$ into X where $n \in \mathbb{N}$.) If $f : \mathbb{N} \to X$ is a sequence and $g : \mathbb{N} \to \mathbb{N}$ satisfies $g(n) < g(m)$ whenever $n < m$, the composition $f \circ g$ is called a **subsequence** of f. It is common, and often convenient, to be careless about distinguishing between sequences and their ranges, which are subsets of X indexed by $\mathbb{N}$. Thus, if $f(n) = x_n$, we speak of the sequence $\{x_n\}_1^\infty$; whether we mean a mapping from $\mathbb{N}$ to X or a subset of X will be clear from the context.

Earlier we defined the Cartesian product of two sets. Similarly one can define the Cartesian product of n sets in terms of ordered n-tuples. However, this definition becomes awkward for infinite families of sets, so the following approach is used instead. If $\{X_\alpha\}_{\alpha\in A}$ is an indexed family of sets, their **Cartesian product** $\prod_{\alpha\in A} X_\alpha$ is the set of all maps $f : A \to \bigcup_{\alpha\in A} X_\alpha$ such that $f(\alpha) \in X_\alpha$ for every $\alpha \in A$. (It should be noted, and then promptly forgotten, that when $A = \{1, 2\}$, the previous definition of $X_1 \times X_2$ is set-theoretically different from the present definition of $\prod_1^2 X_j$. Indeed, the latter concept depends on mappings, which are defined in terms of the former one.) If $X = \prod_{\alpha\in A} X_\alpha$ and $\alpha \in A$, we define the αth **projection** or **coordinate map** $\pi_\alpha : X \to X_\alpha$ by $\pi_\alpha(f) = f(\alpha)$. We also frequently write x and x_α instead of f and $f(\alpha)$ and call x_α the αth **coordinate** of x.

If the sets X_α are all equal to some fixed set Y, we denote $\prod_{\alpha\in A} X_\alpha$ by Y^A:

$$Y^A = \text{ the set of all mappings from } A \text{ to } Y.$$

If $A = \{1, \ldots, n\}$, Y^A is denoted by Y^n and may be identified with the set of ordered n-tuples of elements of Y.

0.2 ORDERINGS

A **partial ordering** on a nonempty set X is a relation R on X with the following properties:

- if xRy and yRz, then xRz;
- if xRy and yRx, then $x = y$;
- xRx for all x.

If R also satisfies

- if $x, y \in X$, then either xRy or yRx,

then R is called a **linear** (or **total**) ordering. For example, if E is any set, then $\mathcal{P}(E)$ is partially ordered by inclusion, and $\mathbb{R}$ is linearly ordered by its usual ordering. Taking this last example as a model, we shall usually denote partial orderings by $\leq$, and we write $x < y$ to mean that $x \leq y$ but $x \neq y$. We observe that a partial ordering on X naturally induces a partial ordering on every nonempty subset of X. Two partially ordered sets X and Y are said to be **order isomorphic** if there is a bijection $f : X \to Y$ such that $x_1 \leq x_2$ iff $f(x_1) \leq f(x_2)$.

If X is partially ordered by $\leq$, a **maximal** (resp. **minimal**) **element** of X is an element $x \in X$ such that the only $y \in X$ satisfying $x \leq y$ (resp. $x \geq y$) is x itself. Maximal and minimal elements may or may not exist, and they need not be unique unless the ordering is linear. If $E \subset X$, an **upper** (resp. **lower**) **bound** for E is an element $x \in X$ such that $y \leq x$ (resp. $x \leq y$) for all $y \in E$. An upper bound for E need not be an element of E, and unless E is linearly ordered, a maximal element of E need not be an upper bound for E. (The reader should think up some examples.)

If X is linearly ordered by $\leq$ and every nonempty subset of X has a (necessarily unique) minimal element, X is said to be **well ordered** by $\leq$, and (in defiance of the laws of grammar) $\leq$ is called a **well ordering** on X. For example, $\mathbb{N}$ is well ordered by its natural ordering.

We now state a fundamental principle of set theory and derive some consequences of it.

0.1 The Hausdorff Maximal Principle. *Every partially ordered set has a maximal linearly ordered subset.*

In more detail, this means that if X is partially ordered by $\leq$, there is a set $E \subset X$ that is linearly ordered by $\leq$, such that no subset of X that properly includes E is linearly ordered by $\leq$. Another version of this principle is the following:

0.2 Zorn's Lemma. *If X is a partially ordered set and every linearly ordered subset of X has an upper bound, then X has a maximal element.*

Clearly the Hausdorff maximal principle implies Zorn's lemma: An upper bound for a maximal linearly ordered subset of X is a maximal element of X. It is also not difficult to see that Zorn's lemma implies the Hausdorff maximal principle. (Apply Zorn's lemma to the collection of linearly ordered subsets of X, which is partially ordered by inclusion.)

0.3 The Well Ordering Principle. *Every nonempty set X can be well ordered.*

Proof. Let $\mathcal{W}$ be the collection of well orderings of subsets of X, and define a partial ordering on $\mathcal{W}$ as follows. If $\leq_1$ and $\leq_2$ are well orderings on the subsets E_1 and E_2, then $\leq_1$ precedes $\leq_2$ in the partial ordering if (i) $\leq_2$ extends $\leq_1$, i.e., $E_1 \subset E_2$ and $\leq_1$ and $\leq_2$ agree on E_1, and (ii) if $x \in E_2 \setminus E_1$ then $y \leq_2 x$ for all $y \in E_1$. The reader may verify that the hypotheses of Zorn's lemma are satisfied, so that $\mathcal{W}$ has a maximal element. This must be a well ordering on X itself, for if $\leq$ is a well ordering on a proper subset E of X and $x_0 \in X \setminus E$, then $\leq$ can be extended to a well ordering on $E \cup \{x_0\}$ by declaring that $x \leq x_0$ for all $x \in E$. ∎

0.4 The Axiom of Choice. *If $\{X_\alpha\}_{\alpha \in A}$ is a nonempty collection of nonempty sets, then $\prod_{\alpha \in A} X_\alpha$ is nonempty.*

Proof. Let $X = \bigcup_{\alpha \in A} X_\alpha$. Pick a well ordering on X and, for $\alpha \in A$, let $f(\alpha)$ be the minimal element of X_α. Then $f \in \prod_{\alpha \in A} X_\alpha$. ∎

0.5 Corollary. *If $\{X_\alpha\}_{\alpha \in A}$ is a disjoint collection of nonempty sets, there is a set $Y \subset \bigcup_{\alpha \in A} X_\alpha$ such that $Y \cap X_\alpha$ contains precisely one element for each $\alpha \in A$.*

Proof. Take $Y = f(A)$ where $f \in \prod_{\alpha \in A} X_\alpha$. ∎

We have deduced the axiom of choice from the Hausdorff maximal principle; in fact, it can be shown that the two are logically equivalent.

0.3 CARDINALITY

If X and Y are nonempty sets, we define the expressions

$$\text{card}(X) \leq \text{card}(Y), \qquad \text{card}(X) = \text{card}(Y), \qquad \text{card}(X) \geq \text{card}(Y)$$

to mean that there exists $f : X \to Y$ which is injective, bijective, or surjective, respectively. We also define

$$\text{card}(X) < \text{card}(Y), \qquad \text{card}(X) > \text{card}(Y)$$

to mean that there is an injection but no bijection, or a surjection but no bijection, from X to Y. Observe that we attach no meaning to the expression "$\text{card}(X)$" when it stands alone; there are various ways of doing so, but they are irrelevant for our purposes (except when X is finite — see below). These relationships can be extended to the empty set by declaring that

$$\text{card}(\varnothing) < \text{card}(X) \text{ and } \text{card}(X) > \text{card}(\varnothing) \text{ for all } X \neq \varnothing.$$

For the remainder of this section we assume implicitly that all sets in question are nonempty in order to avoid special arguments for $\varnothing$. Our first task is to prove that the relationships defined above enjoy the properties that the notation suggests.

0.6 Proposition. $\text{card}(X) \le \text{card}(Y)$ *iff* $\text{card}(Y) \ge \text{card}(X)$.

Proof. If $f : X \to Y$ is injective, pick $x_0 \in X$ and define $g : Y \to X$ by $g(y) = f^{-1}(y)$ if $y \in f(X)$, $g(y) = x_0$ otherwise. Then g is surjective. Conversely, if $g : Y \to X$ is surjective, the sets $g^{-1}(\{x\})$ $(x \in X)$ are nonempty and disjoint, so any $f \in \prod_{x\in X} g^{-1}(\{x\})$ is an injection from X to Y. ∎

0.7 Proposition. *For any sets* X *and* Y, *either* $\text{card}(X) \le \text{card}(Y)$ *or* $\text{card}(Y) \le \text{card}(X)$.

Proof. Consider the set $\mathcal{I}$ of all injections from subsets of X to Y. The members of $\mathcal{I}$ can be regarded as subsets of $X \times Y$, so $\mathcal{I}$ is partially ordered by inclusion. It is easily verified that Zorn's lemma applies, so $\mathcal{I}$ has a maximal element f, with (say) domain A and range B. If $x_0 \in X \setminus A$ and $y_0 \in Y \setminus B$, then f can be extended to an injection from $A \cup \{x_0\}$ to $Y \cup \{y_0\}$ by setting $f(x_0) = y_0$, contradicting maximality. Hence either $A = X$, in which case $\text{card}(X) \le \text{card}(Y)$, or $B = Y$, in which case f^{-1} is an injection from Y to X and $\text{card}(Y) \le \text{card}(X)$. ∎

0.8 The Schröder-Bernstein Theorem. *If* $\text{card}(X) \le \text{card}(Y)$ *and* $\text{card}(Y) \le \text{card}(X)$ *then* $\text{card}(X) = \text{card}(Y)$.

Proof. Let $f : X \to Y$ and $g : Y \to X$ be injections. Consider a point $x \in X$: If $x \in g(Y)$, we form $g^{-1}(x) \in Y$; if $g^{-1}(x) \in f(X)$, we form $f^{-1}(g^{-1}(x))$; and so forth. Either this process can be continued indefinitely, or it terminates with an element of $X \setminus g(Y)$ (perhaps x itself), or it terminates with an element of $Y \setminus f(X)$. In these three cases we say that x is in X_∞, X_X, or X_Y; thus X is the disjoint union of X_∞, X_X, and X_Y. In the same way, Y is the disjoint union of three sets Y_∞, Y_X, and Y_Y. Clearly f maps X_∞ onto Y_∞ and X_X onto Y_X, whereas g maps Y_Y onto X_Y. Therefore, if we define $h : X \to Y$ by $h(x) = f(x)$ if $X \in X_\infty \cup X_X$ and $h(x) = g^{-1}(x)$ if $x \in X_Y$, then h is bijective. ∎

0.9 Proposition. *For any set* X, $\text{card}(X) < \text{card}(\mathcal{P}(X))$.

Proof. On the one hand, the map $f(x) = \{x\}$ is an injection from X to $\mathcal{P}(X)$. On the other, if $g : X \to \mathcal{P}(X)$, let $Y = \{x \in X : x \notin g(x)\}$. Then $Y \notin g(X)$, for if $Y = g(x_0)$ for some $x_0 \in X$, any attempt to answer the question "Is $x_0 \in Y$?" quickly leads to an absurdity. Hence g cannot be surjective. ∎

A set X is called **countable** (or **denumerable**) if $\text{card}(X) \le \text{card}(\mathbb{N})$. In particular, all finite sets are countable, and for these it is convenient to interpret "$\text{card}(X)$" as the number of elements in X:

$$\text{card}(X) = n \text{ iff } \text{card}(X) = \text{card}(\{1, \ldots, n\}).$$

If X is countable but not finite, we say that X is **countably infinite**.

0.10 Proposition.

a. *If X and Y are countable, so is $X \times Y$.*

b. *If A is countable and X_α is countable for every $\alpha \in A$, then $\bigcup_{\alpha \in A} X_\alpha$ is countable.*

c. *If X is countably infinite, then* $\operatorname{card}(X) = \operatorname{card}(\mathbb{N})$.

Proof. To prove (a) it suffices to prove that $\mathbb{N}^2$ is countable. But we can define a bijection from $\mathbb{N}$ to $\mathbb{N}^2$ by listing, for n successively equal to $2, 3, 4, \ldots$, those elements $(j, k) \in \mathbb{N}^2$ such that $j + k = n$ in order of increasing j, thus:

$$(1,1),\ (1,2),\ (2,1),\ (1,3),\ (2,2),\ (3,1),\ (1,4),\ (2,3),\ (3,2),\ (4,1), \ldots$$

As for (b), for each $\alpha \in A$ there is a surjective $f_\alpha : \mathbb{N} \to X_\alpha$, and then the map $f : \mathbb{N} \times A \to \bigcup_{\alpha \in A} X_\alpha$ defined by $f(n, \alpha) = f_\alpha(n)$ is surjective; the result therefore follows from (a). Finally, for (c) it suffices to assume that X is an infinite subset of $\mathbb{N}$. Let $f(1)$ be the smallest element of X, and define $f(n)$ inductively to be the smallest element of $E \setminus \{f(1), \ldots, f(n-1)\}$. Then f is easily seen to be a bijection from $\mathbb{N}$ to X. ∎

0.11 Corollary. *$\mathbb{Z}$ and $\mathbb{Q}$ are countable.*

Proof. $\mathbb{Z}$ is the union of the countable sets $\mathbb{N}$, $\{-n : n \in \mathbb{N}\}$, and $\{0\}$, and one can define a surjection $f : \mathbb{Z}^2 \to \mathbb{Q}$ by $f(m, n) = m/n$ if $n \neq 0$ and $f(m, 0) = 0$. ∎

A set X is said to have the **cardinality of the continuum** if $\operatorname{card}(X) = \operatorname{card}(\mathbb{R})$. We shall use the letter $\mathfrak{c}$ as an abbreviation for $\operatorname{card}(\mathbb{R})$:

$$\operatorname{card}(X) = \mathfrak{c} \text{ iff } \operatorname{card}(X) = \operatorname{card}(\mathbb{R}).$$

0.12 Proposition. $\operatorname{card}(\mathcal{P}(\mathbb{N})) = \mathfrak{c}$.

Proof. If $A \subset \mathbb{N}$, define $f(A) \in \mathbb{R}$ to be $\sum_{n \in A} 2^{-n}$ if $\mathbb{N} \setminus A$ is infinite and $1 + \sum_{n \in A} 2^{-n}$ if $\mathbb{N} \setminus A$ is finite. (In the two cases, $f(A)$ is the number whose base-2 decimal expansion is $0.a_1 a_2 \cdots$ or $1.a_1 a_2 \cdots$, where $a_n = 1$ if $n \in A$ and $a_n = 0$ otherwise.) Then $f : \mathcal{P}(\mathbb{N}) \to \mathbb{R}$ is injective. On the other hand, define $g : \mathcal{P}(\mathbb{Z}) \to \mathbb{R}$ by $g(A) = \log(\sum_{n \in A} 2^{-n})$ if A is bounded below and $g(A) = 0$ otherwise. Then g is surjective since every positive real number has a base-2 decimal expansion. Since $\operatorname{card}(\mathcal{P}(\mathbb{Z})) = \operatorname{card}(\mathcal{P}(\mathbb{N}))$, the result follows from the Schröder-Bernstein theorem. ∎

0.13 Corollary. *If* $\operatorname{card}(X) \geq \mathfrak{c}$, *then X is uncountable.*

Proof. Apply Proposition 0.9. ∎

The converse of this corollary is the so-called continuum hypothesis, whose validity is one of the famous undecidable problems of set theory; see §0.7.

0.14 Proposition.

a. If $\text{card}(X) \leq \mathfrak{c}$ *and* $\text{card}(Y) \leq \mathfrak{c}$, *then* $\text{card}(X \times Y) \leq \mathfrak{c}$.

b. If $\text{card}(A) \leq \mathfrak{c}$ *and* $\text{card}(X_\alpha) \leq \mathfrak{c}$ *for all* $\alpha \in A$, *then* $\text{card}(\bigcup_{\alpha \in A} X_\alpha) \leq \mathfrak{c}$.

Proof. For (a) it suffices to take $X = Y = \mathcal{P}(\mathbb{N})$. Define $\phi, \psi : \mathbb{N} \to \mathbb{N}$ by $\phi(n) = 2n$ and $\psi(n) = 2n - 1$. It is then easy to check that the map $f : \mathcal{P}(\mathbb{N})^2 \to \mathcal{P}(\mathbb{N})$ defined by $f(A, B) = \phi(A) \cup \psi(B)$ is bijective. (b) follows from (a) as in the proof of Proposition 0.10. ∎

0.4 MORE ABOUT WELL ORDERED SETS

The material in this section is optional; it is used only in a few exercises and in some notes at the ends of chapters.

Let X be a well ordered set. If $A \subset X$ is nonempty, A has a minimal element, which is its maximal lower bound or **infimum**; we shall denote it by $\inf A$. If A is bounded above, it also has a minimal upper bound or **supremum**, denoted by $\sup A$. If $x \in X$, we define the **initial segment** of x to be

$$I_x = \{y \in X : y < x\}.$$

The elements of I_x are called **predecessors** of x.

The principle of mathematical induction is equivalent to the fact that $\mathbb{N}$ is well ordered. It can be extended to arbitrary well ordered sets as follows:

0.15 The Principle of Transfinite Induction. *Let X be a well ordered set. If A is a subset of X such that $x \in A$ whenever $I_x \subset A$, then $A = X$.*

Proof. If $X \neq A$, let $x = \inf(X \setminus A)$. Then $I_x \subset A$ but $x \notin A$. ∎

0.16 Proposition. *If X is well ordered and $A \subset X$, then $\bigcup_{x \in A} I_x$ is either an initial segment or X itself.*

Proof. Let $J = \bigcup_{x \in A} I_x$. If $J \neq X$, let $b = \inf(X \setminus J)$. If there existed $y \in J$ with $y > b$, we would have $y \in I_x$ for some $x \in A$ and hence $b \in I_x$, contrary to construction. Hence $J \subset I_b$, and it is obvious that $I_b \subset J$. ∎

0.17 Proposition. *If X and Y are well ordered, then either X is order isomorphic to Y, or X is order isomorphic to an initial segment in Y, or Y is order isomorphic to an initial segment in X.*

Proof. Consider the set $\mathcal{F}$ of order isomorphisms whose domains are initial segments in X or X itself and whose ranges are initial segments in Y or Y itself. $\mathcal{F}$ is nonempty since the unique $f : \{\inf X\} \to \{\inf Y\}$ belongs to $\mathcal{F}$, and $\mathcal{F}$ is partially ordered by inclusion (its members being regarded as subsets of $X \times Y$).