5. Well-founded Relations and recursion

We now extract a property of the ϵ-relation on $\mathbb{V}$, because of its importance and use in other situation. For a relation R, we write 'yRx' for '$(y,x) \in R$'.

Definition 5.1 Let R be a Relation on the class A ($R \subset A \times A$), and for any $x \in A$, let us denote the class $\{y : yRx\}$ by R_x; elements of R_x are called R-predecessor of x. R is called well-founded (wf) if the following holds:

> whenever $X \subset A$, and for all $x \in A$, $R_x \subset X$ implies $x \in X$, then $X = A$.

As an example, consider $A = \mathbb{V}$, and let R be the ϵ-relation on $\mathbb{V}$: $R = \{(y,x) : y \in x\}$. Now we have $R_x = x$. R is well-founded, and this fact precisely expresses the fact V is the smallest class closed under set-formation.

It is easy to see that the restriction of a wf relation to a subclass is again wf. (If $R \subset A \times A$, and $B \subset A$, then the restriction of R to B is $R \cap (B \times B)$). It follows that the ϵ-relation on $\mathbb{O}rd$, $< = \{(\alpha,\beta) : \alpha,\beta \in \mathbb{O}rd, \alpha \in \beta\}$ is again wf. Again, we have $<_\alpha = \alpha$. As an even more special case, we obtain that $<$ on $\mathbb{N}$ is wf.

Another example is the following relation R on $\mathbb{N}$ (or, on $\mathbb{N}$, in any Peano system $(\mathbb{N},0,S)$):

$$mRn \underset{df}{\Leftrightarrow} Sm = n.$$

Now, $R_n = \{n-1\}$ if $n \neq 0$, $R_0 = \emptyset$ (see 4.1'). The fact that

the latter relation is wf also comes from a general principle; this and the one about 'restrictions' are both contained in

Proposition 5.2 If $R \subset A \times A$, and R is wf on A (note that the definition 'wf' depends on A too!), then for any $R' \subset R$ and B such that $R' \subset B \times B$, R' is wf on B.

Proof: exercise.*

So, in particular, R being wf does not depend on which class A is taken (such that $R \subset A \times A$) in the definition. Now, the last example comes from the proposition, since $Sm = n$ implies $m < n$.

Every wf relation R on A has a corresponding induction principle associated to it, which is merely a reformulation of the definition of 'wf', and which reads as follows. Assuming that for an assertion $P(x)$ for elements x of A we have that from the fact that $P(y)$ holds for all y such that yRx, it follows that $P(x)$, we may conclude that $P(x)$ holds for all $x \in A$. Note carefully that for the first example, this is $\in$-induction, and for the last, ordinary (complete) induction. The case for ordinals deserves separate mention.

Principle of Transfinite Induction. Assuming that an assertion $P(\alpha)$ concerning ordinals α holds whenever it holds for all $\beta < \alpha$, we may conclude that $P(\alpha)$ is true for all ordinals α.

An easily verified equivalent definition of 'wf' is contained in

* It is preferable to do this after Proposition 5.7

Proposition 5.3 R is wf on A iff the following holds: whenever $X \subset A$ and $X \neq \emptyset$, then there is an R-minimal element x in X: $x \in X$ but no R-predecessor of x is in X.

Proof: exercise.

Let us see what this proposition means in the case of $A = \mathbb{O}rd$, $R = \in$ on $\mathbb{O}rd$. Let $X \subset \mathbb{O}rd$ be non-empty, let $x \in X$ $\in$-minimal ($<$-minimal) in X. Since trichotomy holds (4.9), every $y \in X$ is comparable to x; since x is minimal, $y < x$ is impossible. Hence, for all $y \in X$, we have $x \leq y$. In other words, x is the least element of X; clearly, only one such is possible, since $y \leq x$ and $x \leq y$ imply $y = x$ (why?). We have obtained: for every non-empty class of ordinals, there is a unique least member of the class.

One way we use this principle in practice is a method of definition of functions with ordinal values. Let A be a class, and assume that for each $a \in A$ we have a property $P(-,a)$ of ordinals (for the blank), the property itself depending on a, and assume that we have established that, for all $a \in A$, there is at least one ordinal x satisfying $P(x,a)$. Then we can pick the least such x, for any fixed $a \in A$, and define a Function with Domain A,

$$F : A \longrightarrow Ord$$

such that $F(a)$ = the least ordinal y such that $P(y,a)$ holds. In terms of class comprehension:

$$F = \{(a,x) : a \in A, x \in \mathbb{O}rd, P(x,a), \text{ and for all } y \in x, \text{ not } P(y,a)\}.$$

Corollary 5.3' For any non-empty set x, there is $y \in x$ such that $x \cap y = \emptyset$.

Proof: exercise.

The statement in 5.3' is called the axiom of foundation, in 'set-form'. It expresses precisely the fact that $\epsilon = \{(x,y) : x \in y\}$ is a wf Relation on V. Namely, let $x = X$ be a non-empty set; then an element y of x with $x \cap y = \emptyset$ is the same as an ϵ-minimal element of x. It can be shown that the well-foundedness of ϵ, in the formulation of 5.3, is a consequence of 5.3' (i.e., we can show the general form of 5.3, with X a class, from the special case with X restricted to be a set).

Given any wf relation R on a class A, we can uniquely define a Function F with domain A once we have a prescription how to compute $F(x)$ $(x \in A)$ from the values of F at R-predecessors of x. This fact is called the recursion principle (for R). This principle is one of the most important theorems of set theory. Definition of functions by recursion along various wf relations is an important ingredient of constructions whenever set theory is used, and in fact, even in more common mathematics, in the case of recursions on natural numbers.

To give an exact form to the principle, we have to formulate what we mean by "computing from the values of F at R-predecessors of x". Given a function $F : A \longrightarrow B$, the object that contains the total information concerning the values of F at R-predecessors of x is, clearly, the restriction $F \restriction R_x$.

Thus, the data in a definition by recursion should include a rule that assigns to any x in A and any object of the form $F \restriction R_x$, preferably to any Function $F' : R_x \longrightarrow B$, an element b of B; the intention is that $F(x)$ be b. In order to formulate this in our present framework, we have to assume that

for all x, R_x is a set.

This last condition will be assumed and, in fact, be considered as a part of the definition of well-foundedness. Note that the examples all satisfy the additional condition. If, under these conditions, $F : A \longrightarrow B$ is any Function, then $F \restriction R_x$ is in fact a set, by the axiom of replacement.

Theorem 5.4 (Principle of Recursion) Suppose R is a wf Relation on A; B another class, and C the class of all pairs (x,f) where $x \in A$ and f is a function of the form $f : R_x \longrightarrow B$. Suppose

$$G : C \longrightarrow B$$

is given. Then there is a unique Function $F : A \longrightarrow B$ that satisfies the identity

$$F(x) = G(x, F \restriction R_x)$$

for all $x \in A$.

Note carefully that 5.4 specializes to 4.3 in case $A = \mathbb{N}$, and R is defined by $mRn \Leftrightarrow Sm = n$.

Proof: The recursion principle in the case of the natural numbers (in the formulation of 4.3) is justified, intuitively, by saying that the recursion equations allow us to compute all values of f step-by-step:

$f(0) = a$,
$f(1) = f(S0) = g(f(0)) = g(a)$,
$f(2) = f(SS0) = g(f(S0)) = g(g(a))$,
etc.

In fact, at any finite stage n , only a finite *approximation* of f is computed, namely $f\restriction\{k : k\leq n\}$; the total function f is the 'limit' , formally the union of all of its finite approximations:

$$f = \bigcup \{ f\restriction\{k : k\leq n\} : n\in\mathbb{N} \}$$

This description is not far from the actual proof of the general case. On the other hand, note that in the proof of the general case, we use the word "Approximation" (the capital is because Approximations may be proper classes) in a generalized sense so that the total function F itself is an Approximation. Now, we turn to the proof proper.

We are given the data as in the statement of the principle: the classes A, R, B, and G satisfying various hypotheses. Let us call a subclass D of A *closed* if for all $x\in A$, $x\in D$ implies $R_x \subset D$. In other words, D is a closed subclass of A is $x\in D$ and yRx imply that $y\in D$.

Note, first of all, that A itself is a closed class. Next, note that if R is, in particular, a transitive relation, then for any $x\in A$, the set R_x is closed, and it is the least closed subclass of A containing x as an element.

By definition, an *Approximation* (of the Function F to be defined) is a Function $E : \mathrm{Dom}(E) \longrightarrow B$ such that the following conditions are satisfied:

(i) $D \underset{\mathrm{def}}{=} \mathrm{Dom}(E)$ is a closed subclass of A ;

(ii) for every $x\in D$, the recursion equation

$$E(x) = G(x, E\restriction R_x)$$

holds true. (Since D is closed, and $x\in D$, we have $R_x \subset D$, thus $E\restriction R_x$ is a function of the

form $R_x \rightarrow B$; thus $E \restriction R_x \in C$, and so $G(x, E \restriction R_x)$ makes sense.)

Claim 1. Any two approximation E_1 and E_2 coincide on the common part of their domains: $E_1(x) = E_2(x)$ for all $x \in D \underset{\text{def}}{=} \text{Dom}(E_1) \cap \text{Dom}(E_2)$.

Proof of Claim 1. The proof is an induction on x along R (that is, it is an application of the induction principle associated with R). Suppose $x \in D$. The induction hypothesis is that for all $y \in D$ such that yRx, if $y \in D$, then $E_1(y) = E_2(y)$. But, since D is closed (it is easy to see that the intersection of closed classes is closed), we have that $R_x \subset D$, and so for all $y \in R_x$, $E_1(y) = E_2(y)$. This says that $E_1 \restriction R_x = E_2 \restriction R_x$. But then, by (ii),

$$E_1(x) = G(x, E_1 \restriction R_x) = G(x, E_2 \restriction R_x) = E_2(x)$$

as required.

end of proof of Claim 1.

We make the following definition:

$$F \underset{\text{def}}{=} \{a: \text{there is an Approximation } E \text{ such that } a \in E\}. \quad (1)$$

Remarks. The definition of the class F says that F is the union of all the Approximations; in particular, for all Approximations E, we have $E \subset F$. However, an Approximation is not necessarily a set; therefore, we cannot talk about the class of all Approximations; if we could, and that class were $Approx$, then the short description of F would be $F = \bigcup Approx$. When the originally given classes A and B are, in particular, sets, then all Approximations are sets, and the class $Approx$ exists and can be used as shown.

Note that since all elements of an Approximation are ordered pairs, so are all elements of F.

Claim 2. F is an approximation.

Proof of Claim 2. *First of all*, we show that F is a Function. As we noted, all

elements of F are ordered pairs (so, F is a Relation). If (x, u_1), (x, u_2) are both in F, then there are approximations E_1, E_2 with $(x, u_1) \in E_1$, $(x, u_2) \in E_2$; by Claim 1, it follows that $u_1 = u_2$, as desired.

Secondly, $\mathrm{Dom}(F)$ is closed. Indeed, if $x \in \mathrm{Dom}(F)$, there is an approximation E with $x \in \mathrm{Dom}(E)$; by the definition of "approximation", $R_x \subset \mathrm{Dom}(E)$; but $\mathrm{Dom}(E) \subset \mathrm{Dom}(F)$; so, $R_x \subset \mathrm{Dom}(F)$.

Thirdly, we have the functional equation:

$$F(x) = G(x, F \restriction R_x) \quad (x \in \mathrm{Dom}(F)). \qquad (2)$$

The reason is that if $x \in \mathrm{Dom}(F)$, there is an approximation E such that $x \in \mathrm{Dom}(E)$; since E is an approximation, we have that

$$E(x) = G(x, E \restriction R_x);$$

but $E \subset F$; thus, $F(x) = E(x)$ and $F \restriction R_x = E \restriction R_x$ (note that $R_x \subset \mathrm{Dom}(E)$); (2) follows.

Claim(2) done.

Claim 3. For every $x \in A$, there is an approximation E with $x \in \mathrm{Dom}(E)$.

Proof of Claim 3. The proof is by R-induction (see middle of page 52). Let $x \in A$, and suppose the assertion holds for all $y \in R_x$. Let $y \in R_x$; by the induction hypothesis, there is an approximation E such that $y \in \mathrm{Dom}(E)$. By the definition of F (see (1)), we have $\mathrm{Dom}(E) \subset \mathrm{Dom}(F)$; thus, $y \in \mathrm{Dom}(F)$. (In other words, the one and the same approximation, namely F, works for all y's in R_x!) Now, consider

$$F^* = F \cup \{ (x, G(x, F \restriction R_x)) \}.$$

We claim that F^* is an approximation. F^* is a function; indeed, since F is a Function, F^* could possibly fail to be a function only if $x \in \mathrm{Dom}(F)$; in that case, since F is an approximation, $F(x) = G(x, F \restriction R_x)$, which is the value of F^* on x "added" to F; so,

after all, there is no "conflict", and F^* is a Function. $\mathrm{Dom}(F^*)=\mathrm{Dom}(F)\cup\{x\}$; thus, it is closed, since $\mathrm{Dom}(F)$ is closed, and $R_x \subset \mathrm{Dom}(F) \subset \mathrm{Dom}(F^*)$. Finally, the functional equation holds for F^* since it holds for all $y \in \mathrm{Dom}(F)$ (F is an approximation), and the definition *makes it hold* for x itself. Of course, $x \in \mathrm{Dom}(F^*)$.

Claim 3 done.

Remark. Note that we can see that F^* in the previous proof is in fact the same as F , since, being an Approximation, $F^* \subset F$, and, by definition, $F \subset F^*$.

Proof of the Principle of Recursion. Returning to the definition of F , we see that Claim 3 says that $\mathrm{Dom}(F) = A$. Since F was already shown to be an approximation, F satisfies all the requirements of the theorem. The uniqueness of F is a consequence of Claim 2.

end of the proof "Principle of Recursion"

Let us make a remark on a possible simplified formulation of the recursion principle. Suppose, in addition to the conditions of the statement of the principle, that $R_x=R_y$ implies that $x=y$. Then the recursion scheme can be equivalently rewritten as

$$F(x) = G(F \restriction R_x) \ ;$$

that is, we may take G to be a one-variable function instead of the original two-variable one. The reason is that, in this case, x can be recaptured from R_x , hence the additional argument x in the original formulation does not carry additional information. All the examples in this section satisfy the additional condition.

We have already mentioned that recursion for natural numbers, 4.3, is a particular case of the general principle. Let us note that if we again take N for A, but this time for the relation R we take $<$, we obtain the following form of recursion. Since now $R_n = n = \{0,\ldots,n-1\}$, the recursion equation

$$F(n) = G(F \restriction \{0,\ldots,n-1\}) \qquad (3)$$

tells us that $F(n)$ is determined by G once we know the values $F(0),\ldots,F(n-1)$; the <u>previous</u> values of F. In the version of 4.3 $F(n)$ has to be determined from $F(n-1)$ alone (if $n > 0$, and it is determined outright if $n = 0$). Thus, in a sense, the second version is more powerful; it allows more complicated conditions to govern the values of F. Note that for $n = 0$, (3) gives $F(0) = G(\emptyset)$, thus again, $F(0)$ is, in fact, determined outright (despite the fact that this is not mentioned separately in the definition).

We can use recursion on natural numbers to generate many ordinals. By recursion, we may define the function $f : \omega \longrightarrow \mathbb{O}rd$ by the requirements

$$f(0) = \omega$$

$$f(Sn) = S(f(n)).$$

We obtain $f(1) = S\omega$, $f(2) = SS\omega$, etc. Notice that if $\alpha < f(n)$ for some $n \in \omega$, then either $\alpha \in N$, or $\alpha = f(k)$ for some $k < n$ (exercise; use induction on n). In other words

$$\omega \cup \text{range}(f)$$

is a transitive set; therefore it is an ordinal itself. At the end of Section 2, what we called $\omega+n$ is $f(n)$, and what we called $\omega+\omega$, or $\omega \cdot 2$, is the last defined ordinal.

By transfinite recursion we mean recursion on Ord, or a subclass of Ord, with respect to the wf relation $<$. To repeat, this takes the form

$$F(\alpha) = G(F \restriction \alpha) \tag{4}$$

(now, $R_\alpha = \{\beta : \beta < \alpha\} = \alpha$); with a given function G. Usually, G is given partly verbally, by spelling out in words and symbols how to obtain $F(\alpha)$ from $F \restriction \alpha$. Class-comprehension can always be invoked to get a Function G so that the verbal recursion will be identified with the formal one as in (4). One frequent occurrence is the distinction of three cases: $\alpha = 0$, α successor, α limit, in giving $F(\alpha)$ from previous values. Consider the following definition:

$$\left.\begin{aligned} \mathbb{V}_0 &= \emptyset \\ \mathbb{V}_{\beta+1} &= P(\mathbb{V}_\beta) \\ \mathbb{V}_\alpha &= \bigcup_{\beta<\alpha} \mathbb{V}_\beta \qquad (\alpha \text{ limit}) \end{aligned}\right\} \tag{5}$$

First of all, from now on we write $\beta+1$ for $S\beta$ even though we have not yet defined $+$ for ordinals. The equations (5) are intended to define the Function $F : \mathbb{O}rd \longrightarrow \mathbb{V}$; we wrote $\mathbb{V}_\alpha$ for $F(\alpha)$. It is a proper recursion since every ordinal α falls in exactly one of the three categories: α is 0, or α is $\beta+1$ for a unique β, or α is limit. In each case, the function value is given in terms of function values of arguments less than α. Formally, we can define

$$G : C \longrightarrow \mathbb{V},$$

with C = the class of functions f with domain some ordinal, as follows

$\langle f,a\rangle \in G \Leftrightarrow f \in C$ and either (dom $f = 0$ and $a = \emptyset$)
or for some $\beta \in \mathbb{O}rd$, (dom $f = \beta+1$ and $a = P(f(\beta))$)
or dom f is a limit ordinal and $a = \cup$ range(f).

Our preceding arguments amount to a proof of the fact that G is a Function (exercise), and then clearly, equation (4) will be equivalent to the system (5) (exercise).

Note that we have now defined precisely the so-called cumulative hierarchy of pure sets, begun at the end of Section 2.

Proposition 5.5 The cumulative hierarchy is increasing: $\mathbb{V}_\alpha \subset \mathbb{V}_\beta$ for $\alpha < \beta$. Moreover, we have

$$\mathbb{V} = \bigcup_{\alpha \in \mathbb{O}rd} \mathbb{V}_\alpha$$

Also: each $\mathbb{V}_\alpha$ is a transitive set [should be proved first!]

i.e. for every x there is $\alpha \in Ord$ such that $x \in \mathbb{V}_\alpha$.

Proof: The proof of the first part is left to the reader (exercise). To prove the second part it suffices to show that $\bigcup_{\alpha \in Ord} \mathbb{V}_\alpha$ $(= \bigcup \text{ range } \mathbb{V}_{(\cdot)})$ is closed under set-formation. So, suppose that $x \subset \bigcup_{\alpha \in Ord} \mathbb{V}_\alpha$. This means that for every $y \in x$ there is an ordinal α such that $y \in \mathbb{V}_\alpha$. Define the function $r : x \longrightarrow Ord$ by the formula:

$r(y)$ = the smallest ordinal α for which $y \in \mathbb{V}_\alpha$.

Consider the set $\{r(y) : y \in x\} = \text{range}(r)$. Apply 4.11 to obtain an ordinal β such that $r(y) < \beta$ for all $y \in x$. Then $y \in \mathbb{V}_{\alpha_y} \subset \mathbb{V}_\beta$ for all $y \in x$, i.e. $x \subset \mathbb{V}_\beta$, and thus $x \in P(\mathbb{V}_\beta) = \mathbb{V}_{\beta+1} \subset \bigcup_{\alpha \in Ord} \mathbb{V}_\alpha$. □

In more advanced set theory, one encounters complicated recursive constructions. E.g., we may construct several objects at the same 'stage' α, with α ranging over all ordinals, or all ordinals less than a fixed one. In the construction, the items constructed at 'stages' $\beta < \alpha$ are used. Such a recursion will fall under our general scheme of transfinite recursion if we form a single object using ordered pairs, triples, etc., of the ones to be constructed. A more important aspect of some of these constructions is the fact that the construction of the items in stage α depends on certain properties of the items constructed at previous stages; simply, the construction cannot be carried out without the previous items having these

properties. Then, of course, it is an integral part of the construction to show that the items constructed at stage α continue to have these necessary properties. Such a recursion could be called an instance of transfinite recursion with induction hypotheses. We will encounter the first such recursion in the proof of the compactness theorem for propositional logic, Theorem 1.6.5 in Part II.

It turns out that recursion with induction hypotheses can also be subsumed under the general framework of our formal principle. The idea is simply that we may "ignore" the induction hypotheses and, at stage α, say that if these hypotheses do not hold then the items to be constructed at stage α are taken to be some arbitrary (irrelevant) objects. The function defined by the modified recursion (without induction hypotheses) will be the same as the one intended by the original recursion. The reason for this is that, as a matter of fact, the induction hypotheses will hold at every stage (and thus, we'll never have had to resort to the "irrelevant objects"), as a consequence of the fact that the induction hypotheses 'propagate themselves' from earlier stages to later stages.

Here is a formal way of making the above somewhat speculative considerations precise.

Let us confine ourselves to the case when R is a transitive well-founded relation on A (which is the case when $A = \mathbb{O}rd$ and $R = \epsilon$). We consider a Function G whose domain is a subclass of the class of all pairs (x,f) with functions f of the form $f : R_x \longrightarrow B$ and for which $G(x,f) \epsilon B$. This reflects

the situation that the 'construction' of $F(x) = G(x, F \restriction R_x)$ may be carried out only if $f = F \restriction R_x$ satisfies certain conditions (ensuring that $(x,f) \in \mathrm{Dom}\ G$). We also consider a relation $I \subset \mathbb{V} \times \mathbb{V}$ representing the "induction hypothesis" (the sum total of all the induction hypotheses). The essential conditions that we have to assume are the following.

(i) Whenever $f : R_x \longrightarrow B$ is such that $(y, f \restriction (R_y \cup \{y\})) \in I$ for all y with yRx, then $(x,f) \in \mathrm{Dom}(G)$; and

(ii) Whenever $\hat{f} : R_x \cup \{x\} \longrightarrow B$ is such that $(x, \hat{f} \restriction R_x) \in \mathrm{Dom}\ G$, $\hat{f}(x) = G(x, \hat{f} \restriction R_x)$ and $(y, \hat{f} \restriction (R_y \cup \{y\})) \in I$ for all y with yRx, then $(x,\hat{f}) \in I$.

In those conditions, "$(y, F \restriction (R_y \cup \{y\})) \in I$" expresses the "induction hypothesis at y", for F the function to be defined. Of course, the conditions have to be expressed <u>without</u> referring to F the proof of whose existence is the final aim. (i) expresses that the induction hypothesis suffices for being able to carry out the construction given by G; (ii) expresses that the induction hypothesis propogates from earlier stages to later stages.

<u>Theorem</u> 5.6 (Recursion with induction hypotheses)

Under the above notations and conditions (including (i) and (ii)), there is a unique Function $F : A \longrightarrow B$ such that for all $x \in A$, $(x, F \restriction x) \in \mathrm{Dom}\ G$, and $F(x) = G(x, F \restriction x)$.

<u>Proof</u>. Let us define G^* to be any function $G^* : \mathbb{V} \longrightarrow B$

extending G (e.g., for any $x \notin \mathrm{Dom}\ G$, let us put $G^*(x) = b$, for a fixed element b of B). Using 5.4, we let $F : A \longrightarrow B$ be a Function such that

$$F(x) = G^*(x, F \restriction R_x) \tag{6}$$

for all $x \in A$. We claim that for F so defined, we have $(x, F \restriction (R_x \cup \{x\})) \in I$ for all $x \in A$. Indeed, suppose $x \in A$ and that this holds for all y (y in place of x) such that $y \in R_x$ (induction hypothesis). Then for $\hat{f} = F \restriction (R_x \cup \{x\})$ and $f = \hat{f} \restriction R_x$ we have $f \restriction (R_y \cup \{y\}) = F \restriction (R_y \cup \{y\})$ ($y \in R_x$), thus $(y, f \restriction (R_y \cup \{y\})) \in I$ by the induction hypothesis. Hence, by condition (i), we have $(x,f) \in \mathrm{Dom}\ G$. Note that, since $G^*(a) = G(a)$ for any $a \in \mathrm{Dom}\ G$, we have, by (6), that $\hat{f}(x) = G(x, \hat{f} \restriction R_x)$. Thus, by (ii), we conclude that $(x,\hat{f}) \in I$, completing the proof of the claim.

Now, applying (i) again, by the claim it follows that for every $x \in A$, $(x, F \restriction R_x) \in \mathrm{Dom}\ G$, and thus, by (6), $F(x) = G(x, F \restriction R_x)$ as desired. □

6. Indexing by ordinals and the axiom of choice

Counting a set x means calling an element of it the first, another the second, another the third, etc. We change this slightly by starting with 0 instead of 1. The counting of x is completed, and x is shown to be finite, if for some natural number n every member of x has been called the k^{th} for some $k < n$. In a formal language, this means a function

$$f : \{k : k < n\} = n \longrightarrow x$$

which is one-to-one and onto; $f(k)$ is the k^{th} element of x. Having such an n-indexing of x we say that "x has n elements".

Ordinals are used to count possibly infinite sets. For an ordinal, an α-indexing (or: α-counting) of x is a function

$$f : \{\beta : \beta < \alpha\} = \alpha \longrightarrow x$$

which is one-to-one and onto. Usually, an α-indexing is denoted by a notation like

$$\langle y_\beta \rangle_{\beta<\alpha}$$

here $y_\beta = f(\beta)$ with f from the previous indexing.

The following defines an ω-indexing of the (positive, negative, and zero) integers:

$$y_{2n} = +n \qquad (n \in \mathbb{N})$$
$$y_{2n+1} = -(n+1) \qquad (n \in \mathbb{N})$$

Of course, the identity function is an ω-indexing of $\mathbb{N}$. But there is an $\omega+1$-indexing of $\mathbb{N}$ as well:

$$y_n = n+1 \qquad (n \in \mathbb{N})$$
$$y_\omega = 0.$$

This possibility of having indexings of the same set with different ordinals as domains is an anomaly one is not used to in the realm of finite things. Certainly, if the two countings induce us to say that $\mathbb{N}$ has ω elements as well as $\omega+1$ elements, then there is something left to be clarified. This will be done in later sections in the theory of cardinal numbers.

The main question we are interested in here is whether every set can be counted. In fact, we will prove

Theorem 6.1 For every set X, there is an ordinal α and an α-indexing of X.

The idea of the proof of this theorem is very simple. If X is empty, then $\alpha = 0$ is appropriate. Otherwise, X has an element. We pick one, and call it the 'zeroth', y_0. If there are no more elements of X, we stop and put $\alpha = 1$. Otherwise, we pick another element $y_1 \in X - \{y_0\}$, and 'continue'. E.g., if we have defined y_n for all $n < \omega$, we ask whether $X = \{y_n : n < \omega\}$; if not we pick a 'next' element y_ω in the difference. This means that we define y_α for increasing ordinals α as long as X is not exhausted. Since there are 'many ordinals' (4.11, 4.12), it turns out that X has to be exhausted sooner or later. At that point, we have completed the definition of the indexing.

Ostensibly, the above construction is one by transfinite

recursion. Let us formalize it in the form (4). First of all, we will construe the indexing as a function on the whole class Ord, since we do not know in advance what the domain will be. We take an entity $*$ (e.g. X itself) that is not in X and we'll define the value of the function F to be $*$ once we have exhausted X. G now should be defined so that, as a consequence of (4), we would have that

$$F(\alpha) = \text{an element of } X - \text{range}(F \restriction \alpha)$$
$$\text{if this difference is non-empty;}$$
$$* \text{ otherwise.}$$

In other words, we need G, a Function defined on the class C of all functions with domains ordinals, such that for $f \in C$.

$$G(f) \in X - \text{range}(f) \quad \text{if this difference is non-empty}$$

and $\quad G(f) = *$ otherwise.

It turns out that without postulating a new principle of our set theory, we cannot prove the existence of such a G. Clearly, what we need is a Function H that assigns an element of x to every non-empty set x.

Axiom of Choice. There is a Function $H : \mathbb{V} - \{\emptyset\} \longrightarrow \mathbb{V}$ such that for all $x \in \mathbb{V} - \{\emptyset\}$, $H(x) \in x$.

For future reference, and to avoid repetition, we now formulate and prove a more detailed version of 6.1.

A *choice function for non-empty subsets* of a set X is a function

$$f : \mathcal{P}(X) - \{\emptyset\} \longrightarrow X$$

such that for any $A \subset X$, $A \neq \emptyset$, we have $f(A) \in A$. When H is a "global" choice function as in the formulation of the Axiom of Choice on p. 65, then $f \underset{def}{=} H \restriction \mathcal{P}(X) - \{\emptyset\}$ is a choice function for non-empty subsets of X.

Proposition 6.1'. Let f be a choice function for non-empty subsets of X. Define the Function $F: \mathbf{Ord} \longrightarrow \mathbf{V}$ by transfinite recursion by the formula

$$F(\alpha) = \begin{cases} f(X\text{-range}(F \restriction \alpha)) & \text{when} \quad X\text{-range}(F \restriction \alpha) \neq \emptyset \\ X & \text{otherwise} \end{cases}$$

Then there is an ordinal α such that $F(\alpha) = X$, and for *the least* such α, we have that

$$F \restriction \alpha \; : \; \alpha \xrightarrow{\cong} X \; :$$

that is, $F \restriction \alpha$ is an α-indexing of X.

Proof. First, we make a couple of remarks. For any ordinal γ,

either $F(\gamma) \in X$, or $F(\gamma) = X$, and of course, not both.

This is because the first clause of the definition of F makes $F(\gamma)$ an element of X, since all values of f are elements of X. Next,

if $\beta < \gamma$ *and* $F(\gamma) \in X$, *then* $F(\beta) \in X$ *as well, and* $F(\beta) \neq F(\gamma)$. (1)

The reason is that, assuming $\beta<\gamma$ and $F(\gamma)\in X$, we have that $X-\mathtt{range}(F\restriction\gamma)\neq\emptyset$, hence, since $X-\mathtt{range}(F\restriction\gamma)\subset X-\mathtt{range}(F\restriction\beta)$ (right?), also $X-\mathtt{range}(F\restriction\beta)\neq\emptyset$, therefore, for β too, the first clause applies, making $F(\beta)\in f(X-\mathtt{range}(F\restriction\beta))$. In particular, $F(\beta)\in X$. But also, since $F(\gamma)=f(X-\mathtt{range}(F\restriction\gamma))\in X-\mathtt{range}(F\restriction\gamma)$, and $F(\beta)\in\mathtt{range}(F\restriction\gamma)$ (right?), it follows that $F(\beta)\neq F(\gamma)$.

Let us put $Y\underset{\mathtt{def}}{=}\{\beta\in\mathbf{Ord}: F(\beta)\in X\}$. By (1) the class Y is transitive: $\beta<\gamma$ and $\gamma\in Y$ imply $x\in Y$. Moreover, $F\restriction Y$ is one-to-one: for $\beta\neq\gamma$, both in Y, either $\beta<\gamma$ or $\gamma<\beta$, and so by (1) again, $F(\beta)\neq F(\gamma)$. We have that

$$F\restriction Y : Y\xrightarrow{\cong}\mathtt{range}(F\restriction Y)\ \underset{(!)}{\subset}\ X \qquad (2)$$

from which, the inverse Function $(F\restriction Y)^{-1}$ exists, $\mathtt{range}(F\restriction Y)$ is a set, and

$$(F\restriction Y)^{-1} : \mathtt{range}(F\restriction Y)\xrightarrow{\cong} Y\ ,$$

and so, by Replacement, Y *is a set.* Y, being a transitive set of ordinals, is an ordinal; we write α for Y. Finally, I claim that $\mathtt{range}(F\restriction\alpha)=X$. Indeed, otherwise $\mathtt{range}(F\restriction\alpha)\underset{\neq}{\subset}X$, which, according to the definition of F, makes $F(\alpha)=f(X-\mathtt{range}(F\restriction\alpha))\in X$, therefore, by the definition of Y, $\alpha\in Y$; but this means $\alpha\in\alpha$: contradiction.

We have proved that $F\restriction\alpha : \alpha\xrightarrow{\cong}X$ as desired. It is also clear that α is indeed the least ordinal for which $F(\alpha)=X$ (why?).

The axiom of choice is a principle entirely different from the other principles of set theory. The form we gave it above is called the principle of global choice. A somewhat weaker form says that for any set X of non-empty sets, there is a function h with domain X such that $h(x) \in x$ for all $x \in X$.

This principle is intuitively clear. If X is, in particular, a finite set, the statement can be proved without using any axiom of choice. But already for the case of a set

$$X = \{x_n : n \in \omega\}$$

indexed by ω, the axiom cannot be proved from the other principles of set theory.

7. Well-orderings

The conclusion of the ordinal indexing theorem 6.1 is usually expressed in another way. This way does not mention ordinals; rather it starts with an abstract treatment of the 'ordering' relation obtained from an indexing: y_α is less than y_β iff $\alpha < \beta$. We step back and introduce a series of commonly used concepts.

A Relation* R on a class A is called a quasi-ordering of A if the following are satisfied for all a, b, c in A (we write aRb for $\langle a,b\rangle \in R$):

aRa (reflexivity)

aRb and bRc imply aRc (transitivity).

If, in addition, we also have

aRb and bRa imply $a = b$,

R is a (reflexive) partial ordering of A.

An example for a partial ordering on V is containment:

$aRb \underset{df}{\Leftrightarrow} a \subset b$.

An equivalence, or equivalence relation, on A is a quasi-ordering which is also symmetric:

aRb implies bRa.

If R is a quasi-ordering on A, we can define E by

$$aEb \underset{df}{\Leftrightarrow} aRb \text{ and } bRa;$$

E will be an equivalence relation (exercise).

Equivalence relations are "generalized equality relations". Let E be an equivalence relation on A, and for $a \in A$, let

* From now on, we abandon the use of capitals, such as in "Relation," for indicating that we have a possibly proper class in mind. In other words, when we say: 'let R be a relation on a class A', we mean that

a/E denote $\{b \in A : aEb\}$; a/E is called the equivalence class of a. We have the following fact:

$$aEb \Leftrightarrow a/E = b/E$$

(exercise). "Passing to equivalence classes changes equivalence into equality." Taking equivalence classes is a common device in mathematics.

Suppose R is a quasi-order on A, E the associated equivalence relation. Assume that each a/E $(a \in A)$ is a set. Then on the class $A/E = \{a/E : a \in A\}$ we can define $\hat{R}$:

$$(a/E)\hat{R}(b/E) \underset{df}{\Leftrightarrow} aRb$$

(exercise; one has to verify that this definition is legitimate) and the relation $\hat{R}$ on A/E so obtained is a partial ordering (exercise). The partial ordering $\hat{R}$ is said to be derived from R. If the equivalence classes a/E are not necessarily sets, the above constructions are not permitted in our framework; we cannot take a class of possibly proper classes. In that case still we can say the following. If we let A' be any subclass of A such that for the restriction of $\leq$ to A' and E' derived from this restriction, $a/E' = \{b \in A' : bE'a(\Leftrightarrow bEa)\}$ is a set for any $a \in A'$, then the above construction yields a partial ordering on $A'/E' = \{a/E' : a \in A'\}$. In particular, this is true if A' itself is a set (why?). In short, the construction is valid when one restricts attention to any fixed subset of A.

A total, or linear, ordering of A is a partial ordering R

that satisfies

either aRb, or bRa.

Linear orderings usually are given in the irreflexive form. In fact, partial orderings can also be given that way. Suppose S is a relation on A satisfying:

not aSa

aSb and bSc imply aSc.

Such an S can be called an irreflexive partial ordering. Then the relation* R on A defined by

$$aRb \underset{df}{\Leftrightarrow} aSb \text{ or } a = b$$

is a (reflexive) partial ordering (exercise). Conversely, if R is a reflexive partial ordering on A, and we define S by

$$aSb \underset{df}{\Leftrightarrow} aRb \text{ and } a \neq b$$

then S is an irreflexive partial ordering (exercise). In fact, if we now define R' as R was defined above from S, we get back our R itself (exercise). Also, if we start with an irreflexive S, pass to the corresponding reflexive R, and then again, to the corresponding irreflexive S', then S' = S (exercise). In other words, it is practically the same to talk about the two kinds of partial orderings.

With partial orderings, it is more customary to deal with the reflexive formulation; with linear orderings, with the irreflexive one. Let us denote an irreflexive linear ordering on A by <. Then the requirements on < become:

$a \nless a$

* see the footnote on p. 68

$a < b$ and $b < c$ imply $a < c$

either $a < b$, or $a = b$, or $b < a$

(exercise).

If a quasi-ordering R on A satisfies

either aRb or bRa

($a,b \in A$), then the derived partial ordering is a linear ordering (exercise).

A well-founded* linear ordering is called a well-ordering. The prime example of a well-ordering is the one we defined on $\mathbb{O}rd$: $\in$ restricted to $\mathbb{O}rd$ (see the last two Sections). Recall that in 5.3 we talked about R-minimal elements in a subclass X of A. If R is a linear ordering, an R-minimal element in X, if it exists, is unique (exercise). Therefore, a linear ordering $<$ on A is a well-ordering means that for every non-empty $X \subset A$, there is a unique least (with respect to $<$) element in X.

If $<$ is a linear ordering of A, then for any $B \subset A$, $<$ restricted to B ($= <\cap(B\times B)$) is a linear ordering of B; if the first is a well-ordering, so is the second (exercise). In particular, any set of ordinals, and any ordinal itself, is well-ordered by the $\in$-relation.

---

* Recall (see p. 55) that 'well-founded' includes the condition that "R_x is a set".

Since the last-mentioned class of well-orderings is fundamental, we emphasize them. Any ordinal α has a natural well-ordering on it: the $\in$-relation restricted to it: $\in \restriction \alpha = \in \cap (\alpha \times \alpha)$; we have the well-ordering $(\alpha, \in \restriction \alpha)$. Sometimes, by an 'ordinal' we mean an ordinal with its natural well-ordering; thus, when we say that "every well-ordering (on a set) is similar to an ordinal" (see below), we have the natural wellordering of an ordinal in mind.

In a definite sense, talking about indexings by ordinals is the same as talking about well-orderings. Suppose $<$ is a well-ordering of A, and $F : A \longrightarrow B$ is a one-to-one and onto Function. Then the relation $<<$ on B defined by

$$F(a) << F(a') \Leftrightarrow a < a'$$

is a well-ordering as well (exercise). Thus if $\langle y_\beta \rangle_{\beta<\alpha}$ is an

α-indexing of the set x, then the relation $<<$ on x defined by

$$y_\beta << y_\gamma \Leftrightarrow \beta < \gamma$$

is a well-ordering of x. Conversely, we have

Proposition 7.1 For any well-ordering $<<$ of a set x, there is a unique α-indexing $\langle y_\beta \rangle_{\beta<\alpha}$ of x, for a unique ordinal α, such that

$$y_\beta << y_\gamma \Leftrightarrow \beta < \gamma$$

for all $\beta,\gamma < \alpha$.

Proof: The proof will be largely left as an exercise I'll say this much: for the existence of the required indexing, use Prop. 6.1′ with f defined by: $f(A) =_{def}$ the unique $<<$-least element of A ($A \subset X$, $A \neq \emptyset$). One has to verify that for $y_\beta =_{def} (F\Gamma\alpha)(\beta)$ the required biconditional "$\Leftrightarrow$" holds (here, $F\Gamma\alpha$ is the indexing that is obtained by 6.1′)

Another approach to ordinals is to define them as types of well-orderings.* Given a relation R on a set A, the pair (A,R) is called a 'relational structure'. For two relational structures (A,R), (B,S), an isomorphism between them, in notation

$$h : (A,R) \xrightarrow{\sim} (B,S),$$

is a one-to-one and onto function $h : A \longrightarrow B$ such that $a_1 R a_2$ iff $h(a_1) S h(a_2)$ for all $a_1, a_2 \in A$. If there exists an isomorphism between (A,R) and (B,S), we call the two structures isomorphic, and we write $(A,R) \simeq (B,S)$. If R and S are

* What we called 'ordinals' are sometimes called the von-Neumann ordinals, after John von Neumann who introduced them, around 1928. [illegible] is the one we are describing here.

linear orderings of A and B, respectively, (in which case (A,R) is called a linearly ordered set, or a linear ordering) then instead of 'isomorphic' it is customary to say 'similar'.

The relation of isomorphism is an equivalence on the class of all relational structures (exercise). If we consider the equivalence classes with respect to this relation $\simeq$, then we have

$$(A,R) \simeq (B,S) \Leftrightarrow (A,R)/\simeq = (B,S)/\simeq.$$

An equivalence class $(A,R)/\simeq$ is considered the type of the structure (A,R); the last fact says that two structures are isomorphic iff they have the same type. Now, ordinals could be construed as types of well-ordered sets. We now proceed to explain this.

From now on, R and S are always irreflexive linear orderings of the sets A and B, respectively. An embedding $h : (A,R) \longrightarrow (B,S)$ is a one-to-one map $h : A \longrightarrow B$ such that aRa' implies h(a)Sh(a') for all a, a' in A (notice that then, conversely, h(a)Sh(a') implies aRa' (exercise)). An initial segment of (A,R) is a subset X of A such that $a \in X$ and bRa implies that $b \in X$; X is a proper initial segment if, in addition, $X \neq A$. For any a, $R_a = \{b \in A : bRa\}$ is an initial segment (exercise). If (A,R) is a well-ordering, then the initial segments are the sets of the form R_a, and A itself; no more* (exercise). A comparison map $h : (A,R) \longrightarrow (B,S)$ is an embedding such that range(h) is an initial segment of (B,S). Using a rather temporary notation, we write

$$(A,R) \lesssim (B,S)$$

* In particular, if (A,R) is an ordinal, $(A,R) = (\alpha, \in \restriction \alpha)$, then the proper initial segments of (A,R) are exactly the ordinals $\beta < \alpha$

if there exists a comparison map from (A,R) to (B,S). It is easy to check that $\lesssim$ is a quasi-ordering on the class of linear orderings (exercise).

From now on, $(A,<)$, $(B,<')$ denote well-orderings.

Proposition 7.2 (i) If $h : (A,<) \longrightarrow (A,<)$ is an embedding of a well-ordering into itself, then for all $a \in A$, $a \leq f(a)$.

(ii) For well-orderings $(A,<)$, $(B,<')$, $(A,<) \lesssim (B,<')$ and $(B,<') \lesssim (A,<)$ imply that $(A,<) \simeq (B,<')$.

(iii) For any two well-orderings $(A,<)$ and $(B,<')$, there is at most one comparison map from $(A,<)$ to $(B,<')$.

Proof: (i) We employ induction with respect to the wf relation $<$. Suppose that $b \leq f(b)$ holds for all $b < a$. Since $b < a$ implies $f(b) < f(a)$, it follows that for all $b < a$, we have $b < f(a)$. So, if we had $f(a) < a$, then, with $b = f(a)$, we would get $f(a) < f(a)$, which is absurd. Since $f(a) \nless a$, we must have $a \leq f(a)$ as required.

(ii) Suppose that $h : (A,<) \longrightarrow (B,<')$, $h' : (B,<') \longrightarrow (A,<)$ are comparison maps. Consider the composite function, $h'' = h' \circ h : A \longrightarrow A$. It is easy to check that it must be a comparison map of $(A,<)$ to itself (exercise). If its image were a proper initial segment of $(A,<)$, it would be of the form $<_a$ for some $a \in A$. But then $h''(a) \in <_a$, i.e., $h''(a) < a$, in contradiction to part (i). Hence, range$(h'') = A$. But then $h' : B \longrightarrow A$ is an onto map as well: if $a \in A$, then there is $a' \in A$ such that $a = h''(a') = h'(h(a'))$, i.e., $a = h'(b)$ for $b = h(a')$.

(iii) We assume that h_1, h_2 are comparison maps from $(A,<)$ to $(B,<')$, and we prove by induction on $a \in A$ (along the well-ordering $<$) that $h_1(a) = h_2(a)$. So, suppose we know that $h_1(a') = h_2(a')$ for all a' such that $a' < a$. I claim that $h(a)$ is the least element b of $B - \text{range}(h \restriction <_a)$ (now, h is either h_1 or h_2). Indeed, since h is one-to-one, $h(a)$ certainly belongs to this set; hence $b \leq' h(a)$. Since h is a comparison map, it follows that $b = h(a')$ for some a', and necessarily $a' \leq a$. But $a' < a$ is impossible since then $b = h(a') \in \text{range}(h \restriction <_a)$. It follows that $a' = a$, i.e. $b = h(a)$, as claimed.

Now, applying this fact to both h_1 and h_2, we get $h_1(a) = $ minimal element of $B - \text{range}(h_1 \restriction <_a) = $ minimal element of $B - \text{range}(h_2 \restriction <_a)$ (since $h_1 \restriction <_a = h_2 \restriction <_a$) $= h_2(a)$, as required. □

If we write, again temporarily, $(A,<) \lesssim (B,<')$ for "there is a comparison map from $(A,<)$ to $(B,<')$ onto a proper initial segment of $(B,<')$", then 7.2(iii) implies that

$$(A,<) \lesssim (B,<') \Leftrightarrow (A,<) \leq (B,<') \text{ and } (A,<) \not\cong (B,<')$$

(verify).

Note that the uniqueness assertions in 7.1 follow from 7.2 (exercise).

Proposition 7.2′ Every well-ordered set is similar to an ordinal (that is, an ordinal with the natural well-ordering of its elements). In symbols, for any well-ordered set (A, R) (A is a set now!), there is a unique ordinal α, and a unique similarity (isomorphism)

$$(\alpha, \in \restriction \alpha) \xrightarrow{\simeq} (A, R).$$

Proof. This is just a restatement of Prop. 7.1 (p. 72) (exercise).

Proposition 7.2″ (Comparability of well-orderings). For any well-ordered sets $(A, <)$ and $(B, <')$, either $(A, <) \lesssim (B, <')$ or $(B, <') \lesssim (A, <)$.

Proof. When $(A, <)$, $(B, <')$ are ordinals, then the assertion follows from the "trichotomy of $\in$ on Ord", Prop. 4.9 (p. 46): in fact, the required comparison map will be an inclusion: $\alpha \longrightarrow \beta$ mapping $\gamma (< \alpha)$ to γ, or vice versa. The general case of 7.2″ now follows by a use of 7.2′ (exercise).

The last result does not mention ordinals, and in fact, it is not hard (after having seen the techniques so far) to prove it directly.

The relation $\lesssim$ among well-orderings is a quasi-ordering with the additional comparability property (5) ('linear ordering'). The partial ordering derived from it is defined over the equivalence classes of the relation

$$(A,<)E(B,<') \Leftrightarrow (A,<) \lesssim (B,<') \text{ and } (B,<') \lesssim (A,<).$$

(Strictly speaking, this is legitimate in our theory only if we have restricted the class of well-orderings to a fixed but otherwise arbitrary set. We suppress the explicit mention of this set, but strictly speaking, it is there.) By 6.3(ii), E is nothing but similarity of well-orderings. Traditionally, the equivalence classes of the similarity relation are the ordinals. The trouble is that ordinals so construed are proper classes (at least, without the 'restriction' mentioned above); let us call them Ordinals in recognition of this fact. The important connection between ordinals and Ordinals is that every Ordinal contains exactly one ordinal and every ordinal is contained in exactly one Ordinal. The fact that an Ordinal contains only one ordinal is the same as to say that two distinct ordinals (with their natural well-orderings) cannot be similar: indeed, if α and β are distinct

then either $\alpha \in \beta$, or $\beta \in \alpha$; in the first case, the identity map on α is a comparison map from α to β, with range $\alpha = <_\alpha$, where $<$ is the well-ordering of β, hence it is a comparison map onto a proper subset of β ($\alpha \notin \alpha$); thus, by (iii), α and β cannot be similar; the second case is symmetric.

The unique ordinal that (with its natural ordering) is similar to the well-ordered set $(A,<)$* is called the order-type of $(A,<)$, and it is denoted by $|(A,<)|$. Thus, we have

$$|(A,<)| \leq |(B,<')| \Leftrightarrow (A,<) \simeq (B,<')$$
$$|(A,<)| < |(B,<')| \Leftrightarrow (A,<) \tilde{\not\simeq} (B,<')$$

We also have

Proposition 7.2'''

$(A,<) \lesssim (B,<') \Leftrightarrow$ there is an embedding of $(A,<)$ into $(B,<')$.

Proof: The implication from left to right is obvious. Suppose, conversely, that h is an embedding of $(A,<)$ into $(B,<')$. On the other hand, we have either $(A,<) \lesssim (B,<')$, or $(B,<') \lesssim (A,<)$ by (5). If the first case holds, we are done. In the second case, we have $h' : B \longrightarrow A$, an embedding of $(B,<')$ into $(A,<)$. If h' is onto, then h' is a similarity map, hence $(A,<) \sim (B,<')$, and we are again done. If, however, range(h') is a proper initial segment of $(A,<)$, then $h' \circ h$ is a comparison map of $(A,<)$ into itself with image a proper initial segment. Hence $h' \circ h \neq Id_A$, a contradiction to 7.2. □

---

* see Proposition 7.2' on p. 75.1

In the language of well-orderings, 6.1 becomes

Theorem 7.3 (The well-ordering theorem). Every set can be well-ordered; for every set X there is at least one well-ordering of X.

Proof: Let $\langle y_\beta \rangle_{\beta<\alpha}$ be an α-indexing of X, for some ordinal α (see 6.1). Then the relation $<<$ on X defined by

$$y_\alpha << y_\beta \Leftrightarrow \alpha < \beta$$

is a well-ordering of X. □

Let $(X,<)$ be a well-ordered set. A subset Y of X is cofinal in $(X,<)$ if for all $x \in X$ there is $y \in Y$ such that $x \leq y$. Let us denote the order type of the initial segment $<_y$ with the ordering $<$ restricted to $<_y$ by $|y|$.

Proposition 7.4 Y is cofinal in $(X,<)$ if and only if $|(X,<)| = \ell.s.u.b.\{|y| : y \in Y\}$. *

Proof: exercise.

---

* For the 'least strict upper bound', l.s.u.b., see p. 49.

8. Zorn's lemma

Zorn's lemma is a principle widely used in mathematics. It is closely related to the principle of transfinite recursion. The advantage of Zorn's lemma is its abstract simplicity; it does not mention well-orderings or ordinals. Sometimes, however, the use of Zorn's lemma in place of transfinite recursion is forced and counter-intuitive. Nevertheless, in many cases it provides elegant and short proofs.

Zorn's lemma deals with a partial ordering on a set X. Let us denote the partial ordering by $\leq$. A chain in $(X,\leq)$ is a subset C of X such that any two elements of C are comparable: $a \leq b$ or $b \leq a$ for any a, b in C. An upper bound of a subset C of X is an element c in X such that $a \leq c$ for all $a \in C$. A maximal element of X is any $a \in X$ such that $a < b$ (meaning $a \leq b$ and $a \neq b$) is impossible. Thus, a maximal element is either greater or equal, or incomparable, to any other element. Now, we have

Theorem 8.1 (Zorn's lemma). Let $\leq$ be a partial ordering of the set X. Suppose that every chain in $(X,\leq)$ has an upper bound in X. Then X has at least one maximal element.

Proof: We start by picking an indexing $\langle x_\beta \rangle_{\beta<\alpha}$ of X. The proof consists in constructing a 'maximal' chain, by systematically going through the x_β, and throwing in as many of the elements into the chain as possible without destroying the chain property. An upper bound of this 'maximal' chain will then be our

maximal element.

We decide, by transfinite recursion on β, if x_β should, or should not, be in the chain C to be constructed. Formally, we define a function ("characteristic function")

$$f : \alpha \longrightarrow \{0,1\};$$

we'll define $C = \{x_\beta : \beta < \alpha \text{ and } f(\beta) = 1\}$. The definition of f is as follows. Suppose $\beta < \alpha$ and $f(\gamma)$ has been defined for all $\gamma < \beta$. Then, by definition, $f(\beta) = 1$ if and only if x_β is comparable to every x_γ such that $\gamma < \beta$ and $f(\gamma) = 1$: $x_\beta \leq x_\gamma$ or $x_\gamma \leq x_\beta$ for all $\gamma < \beta$ such that $f(\gamma) = 1$. (It is left to the reader to cast this definition, if he desires, into the form of the 'official' recursion principle). Now, let us define C from f as indicated. C is a chain, since if $\beta, \gamma < \alpha$ and $f(\beta) = f(\gamma) = 1$, and, say, $\gamma < \beta$, then by the definition of $f(\beta)$, we have that x_β must be comparable to x_γ. I claim that C is a maximal chain in the sense that no element of X not already in C can be added to C so that C remains a chain. Indeed, let x be an arbitrary element of X; we have $x = x_\beta$ for some $\beta < \alpha$; if $C \cup \{x_\beta\}$ is a chain, then in particular, x_β is comparable to every x_γ with $f(\gamma) = 1$, hence, by the definition of $f(\beta)$, $f(\beta) = 1$ and $x_\beta \in C$; i.e., x_β was in C to begin with.

Let x be an upper bound of C. If $y \in X$ such that $x \leq y$, then $C \cup \{y\}$ is a chain, since for every $u \in C$, $u \leq x \leq y$, hence, in particular, y is comparable to all elements of C. By what we said above, this implies that $y \in C$. But then, since x is

an upper bound of C, we have $y \leq x$; thus $x = y$. We have shown that x is such that for any y, $x \leq y$ implies $x = y$. This means that x is maximal. □

A common situation for Zorn's lemma is when X is a family of certain subsets of a given set A, and the partial ordering is ordinary containment ($\subset$). A chain is a subfamily of X in which any two sets are comparable with respect to containment. The 'upper bound' condition is usually satisfied because for a chain $C \subset X$, $\cup C$ turns out to be a member of X as well; of course, then $\cup C$ is an upper bound of C in X.

We will see applications of Zorn's lemma in the theory of Boolean algebras, among others.

9. Cardinal numbers

We develop here the theory of 'size', or 'cardinality', of sets. The familiar case is that of finite sets, where

Definition 9.1 A set x is finite if there exists an $n \in \omega$ and an n-indexing of x. A set is infinite if it is not finite.

It turns out that if a set has an n-indexing for some $n \in \omega$, then it cannot have an α-indexing for any other ordinal $\alpha \neq n$. This justifies our saying that the set has cardinality n: "any way of counting the set gives the same result, namely n". As we indicated above, the situation in general is not this simple. Still, it is possible to develop a useful theory of cardinality of infinite sets.

We start by some considerations not involving ordinals. Possibly the most natural beginning is

Definition 9.2 Two sets A and B* have the same cardinality (an indivisible phrase!), or are equinumerous if there is a one-to-one and onto function

$$h : A \longrightarrow B.$$

We write $A \sim B$ for 'A and B are equinumerous".

This definition expresses the simple idea that if the elements of two sets can be exactly paired, one from one set, the other from the other set, then the two sets have the same number of elements.

---

* In this section, the letters A, B, possibly with subscripts, always denote sets.

Next, we would like to say that a proper subset of a set has a smaller cardinality. This, however, we cannot reasonably maintain, since, e.g. the proper subset $\omega-\{0\}$ of ω is equinumerous with ω, by the function

$$h : \omega \longrightarrow \omega-\{0\}$$
$$n \longmapsto n+1.$$

Therefore, we relax 'smaller' to 'smaller or equal'. Then, of course, we may drop properness of the subset as a requirement. It is reasonable to consider sets with a one-to-one correspondence with a subset instead of just subsets; so we obtain:

Definition 9.3 A has cardinality less than or equal to that of B, or in short, A is dominated by B, in notation $A \lesssim B$, if there is a one-to-one function $h : A \longrightarrow B$.

Proposition 9.4 The relation $\lesssim$ is a quasi-ordering on $\mathbb{V}$. Equinumerosity is an equivalence on $\mathbb{V}$.

Proof: exercise.

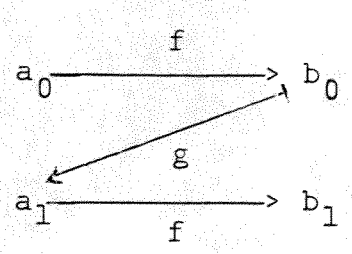
Theorem 9.5 (Cantor-Bernstein theorem)

$A \lesssim B$ and $B \lesssim A$ imply that $A \sim B$.

Proof: Suppose that $f : A \longrightarrow B$, $g : B \longrightarrow A$ are one-to-one functions. Our task is to construct, using f and g, a one-to-one and onto function $h : A \longrightarrow B$. The function h will be built up from f and g in the precise sense that $h \subset f \cup g^{-1}$; i.e., if $h(a) = b$, then either $f(a) = b$, or $g(b) = a$.

Once we know that we are looking for h of this kind, the construction becomes determined to a large extent.

Consider the set $A_0 = A\text{-range}(g)$. For $a \in A_0$, there is no b with $g(b) = a$, hence we must define $h(a) = f(a)$. Furthermore, let $A_1 = g[f[A_0]]$. Note that $A_1 \cap A_0 = \emptyset$. For $a_1 \in A_1$, we have $a_0 \in A_0$ such that for $b_0 = f(a_0)$, we have $a_1 = g(b_0)$:



Note that $a_1 \neq a_0$. Since $h(a_0) = b_0$ by the above, we cannot define $h(a_1) = b_0$ without violating one-to-oneness. So, we are forced again to define $h(a_1) = f(a_1)$. In fact, by recursion on $\mathbb{N}$, we define:

$$A_{n+1} = g[f[A_n]];$$

then we put $A_\omega \underset{df}{=} \bigcup_{n<\omega} A_n$; finally, we define h to coincide with f on A_ω, and g^{-1} on $A\text{-}A_\omega$:

$$h(a) = f(a) \qquad (a \in A_\omega),$$
$$h(a) = g^{-1}(a) \qquad (a \in A\text{-}A_\omega).$$

Let us verify the required properties of h. First of all, since $A_0 \subset A_\omega$, i.e. $A\text{-}A_\omega \subset \text{range}(g)$, and g is one-to-one, $g^{-1}(a)$ is well-defined for $a \in A\text{-}A_\omega$. Clearly, h is one-to-one on the set A_ω, as well as on the set $A\text{-}A_\omega$, separately. To complete the proof of h being one-to-one, assume (for contradiction)

that $a_1 \in A_\omega$, $a_2 \in A\text{-}A_\omega$, and $f(a_1) = g^{-1}(a_2)$. This means that $g(f(a_1)) = a_2$. Also, $a_1 \in A_n$ for some $n < \omega$, hence $a_2 = g(f(a_1))$ belongs to $g[f[A_n]] = A_{n+1}$, in contradiction to $a_2 \in A\text{-}A_\omega$.

It remains to show that h is onto. Let b be an arbitrary element of B. Consider $a = g(b)$. If $a \in A\text{-}A_\omega$, we are done, since then $b = g^{-1}(a) = h(a)$. Otherwise, $a \in A_\omega$, hence $a \in A_n$ for some n. n cannot be 0, since no a in A_0 is of the form $g(b)$. So, $a \in g[f[A_{n-1}]]$, i.e., for some $a' \in A_{n-1}$, we have $a = g(f(a'))$. But also, $a = g(b)$; since g is one-to-one, we must have that $b = f(a')$; since also $a' \in A_\omega$, $b = h(a')$ as required. □

Theorem 9.6 (Comparability theorem).

For any sets A and B, either $A \lesssim B$, or $B \lesssim A$.

Proof: Unlike the previous theorem, this one uses the axiom of choice, through the ordinal indexing theorem. Let α, β be ordinals and $\langle a_\gamma \rangle_{\gamma<\alpha}$, $\langle b_\delta \rangle_{\delta<\beta}$ indexings of A and B, respectively. Now, either $\alpha \leq \beta$, or $\beta \leq \alpha$. E.g., in the first case, we may consider the map

$$a_\gamma \longmapsto b_\gamma \qquad (\gamma < \alpha \leq \beta);$$

clearly, it is a one-to-one map of A into B, hence $A \lesssim B$. The other case similarly yields $B \lesssim A$. □

This theorem does not mention ordinals, but its proof uses them in an essential way. It may be said that ordinals are just the right notion to make the following naive proof of 9.6 precise:

"Start simultaneously counting the elements of A as well as those of B. The set that runs out of elements before or at the same time as the other is the one that is no bigger than the other."

A Cardinal number, or simply Cardinal, is an equivalence class $A/\sim$ under the equivalence 'equinumerosity'. The Cardinals are linearly ordered by the ordering derived from $\lesssim$ (after a restriction of all sets considered to a subset of $\mathbb{V}$); notice that by the Cantor-Bernstein theorem, 'equinumerosity' is the same as the equivalence relation induced by $\lesssim$. Explicitly:

$A/\sim \leq B/\sim \Leftrightarrow$ there is a one-to-one map from A to B;

here we have introduced the notation $\leq$ for the ordering of Cardinals. Of course

$A/\sim < B/\sim \Leftrightarrow A/\sim \leq B/\sim$ and $A/\sim \neq B/\sim$.

We can, in fact, say much more about the ordering of Cardinals. First of all, notice that we have

Proposition 9.7 Every set is equinumerous with an ordinal. □

This fact is contained in the indexing theorem, and appeared already in the proof of 9.6.

In other words, 9.7 says that every Cardinal contains at least one ordinal. The 'anomaly' we noted above is that a Cardinal may contain more than one ordinal; e.g., ω and $\omega+1$ are in the same Cardinal (why?). But, still, we can pick out a best ordinal in every Cardinal: take its smallest ordinal element. Thus, we say

that a cardinal number, or cardinal, is an ordinal α such that α is the smallest ordinal in $\alpha/\sim$. In a less abstract language, we can phrase this as

Definition 9.8 A cardinal is an ordinal which is not equinumerous with any smaller ordinal.

Thus, we have that cardinals and Cardinals are in one-to-one correspondence: every Cardinal contains exactly one cardinal, and every cardinal is an element of exactly one Cardinal. Moreover, we can transfer the ordering of Cardinals to cardinals: for cardinals κ and λ, we define

$$\kappa < \lambda \Leftrightarrow \kappa/\sim < \lambda/\sim.$$

Note that, in fact, for cardinals κ and λ, $\kappa < \lambda$ in the 'cardinal sense' if and only if $\kappa < \lambda$ in the 'ordinal sense'. First of all, if $\kappa < \lambda$ in the ordinal sense and κ, λ are cardinals, then clearly, $\kappa \leq \lambda$ in the cardinal sense (why?); but $\kappa/\sim = \lambda/\sim$ is impossible, since this would contradict λ being a cardinal (why?). Thus, we see that, for cardinals κ and λ (!), $\kappa < \lambda$ in the ordinal sense implies $\kappa < \lambda$ in the cardinal sense. A moment's reflection shows that the converse of this implication is an automatic consequence. Thus, there is no danger of confusion arising from using the notation $<$ for the two different notions. Moreover, since cardinals form a subclass of $\mathbb{O}rd$, and their cardinal-ordering is the same as their ordinal-ordering, we obtain that the ordering of cardinals is a well-ordering. Of course, it follows that the natural ordering of Cardinals is a well-ordering. Note that the

latter statement is one that does not involve the notion of ordinal at all.

Let us use the notation $|A|$, or sometimes card A, for the cardinal of A (A a set); $|A|$ is the unique cardinal equinumerous with A. Clearly, $|A|$ can be defined as the smallest ordinal κ such that A has a κ-indexing (exercise).

Also, clearly,

$$|A| = |B| \Leftrightarrow A \sim B$$

$$|A| < |B| \Leftrightarrow A \lesssim B \quad \text{and} \quad A \not\sim B.$$

$|A| \leq |B| \Leftrightarrow A \lesssim B$ (Cantor-Bernstein ...) (exercise).

Note that for every ordinal α, $|\alpha| \leq \alpha$, and $|\alpha| = \alpha$ if and only if α is a cardinal. It is easily seen that $\alpha < \beta$ implies $|\alpha| \leq |\beta|$. Thus, if κ is a cardinal, and $\kappa < \alpha$, then $\kappa \leq |\alpha|$. As a consequence, if κ is a cardinal and $|\alpha| < \kappa$, then $\alpha < \kappa$.

Proposition 9.9 Every natural number is a cardinal. ω is a cardinal.

Proof: an instructive exercise.

The main result that shows that the theory of infinite cardinals does not collapse to a triviality is

Theorem 9.10 (Cantor's theorem). For every set A,

$$|A| < |P(A)|.$$

Proof: The map $h : A \longrightarrow P(A)$ defined by $h(a) = \{a\}$

shows that $|A| \leq |P(A)|$ (why?). It remains to show that $|A| \neq |P(A)|$. Suppose, for a contradiction, that there is a (one-to-one and) onto function $h : A \longrightarrow P(A)$. Consider the 'paradoxical set':

$$B = \{a \in A : a \notin h(a)\}.$$

Now, since h is onto, $B = h(b)$ for some $b \in A$. But then

$$b \in h(b) \quad \Leftrightarrow \quad b \in B \quad \Leftrightarrow \quad b \notin h(b),$$

$\uparrow$ since $h(b) = B$ $\qquad$ $\uparrow$ by the definition of B

i.e. $$b \in h(b) \quad \Leftrightarrow \quad b \notin h(b)$$

which is absurd. □

Note the similarity of this proof to that of the "Russel paradox": $\mathbb{V}$ is not a set.

The method of proof of Cantor's theorem is called the diagonal method. Here is another illustration of the method.

Consider the set ${}^{\mathbb{N}}B$, the set of all functions $f : \mathbb{N} \longrightarrow B$, with B any set containing at least two distinct elements. Suppose we have a subset F of ${}^{\mathbb{N}}B$ indexed by $\mathbb{N}$: $F = \{f_n : n \in \mathbb{N}\}$. We can construct a function $f : \mathbb{N} \longrightarrow B$ not in F as follows. We specify the value f at any n so that it is distinct from $f_n(n)$;

$$f(n) \neq f_n(n);$$

if $|B| \geq 2$, this is possible. But then $f \notin F$. Namely, if f were in F, it would have to be the same as f_n for some n; but for any n, we have made it explicitly true that f be

different from f_n at a specific argument, namely n itself.

Upon identifying subsets of $\mathbb{N}$ by their characteristic functions, this argument with $B = 2$ is seen to be identical to the one in the proof of 9.10 (for $A = \mathbb{N}$).

As a consequence of Cantor's theorem, we have now at least the following distinct infinite cardinals:

$$\omega$$
$$|P(\omega)|$$
$$|PP(\omega)|$$
$$\vdots$$
$$|\underbrace{P \ldots P}_{n\text{-times}}(\omega)|$$
$$\vdots$$

This sequence is a strictly increasing sequence of cardinals.

Corollary 9.11 The class of cardinals, Card, is not a set. For every set X of cardinals, there is a cardinal strictly greater than any member of X.

Proof: Clearly, the second statement implies the first (why?). To prove the second statement, let $Y = \cup X$. Since for every $\kappa \in X$, we have $\kappa \subset Y$, we have $\kappa = |\kappa| \leq |Y|$. Then consider $\lambda = |P(Y)|$. We have, for $\kappa \in X$, $\kappa \leq |Y| < \lambda$, hence $\kappa < \lambda$. □

Denoting the above sequence of cardinals by $\beth_0^*, \beth_1, \beth_2, \ldots, \beth_n, \ldots$, by replacement and the axiom of infinity, we have that

$$\{\beth_n : n < \omega\}$$

---

* ב : Hebrew BETH

is a set. The last corollary implies that there is a cardinal strictly greater than any $\beth_n$, $n < \omega$. In fact, it is easy to see that

$$\beth_\omega \underset{df}{=} \bigcup_{h<\omega} \beth_n$$

qualifies (exercise).

The 'aleph-sequence' of cardinals is defined to contain all cardinals. By recursion on Ord, define

$\aleph_\alpha$ = the smallest infinite cardinal greater than all $\aleph_\beta$, for $\beta < \alpha$

(verify that this is a legitimate definition!). Thus, $\aleph_0 = \omega$, $\aleph_1$ is the smallest cardinal for which $\aleph_0 < \aleph_1$. Also, we have $\aleph_0 < \beth_1$; hence $\aleph_1 \leq \beth_1$. The question whether $\aleph_1 = \beth_1$ is the famous Continuum Problem; it is not decidable either way using the principles of set theory we have listed so far.

Note the following obvious relations:

$$\aleph_\alpha < \aleph_\beta \Leftrightarrow \alpha < \beta$$
$$\alpha \leq \aleph_\alpha$$

(exercise).

It is important to note that every infinite cardinal is of the form $\aleph_\alpha$, for a uniquely determined $\alpha \in$ Ord. In fact, let κ be an infinite cardinal, and let $X = \{\beta \in \text{Ord} : \aleph_\beta < \kappa\}$. X is a set simply because $X \subset \kappa$ (why?). Thus, there is an ordinal not in X; let α be the smallest such. Since $\aleph_\alpha$ is the smallest cardinal greater than all the $\aleph_\beta$ for $\beta \in X$ (why?),

we have that $\aleph_\alpha \leq \kappa$. But $\aleph_\alpha < \kappa$ is impossible, since then $\alpha \in X$ would follow. Hence $\aleph_\alpha = \kappa$, as desired. The uniqueness of α is clear (why?).

The following proposition involves a simple, but essential, use of the axiom of choice.

Proposition 9.12 Let A, B be sets, and assume that $A \neq \emptyset$. Then $|A| \leq |B|$ if and only if there is an onto function $B \longrightarrow A$.

Proof: exercise.

A set is called countable if its cardinality is $\leq \aleph_0$, i.e. it is either finite, or equinumerous with $\mathbb{N}$.

Corollary 9.13 A set A is countable if and only if it is either empty or it has an enumeration by the natural numbers: an onto function

$$\begin{aligned} \mathbb{N} &\longrightarrow A \\ n &\longmapsto a_n \end{aligned}$$

(Thus, A is given as the range of a sequence

$$a_0, a_1, \ldots, a_n, \ldots)$$

Proof: exercise.

10. Cardinal arithmetic

We can define addition, multiplication, and exponentiation of cardinals, directly generalizing those operations on the natural numbers. Also, having gotten rid of the limitation of finiteness, we can define infinite sums and products as well. The basic definitions have no use of well-orderings, and in fact, they would be more naturally stated for Cardinals than cardinals. Of course, because of their one-to-one correspondence, whatever operation we define for Cardinals, it is automatically transferred to cardinals. For this reason, we will not explicitly mention Cardinals at all.

The idea of addition is contained in the observation that if a set C is the union of two disjoint subsets A and B, then the cardinality of C should be the sum of those of A and B. This statement can be turned into a definition of addition as follows. Let κ and λ be two cardinals, let A and B be arbitrary sets such that $|A| = \kappa$, $|B| = \lambda$ (in fact, $A = \kappa$ and $B = \lambda$ qualify). Then consider $A' = \{0\}\times A = \{\langle 0,a\rangle : a \in A\}$, $B' = \{1\}\times B = \{\langle 1,b\rangle : b \in B\}$. Of course, we still have $|A'| = \kappa$ and $|B'| = \lambda$ (why?), but we now also have that $A'\cap B' = \emptyset$ (why?). Consider $C = A'\cup B'$. We define

$$\kappa + \lambda = |A'\cup B'|.$$

To have this as a legitimate definition, we have to verify that if A'', B'' are other sets such that $|A''| = \kappa$, $|B''| = \lambda$ and $A''\cap B'' = \emptyset$, we obtain, by using A'' and B'' instead of A' and B', the same result for $|A''\cup B''|$; in other words,

$$A' \cup B' \sim A'' \cup B''.$$

This relation is almost obvious: by assumption, we have one-to-one, onto maps $f : A' \longrightarrow A''$, $g : B' \longrightarrow B''$; because of the disjointness assumptions, $f \cup g$ is a one-to-one, onto function from $A' \cup B'$ to $A'' \cup B''$.

Thus, we have defined $\kappa + \lambda$ unambiguously as the cardinal of any disjoint union of sets the first of which is of cardinality κ, the second λ.

The definition can be expressed concisely by the equality

$$|A \dot{\cup} B| = |A| + |B|$$

where the notation $A \dot{\cup} B$ refers to a union $A \cup B$ where $A \cap B = \emptyset$. *

Certain identities of cardinal addition are immediate from the definition:

$$\kappa + \lambda = \lambda + \kappa$$

$$(\kappa + \lambda) + \beta = \kappa + (\lambda + \beta)$$

$$\kappa + 0 = \kappa$$

(exercise).

It is just as easy and natural to define the sum of an arbitrary set of cardinals. Let $\langle \kappa_i \rangle_{i \in I}$ be an indexed family of cardinals (i.e., we have a function $I \longrightarrow \mathrm{Card}$, $i \longmapsto \kappa_i$ with I a set). We may define

$$\sum_{i \in I} \kappa_i \underset{df}{=} \left| \bigcup_{i \in I} (\{i\} \times \kappa_i) \right|$$

which is but a short statement of the kind of definition we gave above for the special case $I = \{0,1\}$. The commutative and associative laws are jointly generalized by the following law.

---

* $A \dot{\cup} B$ is called the disjoint union of A and B; it is defined only if $A \cap B = \emptyset$.

Assume that $I = \dot{\bigcup}_{j \in J} I_j$; the dot means that the union is disjoint, i.e. it is assumed that $I_j \cap I_{j'} = \emptyset$ for $j \neq j'$. Then we have

$$\sum_{i \in \dot{\bigcup}_{j \in J} I_j} \kappa_i = \sum_{j \in J} \left(\sum_{i \in I_j} \kappa_i \right)$$

(exercise).

Turning to multiplication, we define

$$\kappa \cdot \lambda = |\kappa \times \lambda|.$$

It is important to note that then we have

$$|A \times B| = |A| \cdot |B|$$

for any sets A and B (exercise). Multiplication is repeated addition in the sense that

$$\kappa \cdot \lambda = \sum_{i \in \lambda} \kappa$$

(the right-hand-side is the sum of the family $\langle \kappa_i \rangle_{i \in \lambda}$ where each $\kappa_i = \kappa$) (exercise).

We have the following laws:

$$\kappa \cdot \lambda = \lambda \cdot \kappa$$

$$(\kappa \cdot \lambda) \cdot \mu = \kappa \cdot (\lambda \cdot \mu)$$

$$(\kappa + \lambda)\mu = \kappa\mu + \lambda\mu$$

$$\mu \cdot 0 = 0$$

$$\mu \cdot 1 = \mu$$

(exercise).

The infinite version of multiplication is related to 'infinite' Cartesian product. We define

$$\prod_{i\in I} \kappa_i \underset{df}{=} \left| \bigtimes_{i\in I} \kappa_i \right| \quad *$$

We have the more general equality

$$\left| \bigtimes_{i\in I} A_i \right| = \prod_{i\in I} |A_i|$$

as a consequence (exercise).

We have the combined commutativity/associativity law

$$\prod_{i\in \dot{\bigcup}_{j\in J} I_j} \kappa_i = \prod_{j\in J} \left(\prod_{i\in I_j} \kappa_i \right)$$

(exercise). The distributive law, in its full generality, says

$$\prod_{j\in J} \left(\sum_{i\in I_j} \kappa_i \right) = \sum_{f \in \bigtimes_{j\in J} I_j} \prod_{j\in J} \kappa_{f(j)}$$

(exercise). **

Exponentiation is introduced by the definition

$$\kappa^\lambda \underset{df}{=} |{}^\lambda \kappa|$$

In other words, we have

$$|{}^B A| = |A|^{|B|}$$

Some laws:

$$(\kappa\lambda)^\mu = \kappa^\mu \lambda^\mu$$

$$\kappa^\lambda \cdot \kappa^\mu = \kappa^{\lambda+\mu}$$

(exercise).

In general forms:

$$\left(\prod_{i\in I} \kappa_i \right)^\mu = \prod_{i\in I} \kappa_i^\mu$$

$$\prod_{i\in I} \kappa^{\lambda_i} = \kappa^{\sum_{i\in I} \lambda_i}$$

(exercise).

---

* For the definition of $\bigtimes_{i\in I} \kappa_i$, see p. 38, last line.

** see the next page, 96.1.

**

Recall Problem 3 (ii) of Assignment 1. [What is written there as $\prod_{j \in J} I_j$ is now written $X_{j \in J} I_j$.

The reason is that we should be able to make a distinction between $\prod_{j \in J} \kappa_j$, which is a cardinal, and the set $X_{j \in J} \kappa_j$, which is, in general, not a cardinal, but a set of functions satisfying certain properties.] The equality of cardinal arithmetic stated in the text now (on p. 96) is a direct consequence of the stated exercise in Assmnt 1. (When checking this, you should be careful about 'disjointness' of unions)

Since we have that $P(A) \sim 2^A$ (see Section 3), we have that $|P(A)| = 2^{|A|}$. With the notation above, we also see that $\beth_0 = \aleph_0$, $\beth_1 = 2^{\aleph_0}$, and in general, $\beth_{n+1} = 2^{\beth_n}$. Furthermore, Cantor's theorem says that $\kappa < 2^\kappa$.

There are laws relating the operations to the ordering of cardinals. E.g.,

$\kappa \leq \kappa'$ and $\lambda \leq \lambda'$ imply each of $\kappa+\lambda \leq \kappa'+\lambda'$,

$\kappa \cdot \lambda \leq \kappa' \cdot \lambda'$, and $\kappa^\lambda \leq (\kappa')^{\lambda'}$.

In fact, $\kappa_i \leq \kappa_i'$ for all $i \in I$ implies that

$$\sum_{i \in I} \kappa_i \leq \sum_{i \in I} \kappa_i' \quad \text{and} \quad \prod_{i \in I} \kappa_i \leq \prod_{i \in I} \kappa_i'$$

(exercise).

The following simple proposition states a characteristic property of infinite cardinals.

Proposition 10.1 For any infinite ordinal α, $|S\alpha| = |\alpha|$. As a consequence, every infinite cardinal is a limit ordinal.

Proof: Since the ordinals $< \omega$ are precisely the finite ordinals, α being infinite means that $\omega \leq \alpha$, i.e., $\omega \subset \kappa$. Define the function

$$h : S\alpha = \{\alpha\} \cup \alpha \longrightarrow \alpha$$

by

$$\alpha \longmapsto 0 ,$$

$$n \longmapsto n+1 \qquad (n \in \omega)$$

$$\beta \longmapsto \beta \qquad (\beta \in \alpha - \omega)$$

h is one-to-one, and onto; hence $S\alpha \sim \alpha$.

Suppose α is an infinite successor ordinal, $\alpha = S\beta$. Then β is infinite too; otherwise, α would be finite. Hence, by the first part, $|\alpha| = |S\beta| = |\beta|$. We have $\beta < \alpha$, therefore α cannot be a cardinal: it is equinumerous to a strictly smaller ordinal, namely β. □

Cardinal arithmetic, at least its 'finitary' part, is greatly simplified by certain 'absorbtion' laws, the first example of which is the last result. The main such result is

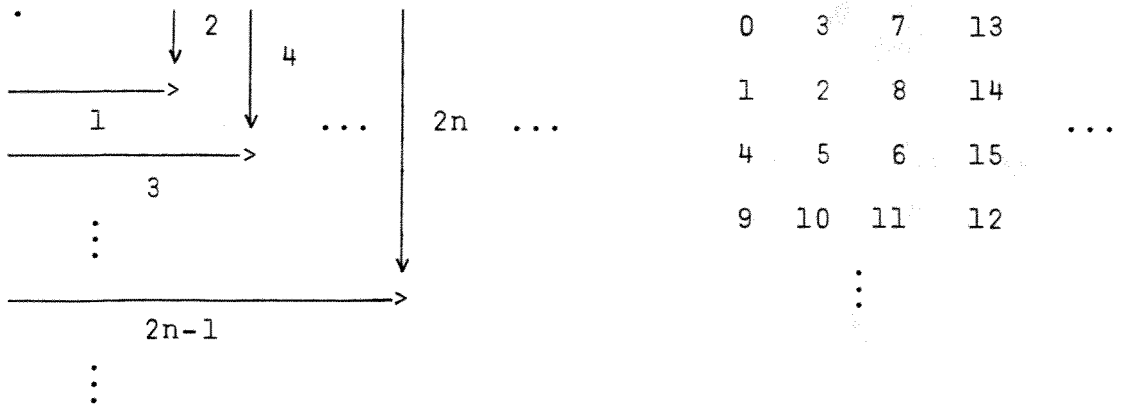
Theorem 10.2 For any infinite cardinal κ, we have that $\kappa \cdot \kappa = \kappa$.

Proof: The assertion, of course, is equivalent to saying that there is a κ-indexing of the set $\kappa \times \kappa$. It is helpful to look at the simplest case $\kappa = \aleph_0$ first.

The set $N \times N$ can be displayed in an "infinite matrix" as follows:

$$\begin{array}{cccccc}
(0,0) & (0,1) & (0,2) & \dots & (0,n) & \dots \\
(1,0) & (1,1) & (1,2) & \dots & (1,n) & \dots \\
(2,0) & (2,1) & (2,2) & \dots & (2,n) & \dots \\
\vdots & & & & & \\
(m,0) & (m,1) & (m,2) & \dots & (m,n) & \dots \\
\vdots & \vdots & \vdots & & \vdots &
\end{array}$$

We can index the entries of this matrix by ω in many ways, e.g., according to the following pattern



This indexing shows that $\aleph_0 \cdot \aleph_0 = \aleph_0$.

The proof of the general case is similar. Instead of an indexing, we define a well-ordering $\lhd$ on $\kappa\times\kappa$, and in fact, on $\mathrm{Ord}\times\mathrm{Ord}$. The definition (imitating the ordering obtained from the indexing in the above picture) of $\lhd$ is as follows:

$(\beta,\gamma) \lhd (\beta',\gamma')$ iff either $\max(\beta,\gamma) < \max(\beta',\gamma')$ or $(\gamma \le \beta = \beta' \ge \gamma'$ and $\gamma < \gamma')$ or $(\beta < \gamma = \gamma' > \beta'$ and $\beta < \beta')$ or $(\gamma \le \beta = \gamma' > \beta')$

[$\max(\beta,\gamma)$ denotes the greater of the two ordinals β, γ or their common value if they are equal.]

We leave it to the reader to verify that $\lhd$ is an (irreflexive) linear ordering (exercise). Let us verify that $\lhd$ is well-founded. Let $X\subset\mathrm{Ord}$ be non-empty; consider the smallest ordinal α for which there is $(\beta,\gamma) \in X$ with $\max(\beta,\gamma) = \alpha$; and let $X_1 = \{(\beta,\gamma) \in X : \max(\beta,\gamma) = \alpha\}$. Let $X_2 = \{(\beta,\gamma) \in X_1 : \gamma \le \beta\ (= \alpha)\}$, $X_3 = \{(\beta,\gamma) \in X_1 : \beta < \gamma\ (= \alpha)\}$. We have $X_2 \cup X_3 = X_1 \neq \emptyset$, and for any $x \in X_2$, $y \in X_3$, $z \in X - (X_2 \cup X_3)$ we have $x \lhd y \lhd z$. Thus, if X_2 is non-empty,

and γ is the minimal ordinal for which $(\alpha,\gamma) \in X_1$, then (α,γ) is the $\lhd$-minimal element of X; if X_2 is empty, then X_3 is not, and for the minimal β for which $(\beta,\gamma) \in X_1$ we have that (β,γ) is the $\lhd$-minimal element of X.

Another easily verified fact is that, for each $\alpha \in \mathrm{Ord}$, $\alpha\times\alpha$ is an initial segment of $(\mathrm{Ord}\times\mathrm{Ord}, \lhd)$; actually, $\alpha\times\alpha = \lhd_{(\alpha,0)}$. Let λ be a limit ordinal, and consider $\lambda\times\lambda$ with $\lhd$ restricted to it. Then the set $\{(\alpha,0) : \alpha < \lambda\}$ is cofinal in $\lambda\times\lambda$: if $(\beta,\gamma) \in \lambda\times\lambda$, let $\alpha = S(\max(\beta,\gamma))$; we have $(\beta,\gamma) \lhd (\alpha,0) \in \lambda\times\lambda$. Let $\alpha^* = |(\alpha,0)|$ = the order type of $\lhd_{(\alpha,0)}$, ordered by $\lhd$, in the notation of 7.4. Then, by 7.4, we have that

$$\lambda^* = \ell.s.u.\{\alpha^* : \alpha < \lambda\} \qquad (1)$$

for any limit ordinal λ.

Note that

$$|\alpha^*| = |\alpha\times\alpha| = |\alpha|\cdot|\alpha|$$

hence for κ a cardinal,

$$|\kappa^*| = \kappa\cdot\kappa.$$

The well-ordering of $\kappa\times\kappa$ so defined represents a 'counting' of $\kappa\times\kappa$ by ordinals $< \kappa^*$. We prove that, in fact, $\kappa^* = \kappa$ for all infinite cardinals κ. This will, of course, establish that $|\kappa\times\kappa| = \kappa$. Note that, since $|\kappa\times\kappa| \geq \kappa$ (why?), $\kappa^* < \kappa$ is impossible; we have to see that $\kappa^* \leq \kappa$.

To prove the claimed equality, we employ induction on cardinals; we take an infinite cardinal κ and we assume that $\lambda^* = \lambda$ is true for all infinite cardinals $\lambda < \kappa$. Consider an arbitrary ordinal $\alpha < \kappa$. If α is finite, then α^* is finite

as well, in particular $\alpha^* < \kappa$; this is easily seen (practically obvious), and in fact, this is precisely the content of our proof for the special case $\kappa = \aleph_0$ described at the beginning. If α is infinite, then $|\alpha|$ is infinite as well; of course $(\lambda =)\,|\alpha| \leq \alpha < \kappa$. We have $|\alpha^*| = |\alpha| \cdot |\alpha|$; hence $|\alpha^*| = |\alpha|$ by the induction hypothesis.(*) It follows that $|\alpha^*| < \kappa$. Thus, since κ is a cardinal, we must have $\alpha^* < \kappa$. We have shown that $\alpha < \kappa$ implies $\alpha^* < \kappa$ (in words: completing the described counting of $\kappa \times \kappa$ up to but not including $\alpha \times \alpha$, we come up with a 'count' α^* strictly less than κ). Now, employ the fact that κ is a limit ordinal (10.1), and the equality (1). We get $\kappa^* = \ell.s.u.b.\{\alpha^* : \alpha < \kappa\} \leq \kappa$, since each $\alpha^* < \kappa$. This completes the proof. □

Corollary 10.3 If at least one of the cardinals κ, λ is infinite, and both non-zero, we have

$$\kappa \cdot \lambda = \kappa + \lambda = \max(\kappa, \lambda).$$

Proof: We have, with $\mu = \max(\kappa, \lambda)$, that

$$\mu \leq \kappa + \lambda \leq \mu + \mu = \mu \cdot 2 \leq \mu \cdot \mu = \mu, \quad \text{i.e.,} \quad \kappa + \lambda = \mu.$$

Also, $\mu = \mu \cdot 1 \leq \kappa \cdot \lambda \leq \mu \cdot \mu = \mu$, i.e., $\kappa \cdot \lambda = \mu$. □

Corollary 10.4 The union of countably many countable sets is countable. If $|I| \leq \aleph_0$, and $|A_i| \leq \aleph_0$ for each $i \in I$, then $|\bigcup_{i \in I} A_i| \leq \aleph_0$.

Proof: exercise.

(*) Let $\lambda = |\alpha|$. Since $\alpha \sim \lambda$, we have $\alpha \times \alpha \sim \lambda \times \lambda$. Also, $\alpha^* \sim \alpha \times \alpha$, and $\lambda^* \sim \lambda \times \lambda$. Therefore, $\alpha^* \sim \lambda^*$. By the induction hypothesis, $\lambda^* = \lambda$. Therefore, $\alpha^* \sim \lambda$, i.e., $|\alpha^*| = |\lambda| = \lambda = |\alpha|$.