Logic for Computer Science

September 16, 2005

Chapter 1

Appendix

This chapter defines and clarifies the general notions used in the preceding chapters. It presents sets, relations, functions, cardinalities, strings, graphs, and trees. We present the topics in a self serving fashion, describing only the definitions and the properties used in the book. For the sake of completion, we include proofs for all propositions, though we do not dwell on them. The reader who is familiar with the concepts or is not particularly interested in proofs, should concentrate on the definitions and on the examples, since they may differ from the ones that he/she is familiar with.

1.1 Sets

Axiomatic set theory is the foundations of mathematics. It serves as a common language for its many branches and as an investigating tool. We would have a hard time describing the semantics of logics without set theory. In this chapter we present the Zermelo-Fraenkel + Choice axiomatization, abbreviated as ZFC. The reader who wants a more detailed introduction can read either the book writen by Hrbacek and Jech or the one authored by Halmos.

We are not going to give a formal definition of a set; instead we will provide an intuitive definition. A *set* is a collection of distinct objects; these objects are called the *members*, or the *elements* of the set. We write $x \in A$ to show that x is an element of A; $y \notin B$ means that y is not a member of B. If all elements of the set A are also in the set B we say that A *is included in* B, and write $A \subseteq B$. For example, set of my ancestors is included in the set of the ancestors of my son.

Axiomatic set theory does not deal with concrete sets like the collection of my ancestors or the books in the university library; it deals only with the sets whose members are also sets. Here, all elements of a set share a common *property*. Properties are statements about sets that have a well defined syntax. The *basic* properties are either memberships $x \in A$, or equalities $X = Y$. In these statements the symbols x, A, X, Y represent sets. They are called *variables*

because they can take a whole range of values. We will represent properties by bold letters. At times we will write $\mathbf{P}[x, A]$ to show that $\mathbf{P}$ is a property that relates the sets x and A. The can now formalize the notion of property.

Definition 1.1.1 *1. The basic properties $\in$ and $=$ are properties.*
2. If $\mathbf{P}$ *is a property then not* $\mathbf{P}$ *is also a property.*
3. If $\mathbf{P}$ *and* $\mathbf{Q}$ *are properties, then*
$\mathbf{P}$ *or* $\mathbf{Q}$, $\mathbf{P}$ *and* $\mathbf{Q}$, *if* $\mathbf{P}$ *then* $\mathbf{Q}$, $\mathbf{P}$ *if and only if* $\mathbf{Q}$
are also properties.
4. If $\mathbf{P}$ *is a property, then*
for all x $\mathbf{P}$, *there exists some* x *such that* $\mathbf{P}$
are also properties.

We will use parentheses or commas, to clear up the ambiguity arising from specifications like *for all* x $\mathbf{P}(x)$ *or* $\mathbf{Q}(x)$ *and* $\mathbf{R}(x)$.

Example 1.1.2 The statements 1-6 below are properties.
1. $x \in A$
2. not $x \in A$
3. $A = B$
4. $x \in A$ or $x \in B$
5. for all x, $x \in U$
6. there is some x such that $x \in y$

We will write $x \notin A$ instead of *not* $x \in A$, and $A \neq B$, instead of *not* $A = B$.

The notation $A \subseteq B$ abbreviates the property *for all* x, *if* $x \in A$ *then* $x \in B$. Axiomatic set theory contains *axioms* and *theorems*. The axioms are properties that are assumed, or postulated, to be true for all sets, while the theorems are the properties deduced from the axioms by logical rules.

The Axiom of Existence: There exists a set that has no elements.

This axiom does not tell us that the set is unique; it simply states that there is at least one such set.

The Axiom of Extensionality: If two sets have the same elements, then they are equal.

This axiom states that whenever $A \subseteq B$ and $B \subseteq A$, $A = B$.

Lemma **??** is a consequence of these two axioms.

Lemma 1.1.3 *There is only one set that has no elements.*

Proof: Let A and B be two sets with no elements. Then the assertions
if $x \in A$ then $x \in B$
and
if $x \in B$ then $x \in A$
are true because the antecedents $x \in A$ and $x \in B$ of the two implications are never true.

So, $A \subseteq B$ and $B \subseteq A$. By the Axiom of Extentionality they are equal. **Q.E.D.**

We will call this set *the empty set* and denote it by ϕ.

The Axiom Schema of Comprehension: Let $\mathbf{P}(x)$ be a property of x. For any set A there is set B such that $x \in B$ iff $x \in A$ and $\mathbf{P}(x)$.

This property is called an axiom schema because it is a collection of axioms. If we choose $\mathbf{P}(x)$ to be $x = x$ we get one axiom; if $\mathbf{P}(x)$ is $x \neq x$ we get another axiom, and so on.

This axiom schema does not state that the set B is unique, only that there is at least one such set.

Lemma 1.1.4 *For every set A there is only one set B such that $x \in B$ iff $x \in A$ and $\mathbf{P}(x)$.*

Proof: Let C be another set that satisfies the condition that $x \in C$ iff $x \in A$ and $\mathbf{P}(x)$. Then

$x \in B$ iff $x \in A$ and $\mathbf{P}(x)$ from the construction of B

iff $x \in C$ from the construction of C

So, $x \in B$ iff $x \in C$, i.e. B and C have the same elements. By the Axiom of Extensionality they are equal. **Q.E.D.**

Definition 1.1.5 *$\{x \in A | \mathbf{P}(x)\}$ is the set of all $x \in A$ with the property $\mathbf{P}(x)$.*

Observations 1.1.6 Let $\mathbf{P}[x]$ be a property.

1. The statement
there is a unique x such that $\mathbf{P}[x]$ holds
stands for the property
there is some x such that $\mathbf{P}[x]$, and
for all x and for all y,
$\mathbf{P}[x]$ and $\mathbf{P}[y]$ imply $x = y$.
2. The notation $z = \{y \in x | \mathbf{P}[y]\}$ corresponds to the property
for all x there is a unique z such that for all y,
$y \in z$ iff ($y \in x$ and $\mathbf{P}[y]$).

We can easily prove that for any set x, $\{y \in x | y = y\} = x$ and $\{y \in x | x \neq x\} = \phi$. We leave these proofs as exercises.

The Axiom of Pair: For all sets A, B there is a set C such that $A \in C$ and $B \in C$.

Lemma 1.1.7 *There is a unique set that contains only A and B. We denote this set by $\{A, B\}$.*

Proof: Let C be a set that satisfies the axiom of pair for A and B. We apply The Axiom of Comprehension to C and get the set $Z = \{x \in C | x = A \text{ or } x = B\}$. This set contains A and B because these elements are in C and they satisfy the property $x = A$ or $x = B$. If $x \neq A$ and $x \neq B$, then x does not satisfies the property $\mathbf{P}[x] : x = A$ or $x = B$, regardless of whether x is in C or not. So, $x \in Z$ iff $x = A$ or $x = B$.

We now show that the set Z is the same regardless of the choice of the set C. For this let C_1 be another set that satisfies the axiom of pair for A and B, and let $Z_1 = \{x \in C_1 | x = A \text{ or } x = B\}$. Again, $x \in Z_1$ iff $x = A$ or $x = B$.

Then $x \in Z$ iff $x = A$ or $x = B$ iff $x \in Z_1$.

So, $x \in Z$ iff $x \in Z_1$. By The Axiom of Extensionality, $Z = Z_1$.**Q.E.D.**

We write $\{A, B\}$ for the set that contains only A and B and call it the *unordered pair of* A and B. If $A = B$, then $\{A, B\} = \{A\}$.

Kuratowski defines the *ordered pair* of A and B as $\{\{A\}, \{A, B\}\}$. We denote the ordered pair of A and B by $< A, B >$.

Lemma 1.1.8 *$< A, B >=< C, D >$ implies that $A = C$ and $B = D$.*

Proof: We express the equality $< A, B >=< C, D >$ as an equality of unordered pairs and get

(1) $\{\{A\}, \{A, B\}\} = \{\{C\}, \{C, D\}\}$

Since $\{A\} \in \{\{C\}, \{C, D\}\}$, $\{A\} = \{C\}$ or $\{A\} = \{C, D\}$. In the first case we get $A = C$; in the second that $C = A$ and $D = A$. In both cases $A = C$.

Now we show that $B = D$. We have two cases, depending on whether $A = B$ or $A \neq B$.

Case 1: $A = B$. Then $\{\{A\}, \{A, B\}\} = \{\{A\}, \{A, A\}\} = \{\{A\}, \{A\}\} = \{\{A\}\}$. So, equation (1) becomes

(2) $\{\{A\}\} = \{\{C\}, \{C, D\}\}$.

From (2) we get that $\{C, D\} = \{A\}$, which yields $D = A$. Since $A = B$, $B = D$.

Case 2: $A \neq B$. From (1) we get that $\{A, B\} = \{C\}$ or $\{A, B\} = \{C, D\}$. The equality $\{A, B\} = \{C\}$ implies that $A = B = C$, contradicting $A \neq B$. So, $\{A, B\} = \{C, D\}$. Then $B = C$ or $B = D$. The equality $B = C$ is not possible since $C = A$. So we are left with $B = D$.

Since both cases yield $B = D$, we conclude the proof. **Q.E.D.**

We can also define ordered n-tuples as

$$< A_1, \ldots, A_n >= \begin{cases} < A_1, A_2 > & \text{if } n = 2 \\ << A_1, \ldots, A_{n-1} >, A_n > & \text{if } n > 2 \end{cases}$$

The Axiom of Union: For every set S, there is a set U that contains the members of the members of S.

Example 1.1.9 Let $S = \{\{A, B\}, \{C, D\}\}$. The members of S are $\{A, B\}$ and $\{C, D\}$. The members of $\{A, B\}$ are A and B, and the members of $\{C, D\}$ are C and D. So, the members of the members of S are A, B, C, and D. The Axiom of Union states that there is a set that contains A, B, C, and D and maybe other sets.

The purists may object to the statement of the axiom because it does not conform to our definition of property. We can restate it as

For every set S there is a set U such that for all x and for all y,

if $x \in y$ and $y \in S$, then $x \in U$.

This property has the same meaning as the statement of the axiom, but it lost the clarity of the the first sentence. We will continue to sacrifice the formalism in favor of clarity.

We can prove that there is a unique set that contains only the members of the members of S.

Lemma 1.1.10 *For every sets S there is a unique set U such that U contains only the members of the members of S.*

We leave the proof of the lemma as exercise.

We call U *the union of* S and we write it $\cup S$. We write $M \cup N$ instead of $\cup\{M, N\}$.

The axioms of union and pairing allow us to define the finite sets $\{A_1, A_2, \ldots, A_n\}$, whose only members are $A_1, A_2, \ldots, A_n$. For example $\{A_1, A_2, A_3\} = \cup\{\{A_1, A_2\}, \{A_3\}$. The sets $\{A_1, A_2\}$ and $\{A_3\}$ exist from The Axiom of Pairing. We apply The Axiom of Pairing to the last two sets and get the set $\{\{A_1, A_2\}, \{A_3\}$. Then we use The Axiom of Union to get the set $\{A_1, A_2, A_3\}$.

The next axiom deals with the subsets of a set.

The Axiom of Power Set: For any set S there is a set P such that P contains all subsets of S.

Lemma 1.1.11 *There is a unique set P such that $x \in P$ iff $x \subseteq S$.*

We leave the proof of this lemma as exercise. We call P the power set of S, and write it as $\mathcal{P}[S]$.

Example 1.1.12 Let us find the power set of $S = \{\phi, \{\phi\}\}$. The subsets of S are ϕ, $\{\phi\}$, $\{\{\phi\}\}$ and S itself. So, $\mathcal{P}[S] = \{\phi, \{\phi\}, \{\{\phi\}\}, \{\phi, \{\phi\}\}\}$.

The axioms listed so far allow us to define the familiar operations of union, intersection and set difference.

$A \cup B = \cup\{A, B\}$

$A \cap B = \{x \in A | x \in B\}$

$A - B = \{x \in A | x \notin B\}$

The first definition uses the axioms of pair, union, and comprehension while the last two employ the axiom of comprehension.

However, all sets produced so far are finite, and every discipline that claims to be a foundation of mathematics must be able to handle infinite structures like the set of natural numbers.

Set theory represents the natural numbers as follows: 0 is the empty set ϕ, 1 is $0 \cup \{0\} = \{\phi\}$, 2 is $1 \cup \{1\} = \{\phi\} \cup \{\{\phi\}\} = \{\phi, \{\phi\}\}$, $\ldots$,$n + 1 = n \cup \{n\}$ and so on. In this representation $1 = \{0\}$, $2 = \{0, 1\}$, $3 = \{0, 1, 2\}$, $\ldots$, $n+1 = \{0, 1, \ldots, n\}$. So, each natural number contains all natural numbers less than itself.

We describe this construction in terms of the *successor* operation, $S(x) = x \cup \{x\}$; 1 is the successor of 0, 2 is the successor of 1, $\ldots$, $n+1$ is the successor of n. The successor operation allows us to define the natural numbers, but this

does not mean that there is a set that contains all of them. Its existence is stated by the Axiom of Infinity.
The Axiom of Infinity: There is a set I such that $0 \in I$ and for all $x \in I$, $S(x) \in I$.

The sets I that satisfy the properties that $0 \in I$ and for all $x \in I$, the successor of x is also in I are called *inductive sets*. So, the Axiom of Infinity states that there are inductive sets. However, the inductive sets may contain sets that are not natural numbers, like $\{\{\phi\}\}$.

Definition 1.1.13 *Let A be any inductive set. Then $N = \{x \in A|$ for all inductive sets B, $x \in B\}$.*

Lemma 1.1.14 *N is an inductive set.*

Proof: The set 0 is in N because it belongs to all inductive sets, so it is in their intersection.

Now assume that $n \in N$ and let I be an inductive set. Then $n \in I$, because N is the intersection of all inductive sets. Since I is inductive and $n \in I$, $S(n) \in I$. So, $S(n)$ belongs to all inductive sets I. Then it belongs to their intersection, i.e. to N. **Q.E.D.**

Now we can prove that we can do mathematical induction on N.

Theorem 1.1.15 *[The Induction Principle]Let $\mathbf{P}[n]$ be a property such that*
1. *$\mathbf{P}[0]$ is true, and*
2. *for all $n \in N$, $\mathbf{P}[n]$ implies $\mathbf{P}[n+1]$.*

Proof: Let $A = \{n \in N|\mathbf{P}[n+1]\}$. The 2 properties listed above tell us that A is an inductive set. Since N is the intersection of all inductive sets, $N \subseteq A$. But $A \subseteq N$ from the definition of A. So, $N = A$. **Q.E.D.**

The next axiom tells us that every set contains *minimal elements*, i.e. members whose intersection with the set is empty.
The Axiom of Foundation: Every non-empty set A has an element x such that $A \cap x = \phi$.

Proposition 1.1.16 *For all sets x, y, at least one of the relations $x \notin y$, $y \notin x$ is true.*

Proof: Assume that both $x \in y$ and $y \in x$ are true. Then we form the set $\{x, y\}$ by using the Axiom of Pairing. By the Axiom of Foundation either $\{x, y\} \cap x = \phi$ or$\{x, y\} \cap x = \phi$. But this this is impossible since $y \in \{x, y\} \cap x$ and $x \in \{x, y\} \cap y$. **Q.E.D.**

Corollary 1.1.17 *$x \notin x$.*

Proof: In Proposition 1.1.16 we make $y = x$.

We will call the property $\mathbf{P}(x, y)$ *functional* if for every set x there is exactly one set y that makes $\mathbf{P}(x, y)$ true. For example, the properties $\mathbf{R}(x, y) : y = \{x, z\}$ and $\mathbf{Q}(x, y) : y = \mathcal{P}[x]$ are functional because the set y is uniquely defined by the set x. We will call *y the value of x*.

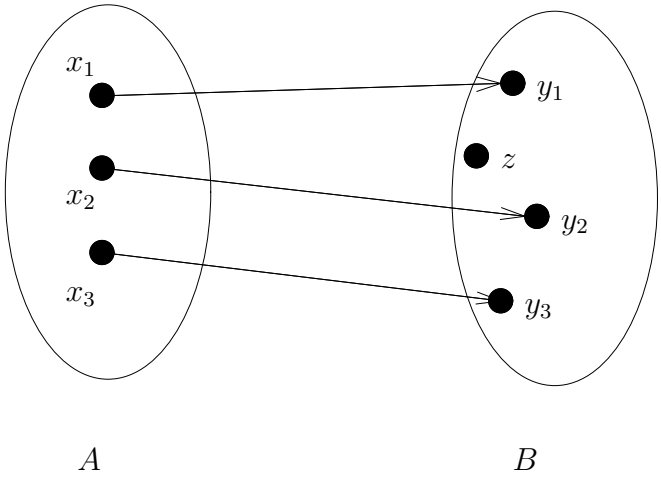


Figure 1.1: The Axiom Schema of Replacement

The next axiom tells us that for every functional property $\mathbf{P}(x, y)$ and set A, there is a set B that contains all values of the members of A.

The Axiom Schema of Replacement: Let $\mathbf{P}(x, y)$ be a functional property. Then, for every set A there is a set B such that for every $x \in A$, B contains a y such that $\mathbf{P}(x, y)$ holds.

This is an axiom schema because we can have an infinity of properties $\mathbf{P}(x, y)$. We illustrate the axiom in Figure 1.1. We represent the pairs x, y that satisfy the property $\mathbf{P}(x, y)$ by an arrow with the feathers in A and the tip in B. In our picture, A has only 3 elements, x_1, x_2, x_3. Their correponding y's, y_1, y_2, y_3 are in B. The set B may contain other elements, like the set z.

We can get rid of the extra elements of B by using the Axiom Schema of Comprehension and get the set $\{y \in B|$ there is some $x \in A$ such that $\mathbf{P}(x, y)\}$. We will write the reduced set as $C = \{y|$ there is some $x \in A$ such that $\mathbf{P}(x, y)\}$. We will call *A the index set*, and *C the indexed set*.

Example 1.1.18 Let $\mathbf{P}(x, y)$ be the property $y = \mathcal{P}[x]$, i.e. y is the power set of x and N be the set of indices. The Axiom Schemas of Replacement and Comprehension tell us that the powers of all natural numbers form a set. We write it as $\{\mathcal{P}[n]|n \in N\}$.

We will use many N-indexed sets in this book. We will write y_i for the y that satisfies the property $\mathbf{P}(i, y)$ and $\bigcup_{i=0}^{\infty} y_i$ for $\cup\{y_i|i \in N\}$.

Now we list the last axiom of ZFC.

The Axiom of Choice: Let I be a non-empty index set and $\{X_i|i \in I\}$ an I-indexed set of non-empty sets. Then there is an I indexed set $\{x_i|i \in I\}$ such that for all $i \in I$,$x_i \in X_i$.

This axiom is illustrated in Figure 1.2. Here $I = \{0, 1, 2\}$ and the elements of the sets X_0, X_1, X_2 are represented by points. The axiom of choice states

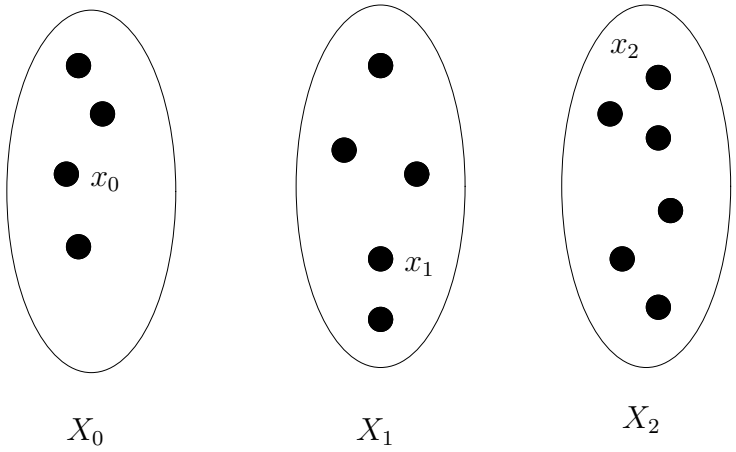


Figure 1.2: The Axiom of Choice

that there is the existence of the sequence x_0, x_1, x_2 with $x_0 \in X_0$, $x_1 \in X_1$, and $x_2 \in X_2$.

Not all properties are true. For example, Russell's paradox, that states that there is a set that contains all sets, is false.

Russell's Paradox: for all $x, x \in U$

Proposition 1.1.19 *There is no set that contains all sets.*

Proof: We can simply apply The Axiom of Foundation, but we do not need it to prove this paradox. Let us assume that there is a set U that contains all sets. We use the Axiom of Comprehension to define the set $S = \{x \in U | x \notin x\}$, i.e. a set is in S iff it is not a member of itself.

Now we ask the question *Does* $S \in S$*?*. If the answer is yes, then $S \notin S$ since S contains only the sets that are not members of themselves.

If the answer is no, i.e. $S \notin S$, then S must be S because S contains all sets that are not members of themselves.

In either case we get a contradiction, that $S \in S$ iff $S \notin S$. This is a paradox, and we got it because we assumed that there is a universal set. **Q.E.D.**

Some properties, like

$x \in y$ iff $x = A$ or $x = B$

define sets, in this case $x = \{A, B\}$. On the other side, the property $\mathbf{P} : x = x$ is satisfied by all sets, so it is true. But the collection of all sets that satisfy this property is not a set because there is no universal set. We say that this property define a *class*. Classes are collections of sets, but are not sets.

We will use this result when we talk about models. The models are sets, but there may not be a set that contains all models. In this case we will talk about the *class of models*, and not the *set of models*.

Exercises

Exercise 1.1.1 *Prove that* $\{y \in x | y = y\} = x$ *and* $\{y \in x | x \neq x\} = \phi$.

Exercise 1.1.2 *Write the set* $< A, B, C >$ *using only pairs of unordered sets.*

Exercise 1.1.3 *Rewrite the set* $\cup\{\{\{A\}\}, \{B\}, \{\{C, D\}\}, \{E\}\}$ *as a list of elements.*

Exercise 1.1.4 *Prove Lemma 1.1.10.*

Exercise 1.1.5 *Prove Lemma 1.1.11.*

Exercise 1.1.6 *Compute* $\mathcal{P}[\{A, B, C\}]$.

Exercise 1.1.7 *Show that the property for all inductive sets sets* B, $x \in B$ *can be specified in the laguage specified by Definition 1.1.1.*

Exercise 1.1.8 *Write the set representation of the numbers 3, 4, and 5.*

Exercise 1.1.9 *Use the induction principle to show that* $x \in N$ *iff* ($x = 0$ *or there is some* $m \in N$ *such that* $x = S(m)$).

Exercise 1.1.10 *Show that the property*
$\mathbf{P}[x_1, x_2, \ldots, x_n] : x_1 \in x_2 \in x_3 \in \ldots \in x_n \in x_1\}$
is not true.

Exercise 1.1.11 *Show that for all sets* x, $x \neq S(x)$

Exercise 1.1.12 *Describe the sets* $\cup N$, *and* $\cup \cup N$.

Exercise 1.1.13 *Show that the collection of all sets* y *satisfying* $\phi \in y$ *is a class.*

1.2 Relations

This section defines binary relations and basic relation operations like composition, restriction, union, intersection, and difference. Then, it takes a closer look at relations on a set, and examines properties like irreflexivity, antisymmetry, transitivity, and totality. The stress is on orderings, particularly well orderings and well founded orderings. The last propositions relate these two concepts and present a characterization of the well founded orderings. These concepts are useful in providing termination proofs for many algorithms.

The section presents the concepts in a rather informal way. The readers who are interested in the set theoretic formalism can read the observations that follow the definitions. The rest can skip them without great loss.

Definition 1.2.1 (binary relation, domain, range, field) *1. A binary relation is a set of ordered pairs.*

2. The domain of the relation R is the set $dom(R) = \{x|$ there is some y such that $< x, y > \in R\}$.

3. The range of R is the set $ran(R) = \{y|$ there is some x such that $< x, y > \in R\}$.

4. The field of R is the set $dom(R) \cup ran(R)$.

5. If the field of R is included in A we say that R is a relation on A.

Observation 1.2.2 The reader who read the preceding section will raise objections to our definitions of domain and range. He/she recalls that any definition that uses the axiom of comprehension must have the form $\{x \in$ Some-Set $|$ Some-Property-of-$x\}$. Or, we did not specify Some-Set. We can easily rectify this omission. The ordered pair $< x, y >$ is the set $\{\{x\}, \{y\}\}$. If $\{\{x\}, \{x, y\}\} \in R$, then both $\{x\}$ and $\{x, y\}$ are in $\cup R$, because $\cup R$ contains the members of the members of R. Since $\{x\}$ and $\{x, y\}$ are in $\cup R$, $x, y \in \cup \cup R$. So, the complete definitions of $dom(R)$ and $ran(R)$ are

$dom(R) = \{x \in \cup \cup R|$ there is some y such that $< x, y > \in R\}$

and

$ran(R) = \{y \in \cup \cup R|$ there is some x such that $< x, y > \in R\}$.

We were intentionally sloppy in order to focus on the concept, instead of the formalism.

Example 1.2.3 The relation $R = \{< 1, a >, < 2, b >, < 1, c >, < 3, a >\}$ has domain $\{1, 2, 3\}$ and range $\{a, b, c\}$.

Definition 1.2.4 (Cartesian Product) *Let A and B be two sets. The cartesian product of A and B is the set $A \times B = \{< x, y > | x \in A$ and $y \in B\}$.*

So, $A \times B$ contains all ordered pairs that have the first element in A and the second one in B.

Observation 1.2.5 Again, this definition is informal. We use the definition of the ordered pair to get formalize it. Let $a \in A$ and $b \in B$. Then both $\{a\}$ and $\{a, b\}$ are elements of the power set of $A \cup B$. So, $< a, b >= \{\{a\}, \{a, b\}\}$ is an element of the power set of the power set of $A \cup B$. So, the formal definition is

$A \times B = \{z \in \mathcal{P}[\mathcal{P}[\cup\{A, B\}]]|$ there are $x \in A$ and $y \in B$ such that $z =< a, b >\}$.

Proposition 1.2.6 *The relation R is a subset of the cartesian product of its domain and its range, i.e. $R \subseteq dom(R) \times ran(R)$.*

Proof: Let $z \in R$. Since R is a set of ordered pairs, $z =< x, y >$ for some sets x and y. From the definition of $dom(R)$ and $ran(R)$ we get that $x \in dom(R)$ and $y \in ran(R)$. Then $z =< x, y >$ is a member of $dom(R) \times ran(R)$ because $dom(R) \times ran(R)$ contains all pairs $< x, y >$ with $x \in dom(R)$ and $y \in ran(R)$. So, $R \subseteq dom(R) \times ran(R)$. **Q.E.D.**

If $dom(R) \subseteq A$ and $ran(R) \subseteq B$, then R is also a subset of $A \times B$. In many cases we will focus on the interaction between R and the sets A and B. In those cases we will say that *R is a relation from A to B*.

Note 1.2.7 In order to simplify the notation, we will frequently write aRb or $R(a,b)$ instead of the more formal notation $< a,b > \in R$.

The next concept will be of great importance in the next section.

Definition 1.2.8 (inverse relation) *Let R be a relation. The inverse of R is the relation $R^{-1} = \{< y,x > \mid < x,y > \in R\}$.*

So, each pair of R^{-1} is obtained from a pair of R by switching the order of the elements.

Example 1.2.9 Let $R = \{< a,1 >, < a,3 >, < b,1 >, < c,2 >\}$. Then $R^{-1} = \{< 1,a >, < 3,a >, < 1,b >, < 2,c >\}$.

The proof of the next proposition is left as exercise.

Proposition 1.2.10 *1. $R^{-1} \subseteq ran(R) \times dom(R)$.*
2. $(R^{-1})^{-1} = R$.

Definition 1.2.11 (products, powers) *Let $n \geq 1$ and $A_1, \ldots, A_n$ be n sets. The product $A_1 \times A_2 \times \ldots \times A_n$ is defined as follows:*
if $n = 1$, the product is A_1,
if $n = 2$, the product is $A_1 \times A_2$,
if $n > 2$, the product is $A_1 \times \ldots \times A_{n-1} \times A_n = (A_1 \times \ldots \times A_{n-1}) \times A_n$.
If $A_1 = A_2 = \ldots = A_n = A$, the product

$$\underbrace{A \times \ldots \times A}_{n\text{times}}$$

, written A^n, is called the n-th power of A.

Definition 1.2.12 (n-ary relations) *Let $n \geq 1$. A n-ary relation among $A_1, \ldots, A_n$ is a subset R of $A_1 \times A_2 \times \ldots \times A_n$.*
If $n = 1$ the relation is called unary. In this case $R \subseteq A_1$. If $n = 2$ the relation is called binary; if $n = 3$ the relation is called ternary, and so on.
The number n is called the arity of the relation.

Example 1.2.13 1. $R = \{n \in N | n \text{ is even}\}$ is a unary relation on the set of natural numbers $N = \{0, 1, \ldots, n, \ldots\}$.
2. $R = \{<< a,\alpha >,1 >, << a,\beta >,2 >, << b,\gamma >,2 >, << b,\gamma >,3 >\}$ is a ternary relation among $A = \{a,b\}$, $B = \{\alpha,\beta,\gamma\}$, and $C = \{1,2,3\}$.

We can use the set operations $\cup$, $\cap$, $-$ to define the *union*, *intersection*, and *difference* of two relations.

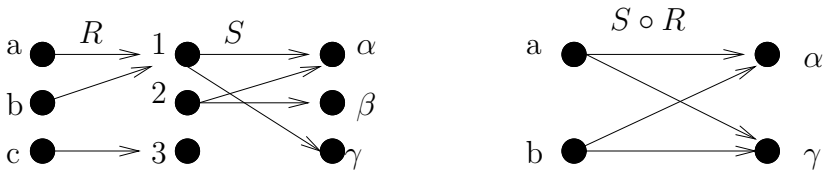


Figure 1.3: The composition of two relations

Definition 1.2.14 (set operations on relations) *Let R and S be two relations among $A_1,\ldots,A_n$. Then*

$R \cup S = \{x \in A_1 \times A_2 \times \ldots \times A_n | x \in R$ or $x \in S\}$ is the union of R and S.

$R \cap S = \{x \in A_1 \times A_2 \times \ldots \times A_n | x \in R$ and $x \in S\}$ is the intersection of R and S.

$R - S = \{x \in A_1 \times A_2 \times \ldots \times A_n | x \in R$ and $x \notin S\}$ is the difference of R and S.

Example 1.2.15 Let R be the set of natural numbers divisible by 3 and S be the set of natural numbers divisible by 5.

Then $R \cup S$ is the set of natural numbers divisible by 3 or divisible by 5, $R \cap S$ is the set of natural numbers divisible by 15, and $R - S$ is the set of natural numbers divisible by 3 and not divisible by 5.

Definition 1.2.16 (relation composition) *Let R and S be two binary relations. The composition of R and S, writen $S \circ R$ or $R.S$, is the relation*

$S \circ R = \{< x, y > \in dom(R) \times ran(S) |$ for some $z \in ran(R)$, $< x, z > \in R$ and $< z, y > \in S\}$.

Example 1.2.17 Let us compose the relations $R = \{< a, 1 >, < b, 1 >, < c, 3 >\}$ and $S = \{< 1, \alpha >, < 1, \gamma >, < 2, \alpha >, < 2, \beta >\}$. We represent the pairs $< x, y >$ as arrows $x \longrightarrow y$. We call x the *source* and y the target of the arrow.

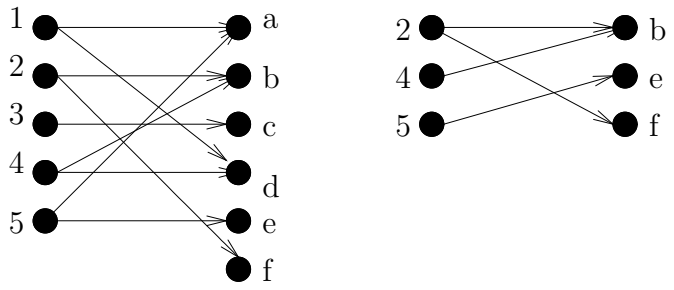
Figure **??** shows the arrow representations of the relations R, S, and $S \circ R$. We compute $S \circ R$ by matching all arrows in R with all arrows in S. If the *target* of an R-arrow is the same as the source of an S-arrow, then we draw an $S \circ R$ arrow from the source of the R-arrow to the target of the S-arrow.

For example, $< a, \alpha >$ is obtained by matching the R-arrow $a \longrightarrow 1$ with the S-arrow $1 \longrightarrow \alpha$.

Now let us compute $R \circ S$ by identifying all sequences $x \longrightarrow y \longrightarrow z$ having the first arrow in S and the second one in R. But there are no such sequences since the targets of the S arrow are Greek letters and the souces of the R arrows are Roman letters. So, $S \circ R = \phi$.

The next proposition tells us that the result of a sequence of compositions is independent of the order in which we do the compositions.

Proposition 1.2.18 *[the associativity of composition] $(T \circ S) \circ R = T \circ (S \circ R)$.*



R $\{1,2,4,5\} \uparrow R \uparrow \{b,c,e,f\}$

Figure 1.4: Restriction of a relation

Proof: We need to show that the sets $(T \circ S) \circ R$ and $T \circ (S \circ R)$ are equal. So, we show the inclusions (1) and (2).

(1) $(T \circ S) \circ R \subseteq T \circ (S \circ R)$

(2) $T \circ (S \circ R) \subseteq (T \circ S) \circ R$

We will prove (1) first. Let $< x, y > \in (T \circ S) \circ R$. From the definition of composition there is some z, such that

(3) $< x, z > \in R$, and

(4) $< z, y > \in T \circ S$.

Since $< z, y > \in T \circ S$, there is some u such that

(5) $< z, u > \in S$, and

(6) $< u, y > \in T$.

From (3) and (5) we get

(7) $< x, u > \in S \circ T$.

From (7) and (6) we get that

(8) $< x, y > \in T \circ (S \circ R)$.

Since $< x, y >$ is arbitrary in $(T \circ S) \circ R$, $(T \circ S) \circ R \subseteq T \circ (S \circ R)$.

The inclusion (2) is proved the same way. We leave the proof as exercise. **Q.E.D.**

Definition 1.2.19 *[restrictions and extensions of relations] Let R be a binary relation and A, B, be two sets. The restriction of R to A and B, writen $A \uparrow R \uparrow B$, is the relation $R \cap A \times B$.*

Any relation S such that $R \subseteq S$ is an extension of R.

Examples 1.2.20 1. Figure 1.4 shows the restriction of R to $\{1,2,3,4\} \times \{b,c,e,f\}$. The arrows of $\{1,2,3,4\} \uparrow R \uparrow \{b,c,e,f\}$ are the arrows of R that have their source in $A = \{1,2,3,4\}$ and their target in $B = \{b,c,e,f\}$.

2. Let $> \subseteq R \times R$ be the greater than relation on the real numbers. Then the restriction $N \uparrow > \uparrow N$ is the greater than relation on the set of natural numbers.

3. The relation $S = \{< a,1 >, < a,2 >, < b,2 >, < c,3 >\}$ is an extention of $R = \{< a,1 >, < b,2 >, < c,3 >\}$ because every element of R is in S.

Definition 1.2.21 *Let R be a relation on A.*

1. R is said to be reflexive if for all $a \in A$, $< a, a > \in R$.

2. R is said to be irreflexive if for all $a \in A$, $< a, a > \notin R$.

3a. R is said to be right connected if for all $a \in A$ there is some $b \in A$ such that $< a, b > \in R$.

3b. R is said to be left connected if for all $a \in A$ there is some $b \in A$ such that $< b, a > \in R$.

4. R is said to be symmetric if for all $a, b \in A$, $< a, b > \in R$ implies that $< b, a > \in R$.

5. R is said to be antisymmetric if for all $a, b \in A$, $< a, b > \in R$ and $< b, a > \in R$ imply that $a = b$.

6. R is said to be transitive if for all $a, b, c \in A$, $< a, b > \in R$ and $< b, c > \in R$ imply that $< a, c > \in R$.

Examples 1.2.22 First of all, let us recall that R is a relation on A when its field is a subset of A, i.e. $dom(R) \cup ran(R) \subseteq A$. Now, let us look at two familiar relations and check if they have any of the the properties listed in the preceding definition.

I. Let $<$ be the relation *less than* defined on the set of natural numbers.

1. $<$ is not reflexive because $1 \not< 1$.

2. $<$ is irreflexive because for all natural numbers n, $n \not< n$.

3. $<$ is connected because for all natural numbers n, $n < n + 1$.

4. $<$ is not symmetric because $1 < 2$, but $2 \not< 1$.

5. $<$ is antisymmetric because for all natural numbers m, n, $m < n$ and $n < m$ imply $m = n$.

6. $<$ is transitive because for all natural numbers m,n, p, $n < m$ and $m < p$ imply $n < p$.

II. Let A be a nonempty set and ϕ be the empty relation on A. Then

1.ϕ is not reflexive because there are elements a in A, and for none of these elements, $< a, a > \in \phi$.

2.ϕ is irreflexive because for all $a \in A$, $< a, a > \notin A$.

3. ϕ is not connected. Since ϕ is empty, there are elements $a \in A$ that are not related to any element of A.

4.ϕ is symmetric. Since there are no pairs $< a, b >$ in ϕ, the assertion
for all $a, b \in A$, $< a, b > \in \phi$ implies $< b, a > \in \phi$
is true by default.

5.ϕ is antisymmetric. Since there are no pairs $< a, b >$ in ϕ, the assertion
for all $a, b \in A$, $< a, b > \in \phi$ and $< b, a > \in \phi$, implies $< b, a > \in \phi$
is true by default.

6. ϕ is transitive. Since there are no pairs $< a, b >$, $< b, c >$ in ϕ, the assertion
for all $a, b, c \in A$, $< a, b >, < b, c > \in \phi$ imply $< a, c > \in \phi$
is true by default.

Observation 1.2.23

- Every reflexive relation is both left and right connected, because for all $a \in A$, $< a, a > \in R$.

- There are relations that are neither reflexive nor irreflexive. For example $R = \{< 1,1 >, < 1,2 >\}$ is a relation on $A = \{1,2\}$ that is neither reflexive nor irreflexive. It is not reflexive because $< 2,2 > \notin R$, and it is not irreflexive because $< 1,1 > \in R$.

- If $A \neq \phi$, then no relations on A is **both** reflexive and irreflexive. This is easy to show. Let $a \in A$. The reflexivity implies that $< a,a > \in A$, while irreflexivity requires that $< a,a > \notin A$, and we cannot have these two conditions met at the same time.

- The antisymmetry property is *NOT* the negation of the symmetry property. There are relations that are both symmetric and antisymmetric. The relation $R = \{< 1,1 >, < 2,2 >, < 3,3 >\}$ is a relation on $A = \{1,2,3\}$ that is both symmetric and antisymmetric. Let us prove it.

  We first show that R is symmetric. If $< a,b > \in R$ then $a = b$. So, $< b,a > = < a,b >$ and we have that $< b,a > \in R$.

  Now, let us check the antisymmetry. Assume that $< a,b > \in R$ and $< b,a > \in R$. If $< a,b > \in R$ then $a = b$. So, the assertion

  for all $a,b \in A$, $< a,b > \in R$ and $< b,a > \in R$ implies $a = b$

  is true.

Definition 1.2.24 (partial orders, strict orders) *The relation R on A is called a partial order (or ordering) if it is reflexive, antisymmetric , and transitive. The relation R on A is called a strict order (ordering) if it is transitive and irreflexive.*

Proposition 1.2.25 *All strict orderings are antisymmetric.*

Proof: Assume the ordering R is not antisymmetric and let A be the field of R. Then there are two elements $a \neq b$ of A such that $< a,b > \in R$ and $< b,a > \in R$.

Since R is transitive, $< a,b > \in R$ and $< b,a > \in R$ imply $< a,a > \in R$. But this contradicts the irreflexivity of R. **Q.E.D.**

Examples 1.2.26 1. The relation $>$ is a strict ordering on the set of natural numbers.

2. The relation $<$ is a strict ordering on the set of integers.

Proposition 1.2.27 relates the partial orders to the strict orders.

Proposition 1.2.27 *Let R be a relation on the set A and $1_A = \{< a,a > | a \in A\}$.*

1. If R is a partial order then $R - 1_A$ is a strict order.

2. If R is a strict order, then $R \cup 1_A$ is a partial order.

Proof: 1. Assume that R is a partial order on A and let $S = R - 1_A$. Then, S is irreflexive because for all $a \in A, < a,a > \notin R - 1_A = S$.

Now let $< x,y >, < y,z > \in S$. Then $< x,z > \in R$ because R is transitive. We need to show that $< x,z > \in S$, i.e. $x \neq z$. Assume that $x = z$.

Since $< x, y >, < y, x > \in S$, $x \neq y$. But then, R is no longer antisymmetric, contradicting the fact that R is a partial order. So, $x \neq z$ and S is transitive.
2. Assume that R is a strict order on A and let $S = R \cup 1_A$.
S is reflexive: For all $a \in A$, $< a, a > \in 1_A \subseteq S$.
S is transitive: Let $< x, y >, < y, z > \in S$.

Case 1: $x \neq y$ and $y \neq z$. Then $< x, y >, < y, z > \in R$, so $< x, y > \in R$ by the transitivity of R. Since $R \subseteq S$, $< x, y > \in S$.

Case 2: $x = y$ and $y \neq z$. Then $< x, z > \in R$. Since $R \subseteq S$, $< x, y > \in S$.

Case 3: $x \neq y$ and $y = z$. Then $< x, z > \in R$. Since $R \subseteq S$, $< x, y > \in S$.

Case 4: $x = y$ and $y = z$. Then $x = z$ and $< x, z > \in 1_A$, so $< x, y > \in S$.

In all 4 cases, $< x, z > \in S$, so S is transitive.
S is antisymmetric. Assume that $< x, y >, < y, x > \in S$. If $x \neq y$, then $< x, y >, < y, x > \in R$, contradicting Proposition 1.2.25. So, $x = y$. **Q.E.D.**

Proposition **??** allows us to go from a strict order to the corresponding partial order and back by adding, respectively subtracting 1_A.

Notation 1.2.28 *We write $\geq$ or $\leq$ for partial orders. Their corresponding strict orders are denoted by $>$, respectively $<$.*

Definition 1.2.29 (total order) *Let R be a strict order on A. R is a total order if for all elements a,b of A, either aRb or $a = b$, or bRa.*

The or in the above sentence is exclusive, i.e. one and only one of the choices aRb, $a = b$, bRa must be satisfied.

Examples 1.2.30 *1. The relation $>$ is total on the set of natural numbers, because for all pairs m, n either $m > n$, or $m = n$, or $n > m$.*

2. Let A be the set of all strings that can be formed with the letters a, and b and let $\succ$ be the relation defined by $s \succ t$ if the string s is longer than the string t. The relation $\succ$ is not total because the strings ab and ba are not equal and neither one of them is longer than the other.

Definition 1.2.31 (least, minimal, greatest, maximal) *Let $\geq$ be partial ordering on A and B be a non-empty subset of A.*

1. $b \in B$ is the least element of B if for all $x \in B$, $x \geq b$.

2. $b \in B$ is a minimal element of B if for all $x \in B$, $b \not> x$.

3. $b \in B$ is the greatest element of B if for all $x \in B$, $b \geq x$.

4. $b \in B$ is a maximal element of B if for all $x \in B$, $x \not> b$.

Example 1.2.32 Let A be the set of all finite strings that can be formed with the letters a and b. On this set we define the ordering $>$ by $s > t$ if the string s is longer than the string t.

I. Let $B = \{aba, ba, ab, bbb\}$. The restriction of $>$ to B is $\{aba > ba, aba > ab, bbb > ba, bbb > ab\}$.

1. B has no least element. The least element of B must be a string with the smallest length, so we eliminate aba and bbb. The strings with the smallest lengths are ab and ba. They are distinct and neither $ab > ba$ nor $ba > ab$. So, neither one is the least element.

2. B has minimal elements. They are the strings of minimal length, in our case 2. So, the minimal elements are ab and ba.

3. B has no greatest element. The greatest element must be a string of greatest length, in our case 3. But we have 2 strings,bbb and aba of length 3. Neither one is greater than the other, so we have no greatest element.

4. B has maximal elements. These are the longest strings, bbb and aba.

II. Let $C = \{a, ba, ab, bbab\}$. The restriction of $>$ to C, $C \uparrow > \uparrow C$, is $\{bbab > ba, bbab > ab, bbab > a, ba > a, ab > a\}$. - 1. The least element is a because all other elements of C are longer than it.

2. The minimal element of C is also a. For all the other strings we can find a shorter string in C.

3. The greatest element of C is $bbab$ because it is longer than all the other strings.

4. The only maximal element of C is $bbab$. For all the other elements of C we can find a string, namely $bbab$, that is longer.

The next relates the four concepts described above. Its proof is straightforward and is left as exercise.

Proposition 1.2.33 *1. Every least element is minimal.*

2. Every greatest element is maximal.

3. If $>$ is total, then every minimal element is a least element.

4. If $>$ is total, then every maximal element is a greatest element.

The next concept plays a key role in set theory.

Definition 1.2.34 (well order) *Let $>$ be a strict order on A. The relation $>$ is called a well order if every non-empty subset of A contains a least element.*

Examples 1.2.35 1. $>$ is a well order on the set of natural numbers.

Every non-empty subset S of natural numbers contains a least element, i.e. a number less than or equal to every element in the subset.

2. The relation $>$ defined on the set of rational numbers is not a well-order. The subset $S = \{1/1, 1/2, 1/3, \ldots, 1/n, \ldots\}$ has no least element since for every element $1/n$ of S, the next element $1/(n+1)$ is smaller than it.

Proposition 1.2.36 *Every well order is total.*

Proof: Let $>$ be a well order on A, and x, y be two elements of A. We need to show that only one of $x > y$, $x = y$, and $y > x$ holds.

Case 1: $x = y$. Then $x > y$ and $y > x$ reduce to $x > x$, and $x > x$ is false because $>$ is irreflexive.

Case 2: $x \neq y$. Then the set $\{x, y\}$ has a least element. If that element is x then $y > x$. If that element is y then $x > y$. So, at least one of $x > y$, $y > x$ is true. If both are true then we get $x > x$ by transitivity. The last realtion contradicts the irreflexivity of $>$. So, only one is true. **Q. E. D.**

The next concept is useful in proving program termination.

Definition 1.2.37 (well founded order) *Let $>$ be a strict order on A. We say that $>$ is a well founded if every non-empty subset of A has minimal elements.*

Examples 1.2.38 1. Let $\{\ a,\ b\ \}^*$ be the set of all finite strings that can be formed with the letters a and b. On this set we define the relation $\succ$ as

$u \succ v$ if u has more a letters than v.

The relation $\succ$ is irrefexive since no string can have more a's than itself.

The relation $\succ$ is transitive because

if u has more a 's than v and v has more a's than w, then u has more a's than w.

So, $\succ$ is a strict order. At the same time every non-empty subset of $S \subseteq \{\ a,\ b\ \}^*$ has minimal elements. Those elements are precisely the strings with the smallest number of **a**'s.

The relation $\succ$ is not total because none of the relations

$ab = ab$, $ab \succ ba$, $ba \succ ab$, is true.

Moreover, there are subsets of $\{a,\ b\ \}^*$ that have more than one minimal element. In the set $\{\ ab,\ ba\ \}$, both ab and **ba** are minimal.

2. The relation $>$ is not a well founded ordering on the set of integers because the subset of negative integers does not contain a minimal element.

Proposition 1.2.39 *1. Every well order is a well founded order.*

2. Every well founded order that is total is a well order.

Proof: 1. Let $>$ be a well order on A. Let S be a non-empty subset of A. Then S has a least element a. By Proposition 1.2.33, a is a minimal element. So, S has minimal elements.

3. Let $>$ be a total order that is well founded on A. Let S be a non-empty subset of A. Since $>$ is well founded, S has minimal elements. Let a be a minimal element of S. By Proposition 1.2.33 a is a least element of S. **Q.E.D.**

Now we are ready to give a theorem that characterizes the well founded orderings.

Theorem 1.2.40 *Let $>$ be a strict order on A. Then $>$ is a well founded ordering iff it does not contain an infinitely descending sequence $a_0 > a_1 > a_2 > \ldots > a_n > \ldots$.*

Proof: First of all let us assume that $>$ is an ordering on A.

Since the theorem has the form

Statement1 iff Statement2 we need to show the two implications,

if Statement1 then Statement2 and

if Statement2 then Statement1.

$\Longrightarrow$: We will prove that no well founded relation has an infinitely descending chain. We do it by contradiction. Assume that the well founded order $>$ contains the chain $a_0 > a_1 > a_2 > ... > a_n > \ldots$. Let $S = \{a_0, \ldots, a_n, \ldots, \ldots\}$. Then S

has no minimal element, because for a_n there $a_n > a_{n+1}$. This contradicts the fact that $>$ is a well founded ordering on A.
$\Longleftarrow$: We will show that whenever $>$ has no infinitely descending sequences, $>$ is well founded. We will actually show prove a statement equivalent to it, that every ordering that is not well founded contains an infinitely descending sequence. (This is called the *contrapositive* of the statement.)

So, assume that $>$ is not a well founded ordering. Then there is a subset S of A that does not have a minimal element. Let a_1 be an element of S. Since a_0 is not minimal then there is an element a_1 such that $a_0 > a_1$. Now a_1 is not minimal, so there is some $a_2 \in A$ such that $a_1 > a_2$, and so on. We can formalize this argument by using induction on the set of natural numbers, N, to get an infinitely descending sequence $a_0 > a_1 > a_2 > \ldots > a_n > \ldots$. **Q.E.D.**

Note 1.2.41 *In the above proof we used the axiom of choice. Where we did it?*

Exercises

Exercise 1.2.1 *Show that* $field(R) \subseteq \cup \cup R$.

Exercise 1.2.2 *Show that* $R^{-1} = \{< y, x > | < x, y > \in R\}$ *is a set by rewriting it as* $R^{-1} = \{z \in SomeSet | \mathbf{P}[z]\}$.

Exercise 1.2.3 *Prove the existence of* R^{-1} *using the axioms of replacement and comprehension.*

Exercise 1.2.4 *Prove Proposition 1.2.10.*

Exercise 1.2.5 *Let* R *and* S *be two relations. Show that* $S \circ R = \phi$ *iff* $ran(R) \cap dom(S) = \phi$.

Exercise 1.2.6 *The relations below are defined on the set of natural numbers. For each one specify whether or not it is reflexive, irreflexive, left and right connected, symmetric, antisymmetric, and transitive.*

1. $\leq$
2. $\geq$
3. R *where* $< m, n > \in R$ *if* m *and* n *have a common divisor greater than 1. We say that* m *is a divisor of* n *if* $m > 0$ *and the remainder of* n *divided by* m *is 0.*

Exercise 1.2.7 *Asumme that* A *is a nonempty set. Let us define the relation* $1_A = \{< a, a > | a \in A\}$. *Find out if if the relation* 1_A *is reflexive, irreflexive, connected, symmetric, antisymmetric, and transitive. Prove your assertions.*

Exercise 1.2.8 *Let* A *be the empty set and* $R = \phi$ *be the empty relation on* A. *Check if* R *is reflexive, irreflexive, connected, symmetric, antisymmetric, and transitive. Prove your assertions.*

Exercise 1.2.9 *Let* R *be a relation on a set* A *and* $1_A = \{< a, a > | a \in A\}$. *Show that if* R *is both symmetric and antisymmetric, then* $R \subseteq 1_A$.

Exercise 1.2.10 *Let R be a relation on a set A. Show that R is reflexive iff $1_A \subseteq R$.*

Exercise 1.2.11 *Now let S be a relation on A. Show that S is antisymmetric iff $S \cap S^{-1} \subseteq 1_A$.*

Exercise 1.2.12 *We say that an operation preserves a property if whenever the operands have the property, the result also has that property. For example, the union preserve the reflexity. This means that when the relations R and S, both defined on the same set A, are reflexive, so is $R \cup S$.*

Prove that the union preserves the following properties: reflexivity, irreflexivity, connectiveness, and symmetry. Give counterexamples to show that the union does not preserve the antisymmetry and the transitivity.

Exercise 1.2.13 *Specify whether the properties reflexivity, irreflexivity, connectiveness, symmetry, antisymmetry and transitivity are preserved by the intersection.*

Prove your answers.

Exercise 1.2.14 *Let x be a set. We define the membership relation on x as $\in_x = \{< y, z >\in x \times x | y \in z\}$. Show that the foundation axiom can also be stated as*

for all sets A, $\in_A$ is a well founded relation, hence the name The Axiom of Foundation.

Exercise 1.2.15 *Prove Proposition 1.2.33.*

Exercise 1.2.16 *Let A be a nonempty set and ϕ be the empty relation on A. We know, from Example 1.2.22, that ϕ is a strict ordering. Is ϕ a well ordering? Is ϕ a well founded ordering?*

Prove your answers.

Exercise 1.2.17 *Check if these orders are well founded.*

1. The relation $<$ on the set of natural numbers.

2. The alphabetical ordering on the set of strings formed with the letters a, b, c.

3. On the set of all strings that can be formed with the letters a and b, we define the relation $\succ$ by $s \succ t$ if $s \neq t$ and t is a substring of s.

Exercise 1.2.18 *Let R be a relation, and A and B be two sets such that $dom(A) \subseteq A$, and $ran(R) \subseteq B$, then $R \circ 1_A = 1_B \circ R = R$, where 1_X, called the identity on X, is the relation, $\{< x, x > | x \in X\}$.*

Exercise 1.2.19 *Let R, S and T be 3 relations. Show that*

1. $S \subseteq T$ implies $S \circ R \subseteq T \circ R$, and

2. $S \subseteq T$ implies $R \circ S \subseteq R \circ T$.

Exercise 1.2.20 *Let R be a relation on A. Show that R is transitive iff $R \circ R \subseteq R$.*

Exercise 1.2.21 *Let R be a relation on A. We define R^n as follows:*
$R^1 = R$
$R^{n+1} = R^n \circ R$
Now we define $R^\dagger = R^0 \cup R^1 \cup R^2 \cup ... \cup R^n \cup .....$ Show that
1. $R^\dagger$ is transitive
2. if S is a transitive relation that includes R then $R^\dagger \subseteq S$.

1.3 Functions

The functions are of the most important concepts in mathematics. It is next to impossible to do mathematics without using it. Intuitively, a function is a procedure, a rule, a law of correspondence, that assigns to each each element of the domain a unique value in the range. So, the functions are a special case of binary relations, $f \subseteq A \times B$. Here, we do not focus on the law of correspondence that defines the pairs of f, but on the properties that relate the sets R, A, and B.

Definition 1.3.1 (function) *A relation F is a function if for all $x \in dom(F)$ and for all $y, z \in ran(F)$, xFy and xFz imply $y = z$.*

So, a function assigns to for every element of the domain a unique member of the range. For each pair xFy, we call *y the value of F* at x and write $y = F(x)$. The element x of the pair xFy is called an *F-pre-image* of y. If the element x is not in the domain of F we say that F is *undefined* at x.

Examples 1.3.2 1. The relation $F = \{< a, 1 >, < b, 2 >, < c, 1 >\}$ has domain $A = \{a, b, c\}$ and range $B = \{1, 2\}$. R is a function because every element of the domain is related to a single element of the range.

2. The relation $G = \{< a, 1 >, < b, 2 >, < a, 3 >, < c, 3 >\}$ is not a function because the element a of its domain is related to two elements of the range, namely 1 and 3.

Definition 1.3.3 *[on, into, onto functions] Let f be a function and A, B be sets.*
1. We say that f is a function from A if $dom(f) \subseteq A$.
2. We say that f is a function on A if $dom(f) = A$.
3. We say that f is function into B if $ran(f) \subseteq B$.
4. We say that f is function onto B if $ran(f) \subseteq B$.

Examples 1.3.4 Let $F = \{< a, 1 >, < b, 2 >< c, 3 >\}$, $A_1 = \{a, b, c\}$, $A_2 = \{a, b, c, d\}$, $A_3 = \{a, c, d\}$, $B_1 = \{1, 2, 3\}$, $B_2 = \{1, 2, 3, 4\}$, $B_3 = \{2, 3, 5\}$. The function F has domain $dom(F) = \{a, b, c\}$ and range $ran(F) = \{1, 2, 3\}$.

1. F is a function on A_1 onto B_1 because $dom(F) = A_1$ and $ran(F) = B_1$.

2. F is a function on $A_1 = \{a, b, c\}$ into B_2. It is not onto because $ran(F) \neq B_2$.

3. F is a function from A_2 onto B_1. It is not on because $A_2 \neq dom(F)$.

4. F is a function from A_2 into (or to) B_2.

5. F is not a function from A_3 because $dom(F) \not\subseteq A_3$.
6. F is not a function into B_3 because $ran(F) \not\subseteq B_3$.

We write $f : A \longrightarrow B$ to show that f is a function from A to B. Most of the time we will denote functions by the lower case letters f,g,h, at times with subscripts and/or superscripts.

Since functions are relations, we can apply the concepts of *restriction* and *extension* from in the preceding section. We are particularly interested in restricting the domain of the function.

Definition 1.3.5 (domain restriction) *Let f be a function. The restriction of f to C is the relation $C \uparrow f = f \cap (C \times ran(F))$.*

Example 1.3.6 The function $f = \{< a, 1 >, < b, 2 >, < c, 2 >\}$ has domain $dom(f) = \{a, b, c\}$ and range $ran(F) = \{1, 2\}$. Let $C = \{a, b\}$. Then $C \uparrow f$ is the relation
$\{< a, 1 >, < b, 2 >, < c, 2 >\} \cap (\{a, b\} \times \{1, 2\})$
$= \{< a, 1 >, < b, 2 >, < c, 2 >\} \cap \{< a, 1 >, < a, 2 >, < b, 1 >, < b, 2 >\}$
$= \{< a, 1 >, < b, 2 >\}$.

Proposition 1.3.7 *$C \uparrow f$ is a function.*

Proof: $C \uparrow f$ is a subset of f, so its members are ordered pairs. If $C \uparrow f$ is not a function, then it must have two members, $< x, y >$ and $< x, z >$ with $y \neq z$. But then, $< x, y >$ and $< x, z >$ must be in f, contradicting the assumption that f is a function.**Q.E.D.**

Now, let us look at the composition of functions. The functions are relations, so we can compose them. We recall that the composition of R and S, in this order, is $S \circ R = \{< x, y > \in dom(R) \times ran(S) |$ for some $z \in ran(R), < x, z > \in R$ and $< z, y > \in S\}$.

Proposition 1.3.8 *1. If f and g are functions, so is $g \circ f$.*

2. For all $x \in dom(x)$, $g \circ f$ is defined at x iff f is defined at x and g is defined at $f(x)$. In that case, $(g \circ f)(x) = g(f(x))$.

Proof: 1. Assume that $g \circ f$ is not a function. Then it has two members $< x, y_1 >$ and $< x, y_2 >$ with $y_1 \neq y_2$. By the definition of composition there are two sets, z_1 and z_2, such that $< x, z_1 > \in f$, $< z_1, y_1 > \in g$, $< x, z_2 > \in f$, and $< z_2, y_2 > \in g$. The function f can have only one value at x, so $z_1 = z_2$. Since g is also a function, it can have only one value at $z_1 = z_2$. So, $y_1 = y_2$. But this contradicts the assumption that $y_1 \neq y_2$. So, $g \circ f$ is a function.

2. Assume that $(g \circ f)$ is defined at x. So, there is some z such that $< x, z > \in g \circ f$. By the definition of composition there is a y such that $< x, y > \in f$ and $< y, z > \in g$. Since f, g, $g \circ f$ are functions, $< x, y > \in f$ means that f is defined at x and $y = f(x)$, $< y, z > \in g$ means that g is defined at $y = f(x)$ and $z = g(f(x))$, and $< x, z > \in g \circ f$ means that $(g \circ f)(x) = z = g(f(x))$.

Now we must go the other way and assume that f is defined at x, and g is defined at $f(x)$. Then $< x, f(x) > \in f$, $< f(x), g(f(x)) > \in g$. By the definition

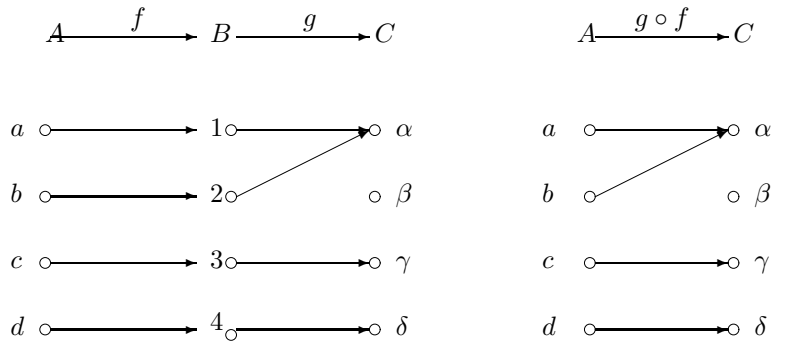


Figure 1.5: Example of composition

of composition, $< x, g(f(x)) > \in g \circ f$. This means that $g \circ f$ is defined at x, and $(g \circ f)(x) = g(f(x))$. **Q.E.D.**

The next two examples show how to compose functions.

Examples 1.3.9 1. Let $\mathbf{R}$ be the set of real numbers and $f, g : \mathbf{R} \longrightarrow \mathbf{R}$ be the functions $f(x) = x + 1$ and $g(x) = x^2$. Then $g \circ f : \mathbf{R} \longrightarrow \mathbf{R}$ is defined as $(g \circ f)(x) = g(f(x)) = (f(x))^2 = (x + 1)^2 = x^2 + 2x + 1$. The composition $f \circ g : \mathbf{R} \longrightarrow \mathbf{R}$ is defined as $(f \circ g)(x) = f(g(x)) = g(x) + 1 = x^2 + 1$. We see that $g \circ f \neq f \circ g$.

2. Let f and g be the functions shown in Figure 1.5. Then $g \circ h$ is the function displayed in the right half of the figure.

Since the composition of relations is associative, so is the composition of functions. Now we recall that every relation has an inverse. The inverse of the function f is the relation $f^{-1} = \{< y, x > | xfy\}$. The next example shows that the inverse of a function is not always a function.

Example 1.3.10 Let $f = \{< a, 1 >, < b, 2 >, < c, 1 >\}$. The inverse $f^{-1} = \{< 1, a >, < 2, b >, < 1, c >\}$ is not a function because it has two values, a and c, at 1.

Definition 1.3.11 (one-to-one function) *The function f is one-to-one or injective, if for all the values x and y in its domain, $f(x) = f(y)$ implies $x = y$. This means that every object in the range has a unique pre-image.*

Examples 1.3.12 The function $f = \{< a, 1 >, < b, 2 >, < c, 3 >\}$ is one-to-one because every object in the range $\{1, 2, 3\}$ has a unique pre-image.

The function $g = \{< a, 1 >, < b, 2 >, < c, 1 >\}$ is not one-to-one because the object 1 of the range has two pre-images, a and c.

Now we can show that the inverse of a one-to-one function is also a function.

Proposition 1.3.13 *Let f be a function. Then f^{-1} is a function iff f is one-to-one.*

Proof: $\Longrightarrow$ Let us assume that f^{-1} is a function and let $f(x) = f(y)$. Then $< f(x), x >, < f(y), y > \in f^{-1}$. Since f^{-1} is a function, $f(x) = f(y)$ implies $x = y$ because a function cannot have two values at the same point. So, f is one-to-one.

$\Longleftarrow$ Assume now that f is one-to-one and let $< x, y >, < x, z > \in f^{-1}$. Then $< y, x >, < z, x > \in f$, i.e. $f(y) = f(z)$. Since f is one-to-one, $y = z$. So, f^{-1} has no pairs $< x, y >, < x, z >$ with $y \neq z$. But this means that f^{-1} is a function. **Q.E.D.**

This proposition tells us not only that the inverse of a one-to-one function is a function, but it is also a one-to-one function. Let us see why. Let f be a one-to-one function. By Proposition 1.3.13, f^{-1} is a function. Now, the inverse of f^{-1} is f, a function. So, the inverse of the function f^{-1} is a function. By the same proposition, f^{-1} is one-to-one.

The composition of functions preserves injectiveness.

Proposition 1.3.14 *If f, g are one-to-one, then $g \circ f$ is one-to-one.*

Proof: Let x and y be two elements of the domain of $(g \circ f)$ and let us assume that $(g \circ f)(x) = (g \circ f)(y)$. We use Proposition 1.3.8 to compute the two sides and get (1).

(1) $g(f(x)) = g(f(y))$.

By the same proposition, both $f(x)$ and $f(y)$ are defined, i.e. they belong to the domain of g. Since g is one-to-one, (1) reduces to (2).

(2) $f(x) = f(y)$

Proposition 1.3.8 tells us that f is defined at x and at y, i.e. they are in the domain of f. Since f is one-to-one, (2) reduces to (3).

(3) $x = y$.

So, $(g \circ f)(x) = (g \circ f)(y)$ implies that $x = y$, i.e. $(g \circ f)$ is one-to-one. **Q.E.D.**

From the definition of composition and Proposition 1.3.8 we know that the composition of $f : A \longrightarrow B$ and $g : B \longrightarrow C$ is a function $g \circ f : A \longrightarrow C$. But this does not tell us much about the domain and the range of $g \circ f$. The next proposition gives a more exact characterization.

Proposition 1.3.15 *Let f be a function on A and onto B and g be a function on B. Then $g \circ f$ is a function on A that has the same range as g.*

Proof: We will show first that $dom(g \circ f) = A$. Let $x \in A$. Since $dom(f) = A$, f is defined at x, and $f(x) \in B$. Since $f(x) \in B = dom(g)$, g is defined at $f(x)$. By Proposition 1.3.8 $g \circ f$ is defined at x, i.e. $x \in dom(g \circ f)$. So, $A \subseteq dom(g \circ f)$. Since the domain of $g \circ f$ is a subset of the domain of f, $dom(g \circ f) = A$.

Now let us show that $ran(g \circ f) = ran(g)$. From the definition of composition we know that $ran(g \circ f) \subseteq ran(g)$, so we need to show that $ran(g) \subseteq ran(g \circ f)$. So, let $z \in ran(g)$. Then there is some $y \in dom(g) = B$ such that $g(y) = z$. Since $y \in B = ran(f)$, there is some $x \in dom(f)$ such that $y = f(x)$. Since both $y = f(x)$ and $z = g(f(x))$ exists, $(g \circ f)(x)$ exists and is equal to $g(f(x)) = z$. So, $z \in ran(g \circ f)$. **Q.E.D.**

Now we will talk about a particular type of functions, called operations.

Definition 1.3.16 (n-ary operations) *An $n-ary$ operation on the set A is a function $f : A^n \longrightarrow A$, A^n being the n-th power of A.*

If $n = 0$, the operation f is an element of A; if $n = 1$, the operation is a function $f : A \longrightarrow A$; if $n = 2$ the operation is a function $f : A \times A \longrightarrow A$, etc.

The number n is called the *arity* of the operation. If $n = 0$ the operation is called *zero-ary*, if $n = 1$ the operation is called *unary*, if $n = 2$ the operation is *binary*, and so on.

Now we will describe some properties of the binary operations. For these operations we will use the *infix* notation, i.e. we will write (xfy) instead of f(x,y). Frequently we will drop the outer parentheses.

Definition 1.3.17 (associativity, commutativity, idempotency) *Let $*$ be a binary operation on A.*

1. $*$ is associative if for all $x, y, z \in A$, $x * (y * z) = (x * y) * z$.

2. $*$ is commutative if for all $x, y \in A$, $x * y = y * x$.

3. $*$ is idempotent if for all $x \in A$, $x * x = x$.

Since $*$ may not be define for all pairs of A^2, we need to clarify these equations. The meaning is that whenever one side exists, the other side exists and they are equal.

Examples 1.3.18 1. The addition of real numbers is associative and commutative. It is not idempotent because $1 + 1 \neq 1$.

2. Let A be a set. The operation $\cap$ is associative, commutative, and idempotent on $\mathcal{P}[A]$, the power set of A.

3. We define the / operation as being the division operation defined on the set on non-zero real numbers.

Then / is not associative because $(16/8)/2 = 1$ while $16/(8/2) = 4$.

The operation is not commutative because $10/5 \neq 5/10$.

It is not not idempotent because $2/2 \neq 2$.

Definition 1.3.19 (distributivity) *Let $\circ$ and $*$ be two operations over A. We say that $*$ distributes over $\circ$ if for all $x, y, z \in A$,*

(1) $(x \circ y) * z = (x * z) \circ (y * z)$, and

(2) $z * (x \circ y) = (z * x) \circ (z * y)$.

The property (1) is called left distributivity and (2) is called right distributivity.

Examples 1.3.20 1. The multiplication of real numbers distributes over the addition of real numbers because

$(x + y)z = (xz) + (yz)$, and

$z(x + y) = (zx) + (zy)$.

2. Let $\cup$ and $\cap$ be the union, respectively the intersection of two subsets of the set A. Then $\cup$ distributes over $\cap$ because

$(x \cap y) \cup z = (x \cup z) \cap (y \cup z)$, and

$z \cup (x \cap y) = (z \cup x) \cap (z \cup y)$.

Note 1.3.21 In our book the notation $f : A \longrightarrow B$ will denote a function with domain A, unless we say otherwise.

Exercises

Exercise 1.3.1 *Let f, g be two relations such that $ran(f) \subseteq dom(g)$. Show that $dom(g \circ f) = dom(f)$.*

Exercise 1.3.2 *Let f, g be two relations such that $dom(g) \subseteq ran(f)$. Show that $ran(g \circ f) = ran(g)$.*

Exercise 1.3.3 *Use the two preceding exercises to prove Proposition 1.3.15.*

Exercise 1.3.4 *How many functions $f : \phi \longrightarrow A$ can be defined? What are their ranges? Are any of them one-to-one?*

Exercise 1.3.5 *How many functions $f : A \longrightarrow \phi$ can defined? Is any of them one-to-one?*

Exercise 1.3.6 *Let $A = \{a\}$. How many functions $f : A \longrightarrow B$ can be defined? Is any of then one-to-one?*

Exercise 1.3.7 *Let f be a function with domain A, and let $1_A = \{< a, a > |a \in A\}$. Show that f is one-to-one iff there is a function g such that $g \circ f = 1_A$.*

Exercise 1.3.8 *Let f be a function with range B. Show that there is a function g such that $f \circ g = 1_B$, where $1_B = \{< b, b > |b \in B\}$.*

Exercise 1.3.9 *Show that every function f can be written as composition $f = h \circ g$ where g is one-to-one.*

Exercise 1.3.10 *Let f be a one-to-one function and g, h be two functions whose ranges are subsets of the domain of f. Show that $f \circ g = f \circ h$ implies $g = h$.*

Exercise 1.3.11 *Let g, h be two functions whose domains are subsets of the range of the function f. Show that $g \circ f = h \circ f$ implies $g = h$.*

Exercise 1.3.12 *Let $\mathcal{P}[A]$ be the power set of A. On $\mathcal{P}[A]$ we define the binary operations $\odot$ and $\diamond$ by $S \odot T = \{x \in A | x \ /S \text{ and } x \ /inT\}$ and $S \diamond T = (S - T) \cup (T - S)$. Check whether these operations are idempotent, associative, or commutative.*

1.4 Congruences

This section is a natural continuation of the relations and functions sections. It begins by defining the concepts of equivalence and partition and then it shows that every equivalence defines a partition and vice-versa.

Definition 1.4.1 (equivalence relation) *Let A be a set. An equivalence relation on A is a relation that is reflexive, symmetric and transitive.*

Examples 1.4.2 1. Let $\equiv_3$ be the relation defined on the set of integers by the relation $m \equiv_3 n$ if 3 divides $m - n$. This relation is reflexive, symmetric, and transitive, so it is an equivalences. It is called *the conguence modulo 3*.

2. Let $A = \{a, b\}^*$ be the set of all strings that can be formed with the letters a and b. Let $\equiv$ be the relation on A defined by

$s \equiv t$ if s and t have the same length.

Again, $\equiv$ is reflexive, symmetric and transitive, hence an equivalence.

Definition 1.4.3 (Partition) *Let A be a non-empty set. A partition of A is a set Π, pronounced big pie, such that*

1. every element $x \in \Pi$ is non-empty,

2. if $x, y \in P_i$ and $x \neq y$, then $x \cap y = \phi$, and

3. $\cup\Pi = A$.

The elements of A are called the blocks of the partition.

Observation 1.4.4 The first condition of Definition 1.4.3 tells us that all blocks are non-empty. Conditions 2 and 3 state that every element of A belongs to one and only one block of the partition.

Examples 1.4.5 1. Let $A = \{1, 2, 3, 4, 5, 6, 7\}$ and $\Pi = \{\{1, 2, 3\}, \{4\}, \{5, 6, 7\}\}$. Then Π is a partition because every block is non-empty and every element of A belongs to one and only one block.

2. Let Z be the set of integers and $\Pi = \{\{n \in Z | z \equiv_3 0\}, \{n \in Z | z \equiv_3 1\}, \{n \in Z | z \equiv_3 0\}\}$, i.e. the blocks of Π are the sets of integers that give remainders 0, 1, and 2, when divided by 3. Again, the blocks of Π are non-empty and each integer belongs to one and only one block.

Definition 1.4.6 (equivalence classes, equivalence index) *Let A be a non-empty set and let R be an equivalence relation on A. Let $a \epsilon A$. The equivalence class of a is the set $[a]_R = \{b \in A | aRb\}$.*

The set of equivalence classes is called the index of R.

Proposition 1.4.7 *Let A be a non-empty set and let R be an equivalence relation on A. Then the set $\pi(R) = \{[a]_R | a \in A\}$ is a partition of A.*

Proof: We need to show that

1. every block is non-empty,
2. the intersection of two distinct blocks is empty, and
3. every element of A belongs to a block.

The conditions 1 and 3 are easy to prove. We know that $a \in [a]_R$ because R is reflexive. So, every block is non-empty, and every element $a \in A$ belongs to one block, namely $[a]_R$.

Now we prove 2, $[a]_R \neq [b]_R$ implies $[a]_R \cap [b]_R = \phi$.

We will show the contrary of 2, i.e. that $[a]_R \cap [b]_R \neq \phi$ implies $[a]_R = [b]_R$.

So, let us assume that
(1) $c \in [a]_R \cap [b]_R$.
We will show that $[a]_R = [b]_R$ by proving the two inclusions, $[a]_R \subseteq [b]_R$ and $[b]_R \subseteq [b]_R$.
(1) tells us that $c \in [a]_R$, so
(2) aRc
from the definition of $[a]_R$. At the same time (1) tells us that
(3) bRc
Now let $x \in [a]_R$. From the definition of the equivalence classes,
(4) aRx
Since R is symmetric, (2) produces
(5) cRa.
Now we apply the transitivity of R to (3) and (5) to get
(6) bRa.
again we apply the transitivity of R to (6) and (4) and get
(7) bRx,
that tells us that $x \in [b]_R$. So, every element of $[a]_R$ is in $[b]_R$, i.e. $[a]_R \subseteq [b]_R$. In a similar way, we prove that $[b]_R \subseteq [b]_R$. From $[a]_R \subseteq [b]_R$ and $[b]_R \subseteq [b]_R$ we conclude that $[a]_R = [b]_R$. **Q.E.D.**

Observation 1.4.8 The proof of Proposition 1.4.7 tells us that $[a]_R = [b]_R$ iff aRb.

Now we can also show that every partition Π of a non-empty set A determines an equivalence on A.

Proposition 1.4.9 *Let A be a non-empty set and Π be a partition of A. We define the relation $\eta(\Pi)$, pronounced a-ta of pie , on A by $a\eta(\Pi)b$ if a and b belong to the same block of Π. Then $\eta(\Pi)$ is an equivalence relation.*

Proof: We need to show that $\eta(\Pi)$ is reflexive, symmetric and transitive.
1. $\eta(\Pi)$ is reflexive, because for all $a \in A$, a belongs to a block of Π and the pair a, a belongs to the same block.
2. $\eta(\Pi)$ is symmetric because if a and b belong to the same block, then of course b and a belong to the same block.
3. $\eta(\Pi)$ is transitive.

Assume that $a\eta(\Pi)b$ and $b\eta(\Pi)c$. From the definition of $\eta(\Pi)$, a and b belong to a block, B_1 of Π. Since $b\eta(\Pi)c$, b and c belong to a block, B_2, of Π. But b belongs to a single block, so $B_1 = B_2$. Then a and c belong to the same block, i.e. $a\eta(\Pi)c$. **Q.E.D.**

So, far we gave two functions, π that maps each equivalence R of A into the partition $\pi(R)$, and η which maps every partition of A into the equivalence $\eta(\Pi)$. Now we will show that $(\pi \circ \eta)(\Pi) = \Pi$ and $(\eta \circ \pi)(R) = R$.

Proposition 1.4.10 *Let A be a non-empty set. Then*
for all partitions Π of A, $(\pi \circ \eta)(\Pi) = \Pi$, and
for all equivalences R of A, $(\eta \circ \pi)(R) = R$.

Proof:
First, we will show that $(\pi \circ \eta)(\Pi) = \Pi$.

We will show that the blocks of the partitions of Π and $\pi(\eta(\Pi))$ are the same. The blocks of $\pi(\eta(\Pi))$ are the sets $[a]_{\eta(\Pi)}$, a being an element of A.

We will show that for every $b \epsilon A$, the Π block that contains b is the same as the $\pi(\eta(\Pi))$ block that has b.

$x \in [b]_{\eta(\Pi)}$ the $\pi(\eta(\Pi))$ block that has b
iff $b\eta(\Pi)x$ from the definition of the equivalence classes
iff b and x belong to the same block of Π from the construction of η
iff x belongs to the Π block of b.

Now, we can show that $(\eta \circ \pi)(R) = R$. Let a, b be two elements of A. Then,
$a\eta(\pi(R))b$
iff a, b belong to the same block of $\pi(R)$ from the construction of η
iff $b \in [a]_R$ from the construction of π
iff aRb from the construction of the equivalence classes of R.
Q.E.D.

Proposition 1.4.11 *Let A be a non-empty set. Let $E(A)$ and $\Pi(A)$ be the set of all equivalences, respectively all partitions, that can be defined on A. Then the functions $\pi : E(A) \longrightarrow \Pi(A)$ and $\eta : \Pi(E) \longrightarrow E(A)$ are on, onto and injective.*

Proof: The mapping π is defined for all equivalences and η assigns an equivalence to every partition. So, both functions are on.

We will show that π is one-to-one and onto and we will leave the proof that η is onto to the reader.

Let R, S, be two equivalences on A. We will show that
(1) $\pi(R) = \pi(S)$ implies that $R = S$.
Let us assume that
(2) $\pi(R) = \pi(S)$.
(2) implies that
(3) $\eta(\pi(R)) = \eta(\pi(S))$.
By Proposition 10, $\eta(\pi(R)) = R$, and $\eta(\pi(S)) = S$. In (3) we replace $\eta(\pi(R))$ by R and $\eta(\pi(S))$ by S and we get
(4) $R = S$.
Q.E.D.

Definition 1.4.12 (the finer than relation) *Let R and S be two equivalences on a non-empty set A. We say that R is finer than S if $R \subseteq S$.*

Observation 1.4.13 If R is finer S, then each block of $\pi(R)$ is a non-empty subset of a block of $\pi(S)$.

Example 1.4.14 Let $A = \{1, 2, 3\}$. We can verify that $R = \{< 1, 1 >, < 2, 2 > , < 3, 3 >\}$ and $S = \{< 1, 1 >, < 1, 2 >, < 2, 1 >, < 2, 2 >, < 3, 3 >\}$ are two equivalences on A, and that $R \subseteq S$.

Then $\pi(R) = \{\{1\},\{2\},\{3\}\}$, and $\pi(S) = \{\{1,2\},\{3\}\}$. The first block of $\pi(R)$ is included in the first block of $\pi(S)$, the second block of $\pi(R)$ is included in the first block of $\pi(S)$, and the third block of $\pi(R)$ is included in the second block of $\pi(S)$.

It is easy to show that the identity on A, $1_A = \{< a, a > | a \in A\}$, and the cartesian product $A \times A$ are equivalences on A. We can also show that for all equivalences R on A, $1_A \subseteq R \subseteq A \times A$. So, 1_A is the finest equivalence on A, and $A \times A$ is the coarsest, i.e. the least fine. All blocks of the partition $\pi(1_A)$ have only one element, while $\pi(A \times A)$ has only one block, A.

Now let us look at the interaction of the equivalence relations with the operations of A.

Definition 1.4.15 (congruence) *Let $n > 0$ and $f : A^n \longrightarrow A$ be an operation[1] on A. Let E be an equivalence on A. We say that E is a congruence for f, if for all positions $1 \leq i \leq n$ and all elements $a_1, \ldots, a_{i-1}, b, c, a_{i+1}, \ldots, a_n$ in A,*

bEc implies that $f(a_1, \ldots, a_{i-1}, b, a_{i+1}, \ldots, a_n)Ef(a_1, \ldots, a_{i-1}, c, a_{i+1}, \ldots, a_n)$.

Observations 1.4.16 1. If the function f from Definition 1.4.15 has arity n, then there are n implications

bEc implies that $f(a_1, \ldots, a_{i-1}, b, a_{i+1}, \ldots, a_n)Ef(a_1, \ldots, a_{i-1}, c, a_{i+1}, \ldots, a_n)$

, one for each position i. For example, if $n = 3$, then we have the implications

bEc implies $f(b, a_1, a_2)Ef(c, a_1, a_2)$,

bEc implies $f(a_1, b, a_2)Ef(a_1, c, a_2)$, and

bEc implies $f(a_1, a_2, b)Ef(a_1, a_2, c)$.

2. Definition 1.4.15 also tells us that replacing an element by an element congruent to it preverses the congruence.

Example 1.4.17 Let $A = \{1,2,3\}$ and $E = \{< 1,1 >, < 1,2 >, < 2,1 >, < 2,2 >, < 3,3 >\}$. Now let $f : A \longrightarrow A$ be the function defined by

$f(1) = f(2) = 3$ and $f(3) = 1$.

Let us check that E is a congruence for f. We need to check that for all $x, y \in A$,

(*) xEy implies $f(x)Ef(y)$.

For $x = y$ condition (*) becomes

(**) xEx implies that $f(x)Ef(x)$.

This condition is satisfied since $f(x)Ef(x)$ is always true by the reflexivity of E.

So, all we need to check is that for all $x, y \in A$,

(†) ($x \neq y$ and xEy) imply $f(x)Ef(y)$.

Only two pairs, $< 1,2 >$ and $< 2,1 >$, satisfy the conditions of †. For $x = 1$ and $y = 2$ † becomes

(1) $1E2$ imply $f(1)Ef(2)$.

For the second pair, $x = 2$ and $y = 1$, † produces

[1] In this section all functions $g : X \longrightarrow Y$ have domain X.

(2) $2E1$ imply $f(2)EF(1)$.

Both (3) and (4) are satisfied because their consequent $3E3$ is true by the reflexivity of E.

Proposition 1.4.18 *Let E be a congruence for $f : A^n \longrightarrow A$. Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be two n-tuples of elements of A such that $x_1Ey_1, \ldots, x_nEy_n$. Then $f(x_1, \ldots, x_n)Ef(y_1, \ldots, y_n)$.*

Proof: We will prove by finite induction on k, $1 \leq k \leq n$, that

$f(x_1, \ldots, x_{k-1}, x_k, x_{k+1}, \ldots, x_n)Ef(y_1, \ldots, y_{k-1}, y_k, x_{k+1}, \ldots, x_n)$.

We notice that for $k = n$ the above formula is exactly the one that we want to prove.

Basis. $(k = 1)$

We need to show that

(1) $f(x_1, x_2, \ldots, x_n)Ef(y_1, x_2, \ldots, x_n)$.

Since x_1Ey_1 and E is congruent for f we can replace x_1 by y_1 in $f(x_1, x_2, \ldots, x_n)$ and get formula (1).

Inductive step

Assume that

(IH) $f(x_1, \ldots, x_{k-1}, x_k, x_{k+1}, \ldots, x_n)Ef(y_1, \ldots, y_{k-1}, y_k, x_{k+1}, \ldots, x_n)$.

We need to show that

(2) $f(x_1, \ldots, x_{k-1}, x_k, x_{k+1}, x_{k+2}, \ldots, x_n)Ef(y_1, \ldots, y_{k-1}, y_k, y_{k+1}, x_{k+2}, \ldots, x_n)$.

Since E is a congruence for f and $x_{k+1}Ey_{k+1}$, we can replace x_{k+1} by y_{k+1} in $f(y_1, \ldots, y_k, x_{k+1}, x_{k+2}, \ldots, x_n)$ and obtain (3).

(3) $f(y_1, \ldots, y_{k-1}, y_k, x_{k+1}, x_{k+2}, \ldots, x_n)Ef(y_1, \ldots, y_{k-1}, y_k, y_{k+1}, x_{k+2}, \ldots, x_n)$.

Now we apply the transitivity of E to (IH) and (3) and obtain

$f(x_1, \ldots, x_{k-1}, x_k, x_{k+1}, x_{k+2}, \ldots, x_n)Ef(y_1, \ldots, y_{k-1}, y_k, y_{k+1}, x_{k+2}, \ldots, x_n)$,

which is exactly the formula that we wanted to prove. **Q.E.D.**

Now let us characterize congruences by using the partition associated with the congruence.

Proposition 1.4.19 *Let E be an equivalence relation on A, $\Pi = \pi(E)$ the partition corresponding to the equivalence E, and $f : A^n \longrightarrow A$ a function of arity $n > 0$.*

Then E is a congruence for f iff for all n-tuples $B_1, \ldots, B_n$ of blocks of Π there is a block $B \in \Pi$ such that $f(B_1, B_2, \ldots, B_n) \subseteq B$, where $f(B_1, \ldots, B_n) = \{f(x_1, \ldots, x_n) | x_1 \in B_1, \ldots, x_n \in B_n\}$.

Examples 1.4.20 Let us illustrate what Proposition 1.4.19 says by two examples.

1. Let $A = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ and let E be the equivalence that has the partition $\Pi = \{\{1, 2, 3, 4\}, \{5, 6, 7, 8\}, \{9, 10\}\}$. Let $f : A \times A \longrightarrow A$ be an operation on A of arity 2. Then E is a congruence for f iff the 9 conditions listed below are satisfied.

(1) there is a block B of Π such that $f(\{1, 2, 3, 4\}, \{1, 2, 3, 4\}) \subseteq B$

(2) there is a block B of Π such that $f(\{1, 2, 3, 4\}, \{5, 6, 7, 8\}) \subseteq B$

(3) there is a block B of Π such that $f(\{1, 2, 3, 4\}, \{8, 9\}) \subseteq B$

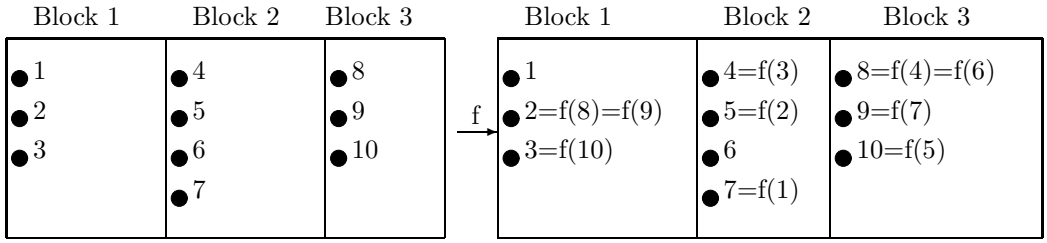


Figure 1.6: The block mapping of Example 1.4.20

(4) there is a block B of Π such that $f(\{5,6,7,8\},\{1,2,3,4\}) \subseteq B$
(5) there is a block B of Π such that $f(\{5,6,7,8\},\{5,6,7,8\}) \subseteq B$
(6) there is a block B of Π such that $f(\{5,6,7,8\},\{9,10\}) \subseteq B$
(7) there is a block B of Π such that $f(\{9,10\},\{1,2,3,4\}) \subseteq B$
(8) there is a block B of Π such that $f(\{9,10\},\{5,6,7,8\}) \subseteq B$
(9) there is a block B of Π such that $f(\{9,10\},\{9,10\}) \subseteq B$

So, if f has the arity n and Π has m blocks, then we must check m^n conditions because $f(B_1,\ldots,B_n)$ has n positions, and each position can be filled by any of the m blocks of Π.

2. Let $A = \{1,2,3,4,5,6,7,8,910\}$, $f : A \longrightarrow A$ be the function defined by $f(1) = 7, f(2) = 5, f(3) = 4, f(4) = 8, f(5) = 10, f(6) = 8, f(7) = 9, f(8) = 2, f(9) = 2, f(10) = 3$,

and E be the equivalence defined by the partition

$\Pi = \{\{1,2,3\},\{4,5,6,7\},\{8,9,10\}\}$.

We want to know if E is a congruence for f.

The blocks of E are Block 1 $= \{1,2,3\}$, Block 2 $= \{4,5,6,7\}$, and Block 3$= \{8,9,10\}$.

The proposition tells us that the image of each block must be included in a block, i.e.

f(Block 1), f(Block 2), and f(Block 3) must be included in one of the 3 blocks, Block 1, Block 2, or Block 3.

Figure 1.6 shows that f(Block 1) $= \{f(1), f(2), f(3)\} = \{7,5,4\}$ is included in Block 2, f(Block 2) $=\{f(4), f(5), f(6), f(7)\} = \{8,10,9\}$ is included in Block 3, and f(Block 3) $= \{f(8), f(9), f(10)\} = \{2,3\}$ is included in Block 1.

So, E is a congruence for f.

Now let us return to the proof of Proposition 1.4.19.

Proof: We need to show two things.

(1) If E is congruent for f then for all n-tuples of blocks $B_1,\ldots, B_n$ of $\pi(E)$ there is a block B of $\pi(E)$ such that $f(B_1,\ldots,B_n) \subseteq B$, and

(2) if for all n-tuples of blocks $B_1,\ldots, B_n$ of $\pi(E)$ there is a block B of $\pi(E)$ such that $f(B_1,\ldots,B_n) \subseteq B$, the E is congruent for f.

$\Longrightarrow$ We will prove (1) first. Assume that E is congruent for f and that $< B_1, \ldots, B_n >$ is an n-tuple of blocks of $\pi(E)$. We need to show that there is a block B such that $f(B_1, \ldots, B_n) \subseteq B$.

Since the blocks are non-empty, we take a value from each block, $x_1 \in B_1$,..., $x_n \in B_n$. Let B be the block that contains $f(x_1, \ldots, x_n)$, i.e. B be the block that contains $f(x_1, x_2, \ldots, x_n)$. The condition that

(3) $f(B_1, \ldots, B_n) \subseteq B$

is equivalent to the statement

(4) for all $y_1 \in B_1, y_2 \in B_2, \ldots, y_n \in B_n$, $f(y_1, y_2, \ldots, y_n) \in B$.

From the construction of $\pi(E)$ we know that $B_1 = [x_1]_E$, $B_2 = [x_2]_E$,..., $B_n = [x_n]$, $B = [f(x_1, \ldots, x_n)]$. So, condition (4) is equivalent to

(5) for all $y_1 \in [x_1]_E, y_2 \in [x_2]_E, \ldots, y_n \in [x_n]_E$, $f(y_1, y_2, \ldots, y_n) \in [f(x_1, x_2, \ldots, x_n)]_E$.

Now we use the fact that $y \in [x]_E$ iff xEy. So, (5) is equivalent to

(6) for all y_1,..., y_n with $x_1Ey_1, x_2Ey_2, \ldots, x_nEy_n$, $f(x_1, x_2, \ldots, x_n)Ef(y_1, y_2, \ldots, y_n)]$.

The last formula is true by Proposition 1.4.18.

$\Longleftarrow$ Now let us assume that

(7) for all n-tuples of blocks B_1,..., B_n of $\pi(E)$ there is a block B of $\pi(E)$ such that $f(B_1, \ldots, B_n) \subseteq B$,

and we will show that E is a congruence on f.

We will have to prove that the equivalences

(8) $f(a_1, \ldots, a_{k-1}, u, a_k, \ldots, a_{n-1})Ef(a_1, \ldots, a_{k-1}, v, a_k, \ldots, a_{n-1})$

hold for all elements a_1,...,a_{n-1}, u, v in A that satisfy uEv.

Let a_1,...,a_{n-1}, u, v be $n+1$ elements, not necessarily distinct, that satisfy uEv. Let B_1 be the block that contains a_1, B_2 the block that contains a_2, ..., B_k the block that contains u, B_{k+1} be the block that contains a_k,..., B_n be the block that contains a_{n-1} and B be the block that contains $f(a_1, \ldots, a_{k-1}, u, a_k, \ldots, a_{n-1})$. Now let us remember that two elements w, z of A belong to the same block iff wEz.

Since uEv and $u \in B_k$, $v \in B_k$. By condition (7),

(9) $f(a_1, \ldots, a_{k-1}, v, a_k, \ldots, a_{n-1}) \in B$

Since $(a_1, \ldots, a_{k-1}, v, a_k, \ldots, a_{n-1})$ is also in B, we get

$f(a_1, \ldots, a_{k-1}, u, a_k, \ldots, a_{n-1})Ef(a_1, \ldots, a_{k-1}, v, a_k, \ldots, a_{n-1})$

from the way we construction of the blocks of $\pi(E)$. **Q.E.D.**

Example 1.4.21 Let us apply Proposition 1.4.19 to check whether the equivalences E_1 and E_2 are congruences for f.

Let A be the set $\{1, 2, 3, 4, 5\}$, E_1 and E_2 be the equivalences that correspond to the partitions

$\Pi_1 = \{\{1, 2\}, \{3\}, \{3, 4\}\}$ respectively $\Pi_2 = \{\{1, 2, 4\}, \{3, 5\}\}$,

and $f : A \times A \longrightarrow A$ be the function below.

$f(1,1) = 5, f(1,2) = 4, f(1,3) = 2, f(1,4) = 2, f(1,5) = 1$
$f(2,1) = 4, f(2,2) = 5, f(2,3) = 1, f(2,4) = 2, f(2,5) = 2$
$f(3,1) = 4, f(3,2) = 5, f(3,3) = 5, f(3,4) = 1, f(3,5) = 1$
$f(4,1) = 5, f(4,2) = 4, f(4,3) = 3, f(4,4) = 1, f(4,5) = 2$
$f(5,1) = 4, f(5,2) = 5, f(5,3) = 3, f(5,4) = 2, f(5,5) = 1.$

We want to check if the equivalences E_1 and E_2 are congruences for f. We apply Proposition 1.4.19.

Let us start with E_1. E_1 has 3 blocks, Block1 = $\{1,2\}$, Block2 = $\{3\}$, and Block3=$\{4,5\}$.

We need to check that each of the sets, f(Block1, Block1), f(Block1, Block2), f(Block1, Block3), f(Block2, Block1), f(Block2, Block2), f(Block2, Block3), f(Block3, Block1), f(Block3, Block2), f(Block3, Block3) is included in one of the 3 blocks.

Let us check these conditions.

f(Block1, Block1) = $\{f(1,1), f(1,2), f(2,1), f(2,2)\} = \{5,4,4,5\} = \{4,5\} \subseteq$ Block3.

f(Block1, Block2) = $\{f(1,3), f(2,3)\} = \{2,1\} = \{1,2\} \subseteq$ Block1.

f(Block1, Block3) = $\{f(1,4), f(1,5), f(2,4), f(2,5)\} = \{2,1,1,2\} = \{1,2\} \subseteq$ Block1.

f(Block2, Block1) = $\{f(3,1), f(3,2)\} = \{4,5\} \subseteq$ Block3.

f(Block2, Block2) = $\{f(3,3)\} = \{5\} \subseteq$ Block3.

f(Block2, Block3) = $\{f(3,4), f(3,5)\} = \{1,1\} = \{1\} \subseteq$ Block1.

f(Block3, Block1) = $\{f(4,1), f(4,2), f(5,1), f(5,2)\} = \{5,4,4,5\} = \{4,5\} \subseteq$ Block3.

f(Block3, Block2) = $\{f(4,3), f(5,3)\} = \{3,3\} = \{3\} \subseteq$ Block2.

f(Block3, Block3) = $\{f(4,4), f(4,5), f(5,4), f(5,5)\} = \{1,2,2,1\} = \{1,2\} \subseteq$ Block1.

Since E satisfies the conditions of Proposition 1.4.19, E is a congruence for f.

Now, let us continue with E_2. This equivalence has 2 blocks, Block1=$\{1,2,4\}$ and Block2=$\{3,5\}$.

We need to check that each of the sets f(Block1, Block1), f(Block1, Block2), f(Block2, Block1), and f(Block2, Block2) are included in one of the two blocks. Let us check the conditions.

f(Block1, Block1) = $\{f(1,1), f(1,2), f(1,4), f(2,1), f(2,2), f(2,4), f(4,1),$

$f(4,2), f(4,4)\} = \{5,4,2,4,5,2,5,4,1\} = \{1,2,4,5\}$ and this set is **not** a subset of any block. So, E_2 fails the conditions of Proposition 1.4.19, and E_2 is not a congruence for f.

Proposition 1.4.22 *Let f be an operation of arity $n > 0$ on A and E be a congruence for f. Let $A/E = \{[a]_E | a \in A\}$ be the set of equivalence classes of E. Then the relation $f^E([a_1]_E, \ldots, [a_n]_E) = [f(a_1, \ldots, a_n)]_E$ is a function from $(A/E)^n$ to A/E.*

Example 1.4.23 The equivalence E_1 of Example 1.4.21 is a congruence for f. The classes of E_1 are the three blocks, Block1 = $\{1,2\}$, Block2 = $\{3\}$ and Block3 = $\{4,5\}$. The 9 conditions of the verification step define the function $f^{E_1} : \{\{1,2\},\{3\},\{4,5\}\}^2 \longrightarrow \{\{1,2\},\{3\},\{4,5\}\}$

as f^{E_1}(Block1, Block2) is the block that includes f(Block1, Block2).

The function is given below.

$f^{E_1}(\{1,2\},\{1,2\}) = \{4,5\}$,

$f^{E_1}(\{1,2\},\{3\}) = \{1,2\}$,
$f^{E_1}(\{1,2\},\{4,5\}) = \{1,2\}$,
$f^{E_1}(\{3\},\{1,2\}) = \{4,5\}$,
$f^{E_1}(\{3\},\{3\}) = \{4,5\}$,
$f^{E_1}(\{3\},\{4,5\}) = \{1,2\}$,
$f^{E_1}(\{4,5\},\{1,2\}) = \{4,5\}$,
$f^{E_1}(\{4,5\},\{3\}) = \{3\}$,
$f^{E_1}(\{4,5\},\{4,5\}) = \{1,2\}$.

Now let us return to the proof of Proposition 1.4.22.

Proof: The relation $f^E([a_1]_E,\ldots,[a_n]_E) = [f(a_1,\ldots,a_n)]_E$ is a function from $(A/E)^n$ to A/E when the value of $f^E([a_1]_E,\ldots,[a_n]_E) = [f(a_1,\ldots,a_n)]_E$ does not depend on the choice of the representatives, $a_1,\ldots, a_n$. So, we have to show that whenever

$b_1Ea_1,\ldots, b_nEa_n$ we have $f(a_1,\ldots,a_n)Ef(b_1,\ldots,b_n)$.

Since E is an f congruence, the above condition is true by Proposition 1.4.18 **Q.E.D.**

1.4.1 Exercises

Exercise 1.4.1 *Give a counter-exemple to show that the union does not preserve equivalences. i.e. construct a set A and two equivalences R and S on A such that $R \cup S$ is not an equivalence.*

Exercise 1.4.2 *Show that if R and S are equivalences, so is $R \cap S$.*

Exercise 1.4.3 *Let $E(A)$ be the set of equivalences that can be defined on the set A. Show that the intersection of* **any** *set of equivalences of $E(A)$ is an equivalence.*

Exercise 1.4.4 *Let E_1 and E_2 be two equivalences in $E(A)$. Show that there is an equivalence relation E such that*

1. $E_1 \subseteq E$ and $E_2 \subseteq E$, and

2. if $\equiv$ is any other equivalence of $E(A)$ such that $E_1 \subseteq \equiv$ and $E_2 \subseteq \equiv$ then $E \subseteq \equiv$.

The equivalence E is called the least upper bound of E_1 and E_2.

How do you go about constructing the least upper bound of two equivalences?

Exercise 1.4.5 *Complete the proof of Proposition 1.4.7 i.e. show that if $c \in [a]_R \cap [b]_R$ then $[b]_R \subseteq [a]_R$.*

Exercise 1.4.6 *Show that $\pi \circ \eta$ and $\eta \circ \pi$ are one-to-one.*

Exercise 1.4.7 *Let A be a non-empty set. Show that finer than is is an equivalence relation on $E(A)$.*

Exercise 1.4.8 *Show that 1_A and $A \times A$ are equivalences on A.*

Exercise 1.4.9 *Show that 1_A and $A \times A$ are congruences for any operations on A of arity greater than 0.*

Exercise 1.4.10 *Let f be an operation on A of arity greater than 0, and let $E_f(A)$ be the set of f congruences that can be defined on A. Show that the intersection of* **any** *congruences of $E_f(A)$ is also a congruence.*

Exercise 1.4.11 *Show that any two congruences of $E_f(A)$ have a least upper bound.*

Exercise 1.4.12 *Let $A = \{1, 2, 3, 4, 5\}$ and $f : A \times A \longrightarrow A$ be the function defined below.*

$f(1,1) = 3, f(1,2) = 4, f(1,3) = 1, f(1,4) = 2, f(1,5) = 5$
$f(2,1) = 4, f(2,2) = 3, f(2,3) = 2, f(2,4) = 2, f(2,5) = 5$
$f(3,1) = 2, f(3,2) = 1, f(3,3) = 5, f(3,4) = 5, f(3,5) = 2$
$f(4,1) = 1, f(4,2) = 2, f(4,3) = 5, f(4,4) = 5, f(4,5) = 1$
$f(5,1) = 3, f(5,2) = 4, f(5,3) = 4, f(5,4) = 3, f(5,5) = 2.$

Which one of the equivalences below are are f congruences?

1. $\Pi_1 = \{\{1,2\},\{3,4,5\}\}$
2. $\Pi_1 = \{\{1\},\{2\},\{3,4,5\}\}$
3. $\Pi_1 = \{\{1,2\},\{3,4\},\{5\}\}$

Exercise 1.4.13 *Use Proposition 1.4.19 to show that the partitions $\Pi_3 = \{\{1,2,3\},\{4,5\}\}$ and $\Pi_3 = \{\{1,2\},\{3,4,5\}\}$ are congruences for the function given in Example 1.4.21.*

1.5 Cardinals

Our focus is on computability, so we investigate only countable sets. The properties of these sets are derived from the features of N, the set of natural numbers. N was defined in Section 1.1, as the smallest inductive set. So, we will use the axioms of the set theory to develop the arithmetic of the natural numbers.

Proposition 1.5.1 *For all $n \in N$, $n = 0$ or there is some $m \in N$ such that $n = S(m)$.*

Proof: By the induction principle.

Basis: $n = 0$, satisfies the proposition.

Inductive Step: Assume that $n \in N$. Then $S(n)$ also satisfies the conditions of the proposition. **Q.E.D.**

We recall that $n = \{0, \ldots, n-1\}$. Then we can define the $>$ relation on N.

Definition 1.5.2 *On N we define the relation $>$ as $m > n$ iff $n \in m$.*

Proposition 1.5.3 *$>$ is a strict order on N.*

Proof: We need to show that $>$ is irreflexive and transitive.
Irreflexivity: $>$ is irreflexive by the axiom of foundation.
Transivity: We show, by induction on n, that $n > m$ and $m > k$ implies $n > k$.

Basis: $n = 0$. Since 0 is the empty set, $0 > m$ is false, so the implication is vacuously true.

Inductive Step: Assume that for all m, k, $n > m$ and $m > k$ imply $n > k$. Let us assume that $S(n) > m$ and $m > k$. Since $S(n) = n \cup \{n\}$, at least one of the conditions $m = n$, $n > m$ must be true.

If $m = n$, then $m > k$ means that $k \in n$, so $k \in S(n)$. Then $S(n) > k$ from the definition of $>$.

If $n > m$, then $n > k$ by the induction hypothesis. So, $k \in S(n)$ and $S(n) > k$. **Q.E.D.**

The next two propositions are needed to prove Proposition 1.5.6.

Proposition 1.5.4 *$n = 0$ or $n > 0$.*

Proof: 1. The proof is by induction on n.

Basis: For $n = 0$ the proposition is true.

Inductive step: Assume that the proposition is true for n.

If $n = 0$, then $0 \in S(0)$, so $S(0) > 0$.

If $n > 0$, then $S(n) > n$ implies $S(n) > 0$ by transitivity.

In both cases $S(n) > 0$. **Q.E.D.**

Proposition 1.5.5 *$m > n$ implies $m = S(n)$ or $m > S(n)$.*

Proof: By induction on n.

Basis: $m = 0$. Then the proposition is vacuously true.

Inductive Step: Assume that it is true for all elements of m. Let $n \in S(m)$. Then either $n \in m$ or $n = m$. In the first case, $S(m) > S(n)$ by the induction hypothesis. In the second, $S(m) = S(n)$.**Q.E.D.**

Proposition 1.5.6 *$>$ is a total order on N.*

Proof: First of all, we can show that only one of the relations $m > n$,$m = n$,$n > m$ hold. If $m > n$ and $m = n$ are both true we get $m > m$. If $m = n$ and $n > m$ again we get $m > m$. Finally, if $m > n$ and $n > m$ hold, we obtain $m > m$ by transitivity. In all 3 cases we contradict the irreflexivity of $>$. We show that $m > n$ or $m = n$ or $n > m$ holds by induction on n.

Basis: $n = 0$. $m = 0$ or $m > 0$ is true by Proposition 1.5.4.

Inductive Step: Assume that $m > n$ or $m = n$ or $n > m$ holds for n.

If $m = n$ or $n > m$ is true, then $S(n) > m$, so $m > S(n)$ or $m = S(n)$ or $S(n) > m$ is true.

If $m > n$ is true, we get $m = S(n)$ or $m > S(n)$ by Proposition 1.5.5. Again, $m > S(n)$ or $m = S(n)$ or $S(n) > m$ holds. **Q.E.D.**

Corollary 1.5.7 *$(N, >)$ is a well order.*

Proof: The relation $>$ on N is exactly $\in_N$. By the axiom of foundation, $>$ is a well founded order. Since $>$ is total, $>$ is a well order, by Proposition 1.2.39. **Q.E.D.**

Next we prove The Recursion Theorem, that allows us to define the addition and the multiplication. We will use the more familiar notation $m+1$ for the successor $S(m)$.

Theorem 1.5.8 (The Recursion Theorem) *Let A be a set, a an element of A, and $g : A \times N \longrightarrow A$ a function on $A \times N$. Then there is a unique function $f : N \longrightarrow A$ on N such that*

1. *$f(0) = a$, and*
2. *for all $n \in N$, $f(n+1) = g(f(n), n)$.*

Before we go on with the proof, let us look at the function f.

$f(0) = a$,
$f(1) = g(f(0), 0) = g(a, 0)$,
$f(2) = g(f(1), 1) = g(g(a, 0), 1)$,
$f(3) = g(f(2), 2) = g(g(g(a, 0), 1), 2)$,
and so on. So, the value of $f(n+1)$ depends on the value of $f(n)$ and n.

We have to show that such a function exists and it is unique. For the persons who wrote recursive programs, the existence of f seems so obvious that it is not worth any further discussion. However, we must work within the framework of set theory, and prove that its existence follows from these postulates. The reader who is not interested in these details, can skip this discussion and go on to cardinals.

Proof: We will start by defining *the computations of length m.*

($\top$) A computation of length m is a function s with domain $m+1$ and range included in A, that has the property that $< 0, a >\in s$, and for all $k \in m$, $s(k+1) = g(s(k), k)$.

We can rewrite this definition as a property $\mathbf{P}[t, m+1]$ constructed according to the rules of Section 1.1. We leave this task as exercise for the interested reader. Now let us define $f = \cup\{s \in \mathcal{P}[N \times A] |$ there is some $m \in N$ such that s is a computation of length $m\}$.

Lemma 1.5.9 *f is function.*

Proof: Let $S = \{s \in \mathcal{P}[N \times A] |$ there is some $m \in N$ such that s is a computation of length $m\}$. The elements of every $s \in S$ are ordered pairs, so f is also a set of ordered pairs. It remains only to show that f does not contain pairs $< k, y >, < k, z >$ with $y \neq z$. So, let us assume that $< k, y >, < k, z >\in f$. Then there are two computations, s of length m and t of length n, such that $< k, y >\in s$ and $< k, z >\in t$. By Proposition 1.5.6, $m+1 \subseteq n+1$ or $n+1 \subseteq m+1$. Let us assume, without loss of generality, that $m+1 \subseteq n+1$. We will show, by finite induction, that for all $i \in m+1$, $s(i) = t(i)$.

If $i = 0$, then $s(0) = a = t(0)$.

Assume that $s(i) = t(i)$ and $i \in m$. Then $i+1 \in m+1$, so both the function s and the function t are defined at $i+1$. $s(i+1) = g(s(i), i+1)$ and $t(i+1) = g(t(i), i+1)$. Since $s(i) = t(i)$, $s(i+1) = t(i+1)$.

So, by finite induction, we conclude that $s(i) = t(i)$ for all $i \in m+1$. Then $y = s(k) = t(k) = z$. **Q.E.D. Lemma 1.5.9**

Next, we have to show that the function f has domain N and range included in A.

Lemma 1.5.10 *$dom(f) = N$ and $ran(f) \subseteq A$.*

Proof: Every computation s has as domain a natural number and range a subset of A, so the domain of f is a subset of N and its range a subset of A. We will show, by induction, that for every m there is a computation of length m.

If $m = 0$ then $\{< 0, a >\}$ is a computation of step 0.

Now let s be a computation of length m. Then $dom(s) = m + 1$, and $m \in dom(s)$. The domain of g is $A \times N$, so g is defined at $< s(m), m >$. Now let $t = s \cup \{< m + 1, g(s(m), m) >\}$. Since t is a union of two relations, it is a relation. The domain of the union is $dom(s) \cup \{m+1\} = m+1 \cup \{m+1\} = m+2$. Since $m+1 \notin dom(s)$, t is a function. We can also check that t is a computation.

1. $< 0, a > \in t$ because s is a computation sequence and $s \subseteq t$.

2. Let $k \in dom(t) = m+2$. If $k \in m$, then $t(k+1) = s(k+1) = g(s(k), k) = g(t(k), k)$. If $k = m$, $t(m + 1) = g(s(m), m) = g(t(m), m)$.

So, $m + 1 \in dom(f)$. **Q.E.D. Lemma 1.5.10**

Now we can easily show that $f(0) = a$ and for all $n \in N$, $f(n + 1) = g(f(n), n)$.

Since $0 \in dom(f)$, there is a computation sequence of length $m \geq 0$. Let s be one of these sequences. Then $< 0, a > \in s$. Since f is a function and $< 0, a > \in f$, $f(0) = a$.

Now let $n \in N$. Since $n + 1 \in dom(f)$, there is a computation sequence of length $m \geq n + 1$. Let s be one of these sequences. Then $s(n + 1) = g(s(n), n)$. Since f is a function and $< n, s(n) >, < n + 1, s(n + 1) > \in f$, $f(n) = s(n)$ and $f(n + 1) = s(n + 1)$. Then $s(n + 1) = g(s(n), n)$ implies $f(n + 1) = g(f(n), n)$. These remarks tell us that there is a function that satisfies the conditions of the theorem.

Now we have to prove its uniqueness.

Lemma 1.5.11 *The function f is unique.*

Proof: Assume that there is another function h, different from f, that satisfies the conditions of The Recursion Theorem. Let $A \neq \phi$ be the set of natural numbers where $f(n) \neq h(n)$. Since N is well ordered by Corollary 1.5.7, A has a least element. Let m be that element.

Case 1: $m = 0$. Then, $f(0) = a$ and $h(0) = a$, so $f(0) = h(0)$, a contradiction.

Case 2. $m \neq 0$. Then, by Proposition 1.5.1, there is some $p \in N$ such that $m = p+1$. Since m is the least point where the two functions differ, $f(p) = h(p)$. From the construction of f and h we get that $f(m) = g(f(p), p) = g(h(p), p) = h(m)$. Again, we get a contradiction.

In both cases we have contradictions. So, $A = \phi$ and $f = h$.

Q.E.D. Lemma 1.5.11
This concludes the proof of the theorem.
Q.E.D. The Recursion Theorem

Corollary 1.5.12 (The Parameterized Recursion Theorem) *Let $g : P \times A \times N \longrightarrow A$ and $a : P \longrightarrow A$ be functions. Then there is a function $f : P \times N \longrightarrow A$ such that for all $n \in N$ and $p \in P$, $f(p,0) = a(p)$ and $f(p,n+1) = g(p, f(p,n), n)$.*

Proof: Let A^P be the set of all functions with domain P and range in A. Let $G : A^P \times N \longrightarrow A^P$ be defined defined as $G(h,n)[p] = g(p, h(p), n)$ [2]. By The Recursion Theorem there is a unique function $F : N \longrightarrow A^P$ such that $F(0) = a$ and for all $n \in N$, $F(n+1) = G(F(n), n)$. Let us define $f : P \times N \longrightarrow A$ as $f(p,n) = F(n)[p]$.

Let us show that f satisfies the conditions of the corollary. First, for all $p \in P$

$f(p,0) = F(0)[p]$ by the definition of f
$= a(p)$. since $F(0) = a$.
Second, for all $n \in N$ and $p \in P$,
$f(p,n+1) = F(n+1)[p]$ by the definition of f
$= G(F(n), n)[p]$ a property of $F(n+1)$
$= g(p, F(n)[p], n)$ from the construction of G
$= g(p, f(p,n), n)$. from the definition of f

So we proved the existence of f we need to show that f is unique. Let $i : P \times N \longrightarrow A$ be a function satisfying the equalities $i(p,0) = a(p)$ and $i(p,n+1) = g(p, i(p,n), n)$ for all $n \in N$ and $p \in P$. If $i \neq f$ there are some $n_0 \in N$ and $p_0 \in P$ such that $i(p_0, n_0) \neq f(p_0, n_0)$. We define $H : N \longrightarrow A^P$ as $H(n)[p] = i(p,n)$. Then, $H(n_0) \neq F(n_0)$ because
$H(n_0)[p_0] = i(n_0, p_0) \neq f(n_0, p_0) = F(n_0)[p_0]$.
Now,
(1) $H(0)[p] = i(p,0) = a(p)$
and
(2) $H(n+1)[p] = i(p, n+1)$
$= g(p, i(p,n), n)$ a property of i
$= g(p, H(n)[p], n)$ from the definition of H
$= G(H(n), n)$ from the definition of G.

So, for we have two different recursive functions for the function G and element $a \in A^P$, contradicting The Recursion Theorem. **Q.E.D.**

The Parameterized Recursion Theorem allows us to define addition, multiplication and exponentiation. Let $P = A = N$, $a : N \longrightarrow N$ be the identity on N, and $g : N \times N \times N \longrightarrow N$ be defined as $g(m,n,p) = n+1$. The Parameterized Recursion Theorem gives us the function $f : N \times N \longrightarrow N$ satifying the equalities below.
$f(n,0) = n$

[2] h and $G(h,n)$ are functions with domain P and range in A. The value of $G(h,n)$ at p is $g(p, h(p), n)$.

$$1.\ n + 0 = n$$
$$2. n + (m + 1) = (n + m) + 1$$

Figure 1.7: The definition of +

$$1.\ m \cdot 0 = 0$$
$$2.\ m \cdot (n + 1) = (m \cdot n) + m$$

Figure 1.8: The definition of $\cdot$

$f(n, m + 1) = f(n, m) + 1$.

We denote this function by +. When we write the above equations in the infix notation, we get the definitions from Figure 1.7.

The operation + is associative and commutative. We leave the proofs of these properties as exercises.

We can go further and define the multiplication. In The Parameterized Recursion Theorem, we take $P = A = N$, the function a to be $0_N : N \longrightarrow N$ defined by $0_N(n) = 0$, and $g(m, n, p) = n+m$. We obtain the operation $\cdot$ defined in Figure 1.8.

Again, we leave the proofs that $\cdot$ is associative and commutative as exercises.

Next, we use The Parameterized Recursion Theorem to define the exponentiation m^n, with $m \neq 0$. We take $P = A = N$, $a[n] = 1$ for all $n \in N$, and $g(p, q, r) = q \cdot m$. We have the operation defined in Figure 1.9. We can prove, by induction, the properties of m^n from Figure 1.10.

Definition 1.5.13 (equipotent) *We say that two sets A and B are equipotent or have the same cardinality if there is a one-to-one function with domain A and range B. We denote this relation by $|A| = |B|$.*

Examples 1.5.14 1. $A = \{1, 2, 3\}$ and $B = \{3, 6, 8\}$ are equipotent because the function $f = \{< 1, 3 >, < 2, 6 >, < 3, 8 >\}$ is one-to-one, has domain $\{1, 2, 3\} = A$ and range $\{3, 6, 8\} = B$.

2. $A = \{1, 2\}$ and $B = \{3\}$ are not equipotent. The only function with domain A and range B is $f = \{< 1, 3 >, < 2, 3 >\}$ and that one is not one-to-one because $f(1) = f(2) = 3$.

Proposition 1.5.15 *1. $|A| = |A|$*
2. If $|A| = |B|$ then $|B| = |A|$.
3. If $|A| = |B|$ and $|B| = |C|$, then $|A| = |C|$.

Proof: 1. We use the function $1_A = \{< a, a > | a \in A\}$. This function is one-to-one, has domain A and range A.

$$1.\ m^0 = 1$$
$$2.\ m^{n+1} = m^n \cdot m$$

Figure 1.9: The definition of m^n

1. $m^{p+q} = m^p \cdot m^q$
2. $m^{p \cdot q} = (m^p)^q$

Figure 1.10: Properties of m^n

2. Assume that f is a one-to-one function with domain A and range B. Let f^{-1} be its inverse. From Section 1.3 we know that f^{-1} is also one-to-one function. Its domain is the $ran(f) = B$ and its range is $dom(f) = A$.

3. Assume that f is a one to one function with domain A and range B and g is a one-to-one function with domain B and range C. From Section 1.3 we know that the composition $g \circ f$ is a one-to-one function. Since $dom(g) = ran(f)$, the domain of $g \circ f$ is $dom(f) = A$ and the range of $g \circ f = ran(g) = C$. **Q.E.D.**

Now, we can define the weaker notion that the cardinal of A is less than the cardinal of B.

Definition 1.5.16 ($|A| \leq |B|$) *We say that the cardinality of A is less than the cardinality of B if there is a one-to-one function with domain A and range included in B.*

Examples 1.5.17 1. The cardinal of $A = \{1, 2, 3\}$ is less than or equal to the cardinal of $B = \{6, 7, 8, 9\}$ because the function $f : A \longrightarrow B$ defined by $f(1) = 6, f(2) = 7, f(3) = 8$ is one-to-one, has domain $\{1, 2, 3\}$ and range $\{6, 7, 8\} \subseteq B$.

2. If A is a subset of B then the cardinality of A is less than or equal to the cardinality of B. Let us see why. The function 1_A is one-to-one, has domain A and range $A \subseteq B$. So, it meets the conditions of Definition 1.5.16.

3. The cardinal of $A = \{1, 2\}$ is not less than or equal to the cardinal of $B = \{4\}$. The only function with domain A and range in B is $f = \{< 1, 4 >, < 2, 4 >\}$. But this function is not one-to-one.

Proposition 1.5.18 *1. If $|A| \leq |B|$ and $|A| = |C|$, then $|C| \leq |B|$.*
2. If $|A| \leq |B|$ and $|B| = |C|$, then $|A| \leq |C|$.
3. $|A| \leq |A|$.
4. If $|A| \leq |B|$ and $|B| \leq |C|$ then $|A| \leq |C|$.

The proof of this proposition is left as exercise. Now we will show that $|A| \leq |B|$ and $|B| \leq |A|$ imply $|A| = |B|$.

We first prove Lemma 1.5.19.

Lemma 1.5.19 (Cantor-Bernstein Lemma) *Let $A_1 \subseteq B \subseteq A$ and $|A_1| = |A|$. Then $|B| = |A|$.*

This lemma tells us that whenever a subset has the same cardinality as the set, all subsets that contain that subset are equipotent to the set.

Proof: Since $|A_1| = |A|$ there are one-to-one functions with domain A and range A_1. Let f be such a function. We apply the recursion theorem to the function $h : \mathcal{P}[A] \times N \longrightarrow \mathcal{P}[A]$ that assigns to each subset $X \subseteq A$ the set of its f-values, i.e. $h(X) = f(X) = \{f(x) \in A | x \in X\}$.

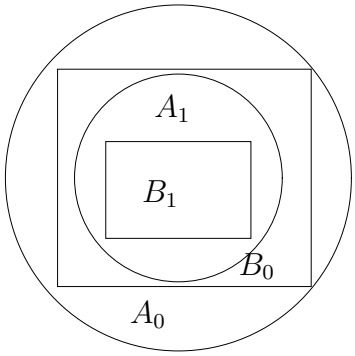


Figure 1.11: The Sets from Cantor-Bernstein' s Lemma

We define the sets
$A_0, A_1, \ldots, A_n \ldots$
and
$B_0, B_1, \ldots, B_n, \ldots$
as $A_0 = A$, $B_0 = B$,
and for each $n \in N$, $A_{n+1} = f[A_n]$ and $B_{n+1} = f[B_n]$. Since $A_1 \subseteq B_0 \subseteq A_0$, we can prove, by induction on n, that
$A_{n+1} \subseteq B_n \subseteq A_n$.
Now, let us define $C_n = A_n - B_n$ and $C = \bigcup_{n=0}^{\infty} C_n$.

In Figure 1.11, the A_i sets are represented as circles and the B_i sets as rectangles. The C_i's are the areas between the circles and the largest rectangle contained in the circle. Now let $D = A - C$. In the figure D is the union of all areas between a rectangle and the largest circle contained in it.

Now, $f(C_n) = f(A_n - B_n) = f(A_n) - f(B_n)$ because f is one-to-one
But $f(A_n) - f(B_n) = A_{n+1} - B_{n+1} = C_{n+1}$.
So, $f(C) = \bigcup_{n=1}^{\infty} C_n$.
Now we define a map g from A to B.

$$g(x) = \begin{cases} f(x) & \text{if } x \in C \\ x & \text{if } x \in D \end{cases}$$

$C \upharpoonright g$ is one-to-one and its range is included in C. $D \upharpoonright g$ is one-to-one and its range is D. Since the ranges of $C \upharpoonright g$ and $D \upharpoonright g$ do not intersect, $g = C \upharpoonright g \cup D \upharpoonright g$ is one-to-one. Its range is $f(C) \cup D = \bigcup_{n=1}^{\infty} C_n \cup D = (C - C_0) \cup D = A - C_0 = A - (A - B) = B$. **Q.E.D.**

Now we can prove The Cantor-Bernstein Theorem.

Theorem 1.5.20 (The Cantor-Bernstein-Theorem) $|A| \leq |B|$ *and* $|B| \leq |A|$ *imply* $|A| = |B|$.

Proof: Since $|A| \leq |B|$, there is function f with domain A and range included in B. The relation $|B| \leq |A|$ implies that there is a one-to-one function g with

domain B and range included in A. From Section 1.3 we know that $h = g \circ f$ is a one-to-one function with domain A and range included in A. Let A_1 be the range of h and B_0 be the range of g. We recall, from Section 1.3, that $A_1 \subseteq B_0$. Then A, B_0 and A_1 satisfy the conditions of the Cantor-Bernstein Lemma, so $|B_0| = |A|$. Since $|B_0| = |B|$, $|A| = |B|$. **Q.E.D.**

Now, let us define *finite* and *infinite sets*.

Definition 1.5.21 (finite sets) *A set S is finite if it is equipotent to some natural number $n \in N$. A set S is infinite if it is not finite.*

Now we will show that no finite set is equipotent a proper subset.

Proposition 1.5.22 *If $n \in N$ there is no one-to-one mapping with domain n and range a proper subset of n.*

Proof: The proof is by induction on n.

Basis: The set $n = \phi$ has no proper subsets, so the statement is true by default.

Inductive Step: Assume that the statement is true for n and let us assume that there is a one-to-one function f with range $n + 1 = n \cup n$ and range a proper subset of $n + 1$. We have 2 according to whether $n \in ran(f)$.

Case 1: $n \notin ran(f)$. Then the restriction $n \upharpoonright f$ is one-to-one function with domain n and range a subset of $n - \{f(n)\}$. This contradicts our assumption that the statement holds for n.

Case 2: $n \in ran(f)$. Then, for some $m \in n + 1$, $f(m) = n$. We define the function g that is exactly like f except that $g(m) = f(n)$ and $g(n) = n$. Now g is also one-to-one, and has the same domain and range as f. Since the range of f is a proper subset of $n + 1$, $ran(n \upharpoonright g) = ran(f) - \{n\}$ must be a proper subset of n. But this contradicts our assumption. **Q.E.D.**

Corollary 1.5.23 *1. If $n \neq m$ there is no one-to-one mapping with domain n and range m.*

2. N is infinite.

Proof: 1. We apply Proposition 1.5.6, which tells us that $>$ is a total relation on N. Since $m \neq n$ either $m > n$ or $n > m$. Let us assume that $n > m$. By Exercise 1.5.22, m is a proper subset of n. By Proposition 1.5.22 there is no one-to-one mapping with domain n and range m.

Now assume that there is a one-to-one mapping with domain m and range n. We recall, from Section 1.3, that the inverse of this function is a one-to-one mapping with domain n and range m. Again, this contradicts Proposition 1.5.22.

2. Let us define the function $f : N \longrightarrow N$ by $f(n) = n + 1$. Let us show that f is one-to-one. Let us assume that $n + 1 = m + 1$. By Exercise 1.5.24, $m = n$. So, f is one-to-one. Its domain is N and its range is $N - \{0\}$.**Q.E.D.**

The recursion theorem deals with functions $g : A \times N \longrightarrow A$ defined on $A \times N$. We need to extend it to functions whose domain is a subset of $A \times N$.

Theorem 1.5.24 *Let A be a non-empty set, $a \in A$, and g a function having as domain a subset of $A \times N$ and range a subset of A. Then there is a unique function $f : N \longrightarrow A$ such that*

1. $f(0) = a$

2. $f(n+1) = g(f(n), n)$ for all $n \in N$ such that $(n+1) \in dom(f)$.

3. The domain of f is either N or $n+1$. In the later case, $g(f(n), n)$ is undefined.

Proof: Let b be a set that is not in A. We extend g to a function h with domain $(A \cup \{b\}) \times N$ and range in $A \cup \{b\}$.

$$h(x,n) = \begin{cases} g(x,n) & \text{if } g(x,n) \text{ is defined} \\ b & \text{if } g(x,n) \text{ is not defined} \end{cases}$$

We notice that $g(b, n)$ is never defined. Now, we apply the recursion theorem to h and get a function $i : N \longrightarrow (A \cup \{b\})$.

We define the set k as follows:

If $b \in ran(h)$ then k is the smallest $l \in N$ such that $i(l) = b$

If $b \notin ran(h)$ then $k = N$.

The set k is well defined because N is well ordered by $>$ (Corollary 1.5.7). Now we define f to be the restriction of i to k. Let us check that f satisfies the conditions of the theorem.

1. $f(0) = i(0) = a$

2. From the construction of h we know that whenever $i(n+1) \neq b$, $i(n) \neq b$ and $g(i(n), n)$ is defined. So, for $(n+1) \in dom(f)$

$f(n+1) = i(n+1)$ since $(n+1) \in dom(f)$

$= h(i(n), n)$ from the construction of i

$= g(i(n), n)$ because $g(i(n), n)$ is defined

$= g(f(n), n)$ since $n \in dom(f)$

3. Assume that k, the domain of f, is not equal to N. Since $0 \in k$, $k = n+1$ for some $n \in N$ (Proposition 1.5.1). So, $f(n)$ is defined but $f(n+1)$ is not. From the construction of k, we know that $f(n+1)$ is undefined in two cases:

Case 1: $f(l)$ is undefined for some $l < n+1$. But $l < n+1$ implies that $l \leq n$, so $f(n)$ would have to be undefined. This contradicts the fact that $f(n)$ is defined.

Case 2: $g(f(n), n)$ is undefined.

Since Case 1 is not possible, we get that $g(f(n), n)$ is undefined.

Q.E.D.

Now we can list some of the properties of the finite sets.

Theorem 1.5.25 *1. If A is finite and $B \subseteq A$, then B is finite. Moreover, $|B| \leq |A|$.*

2. If A is finite and f is a function with domain A, $f(A)$ is finite. Moreover, $|f(A)| \leq |A|$.

Proof: Part 1. If $B = 0$ then $|B| = 0$ and $0 \leq n$ for all $n \in N$.

So, assume that $B \neq 0$. Let $x : n \longrightarrow A$ be a bijection from n to A. We define $g : A \times N \longrightarrow A$ below.

$a = x(j)$ where $j \in n$ is the least index such that $x(j) \in B$

$g(u, i) = x(j)$ where $j \in N$ is the least index $j > x^{-1}[u]$ such that $x(j) \in B$ when such a j exists.

We notice that a is defined because B is not empty. So, we can apply Theorem 1.5.24 to get the function y. We notice that $a \in B$ and the range of g is a subset of B, so $ran(y) \subseteq B$.

Lemma 1.5.26 *Whenver $y(i)$ is defined, it is equal to some $x(j) \in A$ with $j \geq i$.*

Basis: $i = 0$. Then $y(0) = a$. Since $a \in ran(x)$, $a = x(j)$ for some $j \in n$. Since 0 is the least element of N, $j \geq 0$.

Inductive Step: Assume that $y(i + 1)$ is defined. Then $y(i)$ is defined, so $y(i) = x(k)$ for some $k \geq i$. Since $y(i + 1)$ is the least element of $\{j \in n | j > k$ and $x(j) \in B\}$, $j > k$. From $j > k$ and $k \geq i$ we get $j \geq i + 1$. (Exercise 1.5.23) **Q.E.D. lemma**

By Lemma 1.5.26 $y(n)$ is undefined. So, the domain of y is $m + 1 \leq n + 1$.

Lemma 1.5.27 *Let $i, j \in m + 1$, $y(i) = x(k)$, $y(j) = x(l)$, and $j > i$. Then $l > k$.*

Proof: By induction on j.

Basis: $j = 0$. Then the assertion is vacuously true.

Inductive Step: Assume that the assertion is true for j and let assume that $j + 1 \in dom(y)$. By Lemma 1.5.26, $y(j + 1) = x(l)$ for some $l \in n$. Let $i < j + 1$ and $y(i) = x(k)$.

Case 1: $i < j$. Then there is some p such that $y(j) = x(p)$. Since $y(j + 1) = g(y(j), j)$, $l > p$. By induction hypothesis, $p > k$, so $l > k$ by the transitivity of $>$.

Case 2: $i = j$. Since $y(i + 1) = g(y(j), j)$, $l > k$. **Q.E.D. lemma**

Let us show that y is one to one. Assume that for some $i, j \in dom(y)$, $y(i) = y(j)$. By Lemma 1.5.26 there are $y(i) = x(k)$ and $y(j) = x(l)$. If $i \neq j$, then $k \neq l$ by Lemma 1.5.27. Then $k \neq l$ and $x(k) = x(l)$, contradicting the injectivity of x.

It remains to show that y is onto B. Assume that for some $b \in B$, $b \notin ran(y)$. Since $B \subseteq A$, let l be the smallest index such that $x(l) \notin B$. We have two cases, that correspond to the cases when g does not get to b, and skips b.

Case 1: $y(m) = x(p)$ and $p < l$. Then $y(m + 1) = g(y(m), m)$ is defined because there are indices $j > p$ with $x(j) \in B$. This contradicts the fact that the domain of y is $m + 1$.

Case 2: There are indices $i \in m + 1$ with $y(i) = x(p)$ and $p > l$. Let j be the smallest such index. If $j = 0$, this contradicts the definition of a. If $j = q + 1$, then $y(q) = x(k)$ and $k < l$. At the same time, $y(q + 1) = x(r)$ and $r > l$. Then, $y(q + 1) \neq g(y(q), q)$ because $x(l)$ is in B and has a lower index than $x(r)$.

Part 2. The proof is similar to Part 1. If $A = 0$, then $ran(f) = 0$. So, assume that $x : n \longrightarrow A$ is a bijection from n to A. Let $B = f(A)$. Let $p : B \longrightarrow n$ be the function below.

$p(f(x(i))$ is the smallest $j < n$ with $f(x(j)) = f(x(i))$.

The function p satisfies the two properties below.

(1) $p(f(x(i))) \leq i$

(2) $p(f(x(i))) = i$ iff for all $j < i$, $f(x(j)) \neq f(x(i))$.

We take $a = f(x(0))$ and $g : B \times N \longrightarrow B$ to be the function below.

$g(b, i) = f(x(j))$ where $j \in n$ is the smallest number greater than $p(b)$ satisfying $p(f(x(j))) = j$, if such a number exists.

Let $y : m \longrightarrow B$ be the function generated by Theorem 1.5.24.

Lemma 1.5.28 *For all $i \in m$, $p(y(i)) \geq i$.*

Proof: By induction on i.

Basis: $p(y(0)) = p(f(x(0))) = 0$.

Inductive Step: Assume that $p(y(i)) \geq i$ and $y(i+1)$ exists. Then $y(i+1) = f(x(j))$ with $p(y(i+1)) = j > p(y(i))$. Since $p(y(i)) \geq i$, $j \geq i+1$. **Q.E.D. lemma**

Since the range of p is a subset of n, Lemma 1.5.28 tells us that $y(n)$ is undefined. So, $m \leq n$.

We need Lemma 1.5.29 to show that y is one-to-one.

Lemma 1.5.29 *For all $i < j \in m$, $p(y(i)) < p(y(j))$.*

The proof of this lemma is left as an exercise.

Now, let us assume that for some $i, j \in m$, $y(i) = y(j)$. Then $p(y(i)) = p(y(j))$, so $i = j$ by Lemma 1.5.29. It remains to show that all elements of B are in the range of y. We use an argument similar to the one used in Part 1. Assume that $ran(y) \neq B$ and let r be the least element of $S = \{p(b) | b \in B - ran(y)\}$. Since $m \neq 0$, $m = q + 1$ for some $q \in N$.

Case 1: Assume that $T = \{i \in m | p(y(i)) > r\}$ is empty. Then $p(y(q)) < r$, and $p(y(q+1))$ is defined because there are $j > r$ with $p(f(x(j))) = j$. This contradicts the fact that the domain of y is $m = q + 1$.

Case 2: If $T \neq 0$, it has a least element k. $p(y(k))$ cannot be r because this implies that $f(x(r)) \in ran(y)$. So, $p(y(k)) > r$. Since $p(y(0)) = 0$, $k \neq 0$. Then $k = l + 1$. Since k is the least element of T, $l \notin T$. Then either $p(y(l)) = r$ or $p(y(l)) < r$. The equality is not possible since this implies that $r \notin S$. So, $p(y(l)) < r$. Then $g(y(l), l) \neq y(k+1)$ because $p(y(l)) < r < n$ and $p(f(x(r)) = r < p(y(k+1))$. **Q.E.D.** Theorem 1.5.24

Theorem 1.5.30 *1. If A and B are finite, then $A \cup B$ is finite. Moreover, $|A \cup B| \leq |A| + |B|$.*

2. If A is finite, $\mathcal{P}[A]$ is also finite. Moreover, $|\mathcal{P}[A]| = 2^{|A|}$.

Proof: Part 1. If $A = 0$, then $A \cup B = B$. If $B = 0$, then $A \cup B = A$. In both cases, $A \cup B$ is finite. Now let us assume that neither A nor B is empty. Let

$x : n \longrightarrow A$ and $y : m \longrightarrow B$ be two bijections. Let $z = x \cup \{< j+n, y(j) > | j \in m\}$. We will show that z is a function with domain $m+n$ and range $A \cup B$.

1. z is a function. Assume that $< i, a >, < i, b > \in z$. By Exercise 1.5.7 $j + n \geq n$, so $i < n$ and $i = j + n$ cannot be true at the same time. So, either $< i, a >, < i, b > \in x$ or $< i, a >, < i, b > \in \{< j + n, y(j) > | j \in m\}$. If $< i, a >, < i, b > \in x$, $a = b$ since x is a function. If $j + n = k + n$, then $j = k$ by Exercise 1.5.6. So, $\{< j + n, y(j) > | j \in m\}$ is a function, and $< i, a >, < i, b > \in \{< j + n, y(j) > | j \in m\}$ implies $a = b$.

In either case, $a = b$, so z is a function.

2. The domain of z is $n+m$. From the definition of z, $dom(z) = n \cup \{j+n | j \in m\}$.

We will show that $n + m \subseteq dom(z)$. If $i < n$, then $i \in dom(z)$. Let $n \geq i < n + m$. By Exercise 1.5.8 there is a j such that $i = n + j$. It remains to show $j < m$. Assume that $j \geq m$. By Exercise 1.5.7, $i = j + n \geq m + n = m + n$, contradicting $i < n + m$. So, $j < m$, and $i \in dom(z)$.

Let us show $dom(z) \subseteq n + m$. By Exercise 1.5.7, $j < m$ implies $j + n < m + n = n + m$. So, $i \in dom(z)$ implies $i < n$ or $i < n + m$. Since $n < m + n = n + m$, $i < n + m$.

3. The range of z is $A \cup B$. $ran(z) = ran(x) \cup \{< j + n, y(j) > | j \in m\} = A \cup \{y(j) | j \in m\} = A \cup B$.

If $A \cap B = 0$, z is one-to-one because x and $\{< j + n, y(j) > | j \in m\}$ are one-to-one and their ranges are distinct. So, $|A \cup B| = n + m$. If z is not one-to-one, then $A \cup B$ is the range of the function z. Since $n + m$ is finite, $ran(z)$ is finite and $|A \cup B| \leq n + m$ by Theorem 1.5.25, part 2.

Part 2: By induction on $|A|$. If $|A| = 0$, then $A = 0$ and $\mathcal{P}[0] = \{0\} = 1$. So, $|\mathcal{P}[0]| = 1$.

Asumme that the proposition is true for $|B| = n$ and let $|A| = n + 1$. Let $x : (n+1) \longrightarrow A$ be a bijection, $a = x(n)$ and $B = x[n]$. Recall that $x(n)$ is the value of x at n and $x[n]$ is the set $\{x(0), \ldots, x(n-1)\}$.

We will prove that $\mathcal{P}[A] = \mathcal{P}[B] \cup \{S \cup \{a\} | S \in \mathcal{P}[B]\}$. The proof is simple. If $T \in \mathcal{P}[A]$, then either $a \in T$ or in $a \notin T$. In the first case, $T \in \mathcal{P}[B]$. In the second case $S = T - \{a\} \in \mathcal{P}[B]$ and $T \in \{S \cup \{a\} | S \in \mathcal{P}[B]\}$. This establishes the inclusion $\mathcal{P}[A] \subseteq \mathcal{P}[B] \cup \{S \cup \{a\} | S \in \mathcal{P}[B]\}$.

For the other inclusion, we notice that $S \in \mathcal{P}[B]$ implies $S \in \mathcal{P}[A]$ and $(S \cup \{a\}) \in \mathcal{P}[A]$.

So, $\mathcal{P}[A] = \mathcal{P}[B] \cup \{S \cup \{a\} | S \in \mathcal{P}[B]\}$. Moreover, $\mathcal{P}[B]$ and $\{S \cup \{a\} | S \in \mathcal{P}[B]\}$ are disjoint. By induction hypothesis, $\mathcal{P}[B]$ is finite. The function $S \rightarrow S \cup \{a\}$ is a bijection between $\mathcal{P}[B]$ and $\{S \cup \{a\} | S \in \mathcal{P}[B]\}$, so $|\mathcal{P}[B]| = |\{S \cup \{a\} | S \in \mathcal{P}[B]\}|$. So, $\{S \cup \{a\} | S \in \mathcal{P}[B]\}$ is finite. By part 1, $\mathcal{P}[A] = \mathcal{P}[B] \cup \{S \cup \{a\} | S \in \mathcal{P}[B]\}$ is finite. Moreover,

$|\mathcal{P}[A]| = |\mathcal{P}[B]| + |\{S \cup \{a\} | S \in \mathcal{P}[B]\}|$ the two sets are disjoint
$= |\mathcal{P}[B]| + |\mathcal{P}[B]|$ $|\{S \cup \{a\} | S \in \mathcal{P}[B]\}| = |\mathcal{P}[B]|$
$= 2^{|B|} + 2^{|B|}$ by the induction hypothesis
$= 2 \cdot 2^{|B|}$ Exercises 1.5.11-1.5.10
$= 2^{|B|+1}$ Exercise 1.5.13
$= 2^{|A|}$ $|A| = |B| + 1$ **Q.E.D.**

Next we prove that a finite union of finite sets is finite, and the cartesian product of two finite sets is finite.

Theorem 1.5.31 *1. If S is finite and every $X \in S$ is finite, then $\cup S$ is finite.*
2. If A, B are finite, so is $A \times B$. Moreover, $|A \times B| = |A| \cdot |B|$.

Proof: Part 1. We do it by induction on $|S|$.

Basis: $|S| = 0$. Then $S = 0$ and $\cup S = 0$.

Inductive Step: Assume that the theorem is true for all sets T with $|T| = n$. Let S be a set of finite sets and $x : (n+1) \longrightarrow S$ be a bijection, $T = x[n]$ and $X = x(n)$. We recall that $x[n]$ is the image of n, i.e the set $\{x(0), \ldots, x(n-1)\}$. Then $|T| = n$. We will prove that $\cup S = (\cup T) \cup X$.

$z \in \cup S$
iff
there is some $y \in S$ such that $z \in y$
iff
there is some $y \in T$ such that $z \in y$, or $z \in X$ $\quad y \in S$ iff $y \in T$ or $y = X$
iff
$z \in \cup T$ or $z \in X$.

By the induction hypothesis, $\cup T$ is finite and X is finite because it is a member of S. By Theorem 1.5.30, Part 1, $\cup T \cup X$ is finite. Since $\cup S = (\cup T) \cup X$, $\cup S$ is finite.

Part 2. By induction on $|A|$.

Basis: Assume that $|A| = 0$. Then $A = 0$ and $A \times B = 0$, a finite set. The equality $|A \times B| = |A| \cdot |B|$ reduces to $0 = 0 \cdot |B|$ which is true.

Inductive Step: Let $|A| = n + 1$, $x : (n+1) \longrightarrow A$ a bijection, $C = x[n]$ and $D = \{x(n)\}$. Then $A = C \cup D$ and $|C| = n$. By Exercise 1.5.24, $A \times B = (C \times B) \cup (D \times B)$.

The function $f : x(0) \times B \longrightarrow B$ defined by $f(< x(0), b >) = b$ is a bijection, so $|x(0) \times B| = |B|$, a finite set. Since $C \times B$ is also a finite set by the induction hypothesis, $A \times B = (C \times B) \cup (D \times B)$ is finite by Theorem 1.5.30, Part 1.

The element $x(n) \notin C$, so $(C \times B) \cap (D \times B) = 0$. Then
$|A \times B| = |C \times B| + |D \times B|$ $\quad$ Theorem 1.5.30, Part 1
$= (|C| \cdot |B|) + |D \times B|$ $\quad$ by the induction hypothesis
$= (|C| \cdot |B|) + |B|$ $\quad |x(0) \times B| = |B|$
$= (|C| \cdot |B|) + (1 \cdot |B|)$ $\quad$ Exercise 1.5.9, 1.5.10
$= (|C| + 1) \cdot |B|$ $\quad$ Exercise 1.5.11
$= |A| \cdot |B|$ $\quad |A| = |C| + 1$
Q.E.D.

Next, we introduce the notion of countable set.

Definition 1.5.32 *[countable sets] A set A is countable if $|A| = |N|$. A set A is at most countable if $|A| \leq |N|$.*

Theorem 1.5.33 *1. An infinite subset of a countable set is countable.*
2. The range of a countable sequence is at most countable.

Proof: Part 1: Let $x : N \longrightarrow A$ be a bijection and B an infinite subset of A. The definitions of $a \in A$ and $g : A \times N \longrightarrow A$ are given below.

$a = x(j)$ where $j \in n$ is the least index such that $x(j) \in B$

$g(u, i) = x(j)$ where $j \in N$ is the least index $j > x^{-1}[u]$ such that $x(j) \in B$

Since $B \neq 0$, a is defined. Next we prove that g is a total function. Assume that it is not. Then, for some $u \in A$ and $i \in N$, $g(u, i)$ is undefined. Let $k = x^{-1}[u]$. Then, for all $j > k$, $x(j) \notin B$, i.e. $B \subseteq x[k+1]$. By Therem 1.5.25, B is finite, contradicting our assumption. We apply The Recursion Theorem and get the function $y : N \longrightarrow A$. We need to prove that y is one-to-one, and has range B.

Lemmas 1.5.26, 1.5.27 are valid for y, so we have the injectivity. The remarks that follow Lemmas 1.5.27 are also valid for y, so y is onto B.
Part 2: The proof is similar the proof of Part 2 of Theorem 1.5.25. It is left as exercise. **Q.E.D.**

Corollary 1.5.34 *If A is at most countable, then A is finite or countable.*

Proof: $|A| \leq |N|$ means that there is a one-to-one function $f : A \longrightarrow N$. Then, $f[A]$ is a subset of the countable set N. If $f[A]$ is not finite, then it is countable by Theorem 1.5.33. **Q.E.D.**

The next theorem relies on properties of N.

Theorem 1.5.35 *1. The union of two countable sets is countable.*

2. The cartesian product of two countable sets is countable.

3. The set of all finite subsets of a countable set is countable.

Proof: 1. Let $f : N \longrightarrow A$ and $g : N \longrightarrow B$ be two bijections. We define the function $h : N \longrightarrow (A \cup B)$ as

$$h(n) = \begin{cases} f(m) & \text{if } n = 2 \cdot m \\ g(m) & \text{if } n = (2 \cdot m) + 1 \end{cases}$$

Theorem 1.5.36 (Cantor's Theorem) *For any set A there is no bijection $f : A \longrightarrow \mathcal{P}[A]$.*

Proof: Let us assume that there is a bijection $f : A \longrightarrow \mathcal{P}[A]$. Let B be the set $\{x \in A | x \notin f(x)\}$. Clearly, $B \subseteq A$, so $B \in \mathcal{P}[A]$. Since the range of f is $\mathcal{P}[A]$, B has a pre-image $y \in A$. Now we ask the question, does y belong to $B = f(y)$?

If $y \in B$ then $y \notin f(y)$, from the definition of B. Since $B = f(y)$, $y \notin B$. So, we have a contradiction.

If $y \notin B$, then $y \notin f(y)$. By the definition of B, $y \in B$. Again, we have a contradiction.

We got the contradiction because we assumed that B has a pre-image. So, B has no pre-image, and f is not a bijection from A to $\mathcal{P}$.**Q.E.D.**

Exercises

Exercise 1.5.1 *Let P and A be two sets. Show that the collection of all functions with domain P and range in A is a set, A^P. What happens when $A = \phi$? What happens when $P = \phi$?*

Exercise 1.5.2 *Prove by induction on n that $n \subseteq N$.*

Exercise 1.5.3 *Show that the operation $+$ defined in Figure 1.5.11 is comutative, i.e.*

$m + n = n + m$

for all $m, n \in N$.

Hint: We need to show that $m+n = n+m$. We do it by induction on n. For the basis, prove that $0+n = n$. For the inductive step show, by induction on m, that $(n+1)+m = (n+m)+1$ and use it to establish $m+(n+1) = (n+1)+m$.

Exercise 1.5.4 *Prove that $+$ is associative, i.e. for all $m, n, p \in N$,*

$(m + n) + p = m + (n + p)$.

Hint: Use induction on p.

Exercise 1.5.5 *Prove that $m + 1 = n + 1$ implies $m = n$.*

Exercise 1.5.6 *Prove, by induction on p, that $m + p = n + p$ implies $m = n$.*

Hint: Use Exercise 1.5.5.

Exercise 1.5.7 *Prove, by induction on p, that $m > n$ implies $m + p > n + p$.*

Exercise 1.5.8 *Prove that for $n \geq m$ there is some $p \in N$ such that $m+p = n$.*

Hint: Define $A = \{q \in N | m + q \geq n\}$. This set is not empty, because $m + n \geq n$. Since $>$ is a well order on N, A has a least element p. Show that p satisfies the equality $m + p = n$.

Exercise 1.5.9 *Prove that $m \cdot 1 = m$.*

Exercise 1.5.10 *Show that $\cdot$ is commutative, i.e. for all $m, n \in N$, $m \cdot n = n \cdot m$.*

Hint: Use induction on n.

For the basis prove that $0 \cdot n = 0$. For the inductive step, prove by induction on m, that $(n \cdot m) + m = (n+1) \cdot m$. In this proof, use the associativity and the commutativity of $+$.

Exercise 1.5.11 *Show that $\cdot$ distributes over $+$, i.e.*

$(x + y) \cdot z = (x \cdot z) + (y \cdot z)$.

Hint: Use induction on z.

Exercise 1.5.12 *Prove that $\cdot$ is associative.*

Hint: Prove that $(m \cdot n) \cdot p = m \cdot (n \cdot p)$ by induction on p. For the inductive step use the distributivity of $\cdot$ over $+$.

Exercise 1.5.13 *1. Show that $m^1 = m$.*

2. Prove, by induction on q, that $m^{p+q} = m^p \cdot m^q$.

Exercise 1.5.14 *Prove, by induction on q, that $(m^p)^q = m^{p \cdot q}$.*

Exercise 1.5.15 *Prove the inclusions*
$A_{n+1} \subseteq B_n \subseteq A_n$
used in Lemma 1.5.19.

Exercise 1.5.16 *Let $\mathbf{P}[x]$ be a property. Assume that for all $n \in N$, if $\mathbf{P}[k]$ holds for all $k < n$, then $\mathbf{P}[n]$. Show that $\mathbf{P}[n]$ is true for all $n \in N$.*

Exercise 1.5.17 *Prove the following induction principle:*
Let $\mathbf{P}[n]$ be a property and $m \in N$. If
1. $\mathbf{P}[m]$ holds, and
2. for all $n \in N$, if $m \in n$ then $\mathbf{P}[n]$ implies $\mathbf{P}[n+1]$,
then $\mathbf{P}[n]$ is true for all n that include m.

Exercise 1.5.18 *Let A be a non-empty set. Show that there is a set that contains all sequences of elements of A finite or infinite.*

Hint: Finite sequences are functions $n \longrightarrow A$ and infinite sequences are functions $N \longrightarrow A$. In either case, the elements of a sequence are members of $N \times A$.

Exercise 1.5.19 *Prove Proposition 1.5.18.*

Exercise 1.5.20 *Prove Lemma 1.5.29. Hint: use induction on j.*

Exercise 1.5.21 *We say that a set x is transitive if every element of x is a subset of x, i.e. $z \in y \in x$ implies that $z \in x$. Prove that if x is transitive, so is $succ(x) = x \cup \{x\}$. Then prove that all numbers $n \in N$ are transitive.*

Exercise 1.5.22 *Use Exercise 1.5.21 to prove that for all $n, m \in N$, $m > n$ implies $n \subseteq m$.*

Exercise 1.5.23 *Show that for all $i, j, k \in N$, $i \geq j$ and $k > j$ imply $k \geq (i+1)$.*

Exercise 1.5.24 *Show that $\times$ distributes over $\cup$, i.e. $(X \cup Y) \times Z = (X \times Z) \cup (Y \times Z)$.*

Exercise 1.5.25 *Prove Part 2 of Theorem* **??**.

Exercise 1.5.26 *Let A be a set. On it we define the relation $\in_A$ by $x \in_A y$ if $x, y \in A$ and $x \in y$. We say that the set A is an ordinal if A is transitive and $\in_A$ is a well order on A. Show that the natural numbers and N are ordinals.*

Exercise 1.5.27 *Let A be an ordinal and let $x \in A$. Show that x is an ordinal.*

Exercise 1.5.28 *Let A be an ordinal. Show that the successor of A, $S(A) = A \cup \{A\}$ is an ordinal.*

Exercise 1.5.29 *If α and β are ordinal numbers such that $\alpha \subseteq \beta$, then $\alpha = \beta$ or $\alpha \in \beta$*

1.6 Strings

Definition 1.6.1 (strings over A) *Let A be a set of symbols. A string over A is a function $w : n \longrightarrow A$, where $[n]$ is an initial segment of N. The number n is the length of the string, $[n]$ is the set of indices or* **addresses of** *w, and for each $i \in [n]$, $w(i)$ is the value of w at i.*

We say that a string is *empty* if its length is 0, and *non-empty* if its length is greater than 0. We will write λ for an empty string, and $w = w_0 w_1 ... w_{n-1}$ for the string w of length $n > 0$, with $w(0) = w_0$, $w(1) = w_1, \ldots$, $w(n-1) = w_{n-1}$. Throughtout this book we will write $|u|$ for the length of u. We denote by A^* *the set of all strings over A*, empty or not.

Remark 1.6.2 *There is only one empty string.*

Proof

Let A be any alphabet(including the empty one). Then there is only one function, from $[0] = \phi$ to A, and that function is $\lambda = \phi$. So, there is only one empty string in A^*. Moreover, the same λ is a substring of *all alphabets*. **Q.E.D.**

Next we define the fundamental operation on strings, *the concatenation*. The concatenation appends a string at the end of another string. So, if $u = u_0 ... u_{n-1}$ and $v = v_0 ... v_{m-1}$, u concatenated with v, writen $u.v$ or simply uv, is the string $u_0 ... u_{n-1} v_0 ... v_{m-1}$. The length of the concatenation is $n + m$. Now we can give a formal definition of concatenation.

Definition 1.6.3 (concatenation) *Let A be a set of symbols and $u : [n] \longrightarrow A$ and $v : [m] \longrightarrow A$ be two strings over A. Then u concatenated with v, writen $u.v$ or simply uv, is the string $w : [n + m] \longrightarrow A$ defined by the piecewise formula*

$$w(i) = \begin{cases} u(i) & \text{if } 0 \leq i < n \\ v(i-n) & \text{if } n \leq i < n+m \end{cases}$$

The next propositions will show that the concatenation is one-to-one, associative, and every string can be writen as a concatenation of 2 strings.

Proposition 1.6.4 (monotonicity of concatenation) *The concatenation is 1-1 in both arguments, i.e. for all strings u,v, and w, $uv = uw$ implies $v = w$, and $vu = wu$ implies $v = w$.*

Proof: Let us prove that $uv = uw$ first. Since $uv = uw$, $|uv| = |uw|$. Since $|uv| = |u| + |v|$ and $|uw| = |u| + |w|$, we get that $|u| + |v| = |u| + |w|$. The last equality reduces to $|v| = |w|$, i.e. v and w have the same length. The values of the functions uv and uw are given below.

$$uv(i) = \begin{cases} u(i) & \text{if } 0 \leq i < |u| \\ v(i-|u|) & \text{if } |u| \leq i < |u| + |v| \end{cases}$$

$$uw(i) = \begin{cases} u(i) & \text{if } 0 \leq i < |u| \\ w(i-|u|) & \text{if } |u| \leq i < |u| + |w| \end{cases}$$

Since $uv = uw$, $uv(i) = uw(i)$ for all $|u| \leq i < |u| + |v| = |u| + |w|$. So, $v(i - |u|) = w(i - |u|)$ for $|u| \leq i < |u| + |v| = |u| + |w|$. Now let us substitute i by $j + |u|$. We get that $v(j + |u| - |u|) = w(j + |u| - |u|)$ holds for $|u| \leq j + |u| < |u| + |v| = |u| + |w|$. We evaluate the sums and substract $|u|$ from the 4 terms of the inequality and obtain the equality $v(j) = w(j)$ for all j sat$0 \leq j < |v| = |w|$. The last relation tells us that u v are equal.

The second part of the proposition is proved in a similar fashion. We leave it as exercise. **Q.E.D.**

Proposition 1.6.5 *If $uv = xy$ and $|u| = |x|$ then $u = x$ and $v = y$.*

Proof: Let $|u| = |v| = n$. Then for all $0 \leq i < n$, $uv(i) = u(i)$, and $xy(i) = x(i)$. Since $uv = xy$, $u(i) = x(i)$ for all $0 \leq i < n$, i.e. $u = x$. The equality $v = y$ follows from the monotonicty of the concatenation. **Q.E.D.**

Proposition 1.6.6 *1. The concatenation is associative, i.e. for all strings u, v, w, $.(.(u, v), w) = .(u, .(v, w))$.*

2. $.(\lambda, u) = .(u, \lambda) = u$.

3. If A has one element, then . is commutative, otherwise it is not.

Proof

1. We give an intuitive proof and leave the formalization as an exercise to the interested reader.

Let $u = u_0...u_{k-1}$, $v = v_0...v_{l-1}$, $w = w_0...w_{m-1}$. Then $.(u, v) = u_0...u_{k-1}v_0...v_{l-1}$ and $.(v, w) = v_0...v_{l-1}w_0...w_{m-1}$.

Now, $.(.(u, v), w) = .(u_0...u_{k-1}v_0...v_{l-1}, w_0...w_{m-1} = u_0...u_{k-1}v_0...v_{l-1}w_0...w_{m-1}$, and

$.(u, .(v, w)) = .(u_0...u_{k-1}, v_0...v_{l-1}w_0...w_{m-1} = u_0...u_{k-1}v_0...v_{l-1}w_0...w_{m-1}$.

So, $.(.(u, v), w) = .(u, .(v, w))$.

2. Intuitively, by appending an empty string at the beginning or at the end of a string we should get the original string. However this is just our intuition, and we need a formal proof to back it up.

Let $u : [n] \longrightarrow A$ be a string. We know that $\lambda : [0] = \phi \longrightarrow A$.

Since $0 + n = n$, the functions u and $.(\lambda, u)$ have the same domain $[n]$.

The string $.(\lambda, u) : [0 + n] \longrightarrow A$ is defined as

$$.(\lambda, u)(i) = \begin{cases} \lambda(i) & \text{if } 0 \leq i < 0 \\ u(i-0) & \text{if } 0 \leq i < 0 + n \end{cases}$$

Since the condition $0 \leq i < 0$ is never true, $i - 0 = i$, and $0 + n = n$, the above definition reduces to

$.(\lambda, u)(i) = u(i)$ for $0 \leq i < n$.

Since the domain of the functions u and $.(\lambda, u)$ is n and they have the same value for all elements of the domain, the two functions are equal. So, $.(\lambda, u) = u$.

In a similar fashion we show that $.(u, \lambda) = u$.

3. Assume that A has more than 1 symbol. Let $\mathbf{a}$ and $\mathbf{b}$ be two distinct symbols in A. Then $.(\mathbf{a}, \mathbf{b}) = \mathbf{ab} \neq \mathbf{ba} = .(\mathbf{b}, \mathbf{a})$.

So, $.(\mathbf{a}, \mathbf{b}) \neq .(\mathbf{b}, \mathbf{a})$, and . is not commutative.

If A has one element, then $A = \{\mathbf{a}\}$ for some symbol $\mathbf{a}$. Then, every string $w \in A^*$ has the form $\mathbf{a}^n = \underbrace{\mathbf{aa}\ldots\mathbf{a}}_{n\text{times}}$. If $n = 0$ then $\mathbf{a}^n = \lambda$.

Now let u and v be two strings in A^*. Then $u = \mathbf{a}^n$ and $v = \mathbf{a}^m$ for some natural numbers n and m. Then $.(u, v) = .(\mathbf{a}^n, \mathbf{a}^m) = \mathbf{a}^{n+m}$.

The concatenation $.(v, u) = .(\mathbf{a}^m, \mathbf{a}^n) = \mathbf{a}^{m+n}$.

Since $\mathbf{a}^{m+n} = \mathbf{a}^{n+m}$, $.(u, v) = .(v, u)$. So, . is commutative. **Q.E.D.**

The associativity of the concatenation allows us to dispense with the cumbersome notation $.(u, v)$. Further on we will write uv, or $u.v$ when we want to emphasize that the string uv is obtained by concatenating u and v.

Proposition 1.6.7 (string decomposition) *Let $u : [n] \longrightarrow A$ and $0 \leq m \leq n$. Let v_m be the restriction of the function u to the set $[m]$. Then there is string w_m such that $u = v_m w_m$.*

Proof: Let w_m be the function with domain $[n-m]$, defined by $w(i) = u(i+m)$ for all $0 \leq i < n - m$. Let us check that this is a good definition, i.e. there is such a function. First, we must show that $n - m \geq 0$ and second, that $u(i+m)$ is defined for all $0 \leq i < n - m$.

The first conditon is satisfied because $m \leq n$. For the second condition we must show that for $0 \leq i < n - m$, $0 \leq i + m < n$. We add m to the 3 terms of the first inequality and get $m \leq i + m < n - m + m$. With $0 \leq m$, the last inequality becomes $0 \leq m \leq i + m < n$.

So, w_m exists. The domain of $v_m w_m$ is $m+n-m = n$, so u and $v_m w_m$ have the same domain. Let us check that for all $0 \leq i < n$, $v_m w_m(i) = u(i)$.

$$v_m w_m(i) = \begin{cases} v_m(i) & \text{if } 0 \leq i < m \\ w_m(i-m) & \text{if } m \leq i < m+n-m \end{cases}$$

Since $v_m(i) = u(i)$ for all $0 \leq i < m$, we get the relation below.

$$v_m w_m(i) = \begin{cases} u(i) & \text{if } 0 \leq i < m \\ w_m(i-m) & \text{if } m \leq i < n \end{cases}$$

From the construction of w_m we know that $w(j) = u(m + j)$ for all $0 \leq j < n-m$. Let $j = i-m$. Then $w_m(i-m) = u(m+i-m)$ for all $0 \leq i-m < n-m$. By adding m to all 3 terms of the inequality we get $w_m(i-m) = u(i)$ for all $m \leq i < n - m + m$. We use the equality to replace $v_m(i)$.

$$v_m w_m(i) = \begin{cases} u(i) & \text{if } 0 \leq i < m \\ u(i) & \text{if } m \leq i < n \end{cases}$$

Now the formula is reduced to $v_m w_m(i) = u(i)$ for all $0 \leq i < n$, i.e. $v_m w_m = u$.**Q.E.D.**

The strings v_m are called *prefixes* of u and the strings w_m are called *suffixes* of u. The prefixes of $u = \mathbf{abcd}$ are $v_0 = \lambda$, $v_1 = \mathbf{a}$, $v_2 = \mathbf{ab}$, $v_3 = \mathbf{abc}$ and $v_4 = \mathbf{abcd}$. The corresponding suffixes are $w_0 = \mathbf{abcd}$, $w_1 = \mathbf{bcd}$, $w_2 = \mathbf{cd}$, $w_3 = \mathbf{d}$, and $w_4 = \lambda$.

The string $u = u_0 \dots u_{n-1}$ has prefixes $v_0 = \lambda$, $v_1 = u_0$, $\dots$,$v_n = u_0 \dots u_{n-1}$. All these prefixes have different lengths, so u has $n+1$ prefixes. It also has $n+1$ suffixes, $w_0 = u_0 \dots u_{n-1}$, $w_1 = u_1 \dots u_{n-1}$, $\dots w_n = \lambda$.

If $u = \lambda$, then $n = 0$, so u has only one prefix $v_0 = \lambda$ and one suffix $w_0 = \lambda$.

We say that the prefix v of u is *proper* if $v \neq \lambda$ and $v \neq u$. The suffix w of u is said to be *proper* if $w \neq \lambda$ and $w \neq u$.

Now we can ask the question? If $u = vw$ is v a prefix and w a suffix of u?

Proposition 1.6.8 *If $u = vw$ then u is a prefix of u and w is a suffix of w.*

Proof: First of all we notice that the length of v cannot be greater than the length of u. For, if that were the case, $|vw| = |v| + |w| > |u| + |w| > |u|$, so the functions vw and u cannot be equal since they have different domains.

So, $|v| = |u|$. Let m be the length of v. Then $vw = v_m w_m$, and $|v| = |v_m|$. We apply Proposition 1.6.5 and get that $v_m = v$ and $w_m = w$. So, v is a prefix and w a suffix of u.**Q.E.D.**

We can also show that a prefix of a prefix of u is a prefix of u, and a suffix of a suffix of u is a suffix of u.

Corollary 1.6.9 *A prefix of a prefix of u is a prefix of u.*

Proof: Let v be a prefix of u and let w be a prefix of v. Since v is a prefix of u, $u = vy_1$. At the same time $v = wy_2$ because w is a prefix of v. So, $u = v.y_1 = wy_2.y_1 = w.y_2y_1$, i.e. w is a prefix of u.**Q.E.D.**

Corollary 1.6.10 *If x and y are two prefixes of u then one of them is a prefix of the other.*

Proof: Since x and y are prefixes, then there are integers l and m such that $x = v_l$ and $y = v_m$. By comparing l and m we distinguish two cases, $l \leq m$, and $m \leq l$.

If $l \leq m$ then $v_l = [l] \uparrow u = [l] \uparrow ([m] \uparrow u) = [l] \uparrow v_m$. This tells us that v_l is a prefix of v_m.

We use the same argument for the case $m \leq l$. The strings v_m and v_l are the restrictions of u to $[m]$, respectively $[l]$. Since $m \leq l$, $[m] \subseteq [l]$, so v_m is the restriction of v_l to $[m]$, i.e. v_l is a prefix of v_m.**Q.E.D.**

Now we need to define *substrings*.

Definition 1.6.11 (substring) *Let u be a string. The string v is a substring of u if $u = x.v.y$ for some strings x and y.*

We can interpret the equality $u = xvy$ in terms of prefixes and suffixes. $u = xvy$ means that v is a prefix of the suffix vy of u, and v is a suffix of the prefix xv of u. So, we can paraphrase the above definition as *a substring of u is a prefix of a suffix of u* and *a substring of u is a suffix of a prefix of u.*

Examples 1.6.12 *1. The string $v = \mathbf{bc}$ is a substring of $u = \mathbf{abcde}$, because there are two strings, $x = \mathbf{a}$ and $y = \mathbf{de}$, such that $u = x.v.y$.*

2. The string $v = \mathbf{pq}$ is a substring of $u = \mathbf{pqrs}$, because there are two strings, $x = \lambda$ and $y = \mathbf{rs}$ such that $u = x.v.y$.

3. The string $v = \mathbf{345}$ is a substring of $u = \mathbf{12345}$, because there are two strings, $x = \mathbf{12}$ and $y = \lambda$ such that $u = x.v.y$.

The prefixes and suffixes are also substrings.

Proposition 1.6.13 *1. Let v be a prefix of u and w be a suffix of u. The both v and w are substrings of u.*

Proof: Let u be a prefix of u. Then $u = v.y$ for some string y. Let $x = \lambda$. Then $x.v.y = \lambda.v.x = v.x = u$. So v is a substring of u. Now let w be a suffix of u. Then $u = xw$ for some x. Let $y = \lambda$. Then $x.w.y = x.w.\lambda = xw = u$. So w is a substring of u.**Q.E.D.**

Sometimes we need a better specification of the non-empty substrings than the one provided by Definition 1.6.11. For example, the substring $v = \mathbf{abc}$ occurs in $u = \mathbf{abcabc}$ twice, and at times we have to differentiate between these occurrences. For this, we need a definition of *occurence*.

Definition 1.6.14 (occurrence) *An occurrence is a triple $< u, i, v >$ satisfying the conditions below.*

1. u and v are strings

2. i is a natural number

3. there are strings x and y such that $u = xvy$

4. $|x| = i$

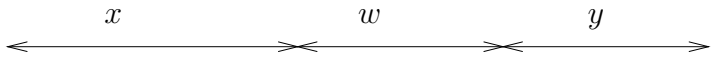
At times we will say that *v occurs in u at position i* instead of $< u, i, v >$ is an occurrence.

Examples 1.6.15 *1. The triple $< \mathbf{abcabc}, 0, \mathbf{abc} >$ is an occurrence because the strings $x = \lambda$ and $y = \mathbf{abc}$,satisfy the conditions $x.\mathbf{abc}.y = \mathbf{abcabc}$ and $|\lambda| = 0$.*

2. The triple $< \mathbf{abcabc}, 3, \mathbf{abc} >$ is an occurrence because the strings $x = \mathbf{abc}$ and $y = \lambda$,satisfy the conditions $x.\mathbf{abc}.y = \mathbf{abcabc}$ and $|\mathbf{abc}| = 3$.

The difference between substring and occurrence is that the occurrence involves the position where the substring occurs. So, we are able the differentiate the multiple occurrences of the same substring. In an occurrence $< u, i, v >$, i can have as values the indices 0,...$|u| - 1$ of u as well as $|u|$.

Throughout the book we will represent the strings as horizontal lines with arrow heads at both ends. The arrow tips indicate the first and the las character of the string. Figure **??** represents the string xwy.



On many occasions we want to point out that certain symbols occur in the string. The notation $F = (G \vee H)$ shows that F obtained by concatenating the strings (, G, $\vee$, H, and). The strings (, $\vee$, and) are symbols, while G, and H are strings. The figure below represents F.

The positions will be indicated by upward pointing vertical arrows. The figure below indicates the occurrence $< u, i, w >$

Let us see what happens when the occurence $< u, i, v >$ has $i = 0$. In the equality $u = xvy$ the string x has length 0, i.e. $x = \lambda$. So, $u = vy$, i.e. v is a prefix of u.

If $i = |u|$ then the string x, of the equality $u = xvy$, has the same length as u. So, the strings v and y must have length 0, i.e. they are empty. So, $v = \lambda$.

If $i + |v| = |u|$ then $|x| + |v| = |u|$. Since $|x| + |v| + |y| = |u|$, the length of y is 0. So, $xvy = u$ become $xv = u$, i.e. v is a suffix of u.

Proposition 1.6.16 *$< u, i, \lambda >$ is an occurrence for all $0 \leq i \leq |u|$.*

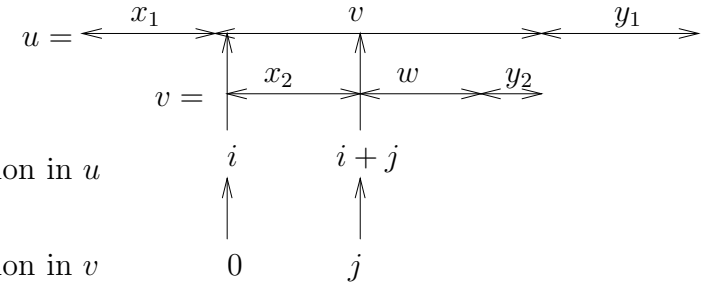
Proof: Let n be the length of u. From Proposition 1.6.7 we know that for any i, $0 \leq i \leq n$, there is a string v_i of length i and a string w_i such that $u = v_i w_i$. Let $x = v_i$ and $y = w_i$. Then $x\lambda y = v_i \lambda w_i = v_i w_i = u$, and $x| = |v_i| = i$. The last two equations tell us that $< u, i, \lambda >$ is an occurrence. **Q.E.D.**

The next proposition shows that an occurrence of an occurence of u is an occurrence of u.

Proposition 1.6.17 *Let $< u, i, v >$ and $< v, j, w >$ be two occurrences. Then $< u, i + j, w >$ is an occurrence.*

Proof:

The occurrence $< u, i, v >$ means that $u = x_1 v y_1$ and $|x_1| = i$. The occurrence $< v, j, w >$ tells us that $v = x_2 w y_2$ and $|x_2| = j$. These properties are captured by Figure .

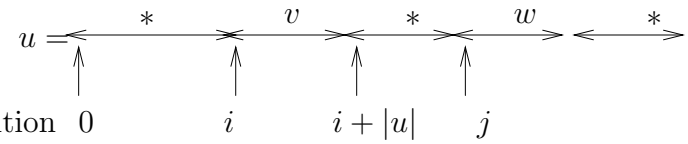


Now we replace v by x_2wy_2 in the equation of u and get $u = x_1x_2wy_2y_1$. Now let $x = x_1x_2$ and $y = y_2y_1$. Then $u = xwy$ and $|x_1x_2| = |x_1| + |x_2| = i + j$. The last two equations tells us that $< u, i + j, w >$ is an occurrence.**Q.E.D.**

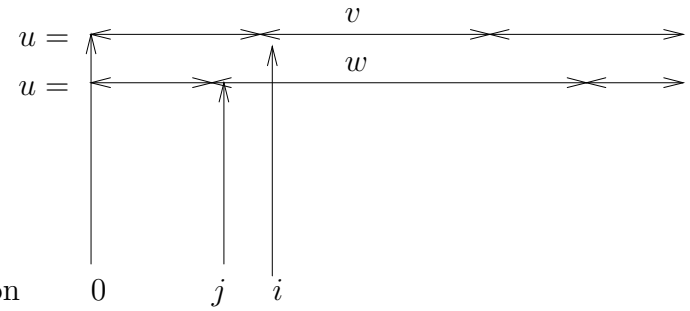
Definition 1.6.18 (no overlap, contains, precedes) *Let $o_1 =< u, i, v >$ and $o_2 =< u, j, w >$ be two occurrences.*

1. *The occurrences o_1 and o_2 do not overlap if either $i+|v| \leq j$ or $j+|w| \leq i$.*
2. *o_1 contains o_2 if $i \leq j$ and $j + |w| \leq i + |v|$.*
3. *o_1 precedes o_2 if $i < j$.*

Let us identify these relations by looking at the next two figures.



The first figure tells us that the occurrence $< u, i, v >$ precedes the occurrence $< u, j, w >$. At the same time, the two occurrence are not overlapping because the v string terminates before w begins, i.e. $i + |v| \leq j$. The * on the top of the three strings means that they can be empty.



Here, the occurrence $< u, i, w >$ precedes $< u, j, v >$. At the same time $< u, i, v >$ contains $< u, j, v >$ because $j \leq i$ and v ends before w.

Now let us define another operation on strings, the substitution.

Definition 1.6.19 (string substitution) *Let $< u, i, v >$ be an occurrence and w a string. Let $u = xvy$ be the equality corresponding to this occurrence. Then, u with v replaced by w at i, writen $u[i \leftarrow v/w]$, is the string xwy.*

So, the substitution replaces a specified occurrence of a substring by another string. Let us look at substitutions.

Examples 1.6.20 *1. Let us compute* $\mathbf{abcabc}[1 \leftarrow \mathbf{bc}/\mathbf{efg}]$. *Here $u = \mathbf{abcabc}$, $v = \mathbf{bc}$, $w = \mathbf{efg}$, and the occurrence is $< u, 1, v >$. The strings x and y corresponding to this occurrence are $x = \mathbf{a}$ and $y = \mathbf{abc}$. The string $u[i \leftarrow v/w] = x.w.y = \mathbf{a.efg.abc} = \mathbf{aefgabc}$. This substitution replaced the first* **bc** *in u by* **efg**.

2. Let us compute $\mathbf{1234}[3 \leftarrow \lambda/\mathbf{ab}]$. *Here $u = \mathbf{1234}$, $v = \lambda$, $w = \mathbf{ab}$. From Proposition 1.6.16 we know that $< u, 3, \lambda >$ is an occurrence. The strings*

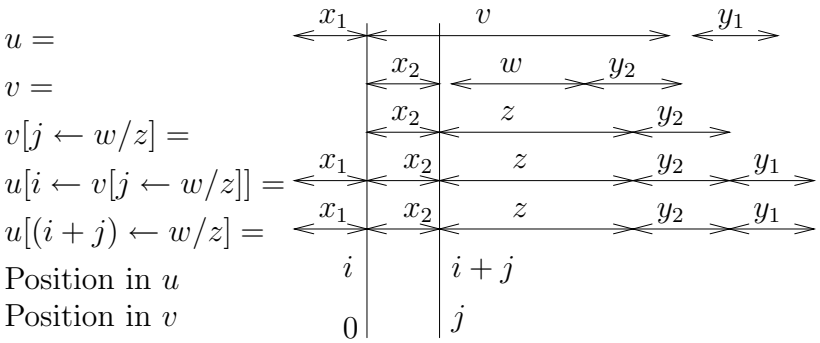
corresponding to this occurrence are $x = \mathbf{123}$ *and* $y = \mathbf{4}$. *Then* $u[3 \leftarrow v/w] = xwy = \mathbf{123.ab.4} = \mathbf{123ab4}$. *This substitution inserted the string* **ab** *in the 3rd position of* u.

3. Now let us evaluate $\mathbf{ABCDE}[0 \leftarrow \mathbf{AB}/\lambda]$. *We identify the strings* $u = \mathbf{ABCDE}$, $v = \mathbf{AB}$, $w = \lambda$, *and the occurrence* $< u, 0, v >$. *The values of* x *and* y *are* λ *and* **CDE**, *respectively. Then* $u[i \leftarrow v/w] = x.w.y = \lambda.\lambda.\mathbf{CDE} = \mathbf{CDE}$. *So, the substitution delete the first occurrence of* v *in* u.

Frequently, we do substitute inside a string v that occurs in a larger string u. We can do this in two ways. We can substitute in v first and then replace the occurrence of v in u, or we can do the substitution directly in u. The next proposition tells us that the result is the same.

Proposition 1.6.21 *Let* $< u, i, v >$ *and* $< v, j, w >$ *be two occurrences, and* z *a string. Then* $u[i + j \leftarrow z] = u[i \leftarrow v/v[j \leftarrow z]]$.

Proof: Before we go on with the proof let us draw the occurrences $< u, i, v >$ and $< v, j, w >$, and the strings $v[j \leftarrow z]$, $u[i + j \leftarrow z]$, and $u[i \leftarrow v/v[j \leftarrow z]]$.



Now let us go on with the proof. The occurrence $< u, i, v >$ tells us that $u = x_1 v y_1$ and $|x_1| = i$. The occurrence $< v, j, w >$ tells us that $v = x_2 w y_2$ and $|x_2| = j$. Now, $v[j \leftarrow w/z] = x_2 z y_2$. So, $u[i \leftarrow v/v[j \leftarrow z]] = u[i \leftarrow v/x_2 z y_2] = x_1 x_2 z y_2 y_1$.

We also know, from Proposition 1.6.17, that $< u, i+j, w >$ is an occurrence. Since $v = x_2 w y_2$, $u = x_1 v y_1 = x_1 x_2 w y_2 y_1$ and $|x_1 x_2| = |x_1| + |x_2| = i + j$. So, the x and the y that correspond to this occurrence are $x = x_1 x_2$ and $y = y_2 y_1$. Then $u[(i+j) \leftarrow w/z] = xzy = x_1 x_2 z y_2 y_1$. So, $u[i+j \leftarrow z] = u[i \leftarrow v/v[j \leftarrow z]]$. **Q.E.D.**

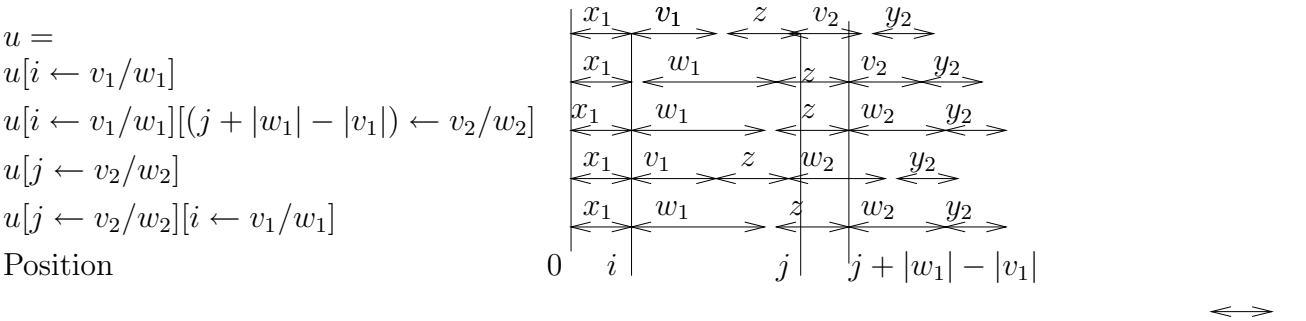
On many occasions we do a sequence of substitions. We substitute the occurrence o_1 in the string u_1 and we get the string u_2. In u_2 we substitute the occurrence o_2 and obtain u_3, and so on. Proposition 1.6.22 tells us that whenever o_1 and o_2 are non-overlapping, we can swith the order of the substitutions without changing the result.

Proposition 1.6.22 *If* $o_1 =< u, i, v_1 >$ *and* $o_2 =< u, j, v_2 >$ *are two non-overlapping occurrences and* o_1 *precedes* o_2, *then* $u[i \leftarrow v_1/w_1][(j+|w_1|-|v_1|) \leftarrow v_2/w_2] = u[j \leftarrow v_2/w_2][i \leftarrow v_1/w_1]$

Proof: Let us look at the decompositions of u that correspond to these two occurrences. We have $u = x_1v_1y_1$, $u = x_2v_2y_2$ and $|x_1| = i$, $|x_2| = j$. Since, o_1 precedes o_2, $i < j$. The non-overlapping of the occurrences together with $i < j$ give $i + |v_1| \leq j$.

Since both x_1v_1 and x_2 are prefixes of u, and the length of x_1v_1 is less than or equal to the length of x_2, x_1v_1 is a prefix of x_2, i.e. $x_2 = x_1v_1z$.

We replace x_2 by x_1v_1z in $u = x_2v_2y_2$ and get $u = x_1v_1zv_2y_2$. Before we go on with the proof let us draw the u and the results of the 4 substitutions.



Now $u[i \leftarrow v_1/w_1] = x_1w_1zv_2y_2$. The address of v_2 is no longer j, but $|x_1|+|w_1|+|z| = |x_1|+|v_1|+|z|+|w_1|-|v_1| = j+|w_1|-|v_1|$. So, the occurrence $< u, j, v_2 >$ became the occurrence $< u[i \leftarrow v_1/w_1, j + |w_1| - |v_1|, v_2 >$. We replace this occurrence in $u[i \leftarrow v_1/w_1]$ and get $u[i \leftarrow v_1/w_1][(j+|w_1|-|v_1|) \leftarrow v_2/w_2] = x_1w_1zw_2y_2$.

The other sequence is even easier. The result of replacing the second occurrence is $u[j \leftarrow v_2/w_2 = x_1v_1zw_2y_2$. The occurrence $< u, i, v_1 >$ becomes the occurrence $< u[j \leftarrow v_2/w_2, i, v_1 >$. We do the second substitution and get the string $u[j \leftarrow v_2/w_2][i \leftarrow v_1/w_1] = x_1w_1zw_2y_2$.

Since both cases yielded the same string, $x_1w_1zw_2y_2$, so we conclude that the order is not relevant. **Q.E.D.**

If the occurrences $< u, i, v_1 >$ and $< u, j, v_2 >$ are overlapping, then the order of the substitutions is relevant, as shown below.

Let $u = \mathbf{abcd}$, $v_1 = \mathbf{bc}$, $w_1 = \mathbf{ab}$, $v_2 = \mathbf{cd}$, and $w_2 = \mathbf{bab}$. If we replace the occurrence $< u, 1, v_1 >$ we get $u[1 \leftarrow \mathbf{bc/ab}] = \mathbf{aabd}$ and the occurrence of $v_2 = \mathbf{cd}$ disappeared.If we replace the occurrence $< u, 2, v_2 >$, we get the string $u[1 \leftarrow \mathbf{cd/bab}] = \mathbf{abbab}$. This time, the occurrence of $v_1 = \mathbf{ab}$ disappeared. Since the occurrences overlapped, they interfered with each other and the results are not the same.

We observe that we need to specify the occurrence. If we just say, replace the string **aa** in **aaab** by **c**, then we get two results, **cab**, and **acb**. The first was obtained by replacing the first occurrence, and the last by changing the second occurrence.

At times we may have to replace a specified set of occurrences, or may be all. Then we will write $u[v/w]$ for the result of substituting all those occurrences of v by w. If the occurrences are non-overlapping, we can perform the substitutions

in any desired order. In most cases, the string v will be a symbol, and the symbol occurrences are always non-overlapping.

Next, we will define 3 orderings that will be useful later on. Let A be a non-empty set of symbols, finite or infinite, and let $>$ be a total order on A.

Definition 1.6.23 (lexicographical order) *On A^* we define the lexicographical order induced by $>$, as $u >_{lex} v$ if*

1. $u = vw$ and $|w| > 0$, or

2. $u = xby$, $v = xaz$, $a, b \in A$, and $b > a$.

We notice that whenever $u_{lex}v$ by the first choice of Definition 1.6.23, v is shorter than u, i.e. $|u| > |v|$.

Example 1.6.24 *Let $A = \{\mathbf{a}, \mathbf{b}\}$, and let us assume that $\mathbf{b} > \mathbf{a}$. Then* $\mathbf{abbb} >_{lex} \mathbf{ab}$ *because* $\mathbf{abbb} = \mathbf{ab}.\mathbf{bb}$ *and* $|\mathbf{bb}| > 0$.

We also have $\mathbf{abbab} >_{lex} \mathbf{ababbb}$ *because for* $x = \mathbf{ab}$ *,*$y = \mathbf{ab}$*, and* $z = \mathbf{bbb}$*, we have* $u = x\mathbf{b}y$*,* $v = x\mathbf{a}z$ *and* $\mathbf{b} > \mathbf{a}$.

We will show that $>_{lex}$ is a total order on A^*. We will show that $>_{lex}$ is irreflexive, transitive, and total.

Lemma 1.6.25 *The lexicographical order is irreflexive.*

Proof: If $u >_{lex} v$ by the first choice of Definition 1.6.23, then $|u| > |v|$, so $u \neq v$.

Let us assume that $u >_{lex} v$ by the second choice, but $u = v$. Then there are strings x, y, z and symbols $\mathbf{a}$ and $\mathbf{b}$ such that $u = xby$, $v = xaz$,$a, b \in A$, and $b > a$. Since $|x\mathbf{a}| = |y\mathbf{b}$ and $u = y$ we can apply Proposition 1.6.5 and get $x\mathbf{a} = x\mathbf{b}$. Now we use the fact that the concatenation is one to one and get $\mathbf{a} = \mathbf{b}$. But this is impossible, since $\mathbf{b} > \mathbf{a}$ and $>$ is irreflexive.**Q.E.D.**

Lemma 1.6.26 *The lexicographical order is total.*

Proof: Let $u \neq v$ be two strings over A. Let us look at the set of all positions k that satisfy the the relation

$k \leq |u|$, $k \leq |v|$, and for all i, $\leq i < k$, $u(i) = v(i)$. (1)

This relation tells us that u and v are identical up to k. The set of all such m's is not empty since 0 satisfies this condition. At the same time the set of these m's is *finite*. So let p be the largest such m. Since p is the largest number that satisfies (1), one of these conditions must be true.

$p = |u|$ (2)

$p = |v|$ (3)

$u(p) \neq v(p)$ (4)

It is easy to see why p must satisfy one of the conditions (2), (3), or (4). For, if all of them are false, $p + 1 \leq |u|$, $p + 1 \leq |v|$, and for all $0 \leq i < p + 1$, $u(i) = v(i)$. So $p + 1$ also satisfies condition (1), contradicting the fact that p is the largest such number. Intuitively, p is the smallest position where u and v are not equal. From the definition of p we get that $[p] \uparrow u = [p] \uparrow v$.

If $p = |u|$, then $[p] \uparrow u = u$, so u is prefix of v, i.e. $v = uw$ for some string w. If $w = \lambda$, $v = uw = u\lambda = u$. But we assumed that $u \neq v$. So, x is non-empty, i.e. $|w| > 0$. Then $v >_{lex} u$ by the first choice of Definition 1.6.23.

By using a similar argument we get that $p = |v|$ implies $u >_{lex} v$.

Let us now assume that p violates condition (4). Since $p < |u|$ and $p < |v|$, both u and v have prefixes with $p + 1$ symbols.Let u_1 and v_1 be these prefixes. The String Decomposition Theorem tells us that $u = u_1 y$ and $v = v_1 z$. Both u_1 and u_2 have prefixes of length p and they are equal because $[p] \uparrow u_1 = [p] \uparrow u = [p] \uparrow v = [p] \uparrow v_1$. Let $x = [p] \uparrow u$. The String Decomposition Theorem tells us that $u_1 = xu(p)$ and $v_1 = xv(p)$. So, $u = u_1 y = xu(p)y$ and $v = xv(p)z$. Since $u(p) \neq v(p)$ and $>$ is a total relation, either $u(p) > v(p)$ or $v(p) > u(p)$. In the first case, the second choice of Definition 1.6.23 tells us that $u > v$. In the second case, we apply the same rule to get $v > u$. **Q.E.D.**

Lemma 1.6.27 *The lexicographical order is transitive.*

Proof: Let us assume that $u >_{lex} v$ and $v >_{lex} w$. Here we have 4 cases, depending on whether the two choices that generated the 2 relations. Let us look at these cases.

Case 1. Both $u_{lex}v$ and $v >_{lex} w$ are obtained by the first choice of Definition 1.6.23. Then $u = vx$, $v = wy$, $|x| > 0$, and $|y| > 0$. So, $u = vx = wyx$ and $|yx| = |x| + |y| > |x| > 0$. So $u >_{lex} w$ by choice 1.

Case 2. $u_{lex}v$ by choice 1 and $v >_{lex} w$ by choice 2. Then

(1) $u = vs$

(2) $|s| > 0$

(3) $v = xby$

(4) $w = xay$

(5) $b > a$

Then $u = vs = xbys$. So $u >_{lex} w$ by the second choice of the definition.

Case 3. $u_{lex}v$ by choice 2 and $v >_{lex} w$ by choice 1. We get the relations below.

(6) $u = xby$

(7) $v = xaz$

(8) $b > a$

(9) $v = ws$

(10)$|s| > 0$

Now we have 2 cases depending on whether $|w| \leq |x|$ or $|w| > |x|$.

If $|w| \leq |x|$ then w is a prefix of x, so $x = wt$ for some string t. Then $u = xby = wtby$, and $|tby| \geq |b| = 1$. So, $u >_{lex} w$ by the first choice of Definition 1.6.23.

If $|w| > |x|$, then $|xa| \leq |w|$, so xa is a prefix of w. Then $w = xas$ for some string s. But, this implies that $u >_{lex} w$ by the second choice of Definition 1.6.23.

Case 4. Both $u_{lex}v$ and $v >_{lex} w$ are obtained by choice 2. We have the relations below.

(11) $u = x_1 d y_1$

(12) $v = x_1 c z_1$

(13) $d > c$
(14) $v = x_2 b y_2$
(15) $w = x_2 a z_2$
(16) $b > a$

We have 3 cases, depending on whether $|x_1| < |x_2|$, $|x_1| = |x_2|$, or $|x_1| > |x_2|$.

Subcase 1. Let us assume that $|x_1| < |x_2|$. Since both $x_1 c$ and x_2 are prefixes of v, and $|x_1 c| \leq |x_2|$, $x_2 = x_1 t$ for some string t. Then $w = x_2 a z_2 = x_1 c t a z_2$. Then $u >_{lex} w$ by the second choice of the definition.

Subcase 2. If $|x_1| = |x_2|$, then $x_1 = x_2$ because both are prefixes of v. So, $v = x_1 c z_1 = x_1 b y_2$. From the monotonicity of the concatenation we get $b = c$. Since $d > c = b > a$, $d > a$ because $>$ is transitive. At the same time $w = x_2 a z_2 = x_1 a z_2$. If we compare this equality with (11), we see that $u >_{lex} w$ by the second choice of the definition.

Subcase 3. Now let us assume that $|x_1| > |x_2|$. Then $x_2 b$ is a prefix of x_1, so $x_1 = x_2 b r$ for some string r. We substitute x_1 in (11) and get $u = x_2 b r d y_1$. When we compare this equality with (11) we see that $u >_{lex} w$ by the second choice of Definition 1.6.23. **Q.E.D.**

Now we ask if the lexicographic order is well founded. The answer depends on the number of symbols in A. If A has only one symbol, say $\mathbf{a}$, then $A^* = \mathbf{a}^n | n \geq 0$. The lexicographic order reduces to choice 1, and becomes $\mathbf{a}^m >_{lex} \mathbf{a}^n$ ii and only if $m > n$. Since $>$ is well founded on N, $>_{lex}$ is well founded on A^*.

Now let us look at the case when A has 2 or more elements. Since $>$ is total on A, there must be two elements $\mathbf{a}$ and $\mathbf{b}$ in A, such that $\mathbf{b} > \mathbf{a}$. Now let us look at the set $S = \{\mathbf{b}^n \mathbf{a} | n \in N\}$. In this set we have the infinitely descending chain below.

$\mathbf{b} >_{lex} \mathbf{ab} >_{lex} \ldots >_{lex} \mathbf{b}^n bfa >_{lex} \ldots$

So, S does not have a minimal element. Hence, it is not a well founded order.

The next theorem summarizes our results about the lexicographic order.

Theorem 1.6.28 *Let A be a set of symbols and $>$ a total order on A. Then the lexicographic order on A^* is a total order. If A has only one element, it is a well order; otherwise it is not. If A has a single element, then $>$ is total; otherwise it is not.*

Now let us define a simpler ordering on the set of strings A^*, *the length ordering.*

We say that $u > v$ iff $|u| > |v|$, i.e. the string u is longer than the string v.

Theorem 1.6.29 *The length ordering is a well founded order on A^*.*

Proof: We will show that $>$ is transitive and every non-empty set of strings has minimal elements.

If $u > v$ and $v > w$, then $|u| > |v|$ and $|v| > |w|$. So, $|u| > |w|$, i.e. $u > v$.

Now let S be a non-empty set of strings. Then the set of the lengths of the strings has a least element, say n, because the set of natural numbers is well ordered by $>$. Let T be the set of all strings of S that have length n. We will

show that any $u \in T$ is a minimal element. Let us assume that there is some $v \in S$ such that $u > v$. Then $|u| = n > |v|$, so n is no longer the least element of the lengths of the strings in S. We got a contradiction. So, S has minimal elements. Since S is arbitrary, A^* is well founded.

If A has a single element, then the length ordering coincide with the lexicographic order, so it is total. Now let us assume that A has more than one element. Let **a** and **b** be two distinct symbols in A. Then $u = \mathbf{ab}$ and $v = \mathbf{ba}$ have the same length, but they are not equal. So, the totality requirement that $u > v$ or $u = v$ or $v > a$ is not satisfied.**Q.E.D.**

The third order combines the length and the lexicographic orders.

Definition 1.6.30 *Let A be a set of symbols and $>$ be a well order on A. We say that $u \succ v$ if either $|u| > |v|$, or $|u| = |v|$ and $u_{lex}v$.*

Remark 1.6.31 *If A is finite then any total order is a well order.*

Let us list the elements of $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}^*$ according to this order. We assume that $\mathbf{c} > \mathbf{b} > \mathbf{a}$.

λ,**a**,**b**, **c**, **aa**,**ab**, **ac**, **ba**, **bb**, **bc**, **ca**, **cb**, **cc**, ...

The first element is the element of length 0, then we list the elements of length 1 in lexicographic order, then the elements of length 2 in lexicographic order, and so on.

Theorem 1.6.32 *The ordering $\succ$ is a well order on A^*.*

Proof: We need to show that every no-nempty set of strings of A^* has a least element. Let S be a non-empty set of strings. Let $L = |u| : u \in S$, i.e., L is the set of the lengths of the strings in S. Since $L \subseteq N$, L has a least element, say n. Now let T be the set of all strings in S that have length n. T is not empty since there are strings of length n in S. Now we will find the least element of $w \in T$ under the lexicographic order. If A is finite, then T is finite, and we can take w to be the smallest element of T in lexicographic order. The problem is what do we do when A is infinite? In this case we construct $w = w_0w_1...w_{n-1}$ recursively as described below.

The symbol w_0 is the least $a \in A$ such that a is a prefix of a string in $T_0 = T$. Such a symbol exists since A is well ordered, and T is not empty. Let T_1 be the set of all strings in T that have prefix w_0. This set is not empty.

Now we define w_{i+1} to be the least $a \in A$ such that $w_0...w_ia$ is a prefix of a string in T_{i+1}. Since T_i, the set of all strings with prefix $w_0 \dots w_i$, is not empty and A is well ordered by $>$, such an a exists. Moreover, the set T_{i+2}, the set of all strings in T_{i+1} that have prefix $w_0 \dots w_{i+1}$ is not empty since at leats one string in T_{i+1} has prefix $w_0 \dots w_{i+1}$.

Now we will show that w is the least element in S. Let $\in S$. Then $|v|geq|w|$, since w has the shortest length of all strings in S. If $|v| > |w|$ then $v \succ w$. So, let us assume that $|v| = |w|$, i.e. $v \in T$. Since the lexicographic order is total, either $v_{lex}w$, or $v = w$, or $w_{lex}v$. We will show that $w >_{lex} v$ is not possible. Assume that it is. Since v and w have the same length, the realtion

was obtained from the second choice of the definition. So, $w = xby$, $v = xaz$, and $b > a$. Let $|x| = i$. From the construction of w, we get that $v \in T_i$. But this means that at step i we did not choose the least w_i. This contradicts the well ordering of A. So, for all $v \in T$, $v = w$ or $v >_{lex} w$, i.e. $v = w$ or $v \succ w$.

The last part of the condition of the least element definition, that there is no $v \in S$ such that $w \succ v$ was already implicitely proved. The condition $w > v$ is impossible when w is longer than v. If the lengths are equal, then $w \succ v$ holds if and only if $w >_{lex} v$ hold s in T, and we showed that this is impossible. **Q.E.D.**

1.6.1 Exercises

Exercise 1.6.1 *Find all prefixes and all suffixes of* **12345**.

Exercise 1.6.2 *List all substrings of* **1234**.

Exercise 1.6.3 *Let u be a string. Are there any substrings that are both prefixes and suffixes of u? Are any of them proper substrings?*

Exercise 1.6.4 *Let $\succeq$ be the relation defined on the set of strings over A^* by $u \succeq v$ if v is a prefix of u. Show that $\succeq$ is reflexive, antisymmetric and transitive.*

Exercise 1.6.5 *Let $\preceq$ be the relation defined on the set of strings over A^* by $u \preceq v$ if u is a suffix of v. Show that $\preceq$ is reflexive, antisymmetric and transitive.*

Exercise 1.6.6 *Prove that a substring of a substring of u is a substring of u.*

Exercise 1.6.7 *Let u is a non-empty string over A and i, j be two indices such that $0 \leq i \leq j < |u|$. We define the string $u[i:j] : [j-i+1] \longrightarrow A$ by $u[i:j](k) = u(k+i)$ for all $0 \leq k < j-i+1$. We call $u[i:j]$ the slice of u that starts at i and ends at j. Use the definition of concatenation to show that whenever $0 < i < |u| - 1$, $u[0:i-1]u[i:|u|-1] = u$.*

Exercise 1.6.8 *Let $v = u[i:j]$ be a slice of u and $w = v[l:m]$ be a slice of v. Show that w is a slice of u.*

Exercise 1.6.9 *Use the two preceding exercises to show that $u = u[0:i-1]u[i,j]u[j+1:|u|-1]$ for all i, j, satisfying the inequality $0 < i \leq j < |u| - 1$.*

Exercise 1.6.10 *Show that v is a non-empty substring of u iff v is a slice of u.*

Exercise 1.6.11 *Give a sufficient conditions in terms of i and/or j for $u[i:j]$ to be a prefix/a suffix of u.*

Exercise 1.6.12 *Show that the occurrence $< u, i, v >$ contains the occurrence $< u, j, w >$ iff $< v, j-i, w >$ is also an occurrence.*

Exercise 1.6.13 *Let x be a string and $Occ(x)$ be the set of all occurrences $< u, j, v >$ with $u = x$. Specify if the relations non-overlap, contains, and precedes restricted to $Occ(x)$ are reflexive, symmetric, antisymmetric, transitive. Prove your assertions.*

Exercise 1.6.14 *We say that two occurrences $o_1 =< u, i, v >$ and $o_2 =< u, j, w >$ overlap if the relation o_1 and o_2 do not overlap is false. Define the overlap relation in terms of i,j,v, and w. Then specify if the reflexive, symmetric, antisymmetric, and transitive properties are valid for this relation.*

1.7 Graphs

Definition 1.7.1 (digraphs) *A directed graph or a digraph is a pair $G =< V, A >$ where V is a set of* **vertices** *or* **nodes** *and $A \subseteq (V \times V - \{< v, v > | v \epsilon V\})$ is a set of* **arcs**. *If $< a, b >$ is an arc of G, we say that a is the* **source** *and b is the* **target** *of the arc.*

Example 1.7.2 *The set $G =< \{a, b, c, d\}, \{< a, b >, < a, c >, < d, a >, < b, c >, < c, d >\} >$ is a graph. The vertices are a, b, c, d and the arcs are $< a, b >$, $< a, c >$, $< d, a >$, $< b, c >$, $< c, d >$. The first vertex of each arc is its source and the second its target.*

In many cases we represent a graph by a picture. The vertices are boxes of different shapes (circles, squares, rectangles, ovals, ellipses) and the arcs as arrows from the source node to the target node. Figure 1.12 represents the graph from Example 1.7.2.

Figure 1.12: Picture of a graph

In many cases we are not interested in the set of vertices, so we will not write their names. Figure 1.13 shows the graph from Figure 1.12 without the names of the vertices.

Figure 1.13: A graph without vertex names

Definition 1.7.3 (walk, cycle, path) *Let $G =< V, A >$ be a graph and $< v_0, ..., v_n >$ be a sequence of vertices such that for all i, $0 \leq i \leq (n-1)$, $< v_i, v_{i+1} >$ is an arc in G. Then we say that the sequence $< v_0, ..., v_n >$ is a* **walk** *from v_0 to v_n of length n.*

If $v_0 = v_n$ then we say that the walk is a **cycle**.

If no arc is repeated then the walk is called a **path**.

Example 1.7.4 *The sequence $< a, b, c, d, a, b >$ from Figure 1.12 is a walk from a to b because $< a, b >$, $< b, c >$, $< c, d >$, $< d, a >$, $< a, b >$ are arcs in G. It is not a path since the arc $< a, b >$ is repeated.*

The walk $< a, b, c, d, a >$ is a path and a cycle. It is a cycle because the two end vertices are the same. It is a path because no arc is repeated.

The walk $< b, c, a >$ is a path from b to a, because no arc is repeated.

Definition 1.7.5 (In-degree, out-degree) *Let $G =< V, A >$ be a graph and v a vertex of G. The in-degree of v is the number of arcs that have v as target. The out-degree of v is the number of arcs that have v as source.*

Example 1.7.6 *Let us find the in and out degrees of node a from Figure 1.12. The in-degree of a is 1 because there is only one arrow pointing to a, namely $< d, a >$. The out-degree of a is 2 because there are two arrow starting from a, $< a, b >$ and $< a, c >$.*

Definition 1.7.7 (labeled graph) *A labeled graph is a graph $G =< V, A >$ together with two labeling functions $f_v : V \longrightarrow L_v$ and $f_a : A \longrightarrow L_a$. The set L_v is the set of vertex labels, L_a is the set of arc labels, f_v is the labeling function for vertices, and f_a is the labeling function for the arcs. The labelings can be partial functions.*

Example 1.7.8 *The figure below shows a labeling of Figure 1.12. The set of vertex labels is $\{1, 2, 3\}$ and the arc labels are α, β, γ.*

The labels are put inside the circles, and the vertex names are writen next to node. We notice that not all vertices and arcs have labels; the ones that have blanks in the label field are unlabeled.

Figure 1.14: Labeled graph

Many times we are not concerned with the vertex names, but with the labels. In those cases we will skip the names and preserve the labels. Figure 1.15 shows the above graph without the vextex names.

Figure 1.15: Labeled graph without vertex names

Definition 1.7.9 (tree) *A tree is a digraph that has a vertex, named the root, such that*

1. *for all other vertices v there is a path from the root to v,*
2. *the in-degree of the root is 0,*
3. *the in-degree of all other vertices is 1.*

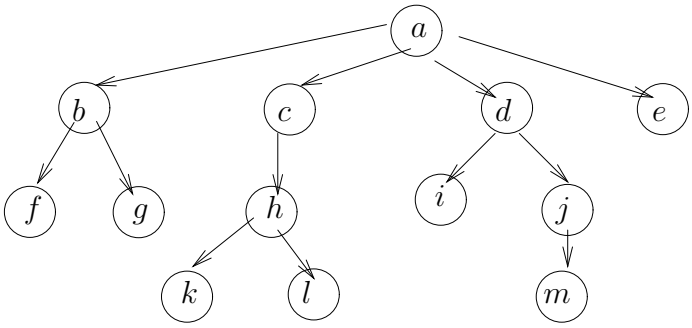


Figure 1.16: A tree

In a tree there is a unique path from the root to any other vertex. The proof of this fact is left as exercise.

Definition 1.7.10 (tree height) *The height of a tree is the length of a longest path in the tree.*

Definition 1.7.11 (branch, leaf, parent, child, descendent, ancestor) *A branch is any vertex with out-degree greater than 0; a leaf is any vertex with out-degree 0.*

If $< u, v >$ is an arc in the tree then u is the parent of v and v is a child of u.

If there is a path from u to v we say that v is a descendent of u and u is an ancestor of v.

Let us illustrate the concepts from Definition 1.7.11 with examples from Figure 1.16.

Vertices f, g, k, l, i,m, e are leaves because their out-degree is 0; the rest of the vertices are branches. The children of a are b, c, d, and e. The descendents of c are h, k, and l. The ancestors of m are all vertices on the path from the root to m, excluding m. So, its ancestors are a,d, and j. The vertex j is also the parent of m.

Definition 1.7.12 (oriented tree) *An* **oriented tree** *is a tree whose arcs are numbered with whole integers as follows:*

if u has n children, say v_1,...,v_n, then the arcs $< u, v_1 >$, $< u, v_2 >$,..., $< u, v_n >$ are labeled respectively 1, 2, ..., n.

Our book deals with oriented trees. In most cases we will not write the labels next to the arc; we will assume that the leftmost arc is labeled 1, the one immediately to its right has label 2, and so on.

Definition 1.7.13 (The Dewey Addressing System) *Let T be an oriented tree. The Dewey numbering system assigns to each vertex of the tree a string over $\{0, 1, 2, \ldots\}$, as follows:*

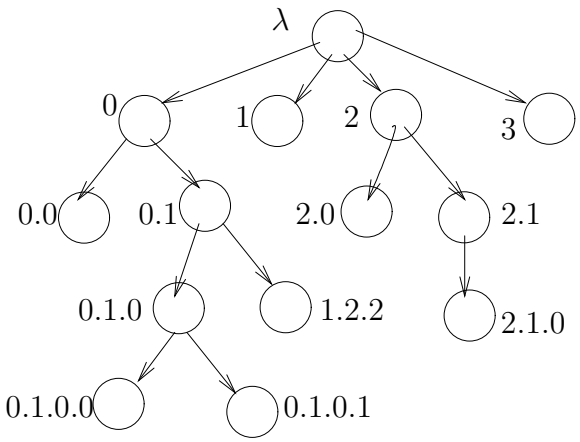


Figure 1.17: Dewey labeling

a. the root gets the empty-string λ.

b. if the vertex u *receives the string* s*, and* v *is the* i*-th child of* u*, then* v *is receives the string* $s.(i-1)$*, the dot being the concatenation operation described in Section* **??**.

The string attached to the vertex is its address. The set of all addresses is the domain of the tree.

Since the alphabet $\{0, 1, 2, 3, \ldots\}$ is infinite, we will use the concatenation operator . to separate the symbols of the strings. So, they serve as markers that allow us to interpret the string. For example, 1.2.3 is the address with 3 symbols, 1, 2, and 3. If we eliminate the periods, then we introduce ambiguity. The same string 123 can represent the addresses 123, 1.23, 12.3, or 1.2.3.

Figure 1.17 shows The Dewey addressing system of a tree. The address of the root is the empty string; its four children have addresses 1, 2, 3, and 4. The addresses of the grandchildren, 1.1, 1.2, 3.1, and 3.2, have length 2. Each address indicate the path from the root to that vertex. For example, to get to the node 1.2.1.2 we must do the following steps:

1. Take the first child of the root to get to node 1.
2. Take the second child of 1 to reach vertex 1.2.
3. Take the first child of 1.2 to arrive at 1.2.1.
4. Take the second child of 1.2.1 and we are at 1.2.1.2.

The tree domain is the set

$\{\lambda, 0, 1, 2, 3, 0.0, 0.1, 2.0, 2.1, 0.1.0, 0.1.1, 2.1.0, 0.1.0.0, 0.1.0.1\}$.

Observation 1.7.14 *Let* $D(T)$ *be the domain of the tree* T*, and* $s.i$ *an address of* T*, with* $i \in N$*. Then*

1. s *is also in* $D(T)$*, and*

2. if $i > 0$ *then* $s.(i-1)$ *is also in* $D(T)$.

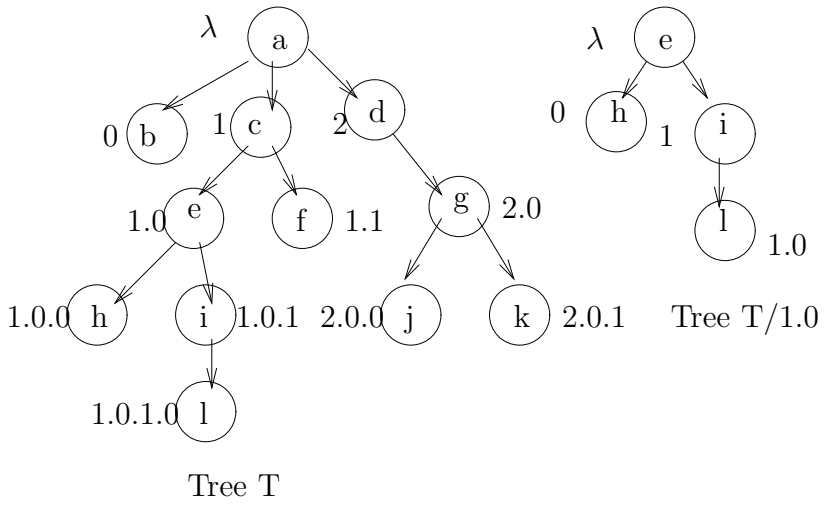


Figure 1.18: Dewey labeling

Definition 1.7.15 (subtrees) *Let T be a tree and $D(T)$ be the domain of T. Let u be an address in $D(T)$. The subtree of T at address u, writen T/u is the tree with domain $D = \{v | u.v \in D(T)\}$, where . is the concatenation operation defined in Section* **??**.

T/u inherits the labeling of T. So, the address v of T/u has the same label as the vertex $u.v$ of T, and the labels of the arcs $v \longrightarrow v.i$ of T/u and $u.v \longrightarrow u.v.i$ of T are identical.

If length of u is equal to 1, i.e. u is a child of the root, then T/u is called a main subtree.

Example 1.7.16 *Figure 1.18 shows the tree T together with the substree $T/1.0$. The root of $T/1.0$ is vertex 1.0 of T and its nodes are the descendents of 1.0. The $T/1.0$ addresses are obtained from the corresponding T addresses by deleting the prefix 1.0.*

Definition 1.7.17 (k-ary tree, full k-ary trees) *Let k be a whole number greater than 0. In a k-ary tree all nodes have at most k children.*

A k-ary tree is full if all nodes have either k children, or no children.

The tree from Figure 1.16 is quaternary (4-ary), since all vertices have an out-degree of 4 or less. It is not full since some vertices have 1 or 2 children. The tree in Figure 1.18 is ternary since the largest out-degree is 3. It is not full since some branches have 1 or 2 children. However, Figure 1.19 shows a full binary tree. A *reversed tree* is obtained from a tree by reversing the direction of the arrows.

Definition 1.7.18 (reversed tree) *A reversed tree is a digraph that has a vertex, named the root, such that*

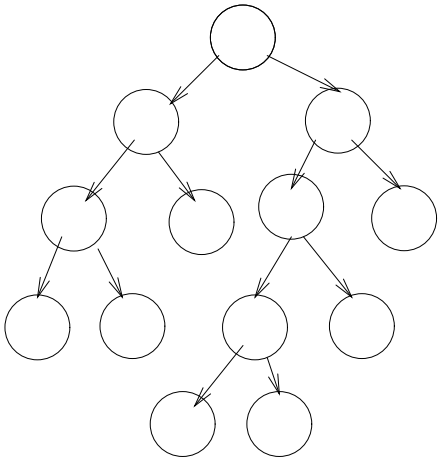


Figure 1.19: A full binary tree

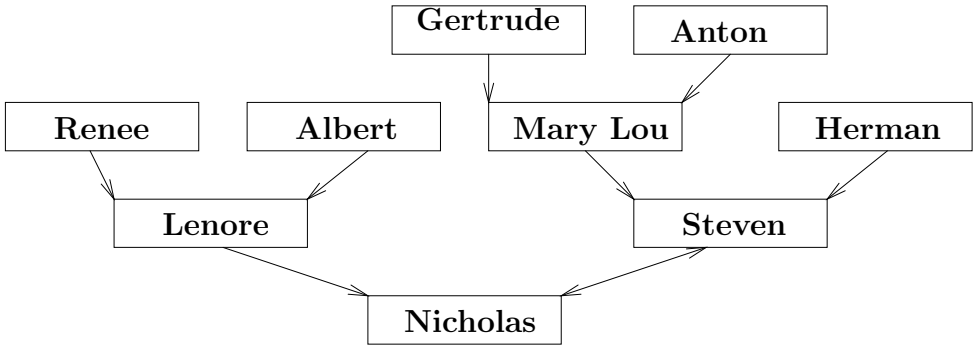


Figure 1.20: A pedegree tree

1. *for all other vertices v there is a path from v to the root,*
2. *the out-degree of the root is 0,*
3. *the out-degree of all other vertices is 1.*

The pedegree tree from Figure 1.20 is a reversed tree.

Definition 1.7.19 (parents, children, ancestors, descendents in a reversed tree) *Let T be a reversed tree.*

If a vertex has in-degree 0, i.e. no arrow points to it, then it is a leaf.
If a vertex has in-degree greater than 0, then it is called a branch.
If a ¡u,v¿ is an arc in T then u is a parent of v and v is the child of u.
If there is path from u to v, u is an ancestor of v and v is a descendent of u.

In Figure 1.20, the parents of **Nicholas** are **Lenore** and **Steven**. His ancestors, besides his parents, are **Renee**, **Albert**, **Mary Lou**, **Gertrude**, **Anton**,

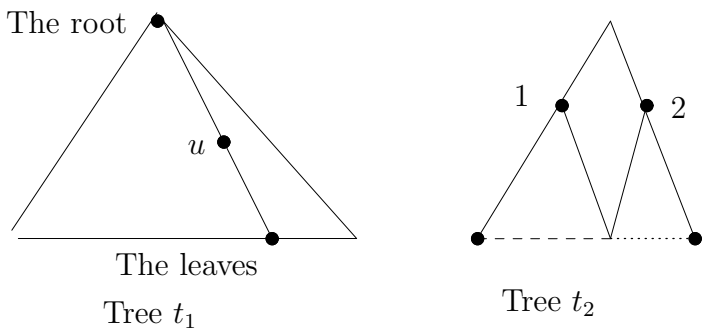


Figure 1.21: Tree representations

and **Herman**. The descendends of **Gertrude** are **Mary Lou**, **Steven**, and **Nicholas**. The vertices **Renee**, **Albert**, **Gertrude**, **Anton**, and **Herman** are the leaves. The rest of the vertices are branches.

We can modify the definitions of the oriented trees, k-ary trees and full k-ary to fit the reverse trees. The only difference is that we replace **children** by **parents**. So, in an oriented tree we label the parents of each node with consecutive whole numbers starting with 1. In a k-ary reverse tree every vertex has at most k parents, and in a full k-ary tree every node has no parents or exactly k parents.

The Dewey numbering system also works for the reverse trees. The root receives the address λ. Its parents receive the the sequences 0, 1,...,k, its grandparents get the two digit sequences, and so on.

The substrees are defined in a similar way. If u is an address in the tree T the subtree T/u has u as the root and its vertices are all ancestors of u, together with their connecting arrows.

We will represent a tree as a triangle with the apex on top and the base on the bottom. The apex represents the root and the base the leaves. A path from the root to a leaf is represented as a line segment connecting the apex to a point on the base. An address is a point on a path from the root to a leaf. Figure 1.21 shows the trees t_1 and an address u in t_1. It also shows the main subtrees of t_2, $t_2/1$ and $t_2/2$. The leaves of t are the union of the leaves in $t_2/1$ represented by dash line and the leaves of $t_2/2$ represented the dotted line.

The reversed trees are represented as upside-down triangle. Figure 1.22 shows the revered trees t_1 and t_2 and the address u of t_2. The triangle with the dash lines is the subtree t_2/u of t_2.

Now we can define *the tree substitution.*

Definition 1.7.20 ($T[u \leftarrow t]$) *Let T and t be trees and u an address of T. Then T with t at u, writen $T[u \leftarrow t]$, is the tree with domain $(D(T) - \{v \in D(T) | u$ is a prefix of $v\}) \cup \{u.w | w \in D(t)\}$.*

Figure 1.23 shows the trees T, t and $T[1.0 \leftarrow t]$. The domain of T is

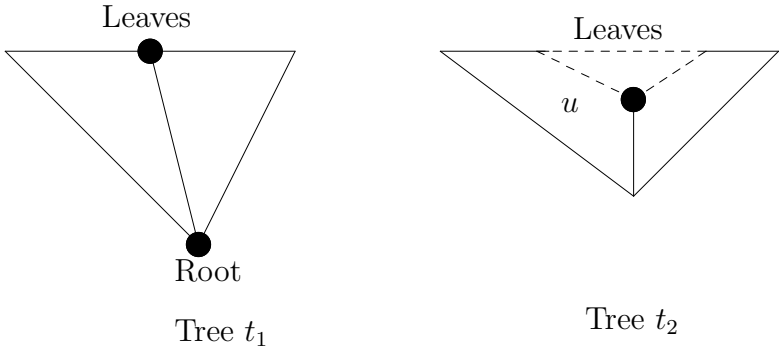


Figure 1.22: Reversed tree representations

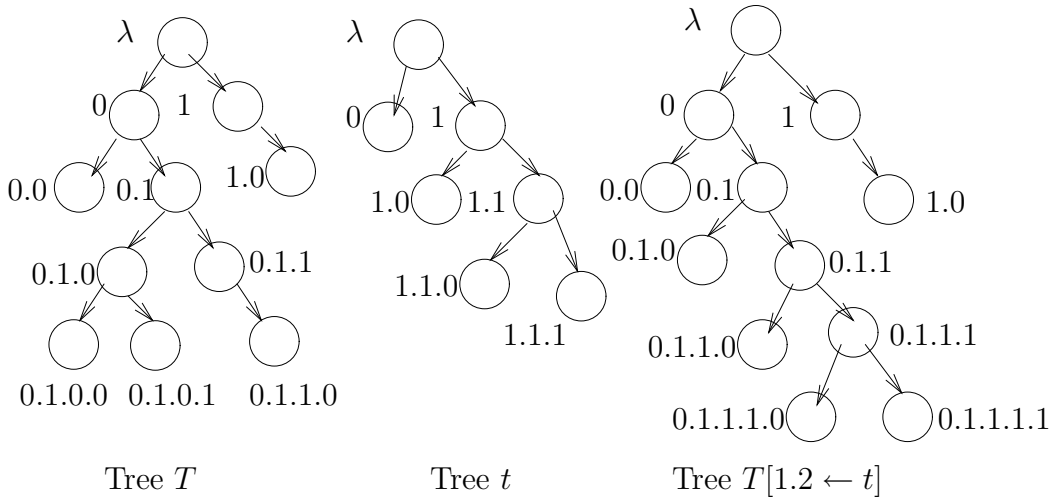


Figure 1.23: Tree substitution

$D(T) = \{\lambda, 0, 1, 0.0, 0.1, 1.0, 0.1.0, 0.1.1, 0.1.0.0, 0.1.0.1, 0.1.1.0\}$.

The domain of t is $D(t) = \{\lambda, 0, 1, 1.0, 1.1, 1.1.0, 1.1.1\}$.

The domain of $T[1.2 \leftarrow t]$ is obtained by removing from $D(T)$ all addresses with prefix u. We get $D(T) - \{v \in D(T) | 1.2 is a prefix of v \in D(T)\} = \{\lambda, 1, 2, 1.1, 2.1\}$. To this set we add the addresses $u.v$ where the v's are the strings in $D(t)$. So, $D(T[1.2 \leftarrow t]) = \{\lambda, 0, 1, 0.0, 1.0\} \cup \{0.1, 0.1.0, 0.1.1, 0.1.1.0, 0.1.1.1, 0.1.1.1.0, 0.1.1.1.1\}$.

We get $D(T[1.2 \leftarrow t]) = \{\lambda, 0, 1, 0.0, 0.1, 0.1, 1.0, 0.1.0, 0.1.1, 0.1.1.0, 0.1.1.1, 0.1.1.1.0, 0.1.1.1.1\}$.

If the trees T and t are labeled, then $T[u \leftarrow t]$ inherits their labelings.

Definition 1.7.21 (labeled trees substitution) *Let T and t be trees and u an address in T. Let f_T and f_t be the labeling functions for T respectively t. Then $T[u \leftarrow t]$ is labeled by the function f defined below.*

$$f(v) = \begin{cases} f_T(v) & \text{if } u \text{ is not a prefix of } v \\ f_t(w) & \text{if } v = uw \end{cases}$$

$$f(< v, v.i >) = \begin{cases} f_T(< v, v.i >) & \text{if } u \text{ is not a prefix of } v \\ f_t(< w, w.i >) & \text{if } v = u.w \end{cases}$$

Figure 1.24 shows the substitution of two labeled trees. The domain of T is $D(T) = \{\lambda, 0, 1, 0.0, 0.1, 1.0\}$. From it we subtract the strings that have prefix 1. We get the set $\{\lambda, 0, 0.0, 0.1\}$. For these nodes, and for the arcs between these nodes, the labels of $T[1 \leftarrow t]$ are the labels of T. So, λ receives the label a, 0 gets b, 0.0 has c , and 0.1 has d. The arrows $< \lambda, 0 >$, $< 0, 0.0 >$, $< 0, 0.1 >$, and $< 0, 1 >$ have the same labels in $T[2 \leftarrow t]$ and T. We notice that even though $u = 1$ is not in the set $D(T) - \{v \in D(T) | \text{u is a prefix of } v\}$, the arrow $< \lambda, 1 >$ receives the label of T because the address λ does not have 1 as a prefix.

The rest of the addresses of $D(T[1 \leftarrow t]$ have prefix 1. They are $\{1, 1.0, 1.1, 1.1.0, 1.1.1\}$. The addresses $1.v$ receive the labels of the corresponding v nodes in t. So, 1 gets A, 1.0 receives B, 1.1 has C, 1.1.0 is labeled C, and 1.1.1 is marked with E. The edges of the form $< u.w, u.w.i >$ get the label of $< w, w.i >$ in t. So, $< 1, 1.1 >$ has the label 0 of $< \lambda, 1 >$ and $< 1.1, 1.1.0 >$ receives the marking 1 of the arrow $< 1, 1.0 >$.

Definition 1.7.22 (disjoint addresses) *Let $D(T)$ be the domain of the tree T. We say that the addresses u and v are disjoint, writen $u \perp v$, if neither string is a prefix of the other.*

There are several pairs of disjoint addresses in tree from Figure 1.25. The addresses of the leaves are pairwise disjoint. So, any pair in the set $\{0.0, 0.1, 2.0, 2.1\}$ is disjoint. The address 0 is disjoint from 1, as well as from 2 and its descendents, 2.0 and 2.1. The address 1 is disjoint from 0, 2, and their descendents, 0.0, 0.1, 2.0, and 2.1. The address 2 is also disjoint from 0, 1, and the descendents of 0, 0.0 and 0.1.

Proposition 1.7.23 (commutative substitutions) *Let T, t_1, and t_2 be tree trees, and u,v be two disjoint addresses in T. Then $T[u \leftarrow t_1][v \leftarrow t_2] = T[v \leftarrow t_2][u \leftarrow t_1]$.*

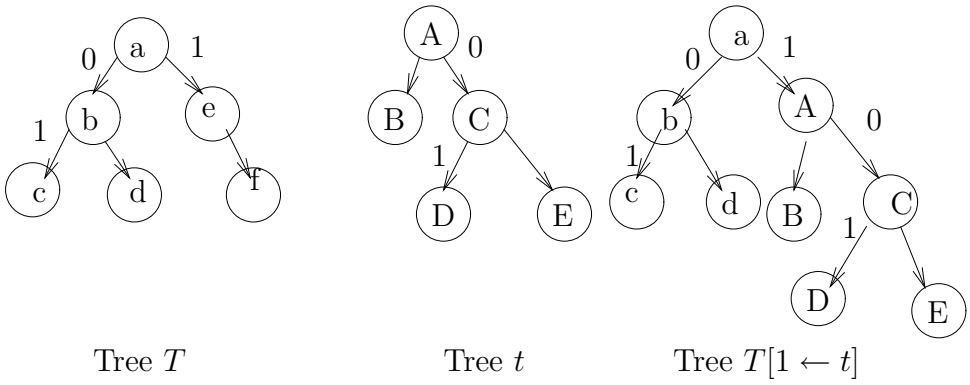


Figure 1.24: Labeled tree substitution

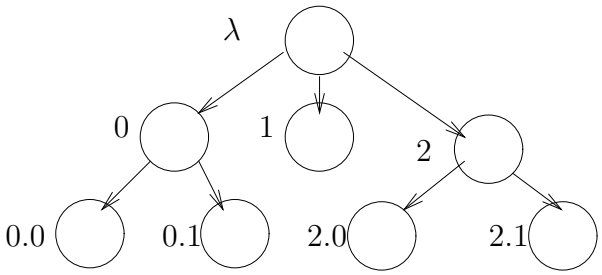


Figure 1.25: Disjoint addresses

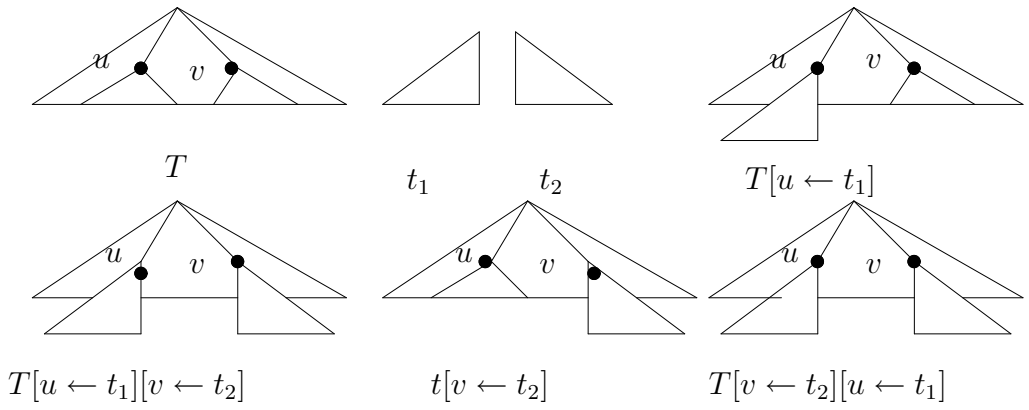


Figure 1.26: Disjoint addresses

Before we go on with the proof, let us draw a picture of what is going on.

The tree t_1 is the right triangle with the hypotenuse to the left and t_2 is the right triangle with the hypotenuse to the right. We see that the two substitutions do not interfere with each other and the end result is the same. Now let us give a formal proof.

Proof: We will show that the trees $T[u \leftarrow t_1][v \leftarrow t_2]$ and $T[v \leftarrow t_2][u \leftarrow t_1]$ have the same domain and the same labeling. The domain of $T[u \leftarrow t_1]$ is $(D(T) - \{w \in D(T) | u \text{ is a prefix of } w\}) \cup \{uw | w \in D(t_1)\}$.

Since u is not a prefix of v, $v \in D(T) - \{w \in D(T) | u \text{ is a prefix of } w\}$ and v is an address in $T[u \leftarrow t_1][v \leftarrow t_2]$. Moreover, the descendents of v in $T[u \leftarrow t_1]$ are precisely the descendents of v in $D(T)$. So,

(1) $\{w \in D(T[u \leftarrow t_1]) | v \text{ is prefix of } w\} = \{w \in D(T) | v \text{ is a prefix of } w\}$.

The domain of $T[u \leftarrow t_1][v \leftarrow t_2]$ is

$(D(T[u \leftarrow t_1]) - \{w \in D(T[u \leftarrow t_1]) | v \text{ is prefix of } w\}) \cup \{vw | w \in D(t_2)\}$

$= ((D(T) - \{w \in D(T) | u \text{ is a prefix of } w\}) \cup \{uw | w \in D(t_1)\}) - \{w \in D(T[u \leftarrow t_1]) | v \text{ is prefix of } w\}) \cup \{vw | w \in D(t_2)\}$

$= ((D(T) - \{w \in D(T) | u \text{ is a prefix of } w\}) \cup \{uw | w \in D(t_1)\}) - \{w \in D(T) | v$ is a prefix of $w\} \cup \{vw | w \in D(t_2)\}$ by (1).

If a string uw has v as a prefix,then u and v are prefixes of the same string, so one of them is a prefix of the other. This contradicts the assumption $u \perp v$. So,

(2) $\{uw | w \in D(t_1)\} \cap \{w \in D(T) | v \text{ is a prefix of } w\} = \phi$.

We use (2) in the expression of $D(T[u \leftarrow t_1][v \leftarrow t_2])$ and get

$D(T[u \leftarrow t_1][v \leftarrow t_2]) = (D(T) - (\{w \in D(T) | u \text{ is a prefix of } u\} \cup \{w \in D(T) | v \text{ is a prefix of } u\})) \cup (\{uw | w \in D(t_1)\} \cup \{vw | w \in D(t_2)\})$.

A similar analysis yields the domain of $T[v \leftarrow t_2][u \leftarrow t_1]$. The domain of $T[v \leftarrow t_2]$ is $(D(T) - \{w \in D(T) | v \text{ is a prefix of } w\}) \cup \{vw | w \in D(t_2)\}$.

If $u.z \in \{w | v \text{ is a prefix of } w\}$ for some z, then v and u are prefixes of the same string, so one must be a prefix of the other. This contradicts $u \perp v$. So,

both u and its descendents are in $D(T) - \{w \in D(T) | v \text{ is a prefix of } w\}$. This tells us that $T[v \leftarrow t_2][u \leftarrow t_1]$ exists and

(3) $\{w \in D(T[v \leftarrow t_2]) | u \text{ is a prefix of } w\} = \{w \in D(T) | u \text{ is a prefix of } w\}$.

Now let us compute $D(T[v \leftarrow t_2][u \leftarrow t_1])$

$D(T[v \leftarrow t_2][u \leftarrow t_1]) = (D(T[v \leftarrow t_2]) - \{w \in D(T[v \leftarrow t_2]) | u \text{ is a prefix of } w\}) \cup \{uw | w \in D(t_1)\}$

$= ((D(T) - \{w \in D(T) | v \text{ is a prefix of } w\}) \cup \{vw | w \in D(t_2)\}) - \{w \in D(T[v \leftarrow t_2]) | u \text{ is a prefix of } w\}) \cup \{uw | w \in D(t_1)\}$

$= ((D(T) - \{w \in D(T) | v \text{ is a prefix of } w\}) \cup \{vw | w \in D(t_2)\}) - \{w \in D(T) | u \text{ is a prefix of } w\}) \cup \{uw | w \in D(t_1)\}$ by (3).

At the time no string of the form vw can have u as a prefix, since $u \perp v$. So,

(4) $\{vw | w \in D(t_2)\} \cap \{w \in D(T) | u \text{ is a prefix of } w\} = \phi$.

With this in mind, the equation of the domain of $T[v \leftarrow t_2][u \leftarrow t_1]$ becomes

$D(T[v \leftarrow t_2][u \leftarrow t_1]) = (D(T) - (\{w \in D(T) | v \text{ is a prefix of } w\} \cup \{w \in D(T) | u \text{ is a prefix of } w\})) \cup (\{vw | w \in D(t_2)\} \cup \{uw | w \in D(t_1)\})$.

So, the domains of the two trees are equal. We must show that the labeling is the same. First of all we observe that the three subsets of the domain of $T[u \leftarrow t_1][v \leftarrow t_2]$, $D(T) - (\{w \in D(T) | v \text{ is a prefix of } w\} \cup \{w \in D(T) | u \text{ is a prefix of } w\})$, $\{vw | w \in D(t_2)\}$, and $\{uw | w \in D(t_1)\}$ are disjoint for the following reasons: the first set has no addresses with prefixes u or v, all addresses of the second one have prefix u, all addresses of the third one have prefix v, and the last two sets are disjoint from the condition $u \perp v$. So, have 3 cases, depending on where we find w.

Case 1. The address w is in $D(T) - (\{w \in D(T) | v \text{ is a prefix of } w\} \cup \{w \in D(T) | u \text{ is a prefix of } w\})$.

It does not matter if we substitute first at u and then at v, these are the original vertices of T that and they keep the T labels.

Case 2. The address w is in $\{vw | w \in D(t_2)\}$. These are the addresses introduced by t_2. Again, it does not matter if these addresses appear during the first or the second substitution, they have the labeling of t_2.

Case 2. The address w is in $\{uw | w \in D(t_1)\}$. These are the addresses introduced by t_1 at during the first or the second substitution. If they are generated by the first substitution, the second one leaves them undisturbed, and they keep their t_1 labels. **Q.E.D.**

If the addresses u and v are not disjoint, then the two substitutions nmay interfere with each other and the results may be different. Figure 1.27 shows such a case.

We do a *tree replacement* when we do the same substitution at a whole set of addresses. In most cases the addresses are leaves, and they are disjoint. So, we can perform the substitutions in any desired order.

Next, we will talk about *tree traversals*.

Definition 1.7.24 (depth-first traversal, breadth-first traversal) *A traversal is a listing of the addresses of a tree based on a total ordering. The vertices are listed in increasing order. The depth-first traversal uses the lexicographic order and the breadth-first uses the $\succ$ order. Both are defined in Section* **??**.

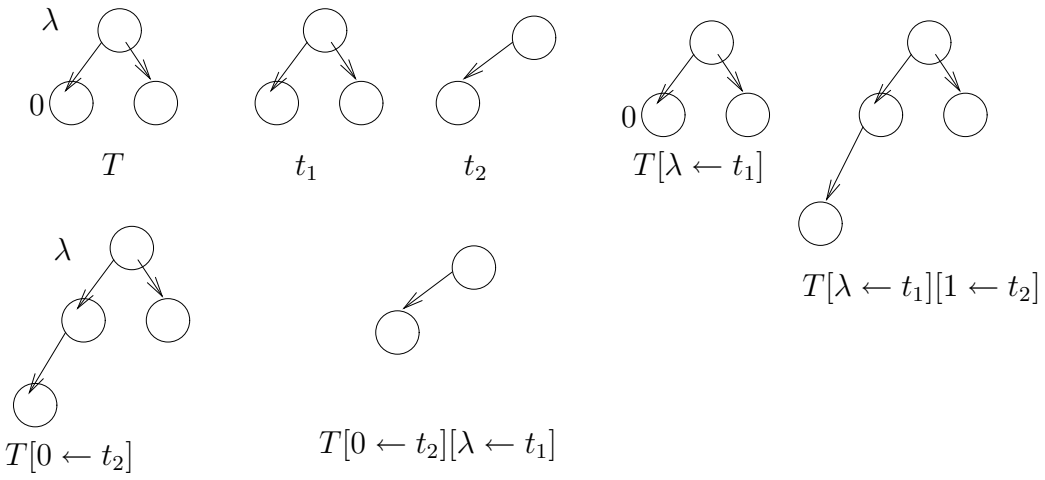


Figure 1.27: Non-commutative substitutions

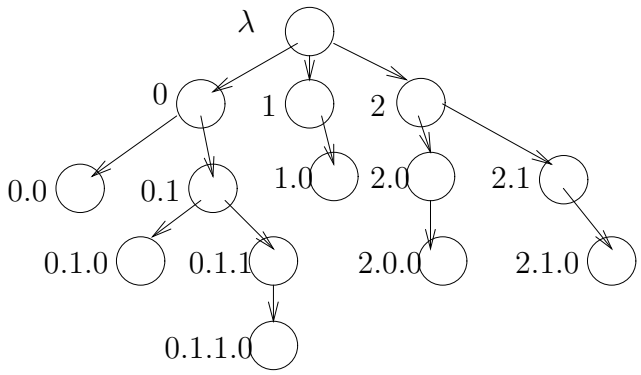


Figure 1.28: Tree traversals

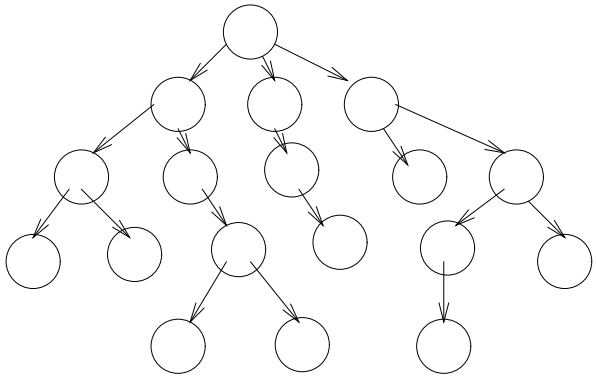


Figure 1.29: A tree

Let us traverse the tree from Figure 1.28 depth-first and breadth-first.

The lexicographical order of the domain of the tree is λ, 0, 0.0, 0.1, 0.1.0, 0.1.1, 0.1.1.0, 1, 1.0, 2, 2.0, 2.0.0, 2.1, 2.1.0.

The ordering $\succ$ lists the addresses in the incresing order of the length of the string, and within the length, they are listed in lexicographic order. We get the listing

λ, 0, 1, 2, 0.0, 0.1, 1.0, 2.0, 2.1, 0.1.0, 0.1.1, 2.0.0, 2.1.0, 0.1.1.0.

1.7.1 Exercises

Exercise 1.7.1 *Show that in a tree there is a unique path from the root to any other node.*

Exercise 1.7.2 *Number the tree below with Dewey numbers. Also, specify the tree domain.*

Exercise 1.7.3 *Assume that the address* 2.3.1.3 *is in the domain of the tree t. Which ones of the adresses below must also be in the domain? Justify your answer.*

a. 2.1
b. 2.3.2
c. 2.3.1.2
d. 1
e. 1.1

Exercise 1.7.4 *Let D be a set of strings over $\{1, 2, \ldots\}$ that satisfies the conditions of Observation 1.7.14. Show that there is a tree with domain D. Notice that the set D can be infinite.*

Exercise 1.7.5 *Draw the substree T/0 and T/2 of the tree from Figure 1.17.*

Exercise 1.7.6 *If the full k-ary tree has L leaves, and $k \geq 2$, how many branches does T have?*

Exercise 1.7.7 *Is the relation $\perp$ reflexive, symmetric, or transitive? Justify your answers.*

Exercise 1.7.8 *List the addresses of the tree from Figure 1.29 in the depth-first and breadth-first order.*