

Chapter 4

Mathematical Data Types

4.1 Sets

We've been assuming that the concepts of sets, sequences, and functions are already familiar ones, and we've mentioned them repeatedly. Now we'll do a quick review of the definitions.

Informally, a *set* is a bunch of objects, which are called the *elements* of the set. The elements of a set can be just about anything: numbers, points in space, or even other sets. The conventional way to write down a set is to list the elements inside curly-braces. For example, here are some sets:

$$\begin{array}{ll} A = \{\text{Alex, Tippy, Shells, Shadow}\} & \text{dead pets} \\ B = \{\text{red, blue, yellow}\} & \text{primary colors} \\ C = \{\{a, b\}, \{a, c\}, \{b, c\}\} & \text{a set of sets} \end{array}$$

This works fine for small finite sets. Other sets might be defined by indicating how to generate a list of them:

$$D = \{1, 2, 4, 8, 16, \dots\} \qquad \text{the powers of 2}$$

The order of elements is not significant, so $\{x, y\}$ and $\{y, x\}$ are the same set written two different ways. Also, any object is, or is not, an element of a given set—there is no notion of an element appearing more than once in a set.¹ So writing $\{x, x\}$ is just indicating the same thing twice, namely, that x is in the set. In particular, $\{x, x\} = \{x\}$.

The expression $e \in S$ asserts that e is an element of set S . For example, $32 \in D$ and $\text{blue} \in B$, but $\text{Tailspin} \notin A$ —yet.

Sets are simple, flexible, and everywhere. You'll find some set mentioned in nearly every section of this text.

¹It's not hard to develop a notion of *multisets* in which elements can occur more than once, but multisets are not ordinary sets.

4.1.1 Some Popular Sets

Mathematicians have devised special symbols to represent some common sets.

symbol	set	elements
\emptyset	the empty set	none
\mathbb{N}	nonnegative integers	$\{0, 1, 2, 3, \dots\}$
\mathbb{Z}	integers	$\{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$
\mathbb{Q}	rational numbers	$\frac{1}{2}, -\frac{5}{3}, 16, \text{ etc.}$
\mathbb{R}	real numbers	$\pi, e, -9, \sqrt{2}, \text{ etc.}$
\mathbb{C}	complex numbers	$i, \frac{19}{2}, \sqrt{2} - 2i, \text{ etc.}$

A superscript “+” restricts a set to its positive elements; for example, \mathbb{R}^+ denotes the set of positive real numbers. Similarly, \mathbb{R}^- denotes the set of negative reals.

4.1.2 Comparing and Combining Sets

The expression $S \subseteq T$ indicates that set S is a *subset* of set T , which means that every element of S is also an element of T (it could be that $S = T$). For example, $\mathbb{N} \subseteq \mathbb{Z}$ and $\mathbb{Q} \subseteq \mathbb{R}$ (every rational number is a real number), but $\mathbb{C} \not\subseteq \mathbb{Z}$ (not every complex number is an integer).

As a memory trick, notice that the \subseteq points to the smaller set, just like a \leq sign points to the smaller number. Actually, this connection goes a little further: there is a symbol \subset analogous to $<$. Thus, $S \subset T$ means that S is a subset of T , but the two are *not* equal. So $A \subseteq A$, but $A \not\subset A$, for every set A .

There are several ways to combine sets. Let’s define a couple of sets for use in examples:

$$X ::= \{1, 2, 3\}$$

$$Y ::= \{2, 3, 4\}$$

- The *union* of sets X and Y (denoted $X \cup Y$) contains all elements appearing in X or Y or both. Thus, $X \cup Y = \{1, 2, 3, 4\}$.
- The *intersection* of X and Y (denoted $X \cap Y$) consists of all elements that appear in *both* X and Y . So $X \cap Y = \{2, 3\}$.
- The *set difference* of X and Y (denoted $X - Y$) consists of all elements that are in X , but not in Y . Therefore, $X - Y = \{1\}$ and $Y - X = \{4\}$.

4.1.3 Complement of a Set

Sometimes we are focused on a particular domain, D . Then for any subset, A , of D , we define \bar{A} to be the set of all elements of D *not* in A . That is, $\bar{A} ::= D - A$. The set \bar{A} is called the *complement* of A .

For example, when the domain we're working with is the real numbers, the complement of the positive real numbers is the set of negative real numbers together with zero. That is,

$$\overline{\mathbb{R}^+} = \mathbb{R}^- \cup \{0\}.$$

It can be helpful to rephrase properties of sets using complements. For example, two sets, A and B , are said to be *disjoint* iff they have no elements in common, that is, $A \cap B = \emptyset$. This is the same as saying that A is a subset of the complement of B , that is, $A \subseteq \overline{B}$.

4.1.4 Power Set

The set of all the subsets of a set, A , is called the *power set*, $\mathcal{P}(A)$, of A . So $B \in \mathcal{P}(A)$ iff $B \subseteq A$. For example, the elements of $\mathcal{P}(\{1, 2\})$ are \emptyset , $\{1\}$, $\{2\}$ and $\{1, 2\}$.

More generally, if A has n elements, then there are 2^n sets in $\mathcal{P}(A)$. For this reason, some authors use the notation 2^A instead of $\mathcal{P}(A)$.

4.1.5 Set Builder Notation

An important use of predicates is in *set builder notation*. We'll often want to talk about sets that cannot be described very well by listing the elements explicitly or by taking unions, intersections, etc., of easily-described sets. Set builder notation often comes to the rescue. The idea is to define a *set* using a *predicate*; in particular, the set consists of all values that make the predicate true. Here are some examples of set builder notation:

$$\begin{aligned} A &::= \{n \in \mathbb{N} \mid n \text{ is a prime and } n = 4k + 1 \text{ for some integer } k\} \\ B &::= \{x \in \mathbb{R} \mid x^3 - 3x + 1 > 0\} \\ C &::= \{a + bi \in \mathbb{C} \mid a^2 + 2b^2 \leq 1\} \end{aligned}$$

The set A consists of all nonnegative integers n for which the predicate

$$\text{"}n \text{ is a prime and } n = 4k + 1 \text{ for some integer } k\text{"}$$

is true. Thus, the smallest elements of A are:

$$5, 13, 17, 29, 37, 41, 53, 57, 61, 73, \dots$$

Trying to indicate the set A by listing these first few elements wouldn't work very well; even after ten terms, the pattern is not obvious! Similarly, the set B consists of all real numbers x for which the predicate

$$x^3 - 3x + 1 > 0$$

is true. In this case, an explicit description of the set B in terms of intervals would require solving a cubic equation. Finally, set C consists of all complex numbers $a + bi$ such that:

$$a^2 + 2b^2 \leq 1$$

This is an oval-shaped region around the origin in the complex plane.

4.1.6 Proving Set Equalities

Two sets are defined to be equal if they contain the same elements. That is, $X = Y$ means that $z \in X$ if and only if $z \in Y$, for all elements, z . (This is actually the first of the ZFC axioms.) So set equalities can be formulated and proved as “iff” theorems. For example:

Theorem 4.1.1 (*Distributive Law for Sets*). *Let A , B , and C be sets. Then:*

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad (4.1)$$

Proof. The equality (4.1) is equivalent to the assertion that

$$z \in A \cap (B \cup C) \quad \text{iff} \quad z \in (A \cap B) \cup (A \cap C) \quad (4.2)$$

for all z . Now we’ll prove (4.2) by a chain of iff’s.

First we need a rule for distributing a propositional AND operation over an OR operation. It’s easy to verify by truth-table that

Lemma 4.1.2. *The propositional formula*

$$P \text{ AND } (Q \text{ OR } R)$$

and

$$(P \text{ AND } Q) \text{ OR } (P \text{ AND } R)$$

are equivalent.

Now we have

$$\begin{aligned} z \in A \cap (B \cup C) & \\ \text{iff } (z \in A) \text{ AND } (z \in B \cup C) & \quad (\text{def of } \cap) \\ \text{iff } (z \in A) \text{ AND } (z \in B \text{ OR } z \in C) & \quad (\text{def of } \cup) \\ \text{iff } (z \in A \text{ AND } z \in B) \text{ OR } (z \in A \text{ AND } z \in C) & \quad (\text{Lemma 4.1.2}) \\ \text{iff } (z \in A \cap B) \text{ OR } (z \in A \cap C) & \quad (\text{def of } \cap) \\ \text{iff } z \in (A \cap B) \cup (A \cap C) & \quad (\text{def of } \cup) \end{aligned}$$

■

4.1.7 Problems

Homework Problems

Problem 4.1.

Let A , B , and C be sets. Prove that:

$$A \cup B \cup C = (A - B) \cup (B - C) \cup (C - A) \cup (A \cap B \cap C). \quad (4.3)$$

Hint: $P \text{ OR } Q \text{ OR } R$ is equivalent to

$$(P \text{ AND } \overline{Q}) \text{ OR } (Q \text{ AND } \overline{R}) \text{ OR } (R \text{ AND } \overline{P}) \text{ OR } (P \text{ AND } Q \text{ AND } R).$$

4.2 Sequences

Sets provide one way to group a collection of objects. Another way is in a *sequence*, which is a list of objects called *terms* or *components*. Short sequences are commonly described by listing the elements between parentheses; for example, (a, b, c) is a sequence with three terms.

While both sets and sequences perform a gathering role, there are several differences.

- The elements of a set are required to be distinct, but terms in a sequence can be the same. Thus, (a, b, a) is a valid sequence of length three, but $\{a, b, a\}$ is a set with two elements —not three.
- The terms in a sequence have a specified order, but the elements of a set do not. For example, (a, b, c) and (a, c, b) are different sequences, but $\{a, b, c\}$ and $\{a, c, b\}$ are the same set.
- Texts differ on notation for the *empty sequence*; we use λ for the empty sequence.

The product operation is one link between sets and sequences. A *product of sets*, $S_1 \times S_2 \times \cdots \times S_n$, is a new set consisting of all sequences where the first component is drawn from S_1 , the second from S_2 , and so forth. For example, $\mathbb{N} \times \{a, b\}$ is the set of all pairs whose first element is a nonnegative integer and whose second element is an a or a b :

$$\mathbb{N} \times \{a, b\} = \{(0, a), (0, b), (1, a), (1, b), (2, a), (2, b), \dots\}$$

A product of n copies of a set S is denoted S^n . For example, $\{0, 1\}^3$ is the set of all 3-bit sequences:

$$\{0, 1\}^3 = \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}$$

4.3 Functions

A *function* assigns an element of one set, called the *domain*, to elements of another set, called the *codomain*. The notation

$$f : A \rightarrow B$$

indicates that f is a function with domain, A , and codomain, B . The familiar notation " $f(a) = b$ " indicates that f assigns the element $b \in B$ to a . Here b would be called the *value* of f at *argument* a .

Functions are often defined by formulas as in:

$$f_1(x) ::= \frac{1}{x^2}$$

where x is a real-valued variable, or

$$f_2(y, z) ::= y10yz$$

where y and z range over binary strings, or

$$f_3(x, n) ::= \text{the pair } (n, x)$$

where n ranges over the nonnegative integers.

A function with a finite domain could be specified by a table that shows the value of the function at each element of the domain. For example, a function $f_4(P, Q)$ where P and Q are propositional variables is specified by:

P	Q	$f_4(P, Q)$
T	T	T
T	F	F
F	T	T
F	F	T

Notice that f_4 could also have been described by a formula:

$$f_4(P, Q) ::= [P \text{ IMPLIES } Q].$$

A function might also be defined by a procedure for computing its value at any element of its domain, or by some other kind of specification. For example, define $f_5(y)$ to be the length of a left to right search of the bits in the binary string y until a 1 appears, so

$$\begin{aligned} f_5(0010) &= 3, \\ f_5(100) &= 1, \\ f_5(0000) &\text{ is undefined.} \end{aligned}$$

Notice that f_5 does not assign a value to any string of just 0's. This illustrates an important fact about functions: they need not assign a value to every element in the domain. In fact this came up in our first example $f_1(x) = 1/x^2$, which does not assign a value to 0. So in general, functions may be *partial functions*, meaning that there may be domain elements for which the function is not defined. If a function is defined on every element of its domain, it is called a *total function*.

It's often useful to find the set of values a function takes when applied to the elements in a set of arguments. So if $f : A \rightarrow B$, and S is a subset of A , we define $f(S)$ to be the set of all the values that f takes when it is applied to elements of S . That is,

$$f(S) ::= \{b \in B \mid f(s) = b \text{ for some } s \in S\}.$$

For example, if we let $[r, s]$ denote the interval from r to s on the real line, then $f_1([1, 2]) = [1/4, 1]$.

For another example, let's take the "search for a 1" function, f_5 . If we let X be the set of binary words which start with an even number of 0's followed by a 1, then $f_5(X)$ would be the odd nonnegative integers.

Applying f to a set, S , of arguments is referred to as "applying f pointwise to S ", and the set $f(S)$ is referred to as the *image* of S under f .² The set of values that arise from applying f to all possible arguments is called the *range* of f . That is,

$$\text{range}(f) ::= f(\text{domain}(f)).$$

Some authors refer to the codomain as the range of a function, but they shouldn't. The distinction between the range and codomain will be important in Sections 4.7 and 4.8 when we relate sizes of sets to properties of functions between them.

4.3.1 Function Composition

Doing things step by step is a universal idea. Taking a walk is a literal example, but so is cooking from a recipe, executing a computer program, evaluating a formula, and recovering from substance abuse.

Abstractly, taking a step amounts to applying a function, and going step by step corresponds to applying functions one after the other. This is captured by the operation of *composing* functions. Composing the functions f and g means that first f applied is to some argument, x , to produce $f(x)$, and then g is applied to that result to produce $g(f(x))$.

Definition 4.3.1. For functions $f : A \rightarrow B$ and $g : B \rightarrow C$, the *composition*, $g \circ f$, of g with f is defined to be the function $h : A \rightarrow C$ defined by the rule:

$$(g \circ f)(x) = h(x) ::= g(f(x)),$$

for all $x \in A$.

Function composition is familiar as a basic concept from elementary calculus, and it plays an equally basic role in discrete mathematics.

4.4 Binary Relations

Relations are another fundamental mathematical data type. Equality and "less-than" are very familiar examples of mathematical relations. These are called *binary relations* because they apply to a pair (a, b) of objects; the equality relation holds for the pair when $a = b$, and less-than holds when a and b are real numbers and $a < b$.

In this chapter we'll define some basic vocabulary and properties of binary relations.

²There is a picky distinction between the function f which applies to elements of A and the function which applies f pointwise to subsets of A , because the domain of f is A , while the domain of pointwise- f is $\mathcal{P}(A)$. It is usually clear from context whether f or pointwise- f is meant, so there is no harm in overloading the symbol f in this way.

4.5 Binary Relations and Functions

Binary relations are far more general than equality or less-than. Here's the official definition:

Definition 4.5.1. A *binary relation*, R , consists of a set, A , called the *domain* of R , a set, B , called the *codomain* of R , and a subset of $A \times B$ called the *graph* of R .

Notice that Definition 4.5.1 is exactly the same as the definition in Section 4.3 of a *function*, except that it doesn't require the functional condition that, for each domain element, a , there is *at most* one pair in the graph whose first coordinate is a . So a function is a special case of a binary relation.

A relation whose domain is A and codomain is B is said to be "between A and B ", or "from A to B ." When the domain and codomain are the same set, A , we simply say the relation is "on A ." It's common to use infix notation " $a R b$ " to mean that the pair (a, b) is in the graph of R .

For example, we can define an "in-charge of" relation, T , for MIT in Spring '10 to have domain equal to the set, F , of names of the faculty and codomain equal to all the set, N , of subject numbers in the current catalogue. The graph of T contains precisely the pairs of the form

$$(\langle \text{instructor-name} \rangle, \langle \text{subject-num} \rangle)$$

such that the faculty member named $\langle \text{instructor-name} \rangle$ is in charge of the subject with number $\langle \text{subject-num} \rangle$ in Spring '10. So graph (T) contains pairs like

(A. R. Meyer, 6.042),
 (A. R. Meyer, 18.062),
 (A. R. Meyer, 6.844),
 (T. Leighton, 6.042),
 (T. Leighton, 18.062),
 (G. Freeman, 6.011),
 (G. Freeman, 6.UAT),
 (G. Freeman, 6.881),
 (G. Freeman, 6.882),
 (T. Eng, 6.UAT),
 (J. Guttag, 6.00)
 ⋮

This is a surprisingly complicated relation: Meyer is in charge of subjects with three numbers. Leighton is also in charge of subjects with two of these three numbers—because the same subject, Mathematics for Computer Science, has two numbers: 6.042 and 18.062, and Meyer and Leighton are co-in-charge of the subject. Freeman is in-charge of even more subjects numbers (around 20), since as Department Education Officer, he is in charge of whole blocks of special subject numbers. Some subjects, like 6.844 and 6.00 have only one person in-charge. Some faculty,

like Guttag, are in charge of only one subject number, and no one else is co-in-charge of his subject, 6.00.

Some subjects in the codomain, N , do not appear in the list—that is, they are not an element of any of the pairs in the graph of T ; these are the Fall term only subjects. Similarly, there are faculty in the domain, F , who do not appear in the list because all their in-charge subjects are Fall term only.

4.6 Images and Inverse Images

The faculty in charge of 6.UAT in Spring '10 can be found by taking the pairs of the form

$$(\langle \text{instructor-name} \rangle, 6.UAT)$$

in the graph of the teaching relation, T , and then just listing the left hand sides of these pairs; these turn out to be just Eng and Freeman.

The introductory course 6 subjects have numbers that start with 6.0. So we can likewise find out all the instructors in-charge of introductory course 6 subjects this term, by taking all the pairs of the form $(\langle \text{instructor-name} \rangle, 6.0\dots)$ and list the left hand sides of these pairs. For example, from the part of the graph of T shown above, we can see that Meyer, Leighton, Freeman, and Guttag are in-charge of introductory subjects this term.

These are all examples of taking an *inverse image* of a set under a relation. If R is a binary relation from A to B , and X is any set, define the inverse image of X under R , written simply as RX to be the set elements of A that are related to something in X .

For example, let D be the set of introductory course 6 subject numbers. So TD , the inverse image of the set D under the relation, T , is the set of all faculty members in-charge of introductory course 6 subjects in Spring '10. Notice that in inverse image notation, D gets written to the right of T because, to find the faculty members in TD , we're looking pairs in the graph of T whose right hand sides are subject numbers in D .

Here's a concise definition of the inverse image of a set X under a relation, R :

$$RX ::= \{a \in A \mid aRx \text{ for some } x \in X\}.$$

Similarly, the *image* of a set Y under R , written YR , is the set of elements of the codomain, B , that are related to some element in Y , namely,

$$YR ::= \{b \in B \mid yRb \text{ for some } y \in Y\}.$$

So, $\{A. Meyer\}T$ gives the subject numbers that Meyer is in charge of in Spring '09. In fact, $\{A. Meyer\}T = \{6.042, 18.062, 6.844\}$. Since the domain, F , is the set of all in-charge faculty, FT is exactly the set of *all* Spring '09 subjects being taught. Similarly, TN is the set of people in-charge of a Spring '09 subject.

It gets interesting when we write composite expressions mixing images, inverse images and set operations. For example, $(TD)T$ is the set of Spring '09 subjects

that have people in-charge who also are in-charge of introductory subjects. So $(TD)T - D$ are the advanced subjects with someone in-charge who is also in-charge of an introductory subject. Similarly, $TD \cap T(N - D)$ is the set of faculty teaching both an introductory *and* an advanced subject in Spring '09.

Warning: When R happens to be a function, the pointwise application, $R(Y)$, of R to a set Y described in Section 4.3 is exactly the same as the image of Y under R . That means that when R is a function, $R(Y) = YR$ —not RY . Both notations are common in math texts, so you'll have to live with the fact that they clash. Sorry about that.

4.7 Surjective and Injective Relations

There are a few properties of relations that will be useful when we take up the topic of counting because they imply certain relations between the *sizes* of domains and codomains. We say a binary relation $R : A \rightarrow B$ is:

- *total* when every element of A is assigned to some element of B ; more concisely, R is total iff $A = RB$.
- *surjective* when every element of B is mapped to *at least once*³; more concisely, R is surjective iff $AR = B$.
- *injective* if every element of B is mapped to *at most once*, and
- *bijective* if R is total, surjective, and injective *function*.

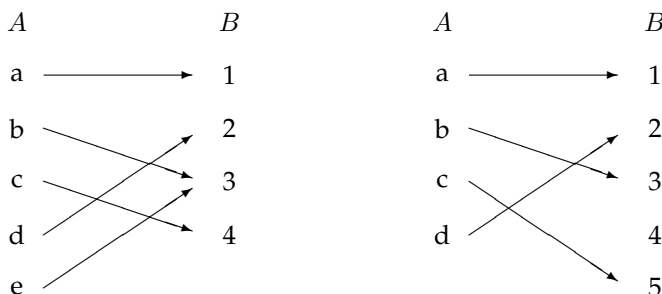
Note that this definition of R being total agrees with the definition in Section 4.3 when R is a function.

If R is a binary relation from A to B , we define AR to be the *range* of R . So a relation is surjective iff its range equals its codomain. Again, in the case that R is a function, these definitions of “range” and “total” agree with the definitions in Section 4.3.

4.7.1 Relation Diagrams

We can explain all these properties of a relation $R : A \rightarrow B$ in terms of a diagram where all the elements of the domain, A , appear in one column (a very long one if A is infinite) and all the elements of the codomain, B , appear in another column, and we draw an arrow from a point a in the first column to a point b in the second column when a is related to b by R . For example, here are diagrams for two functions:

³ The names “surjective” and “injective” are unmemorable and nondescriptive. Some authors use the term *onto* for surjective and *one-to-one* for injective, which are shorter but arguably no more memorable.



Here is what the definitions say about such pictures:

- “ R is a function” means that every point in the domain column, A , has *at most one arrow out of it*.
- “ R is total” means that *every* point in the A column has *at least one arrow out of it*. So if R is a function, being total really means every point in the A column has *exactly one arrow out of it*.
- “ R is surjective” means that *every* point in the codomain column, B , has *at least one arrow into it*.
- “ R is injective” means that every point in the codomain column, B , has *at most one arrow into it*.
- “ R is bijective” means that *every* point in the A column has exactly one arrow out of it, and *every* point in the B column has exactly one arrow into it.

So in the diagrams above, the relation on the left is a total, surjective function (every element in the A column has exactly one arrow out, and every element in the B column has at least one arrow in), but not injective (element 3 has two arrows going into it). The relation on the right is a total, injective function (every element in the A column has exactly one arrow out, and every element in the B column has at most one arrow in), but not surjective (element 4 has no arrow going into it).

Notice that the arrows in a diagram for R precisely correspond to the pairs in the graph of R . But $\text{graph}(R)$ does not determine by itself whether R is total or surjective; we also need to know what the domain is to determine if R is total, and we need to know the codomain to tell if it’s surjective.

Example 4.7.1. The function defined by the formula $1/x^2$ is total if its domain is \mathbb{R}^+ but partial if its domain is some set of real numbers including 0. It is bijective if its domain and codomain are both \mathbb{R}^+ , but neither injective nor surjective if its domain and codomain are both \mathbb{R} .

4.8 The Mapping Rule

The relational properties above are useful in figuring out the relative sizes of domains and codomains.

If A is a finite set, we let $|A|$ be the number of elements in A . A finite set may have no elements (the empty set), or one element, or two elements, ... or any non-negative integer number of elements.

Now suppose $R : A \rightarrow B$ is a function. Then every arrow in the diagram for R comes from exactly one element of A , so the number of arrows is at most the number of elements in A . That is, if R is a function, then

$$|A| \geq \#\text{arrows}.$$

Similarly, if R is surjective, then every element of B has an arrow into it, so there must be at least as many arrows in the diagram as the size of B . That is,

$$\#\text{arrows} \geq |B|.$$

Combining these inequalities implies that if R is a surjective function, then $|A| \geq |B|$. In short, if we write $A \text{ surj } B$ to mean that there is a surjective function from A to B , then we've just proved a lemma: if $A \text{ surj } B$, then $|A| \geq |B|$. The following definition and lemma lists include this statement and three similar rules relating domain and codomain size to relational properties.

Definition 4.8.1. Let A, B be (not necessarily finite) sets. Then

1. $A \text{ surj } B$ iff there is a surjective *function* from A to B .
2. $A \text{ inj } B$ iff there is a total injective *relation* from A to B .
3. $A \text{ bij } B$ iff there is a bijection from A to B .
4. $A \text{ strict } B$ iff $A \text{ surj } B$, but not $B \text{ surj } A$.

Lemma 4.8.2. [Mapping Rules] Let A and B be finite sets.

1. If $A \text{ surj } B$, then $|A| \geq |B|$.
2. If $A \text{ inj } B$, then $|A| \leq |B|$.
3. If $R \text{ bij } B$, then $|A| = |B|$.
4. If $R \text{ strict } B$, then $|A| > |B|$.

Mapping rule 2 can be explained by the same kind of "arrow reasoning" we used for rule 1. Rules 3 and 4 are immediate consequences of these first two mapping rules.

4.9 The sizes of infinite sets

Mapping Rule 1 has a converse: if the size of a finite set, A , is greater than or equal to the size of another finite set, B , then it's always possible to define a surjective

function from A to B . In fact, the surjection can be a total function. To see how this works, suppose for example that

$$\begin{aligned} A &= \{a_0, a_1, a_2, a_3, a_4, a_5\} \\ B &= \{b_0, b_1, b_2, b_3\}. \end{aligned}$$

Then define a total function $f : A \rightarrow B$ by the rules

$$f(a_0) ::= b_0, f(a_1) ::= b_1, f(a_2) ::= b_2, f(a_3) = f(a_4) = f(a_5) ::= b_3.$$

In fact, if A and B are finite sets of the same size, then we could also define a bijection from A to B by this method.

In short, we have figured out if A and B are finite sets, then $|A| \geq |B|$ if and only if $A \text{ surj } B$, and similar iff's hold for all the other Mapping Rules:

Lemma 4.9.1. For finite sets, A, B ,

$$\begin{aligned} |A| \geq |B| & \text{ iff } A \text{ surj } B, \\ |A| \leq |B| & \text{ iff } A \text{ inj } B, \\ |A| = |B| & \text{ iff } A \text{ bij } B, \\ |A| > |B| & \text{ iff } A \text{ strict } B. \end{aligned}$$

This lemma suggests a way to generalize size comparisons to infinite sets, namely, we can think of the relation surj as an “at least as big as” relation between sets, even if they are infinite. Similarly, the relation bij can be regarded as a “same size” relation between (possibly infinite) sets, and strict can be thought of as a “strictly bigger than” relation between sets.

Warning: We haven't, and won't, define what the “size” of an infinite is. The definition of infinite “sizes” is cumbersome and technical, and we can get by just fine without it. All we need are the “as big as” and “same size” relations, surj and bij, between sets.

But there's something else to **watch out for**. We've referred to surj as an “as big as” relation and bij as a “same size” relation on sets. Of course most of the “as big as” and “same size” properties of surj and bij on finite sets do carry over to infinite sets, but *some important ones don't*—as we're about to show. So you have to be careful: don't assume that surj has any particular “as big as” property on *infinite* sets until it's been proved.

Let's begin with some familiar properties of the “as big as” and “same size” relations on finite sets that do carry over exactly to infinite sets:

Lemma 4.9.2. For any sets, A, B, C ,

1. $A \text{ surj } B$ and $B \text{ surj } C$, implies $A \text{ surj } C$.
2. $A \text{ bij } B$ and $B \text{ bij } C$, implies $A \text{ bij } C$.
3. $A \text{ bij } B$ implies $B \text{ bij } A$.

Lemma 4.9.2.1 and 4.9.2.2 follow immediately from the fact that compositions of surjections are surjections, and likewise for bijections, and Lemma 4.9.2.3 follows from the fact that the inverse of a bijection is a bijection. We'll leave a proof of these facts to Problem 4.2.

Another familiar property of finite sets carries over to infinite sets, but this time it's not so obvious:

Theorem 4.9.3 (Schröder-Bernstein). *For any sets A, B , if $A \text{ surj } B$ and $B \text{ surj } A$, then $A \text{ bij } B$.*

That is, the Schröder-Bernstein Theorem says that if A is at least as big as B and conversely, B is at least as big as A , then A is the same size as B . Phrased this way, you might be tempted to take this theorem for granted, but that would be a mistake. For infinite sets A and B , the Schröder-Bernstein Theorem is actually pretty technical. Just because there is a surjective function $f : A \rightarrow B$ —which need not be a bijection—and a surjective function $g : B \rightarrow A$ —which also need not be a bijection—it's not at all clear that there must be a bijection $e : A \rightarrow B$. The idea is to construct e from parts of both f and g . We'll leave the actual construction to Problem 4.7.

Infinity is different

A basic property of finite sets that does *not* carry over to infinite sets is that adding something new makes a set bigger. That is, if A is a finite set and $b \notin A$, then $|A \cup \{b\}| = |A| + 1$, and so A and $A \cup \{b\}$ are not the same size. But if A is infinite, then these two sets *are* the same size!

Lemma 4.9.4. *Let A be a set and $b \notin A$. Then A is infinite iff $A \text{ bij } A \cup \{b\}$.*

Proof. Since A is *not* the same size as $A \cup \{b\}$ when A is finite, we only have to show that $A \cup \{b\}$ is the same size as A when A is infinite.

That is, we have to find a bijection between $A \cup \{b\}$ and A when A is infinite. Here's how: since A is infinite, it certainly has at least one element; call it a_0 . But since A is infinite, it has at least two elements, and one of them must not be equal to a_0 ; call this new element a_1 . But since A is infinite, it has at least three elements, one of which must not equal a_0 or a_1 ; call this new element a_2 . Continuing in the way, we conclude that there is an infinite sequence $a_0, a_1, a_2, \dots, a_n, \dots$ of different elements of A . Now it's easy to define a bijection $e : A \cup \{b\} \rightarrow A$:

$$\begin{aligned} e(b) &::= a_0, \\ e(a_n) &::= a_{n+1} && \text{for } n \in \mathbb{N}, \\ e(a) &::= a && \text{for } a \in A - \{b, a_0, a_1, \dots\}. \end{aligned}$$

■

A set, C , is *countable* iff its elements can be listed in order, that is, the distinct elements is A are precisely

$$c_0, c_1, \dots, c_n, \dots$$

This means that if we defined a function, f , on the nonnegative integers by the rule that $f(i) ::= c_i$, then f would be a bijection from \mathbb{N} to C . More formally,

Definition 4.9.5. A set, C , is *countably infinite* iff $\mathbb{N} \text{ bij } C$. A set is *countable* iff it is finite or countably infinite.

A small modification⁴ of the proof of Lemma 4.9.4 shows that countably infinite sets are the “smallest” infinite sets, namely, if A is a countably infinite set, then $A \text{ surj } \mathbb{N}$.

Since adding one new element to an infinite set doesn’t change its size, it’s obvious that neither will adding any *finite* number of elements. It’s a common mistake to think that this proves that you can throw in countably infinitely many new elements. But just because it’s ok to do something any finite number of times doesn’t make it OK to do an infinite number of times. For example, starting from 3, you can add 1 any finite number of times and the result will be some integer greater than or equal to 3. But if you add add 1 a countably infinite number of times, you don’t get an integer at all.

It turns out you really can add a countably infinite number of new elements to a countable set and still wind up with just a countably infinite set, but another argument is needed to prove this:

Lemma 4.9.6. *If A and B are countable sets, then so is $A \cup B$.*

Proof. Suppose the list of distinct elements of A is a_0, a_1, \dots and the list of B is b_0, b_1, \dots . Then a list of all the elements in $A \cup B$ is just

$$a_0, b_0, a_1, b_1, \dots, a_n, b_n, \dots \quad (4.4)$$

Of course this list will contain duplicates if A and B have elements in common, but then deleting all but the first occurrences of each element in list (4.4) leaves a list of all the distinct elements of A and B . ■

4.9.1 Infinities in Computer Science

We’ve run into a lot of computer science students who wonder why they should care about infinite sets: any data set in a computer memory is limited by the size of memory, and since the universe appears to have finite size, there is a limit on the possible size of computer memory.

The problem with this argument is that universe-size bounds on data items are so big and uncertain (the universe seems to be getting bigger all the time), that it’s simply not helpful to make use of possible bounds. For example, by this argument the physical sciences shouldn’t assume that measurements might yield arbitrary real numbers, because there can only be a finite number of finite measurements in a universe of finite lifetime. What do you think scientific theories would look like without using the infinite set of real numbers?

⁴See Problem 4.3

Similarly, in computer science, it simply isn't plausible that writing a program to add nonnegative integers with up to as many digits as, say, the stars in the sky (billions of galaxies each with billions of stars), would be any different than writing a program that would add any two integers no matter how many digits they had.

That's why basic programming data types like integers or strings, for example, can be defined without imposing any bound on the sizes of data items. Each datum of type `string` has only a finite number of letters, but there are an infinite number of data items of type `string`. When we then consider string procedures of type `string-->string`, not only are there an infinite number of such procedures, but each procedure generally behaves differently on different inputs, so that a single `string-->string` procedure may embody an infinite number of behaviors.

In short, an educated computer scientist can't get around having to understand infinite sets.

4.9.2 Problems

Class Problems

Problem 4.2.

Define a *surjection relation*, `surj`, on sets by the rule

Definition. $A \text{ surj } B$ iff there is a surjective **function** from A to B .

Define the *injection relation*, `inj`, on sets by the rule

Definition. $A \text{ inj } B$ iff there is a total injective *relation* from A to B .

- (a) Prove that if $A \text{ surj } B$ and $B \text{ surj } C$, then $A \text{ surj } C$.
- (b) Explain why $A \text{ surj } B$ iff $B \text{ inj } A$.
- (c) Conclude from (a) and (b) that if $A \text{ inj } B$ and $B \text{ inj } C$, then $A \text{ inj } C$.

Problem 4.3.

Lemma 4.9.4. Let A be a set and $b \notin A$. If A is infinite, then there is a bijection from $A \cup \{b\}$ to A .

Proof. Here's how to define the bijection: since A is infinite, it certainly has at least one element; call it a_0 . But since A is infinite, it has at least two elements, and one of them must not be equal to a_0 ; call this new element a_1 . But since A is infinite, it has at least three elements, one of which must not equal a_0 or a_1 ; call this new element a_2 . Continuing in the way, we conclude that there is an infinite sequence $a_0, a_1, a_2, \dots, a_n, \dots$ of different elements of A . Now we can define a bijection $f : A \cup \{b\} \rightarrow A$:

$$\begin{aligned} f(b) &::= a_0, \\ f(a_n) &::= a_{n+1} && \text{for } n \in \mathbb{N}, \\ f(a) &::= a && \text{for } a \in A - \{b, a_0, a_1, \dots\}. \end{aligned}$$

■

(a) Several students felt the proof of Lemma 4.9.4 was worrisome, if not circular. What do you think?

(b) Use the proof of Lemma 4.9.4 to show that if A is an infinite set, then there is surjective function from A to \mathbb{N} , that is, every infinite set is “as big as” the set of nonnegative integers.

Problem 4.4.

Let $R : A \rightarrow B$ be a binary relation. Use an arrow counting argument to prove the following generalization of the Mapping Rule:

Lemma. If R is a function, and $X \subseteq A$, then

$$|X| \geq |XR|.$$

Problem 4.5.

Let $A = \{a_0, a_1, \dots, a_{n-1}\}$ be a set of size n , and $B = \{b_0, b_1, \dots, b_{m-1}\}$ a set of size m . Prove that $|A \times B| = mn$ by defining a simple bijection from $A \times B$ to the nonnegative integers from 0 to $mn - 1$.

Problem 4.6.

The rational numbers fill in all the spaces between the integers, so a first thought is that there must be more of them than the integers, but it's not true. In this problem

you'll show that there are the same number of nonnegative rational as nonnegative integers. In short, the nonnegative rationals are countable.

(a) Describe a bijection between all the integers, \mathbb{Z} , and the nonnegative integers, \mathbb{N} .

(b) Define a bijection between the nonnegative integers and the set, $\mathbb{N} \times \mathbb{N}$, of all the ordered pairs of nonnegative integers:

$$\begin{aligned} &(0, 0), (0, 1), (0, 2), (0, 3), (0, 4), \dots \\ &(1, 0), (1, 1), (1, 2), (1, 3), (1, 4), \dots \\ &(2, 0), (2, 1), (2, 2), (2, 3), (2, 4), \dots \\ &(3, 0), (3, 1), (3, 2), (3, 3), (3, 4), \dots \\ &\quad \vdots \end{aligned}$$

(c) Conclude that \mathbb{N} is the same size as the set, \mathbb{Q} , of all nonnegative rational numbers.

Problem 4.7.

Suppose sets A and B have no elements in common, and

- A is as small as B because there is a total injective function $f : A \rightarrow B$, and
- B is as small as A because there is a total injective function $g : B \rightarrow A$.

Picturing the diagrams for f and g , there is *exactly one* arrow *out* of each element—a left-to-right f -arrow if the element in A and a right-to-left g -arrow if the element in B . This is because f and g are total functions. Also, there is *at most one* arrow *into* any element, because f and g are injections.

So starting at any element, there is a unique, and unending path of arrows going forwards. There is also a unique path of arrows going backwards, which might be unending, or might end at an element that has no arrow into it. These paths are completely separate: if two ran into each other, there would be two arrows into the element where they ran together.

This divides all the elements into separate paths of four kinds:

- i. paths that are infinite in both directions,
- ii. paths that are infinite going forwards starting from some element of A .
- iii. paths that are infinite going forwards starting from some element of B .
- iv. paths that are unending but finite.

(a) What do the paths of the last type (iv) look like?

(b) Show that for each type of path, either

- the f -arrows define a bijection between the A and B elements on the path, or
- the g -arrows define a bijection between B and A elements on the path, or
- both sets of arrows define bijections.

For which kinds of paths do both sets of arrows define bijections?

(c) Explain how to piece these bijections together to prove that A and B are the same size.

Homework Problems

Problem 4.8.

Let $f : A \rightarrow B$ and $g : B \rightarrow C$ be functions and $h : A \rightarrow C$ be their composition, namely, $h(a) ::= g(f(a))$ for all $a \in A$.

- Prove that if f and g are surjections, then so is h .
- Prove that if f and g are bijections, then so is h .
- If f is a bijection, then define $f' : B \rightarrow A$ so that

$$f'(b) ::= \text{the unique } a \in A \text{ such that } f(a) = b.$$

Prove that f' is a bijection. (The function f' is called the *inverse* of f . The notation f^{-1} is often used for the inverse of f .)

Problem 4.9.

In this problem you will prove a fact that may surprise you —or make you even more convinced that set theory is nonsense: the half-open unit interval is actually the *same size* as the nonnegative quadrant of the real plane!⁵ Namely, there is a bijection from $(0, 1]$ to $[0, \infty)^2$.

- Describe a bijection from $(0, 1]$ to $[0, \infty)$.

Hint: $1/x$ almost works.

- An infinite sequence of the decimal digits $\{0, 1, \dots, 9\}$ will be called *long* if it has infinitely many occurrences of some digit other than 0. Let L be the set of all such long sequences. Describe a bijection from L to the half-open real interval $(0, 1]$.

Hint: Put a decimal point at the beginning of the sequence.

- Describe a surjective function from L to L^2 that involves alternating digits from two long sequences. a *Hint:* The surjection need not be total.

- Prove the following lemma and use it to conclude that there is a bijection from L^2 to $(0, 1]^2$.

⁵The half open unit interval, $(0, 1]$, is $\{r \in \mathbb{R} \mid 0 < r \leq 1\}$. Similarly, $[0, \infty) ::= \{r \in \mathbb{R} \mid r \geq 0\}$.

Lemma 4.9.7. *Let A and B be nonempty sets. If there is a bijection from A to B , then there is also a bijection from $A \times A$ to $B \times B$.*

(e) Conclude from the previous parts that there is a surjection from $(0, 1]$ and $(0, 1]^2$. Then appeal to the Schröder-Bernstein Theorem to show that there is actually a bijection from $(0, 1]$ and $(0, 1]^2$.

(f) Complete the proof that there is a bijection from $(0, 1]$ to $[0, \infty)^2$.

4.10 Glossary of Symbols

symbol	meaning
\in	is a member of
\subseteq	is a subset of
\subset	is a proper subset of
\cup	set union
\cap	set intersection
\bar{A}	complement of a set, A
$\mathcal{P}(A)$	powerset of a set, A
\emptyset	the empty set, $\{\}$
\mathbb{N}	nonnegative integers
\mathbb{Z}	integers
\mathbb{Z}^+	positive integers
\mathbb{Z}^-	negative integers
\mathbb{Q}	rational numbers
\mathbb{R}	real numbers
\mathbb{C}	complex numbers
λ	the empty string/list

MIT OpenCourseWare
<http://ocw.mit.edu>

6.042J / 18.062J Mathematics for Computer Science
Spring 2010

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.