Velleman

Malcolm

Started 6th September 2025

Contents

1 Logic **4**

1.0.1 Logic Factsheet 4
1.0.2 Set operation definitions 5
1.0.3 Distributivity of set operations 5
1.0.4 $x \in A \setminus (B \cap C) = x \in (A \setminus B) \cup (A \setminus C)$ 6
1.0.5 $x \in (A \cup B) \setminus (A \cap B) = x \in (A \setminus B) \cup (B \setminus A)$ 6
1.0.6 $(A \cap B) \cap (A \setminus B) = \emptyset$ 6
1.0.7 Symmetric difference 7
1.0.8 Conditional and Contrapositive laws 7
1.0.9 Biconditional statements 8
1.0.10 Associativity of the conditional connective 9
1.0.11 $(A \to (B \to C)) \leftrightarrow ((A \wedge B) \to C)$ 9
1.1 Quantificational logic 10
1.1.1 Quantifier negation laws 10
1.1.2 Notation 10
1.1.3 Negation law for bounded quantifiers 11
1.1.4 Vacuously true 12
1.1.5 Alternate definition for indexed families 13
1.1.6 Power set 14
1.1.7 Intersection and union of a family of sets 14
1.1.8 More on set notation 15

2 Proof Strategies **16**

2.0.1 Terminology 16
2.1 To prove a conclusion of the form $P \to Q$ 17
2.1.1 Alternative approach 19
2.2 Proofs involving negations and conditionals 21
2.2.1 Reexpress 21
2.2.2 Proof by contradiction 22
2.2.3 To use a given of the form $\neg P$ in proof by contradiction . 24
2.2.4 Given conditionals 27
2.3 Proofs involving quantifiers 30
2.3.1 Goals of form $\forall x P(x)$ 30
2.3.2 Goals of form $\exists x P(x)$ 33

2.3.3 Givens of the form $\exists x P(x)$ and $\forall x P(x)$ . . . 36
2.3.4 Example: For all integers a, b, c, if $a \mid b$ and $b \mid c$ then $a \mid c$ 39
2.4 Proofs involving conjunctions and biconditionals . . . 40
2.4.1 Goals and givens of form $P \wedge Q$ . . . 40
2.4.2 Goals and givens of the form $P \leftrightarrow Q$ (part 1) . . . 42
2.4.3 Example: Suppose x is an integer. Prove that x is even iff x^2 is even . . . 42
2.4.4 Proving $\forall x \neg P(x) \leftrightarrow \neg \exists x P(x)$ . . . 44
2.4.5 Goals and givens of the form $P \leftrightarrow Q$ (part 2) . . . 45
2.4.6 Example: For every integer n, $6 \mid n$ iff $2 \mid n$ and $3 \mid n$ . . . 47
2.5 Proofs involving disjunctions . . . 48
2.5.1 Goal of form $P \vee Q$ (part 1) . . . 50
2.5.2 Example: For integer x, the remainder when x^2 is divided by 4 is either 0 or 1 . . . 51
2.5.3 Goal of form $P \vee Q$ (part 2) . . . 53
2.5.4 Given of form $P \vee Q$ (part 2) . . . 55
2.6 Existence and Uniqueness proofs . . . 56
2.6.1 Strategies and examples . . . 59
2.6.2 Given of the form $\exists! x P(x)$ . . . 62

3 Relations **64**
3.1 Ordered pairs and Cartesian product . . . 64
3.1.1 Properties of the Cartesian product . . . 66
3.1.2 Truth sets . . . 70
3.2 Relations . . . 71
3.2.1 Domain, Range, Inverse, Composition . . . 72
3.2.2 Alternate notation and visual representation . . . 75
3.2.3 Binary relations . . . 76
3.3 Ordering relations . . . 79
3.3.1 Smallest and minimal element . . . 80
3.3.2 Upper and lower bounds . . . 84
3.4 Equivalence relations . . . 86
3.4.1 More definitions . . . 87
3.4.2 Equivalence relations and partitions . . . 90
3.4.3 More on equivalence relations and partitions . . . 92
3.4.4 Even more on equivalence relations and partitions . . . 95
3.4.5 Congruence modulo m . . . 96

4 Functions **97**
4.1 Definition . . . 97
4.1.1 Function equality, Domain/Range/Composition . . . 99
4.2 One-to-One and Onto . . . 102
4.2.1 Composition of functions . . . 106
4.3 Inverses of functions . . . 107
4.3.1 Identity function . . . 108
4.3.2 Circle of implications . . . 110

4.3.3 More implications . . . 111
4.3.4 With regard to equivalence relations . . . 112
4.4 Closures . . . 113
4.4.1 On the union and intersection of empty sets . . . 115
4.4.2 Functions of two variables and closures . . . 116
4.5 Images and Inverse Images . . . 118

5 Mathematical Induction **119**
5.1 Proof by Mathematical Induction . . . 119
5.1.1 Examples . . . 120
5.1.2 More examples . . . 124
5.1.3 Even more examples . . . 128
5.2 Recursion . . . 132
5.2.1 Examples . . . 134
5.2.2 More examples . . . 136
5.3 Strong induction . . . 140
5.3.1 Example: Division algorithm . . . 142
5.3.2 Example: Either prime or a product of two or more primes 143
5.3.3 Example: Fibonacci numbers . . . 144
5.3.4 Example: Well-ordering principle . . . 147
5.4 Closures again . . . 149

6 Number Theory **151**
6.1 Greatest Common Divisors . . . 151
6.1.1 Computation and Euclidean algorithm . . . 154
6.1.2 Linear combinations—Extended Euclidean algorithm . . . 156
6.2 Prime factorization . . . 159
6.2.1 Relatively prime and more . . . 159
6.2.2 Existence and uniqueness of prime factorisations . . . 161
6.3 Modular Arithmetic . . . 165
6.3.1 Sums and products between equivalence classes . . . 168
6.3.2 Properties of operations between equivalence classes . . . 170
6.3.3 Application . . . 174
6.3.4 Alternative approaches . . . 175
6.4 Euler's theorem . . . 176
6.4.1 Generalisation of commutativity and associativity . . . 178
6.4.2 The theorem . . . 180
6.4.3 Phi function computation . . . 181

A Exercises **183**
A.1 Ch 3 . . . 183
A.1.1 $\exists x(P(x) \to Q(x))$ is equivalent to $\forall x P(x) \to \exists x Q(x)$ . . . 183
A.1.2 If $\exists x(P(x) \to Q(x))$, then $\forall x P(x) \to \exists x Q(x)$ . . . 183
A.1.3 . . . 184

Chapter 1

Logic

1.0.1 Logic Factsheet

De Morgan's laws

$$\neg\,(P \wedge Q) \text{ is equivalent to } \neg\,P \vee \neg\,Q$$
$$\neg\,(P \vee Q) \text{ is equivalent to } \neg\,P \wedge \neg\,Q$$

Commutative laws

$$P \wedge Q \text{ is equivalent to } Q \wedge P$$
$$P \vee Q \text{ is equivalent to } Q \vee P$$

Associative laws

$$P \wedge (Q \wedge R) \text{ is equivalent to } (P \wedge Q) \wedge R$$
$$P \vee (Q \vee R) \text{ is equivalent to } (P \vee Q) \vee R$$

Indempotent laws

$$P \wedge P \text{ is equivalent to } P$$
$$P \vee P \text{ is equivalent to } P$$

Distributive laws

$$P \wedge (Q \vee R) \text{ is equivalent to } (P \wedge Q) \vee (P \wedge R)$$
$$P \vee (Q \wedge R) \text{ is equivalent to } (P \vee Q) \wedge (P \vee R)$$

Absorption laws

$$P \vee (P \wedge Q) \text{ is equivalent to } P$$
$$P \wedge (P \vee Q) \text{ is equivalent to } P$$

Double Negation law

$$\neg\neg\,P \text{ is equivalent to } P$$

1.0.2 Set operation definitions

The *intersection* of two sets A and B is the set $A \cap B$ defined as follows:

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}$$

The *union* of A and B is the set $A \cup B$ defined as follows:

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}$$

The *difference* of A and B is the set $A \setminus B$ defined as follows:

$$A \setminus B = \{x \mid x \in A \text{ and } x \notin B\}$$

See that

$$x \in A \cap B = x \in \{y \mid y \in A \text{ and } y \in B\}$$

where y is a dummy variable. So we can also write that

$$x \in A \cap B = x \in A \wedge x \in B$$

The same can be shown for the union and difference.

1.0.3 Distributivity of set operations

We show

$$x \in A \cap (B \cup C) \text{ is equivalent to } x \in (A \cap B) \cup (A \cap C)$$

By analysing their logical forms:

$$\begin{aligned} & x \in A \cap (B \cup C) \\ & = x \in A \wedge x \in (B \cup C) \\ & = x \in A \wedge (x \in B \vee x \in C) \end{aligned}$$

and

$$\begin{aligned} & x \in (A \cap B) \cup (A \cap C) \\ & = x \in (A \cap B) \vee x \in (A \cap C) \\ & = (x \in A \wedge x \in B) \vee (x \in A \wedge x \in C) \\ & = [(x \in A \wedge x \in B) \vee x \in A)] \wedge [(x \in A \wedge x \in B) \vee x \in C)] \\ & = x \in A \wedge [(x \in A \vee x \in C) \wedge (x \in B \vee x \in C)] \\ & = [x \in A \wedge (x \in A \vee x \in C)] \wedge (x \in B \vee x \in C) \\ & = x \in A \wedge (x \in B \vee x \in C) \end{aligned}$$

We can also show, in a similar manner, that

$$x \in A \cup (B \cap C) \text{ is equivalent to } x \in (A \cup B) \cap (A \cup C)$$

1.0.4 $x \in A \setminus (B \cap C) = x \in (A \setminus B) \cup (A \setminus C)$

We can also show

$$x \in A \setminus (B \cap C) = x \in (A \setminus B) \cup (A \setminus C)$$

See that

$$\begin{aligned}
& x \in A \setminus (B \cap C) \\
&= x \in A \wedge \neg (x \in B \cap C) && \text{(Definition of } \setminus\text{)} \\
&= x \in A \wedge \neg (x \in B \wedge x \in C) && \text{(Definition of } \cap\text{)} \\
&= x \in A \wedge (x \notin B \vee x \notin C) && \text{(De Morgan's)} \\
&= (x \in A \wedge x \notin B) \vee (x \in A \wedge x \notin C) && \text{(Distributivity)} \\
&= (x \in A \setminus B) \vee (x \in A \setminus C) && \text{(Definition of } \setminus\text{)} \\
&= x \in (A \setminus B) \cup (A \setminus C) && \text{(Definition of } \cup\text{)}
\end{aligned}$$

1.0.5 $x \in (A \cup B) \setminus (A \cap B) = x \in (A \setminus B) \cup (B \setminus A)$

$$\begin{aligned}
& x \in (A \cup B) \setminus (A \cap B) \\
&= (x \in A \vee x \in B) \wedge \neg (x \in A \wedge x \in B) && \text{(By definition)} \\
&= (x \in A \vee x \in B) \wedge (x \notin A \vee x \notin B) && \text{(De Morgan's)} \\
&= [(x \in A \vee x \in B) \wedge (x \notin A)] \\
&\qquad \vee [(x \in A \vee x \in B) \wedge (x \notin B)] && \text{(Distributivity)} \\
&= [(x \notin A \wedge x \in A) \vee (x \notin A \wedge x \in B)] \\
&\qquad \vee [(x \notin B \wedge x \in A) \vee (x \notin B \wedge x \in B)] && \text{(Distributivity)} \\
&= (x \notin A \wedge x \in B) \vee (x \notin B \wedge x \in A) \\
&= (x \in A \wedge x \notin B) \wedge (x \in B \wedge x \notin A) && \text{(Commutativity)} \\
&= x \in (A \setminus B) \cup (B \setminus A) && \text{(By definition)}
\end{aligned}$$

1.0.6 $(A \cap B) \cap (A \setminus B) = \emptyset$

See that

$$\begin{aligned}
& x \in (A \cap B) \cap (A \setminus B) \\
&= (x \in A \wedge x \in B) \wedge (x \in A \wedge x \notin B) && \text{(Definition)} \\
&= x \in A \wedge \underbrace{(x \in B \wedge x \notin B)}_{\text{Contradiction}} && \text{(Associativity + Commutativity)}
\end{aligned}$$

The last statement is a contradiction, so the statement $x \in (A \cap B) \cap (A \setminus B)$ will always be false, no matter what x is. In other words, nothing can be an element of $(A \cap B) \cap (A \setminus B)$, so it must be the case that $(A \cap B) \cap (A \setminus B) = \emptyset$; $A \cap B$ and $A \setminus B$ are disjoint.

1.0.7 Symmetric difference

For any sets A and B,

$$(A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$$

Both sets will always be the set of objects that are elements of either A or B but not both. This is called the *symmetric difference* of A and B and is written

$$A \Delta B$$

1.0.8 Conditional and Contrapositive laws

Conditional Law

$$P \to Q \text{ is equivalent to } \neg (P \wedge \neg Q)$$

by De Morgan's law we can also say that

$$P \to Q \text{ is equivalent to } \neg P \vee Q$$

Contrapositive law

$$P \to Q \text{ is equivalent to } \neg Q \to \neg P$$

This can be justified using

$$P \to Q = \neg (P \wedge \neg Q) = \neg (\neg Q \wedge P) = \neg Q \to \neg P$$

Intuition
Intuitive ways to think of $P \to Q$ (and equivalently $\neg Q \to \neg P$) include:

- P implies Q.
- Q, if P.
- P only if Q.
- P is a sufficient condition for Q.
- Q is a necessary condition for P.

1.0.9 Biconditional statements

We write

$$P \leftrightarrow Q = (P \rightarrow Q) \wedge (Q \rightarrow P)$$

Note that by the contrapositive law, this is also equivalent to

$$(P \rightarrow Q) \wedge (\neg P \rightarrow \neg Q)$$

Intuition
$Q \rightarrow P$ can be written as 'P if Q' and $P \rightarrow Q$ can be written as 'P only if Q' (since this means $\neg Q \rightarrow \neg P$ which is $P \rightarrow Q$).

Combining the two as $(P \rightarrow Q) \wedge (Q \rightarrow P) = P \leftrightarrow Q$ therefore corresponds to the statement 'P if and only if Q'.

$P \leftrightarrow Q$ means 'P iff Q', or 'P is a necessary and sufficient condition for Q'.

1.0.10 Associativity of the conditional connective

The truth table for the conditional is as follows:

P	Q	$P \to Q$
F	F	T
F	T	T
T	F	F
T	T	T

This may be better understood using the equivalence $(P \to Q) \leftrightarrow \neg(P \wedge \neg Q)$.

Now for a counterexample demonstrating that the conditional connective is not associative:

P	Q	R	$P \to Q$	$Q \to R$	$(P \to Q) \to R$	$P \to (Q \to R)$
F	F	F	T	T	F	T

1.0.11 $(A \to (B \to C)) \leftrightarrow ((A \wedge B) \to C)$

Proof. ($\to$) Suppose $(A \to (B \to C))$. Then suppose $A \wedge B$. Then since A is true, $B \to C$ will be true. Then since B is true C must be true.

($\leftarrow$) Suppose $(A \wedge B) \to C)$. Suppose A. Then suppose B. Then since both A and B are true, C is true. $\square$

1.1 Quantificational logic

1.1.1 Quantifier negation laws

We have

$$\neg\exists x P(x) \text{ is equivalent to } \forall x \neg P(x)$$
$$\neg\forall x P(x) \text{ is equivalent to } \exists x \neg P(x)$$

Intuition
No matter what $P(x)$ stands for, the formula $\neg\exists x P(x)$ means that there is no value of x for which $P(x)$ is true; this is ths same as saying that for every value of x in the universe of discourse, $P(x)$ is false—meaning $\forall x \neg P(x)$.

Similarly, to say that $\neg\forall x P(x)$ means that it is not the case that for all values of x, $P(x)$ is true. This is equivalent to saying that that there is at least one value of x for which $P(x)$ is false—so $\exists x \neg P(x)$.

1.1.2 Notation

'Exactly one' notation
We write

$$\exists! x P(x) = \exists x (P(x) \wedge \neg\exists y (P(y) \wedge y \neq x))$$

As a shorthand way to write 'there is exactly one value of x such that $P(x)$ is true', or 'there is a unique x such that $P(x)$'.

Specifying quantifiers
We write

$$\forall x \in A\, P(x)$$

to mean that *for every value of x in the set A, $P(x)$ is true.* Similarly,

$$\exists x \in A\, P(x)$$

means *there is at least one value of x in the set A such that $P(x)$ is true.*

Formulas containing bounded quantifiers can also be thought of as abbreviations for more complicated formulas containing only normal, unbounded quantifiers. See that

$$\forall x \in A\, P(x) = \forall x (x \in A \to P(x))$$

and

$$\exists x \in A\, P(x) = \exists x (x \in A \wedge P(x))$$

1.1.3 Negation law for bounded quantifiers

We can show

$$\neg\forall x \in A\, P(x) = \exists x \in A\, \neg P(x)$$

See that

$$\begin{aligned}
&\neg\forall x \in A\, P(x) && \\
&= \neg\forall x(x \in A \to P(x)) && \text{(as defined)} \\
&= \exists x\neg(x \in A \to P(x)) && \text{(negation law)} \\
&= \exists x\neg\neg(x \in A \wedge \neg P(x)) && \text{(conditional law)} \\
&= \exists x(x \in A \wedge \neg P(x)) && \\
&= \exists x \in A\, \neg P(x) && \text{(as defined)}
\end{aligned}$$

Similarly we can show

$$\neg\exists x \in A\, P(x) = \forall x \in A\, \neg P(x)$$

See that

$$\begin{aligned}
&\neg\exists x \in A\, P(x) && \\
&= \neg\exists x(x \in A \wedge P(x)) && \text{(as defined)} \\
&= \forall x\neg(x \in A \wedge P(x)) && \text{(negation law)} \\
&= \forall x(x \in A \to \neg P(x)) && \text{(conditional law)} \\
&= \forall x \in A\, \neg P(x) && \text{(as defined)}
\end{aligned}$$

1.1.4 Vacuously true

It is clear that if $A = \emptyset$ then $\exists x \in A\, P(x)$ will be false regardless of $P(x)$, since there is nothing in A that makes $P(x)$ come true (since there is nothing in A to being with).

Now consider $\forall x \in A\, P(x)$. We can reason that

$$\forall x \in A\, P(x) = \neg \exists x \in A\, \neg P(x) \quad \text{(quantifier negation)}$$

See that if $A = \emptyset$ then this formula will be true, no matter what $P(x)$ is. In this case we say that the statement is *vacuously true*.

Another way to see this is to rewrite

$$\forall x \in A\, P(x) = \forall x (x \in A \to P(x))$$

The only way this can be false is if there is some value of x such that $x \in A$ is true but $P(x)$ false; but there is no such value of x. Intuitively, because the condition cannot be met, it is impossible to provide a counterexample to prove something wrong.

An analogy would be me claiming 'i've never lost a race to Usain Bolt'. This is true, but vacuously so.

1.1.5 Alternate definition for indexed families

Say we are looking for the set $\{p_1, p_2, \ldots, p_{100}\}$; another way of describing this set would be to say that it consists of all numbers p_i, for i an element of the set $I = \{1, 2, 3, \ldots, 100\} = \{i \in \mathbb{N} | 1 \leq i \leq 100\}$. We can write

$$P = \{p_i | i \in I\}$$

Each element p_i in this set is identified by $i \in I$, called the *index* of each element. A set defined this way is called an *indexed family*, and I the *index set*. Although the indices for an indexed family are often numbers, they need not be.

In general, see that any indexed family

$$A = \{x_i | i \in I\}$$

Can also be defined as

$$A = \{x | \exists i \in I \, (x = x_i)\}$$

It follows that the statement

$$x \in \{x_i | i \in I\}$$

means the same thing as

$$\exists i \in I \, (x = x_i)$$

1.1.6 Power set

Suppose A is as set. The *power set* of A, denoted $\mathscr{P}(A)$, is the set whose elements are all subsets of A. In other words,

$$\mathscr{P}(A) = \{x | x \subseteq A\}$$

For instance, the set $A = \{7, 12\}$ has four subsets $\emptyset, \{7\}, \{12\}$, and $\{7, 12\}$; thus, $\mathscr{P}(A) = \{\emptyset, \{7\}, \{12\}, \{7, 12\}\}$.

1.1.7 Intersection and union of a family of sets

Suppose $\mathcal{F}$ is a family of sets. The *intersection* and *union* of $\mathcal{F}$ are the sets $\bigcap \mathcal{F}$ and $\bigcup \mathcal{F}$ are defined as follows:

$$\bigcap \mathcal{F} = \{x | \forall A \in \mathcal{F}(x \in A)\} = \{x | \forall A(A \in \mathcal{F} \to x \in A)\}$$
$$\bigcup \mathcal{F} = \{x | \exists A \in \mathcal{F}(x \in A)\} = \{x | \exists A(A \in \mathcal{F} \wedge x \in A)\}$$

Notice that if A and B are any two sets and $\mathcal{F} = \{A, B\}$, then $\bigcap \mathcal{F} = A \cap B$ and $\bigcup \mathcal{F} = A \cup B$; the definitions of intersection and union of a family of sets are generalisations of our old definitions of the intersection and union of two sets.

Alternative notation
An alternative notation is sometimes used for the union or intersection of an indexed family of sets. Suppose $\mathcal{F} = \{A_i | i \in I\}$, where each A_i is a set, then $\bigcap \mathcal{F}$ and $\bigcup \mathcal{F}$ could also be written as $\bigcap_{i \in I} A_i$ and $\bigcup_{i \in I} A_i$; as such

$$\bigcap \mathcal{F} = \bigcap_{i \in I} A_i = \{x | \forall i \in I(x \in A_i)\}$$
$$\bigcup \mathcal{F} = \bigcup_{i \in I} A_i = \{x | \exists i \in I(x \in A_i)\}$$

1.1.8 More on set notation

One generally defines a set using the elementhood test notation

$$\{x|P(x)\}$$

Where the set consists of all x that satisfy the specified condition $P(x)$. Sometimes this notation can be modified to allow the x before the vertical line to be replaced with a more complex expression. For example, suppose we wanted to define S to be the set of all perfect squres, we could write

$$S = \{n^2|n \in \mathbb{N}\}$$

This is the same as

$$S = \{x|\exists n \in \mathbb{N}(x = n^2)\}$$

See therefore that

$$x \in \{n^2|n \in \mathbb{N}\} = \exists n \in \mathbb{N}(x = n^2)$$

Chapter 2

Proof Strategies

2.0.1 Terminology

We want to state the answer to a mathematical question in the form of a *theorem* that says that if certain assumptions called the *hypotheses* of the theorem are true, then some conclusion must also be true.

An assignment of particular values to these variables is called an *instance* of the theorem, and in order for the theorem to be correct it must be the case that for every instance of the theorem that makes the hypotheses come out true, the conclusion is also true.

If there is even one instance in which the hypotheses are true but the conclusion is false, then the theorem is incorrect; such an instance is called a *counterexample* to the theorem.

As in the next section, we will refer to statements that are known or assumed to be true at some point in the course of figuring out the proof as *givens*, and the statements that remains to be proven at that point as the *goal*.

2.1 To prove a conclusion of the form $P \to Q$

To prove a conclusion of the form $P \to Q$, we can *assume P is true and then prove Q*.

Assuming that P is true amounts to adding P to the lists of hypotheses. If the conclusion of the theorem we are trying to prove has the form $P \to Q$, then we can *transform the problem* by adding P to the list of hypotheses and changing the conclusion form $P \to Q$ to Q.

How we solve this new problem will now then be guided by the logical form of the new conclusion Q, and perhaps also that of the new hypothesis P.

This strategy is one that proves the *goal* of $P \to Q$. Even if the conclusion of a theorem is not a conditional statement, if we transform the problem in such a way that the conditional statement becomes the goal, then we can apply this strategy as the next step in figuring out the proof.

Example:

Theorem. *Suppose a and b are real numbers. If $0 < a < b$ then $a^2 < b^2$.*

Proof. Suppose $0 < a < b$. Multiplying the inequality $a < b$ by the positive number a we can conclude that $a^2 < ab$; similarly multiplying by b we get $ab < b^2$. Therefore $a^2 < ab < b^2$, so $a^2 < b^2$. Thus, if $0 < a < b$ then $a^2 < b^2$. □

We were given as a hypothesis that a and b are real numbers with a conclusion of the form $P \to Q$, where P is the statement $0 < a < b$ and Q the statement $a^2 < b^2$. Thus we start with these statements as given and goal:

Givens	*Goal*
a and b are real numbers	$(0 < a < b) \to (a^2 < b^2)$

As per our strategy, we assume that $0 < a < b$ and try to use this assumption to prove $a^2 < b^2$. In other words we add $0 < a < b$ to the list of givenas and make $a^2 < b^2$ our goal:

Givens	*Goal*
a and b are real numbers	$a^2 < b^2$
$0 < a < b$	

(next page)

Generalising
Here's a restatement of the proof strategy we discussed, in the form we will be using to present proof strategies from now on.

To prove a goal of the form $P \to Q$:
Assume P is true and then prove Q.

Scratch work
Before using strategy:

Givens	*Goal*
—	$P \to Q$
—	

After using strategy:

Givens	*Goal*
—	Q
—	
P	

Form of final proof:
Suppose P.
[Proof of Q goes here.]
Therefore $P \to Q$.

(next page)

2.1.1 Alternative approach

Another approach could utilise the contrapositive law, where

$$P \to Q = \neg Q \to \neg P$$

In other words:

> To prove a goal of the form $P \to Q$:
> Assume Q is false and prove that P is false.

Scratch work
Before using strategy:

Givens	*Goal*
—	$P \to Q$
—	

After using strategy:

Givens	*Goal*
—	$\neg P$
—	
$\neg Q$	

Form of final proof:

> Suppose Q is false.
> > [Proof of $\neg P$ goes here.]
>
> Therefore $P \to Q$.

(next page)

Example
Suppose a, b and c are real numbers and $a > b$. Prove that if $ac \leq bc$ then $c \leq 0$.

Scratch work
Before using strategy:

Givens	*Goal*
a, b and c are real numbers $a > b$	$(ac \leq bc) \to (c \leq 0)$

The contrapositive of the goal is $\neg(c \leq 0) \to \neg(ac \leq bc)$, or $(c > 0) \to (ac > bc)$. As per the previous strategy, we can prove this by adding $c > 0$ to the list of givens and making $ac > bc$ our new goal:

Givens	*Goal*
a, b and c are real numbers $a > b$ $c > 0$	$ac > bc$

Form of final proof:
Suppose $c > 0$.
[Proof of $ac > bc$ goes here.]
Therefore, if $ac \leq bc$ then $c \leq 0$.

Solution

Theorem. *Suppose a, b and c are real numbers and $a > b$. If $ac \leq bc$ then $c \leq 0$.*

Proof. We will prove the contrapositive. Suppose $c > 0$. Then we can multiply both sides of the given inequality $a > b$ by c and conclude that $ac > bc$. Therefore if $ac \leq bc$ then $c \leq 0$. □

2.2 Proofs involving negations and conditionals

2.2.1 Reexpress

We now consider proofs in which the goal has the form $\neg P$. Usually it's easier to prove a positive statement than a negative statement, so it is often helpful to reexpress the goal before proving it. Thus our first strategy for proving negated statements is:

> To prove a goal of the form $\neg P$:
> If possible, reexpress the goal in some other form.

Example
Suppose $A \cap C \subseteq B$ and $a \in C$. Prove that $a \notin A \setminus B$.

Scratch Work

Givens	*Goal*
$A \cap C \subseteq B$	$a \notin A \setminus B$
$a \in C$	

We have a negative goal, which we can try to reexpress as a positive statement:

$$\begin{aligned} a \notin A \setminus B &= \neg(a \in A \wedge a \notin B) && \text{(Definition)} \\ &= a \in A \to a \in B && \text{(Conditional law)} \end{aligned}$$

This gives us

Givens	*Goal*
$A \cap C \subseteq B$	$a \in A \to a \in B$
$a \in C$	

We can apply the strategy for conditional goals:

Givens	*Goal*
$A \cap C \subseteq B$	$a \in B$
$a \in C$	
$a \in A$	

See how the proof is now much more straightforward; from the givens $a \in A$ and $a \in C$ we can conclude $a \in A \cap C$. Since $A \cap C \subseteq B$, it follows that $a \in B$.

Solution

Theorem. *Suppose $A \cap C \subseteq B$ and $a \in C$. Then $a \notin A \setminus B$.*

Proof. Suppose $a \in A$. Then since $a \in C$, $a \in A \cap C$. Since $A \cap C \subseteq B$ it follows that $a \in B$. Thus it cannot be the case that a is an element of A but not B, so $a \notin A \setminus B$. □

2.2.2 Proof by contradiction

Say a goal of the form $\neg P$ cannot be reexpressed as a positive statement, in this case one could attempt *proof by contradiction*—start by assuming P is true, and try to use this assumption to prove that something one already knows is false.

Often this is done by proving a statement that contradicts one of the givens; because one knows that the statement proven is false, the assumption that P was true must have been incorrect—the only remaining possibility then is that P is false.

To prove a goal of the form $\neg P$:
Assume P is true and try to reach a contradiction. In which case P must be false.

Scratch Work

Givens	*Goal*
—	$\neg P$
—	

After using strategy:

Givens	*Goal*
—	Contradiction
—	
P	

Form of final proof:
Suppose P is true.
[Proof of contradiction goes here.]
Thus, P is false.

(next page)

Example
Prove that if $x^2 + y = 13$ and $y \neq 4$ then $x \neq 3$.

Scratch work
The goal is a conditional statement. We treat the antecedent as a given and make the consequent our new goal:

Givens	*Goal*
$x^2 + y = 13$	$x \neq 3$
$y \neq 4$	

Our current idea of the proof structure looks like

Suppose $x^2 + y^2 = 13$ and $y \neq 4$.
    [Proof of $x \neq 3$ goes here.]
Thus, if $x^2 + y^2 = 13$ and $y \neq 4$ then $x \neq 3$.

In this sense, each manipulation of our problem dictates the structure of our final proof. At this point the first and last sentences of the final proof have been produced. What remains is to prove $x \neq 3$ given our manipulated problem.

The goal $x \neq 3$ means $\neg(x = 3)$; we try proof by contradiction and transform the problem as follows:

Givens	*Goal*
$x^2 + y = 13$	Contradiction
$y \neq 4$	
$x = 3$	

Once again, the proof strategy that suggested this transformation also tells us how to fill in a few more sentences of the final proof:

Suppose $x^2 + y^2 = 13$ and $y \neq 4$.
    Suppose $x = 3$
        [Proof of contradiction goes here.]
    Therefore $x \neq 3$
Thus, if $x^2 + y^2 = 13$ and $y \neq 4$ then $x \neq 3$.

The first and last lines go together and indicate that we are proving a conditional statement by assuming the antecedent and proving the consequent. Between these lines is a proof of the consequent $x \neq 3$. This inner proof has the form of a proof by contradiction, as indicated by the second first and second last lines; between these lines we still need to fill in a proof of a contradiction.

At this point we don't have a particular statement as a goal; any impossible conclusion will do. We must look closely at the givens to find a contradiction.
(next page)

Example cont.
Solution

Theorem. *If* $x^2 + y = 13$ *and* $y \neq 4$ *then* $x \neq 3$.

Proof. Suppose $x^2 + y = 13$ and $y \neq 4$. Suppose $x = 3$. Substituting this into the equation $x^2 + y = 13$, we get $9 + y = 13$, so $y = 4$. But this contradicts the fact that $y \neq 4$. Therefore $x \neq 3$. Thus, if $x^2 + y = 13$ and $y \neq 4$ then $x \neq 3$. □

2.2.3 To use a given of the form $\neg P$ in proof by contradiction

If attempting a proof by contradiction with a given $\neg P$. Here is a useful strategy;

To use a given of the form $\neg P$:
Try making P the goal. If one can prove P, then the proof is complete, since P contradicts the given $\neg P$.

Scratch Work
Before using strategy:

Givens	*Goal*
$\neg P$	Contradiction
—	
—	

After using strategy:

Givens	*Goal*
$\neg P$	P
—	
—	

Form of final proof:
[Proof of P goes here.]
Since we already know $\neg P$, this is a contradiction.

Note that proof by contradiction is not restricted to goals of the form $\neg P$, and can be used for any goal.
(next page)

Example
Suppose A, B and C are sets, $A \setminus B \subseteq C$, and x is anything at all. Prove that if $x \in A \setminus C$ then $x \in B$.

Scratch work
We're given $A \setminus B \subseteq C$ and our goal is $x \in A \setminus C \to x \in B$. We use the previously discussed strategy for conditionals (assume antecedent and prove consequent):

Givens	*Goal*
$A \setminus B \subseteq C$	$x \in B$
$x \in A \setminus C$	

Our proof currently looks like
Suppose $x \in A \setminus C$.
[Proof of $x \in B$ goes here.]
Thus, if $x \in A \setminus C$ then $x \in B$.

We attempt proof by contradiction:

Givens	*Goal*
$A \setminus B \subseteq C$	Contradiction
$x \in A \setminus C$	
$x \notin B$	

Now our proof looks like
Suppose $x \in A \setminus C$.
Suppose $x \notin B$
[Proof of contradiction goes here.]
Therefore $x \in B$
Thus, if $x \in A \setminus C$ then $x \in B$.

Because we're doing a proof by contradiction and our last given is now a negated statement, we could try our strategy for using givens of the form $\neg P$. Unfortunately, this strategy suggests making $x \in B$ our goal which just gets us back to where we started. We must look at the other givens to try to find the contradiction.
(next page)

Example cont.
We have

Givens	*Goal*
$A \setminus B \subseteq C$	Contradiction
$x \in A \setminus C$	
$x \notin B$	

In this case writing out the definition of the second given is key to the proof. By definition $x \in A \setminus C$ means $x \in A$ and $x \notin C$. Replacing this given by its definition gives us

Givens	*Goal*
$A \setminus B \subseteq C$	Contradiction
$x \in A$	
$x \notin C$	
$x \notin B$	

Now the third given has the form $\neg P$. We can apply the strategy for using givens of the form $\neg P$ and make $x \in C$ our goal, where showing it would complete the proof by contradicting the given $x \notin C$:

Givens	*Goal*
$A \setminus B \subseteq C$	$x \in C$
$x \in A$	
$x \notin C$	
$x \notin B$	

Once again, we can add a little more to the proof we are gradually forming with each step. We add in the fact that we plan to derive our contradiction by proving $x \in C$; we also add the definition of $x \in A \setminus C$ to the proof.

Suppose $x \in A \setminus C$. This means $x \in A$ and $x \notin C$.
  Suppose $x \notin B$
    [Proof of $x \in C$ goes here.]
    This contradicts the fact that $x \notin C$.
  Therefore $x \in B$
Thus, if $x \in A \setminus C$ then $x \in B$.

We have finally reached a point where the goal follows apparently from the givens—from $x \in A$ and $x \notin B$ we conclude that $x \in A \setminus B$. Since $A \setminus B \subseteq C$ it follows that $x \in C$.
(next page)

Example cont.
Solution

Theorem. *Suppose A, B and C are sets, $A \setminus B \subseteq C$, and x is anything at all. If $x \in A \setminus C$ then $x \in B$.*

Proof. Suppose $x \in A \setminus C$. This means that $x \in A$ and $x \notin C$. Suppose $x \notin B$. Then $x \in A \setminus B$, so since $A \setminus B \subseteq C$, $x \in C$. But this contradicts the fact that $x \notin C$. Therefore $x \in B$. Thus, if $x \in A \setminus C$ then $x \in B$. □

2.2.4 Given conditionals

We consider strategies for using givens of the form $P \to Q$:

To use a given of the form $P \to Q$:
If also given P, or if one can prove P, then one can use this to conclude Q. Another approach would be to use its equivalence to $\neg Q \to \neg P$; if one can prove that Q is false, then one can conclude that P is false.

The first rule says that if you know both P and $P \to Q$ are true, then Q must also be true; this rule is called *modus ponens*. The second rule, called *modus tollens*, says that if you know $P \to Q$ is true and Q is false, then you can conclude that P must also be false.

Example
Suppose $P \to (Q \to R)$. Prove that $\neg R \to (P \to \neg Q)$.

Scratch work
We start with the following situation:

Givens	*Goal*
$P \to (Q \to R)$	$\neg R \to (P \to \neg Q)$

If either P or $\neg(Q \to R)$ gets added to the givens list, then we should consider using modus ponens or modus tollens. For now we concentrate on the goal; being a conditional statement, we assume the antecedent and set the consequent as the new goal:

Givens	*Goal*
$P \to (Q \to R)$	$P \to \neg Q$
$\neg R$	

Our proof currently looks like:

Suppose $\neg R$.
[Proof of $P \to \neg Q$ goes here.]
Therefore $\neg R \to (P \to \neg Q)$.
(next page)

Example cont.
We had

Givens	*Goal*
$P \to (Q \to R)$	$P \to \neg Q$
$\neg R$	

The goal is still a conditional. We use the same strategy again:

Givens	*Goal*
$P \to (Q \to R)$	$\neg Q$
$\neg R$	
P	

Now the proof looks like

Suppose $\neg R$.
    Suppose P.
        [Proof for $\neg Q$ goes here.]
    Therefore $P \to \neg Q$.
Therefore $\neg R \to (P \to \neg Q)$.

Since we know $P \to (Q \to R)$ and P, by modus ponens we can infer $Q \to R$:

Givens	*Goal*
$P \to (Q \to R)$	$\neg Q$
$\neg R$	
P	
$Q \to R$	

We add one more line to our proof:

Suppose $\neg R$.
    Suppose P.
        Since P and $P \to (Q \to R)$, it follows that $Q \to R$.
        [Proof of $\neg Q$ goes here.]
    Therefore $P \to \neg Q$.
Therefore $\neg R \to (P \to \neg Q)$.

Finally, our last step is to use modus tollens. We now know $Q \to R$ and $\neg R$, so by modus tollens we can conclude $\neg Q$.
(next page)

Example cont.
Solution

Theorem. *Suppose* $P \to (Q \to R)$. *Then* $\neg R \to (P \to \neg Q)$.

Proof. Suppose $\neg R$. Suppose P. Since P and $P \to (Q \to R)$, it follows that $Q \to R$. But then since $\neg R$, we can conclude $\neg Q$. Thus $P \to \neg Q$. Therefore $\neg R \to (P \to \neg Q)$. □

2.3 Proofs involving quantifiers

2.3.1 Goals of form $\forall x P(x)$

If you can give a proof of the goal $P(x)$ that would work for all x, then you can conclude that $\forall x P(x)$ must be true. Thus it is important to start the proof with no assumptions about x, that is, x must be *arbitrary*; one must not assume that x is already equal to any other object already under discussion in the proof.

If x is already being used in the proof to stand for some particular object, then one cannot use it to stand for an arbitrary object, and must choose a different variable that is not already being used in the proof, say y, and replace $\forall x P(x)$ with the equivalent statement $\forall y P(y)$.

To prove a goal of the form $\forall P(x)$:

Let x stand for an arbitrary object and prove $P(x)$. The letter x must be a new variable in the proof. If x is already being used to stand for something, then choose an unused variable, say y, to stand for the arbitrary object, and prove $P(y)$.

Scratch work
Before using strategy:

Givens	*Goal*
—	$\forall x P(x)$
—	

After using strategy:

Givens	*Goal*
—	$P(x)$
—	

Form of final proof:

Let x be arbitrary.

[Proof of $P(x)$ goes here.]

Since x was arbitrary, we can conclude that $\forall x P(x)$.

(next page)

Example
Consider a proof covered earlier, but phrased slightly differently: Suppose A, B, and C are sets, and $A \setminus B \subseteq C$. Prove that $A \setminus C \subseteq B$.

Scratch work
Before using strategy:

Givens	*Goal*
$A \setminus B \subseteq C$	$A \setminus C \subseteq B$

As usual we consider the logical form of the goal to plan our strategy. In this case we can write out the definition of $\subseteq$:

Givens	*Goal*
$A \setminus B \subseteq C$	$\forall x(x \in A \setminus C \to x \in B)$

Because the goal has the form $\forall x P(x)$, where $P(x)$ is the statement $x \in A \setminus C \to x \in B$, we will introduce a new variable x into the proof to stand for an arbitrary object and then try to prove $x \in A \setminus C \to x \in B$.

Note that x is a new variable in the proof; it appeared in the logical form of the goal as a bound variable, but, being a bound variable, didn't represent anything in particular. We also haven't used it as a free variable in any statement so it doesn't stand for any particular object. To make sure x is arbitrary we must be careful not to add any assumptions about x to the givens column. Our goal becomes

Givens	*Goal*
$A \setminus B \subseteq C$	$x \in A \setminus C \to x \in B$

Our proof looks like

> Let x be arbitrary.
> [Proof of $x \in A \setminus C \to x \in B$ goes here]
> Since x was arbitrary, we can conclude that $\forall x(x \in A \setminus C \to x \in B)$,
> so $A \setminus C \subseteq B$.

The problem is now exactly the same as it was in the previous example.

Solution

Theorem. *Suppose A, B, and C are sets, and $A \setminus B \subseteq C$. Then $A \setminus C \subseteq B$.*

Proof. Let x be arbitrary. Suppose $x \in A \setminus C$. This means that $x \in A$ and $x \notin C$. Suppose $x \notin B$. Then $x \in A \setminus B$, so since $A \setminus B \subseteq C$, $x \in C$. But this contradicts the fact that $x \notin C$. Therefore $x \in B$. Thus, if $x \in A \setminus C$ then $x \in B$. Since x was arbitrary we can conclude that $\forall x(x \in A \setminus C \to x \in B)$, so $A \setminus C \subseteq B$. □

(next page)

Another example
Suppose A and B are sets. Prove that if $A \cap B = A$ then $A \subseteq B$.

Scratch work

Our goal is $A \cap B = A \to A \subseteq B$. We add the antecedent to the givens list and make the consequent the goal. We write out the logical form of this new goal:

Givens	*Goal*
$A \cap B = A$	$\forall x(x \in A \to x \in B)$

We let x be arbitrary, assume $x \in A$, and probe $x \in B$:

Givens	*Goal*
$A \cap B = A$	$x \in B$
$x \in A$	

Our final proof form looks like

Suppose $A \cap B = A$.
    Let x be arbitrary.
        Suppose $x \in A$.
            [Proof of $x \in B$ goes here.]
        Therefore $x \in A \to x \in B$.
    Since x was arbitrary, we can conclude that $\forall x(x \in A \to x \in B)$,
    so $A \subseteq B$.
Therefore, if $A \cap B = A$ then $A \subseteq B$.

Since $x \in A$ and $A \cap B = A$, it follows that $x \in A \cap B$, so $x \in B$. (recall $x \in A \cap B$ means $x \in A$ and $x \in B$)

Oftentimes the statement that x is arbitrary is left out—when a new variable x is introduced into a proof it is usually understood that it is arbitrary and no assumptions are being made about x. Many of the proofs written so far end with multiple concluding sentences; oftentimes these are condensed into a single sentence, or skipped entirely.

Solution

Theorem. *Suppose A and B are sets. If $A \cap B = A$ then $A \subseteq B$.*

Proof. Suppose $A \cap B = A$, and suppose $x \in A$. Then since $A \cap B = A$, $x \in A \cap B$, so $x \in B$. Since x was an arbitrary element of A, we can conclude that $A \subseteq B$. □

2.3.2 Goals of form $\exists x P(x)$

Proving a goal of the form $\exists x P(x)$ also involves introducing a new variable x into the proof and proving $P(x)$. But in this case x will not be arbitrary; we only need to prove that $P(x)$ is true for *at least one* x. It suffices to assign a particular value to x and prove $P(x)$ for this one value of x.

To prove a goal of the form $\exists x P(x)$:

Try to find a value of x for which $P(x)$ would be true. Start with 'Let $x =$ (decided value)' and proceed to prove $P(x)$ for this x. Once again if the variable name x is already being used in the proof for some other purpose then one should choose an unused variable, say y and rewrite the goal in the equivalent form $\exists y P(y)$ and proceed as before.

Scratch work
Before using strategy:

Givens	*Goal*
—	$\exists x P(x)$
—	

After using strategy:

Givens	*Goal*
—	$P(x)$
—	
$x =$ (the value you decided on)	

Form of final proof:

Let $x =$ (the value you decided on).
[Proof of $P(x)$ goes here.]
Thus, $\exists x P(x)$.

Note that *the reasoning used for finding a specific x will not appear in the final proof.* To justify the conclusion that $\exists x P(x)$ it is only necessary to verify that $P(x)$ comes out true when x is assigned some particular value. How one thought of that value is not part of the justification of the conclusion (be it by trial and error etc.).
(next page)

Example
Prove that for every real number x, if $x > 0$ then there is a real number y such that $y(y + 1) = x$.

Scratch work
Logically, our goal is $\forall x(x > 0 \to \exists y[y(y + 1) = x])$, where the variables x and y in this statement are understood to range over $\mathbb{R}$. We start by letting x be an arbitrary real number, and then assume that $x > 0$ and try to prove that $\exists y[y(y + 1) = x]$:

Givens	*Goal*
$x > 0$	$\exists y[y(y + 1) = x]$

Because our goal has the form $\exists y P(y)$, where $P(y)$ is the statement $y(y+1) = x$, according to our strategy we should try to find a value of y for which $P(y)$ is true. In this case we do this by solving the equation $y(y + 1) = x$ for y.

$$y(y + 1) = x \iff y^2 + y - x = 0 \iff y = \frac{-1^2 \pm \sqrt{1 + 4x}}{2}$$

Note that $\sqrt{1 + 4x}$ is defined since we have $x > 0$ as a given. Also see that we have found two solutions for y, but to prove that $\exists y[y(y + 1) = x]$ we only need one valid y, so either of the the two solutions could be used.

Givens	*Goal*
$x > 0$	$y(y + 1) = x$
$y = (-1 + \sqrt{1 + 4x})/2$	

Our final proof looks like

Let x be an arbitrary real number.
  Suppose $x > 0$.
    Let $y = (-1 + \sqrt{1 + 4x})/2$.
      [Proof of $y(y + 1) = x$ goes here.]
    Thus, $\exists y[y(y + 1) = x]$.
  Therefore $x > 0 \to \exists y[y(y + 1) = x]$.
Since x was arbitrary, we can conclude that $\forall x(x > 0 \to \exists y[y(y + 1) = x])$.

(next page)

Example cont.
Solution

Theorem. *For every real number x, if $x > 0$ then there is a real number y such that $y(y+1) = x$.*

Proof. Let x be an arbitrary real number and suppose $x > 0$. Let

$$y = \frac{-1 + \sqrt{1+4x}}{2}$$

which is defined since $x > 0$. Then

$$\begin{aligned} y(y+1) &= \left(\frac{-1+\sqrt{1+4x}}{2}\right) \cdot \left(\frac{-1+\sqrt{1+4x}}{2} + 1\right) \\ &= \left(\frac{\sqrt{1+4x}-1}{2}\right) \cdot \left(\frac{\sqrt{1+4x}+1}{2}\right) \\ &= \frac{1+4x-1}{4} = \frac{4x}{4} = x \end{aligned}$$

□

2.3.3 Givens of the form $\exists x P(x)$ and $\forall x P(x)$

Givens of the form $\exists x P(x)$
This given says that an object with a certain property exists. A good approach would be to introduce a new variable, say x_0, to stand for this object, and add $P(x_0)$ to the givens list.

To use a given of the form $\exists x P(x)$:
Introduce a new variable x_0 into the proof to stand for an object for which $P(x_0)$ is true. This means that one can now assume that $P(x_0)$ is true.

Logicians call this *existential instantiation*. Note that this is very different from our treatment of goals of the same form—we can assume that x_0 stands for some object for which $P(x_0)$ is true, but we can't assume anything else about x_0.

Givens of the form $\forall x P(x)$
This says that $P(x)$ would be true no matter what value is assigned ot x. One can therefore *choose any value* to be plugged in for x and use this to conclude $P(x)$.

To use a given of the form $\forall x P(x)$
Plug in any value, say a, for x and use this given to conclude that $P(a)$ is true.

This is called *universal instantiation*. Usually, if we have a given of the form $\exists x P(x)$, we should apply existential instantiation to it immediately. On the other hand, we won't be able to apply universal instantiation to a given of the form $\forall x P(x)$ unless we had an idea for a particular value a to plug in for x; we would probably wait until a suitable object a appears in the proof.
(next page)

Example
Suppose $\mathcal{F}$ and $\mathcal{G}$ are families of sets and $\mathcal{F} \cap \mathcal{G} \neq \emptyset$. Prove that $\bigcap \mathcal{F} \subseteq \bigcup \mathcal{G}$.

Scratch work
First we analyse the logical form of the goal by writing out the meaning of the subset: $\forall x(x \in \bigcap \mathcal{F} \to x \in \bigcup \mathcal{G})$. We could go further with the logical forms of the union and intersection, but the analysis done so far is sufficient for us to decide how to get started on the proof; their definitions might be required later, but it is usually best to do only as much of the analysis as is needed to determine the next step of the proof. Going further might introduce unnecessary complication.

We let x be arbitrary, assume $x \in \bigcap \mathcal{F}$, and try to prove $x \in \bigcup \mathcal{G}$.

Givens	*Goal*
$\mathcal{F} \cap \mathcal{G} \neq \emptyset$	$x \in \bigcup \mathcal{G}$
$x \in \bigcap \mathcal{F}$	

Now we further analyse the logical forms:

Givens	*Goal*
$\exists A(A \in \mathcal{F} \cap \mathcal{G})$	$\exists A \in \mathcal{G}(x \in A)$
$\forall A \in \mathcal{F}(x \in A)$	

To prove this new goal means we should try to find a value that would 'work' for A. The second given starts with $\forall A$, so we may not be able to use it until a likely value for A pops up during the course of the proof. The first given, however, starts with $\exists A$, so we should use it immediately; by existential instantiation, we introduce a name, say A_0, for this object, and treat $A_0 \in \mathcal{F} \cap \mathcal{G}$ as a given from now on. It would be redundant to continue to discuss the given statement $\exists A(A \in \mathcal{F} \cap \mathcal{G})$ so we will drop it from our list of givens.

Givens	*Goal*
$A_0 \in \mathcal{F}$	$\exists A \in \mathcal{G}(x \in A)$
$A_0 \in \mathcal{G}$	
$\forall A \in \mathcal{F}(x \in A)$	

An element to plug into the third given has now surfaced, plugging A_0 for A we can conclude that $x \in A_0$; we can treat this as a given. See that the goal can now be easily reached.

Although we translated the statements $x \in \bigcap \mathcal{F}, x \in \bigcup \mathcal{G}$, and $\mathcal{F} \cap \mathcal{G} \neq \emptyset$ into logical symbols in order to figure out how to use them in the proof; these translations are not usually written out it in the final proof—left to the reader to work out their logical forms in order to follow the reasoning.
(next page)

Example cont.
Solution

Theorem. *Suppose $\mathcal{F}$ and $\mathcal{G}$ are families of sets, and $\mathcal{F} \cap \mathcal{G} \neq \emptyset$. Then $\bigcap \mathcal{F} \subseteq \bigcup \mathcal{G}$.*

Proof. Suppose $x \in \bigcap \mathcal{F}$. Since $\mathcal{F} \cap \mathcal{G} = \emptyset$, we can let A_0 be an element of $\mathcal{F} \cap \mathcal{G}$. Thus, $A_0 \in \mathcal{F}$ and $A_0 \in \mathcal{G}$. Since $x \in \bigcap \mathcal{F}$ and $A_0 \in \mathcal{F}$, it follows that $x \in A_0$. But we also know that $A_0 \in \mathcal{G}$, so we can conclude that $x \in \bigcup \mathcal{G}$. $\square$

Another example
Suppose B is a set and $\mathcal{F}$ is a family of sets. Prove that if $\bigcup \mathcal{F} \subseteq B$ then $\mathcal{F} \subseteq \mathscr{P}(B)$

Scratch work
We assume $\bigcup \mathcal{F} \subseteq B$ and try to prove $\mathcal{F} \subseteq \mathscr{P}(B)$, which means $\forall x(x \in \mathcal{F} \rightarrow x \in \mathscr{P}(B))$, we let x be arbitrary, assume $x \in \mathcal{F}$, and set $x \in \mathscr{P}(B)$ as our goal. Note that x in this case is a set:

Givens	*Goal*
$\bigcup \mathcal{F} \subseteq B$	$x \in \mathscr{P}(B)$
$x \in \mathcal{F}$	

To further analyse this goal, we use the definition of the power statement $x \in \mathscr{P}(B) = x \subseteq B$; which further means $\forall y(y \in x \rightarrow y \in B)$. We must therefore introduce another arbitrary object into the proof; we let y be arbitrary, assume $y \in x$, and try to prove $y \in B$.

Givens	*Goal*
$\bigcup \mathcal{F} \subseteq B$	$y \in B$
$x \in \mathcal{F}$	
$y \in x$	

The first given can be written as $\forall z(z \in \bigcup \mathcal{F} \rightarrow z \in B)$, which further means $\forall z(\exists A \in \mathcal{F}(z \in A) \rightarrow z \in B)$

Givens	*Goal*
$\forall z(\exists A \in \mathcal{F}(z \in A) \rightarrow z \in B)$	$y \in B$
$x \in \mathcal{F}$	
$y \in x$	

The conclusion is now easily reachable.
(next page)

Example cont.
Solution

Theorem. *Suppose B is a set and $\mathcal{F}$ is a family of sets. If $\bigcup \mathcal{F} \subseteq B$ then $\mathcal{F} \subseteq \mathscr{P}(B)$.*

Proof. Suppose $\bigcup \mathcal{F} \subseteq B$. Let x be an arbitrary element of $\mathcal{F}$. Let y be an arbitrary element of x. Since $y \in x$ and $x \in \mathcal{F}$, $y \in \bigcup \mathcal{F}$. But since $\bigcup \mathcal{F} \subseteq B$, $y \in B$. Since y was an arbitrary element of x, we can conclude that $x \subseteq B$, so $x \in \mathscr{P}(B)$. But x was an arbitrary element of $\mathcal{F}$, so $\mathcal{F} \subseteq \mathscr{P}(B)$, as required. □

2.3.4 Example: For all integers a, b, c, if $a \mid b$ and $b \mid c$ then $a \mid c$

For this proof, we need the following definition:

Definition. For any integers x and y, we'll say that x *divides* y (or y *is divisible by* x) if $\exists k \in \mathbb{Z}(kx = y)$. We use the notation $x \mid y$ to mean 'x divides y' and $x \nmid y$ to mean 'x does not divide y'.

Theorem. *For all integers a, b, and c, if $a \mid b$ and $b \mid c$ then $a \mid c$.*

Proof. Let a, b, and c be arbitrary integers and suppose $a \mid b$ and $b \mid c$. Since $a \mid b$, we can choose some integer m such that $ma = b$. Similarly, since $b \mid c$, we can choose an integer n such that $nb = c$. Therefore $c = nb = nma$, so since nm is an integer, $a \mid c$. □

2.4 Proofs involving conjunctions and biconditionals

2.4.1 Goals and givens of form $P \wedge Q$

This is fairly straightforward:

> To prove a goal of the form $P \wedge Q$:
> Prove P and Q separately.

In other words, a goal of the form $P \wedge Q$ is treated as two separate goals: P and Q. The same is true of givens of the form $P \wedge Q$:

> To use a given of the form $P \wedge Q$:
> Treat this given as two separate givens: P and Q.

We've already used these ideas in previous examples.

Example
Suppose $A \subseteq B$, and A and C are disjoint. Prove that $A \subseteq B \setminus C$.

Scratch work

Givens	*Goal*
$A \subseteq B$	$A \subseteq B \setminus C$
$A \cap C = \emptyset$	

The logical form of the goal has the form $\forall x(x \in A \to x \in B \setminus C)$, so we let x be arbitrary, assume $x \in A$, and try to prove $x \in B \setminus C$, which means $x \in B \wedge x \notin C$. As per our strategy we split this into two goals and prove them separately.

Givens	*Goal*
$A \subseteq B$	$x \in B$
$A \cap C = \emptyset$	$x \notin C$
$x \in A$	

The final proof will have the form

> Let x be arbitrary.
> > Suppose $x \in A$.
> > > [Proof of $x \in B$ goes here.]
> > > [Proof of $x \in B$ goes here.]
> > > Thus, $x \in B \wedge x \notin C$, so $x \in B \setminus C$.
> > Therefore $x \in A \to x \in B \setminus C$.
> Since x was arbitrary, $\forall x(x \in A \to x \in B \setminus C)$, so $A \subseteq B \setminus C$.

(next page)

Example cont.
The first goal, $x \in B$, follows from $x \in A$ and $A \subseteq B$; that much is clear. The second goal, $x \notin C$, follows from $x \in A$ and $A \cap C = \emptyset$; this can be seen from analysing the logical form of $A \cap C = \emptyset$:

$$\begin{aligned} A \cap C = \emptyset \text{ is equivalent to } & \neg\exists y(y \in A \wedge y \in C) && \text{(definition)} \\ =& \forall y \neg(y \in A \wedge y \in C) && \text{(Quantifier negation)} \\ =& \forall y(y \in A \rightarrow y \notin C) && \text{(definition)} \end{aligned}$$

From which we can conclude that $x \in A \rightarrow x \notin C$.

Solution

Theorem. *Suppose $A \subseteq B$, and A and C are disjoint. Then $A \subseteq B \setminus C$.*

Proof. Suppose $x \in A$. Since $A \subseteq B$, it follows that $x \in B$, and since A and C are disjoint, we must have $x \notin C$. Thus, $x \in B \setminus C$. Since x was an arbitrary element of A, we can conclude that $A \subseteq B \setminus C$. □

2.4.2 Goals and givens of the form $P \leftrightarrow Q$ (part 1)

$P \leftrightarrow Q$ is equivalent to $(P \to Q) \wedge (Q \to P)$; as per the previous strategy, this sshould be treated as two separate givens or goals: $P \to Q$ and $Q \to P$.

To prove a goal of the for $P \leftrightarrow Q$:
Prove $P \to Q$ and $Q \to P$ separately.

To use a given of the form $P \leftrightarrow Q$:
Treat this as two separate givens $P \to Q$ and $Q \to P$.

2.4.3 Example: Suppose x is an integer. Prove that x is even iff x^2 is even

We use the definition

Definition. An integer x is *even* if $\exists k \in \mathbb{Z}(x = 2k)$, and x is *odd* if $\exists k \in \mathbb{Z}(x = 2k + 1)$.

We also use the fact that every integer is either even or odd, but not both.

Example
Suppose x is an integer. Prove that x is even iff x^2 is even.

Scratch work
The goal is (x is even)$\leftrightarrow$(x^2 is even), so we prove the two goals (x is even)$\to$(x^2 is even) and (x^2 is even)$\to$(x is even) separately. So first:

Givens	*Goal*
$x \in \mathbb{Z}$	x^2 is even
x is even	

We reveal the logical forms by writing out the definition of *even*:

Givens	*Goal*
$x \in \mathbb{Z}$	$\exists j \in \mathbb{Z}(x^2 = 2j)$
$\exists k \in \mathbb{Z}(x = 2k)$	

The second given starts with $\exists k$, we immediately introduce k_0 to stand for some integer for which the enclosed statement is true: $k_0 \in \mathbb{Z}$, and $x = 2k_0$.

Givens	*Goal*
$x \in \mathbb{Z}$	$\exists j \in \mathbb{Z}(x^2 = 2j)$
$\exists k \in \mathbb{Z}(x = 2k)$	
$k_0 \in \mathbb{Z}$	
$x = 2k_0$	

We need an integer j such that $x^2 = 2j$. Plugging in $2k_0$ for x we get $x^2 = 4k_0^2 = 2(2k_0^2)$ See that $j = 2k_0^2$ is a possible choice.
(next page)

Example cont.
To prove the second goal (x^2 is even)$\rightarrow$(x is even), we'll prove the contrapositive (x is not even)$\rightarrow$(x^2 is not even) instead. Since any integer is either even or odd but not both, this is equivalent to (x is odd)$\rightarrow$(x^2 is odd):

Givens	*Goal*
$x \in \mathbb{Z}$	x^2 is odd
x is odd	

Again we write out the definition of *odd* to reveal the logical forms:

Givens	*Goal*
$x \in \mathbb{Z}$	$\exists j \in \mathbb{Z}(x^2 = 2j + 1)$
$\exists k \in \mathbb{Z}(x = 2k + 1)$	

Considering the second given, we use existential instantiation:

Givens	*Goal*
$x \in \mathbb{Z}$	$\exists j \in \mathbb{Z}(x^2 = 2j + 1)$
$\exists k \in \mathbb{Z}(x = 2k + 1)$	
$k \in \mathbb{Z}$	
$x = 2k_0 + 1$	

We must now find an integer j such that $x^2 = 2j + 1$. Plugging in $2k_0 + 1$ for x we get $x^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$, so $j = 2k^2 + 2k$ is a possible choice.

The two conditional statements we've proven can be thought of as representing the two directions $\rightarrow$ and $\leftarrow$. The two parts of the proof are sometimes labelled with the symbols $\rightarrow$ and $\leftarrow$.

In each part, we prove a statement that asserts the existence of a number with certain properties. We called this number j, but note that j was not mentioned explicily in the statement of the problem, so we choose not to mention it in the final proof.

Solution

Theorem. *Suppose x is an integer. Then x is even iff x^2 is even.*

Proof. ($\rightarrow$) Suppose x is even. Then for some integer k_0, $x = 2k_0$. Therefore $x^2 = 4k^2 = 2(2k^2)$, so since $2k^2$ is an integer, x^2 is even. Thus if x is even then x^2 is even.

($\leftarrow$) Suppose x is odd. Then $x = 2k + 1$ for some integer k. Therefore $x^2 = (2k+1)^2 = 4k^2+4k+1 = 2(2k^2+2k)+1$, so since $(2k^2+2k)$ is an integer, x^2 is odd. Thus if x^2 is even then x is even. $\square$

2.4.4 Proving $\forall x \neg P(x) \leftrightarrow \neg \exists x P(x)$

Scratch work

($\rightarrow$) We first prove $\forall x \neg P(x) \rightarrow \neg \exists x P(x)$, so we assume $\forall x \neg P(x)$ and try to prove $\neg \exists x P(x)$. Our goal is a negated statement, and trying to rexpress it would require the use of the very equivalence we are trying to prove. We therefore try proof by contradiction:

Givens	*Goal*
$\forall x \neg P(x)$	Contradiction
$\exists x P(x)$	

The second given starts with and existential quantifier, so we use it immediately and let x_0 stand for some object for which the statement $P(x_0)$ is true. But now plugging x_0 into the first given we get $\neg P(x_0)$, which gives us the contradiction we need.

($\leftarrow$) For the other direction we assume $\neg \exists x P(x)$ and try to prove $\forall x \neg P(x)$. We let x be arbitrary and try to prove $\neg P(x)$; once again we have a negated goal that can't be reexpressed, so we use proof by contradiction:

Givens	*Goal*
$\neg \exists x P(x)$	Contradiction
$P(x)$	

The first line is a negated statement. This suggests that we could get our contradiction by proving $\exists x P(x)$. We therefore set this as our goal

Givens	*Goal*
$\neg \exists x P(x)$	$\exists x P(x)$
$P(x)$	

Our second given $P(x)$ tells us that our arbitrary x is the value we need to show the goal.

Solution

Theorem. $\forall x \neg P(x) \leftrightarrow \neg \exists x P(x)$.

Proof. ($\rightarrow$) Suppose $\forall x \neg P(x)$, and suppose $\exists x P(x)$. Then we can choose some x_0 such that $P(x_0)$ is true. But since $\forall x \neg P(x)$, we can conclude that $\neg P(x_0)$, which is a contradiction. Therefore $\forall x \neg P(x) \rightarrow \neg \exists x P(x)$.

($\leftarrow$) Suppose $\neg \exists x P(x)$. Let x be arbitrary, and suppose $P(x)$. Since we have a specific x for which $P(x)$ is true, it follows that $\exists x P(x)$, which is a contradiction; therefore $\neg P(x)$. Since x was arbitrary, we can conclude that $\forall x \neg P(x)$, so $\neg \exists x P(x) \rightarrow \forall x \neg P(x)$. $\square$

2.4.5 Goals and givens of the form $P \leftrightarrow Q$ (part 2)

Sometimes in a proof of a goal of the form $P \leftrightarrow Q$ the steps in the proof of $Q \to P$ are the same as the steps used to prove $P \to Q$, just in reverse order. In this case one might be able to simplify the proof by writing it as a string of equivalences, starting with P and ending with Q.

For example suppose one could prove $P \to R \to Q$ (which is $P \to R$ and $R \to Q$), and $Q \to R \to P$ (which is $Q \to R$ and $R \to P$). This would be the same as asserting that $P \leftrightarrow R \leftrightarrow Q$ ($P \leftrightarrow R$ and $R \leftrightarrow Q$); as such both statements imply the goal $P \leftrightarrow Q$. One would present this kind of proof by simply writing the string of equivalences

$$P \text{ iff } R \text{ iff } Q$$

Since this is just an abbreviation for 'P iff R and R iff Q (and therefore P iff Q)'.

Example
Suppose A, B, and C are sets. Prove that $A \cap (B \setminus C) = (A \cap B) \setminus C$.

Scratch work
The equation $A \cap (B \setminus C) = (A \cap B) \setminus C$ means $\forall x(x \in A \cap (B \setminus C) \leftrightarrow x \in (A \cap B) \setminus C)$; it can also mean $[A \cap (B \setminus C) \subseteq (A \cap B) \setminus C] \wedge [(A \cap B) \setminus C \subseteq A \cap (B \setminus C)]$.

This suggests two approaches to the proof: we could let x be arbitrary and then prove $x \in A \cap (B \setminus C) \leftrightarrow x \in (A \cap B) \setminus C)$, or we could prove the two statements $A \cap (B \setminus C) \subseteq (A \cap B) \setminus C$ and $(A \cap B) \setminus C \subseteq A \cap (B \setminus C)$.

In this case we use the first approach. Considering the first half ($\to$) of the equivalence, we assume $x \in A \cap (B \setminus C)$ and try to prove $x \in (A \cap B) \setminus C$:

Givens	*Goal*
$x \in A \cap (B \setminus C)$	$x \in (A \cap B) \setminus C$

To see the logical forms of the given and goal, we write out their definitions as follows:

$$x \in A \cap (B \setminus C) \text{ iff } x \in A \wedge x \in B \setminus C \text{ iff } x \in A \wedge x \in B \wedge x \notin C$$
$$x \in (A \cap B) \setminus C \text{ iff } x \in A \cap B \wedge x \notin C \text{ iff } x \in A \wedge x \in B \wedge x \notin C$$

At this point it is clear that the reasoning involved in the ($\leftarrow$) direction of the proof will be exactly the same, but with the given and goal columns reversed. As such we might try to shorten the proof by writing it as a string of equivalences. (next page)

Example cont.
Solution

Theorem. *Suppose A, B, and C are sets. Then $A \cap (B \setminus C) = (A \cap B) \setminus C$.*

Proof. Let x be arbitrary. Then

$$\begin{aligned} x \in A \cap (B \setminus C) &\text{ iff } x \in A \wedge x \in B \setminus C \\ &\text{iff } x \in A \wedge x \in B \wedge x \notin C \\ &\text{iff } x \in A \cap B \wedge x \notin C \\ &\text{iff } x \in (A \cap B) \setminus C \end{aligned}$$

Thus, $\forall x(x \in A \cap (B \setminus C) \leftrightarrow x \in (A \cap B) \setminus C)$, so $A \cap (B \setminus C) = (A \cap B) \setminus C$. □

Another example
A similar technique can be used when proving equations. Consider proving that for any real numbers a and b,

$$(a+b)^2 - 4(a-b)^2 = (3b-a)(3a-b)$$

Scratch work
The goal has the form $\forall a \forall b[(a+b)^2 - 4(a-b)^2 = (3b-a)(3a-b)]$; so we start by letting a and b be arbitrary real numbers and try to prove the equation. Multiplying our both sides gives us:

$$\begin{aligned} (a+b)^2 - 4(a-b)^2 &= a^2 + 2ab + b^2 - 4(a^2 - 2ab + b^2) \\ &= -3a^2 + 10ab - 3b^2; \\ (3b-a)(3a-b) &= 9ab - 3a^2 - 3b^2 + ab = -3a^2 + 10ab - 3b^2 \end{aligned}$$

Clearly both sides are equal. The simplest way to write the proof is using a string of equalities.

Solution

Theorem. *For any real numbers a and b.*

$$(a+b)^2 - 4(a-b)^2 = (3b-a)(3a-b)$$

Proof. Let a and b be arbitrary real numbers. Then

$$\begin{aligned} (a+b)^2 - 4(a-b)^2 &= a^2 + 2ab + b^2 - 4(a^2 - 2ab + b^2) \\ &= -3a^2 + 10ab - 3b^2 \\ &= ab - 3a^2 - 3b^2 + ab = (3b-a)(3a-b) \end{aligned}$$

□

2.4.6 Example: For every integer n, $6 \mid n$ iff $2 \mid n$ and $3 \mid n$

Theorem. *For every integer n, $6 \mid n$ iff $2 \mid n$ and $3 \mid n$.*

Proof. ($\rightarrow$) Suppose $6 \mid n$. Then we can choose an integer k such that $6k = n$. Therefore $n = 6k = 2(3k)$, so $2 \mid n$, and similarly $n = 6k = 3(2k)$, so $3 \mid n$.

($\leftarrow$) Suppose $2 \mid n$ and $3 \mid n$. Then we can choose integers j and k such that $n = 2j$ and $n = 3k$. Therefore $6(j-k) = 6j - 6k = 3(2j) - 2(3k) = 3n - 2n = n$, so $6 \mid n$. □

2.5 Proofs involving disjunctions

Suppose one of the givens in the proof has the form $P \vee Q$. One way to do the proof would be to consider these two possibilities in turn—first assume that P is true and use this assumption to prove the goal, then assume Q is true and give another proof of the goal.

Although we don't know which of these assumptions is correct, the given $P \vee Q$ means that *one* of then must be correct, and that whichever one it is, we would have then shown that it implies the goal.

The two possibilities considered separately in this type of proof—the possibility that P is true and the possibility that Q is true—are called *cases*. The given $P \vee Q$ justifies the use of these two cases by guaranteeing that these cases cover all of the possibilities; we say in this situation that the cases are *exhaustive*. Any proof can be broken into two or more cases at an time, as long as the cases are exhaustive.

To use a given of the form $P \vee Q$:

Break the proof into cases. For case 1, assume P is true and use this assumption to prove the goal. For case 2, assume Q is true and give another proof of the goal.

Scratch work
Before using strategy:

Givens	*Goal*
$P \vee Q$	—
—	

After using strategy:

Case 1:

Givens	*Goal*
P	—
—	

Case 2:

Givens	*Goal*
Q	—
—	

Form of final proof:

Case 1. P is true.
[Proof of goal goes here.]
Case 2. Q is true.
[Proof of goal goes here.]
Since we know $P \vee Q$, these cases cover all the possibilities. Therefore the goal must be true.

(next page)

Example
Suppose that A, B, and C are sets. Prove that if $A \subseteq C$ and $B \subseteq C$ then $A \cup B \subseteq C$.

Scratch work
We assume that $A \subseteq C$ and $B \subseteq C$ and prove $A \cup B \subseteq C$. Writing out the goal using logical symbols gives us the following givens and goal:

Givens	*Goal*
$A \subseteq C$	$\forall x(x \in A \cup B \to x \in C)$
$B \subseteq C$	

We let x be arbitrary, assume $x \in A \cup B$, and try to prove $x \in C$. Thus we now have a new given $x \in A \cup B$, which we write as $x \in A \vee x \in B$, and our goal is now $x \in C$:

Givens	*Goal*
$A \subseteq C$	$x \in C$
$B \subseteq C$	
$x \in A \vee x \in B$	

The goal cannot be analysed any further at this point, so we look at the givens. The first and second givens are useful if we ever come across objects that are elements of A or B. Moving on to the third given, because it has the form $P \vee Q$, we try proof by cases; for the first case we assume $x \in A$ and for the second $x \in B$. For the first case we have

Givens	*Goal*
$A \subseteq C$	$x \in C$
$B \subseteq C$	
$x \in A$	

We've come across an element of A; using our first given we can conclude that $x \in C$ proving our goal. The reasoning for the second case is similar, using the second given instead of the first.

Solution

Theorem. *Suppose that A, B, and C are sets. If $A \subseteq C$ and $B \subseteq C$ then $A \cup B \subseteq C$.*

Proof. Suppose $A \subseteq C$ and $B \subseteq C$, and let x be an arbitrary element of $A \cup B$. Then either $x \in A$ or $x \in B$.

Case 1. $x \in A$. Then since $A \subseteq C$, $x \in C$.

Case 2. $x \in B$. Then since $B \subseteq C$, $x \in C$.

Since we know that either $x \in A$ or $x \in B$, these cases cover all the possibilities, so we can conclude that $x \in C$. Since x was an arbitrary element of $A \cup B$, this means that $A \cup B \subseteq C$. □

(next page)

Cont.
Note that the cases in this proof are not exclusive—it is possible for both $x \in A$ and $x \in B$ to be true, so some values of x might fall under both cases. There is nothing wrong with this; the cases in a proof by cases must cover all possibilities, there is no harm in covering some possibilities more than once—the cases must be exhaustive, they need not be exclusive.

2.5.1 Goal of form $P \vee Q$ (part 1)

Proof by cases is sometimes also helpful if one is proving a goal of the form $P \vee Q$. If one can prove P in some cases and Q in others, then as long as the cases are exhaustive then one can conclude that $P \vee Q$ is true.

This method is particularly useful if one of the givens also has the form of a disjunction, since that would naturally imply a proof by cases.

> To prove a goal of the form $P \vee Q$:
> Break the proof into cases. In each case, either prove P or Q.

Example
Suppose that A, B and C are sets. Prove that $A \setminus (B \setminus C) \subseteq (A \setminus B) \cup C$.

Scratch work
The goal has the form $\forall x(x \in A \setminus (B \setminus C) \to x \in (A \setminus B) \cup C)$, we let x be arbitrary, assume $x \in A \setminus (B \setminus C)$, and try to prove $x \in (A \setminus B) \cup C$. Writing these statements out in logical symbols gives us

Givens	*Goal*
$x \in A \wedge \neg(x \in B \wedge x \notin C)$	$(x \in A \wedge x \notin B) \vee x \in C$

We split the given into two separate givens $x \in A$ and $\neg(x \in B \wedge x \notin C)$, and since the second is a negated statement we use De Morgan's law to reexpress it as a positive statement $x \notin B \vee x \in C$:

Givens	*Goal*
$x \in A$ $x \notin B \vee x \in C$	$(x \in A \wedge x \notin B) \vee x \in C$

Now the second given and the goal are both disjunctions, so we try considering the two cases $x \notin B$ and $x \in C$ suggested by the second given.
(next page)

Example cont.
If in each case we can either prove $x \in A \wedge x \notin B$ or $x \in C$, then the proof will be complete. For the first case we assume $x \notin B$

Givens	*Goal*
$x \in A$	$(x \in A \wedge x \notin B) \vee x \in C$
$x \notin B$	

In this case the goal is clearly true. For the second case we assume $x \in C$, and once again the goal is clearly true.

Solution

Theorem. *Suppose that A, B, and C are sets. Then $A \setminus (B \setminus C) \subseteq (A \setminus B) \cup C$.*

Proof. Suppose $x \in A \setminus (B \setminus C)$. Then $x \in A$ and $x \notin B \setminus C$. Since $x \notin B \setminus C$ it follows that either $x \notin B$ or $x \in C$. We will consider these cases separately.

Case 1. $x \notin B$. Then since $x \in A$, $x \in A \setminus B$, so $x \in (A \setminus B) \cup C$.
Case 2. $x \in C$. Then clearly $x \in (A \setminus B) \cup C$.

Since x was an arbitrary element of $A \setminus (B \setminus C)$, we can conclude that $A \setminus (B \setminus C) \subseteq (A \setminus B) \cup C$. □

2.5.2 Example: For integer x, the remainder when x^2 is divided by 4 is either 0 or 1

Sometimes one may find it useful to break a proof into cases even if the cases are not naturally suggested by a given of the form $P \vee Q$. Any proof can be broken into cases at any time, as long as the cases exhaust all of the possibilities.

Example
Prove that for every integer x, the remainder when x^2 is divided by 4 is either 0 or 1.

Scratch work
We start by letting x be an arbitrary integer and then try to prove that the remainder when x^2 is divided by 4 is either 0 or 1:

Givens	*Goal*
$x \in \mathbb{Z}$	($x^2 \div 4$ has remainder 0) $\vee$ ($x^2 \div 4$ has remainder 1)

(next page)

Cont.
Since the goal is a disjunction, breaking the proof into cases seems like a probable approach, but there is no given that suggests what cases to use. However, trying out a few values for x suggests the right cases

x	x^2	quotient of $x^2 \div 4$	remainder of $x^2 \div 4$
1	1	0	1
2	4	1	0
3	9	2	1
4	16	4	0
5	25	6	1
6	36	9	0

It appears that the remainder is 0 when x is even and 1 when x is odd. These are the cases we will use. Thus for case 1 we assume x is even and try to prove that the remainder is 0, and for case 2 we assume x is odd and prove that the remainder is 1. Because every integer is either even or odd, these cases are exhaustive.

Filling in the definition of *even*, here are our givens and goal for case 1:

Givens	*Goal*
$x \in \mathbb{Z}$	$x^2 \div 4$ has remainder 0
$\exists k \in \mathbb{Z}(x = 2k)$	

Immediately using the second given and letting k stand for some particular integer in which $x = 2k$, then $x^2 = (2k)^2 = 4k^2$, so clearly when we divide x^2 by 4 the quotient is k^2 and the remainder is 0. Case 2 is quite similar:

Givens	*Goal*
$x \in \mathbb{Z}$	$x^2 \div 4$ has remainder 1
$\exists k \in \mathbb{Z}(x = 2k + 1)$	

Once again we use the second given immediately and let k stand for an integer for which $x = 2k + 1$. Then $x^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 4(k^2 + k) + 1$, so when x^2 is divided by 4 the quotient is $k^2 + k$ and the remainder is 1.

Solution

Theorem. *For every integer x, the remainder when x^2 is divided by 4 is either 0 or 1.*

Proof. Suppose x is an integer, we consider two cases.

Case 1. x is even. Then $x = 2k$ for some integer k, so $x^2 = 4k^2$. Clearly the remainder when x^2 is divided by 4 is 0.

Case 2. x is odd. Then $x = 2k+1$ for some integer k, so $x^2 = 4k^2+4k+1 = 4(k^2 + k) + 1$. Clearly in this case the remainder when x^2 is divided by 4 is 1. □

2.5.3 Goal of form $P \vee Q$ (part 2)

Sometimes in a proof with a goal of the form $P \vee Q$ it may be difficult to figure out how to break the proof into cases. Another strategy would be to simply assume that P is true in case 1 and assume that it is false in case 2; since P is either true or false, these cases are exhaustive.

In the first case, since one assumes P is true, the goal $P \vee Q$ is certainly true and no further reasoning is required. In the second case we assume P is false, and the only way for the goal $P \vee Q$ to be true is if Q is true; thus to complete this case one should try to prove Q.

> To prove a goal of the form $P \vee Q$:
> If P is true, then clearly the goal $P \vee Q$ is true, so one only needs to consider the case where P is false; one can complete the proof by proving that for that case Q is true.

Scratch work
Before using strategy:

Givens	*Goal*
—	$P \vee Q$
—	

After using strategy:

Givens	*Goal*
—	Q
—	
$\neg P$	

Form of final proof:

> If P is true, then of course $P \vee Q$ is true. Now suppose P is false.
> [Proof of Q goes here.]
> Thus $P \vee Q$ is true.

Note that this strategy is the same transformation one would use when proving the goal $\neg P \to Q$. This isn't suprising given that the statements $P \vee Q$ and $\neg P \to Q$ are equivalent.

Note that the roles of P and Q could also be reversed in this strategy—we can also prove $P \vee Q$ by assuming that Q is false and proving P.
(next page)

Example
Prove that for every real number x, if $x^2 \geq x$ then either $x \leq 0$ or $x \geq 1$.

Scratch work
Our goal is $\forall x(x^2 \geq x \to (x \leq 0 \vee x \geq 1))$, so to start we let x be an arbitrary real number, assume $x^2 \geq x$, and set $x \leq 0 \vee x \geq 1$ as our goal:

Givens	*Goal*
$x^2 \geq x$	$x \leq 0 \vee x \geq 1$

As per our strategy, to prove this we can either assume $x > 0$ and prove $x \geq 1$ or assume $x < 1$ and prove $x \leq 0$. We take the first approach.

Givens	*Goal*
$x^2 \geq x$	$x \geq 1$
$x > 0$	

The proof is now easy. Since $x > 0$, we can divide the given inequality $x^2 \geq x$ by x to get the goal $x \geq 1$.

Solution

Theorem. *For every real number x, if $x^2 \geq x$ then either $x \leq 0$ or $x \geq 1$.*

Proof. Suppose $x^2 \geq x$. If $x \leq 0$, then of course $x \leq 0$ or $x \geq 1$. Now suppose $x > 0$. Then we can divide both sides of the inequality $x^2 \geq x$ by x to conclude that $x \geq 1$. Thus either $x \leq 0$ or $x \geq 1$. □

2.5.4 Given of form $P \vee Q$ (part 2)

The equivalence of $P \vee Q$ and $\neg P \rightarrow Q$ suggests a rule of inference called *disjunctive syllogism* for using a given statement of the form $P \vee Q$:

> To use a given of the form $P \vee Q$:
> If one knows $\neg P$, or if one can prove P is false, then one can use this given to conclude that Q is true. Similarly if given $\neg Q$ or if one can prove Q is false, then one can conclude that P is true.

2.6 Existence and Uniqueness proofs

We consider proofs in which the goal has the form $\exists! x P(x)$, meaning 'there is exactly one x such that $P(x)$'. This can be thought as an abbreviation for the formula $\exists x(P(x) \wedge \neg\exists y(P(y) \wedge y \neq x))$.

As discussed, we could prove this goal by finding a particular x for which we could prove both $P(x)$ and $\neg\exists y(P(y) \wedge y \neq x))$. The second part would involve proving a negated statement, but we can reexpress it as an equivalent positive statement:

$$\begin{aligned} &\neg\exists y(P(y) \wedge y \neq x)) \\ &\quad \text{is equivalent to } \forall y \neg(P(y) \wedge y \neq x) && \text{(quantifier negation law)} \\ &\text{which is equivalent to } \forall y(P(y) \rightarrow y = x) && \text{(conditional law)} \end{aligned}$$

Thus see that $\exists! x P(x)$ could also be written as $\exists x(P(x) \wedge \forall y(P(y) \rightarrow y = x))$. We can prove that the following are all equivalent:

1. $\exists x(P(x) \wedge \forall y(P(y) \rightarrow y = x))$
2. $\exists x \forall y(P(y) \leftrightarrow y = x)$
3. $\exists x P(x) \wedge \forall y \forall z((P(y) \wedge P(z)) \rightarrow y = z)$

Scratch work
We have three biconditionals to prove: statement 1 iff statement 2, statement 1 iff statement 3, and statement 2 iff statement 3. Rather that proving 6 different conditionals, we will prove that statement 1 implies statement 2, statement 2 implies statement 3, and statement 3 implies statement 1.

Although we will not give a separate proof of the reverse direction, they are implied by these three proofs; for instance statement 2 implies 3 which implies 1, so statement 2 implies statement 1.

Because we'll be proving three conditional statements, our proof will have three parts, which we will label $1 \rightarrow 2, 2 \rightarrow 3$, and $3 \rightarrow 1$. We'll need to consider each part separately.
(next page)

1→2: $\exists x(P(x) \wedge \forall y(P(y) \to y = x)) \to \exists x \forall y(P(y) \leftrightarrow y = x)$
We assume statement 1 and prove statement 2. Given the existential quantifier in the given we choose a name, say x_0, for some object where both $P(x_0)$ and $\forall y(P(y) \to y = x_0)$:

Givens	*Goal*
$P(x_0)$	$\exists x \forall y(P(y) \leftrightarrow y = x)$
$\forall y(P(y) \to y = x_0)$	

Our goal starts with an existential quantifier, so we should try to find a value of x that makes the rest of the statement come out true. The obvious choice here is $x = x_0$; plugging in x_0, we must now prove $\forall y(P(y) \leftrightarrow y = x_0)$. We let y be arbitrary and prove both directions of the biconditional. $(\to)$ is clear by the second given. For $(\leftarrow)$, suppose $y = x_0$; since we have $P(x_0)$ as a given, we have $P(y)$.

2→3: $\exists x \forall y(P(y) \leftrightarrow y = x) \to \exists x P(x) \wedge \forall y \forall z((P(y) \wedge P(z)) \to y = z)$
As per our usual strategy for existential given statements, we let x_0 be some object such that $\forall y(P(y) \leftrightarrow y = x_0)$. The goal is a conjunction, so we treat it as two separate goals:

Givens	*Goal*
$\forall y(P(y) \leftrightarrow y = x_0)$	$\exists x P(x)$
	$\forall y \forall z((P(y) \wedge P(z)) \to y = z)$

To prove the first goal we choose $x = x_0$ and try to prove $P(x_0)$; plugging in x_0 for y in the given, we get $P(x_0) \leftrightarrow x_0 = x_0$. Since $x_0 = x_0$ is always true, we get $P(x_0)$ by the $\leftarrow$ direction of the bicondidtional. For the second goal, we let y and z be arbitrary, assume $P(y)$ and $P(z)$, and try to prove $y = z$:

Givens	*Goal*
$\forall y(P(y) \leftrightarrow y = x_0)$	$y = z$
$P(y)$	
$P(z)$	

Plugging in each of y and z into the first given we get $P(y) \leftrightarrow y = x_0$ and $P(z) \leftrightarrow z = x_0$. Since we've assumed $P(y)$ and $P(z)$ we use the $\to$ directions of these biconditionals to conclude that $y = x_0$ and $z = x_0$; our goal $y = z$ clearly follows.
(next page)

3→1:$\exists x P(x) \wedge \forall y \forall z((P(y) \wedge P(z)) \to y = z) \to \exists x(P(x) \wedge \forall y(P(y) \to y = x))$
We treat the given conjunction as two separate givens. The first is an existential statement so we let x_0 stand for some object such that $P(x_0)$ is true. The goal is an existential statement, we let $x = x_0$, leaving us with this situation:

Givens	*Goal*
$P(x_0)$	$P(x_0) \wedge \forall y(P(y) \to y = x_0)$
$\forall y \forall z((P(y) \wedge P(z)) \to y = z)$	

The first part of the goal is already proven. We only need to consider the second; we let y be arbitrary, assume $P(y)$ and make $y = x_0$ our goal:

Givens	*Goal*
$P(x_0)$	$y = x_0$
$\forall y \forall z((P(y) \wedge P(z)) \to y = z)$	
$P(y)$	

We know both $P(y)$ and $P(x_0)$. The goal follows from the second given.

Solution

Theorem. *The following are equivalent:*

1. $\exists x(P(x) \wedge \forall y(P(y) \to y = x))$
2. $\exists x \forall y(P(y) \leftrightarrow y = x)$
3. $\exists x P(x) \wedge \forall y \forall z((P(y) \wedge P(z)) \to y = z)$

Proof. $1 \to 2$. By statement 1, we can let x_0 be some object such that $P(x_0)$ and $\forall y(P(y) \to y = x_0)$ To prove statement 2 we will show that $\forall y(P(y) \leftrightarrow y = x_0)$. Let y be arbitrary. We already know the $\to$ direction of the biconditional. For the $\leftarrow$ direction, suppose $y = x_0$. Then since we know $P(x_0)$, we can conclude $P(y)$.

$2 \to 3$. By statement 2, choose x_0 such that $\forall y(P(y) \leftrightarrow y = x_0)$. Then, in particular, $P(x_0) \leftrightarrow x_0 = x_0$, and since clearly $x_0 = x_0$, it follows that $P(x_0)$ is true. Thus $\exists x P(x)$. To prove the second half of statement 3, let y and z be arbitrary and suppose $P(y)$ and $P(z)$. Then by our choice of x_0 (as something for which $\forall y(P(y) \leftrightarrow y = x_0)$ is true), it follows that $y = x_0$ and $z = x_0$, so $y = z$.

$3 \to 1$. By the first half of statement 3, let x_0 be some object such that $P(x_0)$. We will prove that $\forall y(P(y) \to y = x_0)$, so suppose $P(y)$. Since we have both $P(x_0)$ and $P(y)$, by the second half of statement 3 we can conclude that $y = x_0$, as required. □

2.6.1 Strategies and examples

As per the previous section, the following are equivalent:

1. $\exists x(P(x) \wedge \forall y(P(y) \rightarrow y = x))$
2. $\exists x \forall y(P(y) \leftrightarrow y = x)$
3. $\exists x P(x) \wedge \forall y \forall z((P(y) \wedge P(z)) \rightarrow y = z)$

Since all three of the statements in the theorem are equivalent to $\exists! x P(x)$, we can prove a goal of this form by proving any of the three statements in the theorem. The most common approach is using statement 3:

> To prove a goal of the form $\exists! x P(x)$:
> Prove $\exists x P(x)$ and $\forall y \forall z((P(y) \wedge P(z)) \rightarrow y = z)$. The first of these goals shows *existence*, the next *uniqueness*.

Form of final proof:

> Existence: [Proof of $\exists x P(x)$ goes here.]
> Uniqueness: [Proof of $\forall y \forall z((P(y) \wedge P(z)) \rightarrow y = z)$ goes here.]

Example
Prove that there is a unique set A such that for every set B, $A \cup B = B$.

Scratch work
Our goal is $\exists! A P(A)$, where $P(A)$ is the statement $\forall B(A \cup B = B)$. As per our strategy we prove existence and uniqueness separately.

For the existence half of the proof we must show $\exists A P(A)$, so we try to find a value of A that makes $P(A)$ true. There is no formula for finding this A, but one might eventually realise that the right choice is $A = \emptyset$. Plugging in this value we see that to complete the existence half of the proof we must show $\forall B(\emptyset \cup B = B)$, which is clearly true (this can be proven).

For the uniqueness half of the proof we prove $\forall C \forall D((P(C) \wedge P(D)) \rightarrow C = D)$. We let C and D be arbitrary, assume $P(C)$ and $P(D)$, and prove $C = D$:

Givens	*Goal*
$\forall B(C \cup B = B)$	$C = D$
$\forall B(D \cup B = B)$	

We plug in D for B in the first given, and C for B in the second. We get $C \cup D = D$ and $D \cup C = C$. But clearly $C \cup D = D \cup C$ (this can be proven). The goal $C = D$ follows immediately.
(next page)

Example cont.
Solution

Theorem. *There is a unique set A such that for every set B, $A \cup B = B$.*

Proof. Existence: Clearly $\forall B(\emptyset \cup B = B)$, so $\emptyset$ has the required property.

Uniqueness: Suppose $\forall B(C \cup B = B)$ and $\forall B(D \cup B = B)$. Applying the first of these assumptions to D we see that $C \cup D = D$, and applying the second to C we get $D \cup C = C$. But clearly $C \cup D = D \cup C$, so $C = D$. □

Alternative approach
The other equivalences can also be used. Statement 1 leads to the strategy:

> To prove a goal of the for $\exists! x P(x)$:
> Prove $\exists x(P(x) \wedge \forall y(P(y) \to y = x))$, using strategies from previous sections.

Example
Prove that for every real number x, if $x \neq 2$ then there is a unique real number y such that $2y/(y+1) = x$.

Scratch work
Our goal is $\forall x(x \neq 2 \to \exists! y(2y/(y_1) = x))$. We let x be arbitrary, assume $x \neq 2$, and prove $\exists! y(2y/(y+1) = x)$. As per our strategy, we can prove this goal by proving the equivalent statement

$$\exists y \left(\frac{2y}{y+1} = x \wedge \forall z \left(\frac{2z}{z+1} = x \to z = y \right) \right)$$

We start by trying to find some y that will make the equation $2y/(y+1) = x$ come out true. In other words, we solve this equation for y:

$$\frac{2y}{y+1} = x \implies 2y = x(y+1) \implies y(2-x) = x \implies y = \frac{x}{2-x}$$

Note that we have $x \neq 2$ so the last step makes sense. These steps will not appear in the proof; we simply let $y = x/(2-x)$ and try to prove $2/(y+1) = x$ and $\forall z(2z/(z+1) = x \to z = y)$.

Givens	*Goal*
$x \neq 2$	$\frac{2y}{y+1} = x$
$y = \frac{x}{2-x}$	$\forall z \left(\frac{2z}{z+1} = x \to z = y \right)$

The first goal is easy to verify by simply plugging in our candidate y.
(next page)

Example cont.
For the second goal, we let z be arbitrary, assume $2z(z+1) = x$, and prove $z = y$:

Givens	*Goal*
$x \neq 2$	$z = y$
$y = \frac{x}{2-x}$	
$\frac{2z}{z+1} = x$	

We can now show that $z = y$ by solving for z in the third given:

$$\frac{2z}{z+1} = x \implies 2z = x(z+1) \implies z(2-x) = x \implies z = \frac{x}{2-x} = y$$

Note that the steps we use here are exactly the same as the steps we used earlier in solving for y. This is a common pattern in existence and uniqueness proofs. Although the scratch work for figuring out an existence proof should not appear in the proof, this scratch work, or reasoning similar to it, can be sometimes used to demonstrate uniqueness.

Solution

Theorem. *For every real number x, if $x \neq 2$ then there is a unique real number y such that $2y/(y+1) = x$.*

Proof. Let x be an arbitrary real number, and suppose $x \neq 2$. Let $y = x/(2-x)$, which is defined since $x \neq 2$. Then

$$\frac{2y}{y+1} = \frac{\frac{2x}{2-x}}{\frac{x}{2-x}+1} = \frac{\frac{2x}{2-x}}{\frac{2}{2-x}} = \frac{2x}{2} = x$$

To see that this solution is unique, suppose $2z/(z+1) = x$. Then $2z = x(z+1)$, so $z(2-x) = x$. Since $x \neq 2$ we can divide both sides by $2-x$ to get $z = x/(2-x) = y$. □

2.6.2 Given of the form $\exists! x P(x)$

The three statements discussed earlier can also be used for givens of the form $\exists! x P(x)$. Once again, statement 3 of the theorem is the one used most often.

> To use a given of the form $\exists! x P(x)$:
> Treat this as two given statements $\exists x P(x)$ and $\forall y \forall z((P(y) \wedge P(z)) \rightarrow y = z)$. To use the first statement one should probably choose some x_0 to stand for some object where $P(x_0)$ is true. For the second, if one ever comes across two objects y and z such that $P(y)$ and $P(z)$ are both true, then one can conclude that $y = z$.

Example
Suppose A, B and C are sets, A and B are not disjoint, A and C are not disjoint, and A has exactly one element. Prove that B and C are not disjoint.

Scratch work

Givens	*Goal*
$A \cap B = \emptyset$	$B \cap C = \emptyset$
$A \cap C = \emptyset$	
$\exists! x(x \in A)$	

We treat the last given as two separate givens, as suggested by our strategy. We also analyse the logical forms of the other givens and the goal.

Givens	*Goal*
$\exists x(x \in A \wedge x \in B)$	$\exists x(x \in B \wedge x \in C)$
$\exists x(x \in A \wedge x \in C)$	
$\exists x(x \in A)$	
$\forall y \forall z((y \in A \wedge z \in A) \rightarrow y = z)$	

The first given tells use that we can choose a name, say b, for something such that $b \in A$ and $b \in B$. Similarly, by the second given we can let c be something such that $c \in A$ and $c \in C$. At this point the third given is redundant; we already know that there's something in A, so we may as well skip to the last given.

We know $b \in A$ and $c \in A$, so by the last given we can conclude that $b = c$. Since $b \in B$ and $b = c \in C$, we have found something that is an element of both B and C, as is required to prove the goal.
(next page)

Example cont.
Solution

Theorem. *Suppose A, B and C are sets, A and B are not disjoint, A and C are not disjoint, and A has exactly one element. Then B and C are not disjoint.*

Proof. Since A and B are not disjoint, we can let b be something such that $b \in A$ and $b \in B$. Similarly, since A and C are not disjoint, then there is some object c such that $c \in A$ and $c \in C$. Since A has only one element, we must have $b = c$. Thus $b = c \in B \cap C$ and therefore B and C are not disjoint. □

Chapter 3

Relations

3.1 Ordered pairs and Cartesian product

Justification
Suppose $P(x, y)$ is a statement with two free variables x and y. We can't speak of this statement as being true or false until we have specified two values—one for x and y. Thus if we want the truth set to identify which assignment of values make the statement come out true, then the truth set will have to contain not individual variables, but pairs of values. We will specify a pair of values by writing the two values in parantheses separated by a comma.

For example, let $D(x, y)$ mean "x divides y". Then $D(6, 18)$ is true, but $D(18, 6)$ is false. We must therefore distinguish between the pairs $(18, 6)$ and $(6, 18)$.

Since the order of values in the pair makes a difference, we will refer to a pair (a, b) as an *ordered pair*, with *first coordinate* a and *second coordinate* b. Note that ordered pairs can have anything at all as their coordinates, and are not just restricted to real numbers.
(next page)

Cartesian product
In general, if $P(x, y)$ is a statement in which x ranges over some set A and y ranges over a set B, then the only assignments of values to x and y that will make sense in $P(x, y)$ will be ordered pairs in which the first coordinate is an element of A and the second from B. We therefore make the following definition:

Definition. Suppose A and B are sets. Then the *Cartesian product* of A and B, denoted $A \times B$, is the set of all ordered pairs in which the first coordinate is an element of A and the second an element of B. In other words,

$$A \times B = \{(a, b) |\, a \in A \text{ and } b \in B\}$$

Examples
If $A = \{\text{red, green}\}$ and $B = \{2, 3, 5\}$ then

$$A \times B = \{(\text{red, } 2), (\text{red, } 3), (\text{red, } 5), (\text{green, } 2), (\text{green, } 3), (\text{green, } 5)\}$$

for another example:

$$\mathbb{R} \times \mathbb{R} = \{(x, y) |\, x \text{ and } y \text{ are real numbers}\}$$

(These are the coordinates of all the points in the plane. For obvious reasons, this set is sometimes written $\mathbb{R}^2$.)

3.1.1 Properties of the Cartesian product

Theorem. *Suppose A, B, C and D are sets.* $A \times (B \cap C) = (A \times B) \cap (A \times C)$

Proof. Let p be an arbitrary element of $A \times (B \cap C)$. Then by the definition of the Cartesian product, p must be an ordered pair whose first coordinate is an element of A and whose second coordinate is an element of $B \cap C$. In other words, $p = (x, y)$ for some $x \in A$ and $y \in B \cap C$. Since $y \in B \cap C$, $y \in B$ and $y \in C$. Since $x \in A$ and $y \in B$, $p = (x, y) \in A \times B$, and similarly $p \in A \times C$. Since p was an arbitrary element of $A \times (B \cap C)$, it follows that $A \times (B \cap C) \subseteq (A \times B) \cap (A \times C)$.

Now let p be an arbitrary element of $(A \times B) \cap (A \times C)$. Then $p \in A \times B$, so $p = (x, y)$ for some $x \in A$ and $y \in B$. Also $(x, y) = p \in A \times C$, so $y \in C$. Since $y \in B$ and $y \in C$, $y \in B \cap C$. Thus $p = (x, y) \in A \times (B \cap C)$. Since p was an arbitrary element of $(A \times B) \cap (A \times C)$ we can conclude that $(A \times B) \cap (A \times C) \subseteq A \times (B \cap C)$. So $A \times (B \cap C) = (A \times B) \cap (A \times C)$. □

In the first part we let p be an arbitrary element of $A \times (B \cap C)$ and then prove $p \in (A \times B) \cap (A \times C)$. Because $p \in A \times (B \cap C)$ means $\exists x \exists y (x \in A \wedge y \in B \cap C \wedge p = (x, y))$; and we immediately introduct the variables x and y by existential instantiation. The rest of the proof is straightforward.

In both parts of this proof we introduced an arbitrary object p that turned out to be an ordered pair, and we were therefore able to say that $p = (x, y)$ for some objects x and y. In most proofs involving Cartesian products mathematicians suppress this step. If it is clear from the beginning that an object will turn out to be an ordered pair, it is usually just called (x, y) from the outset. We will follow this practice in our proofs.
(next page)

Theorem. *Suppose A, B, C and D are sets.* $A \times (B \cup C) = (A \times B) \cup (A \times C)$.

Proof. Let (x, y) be an arbitrary element of $A \times (B \cup C)$. Then $x \in A$ and $y \in B \cup C$.
Case 1: $y \in B$. Then since $x \in A$, $(x, y) \in A \times B$ and $(x, y) \in (A \times B) \cup (A \times C)$.
Case 2: $y \in C$. Then similarly, $(x, y) \in A \times C$ and $(x, y) \in (A \times B) \cup (A \times C)$.
Since (x, y) is an arbitrary element of $A \times (B \cup C)$, $A \times (B \cup C) \subseteq (A \times B) \cup (A \times C)$.

Let (x, y) be an arbitrary element of $(A \times B) \cup (A \times C)$.
Case 1: $(x, y) \in (A \times B)$. Then $x \in A$, $y \in B$, so $y \in B \cup C$ and $(x, y) \in A \times (B \cup C)$.
Case 2: $(x, y) \in (A \times C)$. Then similarly, $x \in A$, $y \in C$, so $y \in B \cup C$ and $(x, y) \in A \times (B \cup C)$.
Since (x, y) is an arbitrary element of $(A \times B) \cup (A \times C)$, $(A \times B) \cup (A \times C) \subseteq A \times (B \cup C)$. So $A \times (B \cup C) = (A \times B) \cup (A \times C)$. □

Theorem. *Suppose A, B, C and D are sets.* $(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)$.

Proof. Let (x, y) be an arbitrary element of $(A \times B) \cap (C \times D)$. Then $(x, y) \in (A \times B)$ and $(x, y) \in (C \times D)$. So $x \in A, y \in B, x \in C, y \in D$; as such $x \in A \cap C$ and $y \in B \cap D$, and $(x, y) \in (A \cap C) \times (B \cap D)$. Since (x, y) is an arbitrary element of $(A \times B) \cap (C \times D)$, $(A \times B) \cap (C \times D) \subseteq (A \cap C) \times (B \cap D)$.

Let (x, y) be an arbitrary element of $(A \cap C) \times (B \cap D)$. Then $x \in (A \cap C), y \in (B \cap D)$, so $x \in A, x \in C, y \in B, y \in D$. Therefore $(x, y) \in A \times B$, and $(x, y) \in C \times D$, and $(x, y) \in (A \times B) \cap (C \times D)$. Since (x, y) is an arbitrary element of $(A \cap C) \times (B \cap D)$, $(A \cap C) \times (B \cap D) \subseteq (A \times B) \cap (C \times D)$. So $(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)$. □

Theorem. *Suppose A, B, C and D are sets.* $(A \times B) \cup (C \times D) \subseteq (A \cup C) \times (B \cup D)$.

Proof. Let (x, y) be an arbitrary element of $(A \times B) \cup (C \times D)$.
Case 1: $(x, y) \in A \times B$. Then $x \in A$ and $y \in B$, so clearly $x \in A \cup C$ and $y \in B \cup D$. Therefore $(x, y) \in (A \cup C) \times (B \cup D)$.
Case 2: $(x, y) \in C \times D$. Then $x \in C$ and $y \in D$. so similarly $(x, y) \in (A \cup C) \times (B \cup D)$.
Since (x, y) is an arbitrary element of $(A \times B) \cup (C \times D)$, $(A \times B) \cup (C \times D) \subseteq (A \cup C) \times (B \cup D)$. □

(next page)

Theorem. *Suppose A is a set. $A \times \emptyset = \emptyset \times A = \emptyset$.*

Proof. Suppose $A \times \emptyset \neq \emptyset$. Then $A \times \emptyset$ has at least one element, and by the definition of the cartesian product this element must be an ordered pair (x, y) for some $x \in A$ and $y \in \emptyset$. But this is impossible since $\emptyset$ has no elements. Thus $A \times \emptyset = \emptyset$.

Suppose $\emptyset \times A \neq \emptyset$. Then similarly there must be an ordered pair for some $x \in \emptyset$ and $y \in A$. But this is impossible since $\emptyset$ has no elements. Thus $\emptyset \times A = \emptyset$. So we have $A \times \emptyset = \emptyset \times A = \emptyset$. □

The statement we are trying to prove can be interpreted as $A \times \emptyset = \emptyset \wedge \emptyset \times A = \emptyset$. We treat this as two goals and prove $A \times \emptyset = \emptyset$ and $\emptyset \times A = \emptyset$ separately. See that to say that a set equals the empty set is actually a negative statement, since it means that the set does not have any elements. Thus the proof naturally proceeds by contradiction. The assumption that $A \times \emptyset \neq \emptyset$ means $\exists p(p \in A \times \emptyset)$ so we existantially instantiate the ordered pair, leading to a contradiction. The proof that $\emptyset \times A = \emptyset$ is similar.

To summarise

Theorem. *Suppose A, B, C, and D are sets.*

1. $A \times (B \cap C) = (A \times B) \cap (A \times C)$.
2. $A \times (B \cup C) = (A \times B) \cup (A \times C)$.
3. $(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)$.
4. $(A \times B) \cup (C \times D) \subseteq (A \cup C) \times (B \cup D)$.
5. $A \times \emptyset = \emptyset \times A = \emptyset$.

(next page)

Theorem. *Suppose A and B are sets. Then $A \times B = B \times A$ iff either $A = \emptyset, B = \emptyset$, or $A = B$.*

Proof. ($\rightarrow$) Suppose $A \times B = B \times A$. If either $A = \emptyset$ or $B = \emptyset$, then there is nothing more to prove, so suppose $A \neq \emptyset$ and $B \neq \emptyset$; we will show that $A = B$. Let x be arbitrary, and suppose $x \in A$. Since $B \neq \emptyset$ we can choose some $y \in B$. Then $(x, y) \in A \times B = B \times A$, so $x \in B$.

Now suppose $x \in B$. Since $A \neq \emptyset$ we can choose some $z \in A$. Therefore $(x, z) \in B \times A = A \times B$, so $x \in A$. Thus $A = B$ as required.

($\leftarrow$) Suppose either $A = \emptyset, B = \emptyset$, or $A = B$.
Case 1: $A = \emptyset$. Then $A \times B = \emptyset \times B = \emptyset = B \times \emptyset = B \times A$.
Case 2: $B = \emptyset$. Similar to case 1.
Case 3: $A = B$. Then $A \times B = A \times A = B \times A$. □

The statement to be proven here is an iff statement so we prove both directions separately. For the $\rightarrow$ direction, our goal is $A = \emptyset \vee B = \emptyset \vee A = B$, which can also be written as $(A = \emptyset \vee B = \emptyset) \vee A = B$. We proceed using proof by cases, assuming $\neg(A = \emptyset \vee B = \emptyset)$ and proving $A = B$. (By de morgan's laws $\neg(A = \emptyset \vee B = \emptyset)$ is equivalent to $A \neq \emptyset \wedge B \neq \emptyset$). Of course we could have proceeded differently, say by assuming $A \neq B$ and $B \neq \emptyset$ and then proving $A = \emptyset$; but since $A = \emptyset$ and $B = \emptyset$ are actually negative statements, we're better off negating both of them, and then proving the positive statement $A = B$. The assumptions $A \neq \emptyset$ and $B \neq \emptyset$ are existential statements, so they are used in the proof to justify y and z. The proof $A = B$ then proceeds as per usual. For the other direction we use naturally use proof by cases.

3.1.2 Truth sets

Definition. Suppose $P(x, y)$ is a statement with two free variables in which x ranges over a set A and y ranges over another set B. Then $A \times B$ is the set of all assignments of values to x and y that make sense in the statement $P(x, y)$. The *truth set* of $P(x, y)$ is the subset of $A \times B$ consisting of those assignments that make the statement come out true. In other words, the truth set of $P(x, y)$ is the set $\{(a, b) \in A \times B | P(a, b)\}$.

Example 1
Consider the statement 'x is located in y,' where x ranges over the set C of all cities and y ranges over the set N of all countries.

Its truth set is $\{(c, n) \in C \times N | \text{the city } c \text{ is located in the country } n\} = \{$(New York, United States), (Tokyo, Japan), (Paris, France), $\ldots\}$.

Example 2
Consider the statement '$y = 2x - 3$' where x and y range over $\mathbb{R}$.

Its truth set is $\{(x, y) \in \mathbb{R} \times \mathbb{R} | y = 2x - 3\} = \{(0, -3), (1, -1), (2, 1), \ldots\}$. This is just points along the line specified in the statement. One can think of the graph of an equation as a picture of its truth set.

Properties
Many facts about truth sets for statements with one free variable carry over to truth sets for statements with two free variables. For example, suppose T is the truth set of a statement $P(x, y)$, where x ranges over some set A and y ranges over B. Then for any $a \in A$ and $b \in B$ the statement $(a, b) \in T$ means the same thing as $P(a, b)$.

If $P(x, y)$ is true for every $x \in A$ and $y \in B$, then $T = A \times B$; if $P(x, y)$ is false for every $x \in A$ and $y \in B$, then $T = \emptyset$. If S is the truth set of another statement $Q(x, y)$, then the truth set of the statement $P(x, y) \wedge Q(x, y)$ is $T \cap S$, and the truth set of $P(x, y) \vee Q(x, y)$ is $T \cup S$.

More coordinates
Although we may concentrate on ordered pairs, it is possible to work with ordered triples, ordered quadruples, and so on. These might be used to talk about truth sets for statements containing three or more free variables.

3.2 Relations

Suppose $P(x, y)$ is a statement with two free variables x and y. Often such a statement can be thought of as expressing a *relationship* between x and y. The truth set of the statement $P(x, y)$ is a set of ordered pairs that records when this relationship holds. It is often useful to think of any set of ordered pairs in this way, as a record of when some relationship holds. This is the motivation behind the following definition.

Definition. Suppose A and B are sets. Then a set $R \subseteq A \times B$ is called a *relation from A to B*.

If x ranges over A and y ranges over B, then clearly the truth set of any statement $P(x, y)$ will be a relation from A to B. Note however that the definition does not require that the set of ordered pairs be defined as the truth set of some statement in order for that set to be a relation; although truth sets motivated the definition, *any* subset of $A \times B$ is to be called a relation from A to B.

Examples

1. Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, and $R = \{(1, 3), (1, 5), (3, 3)\}$. Then $R \subseteq A \times B$, so R is a relation from A to B.
2. Let $G = \{(x, y) \in \mathbb{R} \times \mathbb{R} | x > y\}$. Then G is a relation from $\mathbb{R}$ to $\mathbb{R}$.
3. Let $A = \{1, 2\}$ and $B = \mathscr{P}(A) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$. Let $E = \{(x, y) \in A \times B | x \in y\}$. Then E is a relation from A to B. In this case $E = \{(1, \{1\}), (1, \{1, 2\}), (2, \{2\}), (2, \{1, 2\})\}$.

3.2.1 Domain, Range, Inverse, Composition

Definition. Suppose R is a relation from A to B. Then the *domain* of R is the set

$$\text{Dom}(R) = \{a \in A | \exists b \in B((a, b) \in R)\}$$

Intuition: The domain of a relation from A to B is the set containing all the first coordinates of ordered pairs in the relation. This will in general be a subset of A, but it need not be all of A.

Definition. The *range* of R is the set

$$\text{Ran}(R) = \{b \in B | \exists a \in A((a, b) \in R)\}$$

Intuition: In a manner similar to the domain, the range of a relation is the set containing all the second coordinates of its ordered pairs, this set need not be all of B.

Definition. The *inverse* of R is the relation R^{-1} from B to A defined as follows:

$$R^{-1} = \{(b, a) \in B \times A | (a, b) \in R\}$$

Intuition: The inverse of a relation contains exactly the same ordered pairs as the original relation, but with the order of the coordinates of each pair reversed.

Definition. Suppose R is a relation from A to B and S is a relation from B to C. Then the *composition* of S and R is the relation $S \circ R$ from A to C defined as

$$S \circ R = \{(a, c) \in A \times C | \exists b \in B((a, b) \in R \text{ and } (b, c) \in S)\}$$

Intuition: $(a, c) \in S \circ R$ iff there is some b such that $(a, b) \in R$ and $(b, c) \in S$. This notation may seem backward at first—if $(a, b) \in R$ and $(b, c) \in S$, one might be tempted to write $(a, c) \in R \circ S$—but according to our definition, the proper notation is $(a, c) \in S \circ R$.
(next page)

Properties

Theorem. *Suppose R is a relation from A to B. Then $(R^{-1})^{-1} = R$.*

Proof. First of all, note that R^{-1} is a relation from B to A, so $(R^{-1})^{-1}$ is a relation from A to B, just like R. To see that $(R^{-1})^{-1} = R$, let (a, b) be an arbitrary ordered pair in $A \times B$. Then

$$(a, b) \in (R^{-1})^{-1} \text{ iff } (b, a) \in R^{-1} \text{ iff } (a, b) \in R \qquad \square$$

The statement we are trying to prove means $\forall p(p \in (R^{-1})^{-1} \leftrightarrow p \in R)$. We introduce an arbitrary ordered pair and prove the equivalence.

Theorem. *Suppose R is a relation from A to B. Then $Dom(R^{-1}) = Ran(R)$.*

Proof. First note that $\text{Dom}(R^{-1})$ and $\text{Ran}(R)$ are both subsets of B. Now let b be an arbitrary element of B. Then

$$\begin{aligned} b \in \text{Dom}(R^{-1}) &\text{ iff } \exists a \in A((b, a) \in R^{-1}) \\ &\text{iff } \exists a \in A((a, b) \in R) \text{ iff } b \in \text{Ran}(R) \end{aligned} \qquad \square$$

Theorem. *Suppose R is a relation from A to B. Then $Ran(R^{-1}) = Dom(R)$.*

Proof. First note that $\text{Ran}(R^{-1})$ and $\text{Dom}(R)$ are both subsets of A. Now let a be an arbitrary element of A. Then

$$\begin{aligned} a \in \text{Ran}(R^{-1}) &\text{ iff } \exists b((b, a) \in R^{-1}) \\ &\text{iff } \exists b((a, b) \in R) \text{ iff } a \in \text{Dom}(R) \end{aligned} \qquad \square$$

Theorem. *Suppose R is a relation from A to B, S is a relation from B to C, and T is a relation from C to D. Then $T \circ (S \circ R) = (T \circ S) \circ R$.*

Proof. Clearly $T \circ (S \circ R)$ and $(T \circ S) \circ R$ are both relations from A to D. Let (a, d) be an arbitrary element of $A \times D$.

First, suppose $(a, d) \in T \circ (S \circ R)$. By the definition of composition, this means that we can choose some $c \in C$ such that $(a, c) \in S \circ R$ and $(c, d) \in T$. Since $(a, c) \in S \circ R$, we can again use the definition of composition and choose some $b \in B$ such that $(a, b) \in R$ and $(b, c) \in S$. Now since $(b, c) \in S$ and $(c, d) \in T$, we can conclude that $(b, d) \in T \circ S$. Similarly, since $(a, b) \in R$ and $(b, d) \in T \circ S$ it follows that $(a, d) \in (T \circ S) \circ R$.

Now suppose $(a, d) \in (T \circ S) \circ R$. We can choose some $b \in B$ such that $(a, b) \in R$ and $(b, d) \in T \circ S$. Again we can declare some $c \in C$ such that $(b, c) \in S$, $(c, d) \in T$. Now since $(b, c) \in S$ and $(a, b) \in R$, $(a, c) \in S \circ R$. Since $(c, d) \in T$, $(a, d) \in T \circ (S \circ R)$. $\square$

(next page)

Theorem. *Suppose R is a relation from A to B, and S is a relation from B to C. Then $(S \circ R)^{-1} = R^{-1} \circ S^{-1}$.*

Proof. $(S \circ R)^{-1}$ and $R^{-1} \circ S^{-1}$ are both relations from C to A. Let (c, a) be an arbitrary element of $C \times A$.

First suppose $(c, a) \in (S \circ R)^{-1}$. Then $(a, c) \in S \circ R$ and we can choose some b such that $(a, b) \in R$ and $(b, c) \in S$. We can then say that $(b, a) \in R^{-1}$ and $(c, b) \in S^{-1}$, and $(c, a) \in R^{-1} \circ S^{-1}$.

Now suppose $(c, a) \in R^{-1} \circ S^{-1}$. Then we can choose some b such that $(c, b) \in S^{-1}$ and $(b, a) \in R^{-1}$, and so $(b, c) \in S$ and $(a, b) \in R$. We can then say that $(a, c) \in S \circ R$ and $(c, a) \in (S \circ R)^{-1}$. Thus $(S \circ R)^{-1} = R^{-1} \circ S^{-1}$. □

To summarise:

Theorem. *Suppose R is a relation from A to B, S is a relation from B to C, and T is a relation from C to D. Then:*

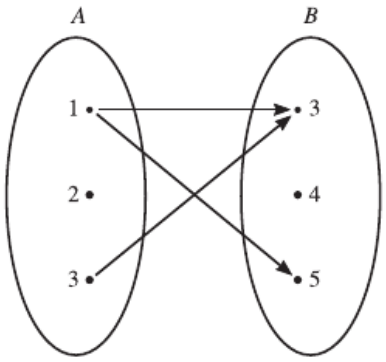
1. $(R^{-1})^{-1} = R$
2. $Dom(R^{-1}) = Ran(R)$
3. $Ran(R^{-1}) = Dom(R)$
4. $T \circ (S \circ R) = (T \circ S) \circ R$
5. $(S \circ R)^{-1} = R^{-1} \circ S^{-1}$

3.2.2 Alternate notation and visual representation

Alternate notation
If R is a relation from A to B, $x \in A$, and $y \in B$, the notation xRy is sometimes written to mean $(x, y) \in R$.

For instance, the definition of composition of relations, rewritten in this notation, looks like $S \circ R = \{(a, c) \in A \times C | \exists b \in B(aRb \wedge bSc)\}$, where R is a relation from A to B and S is a relation from B to C.

Visual representation
Another way to think about relations is to draw pictures of them. Consider a visual representation of the relation $R = \{(1, 3), (1, 5), (3, 3)\}$ from the set $A = \{1, 2, 3\}$ to the set $B = \{3, 4, 5\}$:



Each of these sets is represented by an oval, with the elements of the set represented by dots inside the oval. Each ordered pair $(a, b) \in R$ is represented by an arrow from the dot representing a to the dot representing b.

In general, any relation R from set A to set B can be represented by such a picture. The dots representing the elements of A and B are called *vertices*, and the arrows representing the ordered pairs in R are called *edges*. It doesn't matter exactly how the vertices representing elements of A and B are arranged on the page; what's important is that the edges correspond precisely to the ordered pairs in R.

One should be able to convince oneself that the domain of R can be found by locating the vertices in A that have edges pointing away from them. Similarly the range of R would consist of those elements of B whose vertices have edges pointing toward them. A picture of R^{-1} would look just like a picture of R but with the directions of all the arrows reversed.

3.2.3 Binary relations

A relation R from A to A is called a *relation on* A (or a *binary relation on* A). Relations of this type come up often, for example:

1. Let $A = \{1, 2\}$ and $B = \mathscr{P}(A) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$. If we had

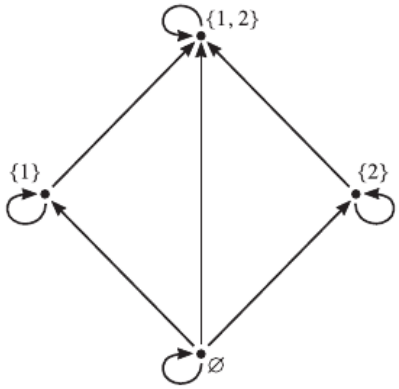
$$\begin{aligned} S &= \{(x, y) \in B \times B | x \subseteq y\} \\ &= \{(\emptyset, \emptyset), (\emptyset, \{1\}), (\emptyset, \{2\}), (\emptyset, \{1, 2\}), (\{1\}, \{1\}), (\{1\}, \{1, 2\}), \\ &\quad (\{2\}, \{2\}), (\{2\}, \{1, 2\}), (\{1, 2\}, \{1, 2\})\} \end{aligned}$$

   Then S is a relation on B.

2. Suppose A is a set. Let $i_A = \{(x, y) \in A \times A | x = y\}$. Then i_A is a relation on A (called the *identity relation* on A.) For instance if $A = \{1, 2, 3\}$, then $i_A = \{(1, 1), (2, 2), (3, 3)\}$. Note that i_A could also be defined by writing $i_A = \{(x, x) | x \in A\}$.

3. For each positive real number r, let $D_r = \{(x, y) \in \mathbb{R} \times \mathbb{R} | x$ and y differ by less than r, or in other words $|x - y| < r\}$. Then D_r is a relation on $\mathbb{R}$.

Suppose R is a relation on set A. If we used the method described earlier to draw a picture of R, we would be drawing two copies of A with edges representing each ordered pair in R. An easier way to draw the picture would be to draw just one copy of A and then connect the vertices represening the elements of A with edges to represent the ordered pairs in R; for instance the first example looks like



Pictures like this one are called *directed graphs*. Note that in this directed graph there is an edge from $\emptyset$ to itself, because $(\emptyset, \emptyset) \in S$. Edges such as this one that go from a vertex to itself are called *loops*. In fact in this case there is a loop at every vertex—S has the property that $\forall x \in B((x, x) \in S)$. We describe this situation by saying that S is *reflexive*.
(next page)

Definition. Suppose R is a relation on A.

1. R is said to be *reflexive* on A (or just *reflexive*, if A is clear from context) if $\forall x \in A(xRx)$, or in other words $\forall x \in A((x,x) \in R)$.

2. R is *symmetric* if $\forall x \in A \forall y \in A(xRy \to yRx)$.

3. R is *transitive* if $\forall x \in A \forall y \in A \forall z \in A((xRy \wedge yRz) \to xRz)$.

Intuition
If R were reflexive on A, then the directed graph representing R would have loops at all vertices. If R were symmetric, then whenever there is an edge from x to y, there will also be an edge from y to x. If R were transitive, then whenever there was an edge from x to y and y to z, there would also be an edge from x to z.

Properties

Theorem. *Suppose R is a relation on set A. R is reflexive iff $i_A \subseteq R$, where i_A is the identity relation on A.*

Proof. ($\to$) Suppose R is reflexive, let (a,b) be an arbitrary element of i_A. Then by the definition of i_A, $(a,b) \in \{(x,y) \in A \times A | x = y\}$, so $a = b \in A$. Since R is reflexive, meaning $\forall x \in A((x,x) \in R)$, $(a,b) = (a,a) \in R$. Since (a,b) is arbitrary, $i_A \subseteq R$.

($\leftarrow$) Suppose $i_A \subseteq R$. Let x be an arbitrary element of A. Then $(x,x) \in i_A$, and since $i_A \subseteq R$, $(x,x) \in R$. Since x was an arbitrary element of A, $\forall x \in A((x,x) \in R)$, meaning R is reflexive. $\square$

Theorem. *Suppose R is a relation on set A. R is symmetric iff $R = R^{-1}$.*

Proof. ($\to$) Suppose R is symmetric. Let (x,y) be an arbitrary element of R. Then xRy, so since R is symmetric, yRx. Thus, $(y,x) \in R$, so by the definition of R^{-1}, $(x,y) \in R^{-1}$. Since (x,y) was arbitrary, it follows that $R \subseteq R^{-1}$.

Now suppose $(x,y) \in R^{-1}$. Then $(y,x) \in R$, so since R is symmetric, $(x,y) \in R$. Thus $R^{-1} \subseteq R$, so $R = R^{-1}$.

($\leftarrow$) Suppose $R = R^{-1}$, and let x and y be arbitrary elements of A. Suppose xRy. Then $(x,y) \in R$, so since $R = R^{-1}$, $(x,y) \in R^{-1}$. By the definition of R^{-1} this means $(y,x) \in R$, so yRx. Thus $\forall x \in A \forall y \in A(xRy \to yRx)$, so R is symmetric. $\square$

(next page)

Theorem. *Suppose R is a relation on set A. R is transitive iff $R \circ R \subseteq R$.*

Proof. ($\rightarrow$) Suppose R is transitive. Now let (a, b) be an arbitrary element of $R \circ R$. By the definition of the composition, we can declare some $c \in A$ such that $(a, c) \in R$ and $(c, b) \in R$. Since R is transitive, $(a, b) \in R$. Since (a, b) was arbitrary, $R \circ R \subseteq R$.

($\leftarrow$) Suppose $R \circ R \subseteq R$. Let x, y, z be arbitrary elements of A. Suppose $(xRy \wedge yRz)$, then by the definition of the composition $(x, z) \in R$. Since x, y, z were arbitrary, $\forall x \in A \forall y \in A \forall z \in A((xRy \wedge yRz) \rightarrow xRz)$, meaning R is transitive. □

To summarise

Theorem. *Suppose R is a relation on set A.*

1. *R is reflexive iff $i_A \subseteq R$, where i_A is the identity relation on A*
2. *R is symmetric iff $R = R^{-1}$*
3. *R is transitive iff $R \circ R \subseteq R$*

3.3 Ordering relations

Definition. Suppose R is a relation on set A. Then R is said to be *antisymmetric* if $\forall x \in A \forall y \in A((xRy \wedge yRx) \rightarrow x = y)$.

Intuition and examples
For a first example, consider the relation $L = \{(x, y) \in \mathbb{R} \times \mathbb{R} | x \leq y\}$. This relation is clearly reflexive and transitive, but not symmetric. Notice further that it fails to be symmetric in a rather extreme way because there are many pairs (x, y) such that xLy is true but yLx is false. In fact, the only way xLy and yLx can both be true is if $x \leq y$ and $y \leq x$, and thus $x = y$. We therefore say that L is *antisymmetric*.

For another example, let $A = \{1, 2\}, B = \mathscr{P}(A)$, and $S = \{(x, y) \in B \times B | x \subseteq y\}$. If x and y are elements of B, then xSy means $x \subseteq y$. S is reflexive and transitive, but not symmetric. In fact, S is also antisymmetric, because for any sets x and y, if $x \subseteq y$ and $y \subseteq x$ then $x = y$.

See that both of these relations specifies an *order* x and y come in (in the sense that if (x, y) is present in the relation, then (y, x) will not be present unless $y = x$). This motivates the next definition.

Definition. Suppose R is a relation on set A. Then R is called a *partial order on* A (or just a *partial order* if A is clear from context) if it is reflexive, transitive, and antisymmetric. It is called a *total order on* A (or just a *total order*) if it is a partial order, and in addition has the following property:

$$\forall x \in A \forall y \in A(xRy \vee yRx)$$

Intuition
Both L and S (the two examples above) are partial orders. S is not a total order, because it is not true that $\forall x \in B \forall y \in B(x \subseteq y \vee y \subseteq x)$; although we can think of the relation S as specifying an order between elements, it does not give us a way of comparing *every* pair of elements of B—for some pairs (such as $\{1\}$ and $\{2\}$), S doesn't pick out either one as being 'at least as large' as the other—this is the sense in which the ordering is *partial*.

On the other hand, L is a total order, because if x and y are any two real numbers, then either $x \leq y$ or $y \leq x$. L gives us a way of comparing *any* two real numbers.

Example
Consider the relation $G = \{(x, y) \in \mathbb{R} \times \mathbb{R} | x \geq y\}$. One should be able to check that G is a total order. Notice that in this case the order goes in the other way—the definition of partial order, though motivated by thinking about orderings that go in one direction, actually applies to orderings in either direction.

3.3.1 Smallest and minimal element

Definition. Suppose R is a partial order on a set A, $B \subseteq A$, and $b \in B$. Then b is called an *R-smallest* element of B (or just a *smallest* element if R is clear from context) if $\forall x \in B(bRx)$. It is called the *R-minimal* element (or just a *minimal* element) if $\neg\exists x \in B(xRb \wedge x \neq b)$.

Example
Let A be the set of words in english, and let $R = \{(x, y) \in A \times A|$ all the letters in the word x appear, consecutively, and in the right order, in the word $y\}$. For example, (can, cannot), (tar, start), and (ball, ball) are all elements of R, but (can, anchor) and (can, carnival) are not.

One can check that R is reflexive, transitive, and antisymmetric, so R is a partial order. Now consider the set $B = \{$me, men, tame, mental$\} \subseteq A$. If we think of xRy as meaning that y is in some sense at least as large as x, then we could say that the word *me* is the *smallest* element of B, in the sense that a relation exists between all other elements and it, and in each relation it is deemed to be 'smaller'.

Now consider the set $C = \{$a, me, men, tame, mental$\} \subseteq A$. Each of the words *men, tame,* and *mental* is 'larger' than at least one word in the set, but neither *a* nor *me* is larger than anything else in the set; they are *minimal* elements of C—neither can be determined to be the *smallest* element of C since the relation doesn't contain an ordered pair with both *me* and *a*, but they are 'smaller' than everything else that they can be compared to.

Another example
Let $S = \{(X, Y) \in \mathscr{P}(\mathbb{N}) \times \mathscr{P}(\mathbb{N})|X \subseteq Y\}$, which is a partial order on the set $\mathscr{P}(\mathbb{N})$. Let $\mathcal{F} = \{X \in \mathscr{P}(\mathbb{N})|2 \in X$ and $3 \in X\}$. Does $\mathcal{F}$ have any S-smallest or S-minimal elements?

The set $\{2, 3\}$ is a smallest element of $\mathcal{F}$, since 2 and 3 are elements of every set in $\mathcal{F}$, and therefore $\forall X \in \mathcal{F}(\{2, 3\} \subseteq X)$. It is also a minimal element, since no other element of $\mathcal{F}$ is a subset of it.

Now consider the set $\mathcal{G} = \{X \in \mathscr{P}(\mathbb{N})|2 \in X$ or $3 \in X\}$.

The set $\mathcal{G}$ has two minimal elements, $\{2\}$ and $\{3\}$. Every other set in $\mathcal{G}$ must contain one of these two as a subset. Neither set is smallest, since neither is a subset of the other.
(next page)

Properties

Theorem. *If B has a smallest element, then this smallest element is unique. Thus we can speak of* ***the*** *smallest element of B rather than* ***a*** *smallest element.*

Proof. Suppose b is a smallest element of B, and suppose c is also a smallest element of B. Since b is a smallest element, $\forall x \in B(bRx)$, so in particular bRc. Similarly, since c is a smallest element, cRb. But now since R is a partial order, it must be antisymmetric, so from bRc and cRb we can conclude $b = c$. □

We want to show (B has a smallest element)→(this smallest element is unique). This translates to $\exists b \in B(\forall x \in B(bRx)) \rightarrow \forall c \in B(\forall x \in B(cRx) \rightarrow b = c)$. The proof proceeds as expected, then we use antisymmetry to complete it.

Theorem. *Suppose b is the smallest element of B. Then b is also a minimal element of B, and it is the only minimal element.*

Proof. Let x be an arbitrary element of B and suppose that xRb. Since b is the smallest element of B, so we must have bRx, and now by antisymmetry it follows that $x = b$. Thus there can be no $x \in B$ such that xRb and $x \neq b$, so b is a minimal element.

To see that it is the only one, suppose c is also a minimal element. Since b is the smallest element of B, bRc. But then since c is minimal we must have $b = c$. Thus b is the only minimal element of B. □

Given that b is the smallest element of B, we want to show (b is a minimal element of B)∧(it is the only minimal element). This can be rewritten as

$$\neg\exists x \in B(xRb \wedge x \neq b) \wedge \forall c \in B(\neg\exists x \in B(xRc \wedge x \neq c) \rightarrow c = b)$$

The key here is to simplify $\neg\exists x \in B(xRb \wedge x \neq b)$:

$$\begin{aligned}\neg\exists x \in B(xRb \wedge x \neq b) &\leftrightarrow \forall x \in B\neg(xRb \wedge x \neq b)\\ &\leftrightarrow \forall x \in B(xRb \rightarrow x = b)\end{aligned}$$

and so our goal now looks like

$$\forall x \in B(xRb \rightarrow x = b) \wedge \forall c \in B(\forall x \in B(xRc \rightarrow x = c) \rightarrow c = b)$$

Both goals are now fairly straightforward, the first goal can be completed using antisymmetry.
(next page)

Theorem. *Suppose $B \subseteq A$. If R is a total order on A and b is a minimal element of B, then b is the smallest element of B.*

Proof. Suppose R is a total order and b is a minimal element of B. Let x be an arbitrary element of B. If $x = b$, then since R is reflexive, bRx. Now suppose $x \neq b$; since R is a total order, we know that either xRb or bRx. However since b is minimal, xRb cannot be true. Thus bRx must be true. Since x was arbitrary, we can conclude that $\forall x \in B(bRx)$, so b is the smallest element of B. □

The proof in the book wants to use a disjunctive syllogism involving the total order property, since we have $xRb \vee bRx$, to show bRx (since $xRb \vee bRx \leftrightarrow \neg xRb \rightarrow bRx$). They do this by using proof by cases $x = b$ and $x \neq b$. The first case is straightforward using reflexivity, and the second uses the disjunctive syllogism, using the minimal property to show $\neg xRb$ and combining that with the total order property to show bRx.

(I felt that it might be more straightforward to immediately use the total order property to show $xRb \vee bRx$, then go from there using proof by cases.)

To summarise

Theorem. *Suppose R is a partial order on set A, and $B \subseteq A$.*

1. *If B has a smallest element, then this smallest element is unique. Thus we can speak of **the** smallest element of B rather than **a** smallest element.*
2. *Suppose b is the smallest element of B. Then b is also a minimal element of B, and it is the only minimal element.*
3. *If R is total order and b is a minimal element of B, then b is the smallest element of B.*

(next page)

Application
When comparing subsets of some set A, mathematicians often use the partial order $S = \{(X, Y) \in \mathscr{P}(A) \times \mathscr{P}(A) | X \subseteq Y\}$, although this is not always made explicit.

In this case, if $\mathcal{F} \subseteq \mathscr{P}(A)$ and $X \in \mathcal{F}$, then X is the S-smallest element of F iff $\forall Y \in \mathcal{F}(X \subseteq Y)$—to say that an element of $\mathcal{F}$ is the smallest element means that it is a subset of every element of $\mathcal{F}$. Similarly, mathematicians sometimes talk of a set being the smallest one with a certain property. Generally this means that the set has the property in question, ant furthermore that it is a subset of every set that has the property.

Example
Say we want to find the smallest set of real numbers X such that $5 \in X$ and for all real numbers x and y, if $x \in X$ and $x < y$ then $y \in X$.

Another way to phrase this question would be to say that we are looking for the smallest element of the family of sets $\mathcal{F} = \{X \subseteq \mathbb{R} | 5 \in X$ and $\forall x \forall y((x \in X \wedge x < y) \rightarrow y \in X)\}$. In particular since $5 \in X$ we can say that $\forall y(5 < y \rightarrow y \in X)$. See that $A = \{y \in R | 5 \leq y\}$ is the smallest element of $\mathcal{F}$.

Maximal and largest elements
Following a similar principle, suppose R is a partial order on A, $B \subseteq A$, and $b \in B$. We say that b is the *largest* element of B if $\forall x \in B(xRb)$, and it is a *maximal* element of B if $\neg \exists x \in B(bRx \wedge b \neq x)$.

3.3.2 Upper and lower bounds

Definition. Suppose R is a partial order on A, $B \subseteq A$, and $a \in A$. Then a is called a *lower bound* for B if $\forall x \in B(aRx)$. Similarly, it is an *upper bound* for B if $\forall x \in B(xRa)$.

Note that a lower bound for B need not be an element of B. This is the only difference between lower bounds and smallest elements. A smallest element of B is just a lower bound that is also an element of B.

Definition. Suppose R is a partial order on A and $B \subseteq A$. Let U be the set of all upper bounds for B, and let L be the set of all lower bounds. If U has a smallest element, then this smallest element is called the *least upper bound* of B. If L has a largest element, then this largest element is called the *greatest lower bound* of B. The phrases *least upper bound* and *greatest lower bound* are sometimes abbreviated as *l.u.b* and *g.l.b.*

Theorem. *Suppose A is a set, $\mathcal{F} \subseteq \mathscr{P}(A)$, and $\mathcal{F} \neq \emptyset$. Then the least upper bound of $\mathcal{F}$ (in the subset partial order) is $\bigcup \mathcal{F}$.*

Proof. (my proof) First we show that $\bigcup \mathcal{F}$ is an upper bound. Let x be an arbitrary of $\mathcal{F}$. Let y be an arbitrary element of x. Since $y \in x \in \mathcal{F}$, then $y \in \bigcup F$. Since y was an arbitrary element of x, then $x \subseteq \bigcup \mathcal{F}$. Since x was arbitrary, $\forall x \in \mathcal{F}(x \subseteq \bigcup \mathcal{F})$, and $\bigcup \mathcal{F}$ is an upper bound.

Now we show that $\bigcup \mathcal{F}$ is the least upper bound. Let x be an arbitrary element of $\mathscr{P}(A)$; suppose x is in the set of upper bounds. Suppose y is an arbitrary element of $\bigcup \mathcal{F}$; then we can declare some B such that $B \in \mathcal{F}$ and $y \in B$. Since x is in the set of upper bounds, $B \subseteq x$, and so $y \in x$. Since y was arbitrary, $\bigcup \mathcal{F} \subseteq x$. Since x was arbitrary, $\bigcup \mathcal{F}$ is the least upper bound. □

The idea here was to prove ($\bigcup \mathcal{F}$ is an upper bound)$\wedge$($\bigcup \mathcal{F}$ is smaller than all the elements in the set of upper bounds (the smallest upper bound)). This translated to

$$\forall x \in \mathcal{F}(xR\bigcup \mathcal{F}) \wedge \forall x \in \{X \in \mathscr{P}(A) | \forall x \in \mathcal{F}(xRX)\}(\bigcup \mathcal{F}Rx)$$

where the second term can be rewritten as

$$\forall x \in \mathscr{P}(A)(\forall y \in \mathcal{F}(yRx) \rightarrow \bigcup \mathcal{F}Rx)$$

and the subset partial order is defined as

$$R = \{(X, Y) \in \mathscr{P}(A) \times \mathscr{P}(A) | X \subseteq Y\}$$

It is also assumed that $\bigcup \mathcal{F} \in \mathscr{P}(A)$. Although this would be fairly straightforward to prove.
(next page)

Theorem. *Suppose A is a set, $\mathcal{F} \subseteq \mathscr{P}(A)$, and $\mathcal{F} \neq \emptyset$. Then the greatest lower bound of $\mathcal{F}$ (in the subset partial order) is $\bigcap \mathcal{F}$.*

Proof. (my proof) First we show that $\bigcap \mathcal{F}$ is a lower bound. Let x be an arbitrary element of $\mathcal{F}$. Suppose y is an arbitrary element of $\bigcap \mathcal{F}$. Then $y \in x$; since y is arbitrary, $\bigcap \mathcal{F} \subseteq x$. Since x is an arbitrary element of $\mathcal{F}$, $\bigcap \mathcal{F}$ is a lower bound.

Now we show that $\bigcap \mathcal{F}$ is the greatest lower bound. Let x be an arbitrary element of $\mathscr{P}(A)$. Suppose x is within the set of lower bounds. Let y be an arbitrary element of x. Since x is a lower bound, $\forall B \in \mathcal{F}(x \subseteq B)$, and since $y \in x$, $\forall B \in \mathcal{F}(y \in B)$, so $y \in \bigcap \mathcal{F}$. Since y is arbitrary, $x \subseteq \bigcap \mathcal{F}$ since x is arbitrary, all elements within the set of lower bounds is a subset of $\bigcap \mathcal{F}$, so $\bigcap \mathcal{F}$ is the greatest lower bound. □

I show ($\bigcap \mathcal{F}$ is a lower bound)$\wedge$($\bigcap \mathcal{F}$ is the greatest element in the set of lower bounds). This translates to

$$\forall x \in \mathcal{F}(\bigcap \mathcal{F} R x) \wedge \forall x \in \{X \in \mathscr{P}(A) | \forall y \in \mathcal{F}(X R y)\}(x R \bigcap \mathcal{F})$$

Where the second term can be rewritten in a manner similar to that in the previous proof.

Theorem. *Suppose R is a partial order on A, $B \in A$, and $b \in B$.*

1. *If b is the smallest element of B, then it is also the greatest lower bound of B.*

2. *If b is the largest element of B, then it is also the least upper bound of B.*

Proof.

1. Suppose b is the smallest element of B. Since $B \subseteq A$, b is in the set of lower bounds. Let x be an arbitrary element in the set of lower bounds. Then x is smaller than all the elements in B. Since $b \in B$, x must be smaller than b. Since x is arbitrary, b is the greatest lower bound.

2. Suppose b is the largest element of B. Since $B \subseteq A$, b is in the set of upper bounds. Let x be an arbitrary element in the set of upper bounds. Then x is larger than all the elements in B. Since $b \in B$, x must be larger than b. Since x is arbitrary, b is the least upper bound. □

3.4 Equivalence relations

Definition. Suppose R is a relation on set A. Then R is called an *equivalence relation on* A. (or just an *equivalence relation* if A is clear from context) if it is reflexive, symmetric, and transitive.

Intuition and examples
For instance, let T be the set of all triangles, and let C be the relation $C = \{(s,t) \in T \times T |$ the triangle s is congruent to the triangle $t\}$ (a triangle is congruent to another if it can be moved without distorting it so that it coincides with the other). Clearly every triangle is congruent to itself, so C is reflexive. If triangle s is congruent to triangle t, then t is congruent to s, so C is symmetric; if r is congruent to s, and s is congruent to t, then r is congruent to t, so C is transitive. Thus C is an equivalence relation on T.

For another example, let P be the set of all people, and let $B = \{(p,q) \in P \times P |$ the person p has the same birthday as the person $q\}$. (By 'same birthday' we mean same month and day, but not necessarily the same year.) It is straightforward to see that this relation is reflexive, symmetric, and transitive—B is an equivalence relation.

Consider B more closely, we can think of this relation as splitting the set P of all people into 366 categories, one for each possible birthday (taking into account Feb 29th on leap years). An ordered pair of people will only be an element of B if they both come from the same category. They will not be an element of B if both people came from different categories. As such we could think of these categories as forming a family of subsets of P, which we could write as an indexed family as follows:

First let D be the set of all possible birthdays $D = \{$Jan. 1, Jan. 2, Jan. 3, $\ldots$, Dec. 30, Dec. 31$\}$. Then the family $\mathcal{F} = \{P_d | d \in D\}$ is an indexed family of subsets of P. The elements of $\mathcal{F}$ (each subset) are called *equivalence classes* for the relation B, and every person is an element of exactly one of these equivalence classes. The relation B consists of those pairs $(p,q) \in P \times P$ such that the people p and q are in the same equivalence class:

$$\begin{aligned} B &= \{(p,q) \in P \times P | \exists d \in D (p \in P_d \text{ and } q \in P_d)\} \\ &= \{(p,q) \in P \times P | \exists d \in D ((p,q) \in P_d \times P_d)\} \\ &= \bigcup_{d \in D} (P_d \times P_d) \end{aligned}$$

We will call the family $\mathcal{F}$ a *partition* because it breaks the set P into disjoint pieces.

3.4.1 More definitions

Definition. Suppose A is a set and $\mathcal{F} \subseteq \mathscr{P}(A)$. We will say that $\mathcal{F}$ is *pairwise disjoint* if every pair of distinct elements of $\mathcal{F}$ are disjoint, or in other words $\forall X \in \mathcal{F} \forall Y \in \mathcal{F}(X \neq Y \rightarrow X \cap Y = \emptyset)$. $\mathcal{F}$ is called a *partition* of A if it has the following properties:

1. $\bigcup \mathcal{F} = A$
2. $\mathcal{F}$ is pairwise disjoint
3. $\forall X \in \mathcal{F}(X \neq \emptyset)$

Example
For instance, suppose $A = \{1, 2, 3, 4\}$ and $\mathcal{F} = \{\{2\}, \{1, 3\}, \{4\}\}$. Then $\bigcup \mathcal{F} = \{2\} \cup \{1, 3\} \cup \{4\} = \{1, 2, 3, 4\} = A$, so $\mathcal{F}$ satisfies the the first clause in the definition of partition. No two sets of $\mathcal{F}$ have any elements in common, so $\mathcal{F}$ is disjoint, and clearly all the sets in $\mathcal{F}$ are nonempty. Thus, $\mathcal{F}$ is a partition of A.

On the other hand, the family $\mathcal{G} = \{\{1, 2\}, \{1, 3\}, \{4\}\}$ is not pairwise disjoint because $\{1, 2\} \cap \{1, 3\} = \{1\} \neq \emptyset$. The family $\mathcal{H} = \{\emptyset, \{2\}, \{1, 3\}, \{4\}\}$ is also not a partition of A because it violates the third requirement in the definition.

Definition. Suppose R is an equivalence relation on set A, and $x \in A$. Then the *equivalence class of x with respect to R* is the set

$$[x]_R = \{y \in A | yRx\}$$

If R is clear from context then we just write $[x]$ instead of $[x]_R$. The set of all equivalence classes of elements of A is called A *modulo* R, and is denoted by A/R. Thus,

$$A/R = \{[x]_R | x \in A\} = \{X \subseteq A | \exists x \in A(X = [x]_R)\}$$

Note to self
(Recall that set notation can be written in multiple ways; for instance $\{n^2 | n \in \mathbb{N}\} = \{x | \exists n \in \mathbb{N}(x = n^2)\}$. Also recall that an element can appear more than once in a set and that set will still be considered to be the same as one with all unique elements. For instance $\{14, 3, 7\}$ and $\{3, 7, 14, 7\}$ are the same set.)
(next page)

Example
Let S be the relation $\mathbb{R}$ defined as follows:

$$S = \{(x, y) \in \mathbb{R} \times \mathbb{R} | x - y \in \mathbb{Z}\}$$

For example, $(5.73, 2.73) \in S$ and $(-1.27, 2.73) \in S$, since $5.73 - 2.73 = 3 \in \mathbb{Z}$ and $-1.27 - 2.73 = -4 \in \mathbb{Z}$, but $(1.27, 2.73) \notin S$ since $1.27 - 2.73 = -1.46 \notin \mathbb{Z}$.

Clearly for any $x \in \mathbb{R}, x - x = 0 \in \mathbb{Z}$, so $(x, x) \in S$ and S is reflexive. To see that S is symmetric, suppose $(x, y) \in S$; by the definition of S this means that $x - y \in \in Z$; but then $y - x = -(x - y) \in \mathbb{Z}$ too, since the negative of any integer is also an integer, so $(y, x) \in S$. Because (x, y) was an arbitrary element of S, S is symmetric. Finally, to see that S is transitive, suppose that $(x, y) \in S$ and $(y, z) \in S$; then $x - y \in \mathbb{Z}$ and $y - z \in \mathbb{Z}$. Because the sum of any two integers is an integer, it follows that $x - z = (x - y) + (y - z) \in \mathbb{Z}$, so $(x, z) \in S$ as required. Thus S is an equivalence relation on $\mathbb{R}$.

What would the equivalence classes for this equivalence relation look like? We know $(5.73, 2.73) \in S$ and $(-1.27, 2.73) \in S$, so $5.73 \in [2.73]$ and $-1.27 \in [2.73]$. Intuitively, the other elements of this equivalence class would look like

$$[2.73] = \{\ldots, -1.27, -0.27, 0.73, 1.73, 2.73, 3.73, 4.73, 5.73, \ldots\}$$

In general, for any real number x, the equivalence class of x will contain all real numbers that differ from x by an integer amount:

$$[x] = \{\ldots, x - 3, x - 2, x - 1, x, x + 1, x + 2, x + 3, \ldots\}$$

It turns out that x is always an element of $[x]$. It also turns out that if we choose any number $x \in [2.73]$, then $[x]$ will be exactly the same as $[2.73]$. For instance taking $x = 4.73$ we find that

$$[4.73] = \{\ldots, -1.27, -0.27, 0.73, 1.73, 2.73, 3.73, 4.73, 5.73, \ldots\} = [2.73]$$

Thus $[4.73]$ and $[2.73]$ are just two different names for the same set. But if we choose $x \notin [2.73]$, then $[x]$ will be different from $[2.73]$. For example,

$$[1.3] = \{\ldots, -1.7, -0.7, 0.3, 1.3, 2.3, 3.3, 4.3, \ldots\}$$

$[1.3]$ is actually *disjoint* from $[2.73]$. In general, for any two real numbers x and y, the equivalence classes $[x]$ and $[y]$ are either identical or disjoint. each equivalence classes has many different names, but different equivalent classes are disjoint.
(next page)

I wrote this proof myself, may contain errors.

Theorem. *Suppose R is an equivalence relation on $\mathbb{R}$. Then $\forall x \in \mathbb{R} \forall y \in \mathbb{R}(([x] = [y]) \vee ([x] \cap [y] = \emptyset)$; that is, the equivalence classes of any two real numbers are either identical or disjoint.*

Proof. Let x and y be arbitrary real numbers.

Case 1: $x \in [y]$, then by the definition of the equivalence class, xRy.

($\rightarrow$) Let z be an arbitrary real number. Suppose $z \in [x]$, then zRx. Since xRy, then by transitivity zRy, and by the definition of the equivalence class $z \in [y]$. Since z was arbitrary, $[x] \subseteq [y]$.

($\leftarrow$) Let z be an arbitrary real number. Suppose $z \in [y]$, then zRy. Since xRy by symmetry yRx and then by transitivity zRx; so $z \in [x]$. Since z was arbitrary, $[y] \subseteq [x]$. So we have $[x] = [y]$.

Case 2: $x \notin [y]$. Suppose $[x] \cap [y] \neq \emptyset$, then $\exists z \in \mathbb{R}(z \in [x] \wedge z \in [y])$, and we can declare some z such that $z \in [x]$ and $z \in [y]$, meaning zRx and zRy. By symmetry xRz, then by transitivity xRy. Since this means $x \in [y]$, this contradicts our initial assumption that $x \notin [y]$, so $[x] \cap [y] = \emptyset$. $\square$

I also had some doubts regarding the definition of A/R:

$$A/R = \{[x]_R | x \in A\} = \{X \subseteq A | \exists x \in A(X = [x]_R)\}$$

Particularly regarding the bit where $X \subseteq A$. Since the existential part of the second equality made sense but not the part where $X \subseteq A$. So I wrote another proof

Theorem. *Suppose R is an equivalence relation on A. Then $\forall X(\exists x \in A(X = [x]) \rightarrow X \subseteq A)$. That is, if X is an equivalence class in A, then $X \subseteq A$.*

Proof. Let X be some arbitrary set, suppose $\exists x \in A(X = [x]) \rightarrow X \subseteq A)$, then we can declare some $x \in A$ such that $X = [x]$. Now let y be some arbitrary element of X, then $y \in [x]$. From the definition of the equivalence class, this means that $y \in A$. Since y was arbitrary, $X \subseteq A$. $\square$

3.4.2 Equivalence relations and partitions

The set of all of the equivalence classes $\mathbb{R}/S$, is a partition of $\mathbb{R}$. The proof of this will be easier to understand if we first prove a few facts about equivalence classes. Facts that are proven primarily for the purpose of using them to prove a theorem are usually called *lemmas*.

Theorem. *Suppose R is an equivalence relation on set A. Then A/R is a partition of A.*

Lemma. *Suppose R is an equivalence relation on A. Then:*

1. *For every $x \in A$, $x \in [x]$.*
2. *For every $x \in A$ and $y \in A$, $y \in [x]$ iff $[y] = [x]$.*

Proof.

1. Let $x \in A$ be arbitrary. Since R is reflexive, xRx. Therefore, by the definition of the equivalence class, $x \in [x]$.
2. ($\rightarrow$) Let x, y be arbitrary elements of A. Suppose $y \in [x]$, then by the definition of the equivalence class, yRx.
   ($\rightarrow$) Now let z be an arbitrary element in $[y]$. Then zRy. Since zRy and yRx, by transitivity, we can conclude that zRx, so $z \in [x]$. Since z was arbitrary, $[y] \subseteq [x]$.
   ($\leftarrow$) Now suppose $z \in [x]$, so zRx. We already know yRx, so by symmetry xRy, then by transitivity zRy, so $z \in [y]$. Therefore $[x] \subseteq [y]$, and $[x] = [y]$.
   ($\leftarrow$) Again let x, y be arbitrary elements of A. Suppose $[y] = [x]$, by part 1 we know that $y \in [y]$, so since $[y] = [x]$, it follows that $y \in [x]$. $\square$

(next page)

Proof. To prove that A/R is a partition of A, we must prove the three properties that define a partition. First we must show that $\bigcup(A/R) = A$, or in other words $\bigcup_{x \in A}[x] = A$. Since every equivalence class in A/R is a subset of A, it should be clear that their union is also a subset of A (Suppose arbitrary $x \in \bigcup(A/R)$, then then use $\exists_B \in A/R(x \in B)$ to declare some B, then use $B \subseteq A$ to show $x \in A$). Thus $\bigcup(A/R) \subseteq A$. Now we show $A \subseteq \bigcup(A/R)$. To prove this, suppose $x \in A$, then by the first lemma $x \in [x]$, and of course $[x] \in A/R$, so $x \in \bigcup(A/R)$. Thus $\bigcup(A/R) = A$.

Next we show that A/R is pairwise disjoint. Let X and Y be two elements of A/R, and suppose $X \neq Y$. Now suppose that $X \cap Y \neq \emptyset$. By the definition of A/R, since $X \in A/R$ and $Y \in A/R$, we can declare some $x, y \in A$ such that $X = [x]$ and $Y = [y]$. Since $X \cap Y \neq \emptyset$, we can choose some z such that $z \in X \cap Y = [x] \cap [y]$, meaning $z \in [x]$ and $z \in [y]$. By the second lemma, it follows that $[x] = [z] = [y]$, so $X = Y$. Since this contradicts our initial assumption that $X \neq Y$, $X \cap Y = \emptyset$. So A/R is pairwise disjoint.

Finally we show that $\forall X \in A/R(X \neq \emptyset)$. Suppose $X \in A/R$. As before, we can then declare some $x \in A$ such that $X = [x]$. Now by the first lemma, $x \in [x] = X$, so $X \neq \emptyset$, as required. □

The book gave an intuitive reasoning for why $\bigcup(A/R) \subseteq A$. I included an outline for a formal proof within the parentheses. For the part where we show that A/R is pairwise disjoint, we are trying to show

$$\forall X \in A/R \forall Y \in A/R(X \neq Y \rightarrow X \cap Y = \emptyset)$$

I used a proof by contradiction, supposing $X \neq Y$ and using a contradiction to show $X \cap Y = \emptyset$. The book considers proving the contrapositive, meaning showing $X \cap Y \neq \emptyset \rightarrow X = Y$. Either way, the important part here is that the definition of A/R,

$$A/R = \{[x]_R | x \in A\} = \{X \subseteq A | \exists x \in A(X = [x])\}$$

So $X \in A/R$ implies the existential statement $\exists x \in A(X = [x])$, and we go from there.

3.4.3 More on equivalence relations and partitions

Theorem. *Suppose A is a set and $\mathcal{F}$ is a partition of A. Then there is an equivalence relation R on A such that $A/R = \mathcal{F}$.*

Intuition
The conclusion of the theorem is an existential statement, we should find an equivalence relation R such that $A/R = \mathcal{F}$. Clearly for different choices of $\mathcal{F}$ we need to choose R differently, so the definition of R should depend on $\mathcal{F}$ in some way.

Recall the same birthday example at the beginning of this section. The equivalence relation B consisted of all pairs of people (p, q) such that p and q were in the same set in the partition $\{P_d | d \in D\}$. In fact, we found that we could also express this by saying that $B = \bigcup_{d \in D}(P_d \times P_d)$.

Consider applying the same principle to this theorem. We should let R be the set of all pairs $(x, y) \in A \times A$ such that x and y are in the same set in the partition $\mathcal{F}$. That is, $R = \bigcup_{X \in \mathcal{F}}(X \times X)$.

Naturally, the reasoning that led us to the formula $R = \bigcup_{X \in \mathcal{F}}(X \times X)$ will not be part of the proof. When we write the proof, we can simply define R in this way and then verify that it is an equivalence relation on A and that $A/R = \mathcal{F}$.
(next page)

Lemma. *Suppose A is a set and $\mathcal{F}$ is a partition of A. Let $R = \bigcup_{X \in \mathcal{F}}(X \times X)$. Then R is an equivalence relation on A. We will call R the equivalence relation determined by $\mathcal{F}$.*

Proof. First we show that R is reflexive. Let x be an arbitrary element of A. Since $\mathcal{F}$ is a partition of A, $\bigcup \mathcal{F} = A$, so $x \in \bigcup \mathcal{F}$. Thus we can choose some $B \in \mathcal{F}$ such that $x \in B$, and so $(x, x) \in B \times B$, and $(x, x) \in \bigcup_{X \in \mathcal{F}}(X \times X) = R$. Since x is arbitrary, R is reflexive.

Now we show that R symmetric. Let x, y be arbitrary elements of A. Suppose $(x, y) \in R = \bigcup_{X \in \mathcal{F}}(X \times X)$. Then we can declare some $B \in \mathcal{F}$ such that $(x, y) \in B \times B$; so by the definition of the cartesian product, $x \in B$ and $y \in B$, and $(y, x) \in B \times B$. So $(y, x) \in \bigcup_{X \in \mathcal{F}}(X \times X) = R$. Since x and y are arbitrary, R is symmetric.

Finally we show that R is transitive. Let x, y, z be arbitrary elements in A. Suppose $(x, y) \in R$ and $(y, z) \in R$, where $R = \bigcup_{X \in \mathcal{F}}(X \times X)$. As such we can declare some B, C in $\mathcal{F}$ such that $(x, y) \in B \times B$ and $(y, z) \in C \times C$. Then $x \in B$, $y \in B$, $y \in C$, $z \in C$; since B and C are both in $\mathcal{F}$, and $\mathcal{F}$ is a partition, B and C are pairwise disjoint, so since $B \cap C \neq \emptyset$, $B = C$. So since $z \in C = B$ and $x \in B$, $(x, z) \in B \times B$ and $(x, z) \in \bigcup_{X \in \mathcal{F}}(X \times X) = R$. Since x, y, z are arbitrary, R is transitive. □

Note that $\bigcup_{X \in \mathcal{F}}(X \times X)$ implies an existential statement.

$$(x, y) \in \bigcup_{X \in \mathcal{F}} (X \times X) \leftrightarrow \exists X \in \mathcal{F}((x, y) \in X \times X)$$

For the symmetry part of the proof, we use the first clause in the definition of the partition $\bigcup \mathcal{F} = A.$, and the definition of the cartesian product:

$$A \times B = \{a, b | a \in A \wedge b \in B\}$$

Finally for the transitive part, we use the fact that $\mathcal{F}$ is a partition on A, meaning it has the property of pairwise disjointness (second clause):

$$\forall X \in \mathcal{F} \forall Y \in \mathcal{F}(X \neq Y \rightarrow X \cap Y = \emptyset)$$

where using the contrapositive, this is equivalent to

$$\forall X \in \mathcal{F} \forall Y \in \mathcal{F}(X \cap Y \neq \emptyset \rightarrow X = Y)$$

(next page)

For clarity, here is a proof written by me that shows that R, defined as $R = \bigcup_{X \in \mathcal{F}}(X \times X)$, is a relation on A. Meaning $R \subseteq A \times A$.

Theorem. *Let $\mathcal{F}$ be a partition on some set A, and let $R = \bigcup_{X \in \mathcal{F}}(X \times X)$. Then R is a relation on A.*

Proof. (by me) Let (x, y) be an arbitrary element in R. Then we can declare some $B \in \mathcal{F}$ such that $x \in B$ and $y \in B$. Since $\bigcup \mathcal{F}$ is defined as the set $\{x | \exists X \in \mathcal{F}(x \in X)\}$, both x and y are in $\bigcup \mathcal{F}$. Since $\mathcal{F}$ is a partition on A, $\bigcup \mathcal{F} = A$. So both x and y are in A, and therefore $(x, y) \in A \times A$. Since (x, y) was arbitrary, $R \subseteq A \times A$. □

Lemma. *Suppose A is a set and $\mathcal{F}$ is a partition of A. Let R be the equivalence determined by $\mathcal{F}$. Suppose $X \in \mathcal{F}$ and $x \in X$. Then $[x]_R = X$.*

Proof. Let y be an arbitrary element of $[x]_R$. Then $(y, x) \in R$, so by the definition of R there must be some $Y \in \mathcal{F}$ such that $y \in Y$ and $x \in Y$. Since $x \in X$ and $x \in Y$, $X \cap Y \neq \emptyset$, and since $\mathcal{F}$ is pairwise disjoint it follows that $X = Y$. Thus since $y \in Y$, $y \in X$. Since y was arbitrary, $[x]_R \subseteq X$.

Now let y be an arbitrary element of X. Then $(y, x) \in X \times X$, so $(y, x) \in R$ and therefore $y \in [x]_R$. Since y was arbitrary, $X \subseteq [x]_R$, and $[x]_R = X$. □

In the first part of this proof we use the contrapositive of the pairwise disjointness clause.

Now we prove the initial theorem.

Theorem. *Suppose A is a set and $\mathcal{F}$ is a partition of A. Then there is an equivalence relation R on A such that $A/R = \mathcal{F}$.*

Proof. Let $R = \bigcup_{X \in \mathcal{F}}(X \times X)$. We have already seen that R is an equivalence relation, so we need only check that $A/R = \mathcal{F}$. To see this, let X be an arbitrary element of A/R. Then this means $X = [x]$ for some $x \in A$. Since $\mathcal{F}$ is a partition, we know that $\bigcup \mathcal{F} = A$, so $x \in \bigcup \mathcal{F}$, and therefore we can choose some $Y \in \mathcal{F}$ such that $x \in Y$. But then by the second lemma, $[x] = Y$. Thus $X = Y \in \mathcal{F}$, so $A/R \subseteq \mathcal{F}$.

Now let X be an arbitrary element of $\mathcal{F}$. Then since $\mathcal{F}$ is a partition, $X \neq \emptyset$, so we can choose some $x \in X$. By the second lemma, $X = [x] \in A/R$ so $\mathcal{F} \subseteq A/R$, and $A/R = \mathcal{F}$. □

We use the third clause in the definition of the parition for the second part of the proof.

3.4.4 Even more on equivalence relations and partitions

We have seen how an equivalence relation R on a set A can be used to define a partition A/R of A; recall the theorem:

Theorem. *Suppose R is an equivalence relation on set A. Then A/R is a partition of A.*

We've also seen how a partition $\mathcal{F}$ of A can be used to define an equivalence relation $\bigcup X \in \mathcal{F}(X \times X)$ on A. The proof of the theorem from the previous section demonstrates an interesting relationship between these operations:

Theorem. *Suppose A is a set and $\mathcal{F}$ is a partition of A. Then there is an equivalence relation R on A such that $A/R = \mathcal{F}$.*

The proof showed that if we start with a partition $\mathcal{F}$ on A, and use $\mathcal{F}$ to define the equivalence relation $Q = \bigcup X \in \mathcal{F}(X \times X)$, and then used Q to define a partition A/Q, this partition would be equal to $\mathcal{F}$.

One might wonder whether if we started with an equivalence relation R, and used R to define a partition $\mathcal{F} = A/R$, and then used $\mathcal{F}$ to define an equivalence relation $S = \bigcup X \in \mathcal{F}(X \times X)$, would this final equivalence relation S be the same as the original equivalence relation R?

Lemma. *Suppose R and S are equivalence relations on A and $A/R = A/S$. Then $R = S$.*

Proof. (my proof) First we show $S \subseteq R$. Let (x, y) be an arbitrary element in S. It was shown earlier that $y \in [y]_S$, and since xSy, then x and y are both in $[y]_S \in A/S$. Since $A/S = A/R$, $[y]_S \in A/R$, so there exists some $z \in A$ such that $[y]_S = [z]_R$. So $x \in [z]_R$ and $y \in [z]_R$, so xRz and yRz. By symmetry zRy, and by transitivity, xRy; so $(x, y) \in R$. Since (x, y) was arbitrary, $S \subseteq R$.

Now we show $R \subseteq S$. Let (x, y) be an arbitrary element of R. Then similarly, $x \in [y]_R$ and $y \in [y]_R$. Since $[y]_R \in A/R = A/S$, we can declare some $z \in A$ such that $[y]_R = [z]_S$. So $x \in [z]_S$ and $y \in [z]_S$, meaning xSz and ySz. By symmetry, zSy, and by transitivity, xSy, so $(x, y) \in S$. Since (x, y) was arbitrary, $R \subseteq S$ and $R = S$. □

Theorem. *Suppose R is an equivalence relation on A. Let $\mathcal{F} = A/R$, and let S be the equivalence relation determined by $\mathcal{F}$. In other words, $S = \bigcup X \in \mathcal{F}(X \times X)$. Then $S = R$.*

Proof. We've shown that $\mathcal{F} = A/R$ is a partition of A. By the theorem derived in the previous section, the equivalence relation determined by $\mathcal{F}$, $\bigcup X \in \mathcal{F}(X \times X) = S$, exists such that $A/S = \mathcal{F}$. So $A/S = \mathcal{F} = A/R$. By the previous lemma, since R and S are both equivalence relations, $S = R$. □

3.4.5 Congruence modulo m

Recall the definition of $x|y$.

Definition. For any integers x and y, we'll say that x *divides* y (or y *is divisible by* x) if $\exists k \in \mathbb{Z}(kx = y)$. We use the notation $x \mid y$ to mean 'x divides y' and $x \nmid y$ to mean 'x does not divide y'.

Definition. Suppose m is a positive integer. For any integers x and y, we will say that x is *congruent to* y *modulo* m if $\exists k \in \mathbb{Z}(x - y = km)$. In other words, x is congruent to y modulo m iff $m|(x - y)$. We will use the notation $x \equiv y \pmod{m}$ to mean that x is congruent to y modulo m.

Example
For example, $12 \equiv 27 \pmod{5}$, since $12 - 27 = -15 = (-3) \cdot 5$.

Intuition
Now for any positive integer m we can consider the relation

$$\{(x, y) \in \mathbb{Z} \times \mathbb{Z} | x \equiv y \pmod{m}\}$$

We will use the symbol $\equiv_m$ to denote this relation. For any integers x and y, $x \equiv_m y$ means the same thing as $x \equiv y \pmod{m}$. It turns out that this relation is an equivalence relation. We will prove the following theorem:

Theorem. *For every positive integer m, $\equiv_m$ is an equivalence relation on $\mathbb{Z}$.*

Lemma. *For integers a, b, c, if $a|b$ and $a|c$, then $a|(b + c)$.*

Proof. Suppose $a|b$ and $a|c$, then we can declare some integers k_0 and k_1 such that $ak_0 = b$ and $ak_1 = c$. Then $a(k_0 + k_1) = b + c$. Since $(k_0 + k_1)$ is an integer, $a|(b + c)$. □

Proof. (my proof) First we consider reflexivity. Let x be an arbitrary integer. Choosing $k = 0$, $km = (x - x) = 0$, and therefore $x \equiv_m x$. Since x is arbitrary, $\equiv_m$ is reflexive.

Now we show symmetry. Let x and y be arbitrary integers. Suppose $x \equiv_m y$; then we can declare some integer k_0 such that $k_0 m = (x - y)$. Then $-k_0 m = (y - x)$; since $-k_0$ is also an integer, $y \equiv_m x$. Since x and y are arbitrary, $\equiv_m$ is symmetric.

Finally we show transitivity. Let x, y, z be arbitrary integers. Suppose that $x \equiv_m y$ and $y \equiv_m z$. This means that $m|(x - y)$ and $m|(y - z)$. Therefore by the previous lemma, $m|[(x - y) + (y - z)]$. But $(x - y) + (y - z) = x - z$, so it follows that $m|(x - z)$, and therefore $x \equiv_m z$. Since x, y, z were arbitrary, $\equiv_m$ is transitive. □

Chapter 4

Functions

4.1 Definition

Definition. Suppose F is a relation from A to B. Then F is called a *function from A to B* if for every $a \in A$ there is exactly one $b \in B$ such that $(a, b) \in F$. In other words, to say that F is a function from A to B means:

$$\forall a \in A \exists! b \in B((a, b) \in F)$$

To indicate that F is a function from A to B, we will write $F : A \to B$.

'Exactly one' notation
To recapitulate, we write

$$\exists! x P(x) = \exists x(P(x) \wedge \neg \exists y(P(y) \wedge y \neq x))$$

As a shorthand way to write 'there is exactly one value of x such that $P(x)$ is true', or 'there is a unique x such that $P(x)$'.

Examples
Let $A = \{1, 2, 3\}$, $B = \{4, 5, 6\}$, and $F = \{(1, 5), (2, 4), (3, 5)\}$. 1 is paired with 5 in the relation F, and is not paired with any other element of B. Similarly, 2 is paired only with 4, and 3 with 5; F is therefore a function. Note that the definition of the function does *not* require that each element of B be paired with exactly one element of A (see that 5 is paired with 1 and 3), only that each element of A maps to only one B.

Now let $A = \{1, 2, 3\}$, $B = \{4, 5, 6\}$, and $G = \{(1, 5), (2, 4), (1, 6)\}$. G is not a function for multiple reasons; firstly 3 isn't paired with any element of B in the relation G, which violates the requirement that every element of A must be paired with some element of B. Second, 1 is paired with two different elements of B, 5 and 6, which violates the requirement that each element of A be paired with *only one* element of B.
(next page)

More examples
Let A be any set, recall that $i_A = \{(a, a)|a \in A\}$ is called the identity relation on A. This is also a function from A to A. Each $a \in A$ is paired in the relation i_A with exactly one element of A, namely a itself.

Let $f = \{(x, y) \in \mathbb{R} \times \mathbb{R}|y = x^2\}$. f is a function from $\mathbb{R}$ to $\mathbb{R}$; for each real number x there is exactly one value of y, namely $y = x^2$, such that $(x, y) \in f$.

Terminology and specification by rule
Suppose $f : A \to B$. If $a \in A$, then we know that there is exactly one $b \in B$ such that $(a, b) \in f$. This unique b is called "the value of f at a" or "the image of a under f" or "the result of applying f to a", or just "f of a", and it is written $f(a)$. That is, for every $a \in A$ and $b \in B$, $b = f(a)$ iff $(a, b) \in f$. For example, for the function $F = \{(1, 5), (2, 4), (3, 5)\}$, we could say that $F(1) = 5$, since $(1, 5) \in F$; similarly $F(2) = 4$ and $F(3) = 5$.

A function f from a set A to a set B is often specified by giving a rule that can be used to determine $f(a)$ from any $a \in A$. For example, if A is the set of all people and $B = \mathbb{R}^+$, we could define a function f from A to B by the rule that for every $a \in A, f(a) = a$'s height in inches. This is another way of defining a function; although this definition doesn't say explicitly which ordered pairs are elements of f, recall that for all $a \in A$ and $b \in B$, $b = f(a)$ iff $(a, b) \in f$. From this equivalence we can say

$$f = \{(a, b) \in A \times B|b = f(a)\}$$

then using the given rule, this becomes

$$= \{(a, b) \in A \times B|b = a\text{'s height in inches}\}$$

This idea of thinking of a function as representing a rule that associates, with each $a \in A$, some corresponding object $b = f(a) \in B$ is useful. However it is important to remember that although a function can be defined by giving such a rule, it need not be defined in this way—any subset of $A \times B$ that satisfies the requirements outlined in the definition is a function from A to B.

For another example, we can define a function g from $\mathbb{Z}$ to $\mathbb{R}$ by the rule that for every $x \in \mathbb{Z}$, $g(x) = 2x + 3$. Then

$$\begin{aligned} g &= \{(x, y) \in \mathbb{Z} \times \mathbb{R}|y = g(x)\} \\ &= \{(x, y) \in \mathbb{Z} \times \mathbb{R}|y = 2x + 3\} \\ &= \{\ldots, (-2, -1), (-1, 1), (0, 3), (1, 5), (2, 7), \ldots\} \end{aligned}$$

Note that when a function f from A to B is specified by giving a rule for finding $f(a)$, the rule must determine the value of $f(a)$ for *every* $a \in A$.

4.1.1 Function equality, Domain/Range/Composition

Since a function f from A to B is completely determined by the rule for finding $f(a)$, two functions that are defined by equivalent rules must be equal:

Theorem. *Suppose f and g are functions from A to B. If $\forall a \in A(f(a) = g(a))$, then $f = g$.*

Proof. Suppose $\forall a \in A(f(a) = g(a))$. Let (a, b) be an arbitrary element of f. Then $b = f(a)$. But by our assumption $f(a) = g(a)$, so $b = g(a)$ and therefore $(a, b) \in g$. Thus, $f \subseteq g$. A similar argument shows $g \subseteq f$, so $f = g$. □

Now that we have proven this, we have another method for showing that two functions f and g from a set A to another set B are equal. In the future, we may prove $\forall a \in A(f(a) = g(a))$ and then apply this theorem.

Domain, Range, Composition
Recall these ideas from the previous chapter

Definition. Suppose R is a relation from A to B. Then the *domain* of R is the set

$$\text{Dom}(R) = \{a \in A | \exists b \in B((a, b) \in R)\}$$

The *range* of R is the set

$$\text{Ran}(R) = \{b \in B | \exists a \in A((a, b) \in R)\}$$

Definition. Suppose R is a relation from A to B and S is a relation from B to C. Then the *composition* of S and R is the relation $S \circ R$ from A to C defined as

$$S \circ R = \{(a, c) \in A \times C | \exists b \in B((a, b) \in R \text{ and } (b, c) \in S)\}$$

Properties
Since functions are relations, these concepts can be applied here as well. For instance suppose $f : A \to B$. Then f is a relation from A to B so it makes sense to talk about the domain of f, which is a subset of A, and the range of f, which is a subset of B.

According to the definition of the function, every element of A must appear as the first coordinate of exactly one ordered pair in f, so the domain of f must be all of A. But the range of f need not be all of B. The elements of the range of f will be the second coordinate of all the ordered pairs in f, also called the image of the first coordinate; thus the range of f could also be described as the set of all images of the elements of A under f:

$$\text{Ran}(f) = \{f(a) | a \in A\}$$

(next page)

Cont.
The definition of composition of relations can also be applied to functions. If $f : A \to B$ and $g : B \to C$, then f is a relation from A to B and g is a relation from B to C. So $g \circ f$ will be a relation from A to C. In fact it turns out that $g \circ f$ is a function from A to C.

Theorem. *Suppose $f : A \to B$ and $g : B \to C$. Then $g \circ f : A \to C$, and for any $a \in A$, the value of $g \circ f$ at a is given by the formula $(g \circ f)(a) = g(f(a))$.*

Existence and uniqueness
Before proving this theorem, recall the definition of a function. To show that $g \circ f : A \to C$ we must prove that $\forall x \in A \exists! c \in C((a, c) \in g \circ f)$ meaning that for all $a \in A$ there exists a unique c (exactly one) such that $(a, c) \in g \circ f$.

The idea of 'exactly one' implies existence and uniqueness:

$$\exists! x P(x) = \exists x(P(x) \wedge \neg \exists y(P(y) \wedge y \neq x)) = \exists x(P(x) \wedge \forall y(P(y) \to y = x))$$

Recall from the section on this that the following are equivalent:

1. $\exists x(P(x) \wedge \forall y(P(y) \to y = x))$
2. $\exists x \forall y(P(y) \leftrightarrow y = x)$
3. $\exists x P(x) \wedge \forall y \forall z((P(y) \wedge P(z)) \to y = z)$

(see that the existence and uniqueness only 'goes one way'; in the sense that there must exist only one b for each a, but nothing is said about there being multiple a for each b.) In this proof we take the third approach; proving existence and uniqueness separately.

We are trying to show that $g \circ f$ is a function from A to C, meaning $\forall a \in A \exists! c \in C((a, c) \in g \circ f)$. Breaking down the 'exactly' one part we get the final expression

$$\begin{aligned} &\forall a \in A(\exists c \in C((a, c) \in g \circ f) \\ &\qquad \wedge \forall c_1 \in C \forall c_2 \in C(((a, c_1) \in g \circ f \wedge (a, c_2) \in g \circ f) \to c_1 = c_2)) \end{aligned}$$

(next page)

Theorem. *Suppose $f : A \to B$ and $g : B \to C$. Then $g \circ f : A \to C$, and for any $a \in A$, the value of $g \circ f$ at a is given by the formula $(g \circ f)(a) = g(f(a))$.*

Proof. Let a be an arbitrary element of A. We wil show existence and uniqueness separately.

Existence: Let $b = f(a) \in B$. Let $c = g(b) \in C$. Then $(a, b) \in f$ and $(b, c) \in g$, so by the definition of composition of relations, $(a, c) \in g \circ f$. Thus, $\exists c \in C((a, c) \in g \circ f)$.

Uniqueness: Suppose $(a, c_1) \in g \circ f$ and $(a, c_2) \in g \circ f$. Then by the definition of composition, we can choose $b_1 \in B$ such that $(a, b_1) \in f$ and $(b_1, c_1) \in g$, and we can choose $b_2 \in B$ such that $(a, b_2) \in f$ and $(b_2, c_2) \in g$. Since f is a function, there can be only one $b \in B$ such that $(a, b) \in f$. Thus since (a, b_1) and (a, b_2) are elements of f, it follows that $b_1 = b_2$. But now applying the same reasoning to g, since $(b_1, c_1) \in g$ and $(b_1, c_2) = (b_2, c_2) \in g$, it follows that $c_1 = c_2$ as required.

This completes the proof that $g \circ f$ is a function from A to C. Finally, to derive the formula for $(g \circ f)(a)$, note that we showed in the existence half of the proof that for any $a \in A$, if we let $b = f(a)$ and $c = g(b)$, then $(a, c) \in g \circ f$. Thus

$$(g \circ f)(a) = c = g(b) = g(f(a)) \qquad \square$$

I felt that the existence part and the last part of this proof were a bit opaque. I've included my own proof of those parts here:

Proof. Let a be an arbitrary...

$\vdots$

Existence: Since $a \in A$ and f is a function from A to B we can declare some $b \in B$ such that $(a, b) \in f$. Then since g is a function from B to C we can declare some $c \in C$ such that $(b, c) \in g$. Then by the definition of composition of functions, $(a, c) \in g \circ f$, thus $\exists c \in C((a, c) \in g \circ f)$.

$\vdots$

Finally to show the formula for $(g \circ f)(a)$, revisit the existence part of this proof, which found that for $a \in A$, we could declare $b \in B$ such that $b = f(a)$, and then that we could declare $c \in C$ such that $c = g(b)$. We then found that $(a, c) \in g \circ f$, which means $(g \circ f)(a) = c$. Thus

$$(g \circ f)(a) = c = g(b) = g(f(a)) \qquad \square$$

Remember that functions are just relations of a special kind. Everything proven about compositions of relations applies to composition of functions. (for instance associativity which was proven in an earlier section)

4.2 One-to-One and Onto

Definition. Suppose $f : A \to B$. We will say that f is *one-to-one* if

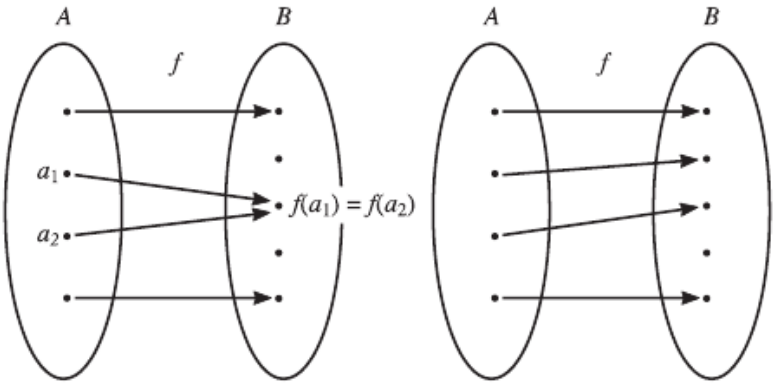
$$\neg\exists a_1 \in A \exists a_2 \in A(f(a_1) = f(a_2) \wedge a_1 \neq a_2)$$

We say that f *maps onto* B (or just *is onto* if B is clear from context) if

$$\forall b \in B \exists a \in A(f(a) = b)$$

One-to-one functions are sometimes also called *injections*, and onto functions are sometimes called *surjections*.

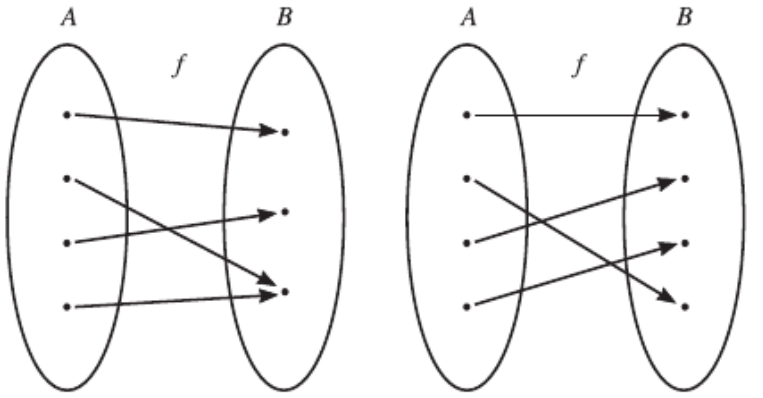
Intuition
Note that our definition of one-to-one starts with the negation symbol. In other words, to say that f is one-to-one means that a certain situation does *not* occur. Namely, it must not be the case that there are two different elements of the domain of f, a_1 and a_2, such that $f(a_1) = f(a_2)$:



(a) f is not one-to-one. (b) f is one-to-one.

(next page)

Cont.
If $f : A \to B$, then to say that f is onto means that every element of B is the image under f of some element of A. In other words, in the diagram of f, every element of B has an edge pointing to it. In the figure on the previous page, neither of the functions are onto, because in both cases there are elements of B without edges pointing to them. The following two functions are onto:



(a) f is onto but not one-to-one. (b) f is both one-to-one and onto.

Example
Consider the function g from $\mathbb{Z}$ to $\mathbb{R}$ defined by $g(x) = 2x+3$. To decide whether g is one-to-one, we must determine whether there are integers n_1 and n_2 such that $g(n_1) = g(n_2)$ and $n_1 \neq n_2$. According to the definition of g, we have

$$\begin{aligned} g(n_1) = g(n_2) &\text{ iff } 2n_1 + 3 = 2n_2 + 3 \\ &\text{iff } 2n_1 = 2n_2 \\ &\text{iff } n_1 = n_2 \end{aligned}$$

Thus there can be no integers n_1 and n_2 for which $g(n_1) = g(n_2)$ and $n_1 \neq n_2$. So g is one-to-one.

However, g is not onto; for example there is no integer n for which $g(n) = 0$. To see why suppose n is an integer and $g(n) = 0$. Then by the definition of g we have $2n + 3 = 0$ and $n = -3/2$, contradicting the fact that n is an integer. Since the domain of g is $\mathbb{Z}$, for g to be onto it must be the case that for every real number y there is an *integer* n such that $g(n) = y$. This counterexample allows us to conclude that g is not onto. (next page)

Theorem. *Suppose* $f : A \to B$.

1. f *is one-to-one iff* $\forall a_1 \in A \forall a_2 \in A(f(a_1) = f(a_2) \to a_1 = a_2)$.
2. f *is onto iff* $Ran(f) = B$.

Proof.

1. We have

$$\begin{aligned} f \text{ is one-to-one iff } & \neg\exists a_1 \in A \exists a_2 \in A(f(a_1) = f(a_2) \wedge a_1 \neq a_2) \\ \text{iff } & \forall a_1 \in A \forall a_2 \in A \neg(f(a_1) = f(a_2) \wedge a_1 \neq a_2) \\ \text{iff } & \forall a_1 \in A \forall a_2 \in A(f(a_1) = f(a_2) \to a_1 = a_2) \end{aligned}$$

2. First see that we have the equivalence:

$$\begin{aligned} f \text{ is onto iff } & \forall b \in B \exists a \in A(f(a) = b) \\ \text{iff } & \forall b \in B \exists a \in A((a, b) \in f) \\ \text{iff } & \forall b \in B(b \in \text{Ran}(f)) \\ \text{iff } & B \subseteq \text{Ran}(f) \end{aligned}$$

   ($\to$) Suppose f is onto. By the equivalence just derived we have $B \subseteq \text{Ran}(f)$. and by the definition of range we have $\text{Ran}(f) \subseteq B$. Thus it follows that $\text{Ran}(f) = B$.
   ($\leftarrow$) Suppose $\text{Ran}(f) = B$. Then certainly $B \subseteq \text{Ran}(f)$, so by the equivalence, f is onto. □

(next page)

Example. Let $A = \mathbb{R}\backslash\{-1\}$, and define $f : A \to \mathbb{R}$ by the formula

$$f(a) = \frac{2a}{a+1}$$

Prove that f is one-to-one but not onto.

Proof. To see that f is one-to-one, let a_1 and a_2 be arbitrary elements of A and suppose that $f(a_1) = f(a_2)$. Applying the definition of f, it follows that $2a_1/(a_1+1) = 2a_2/(a_2+1)$. Multiplying out both sides gives us $2a_1a_2 + 2a_1 = 2a_1a_2 + 2a_2$, so $2a_1 = 2a_2$ and therefore $a_1 = a_2$.

To show that f is not onto we will prove that $\forall a \in A(f(a) \neq 2)$. Suppose $a \in A$ and $f(a) = 2$. Applying the definition of f, we get $2a/(a+1) = 2$. Thus, $2a = 2a + 2$ which is clearly impossible. Thus $f(a) \neq 2$ for arbitrary a and f is not onto. □

Intuition for proof
By the previous theorem we can prove f is one-to-one by proving the equivalent statement $\forall a_1 \in A \forall a_2 \in A(f(a_1) = f(a_2) \to a_1 = a_2)$. The one-to-one part of the proof is fairly straightforward from there.

To show that f is not onto we prove $\neg\forall x \in \mathbb{R}\exists a \in A(f(a) = x)$. This is equivalent to showing $\exists x \in \mathbb{R}\forall a \in A(f(a) \neq x)$, so we should try to find a particular real number x such that $\forall a \in A(f(a) \neq x)$. Unfortunately it is not clear what value we should use for x. We consider a somewhat unusual procedure to get around this: consider trying to prove that f *is* onto; we expect this won't work, but it might hint at the value of x we are trying to find.

We would have to prove $\forall x \in \mathbb{R}\exists a \in A(f(a) = x)$ so we let x be an arbitrary real number and try to find some $a \in A$ such that $f(a) = x$. Using the definition of f, we must find some $a \in A$ such that

$$\frac{2a}{a+1} = x$$

expression in terms of a gives us

$$\frac{2a}{a+1} = x \leftrightarrow 2a = ax + x \leftrightarrow a(2-x) = x \leftrightarrow a = \frac{x}{2-x}$$

Now see that the last step in this derivation wouldn't work if $x = 2$, giving us a candidate value for x. The rest of the proof is fairly straightforward; we use contradiction.

4.2.1 Composition of functions

Theorem. *Suppose $f : A \to B$ and $g : B \to C$. As we saw earlier, it follows that $g \circ f : A \to C$.*

1. *If f and g are both one-to-one, then so is $g \circ f$.*
2. *If f and g are both onto, then so is $g \circ f$.*

Proof.

1. Suppose f and g are both one-to-one. Let a_1 and a_2 be arbitrary elements of A and suppose that $(g \circ f)(a_1) = (g \circ f)(a_2)$. By an earlier theorem this means that $g(f(a_1)) = g(f(a_2))$. Since g is one-to-one it follows that $f(a_1) = f(a_2)$, and similarly since f is one-to-one we can then conclude that $a_1 = a_2$. Thus $g \circ f$ is one-to-one.
2. Suppose f and g are both onto. Let c be an arbitrary element of C. Since g is onto, we can find some $b \in B$ such that $c = g(b)$. Similarly since f is onto, there is some $a \in A$ such that $b = f(a)$. Then $(g \circ f)(a) = g(f(a)) = g(b) = c$. Thus $g \circ f$ is onto. □

Functions that are both one-to-one and onto are sometimes called *one-to-one correspondences* or *bijections*.

4.3 Inverses of functions

One may question whether the inverse of a function from A to B is always a function from B to A. Consider the function F, where $A = \{1, 2, 3\}$, $B = \{4, 5, 6\}$, and $F = \{(1, 5), (2, 4), (3, 5)\}$. F is a function from A to B. But consider F^{-1}, which by definition is $F^{-1} = \{(5, 1), (4, 2), (5, 3)\}$, which is clearly a relation from B to A, but fails to be a function from B to A for two reasons: firstly $6 \in B$ but isn't paired with any element of A in the relation F^{-1}. Second, 5 is paired with two different elements of A, 1 and 3.

The inverse of a function from A to B is not always a function from B to A.

Theorem. *Suppose $f : A \to B$. If f is one-to-one and onto, then $f^{-1} : B \to A$.*

Proof. Suppose f is one-to-one and onto, and let b be an arbitrary element of B. To show f^{-1} is a function form B to A. We must prove that $\exists! a \in A((a, b) \in f^{-1})$, so we prove existence and uniqueness separately.

Existence: Since f is onto, there is some $a \in A$ such that $b = f(a)$. Thus $(a, b) \in f$, so $(b, a) \in f^{-1}$.

Uniqueness: Suppose $(b, a_1) \in f^{-1}$ and $(b, a_2) \in f^{-1}$ for some $a_1, a_2 \in A$. Then $(a_1, b) \in f$ and $(a_2, b) \in f$, so $f(a_1) = b = f(a_2)$. Since f is one-to-one, it follows that $a_1 = a_2$. □

The form of this proof is guided by the logical form of the statement $f^{-1} : B \to A$, meaning $\forall b \in B \exists! a \in A((b, a) \in f^{-1})$. We let b be arbitrary and prove existence and uniqueness separately. Note that the assumption that f is onto is the key to the existence half of the proof, and the assumption that f is one-to-one is the key to the uniqueness half.

Sufficient, Necessary, Converse
Recall that saying that A is a sufficient condition for B means $A \to B$. Saying that A is a necessary condition for B means $\neg A \to \neg B$; by contrapositivity this is equivalent to $B \to A$—the converse of $A \to B$.

Suppose f is any function from the set A to a set B. the previous theorem says that a sufficient condition for f^{-1} to be a function from B to A is that f is one-to-one and onto.

Intuition
Again suppose $f : A \to B$. If $f^{-1} : B \to A$, then by definition, for every $b \in B$ there is exactly one $a \in A$ such that $(b, a) \in f^{-1}$ and

$$\begin{aligned} f^{-1}(b) &= \text{the unique } a \in A \text{ such that } (b, a) \in f^{-1} \\ &= \text{the unique } a \in A \text{ such that } (a, b) \in f \\ &= \text{the unique } a \in A \text{ such that } f(a) = b \end{aligned}$$

4.3.1 Identity function

Recall the theorem proved earlier regarding function equality. To prove that two functions are equal we usually apply this theorem.

Theorem. *Suppose f and g are functions from A to B. If $\forall a \in A(f(a) = g(a))$, then $f = g$.*

Theorem. *Suppose f is a function from A to B, and suppose that f^{-1} is a function from B to A. Then $f^{-1} \circ f = i_A$ and $f \circ f^{-1} = i_B$.*

Proof. Let a be an arbitrary element of A. Then $b = f(a) \in B$. Then $(a, b) \in f$, so $(b, a) \in f^{-1}$ and therefore $f^{-1}(b) = a$. Thus

$$(f^{-1} \circ f)(a) = f^{-1}(f(a)) = f^{-1}(b) = a = i_A(a)$$

Since a was arbitrary, we have shown that $\forall a \in A((f^{-1} \circ f)(a) = i_A(a))$, so $f^{-1} \circ f = i_A$.

Now let b be an arbitrary element of B. Then $a = f^{-1}(b)$ and $(b, a) \in f^{-1}$. Then $(a, b) \in f$ and $f(a) = b$. Thus

$$(f \circ f^{-1})(b) = f(f^{-1}(b)) = f(a) = b = i_B(b)$$

Since b was arbitrary, we have shown that $\forall b \in B((f \circ f^{-1})(b) = i_B(b))$, so $f \circ f^{-1} = i_B$. □

This theorem essentially says that if $f : A \to B$ and $f^{-1} : B \to A$, then each function undoes the effect of the other. For any $a \in A$, applying the function f gives us $f(a) \in B$. According to the theorem, $f^{-1}(f(a)) = (f^{-1} \circ f)(a) = i_A(a) = a$, thus undoing the effect of applying f. Similarly for any $b \in B$, applying f^{-1} we get $f^{-1}(b) \in A$, and we can undo the effect of applying f^{-1} by applying f, since $f(f^{-1}(b)) = b$.
(next page)

Theorem. *Suppose $f : A \to B$.*

1. *If there is a function $g : B \to A$ such that $g \circ f = i_A$ then f is one-to-one.*
2. *If there is a function $g : B \to A$ such that $f \circ g = i_B$ then f is onto.*

Proof.

1. Suppose $g : B \to A$ and $g \circ f = i_A$. Let a_1 and a_2 be arbitrary elements of A, and suppose that $f(a_1) = f(a_2)$. Applying g to both sides of this equation we get $g(f(a_1)) = g(f(a_2))$. But $g(f(a_1)) = (g \circ f)(a_1) = i_A(a_1) = a_1$; and similarly $g(f(a_2)) = a_2$. Thus we can conclude that $a_1 = a_2$, and therefore f is one-to-one.

2. Suppose that $g : B \to A$ and $f \circ g = i_B$. Let b be an arbitrary element of B. Since g is a function there exists $a \in A$ such that $a = g(b)$. Then, applying f to both sides we have $f(a) = f(g(b))$. Since $f(g(b)) = (f \circ g)(b) = i_B(b) = b$, then $f(a) = b$. Since b was arbitrary, this shows that f is onto. □

For the first part of the proof, being able to declare that $g(f(a_1)) = g(f(a_2))$ comes from the existential nature of the given $g : B \to A$, where we can declare some $a_3 = g(f(a_1)) = g(f(a_2))$ exists.

4.3.2 Circle of implications

We have proved the following theorems:

Theorem. *Suppose $f : A \to B$. If f is one-to-one and onto, then $f^{-1} : B \to A$.*

Theorem. *Suppose f is a function from A to B, and suppose that f^{-1} is a function from B to A. Then $f^{-1} \circ f = i_A$ and $f \circ f^{-1} = i_B$.*

Theorem. *Suppose $f : A \to B$.*

1. *If there is a function $g : B \to A$ such that $g \circ f = i_A$ then f is one-to-one.*
2. *If there is a function $g : B \to A$ such that $f \circ g = i_B$ then f is onto.*

Intuition
We have come full circle. In the first theorem we cound that if f is a one-to-one, onto function from A to B, then f^{-1} is a function from B to A. Then we showed that the composition of f with its inverse must be the identity function. Finally we found that when the composition of two functions is the identity, we are led back to the properties of one-to-one and onto. We combine all these into a theorem.

Theorem. *Suppose $f : A \to B$. Then the following statements are equivalent.*

1. *f is one-to-one and onto.*
2. *$f^{-1} : B \to A$.*
3. *There is a function $g : B \to A$ such that $g \circ f = i_A$ and $f \circ g = i_B$.*

Proof. $1 \to 2$. This is precisely what the first theorem says.
$2 \to 3$. Suppose $f^{-1} : B \to A$. Let $g = f^{-1}$ and apply the second theorem.
$3 \to 1$. Apply the third theorem. □

4.3.3 More implications

Intuition
Say some g and f are functions from $\mathbb{R}$ to $\mathbb{R}$, where $g \circ f = i_{\mathbb{R}}$ and $f \circ g = i_R$. It follows from the previous theorem that f must be one-to-one and onto, and f^{-1} must also be a function from $\mathbb{R}$ to $\mathbb{R}$. One may wonder what is f^{-1}; a logical guess would be that $f^{-1} = g$, but this doesn't actually follow from the theorems we've proven. The following theorem shows that f^{-1} must be equal to g.

Lemma. *Suppose A and B are sets,*

1. *For every relation R from A to B, $R \circ i_A = R$.*
2. *For every relation R from A to B, $i_B \circ R = R$.*

Proof.

1. Let R be some relation from A to B. Let (a, b) be an arbitrary element of $R \circ i_A$. Then we can declare some $a_0 \in A$ such that $(a, a_0) \in i_A$ and $(a_0, b) \in R$. Since $(a, a_0) \in i_A$, $a = a_0$. So $(a, b) = (a_0, b) \in R$. Since (a, b) was arbitrary, $R \circ i_A \subseteq R$.
   Now let (a, b) be an arbitrary element of R. Then $a \in A$, so $(a, a) \in i_A$. Since $(a, a) \in i_A$ and $(a, b) \in R$, $(a, b) \in R \circ i_A$. Since (a, b) was arbitrary, $R \subseteq R \circ i_A$ and $R \circ i_A = R$.

2. Let R be some relation from A to B. Let (a, b) be an arbitrary element of $i_B \circ R$. Then we can declare some $b_0 \in B$ such that $(a, b_0) \in R$ and $(b_0, b) \in i_B$. Since $(b_0, b) \in i_B$. $b = b_0$. so $(a, b) = (a, b_0) \in R$. Since (a, b) is arbitrary, $i_B \circ R \subseteq R$.
   Now let (a, b) be an arbitrary element of R. Then $b \in B$ and $(b, b) \in i_B$. Since $(a, b) \in R$ and $(b, b) \in i_B$, $(a, b) \in i_B \circ R$. Since (a, b) was arbitrary, $R \subseteq i_B \circ R$, so $i_B \circ R = R$. □

Theorem. *Suppose $f : A \to B$, $g : B \to A$, $g \circ f = i_A$, and $f \circ g = i_B$. Then $g = f^{-1}$.*

Proof. By the previous theorem, $f^{-1} : B \to A$. By a theorem from the first subsection, $f^{-1} \circ f = i_A$. Thus

$$\begin{aligned} g &= i_A \circ g && \text{(lemma)} \\ &= (f^{-1} \circ f) \circ g \\ &= f^{-1} \circ (f \circ g) && \text{(associativity, proven earlier)} \\ &= f^{-1} \circ i_B \\ &= f^{-1} && \text{(lemma)} \end{aligned}$$

□

4.3.4 With regard to equivalence relations

Recall the following theorems proven in the section on equivalence relations:

Theorem. *Suppose A is a set and $\mathcal{F}$ is a partition of A. Then there is an equivalence relation R on A such that $A/R = \mathcal{F}$.*

Theorem. *Suppose R is an equivalence relation on A. Let $\mathcal{F} = A/R$, and let S be the equivalence relation determined by $\mathcal{F}$. In other words, $S = \bigcup X \in \mathcal{F}(X \times X)$. Then $S = R$.*

Intuition
Suppose A is any set, let $\mathcal{E}$ be the set of all equivalence relations on A, and let $\mathcal{P}$ be the set of all partitions of A. Define a function $f : \mathcal{E} \to \mathcal{P}$ by the formula $f(R) = A/R$, and define another function $g : \mathcal{P} \to \mathcal{E}$ by the formula

$$\begin{aligned} g(\mathcal{F}) &= \text{the equivalence relation determined by } \mathcal{F}. \\ &= \bigcup_{X \in \mathcal{F}} (X \times X) \end{aligned}$$

See that the first theorem shows $f \circ g = i_{\mathcal{P}}$, and the second theorem shows that $g \circ f = i_{\mathcal{E}}$. Thus f is one-to-one and onto, and $g = f^{-1}$.

4.4 Closures

Definition. Suppose $f : A \to A$ and $C \subseteq A$. We wil say that C is closed under f if $\forall x \in C(f(x) \in C)$.

Examples
Say $A = \{a, b, c, d\}$ and $f = \{(a, c), (b, b), (c, d), (d, c)\}$. Then $f : A \to A$. Let $C_1 = \{a, c, d\}$ and $C_2 = \{a, b\}$.

The set C_1 is closed under f because $f(a) = f(d) = c \in C_1$ and $f(c) = d \in C_1$. However C_2 is not closed under f, because $a \in C_2$ but $f(a) = c \notin C_2$.

For another example, let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to R$ be defined by the formulas $f(x) = x + 1$ and $g(x) = x - 1$. Is $\mathbb{N}$ closed under f? Is it closed under g?

For every natural number n, $n+1$ is also a natural number, so $\mathbb{N}$ is closed under f. However $\mathbb{N}$ is not closed under g, because $0 \in N$ but $g(0) = -1 \notin \mathbb{N}$.

Intuition for closures
We saw in the second example that $\mathbb{N}$ is not closed under the function $g : \mathbb{R} \to \mathbb{R}$ defined by the formula $g(x) = x - 1$. Suppose we wanted to add elements to $\mathbb{N}$ to get a set that is closed under g. Since $0 \in \mathbb{N}$ we'd need to add $g(0) = -1$. But if we added -1 to the set, then it would also have to contain $g(-1) = -2$, and if we added -2 in we'd need to add $g(-2) = -3$. Continuing in this way it should be clear that we'd have to add all the negative integers to $\mathbb{N}$, giving us the set of all integers $\mathbb{Z}$. Notice that $\mathbb{Z}$ is closed under g, because for every integer n, $n - 1$ is also an integer; so we have succeeded in enlarging $\mathbb{N}$ to get a set closed under g.

When we enlarged $\mathbb{N}$ to $\mathbb{Z}$, the numbers we added—the negative integers—were numbers that *had* to be added if we wanted the resulting set to be closed under g. It follows that $\mathbb{Z}$ is the smallest set containing $\mathbb{N}$ is closed under g. Note that we use the word *smallest* here in the sense of the subset partial order

$$S = \{(X, Y) \in \mathscr{P}(A) \times \mathscr{P}(A) | X \subseteq Y\}$$

That is, if we let $\mathcal{F} = \{C \subseteq \mathbb{R} | \mathbb{N} \subseteq C \text{ and } C \text{ is closed under } g\}$, then $\mathbb{Z}$ is the smallest element of $\mathcal{F}$, meaning $\mathbb{Z}$ is an element of $\mathcal{F}$, and it is a subset of every element of $\mathcal{F}$. We say that $\mathbb{Z}$ is the *closure* of $\mathbb{N}$ under g.
(next page)

Definition. Suppose $f : A \to A$ and $B \subseteq A$. Then the *closure* of B under f is the smallest set $C \subseteq A$ such that $B \subseteq C$ and C is closed under f, if there is such a smallest set. In other words, a set $C \subseteq A$ is the closure of B under f if it has the following properties:

1. $B \subseteq C$.
2. C is closed under f.
3. For every set $D \subseteq A$, if $B \subseteq D$ and D is closed under f then $C \subseteq D$.

Intuition
Recall from the section on ordering relations that if a set has a smallest element, then it can have only one smallest element. Thus if a set B has a closure under a function f, then this closure must be unique, so it makes sense to call it *the* closure rather than *a* closure. Not all families of sets have smallest elements, so it is not immediately clear if sets always have closures under functions.

We want to prove that closures always exist. Suppose $f : A \to A$ and $B \subseteq A$. We know that the closure of B under f, if it exists, must be the smallest element of the family $\mathcal{F} = \{C \subseteq A | B \subseteq C \wedge C \text{ is closed under } f\}$. Recall from the section on ordering relations that the smallest element of a set is also always the greatest lower bound of the set, and that the g.l.b of any nonempty family of sets $\mathcal{F}$ is $\bigcap \mathcal{F}$. This gives us a starting point for the proof.

Lemma. *If $\mathcal{F}$ is a family of sets and $A \in \mathcal{F}$, then $\bigcap \mathcal{F} \subseteq A$.*

Proof. Suppose $A \in \mathcal{F}$. Let x be an arbitrary element in $\bigcap \mathcal{F}$. Then since $A \in \mathcal{F}$, $x \in A$. Since x was arbitrary, $\bigcap \mathcal{F} \subseteq A$. □

Theorem. *Suppose that $f : A \to A$ and $B \subseteq A$. Then B has a closure under f.*

Proof. Let $\mathcal{F} = \{C \subseteq A | B \subseteq C \text{ and } C \text{ is closed under } f\}$. First we show that $\mathcal{F} \neq \emptyset$. See that $B \subseteq A$, now let a be an arbitrary element of A, since f is a function we can always declare some $b \in A$ such that $f(a) = b \in A$. Since a is arbitrary, $\forall a \in A(f(a) \in A)$, so $A \in \mathcal{F}$ and $\mathcal{F} \neq \emptyset$. See that by the previous lemma, $\bigcap \mathcal{F} \subseteq A$. We wil show that $\bigcap \mathcal{F}$ is the closure under B by proving the three required properties.

First we show $B \subseteq \bigcap \mathcal{F}$. Suppose $x \in B$. Let D be an arbitrary element of $\mathcal{F}$. Then by the definition of $\mathcal{F}$, $B \subseteq D$ so $x \in D$. Since D was arbitrary, this shows that $\forall D \in \mathcal{F}(x \in D)$, so $x \in \bigcap \mathcal{F}$. Thus $B \subseteq \bigcap \mathcal{F}$.

Next we show $\bigcap \mathcal{F}$ is closed under f. Suppose $x \in \bigcap \mathcal{F}$ and let D be an arbitrary element of $\mathcal{F}$. Then since $x \in \bigcap \mathcal{F}$, $x \in D$. Now since $D \in \mathcal{F}$, D is closed under f, so $f(x) \in D$. Since D was arbitrary we can conclude that $f(x) \in \bigcap \mathcal{F}$. Thus $\bigcap \mathcal{F}$ is closed under f.

Finally we show that $\forall D((B \subseteq D \subseteq A \wedge (D \text{ is closed under } f)) \to (\bigcap F \subseteq D))$. Let D be arbitrary, suppose $B \subseteq D \subseteq A$ and D is closed under f. Then $D \in \mathcal{F}$, and by the previous lemma $\bigcap \mathcal{F} \subseteq D$. □

4.4.1 On the union and intersection of empty sets

Union

Theorem. *If $\mathcal{F} = \emptyset$, then the statement $x \in \bigcup \mathcal{F}$ will be false no matter what x is. It follows that $\bigcup \emptyset = \emptyset$.*

Proof. Suppose $\mathcal{F} = \emptyset$. Let x be arbitrary and suppose $x \in \bigcup F$. Then we would be able to declare some $A \in \mathcal{F}$ such that $x \in A$. But this would contradict the statement that $\mathcal{F} = \emptyset$. So $x \notin \bigcup \mathcal{F}$. Since x is arbitrary, $\forall x(x \notin \bigcup \mathcal{F})$. This means $\bigcup \mathcal{F} = \emptyset$. □

Intersection
Recall that the statement

$$\forall x \in \emptyset P(x)$$

is always *vacuously true*. We can show that if $\mathcal{F} = \emptyset$, then the statement $x \in \bigcap \mathcal{F}$ will be true no matter what x is, since

$$\forall A \in \emptyset(x \in A)$$

is vacuously true. In a context where it is clear what the universe of discourse U is, we might therefore want to say that $\bigcap \emptyset = U$. However this means that the notation $\bigcap \emptyset = U$ will mean different things depending on the universe of discourse.

Furthermore, when working with sets whose elements are sets, mathematicians often do not use a universe of discourse at all. To understand this, consider discussing sets whose elements are sets; it may seem natural to consider the universe of discourse to be the set of all sets. Assuming such a set, however, leads to a contradiction:

Suppose U were the set of all sets. Note that in particular U is a set, so we would have $U \in U$ (which, although unusual, isn't a contradiction). But this suggests that the sets in the universe U could be split into two categories: the unusual sets, like U itself, that are elements of themselves, and the more typical sets that are not. Let R be the set of sets in that second category, meaning $R = \{A \in U | A \notin A\}$. This means that for any set A in the universe U, A will be an element of R if $A \notin A$. So

$$\forall A \in U(A \in R \leftrightarrow A \notin A)$$

Now see that applying this last fact to the set R itself leads to a contradiction. This contradiction is known as Russell's paradox. Some mathematicians consider the notation $\bigcap \emptyset$ to therefore be meaningless. In this set of notes we will avoid this problem by using the notation $\bigcup \mathcal{F}$ only in contexts in which we can be sure that $\mathcal{F} \neq \emptyset$.

4.4.2 Functions of two variables and closures

Closed sets and closures also come up in the study of functions of more than one variable. If $f : A \times A \to A$, then f is called a *function of two variables*. An element of the domain of f would be an ordered pair (x, y), where $x, y \in A$. The result of applying f to this pair would be written $f((x, y))$, but it is customary to leave out one pair of parentheses and just write $f(x, y)$.

Definition. Suppose $f : A \times A \to A$ and $C \subseteq A$. We will say that C is *closed under* f if $\forall x \in C \forall y \in C(f(x, y) \in C)$.

As before we can define the closure of a set under a function of two variables to be the smallest closed set containing it, and we can prove that such closures always exist.

Definition. Suppose $f : A \times A \to A$ and $B \subseteq A$. Then the *closure of* B *under* f is the smallest set $C \subseteq A$ such that $B \subseteq C$ and C is closed under f, if there is such a smallest set. In other words, a set $C \subseteq A$ is the closure of B under f if it has the following properties:

1. $B \subseteq C$.
2. C is closed under f.
3. For every set $D \subseteq A$, if $B \subseteq D$ and D is closed under f then $C \subseteq D$.

Theorem. *Suppose that* $f : A \times A \to A$ *and* $B \subseteq A$. *Then* B *has a closure under* f.

Proof. Let $\mathcal{F} = \{C \subseteq A | B \subseteq C$ and C is closed under $f\}$. First we show that $\mathcal{F} \neq \emptyset$. We know $B \subseteq A$. Let x, y be arbitrary elements of A. Since f is a function, there exists some $f(x, y) \in A$. Therefore A is closed under f, and $A \in \mathcal{F}$. So $\mathcal{F} \neq \emptyset$. By the previous lemma, we can also conclude that $\bigcap \mathcal{F} \subseteq A$. We will show that $\bigcap \mathcal{F}$ is the closure under B by proving the three required properties.

First we show that $B \subseteq \bigcap \mathcal{F}$. Suppose $x \in B$. Let D be an arbitrary element of $\mathcal{F}$. Then by the definition of $\mathcal{F}$, $B \subseteq D$ so $x \in D$. Since D was arbitrary this shows that $\forall D \in \mathcal{F}(x \in D)$, so $x \in \bigcap \mathcal{F}$. Thus $B \subseteq \bigcap \mathcal{F}$.

Next we show that $\bigcap \mathcal{F}$ is closed under f. Let x and y be arbitrary elements of $\bigcap \mathcal{F}$. Let D be an arbitrary element of $\mathcal{F}$. Then since $x, y \in \bigcap \mathcal{F}$, $x, y \in D$. By the definition of $\mathcal{F}$, D is closed under f, so $f(x, y) \in D$. Since D was arbitrary we can conclude that $f(x, y) \in \bigcap \mathcal{F}$. Thus $\bigcap \mathcal{F}$ is closed under f.

Finally we show that $\forall D((B \subseteq D \subseteq A \wedge (D$ is closed under $f)) \to \bigcup \mathcal{F} \subseteq D)$. Let D be arbitrary, suppose $B \subseteq D \subseteq A$ and D is closed under f. Then $D \in \mathcal{F}$, and by the previous lemma $\bigcup \mathcal{F} \subseteq D$. □

Convention
A function from $A \times A$ to A could be thought of as an operation that can be applied to a pair of objects $(x, y) \in A \times A$ to produce another element of A. Often an operation to be performed on a pair of mathematical objects (x, y) is represented by a symbol written between x and y; for instance $x + y$ or $x \cup y$. Imitating this notation, when defining a function from $A \times A$ to A, it is sometimes represented with a symbol between x and y rather than a letter in front of the pair. When written in this way, it is usually called a *binary relation* on A.

4.5 Images and Inverse Images

Suppose $f : A \to B$. We have already seen that we can think of f matching each element of A with exactly one element of B. f can also be thought of as matching *subsets* of A with subsets of B and vice-versa.

Definition. Suppose $f : A \to B$ and $X \subseteq A$. Then the *image* of X under f is the set $f(X)$ defined as

$$\begin{aligned} f(X) &= \{f(x) | x \in X\} \\ &= \{b \in B | \exists x \in X (f(x) = b)\} \end{aligned}$$

(Note that the image of the whole domain A under f is $\{f(a) | a \in A\}$, which is the same as the range of f.)

If $Y \subseteq B$, then the *inverse image* Y under f is the set $f^{-1}(Y)$ defined as

$$f^{-1}(Y) = \{a \in A | f(a) \in Y\}$$

Note that the function f may fail to be one-to-one or onto, and as a result f^{-1} may not be a function from B to A, and for $y \in B$, the notation "$f^{-1}(y)$" may be meaningless. Notice that the definition still assigns a meaning to the notation $f^{-1}(Y)$ for $Y \subseteq B$; it doesn't treat f^{-1} as a function, and refers only to the results of applying f to elements of A. Not the results of applying f^{-1} to elements of B.

Theorem. *Suppose $f : A \to B$, and W and X are subsets of A. Then $f(W \cap X) \subseteq f(W) \cap f(X)$. Furthermore, if f is one-to-one, then $f(W \cap X) = f(W) \cap f(X)$.*

Proof. Suppose first that y is an arbitrary element of $f(W \cap X)$. By the definition of $f(W \cap X)$, this means that $y = f(x)$ for some $x \in W \cap X$. Since $x \in W \cap X$, it follows that $x \in W$ and $x \in X$. But now we have $y = f(x)$ and $x \in W$, so we can conclude that $y \in f(W)$. Similarly, since $y = f(x)$ and $x \in X$, it follows that $y \in f(X)$. Thus $y \in f(W) \cap f(X)$. This completes the first half of the proof.

Now suppose that $y \in f(W) \cap f(X)$ and that f is one-to-one. Then $y \in f(W)$, so there is some $w \in W$ such that $f(w) = y$, similarly since $y \in f(X)$, there is some $x \in X$ such that $y = f(x)$. Since $f(w) = y = f(x)$ and f is one-to-one, $w = x$. So $w \in W \cap X$ and $f(w) = y$, meaning $y \in f(W \cap X)$. □

Chapter 5

Mathematical Induction

5.1 Proof by Mathematical Induction

Suppose we want to prove that every natural number $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$ has some property P. Of course there are infinitely many numbers in this list, so one can't check one-by-one that they all have property P.

The key idea behind mathematical induction is that to list all the natural numbers all one has to do is start with 0 and repeatedly add 1. Thus one can show that every natural number has the property P by showing 0 has the property P, and that whenever one adds 1 to a number that has the property P, the resulting number also has the property P. This would gaurantee that, as one goes through the list of all natural numbers, starting with 0 and repeatedly adding 1, every number one encounters must have property P:

To prove a goal of the form $\forall n \in \mathbb{N} P(n)$:
First prove $P(0)$, and then prove $\forall n \in \mathbb{N}(P(n) \to P(n+1))$. The first of these proofs is sometimes called the *base case* and the second the *induction step*.

Form of final proof:
Base case: [Proof of $P(0)$ goes here.]
Induction step: [Proof of $\forall n \in \mathbb{N}(P(n) \to P(n+1))$ goes here.]

5.1.1 Examples

Example 1
The following suggests a pattern:

$$2^0 = 1 = 2^1 - 1$$
$$2^0 + 2^1 = 1 + 2 = 3 = 2^2 - 1$$
$$2^0 + 2^1 + 2^2 = 1 + 2 + 4 = 7 = 2^3 - 1$$
$$2^0 + 2^1 + 2^2 + 2^3 = 1 + 2 + 4 + 8 = 15 = 2^4 - 1$$

Theorem. *For every natural number n, $2^0 + 2^1 + \cdots + 2^n = 2^{n+1} - 1$.*

Scratch work
Our goal is to prove the statement $\forall n \in \mathbb{N} P(n)$, where $P(n)$ is the statement $2^0 + 2^1 + \cdots + 2^n = 2^{n+1} - 1$. According to our strategy, we can do this by proving two other statements, $P(0)$ and $\forall n \in \mathbb{N}(P(n) \to P(n+1))$.

Plugging in 0 for n, we see that $P(0)$ is simply the statement $2^0 = 2^1 - 1$. This part is easy, just verifying that both sides are equal to 1. Often the base case is easy, and the induction step more complex.

For the induction step, we must prove $\forall n \in \mathbb{N}(P(n) \to P(n+1))$. As expected, we do this by letting n be an arbitrary natural number, assuming that $P(n)$ is true, and then proving that $P(n+1)$ is true. In this case, we'll let n be an arbitrary natural number, assume that $2^0 + 2^1 + \cdots + 2^n = 2^{n+1} - 1$, and then prove $2^0 + 2^1 + \cdots + 2^{n+1} = 2^{n+2} - 1$:

Givens	*Goal*
$n \in \mathbb{N}$	$2^0 + 2^1 + \cdots + 2^{n+1} = 2^{n+2} - 1$
$2^0 + 2^1 + \cdots + 2^n = 2^{n+1} - 1$	

The key to this proof is to recognise that the left side of the equation in the goal is exactly the same as the left side of the given, but with the extra term 2^{n+1} added on. So we try adding 2^{n+1} to both sides of the second given:

$$(2^0 + 2^1 + \cdots + 2^n) + 2^{n+1} = (2^{n+1} - 1) + 2^{n+1}$$

or in other words,

$$2^0 + 2^1 + \cdots + 2^{n+1} = 2^{n+2} - 1$$

which is our goal.
(next page)

Cont.

Theorem. *For every natural number n, $2^0 + 2^1 + \cdots + 2^n = 2^{n+1} - 1$.*

Proof. We use mathematical induction.

Base case: Setting $n = 0$, we get $2^0 = 1 = 2^1 - 1$ as required.

Induction step: Let n be an arbitrary natural number and suppose that $2^0 + 2^1 + \cdots + 2^n = 2^{n+1} - 1$. Then

$$\begin{aligned} 2^0 + 2^1 + \cdots + 2^{n+1} &= (2^0 + 2^1 + \cdots + 2^n) + 2^{n+1} \\ &= (2^{n+1} - 1) + 2^{n+1} \\ &= 2 \cdot 2^{n+1} - 1 \\ &= 2^{n+2} - 1 \end{aligned}$$

□

The hardest proof by mathematical induction is usually the induction step $\forall n \in \mathbb{N}(P(n) \to P(n+1))$. Generally we would let n be an arbitrary natural number, assume $P(n)$ is true, and then prove $P(n+1)$ is true. The assumption that $P(n)$ is true is sometimes called the *inductive hypothesis*, and the key to the proof is usually to work out some relationship between the inductive hypothesis $P(n)$ and the goal $P(n+1)$.

Example 2

Theorem. *For every natural number n, $3|(n^3 - n)$.*

Scratch work

The base case is easy. For the induction step, we let n be an arbitrary natural number and assume that $3|(n^3-n)$, and we must prove that $3|((n+1)^3-(n+1))$. Filling in the definition of divides, we have

Givens	*Goal*
$n \in \mathbb{N}$	$\exists j \in \mathbb{Z}(3j = (n+1)^3 - (n+1)$
$\exists k \in \mathbb{Z}(3k = n^3 - n)$	

The second given is the inductive hypothesis. We use it and let k stand for a particular integer such that $3k = n^3 - n$. To complete the proof we need to find some j satisfying the goal:

$$\begin{aligned} (n+1)^3 - (n+1) &= n^3 + 3n^2 + 3n + 1 - n - 1 \\ &= (n^3 - n) + 3n^2 + 3n \\ &= 3k + 3n^2 + 3n \\ &= 3(k + n^2 + n) \end{aligned}$$

It should be clear now that we can complete the proof by letting $j = k + n^2 + n$.
(next page)

Theorem. *For every natural number n, $3|(n^3 - n)$.*

Proof. We use mathematical induction.

Base case: If $n = 0$, then $n^3 - n = 0 = 3 \cdot 0$, so $3|(n^3 - n)$. Induction step: Let n be an arbitrary natural number and suppose $3|(n^3 - n)$. Then we can choose an integer k such that $3k = n^3 - n$. Thus

$$\begin{aligned}(n+1)^3 - (n+1) &= n^3 + 3n^2 + 3n + 1 - n - 1 \\ &= (n^3 - n) + 3n^2 + 3n \\ &= 3k + 3n^2 + 3n \\ &= 3(k + n^2 + n)\end{aligned}$$

Therefore $3|((n+1)^3 - (n+1))$, as required. □

Example 3
Consider the following values

n	n^2	2^n	Which is larger?
0	0	1	2^n
1	1	2	2^n
2	4	4	tie
3	9	8	n^2
4	16	16	tie
5	25	32	2^n
6	36	64	2^n

It seems that for $n \geq 5$, 2^n will be greater than n^2.

Theorem. *For every natural number $n \geq 5$, $2^n > n^2$.*

Scratch work
The base case $n = 5$ has already been checked in the table. For the induction step, we let $n \geq 5$ be arbitrary, assume $2^n > n^2$, and try to prove that $2^{n+1} > (n+1)^2$. How can we relate the inductive hypothesis to the goal? Multiplying both sides of the inductive hypothesis by 2 we get $2^{n+1} > 2n^2$. We want to show $2^{n+1} > (n+1)^2$, so perhaps we can show $2n^2 \geq (n+1)^2$.

Expanding the inequality we want to show, we want to show $2n^2 \geq n^2 + 2n + 1$. Starting from $2n^2$ and the other given $n \geq 5$, we have $2n^2 = n^2 + n^2 \geq n^2 + 5n = n^2 + 2n + 3n > n^2 + 2n + 1$.
(next page)

Theorem. *For every natural number $n \geq 5$, $2^n > n^2$.*

Proof. By mathematical induction.

Base case: When $n = 5$ we have $2^n = 32 > 25 = n^2$.

Induction step: Let $n \geq 5$ be arbitrary, and suppose that $2^n > n^2$. Then

$$\begin{aligned}
2^{n+1} &= 2 \cdot 2^n \\
&> 2n^2 && \text{(inductive hypothesis)} \\
&= n^2 + n^2 \\
&\geq n^2 + 5n && \text{(since } n \geq 5\text{)} \\
&= n^2 + 2n + 3n \\
&> n^2 + 2n + 1 = (n+1)^2
\end{aligned}$$

Thus $2^{n+1} > (n+1)^2$, as required. □

5.1.2 More examples

Example 1

Theorem. *Suppose R is a partial order on a set A. Then every finite, nonempty set $B \subseteq A$ has an R-mininal element.*

Scratch work
It is not immediately obvious that mathematical induction could be used here—the goal doesn't seem to have the form $\forall n \in \mathbb{N} P(n)$. Natural numbers enter into the problem when we recognise that to say that B is finite and nonempty means that it has n elements, for some $n \in \mathbb{N}$, $n \geq 1$ (this will be elaborated on). The goal here is $\forall n \geq 1 \forall B \subseteq A(B$ has n elements $\rightarrow B$ has a minimal element).

In the base case we have $n = 1$, and we must prove that if B has one element then it has a minimal element. It is easy to check that in this case that the one element of B must be minimal.

For the induction step we want to show

$$\begin{aligned}\forall n \geq 1 (&\forall B \subseteq A(B \text{ has } n \text{ elements} \rightarrow B \text{ has a minimal element}) \\ &\rightarrow \forall B \subseteq A(B \text{ has } n+1 \text{ elements} \rightarrow B \text{ has a minimal element}))\end{aligned}$$

So we let $n \geq 1$ be arbitrary, assume that $\forall B \subseteq A(B$ has n elements $\rightarrow$ B has a minimal element), and try to prove that $\forall B \subseteq A(B$ has $n+1$ elements $\rightarrow B$ has a minimal element). Guided by the form of the goal, we let B be an arbitary subset of A, assume B has $n+1$ elements, and try to prove that B has a minimal element.

How can we use the inductive hypothesis to reach our goal? It tells us that if we had a subset of A with n elements, then it would have a minimal element. To apply it we need a subset of A with n elements. Our arbitrary set B is a subset of A, and we have assumed that it has $n+1$ elements. Thus a simple way to produce a subset of A with n elements would be to remove one element from B. It is not clear where this reasoning would lead, but it seems to be the simplest way to make use of the inductive hypothesis.
(next page)

Scratch work continued
Let b be any element of B, and let $B' = B\backslash\{b\}$. Then B' is a subset of A with n elements, so by the inductive hypothesis, B' has a minimal element. This is an existential statement, so we immediately introduce a new variable c to stand for a minimal element of B'.

Our goal is to prove that B has a minimal element, which is also an existential statement, so we should try to come up with a minimal element of B. We only know about two elements of B at this point, b and c, so we should probably try to prove that one of these is a minimal element of B. Which one? It may depend on whether one of them is smaller than the other in R, suggesting proof by cases. In our proof we use the cases bRc and $\neg bRc$, where in the first case we show that b is a minimal element of B, and in the second that c is a minimal element of B. In both cases saying something is a minimal element is a negative statement, so we use proof by contradiction.

Theorem. *Suppose R is a partial order on a set A. Then every finite, nonempty set $B \subseteq A$ has an R-mininal element.*

Proof. We will show by induction that for every natural number $n \geq 1$, every subset of A with n elements has a minimal element.

Base case: $n = 1$. Suppose $B \subseteq A$ and B has one element. Then $B = \{b\}$ for some $b \in A$. Clearly $\neg\exists x \in B(x \neq b)$, so clearly $\neg\exists x \in B(xRb \wedge x \neq b)$. Thus b is minimal. (This is a bit opaque, a better approach may be to suppose b is not R minimal, which would lead to a contradiction.)

Induction step: Suppose $n \geq 1$, and suppose that every subset of A with n elements has a minimal element. Now let B be an arbitrary subset of A with $n + 1$ elements. Let b be any element of B, and let $B' = B\backslash\{b\}$, a subset of A with n elements. By the inductive hypothesis, we can choose a minimal element $c \in B'$.

Case 1: bRc. We will show that b is a minimal element of B. To justify this, suppose it isn't. Then we can choose some $x \in B$ such that xRb and $x \neq b$. Since $x \neq b$, $x \in B'$. Also since xRb and bRc, by transitivity of R it follows that xRc. Since c is a minimal element of B', we must have $x = c$. But then since xRb we have cRb, and we also know bRc, so by antisymmetry of B it follows that $b = c$. But this is clearly impossible since $c \in B' = B\backslash\{b\}$. Thus, b must be a minimal element of B.

Case 2: $\neg bRc$. We will show that c is a minimal element of B. To justify this, suppose it isn't. Then we can choose some $x \in B$ such that xRc and $x \neq c$. Since c is a minimal element of B', we can't have $x \in B'$, so since $x \in B$ the only other possibility is $x = b$. Then since xRc we must have bRc, which contradicts the assumption that $\neg bRc$. Thus c is a minimal element of B. (or show by contradiction that $x \neq b$ so $x \in B'$, but since xRc and $x \neq c$ this would contradict the fact that c is minimal). □

(next page)

Example 2

Theorem. *Suppose A is a finite set and R is a partial order on A. Then there is a total order T on A such that $R \subseteq T$.*

Scratch work
The idea here is to prove by induction that $\forall n \in \mathbb{N} \forall A \forall R[(A \text{ has } n \text{ elements and } R \text{ is a partial order on } A) \to \exists T(T \text{ is a total order on } A \text{ and } R \subseteq T)]$.

The base case is fairly straightforward. The induction step is

$$\begin{aligned} \forall n \in \mathbb{N}[\forall A \forall R[(A \text{ has } n \text{ elements and } R \text{ is a partial order on } A) \\ \to \exists T(T \text{ is a total order on } A \text{ and } R \subseteq T)] \\ \to \forall A \forall R[(A \text{ has } n+1 \text{ elements and } R \text{ is a partial order on } A) \\ \to \exists T(T \text{ is a total order on } A \text{ and } R \subseteq T)]] \end{aligned}$$

R is a partial order on a set A with $n+1$ elements, we remove one element, call it a, from A, and apply the inductive hypothesis to the remaining set $A' = A \backslash \{a\}$, giving us total order T' on A'. To complete the proof we must somehow come up with a total order T on A such that $R \subseteq T$.

The relation T' tells us how to compare two elements of A', but it doesn't tell us how to compare a to elements of A'. This is what we must decide in order to define T; the main diffculty in this step is making this decision in a way such that we end up with $R \subseteq T$. This is done by choosing a to be an R-minimal element of A, and then we define T by making a smaller in T than every element of A'. We us the theorem in the previous example, taking $B = A$, to guarantee that A has an R-minimal element.
(next page)

Theorem. *Suppose A is a finite set and R is a partial order on A. Then there is a total order T on A such that $R \subseteq T$.*

Proof. We will show by induction on n that every partial order on a set with n elements can be extended to a total order.

Base case: $n = 0$. Suppose R is a partial order on A and A has 0 elements. Then clearly $A = R = \emptyset$. Then all the properties making $\emptyset$ a total order on A hold vacuously. So we are done.

Induction step: Let n be an arbitrary natural number, and suppose that every partial order on a set with n elements can be extended to a total order. Now suppose that A has $n + 1$ elements and R is a partial order on A. By the previous theorem, there must be some $a \in A$ such that a is an R-minimal element of A. Let $A' = A \backslash \{a\}$ and let $R' = R \cap (A' \times A')$.

We will now show that $R' = R \cap (A' \times A')$ is a partial order on A'. We must prove that R' is reflexive, transitive, and antisymmetric on A'. For the first, suppose $x \in A'$. Since R is reflexive on A and $x \in A$, $(x, x) \in R \cap (A' \times A') = R'$. So R' is reflexive.

Next suppose that $(x, y) \in R'$ and $(y, z) \in R'$. Then $(x, y) \in R$, $(y, z) \in R$, and $x, y, z \in A'$. Since R is transitive, $(x, z) \in R$, so $(x, z) \in R \cap (A' \times A') = R'$. Therefore R' is transitive.

Finally suppose that $(x, y) \in R'$ and $(y, x) \in R'$. Then $(x, y) \in R$ and $(y, x) \in R$, so since R is antisymmetric, $x = y$. Thus R' is antisymmetric. R' is a partial order on A'.

Now by the inductive hypothesis, we can let T' be a total order on A' such that $R' \subseteq T'$. Now let $T = T' \cup (\{a\} \times A)$. We will now show that T is a total order on A and $R \subseteq T$.

To see that T is reflexive, suppose $x \in A$. If $x = a$, then $(x, x) = (a, a) \in (\{a\} \times A) \subseteq T$. Now if $x \neq a$. Then $x \in A'$, so since R' is reflexive, $(x, x) \in R' \subseteq T' \subseteq T$.

For transitivity, suppose that $(x, y) \in T$ and $(y, z) \in T$. If $x = a$ then $(x, z) = (a, z) \in \{a\} \times A \subseteq T$. Now suppose $x \neq a$. Then $(x, y) \notin \{a\} \times A$, so since $(x, y) \in T = T' \cup (\{a\} \times A)$, we must have $(x, y) \in T'$. But $T' \subseteq A' \times A'$, so $y \in A'$ and therefore $y \neq a$. Similar reasoning now shows that $(y, z) \in T'$. Since T' is transitive, it follows that $(x, z) \in T' \subseteq T$.

For antisymmetry, suppose $(x, y) \in T$ and $(y, x) \in T$. If $x \in a$ then $(y, x) \notin T'$, so $(y, x) \in \{a\} \times A$ and therefore $y = a = x$. Similarly if $y = a$ then $x = y$. Now suppose $x \neq a$ and $y \neq a$. Then as in the proof of transitivity it follows that $(x, y) \in T'$ and $(y, x) \in T'$, so by antisymmetry of T', $x = y$.

We now know that T is a partial order. To see that it is total, suppose $x \in A$ and $y \in A$. If $x = a$ then $(x, y) \in \{a\} \times A \subseteq T$. Similarly if $y = a$ then $(y, x) \in T$. Now suppose that $x \neq a$ and $y \neq a$. Then $x \in A'$ and $y \in A'$, so since T' is a total order, either $(x, y) \in T' \subseteq T$ or $(y, x) \in T' \subseteq T$.

Finally, to see that $R \subseteq T$, suppose that $(x, y) \in R$. If $x = a$ then $(x, y) \in \{a\} \times A \subseteq T$. Now suppose $x \neq a$. If $y = a$ then the fact that $(x, y) \in R$ would contradict the R-minimality of a. Therefore $y \neq a$. But then $(x, y) \in R \cap (A' \times A') = R' \subseteq T' \subseteq T$. □

5.1.3 Even more examples

Theorem. *For all $n \geq 3$, if n distinct points on a circle are connected in consecutive order with straight lines, then the interior angles of the resulting polygon add up to $(n-2)180°$. For example, here is $n = 4$:*

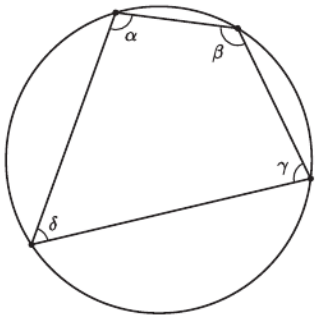
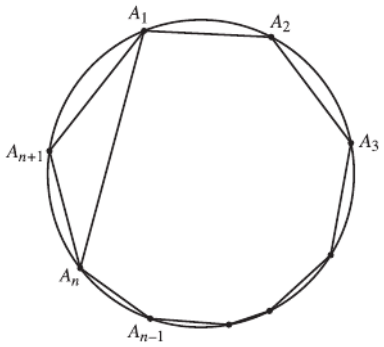


Figure 6.1. $\alpha + \beta + \gamma + \delta = (4-2)180^\circ = 360^\circ$.

Proof. We use induction on n.

Base case: Suppose $n = 3$. Then the polygon is a triangle, and it is well known that the interior angles of a triangle add up to 180°.

Induction step: Let n be an arbitrary natural number, $n \geq 3$, and assume that the statement is true for n. Now consider the polygon P formed by connecting some $n+1$ distinct points $A_1, A_2, \ldots, A_{n+1}$ on a circle. If we skip the last point A_{n+1} then we get a polygon P' with only n vertices, and by the inductive hypothesis the interior angles of this polygon add up to $(n-2)180°$.



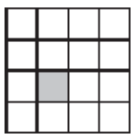
But now the sum of the interior angles of P is equal to the sum of the interior angles of P' plus the sum of the interior angles of the triangle $A_1A_nA_{n+1}$. Since the sum of the interior angles of the triangle is 180°, we can conclude that the sum of the interior angles of P is

$$(n-2)180^\circ + 180^\circ = ((n+1)-2)180^\circ$$

As required. □

(next page)

Theorem. *For any positive integer n, a $2^n \times 2^n$ square grid with any one square removed can be covered with L-shaped tiles that look like* ⊔. *For instance here is the $n = 2$ case*

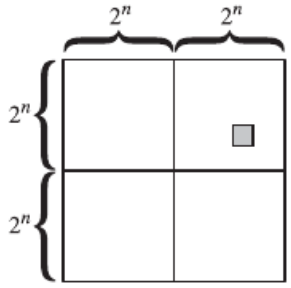


(a) 4×4 grid with one square removed.

(b) Grid covered with L-shaped tiles.

Scratch work
We'll use induction. Since we're only interested in positive n, the base case will be $n = 1$. In this case we have a 2×2 grid with one square removed. This can clearly be covered with one L-shaped tile.

For the induction step, we let n be an arbitrary positive integer and assume that a $2^n \times 2^n$ grid with any one square removed can be covered with L-shaped tiles. Now suppose we have a $2^{n+1} \times 2^{n+1}$ with any one square removed. To use our inductive hypothesis we must somehow relate this to the $2^n \times 2^n$ grid. Since $2^{n+1} = 2^n \cdot 2$, the $2^{n+1} \times 2^{n+1}$ grid is twice as wide and twice as high as the $2^n \times 2^n$ grid. In other words, by dividing the $2^{n+1} \times 2^{n+1}$ grid in half both vertically and horizontally, we can split it into four $2^n \times 2^n$ "subgrids":



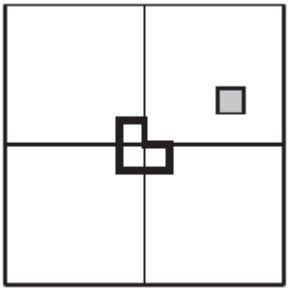
The one square that has been removed will be in one of the subgrids, as shown. The inductive hypothesis tells us that it is possible to cover the upper right subgrid. But what about the other three?
(next page)

Theorem. *For any positive integer n, a $2^n \times 2^n$ square grid with any one square removed can be covered with L-shaped tiles.*

Proof. We use induction on n.

Base case: Suppose $n = 1$. Then the grid is a 2×2 grid with one square removed, which can clearly be covered with one L-shaped tile.

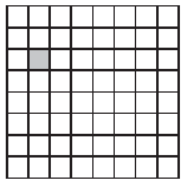
Induction step: Let n be an arbitrary positive integer and suppose that a $2^n \times 2^n$ grid with any one square removed can be covered with L-shaped tiles. Now consider a $2^{n+1} \times 2^{n+1}$ grid with one square removed. Cut the grid in half both vertically and horizontally, splitting it into four $2^n \times 2^n$ subgrids. The one square that has been removed comes from one of these subgrids, so by the inductive hypothesis the rest of that subgrid can be covered with L-shaped tiles. To cover the other three subgrids, first place one L-shaped tile in the center so that it covers one square from each of the three remaining subgrids:



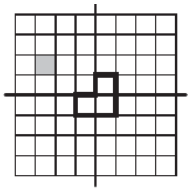
The area remaining to be covered now contains every square except one in each of the subgrids, so by applying the inductive hypothesis to each subgrid we can see that this area can be covered with tiles. □

(next page)

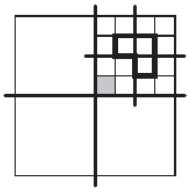
Intuition
It is interesting to note that the previous proof can be used to figure out how to place tiles on a particular grid. For example consider the 8×8 case:



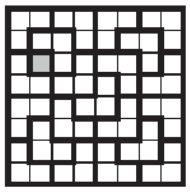
According to the proof, the first step is to split the grid into four 4×4 subgrids and place one tile in the center, covering one square from each subgrid except the upper left:



The area remaining now consists of four 4×4 subgrids each with one square removed. Now to cover the remaining four subgrids we use the same method; we divide each subgrid again into four 2×2 subgrids and put one tile in the middle. Now we've reached the base case for each subgrid, where each can be covered with one tile. For instance the upper right subgrid looks like



The remaining three 4×4 subgrids are completed by a similar procedure:



This is an example of a *recursive* procedure, where we repeat algorithmic steps until we reach a base case.

5.2 Recursion

Intuition
The principle behind induction can be used when defining functions. Recall that we usually defined a function f by specifying how to compute $f(n)$ for any n in the domain of f; if the domain of f is the set of all natural numbers, an alternative method to define f would be to say what $f(0)$ is and then, for any natural number n, say how we could compute $f(n+1)$ if we already knew the value of $f(n)$. Such a definition would enable us to run through all natural numbers in order computing the image of each one under f.

For example, we might use the following equations to define a function f with domain $\mathbb{N}$:

$$f(0) = 1;$$
$$\text{for every } n \in \mathbb{N}, f(n+1) = (n+1) \cdot f(n)$$

The second equation tells us how to compute $f(n+1)$ given $f(n)$. Although we cannot use this equation to tell us directly what the image of any number is under f, we *can* use it to run through all the natural numbers in order and compute their images.

In this case we start with $f(0)$, which we know from the first equation is equal to 1. Plugging in $n = 0$ in the second equation, we see that $f(1) = 1 \cdot f(0) = 1 \cdot 1 = 1$, so we've determined the value of $f(1)$; now that we know $f(1) = 1$ we find that $f(2) = 2 \cdot f(1) = 2 \cdot 1 = 2$. Similarly, setting $n = 2$ in the second equation we get $f(3) = 3 \cdot f(2) = 3 \cdot 2 = 6$. Continuing this way we can compute $f(n)$ for any natural number n.

The two equations give us a rule that determines a unique value $f(n)$ for each natural number n, so they define a function f with domain $\mathbb{N}$. Definitions of thie kind are called *recursive* definitions.
(next page)

Cont.
Sometimes we'll work backwards using a recursive definition to evaluate a function. For instance for the function from the previous page:

$$f(0) = 1;$$
$$\text{for every } n \in \mathbb{N}, f(n+1) = (n+1) \cdot f(n)$$

Suppose we want to compute $f(6)$. We then have the following calculation:

$$\begin{aligned} f(6) &= 6 \cdot f(5) \\ &= 6 \cdot 5 \cdot f(4) \\ &= 6 \cdot 5 \cdot 4 \cdot f(3) \\ &= 6 \cdot 5 \cdot 4 \cdot 3 \cdot f(2) \\ &= 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot f(1) \\ &= 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 \cdot f(0) \\ &= 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 \cdot 1 \\ &= 720 \end{aligned}$$

One might recognise this function, where for any positive integer $f(n) = n \cdot (n-1) \cdot (n-2) \cdots 1$, and $f(0) = 1$. The number $f(n)$ is called n *factorial*, and is denoted $n!$; for example $6! = 720$. Often, if a function can be written as a formula with an ellipsis $(\ldots)$ in it, then the ellipsis can be avoided by using a recursive definition for the function. Such a definition is usually easier to work with.

5.2.1 Examples

Many familiar functions are most easily defined using recursive definitions.

For a first example, for any number a, we could determine a^n with the following recursive definition:

$$\begin{aligned}&a^0 = 1;\\&\text{for every } n \in \mathbb{N}, a^{n+1} = a^n \cdot a\end{aligned}$$

Where using this definition we would compute a^4 like:

$$\begin{aligned}a^4 &= a^3 \cdot a\\&= a^2 \cdot a \cdot a\\&= a^1 \cdot a \cdot a \cdot a\\&= a^0 \cdot a \cdot a \cdot a \cdot a\\&= 1 \cdot a \cdot a \cdot a \cdot a\end{aligned}$$

For another example, consider the sum $2^0 + 2^1 + 2^2 + \cdots + 2^n$. The ellipsis suggests that we might be able to use a recursive definition. If we let $f(n) = 2^0 + 2^1 + 2^2 + \cdots + 2^n$, then notice that for every $n \in \mathbb{N}$, $f(n+1) = 2^0 + 2^1 + 2^2 + \cdots + 2^n + 2^{n+1} = f(n) + 2^{n+1}$. Thus we could define f recursively as follows:

$$\begin{aligned}&f(0) = 2^0 = 1;\\&\text{for every } n \in \mathbb{N}, f(n+1) = f(n) + 2^{n+1}\end{aligned}$$

As a check that this definition is right, consider the $n = 3$ case:

$$\begin{aligned}f(3) &= f(2) + 2^3\\&= f(1) + 2^2 + 2^3\\&= f(0) + 2^1 + 2^2 + 2^3\\&= 2^0 + 2^1 + 2^2 + 2^3 = 15\end{aligned}$$

(next page)

Cont.
If $a_0, a_1, \ldots, a_n$ is a list of numbers, then the sum of these numbers is written $\sum_{i=0}^{n} a_i$. This is read "the sum as i goes from 0 to n of a_i". For example:

$$\sum_{i=0}^{n} 2^i = 2^0 + 2^1 + 2^2 + \cdots + 2^n$$

More generaly, if $n \geq m$, then

$$\sum_{i=m}^{n} a_i = a_m + a_{m+1} + a_{m+2} + \cdots + a_n$$

For example,

$$\sum_{i=3}^{6} i^2 = 3^2 + 4^2 + 5^2 + 6^2 = 86$$

The letter i in these formulas is a bound variable and therefore can be replaced by a new variable without changing the meaning of the formula. Now we try to give a recursive definition for this notation. We let m be an arbitrary integer, and then proceed by recursion on n. Just as the base case for an induction proof need not be $n = 0$, the base case for a recursive definition can also be a number other than 0. In this case we are only interested in $n \geq m$, so we take $n = m$ as the base for our recursion:

$$\sum_{i=m}^{m} a_i = a_m;$$

$$\text{for every } n \geq m, \sum_{i=m}^{n+1} a_i = \sum_{i=m}^{n} a_i + a_{n+1}$$

Trying this definition out on the previous example we get

$$\begin{aligned}
\sum_{i=3}^{6} i^2 &= \sum_{i=3}^{5} i^2 + 6^2 \\
&= \sum_{i=3}^{4} i^2 + 5^2 + 6^2 \\
&= \sum_{i=3}^{3} i^2 + 4^2 + 5^2 + 6^2 \\
&= 3^2 + 4^2 + 5^2 + 6^2 = 86
\end{aligned}$$

Just as we wanted.

5.2.2 More examples

Theorem. *For every* $n \geq 4, n! > 2^n$.

Scratch work
Because the problem involves factorial and exponentiation, both of which are defined recursively, induction seems apt here. Recall the recursive definitions; first the factorial

$$0! = f(0) = 1;$$
$$\text{for every } n \in \mathbb{N}, (n+1)! = f(n+1) = (n+1) \cdot f(n) = (n+1) \cdot n!$$

and then exponentiation

$$a^0 = 1;$$
$$\text{for every } n \in \mathbb{N}, a^{n+1} = a^n \cdot a$$

The base case will be $n = 4$. For the induction step, our inductive hypothesis will be $n! > 2^n$, and we must prove that $(n+1)! > 2^{n+1}$. Naturally, the way to relate the inductive hypothesis to the goal is to use the recursive definitions of factorial and exponentiation.

Proof. By mathematical induction.
Base case: When $n = 4$ we have $n! = 24 > 16 = 2^n$.
Induction step: Let $n \geq 4$ be arbitrary and suppose that $n! > 2^n$. Then

$$\begin{aligned}(n+1)! &= (n+1) \cdot n! \\ &> (n+1) \cdot 2^n && \text{(inductive hypothesis)} \\ &> 2 \cdot 2^n = 2^{n+1}\end{aligned}$$

Which is what we wanted. □

(next page)

Theorem. *Prove that for every real number a and all natural numbers m and n, $a^{m+n} = a^m \cdot a^n$.*

Scratch work
There are three universal quantifiers here. We'll treat the first two differently from the third. We'll let a and m be arbitrary and then use mathematical induction to prove that $\forall n \in \mathbb{N}(a^{m+n} = a^m \cdot a^n)$. The key part of this proof is the formula $a^{n+1} = a^n \cdot a$ from the recursive definition of exponentiation.

Proof. Let a be an arbitrary real number and m an arbitrary natural number. We now proceed by induction on n.

Base case: When $n = 0$, we have $a^{m+n} = a^{m+0} = a^m = a^m \cdot 1 = a^m \cdot a^0 = a^m \cdot a^n$.

Induction step: Suppose $a^{m+n} = a^m \cdot a^n$. Then

$$\begin{aligned} a^{m+(n+1)} &= a^{(m+n)+1} \\ &= a^{m+n} \cdot a && \text{(definition of exponentiation)} \\ &= a^m \cdot a^n \cdot a && \text{inductive hypothesis} \\ &= a^m \cdot a^{n+1} && \text{(definition of exponentiation)} \end{aligned}$$

Which is what we wanted. □

Example
Say we have a sequence of numbers $a_0, a_1, a_2, \ldots$ defined recursively as follows:

$$\begin{aligned} &a_0 = 0; \\ &\text{for every } n \in \mathbb{N}, a_{n+1} = 2a_n + 1 \end{aligned}$$

and we want to find a formula for a_n and prove it is correct.

It may be a good idea to compute the first few terms of a sequence for intuition. We know $a_0 = 0$, so plugging in $n = 0$ in the second equation we get $a_1 = 1$. Continuing this way we get the table of values:

n	0	1	2	3	4	5	6	$\cdots$
a_n	0	1	3	7	15	31	63	$\cdots$

A plausible formula might be $a_n = 2^n - 1$. We will prove this by induction.
(next page)

Theorem. *For a sequence of numbers $a_0, a_1, a_2, \ldots$ defined recursively as follows:*

$$a_0 = 0;$$
$$\text{for every } n \in \mathbb{N}, a_{n+1} = 2a_n + 1$$

for every natural number n, $a_n = 2^n - 1$.

Proof. By induction,
Base case: $a_0 = 0 = 2^0 - 1$.
Induction step: Suppose $a_n = 2^n - 1$. Then

$$\begin{aligned} a_{n+1} &= 2a_n + 1 && \text{(definition of } a_{n+1}) \\ &= 2(2^n - 1) + 1 && \text{(inductive hypothesis)} \\ &= 2^{n+1} - 2 + 1 = 2^{n+1} - 1 \end{aligned}$$

Which is what we wanted. □

Theorem. *For every $x > -1$ and every natural number n, $(1 + x)^n > nx$.*

Scratch work
A natural way to proceed would be to let $x > -1$ be arbitrary, and then use induction on n. In the induction step we assume that $(1 + x)^n > nx$ and then try to prove $(1 + x)^{n+1} > (n + 1)x$. Because we've assumed $x > -1$, we have $(1 + x) > 0$ so we can multiply both sides of the inductive hypothesis to get

$$\begin{aligned} (1 + x)^{n+1} &> (1 + x)nx \\ &= nx + nx^2 \end{aligned}$$

But the conclusion we want is $(1 + x)^{n+1} > (n + 1)x$, and it is not clear how to get this conclusion from the inequality we've derived.

The key here is to replace the original problem with a problem that appears to be harder but is actually easier. Instead of proving the inequality $(1 + x)^n > nx$ directly, we'll prove $(1 + x)^n \geq 1 + nx$, ad then observe that since $1 + nx > nx$, it follows immediately that $(1 + x)^n > nx$.
(next page)

Theorem. *For every $x > -1$ and every natural number n, $(1+x)^n > nx$.*

Proof. Let $x > -1$ be arbitrary. We will prove by induction that for every natural number n, $(1+x)^n \geq 1+nx$, from which it clearly follows that $(1+x)^n > nx$.

Base case: If $n = 0$, then $(1+x)^n = (1+x)^0 = 1 = 1+0 = 1+nx$.

Induction step: Suppose $(1+x)^n \geq 1+nx$. Then

$$\begin{aligned}
(1+x)^{n+1} &= (1+x)(1+x)^n \\
&\geq (1+x)(1+nx)^n && \text{(inductive hypothesis)} \\
&= 1+x+nx+nx^2 \\
&\geq 1+(n+1)x && \text{(since } nx^2 \geq 0\text{)}
\end{aligned}$$

Which is what we wanted. □

Theorem. *For all positive integers a and b,*

$$(2^b-1)\cdot\sum_{k=0}^{a-1} 2^{kb} = 2^{ab}-1$$

Proof. We let b be an arbitrary positive integer and then proceed by induction on a.

Base case: When $a = 1$ we have

$$\begin{aligned}
(2^b-1)\cdot\sum_{k=0}^{a-1} 2^{kb} &= (2^b-1)\cdot\sum_{k=0}^{0} 2^{kb} \\
&= (2^b-1)\cdot 1 \\
&= 2^{ab}-1
\end{aligned}$$

Induction step: suppose $a \geq 1$ and $(2^b-1)\cdot\sum_{k=0}^{a-1} 2^{kb} = 2^{ab}-1$. Then

$$\begin{aligned}
(2^b-1)\cdot\sum_{k=0}^{a} 2^{kb} &= (2^b-1)\cdot\left(\sum_{k=0}^{a-1} 2^{kb} + 2^{ab}\right) \\
&= (2^b-1)\cdot\sum_{k=0}^{a-1} 2^{kb} + 2^b\cdot 2^{ab} - 2^{ab} \\
&= 2^{ab}-1+2^{b+ab}-2^{ab} \\
&= 2^{(a+1)b}-1
\end{aligned}$$

□

5.3 Strong induction

Intuition
In the induction step of a proof by mathematical induction, we prove that a natural number has some property based on the assumption that the previous number has the same property.

In some cases this assumption isn't strong enough to make the proof work, and we need to assume that *all* smaller numbers have the property. This is the idea behind a variant of mathematical induction sometimes called *strong induction*:

To prove a goal of the form $\forall n \in \mathbb{N} P(n)$:
Prove that $\forall n[(\forall k < n P(k)) \to P(n)]$, where both n and k range over the natural numbers in this statement. The most direct way to prove this is to let n be an arbitrary natural number, assume that $\forall k < n P(k)$, then prove $P(n)$.

No base case is necessary in a proof by strong induction. All that is needed is the modified induction step in which we prove that if every natural number smaller than n has the property P, then n has the property P. In a proof by strong induction, we also refer to this assumption as the *inductive hypothesis*.

In ordinary induction, the base case gets the process started, and the induction step shows that the process can always be continued. See that by the time we check that some n has the property P, we've already checked that *all smaller numbers* have the property, meaning we already know that $\forall k < n P(k)$. The idea behind strong induction is that we should be allowed to use this information in our proof of $P(n)$.

Suppose that we follow the strong induction proof strategy, and we've proven the statement $\forall n[(\forall k < n P(k)) \to P(n)]$. Then plugging in 0 for n, we can conclude that $(\forall k < 0 P(k)) \to P(0)$. But because there are no natural numbers smaller than 0, the statement $\forall k < 0 P(k)$ is vacuously true, and $P(0)$ follows (this explains why the base case doesn't have to be checked separately, since $P(0)$ follows from the inductive hypothesis). Similarly, plugging in 1 for n we can conclude that $(\forall k < 1 P(k)) \to P(1)$. The only natural number less than 1 is 0 and we've just shown that $P(0)$ is true, so the statement $\forall k < 1 P(k)$ is true, allowing us to declare $P(1)$. Now $P(0)$ and $P(1)$ are true, and plugging in 2 for n we get $(\forall k < 2 P(k)) \to P(2)$, and we can declare $P(2)$.

Continuing this way, we can show that $P(n)$ is true for every natural number n, as required.
(next page)

Alternative justification

Lemma. *Suppose all variables range over* $\mathbb{N}$. *Then* $\forall n \forall k < nP(k) \leftrightarrow \forall n P(n)$.

Proof. ($\rightarrow$) Suppose that $\forall n \forall k < nP(k)$. Let n be arbitrary. Then $\forall k < n+1P(k)$. In particular, since $n < n+1$, $P(n)$ is true. Since n was arbitrary, this shows that $\forall n P(n)$.

($\leftarrow$) Suppose that $\forall n P(n)$. Then for any n, it is clearly true that $\forall k < nP(k)$. So $\forall n \forall k < nP(k)$. □

Theorem. *Suppose all variables range over* $\mathbb{N}$. *Suppose* $\forall n[(\forall k < nP(k)) \rightarrow P(n)]$. *Then* $\forall n P(n)$.

Proof. Now we use ordinary induction to show $\forall n \forall k < nP(k)$, which then allows us to use the lemma we just derived.

Base case: $n = 0$. Then $\forall k < 0P(k)$ is vacuously true.

Induction step: Suppose $\forall k < nP(k)$. Then by our initial assumption $P(n)$ is true. Therefore $\forall k < n+1P(k)$. So we are done. □

5.3.1 Example: Division algorithm

Theorem. *(Division algorithm) For all natural numbers n and m, if $m > 0$ then there are natural numbers q and r such that $n = qm + r$ and $r < m$. (The numbers q and r are called the quotient and remainder when n is divided by m.)*

Scratch work
We want to prove

$$\forall m > 0\, \forall n \exists q \exists r (n = qm + r \wedge r < m)$$

We let m be an arbitrary positive integer and then use strong induction to prove that $\forall n \exists q \exists r (n = qm + r \wedge r < m)$. This means that we should let n be an arbitrary natural number, assume that $\forall k < n \exists q \exists r (k = qm + r \wedge r < m)$, and prove that $\exists q \exists r (n = qm + r \wedge r < m)$.

Our goal is existential statement, we should try to come up with values of q and r with the required properties. If $n < m$ then we can let $q = 0$ and $r = n$. But if $n \geq m$, this same strategy won't work, since we must have $r < m$. So we try something different. The inductive hypothesis starts with $\forall k < n$, so we should try to find some natural number smaller than n to plug in for k. If we think of division as repeated subtraction, then dividing n by m involves subtracting m from n repeatedly. The first step in this process would be to compute $n - m$, which is a natural number smaller than n since $m > 0$, giving us a candidate for k. It is not clear where this will lead but its worth a try.

Proof. We let m be an arbitrary positive integer and then proceed by strong induction on n.

Suppose n is a natural number, and for every $k < n$ there are natural numbers q and r such that $k = qm + r$ and $r < m$.

Case 1. $n < m$. Let $q = 0$ and $r = n$. Then clearly $n = qm + r$ and $r < m$.

Case 2. $n \geq m$. Let $k = n - m < n$ and note that since $n \geq m$, k is a natural number. By the inductive hypothesis we can choose q' and r' such that $k = q'm + r'$ and $r' < m$. Then $n - m = q'm + r'$, so $n = q'm + r' + m = (q' + 1)m + r'$. Thus if we let $q = q' + 1$ and $r = r'$ then we have $n = qm + r$ and $r < m$, as required. □

The terminology here is somewhat unfortunate since what we are calling the division algorithm is really a theorem and not an algorithm. Nevertheless this terminology is common.

5.3.2 Example: Either prime or a product of two or more primes

Theorem. *Every integer $n > 1$ is either prime or a product of two or more primes.*

Scratch work
The goal here is $\forall n \in \mathbb{N}[n > 1 \to (n \text{ is prime } \vee\ n \text{ is a product of primes})]$. We use strong induction; we let n be natural, we assume our inductive hypothesis $\forall k < n[k > 1 \to (k \text{ is prime } \vee\ k \text{ is a product of primes})]$ and want to show $n > 1 \to (n \text{ is prime } \vee\ n \text{ is a product of primes})$. We assume $n > 1$. Then we assume n is not prime and prove that it must be a product of primes. The assumption that n is not prime means $\exists a \exists b(n = ab \wedge a < n \wedge b < n)$. We immediately use existential instantiation to introduce the new variables a and b into the proof. Applying the inductive hypothesis to a and b leads to the desired conclusion.

Proof. We use strong induction. Let n be an arbitrary real number and suppose that for every natural k, if $1 < k < n$ then k is either a prime or a product of primes. Suppose $n > 1$, then suppose n is not prime. Then we can choose positive integers a and b such that $n = ab, a < n$ and $b < n$. Note that since $a < n = ab$ (and since n is nonzero), it follows that $b > 1$, and similarly we must have $a > 1$. Thus by the inductive hypothesis, each of a and b is either prime or a product of primes. But then since $n = ab$, n is a product of (two or more) primes. □

5.3.3 Example: Fibonacci numbers

The method of recursion also has a 'strong' form. Consider the following definition of a sequence of numbers, called the *fibonacci numbers*.

$$F_0 = 0;$$
$$F_1 = 1;$$
$$\text{for every } n \geq 2, F_n = F_{n-2} + F_{n-1}$$

For example, plugging in $n = 2$ into the last equation we find $F_2 = F_0 + F_1 = 0 + 1 = 1$. Similarly, $F_3 = F_1 + F_2 = 1 + 1 = 2$ and $F_4 = F_2 + F_3 = 1 + 2 = 3$. Continuing this way:

n	0	1	2	3	4	5	6	7	8	$\cdots$
F_n	0	1	1	2	3	5	8	13	21	$\cdots$

Note that, starting with F_2, each fibonacci number is computed using, not just the previous number in the sequence, but also the one before that. It is in this sense that the recursion is 'strong'. It shouldn't be suprising, therefore, that proofs involving fibonacci numbers often require strong induction rather than ordinary induction.

We'll prove the following formula for the fibonacci numbers:

$$F_n = \frac{\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n}{\sqrt{5}}$$

It is hard to believe that this formula is right. Nevertheless, a proof by strong induction shows that it is correct.

Theorem. *If F_n is the nth Fibonacci number, then*

$$F_n = \frac{\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n}{\sqrt{5}}$$

Scratch work
Because F_0 and F_1 are defined separately from F_n for $n \geq 2$, we check the formula for these cases separately. For $n \geq 2$, the definition of F_n suggests that we should use F_{n-2} and F_{n-1}. Because we need to know that the formula works for *two* previous cases, we must use strong induction rather than ordinary induction. The rest of the proof is straightforward.
(next page)

Proof. We use strong induction. Let n be an arbitrary natural number, and suppose that for all $k < m$,

$$F_k = \frac{\left(\frac{1+\sqrt{5}}{2}\right)^k - \left(\frac{1-\sqrt{5}}{2}\right)^k}{\sqrt{5}}$$

Case 1. $n = 0$. Then

$$\begin{aligned}\frac{\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n}{\sqrt{5}} &= \frac{\left(\frac{1+\sqrt{5}}{2}\right)^0 - \left(\frac{1-\sqrt{5}}{2}\right)^0}{\sqrt{5}} \\ &= \frac{1-1}{\sqrt{5}} = 0 = F_0\end{aligned}$$

Case 2. $n = 1$. Then

$$\begin{aligned}\frac{\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n}{\sqrt{5}} &= \frac{\left(\frac{1+\sqrt{5}}{2}\right)^1 - \left(\frac{1-\sqrt{5}}{2}\right)^1}{\sqrt{5}} \\ &= \frac{\sqrt{5}}{\sqrt{5}} = 1 = F_1\end{aligned}$$

Case 3. $n \geq 2$. Then applying the inductive hypothesis to $n-2$ and $n-1$ we get

$$\begin{aligned}F_n &= F_{n-2} + F_{n-1} \\ &= \frac{\left(\frac{1+\sqrt{5}}{2}\right)^{n-2} - \left(\frac{1-\sqrt{5}}{2}\right)^{n-2}}{\sqrt{5}} + \frac{\left(\frac{1+\sqrt{5}}{2}\right)^{n-1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n-1}}{\sqrt{5}} \\ &= \frac{\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n-2} + \left(\frac{1+\sqrt{5}}{2}\right)^{n-1}\right] - \left[\left(\frac{1-\sqrt{5}}{2}\right)^{n-2} + \left(\frac{1-\sqrt{5}}{2}\right)^{n-1}\right]}{\sqrt{5}} \\ &= \frac{\left(\frac{1+\sqrt{5}}{2}\right)^{n-2}\left[1 + \frac{1+\sqrt{5}}{2}\right] - \left(\frac{1-\sqrt{5}}{2}\right)^{n-2}\left[1 + \frac{1-\sqrt{5}}{2}\right]}{\sqrt{5}}\end{aligned}$$

Now note that

$$\left(\frac{1+\sqrt{5}}{2}\right)^2 = \frac{1+2\sqrt{5}+5}{4} = \frac{6+2\sqrt{5}}{4} = \frac{3+\sqrt{5}}{2} = 1 + \frac{1+\sqrt{5}}{2}$$

and similarly

$$\left(\frac{1-\sqrt{5}}{2}\right)^2 = 1 + \frac{1-\sqrt{5}}{2}$$

(next page)

Substituting into the formula for F_n, we get

$$\begin{aligned} F_n &= \frac{\left(\frac{1+\sqrt{5}}{2}\right)^{n-2}\left(\frac{1+\sqrt{5}}{2}\right)^2 - \left(\frac{1-\sqrt{5}}{2}\right)^{n-2}\left(\frac{1-\sqrt{5}}{2}\right)^2}{\sqrt{5}} \\ &= \frac{\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n}{\sqrt{5}} \end{aligned}$$

□

5.3.4 Example: Well-ordering principle

Theorem. *(Well-ordering principle) Every nonempty set of natural numbers has a smallest element.*

Scratch work
Formally this means

$$\forall S[(S \subseteq \mathbb{N} \wedge S \neq \emptyset) \rightarrow (\exists x \in S \forall y \in A(x \leq y)]$$

which is the same as

$$\forall S \subseteq \mathbb{N}(S \neq \emptyset \rightarrow \exists x \in S \forall y \in A(x \leq y))$$

After letting S be an arbitrary subset of $\mathbb{N}$, we'll prove the contrapositive, assuming S has no smallest element and proving $S = \emptyset$. For a set $S \subseteq \mathbb{N}$, to say that $S = \emptyset$ is the same as saying that $\forall n \in \mathbb{N}(n \notin S)$ (it can be shown that $S \subseteq \mathbb{N} \rightarrow (S = \emptyset \leftrightarrow \forall x \in \mathbb{N}(x \notin S))$.) We'll prove $\forall n \in \mathbb{N}(n \notin S)$ by strong induction.

Proof. Suppose $S \subseteq \mathbb{N}$, and then suppose S does not have a smallest element. We will prove that $\forall n \in \mathbb{N}(n \notin S)$, so $S = \emptyset$. Thus if $S \neq \emptyset$ then S must have a smallest element.

We use strong induction to prove $\forall n \in \mathbb{N}(n \notin S)$. Suppose that $n \in \mathbb{N}$ and $\forall k < n(k \notin S)$. Now suppose that $n \in S$. Using the contrapositive of the inductive hypothesis, n is the smallest element of S, contradicting the assumption that S has no smallest element. Therefore $n \notin S$. □

(next page)

Applying the Well-ordering principle
Sometimes proofs that could be done by induction are written instead as applications of the well-ordering principle.

Theorem. $\sqrt{2}$ *is irrational.*

Irrational means not rational. Our goal is a negative statement, so we proceed using proof by contradiction, assuming that $\sqrt{2}$ is rational. This means that there exists integers p and q such that $p/q = \sqrt{2}$, and since $\sqrt{2}$ is positive, we may as well restrict our attention to positive p and q. Since this is an existential statement, we let p and q stand for positive integers such that $p/q = \sqrt{2}$. Simple algebraic manipulations allow us to conclude that p and q must both be even. Thus in the fraction p/q we can cancel a 2 from both the numerator and denominator, getting a new fraction with smaller numerator and denominator that is equal to $\sqrt{2}$.

The key idea is to note that our reasoning would apply to *any* fraction that is equal to $\sqrt{2}$. Thus in any such fraction we can endlessly cancel a factor of 2 from the numerator and denominator, and therefore there can be no smallest possible numerator or denominator. This would violate the well-ordering principle, leading to our contradiction.

Proof. Suppose that $\sqrt{2}$ is rational. This means that $\exists q \in \mathbb{Z}^+ \exists p \in \mathbb{Z}^+ (p/q = \sqrt{2})$, so the set $S = \{q \in \mathbb{Z}^+ | \exists p \in \mathbb{Z}^+ (p/q = \sqrt{2})\}$ is nonempty. By the well-ordering principle we can declare some q to be the smallest element of S. Since $q \in S$, we can choose some $p \in \mathbb{Z}^+$ such that $p/q = \sqrt{2}$. Therefore $p^2/q^2 = 2$, so $p^2 = 2q^2$ and therefore p^2 is even. In an earlier theorem we proved that for any integer x, x is even iff x^2 is even. Since p^2 is even, p must be even, so we can choose some $p' \in \mathbb{Z}^+$ such that $p = 2p'$. Therefore $p^2 = 4p'^2$, and substituting this into the equation $p^2 = 2q^2$ we get $4p'^2 = 2q^2$ and then $2p'^2 = q^2$, so q^2 is even. Using the earlier theorem again q must then be even. So we can choose some $q' \in \mathbb{Z}^+$ such that $q = 2q'$. But then $\sqrt{2} = p/q = (2p')/(2q') = p'/q'$, so $q' \in S$. Clearly $q' < q$, contradicting the fact that q was chosen to be the smallest element of S. Therefore $\sqrt{2}$ is irrational. $\square$

5.4 Closures again

Recall that if $f : A \to A$ and $B \subseteq A$ then the closure of B under f is the smallest set $C \subseteq A$ such that $B \subseteq C$ and C is closed under f. In this section we'll show that we can find this set C by starting with B and then adding only those elements that *must* be added to end up with a set closed under f.

Theorem. *Suppose $B \subseteq C \subseteq A$, $f : A \to A$, and C is closed under f, then $\forall x \in B(f(x) \in C)$. Furthermore, $\forall x \in B(f(x) \in C) \leftrightarrow f(B) \subseteq C$.*

Proof. First we show $\forall x \in B(f(x) \in C)$. Suppose $x \in B$. Then since $B \subseteq C$, $x \in C$. Since C is closed under f, $f(x) \in C$.

Now we show the equivalence. First note that

$$f(B) \leftrightarrow \{x | \exists y \in B(f(y) = x)\}$$

($\rightarrow$) Suppose $\forall x \in B(f(x) \in C)$. Let $x \in f(B)$, then we can declare some $y \in B$ such that $f(y) = x$. Since $\forall x \in B(f(x) \in C)$, $x = f(y) \in C$.

($\leftarrow$) Now suppose $f(B) \subseteq C$. Let x be an arbitrary element of B. Since $x \in B \subseteq A$, and $f : A \to A$, we can declare some $y \in A$ such that $f(x) = y$. Since $x \in B$ and $f(x) = y$, $y \in f(B)$, so $f(x) = y \in C$. □

If we want to find a set $C \subseteq A$ such that $B \subseteq C$ and C is closed under f, then for every $x \in B$, we must have $f(x) \in C$, or in other words, $f(B) \subseteq C$, where $f(B)$ is the image of B under f.

A similar reasoning implies that the image of $f(B)$ under f must also be a subset of C, that is, $f(f(B)) \subseteq C$. Continuing this way leads to a sequence of sets that must be contained in C: B, $f(B)$, $f(f(B))$, and so on. We will prove that putting these sets together by taking their union will give us the closure of B under f.

The idea here is that if we let $B_0 = B, B_1 = f(B), B_2 = f(f(B)), \ldots$, then the closure of B under f is $B_0 \cup B_1 \cup B_2 \cup \cdots$. The use of ellipses suggests that we could use a recursive definition and induction to prove this.
(next page)

Theorem. *Suppose $f : A \to A$ and $B \subseteq A$. Let the sets B_0, B_1, $B_2, \ldots$ be defined recursively as follows:*

$$B_0 = B;$$
$$\text{for all } n \in \mathbb{N},\ B_{n+1} = f(B_n)$$

Then the closure of B under f is the set $\bigcup_{n\in\mathbb{N}} B_n$.

Proof. Let $C = \bigcup_{n\in\mathbb{N}} B_n$. First we prove by induction that each set B_n is a subset of A.

Base case: $B_0 = B \subseteq A$.

Induction step: Suppose $n \in \mathbb{N}$ and $B_n \in A$. Now suppose x is an element of B_{n+1}, which by definition is equal to $f(B_n)$, so $x \in f(B_n)$. Since $f : A \to A$ and $B_n \subseteq A$, by definition $x \in A$, so $B_{n+1} \subseteq A$.

Now we show that $C \subseteq A$. Let x be an arbitrary element of C. Then for some $n \in \mathbb{N}$ we can declare $x \in B_n$ (union on an indexed family). Since we know $\forall n \in \mathbb{N} B_n \subseteq A$, $x \in A$.

We now have proven $C \subseteq A$. To show that C is a closure, we must check that $B \subseteq C$, C is closed under f, and for every set $D \subseteq A$, if $B \subseteq D$ and D is closed under f then $C \subseteq D$.

For the first requirement, suppose x in B. Then since $x \in B = B_0$, $x \in \bigcup_{n\in\mathbb{N}} B_n = C$. So $B \subseteq C$.

For the second, suppose that $x \in C$, then we can declare some $m \in \mathbb{N}$ such that $x \in B_m$. Then by the definition of the image of a function, $f(x) \in f(B_n) = B_{n+1}$. So $f(x) \in \bigcup_{n\in\mathbb{N}} B_n = C$. Since x was an arbitrary element of C, C is closed under f.

For the final requirement, suppose $B \subseteq D \subseteq A$, and that D is closed under f. We need to show $C \subseteq D$. At this point, we will first show that $\forall n \in \mathbb{N}(B_n \subseteq D)$ implies $C = \bigcup_{n\in\mathbb{N}} B_n \subseteq D$. Supposing that $\forall n \in \mathbb{N}(B_n \subseteq D)$, we let x be an arbitrary element of C. Then we can declare some $n \in \mathbb{N}$ such that $x \in B_n$. By our assumption $\forall n \in \mathbb{N}(B_n \subseteq D)$, $B_n \subseteq D$, so $x \in D$. So $C \subseteq D$, proving the conditional statement.

Now we show $\forall n \in \mathbb{N}(B_n \subseteq D)$. We will prove this by induction. The base case holds since $B_0 = B \subseteq D$ by assumption. For thse induction step, suppose that $n \in \mathbb{N}$ and $B_n \subseteq D$. Now suppose $x \in B_{n+1}$. By definition this means $x \in f(B_n)$, so there is some $b \in B_n$ such that $f(b) = x$. By the inductive hypothesis $B_n \subseteq D$ so $b \in D$, and since D is closed under f, $x = f(b) \in D$. Since x was arbitrary, $B_{n+1} \subseteq D$. □

Chapter 6

Number Theory

6.1 Greatest Common Divisors

Number theory is the study of the positive integers $1, 2, 3, \ldots$. We begin with a fundamental concept, the *greatest common divisor* of a pair of positive integers.

Definition. Suppose a is a positive integer. The *divisors* of a are the positive integers that divide a. We will denote the set of divisors of a by $D(a)$. Thus,

$$D(a) = \{d \in \mathbb{Z}^+ \mid d \text{ divides } a\} = \{d \in \mathbb{Z}^+ \mid \exists k \in \mathbb{Z}(a = kd)\}$$

If a and b are two positive integers, then $D(a) \cap D(b)$ is the set of positive integers that divide both a and b—the *common divisors* of a and b. The largest element of this set is called the *greatest common divisor* of a and b, denoted $\gcd(a, b)$.

For example, $D(18) = \{1, 2, 3, 6, 9, 18\}$ and $D(12) = \{1, 2, 3, 4, 6, 12\}$, so the set of common divisors of 18 and 12 is $D(18) \cap D(12) = \{1, 2, 3, 6\}$. The largest of these common divisors is 6, so $\gcd(18, 12) = 6$.
(next page)

Theorem. *Suppose R is a total order on set A. Prove that every finite, nonempty set $B \subseteq A$ has an R-smallest element and an R-largest element.*

Scratch work

$\forall n \geq 1 \forall B \subset A(B$ has n elements

$\rightarrow B$ has an R-smallest element $\wedge$ B has an R-largest element)

Where n is a natural number.

Proof. We proceed by induction.

Base case: $n = 1$. Let B be an arbitrary subset of A, suppose B has 1 element. Then since B has one element, we can declare some $x \in B$. Then by reflexivity, xRx, since x is the only element in B, $\forall b \in B(xRb)$, and it is the R-smallest element in B. Similarly $\forall b \in B(bRx)$, so it is also the R-largest element in b.

Induction step: Suppose n is an arbitrary natural number where $n \geq 1$. Suppose that for any arbitary subset of A with n elements, that subset has an R-smallest and an R-largest element. Let B be an arbitrary subset of A, and suppose B has $n+1$ elements. We can declare x_0 to be some element of B, and we can denote $B' = B \backslash \{x_0\}$, which has n elements. We need to show that B has an R-smallest element and an R-largest element.

First we show that B has an R-smallest element. Since B' has n elements, by the inductive hypothesis we can declare x_1 to be the R-smallest element of B'. We consider two cases:

Case 1: x_0Rx_1. We will show that x_0 is the smallest element of B. Let b be an arbitrary element of B. Case 1a: $b = x_0$. Then by reflexivity, $x_0Rx_0 = x_0Rb$. Case 1b: $b \neq x_0$. Then $b \in B'$, and x_1Rb; since x_0Rx_1, by transitivity we have x_0Rb. Since b was arbitrary, x_0 is the smallest element in B.

Case 2: $\neg x_0Rx_1$. Now we will show that x_1 is the smallest element of B. Again let b to be some arbitrary element of B. First see that since R is a total order, $(x_0Rx_1 \vee x_1Rx_0)$, so $(\neg x_0Rx_1 \rightarrow x_1Rx_0)$, and x_1Rx_0. Case 2a: $b = x_0$. Then by reflexivity, x_0Rx_0 and x_0Rb. By transitivity, x_1Rb. Case 2b: $b \neq x_0$. Then $b \in B'$ and x_1Rb. Since b was arbitrary, x_1 is the smallest element of b.

Next we show that B has an R-largest element. Since B' has n elements, by the inductive hypothesis we can declare x_2 to be the R-largest element of B'.

Case 1: x_2Rx_0. We will show that x_0 is the largest element of B. Let b be an arbitrary element of B. Case 1a: $b = x_0$. Then by reflexivity, x_0Rx_0 and bRx_0. Case 1b: $b \neq x_0$. Then $b \in B'$ and bRx_2. By transitivity bRx_0. Since b was arbitrary, x_0 is the largest element in B.

Case 2: $\neg x_2Rx_0$. Now we will show that x_2 is the largest element of B. Again let b be some arbitrary element of B. First see that since R is a total order, $(x_2Rx_0 \vee x_0Rx_2)$, meaning $(\neg x_2Rx_0 \rightarrow x_0Rx_2)$, so x_0Rx_2. Case 2a: $b = x_0$. Then by reflexivity $x_0Rx_0 = bRx_0$. By transitivity bRx_2. Case 2b: $b \neq x_0$. Then $b \in B'$ and bRx_2. Since b was arbitrary, x_2 is the largest element in B. $\square$

(next page)

Lemma. *For any positive integer a, 1 and a are always elements of $D(a)$.*

Proof. Suppose a is a positive integer. 1 is an element of $D(a)$ because $a = a \cdot 1$. This expression also shows that a is an element of $D(a)$. □

It is clear that for any two positive integers a and b, $D(a) \cap D(b)$ is nonempty since $1 \in D(a)$ and $1 \in D(b)$. It is also obvious (though not rigorously shown), that $D(a) \cap D(b)$ is a finite set since $D(a)$ is a finite set. By the previous theorem we can therefore say that $D(a) \cap D(b)$ contains a largest element, meaning $\gcd(a, b)$ is always defined.

Lemma. *For a positive integer a, a is the largest element of $D(a)$.*

Proof. Suppose x is an arbitrary element of $D(a)$. Then $x \in \mathbb{Z}^+$ and we can declare some integer k such that $a = kx$. Since a and x are both positive integers, $k = a/x$ is a positive integer, so since $a = kx$, $x \leq a$. □

Since $D(a) \cap D(b) = D(b) \cap D(a)$, $\gcd(a, b) = \gcd(b, a)$. In other words, in the notation for the greatest common divisor of two positive integers, it doesn't matter which integer we list first. We will often find it convenient to list the larger integer first; in particular when compute $\gcd(a, b)$, we will assume that $a \geq b$. Recall the division algorithm mentioned earlier:

Theorem. *(Division algorithm) For all natural numbers n and m, if $m > 0$ then there are natural numbers q and r such that $n = qm + r$ and $r < m$. (The numbers q and r are called the quotient and remainder when n is divided by m.)*

Theorem. *Suppose a and b are positive integers with $a \geq b$. Let r be the remainder when we divide a by b. If $r = 0$ then $\gcd(a, b) = b$ and if $r > 0$ then $\gcd(a, b) = \gcd(b, r)$.*

Proof. For natural numbers q, b, we have $a = qb + r$ (the division algorithm guarantees this). First we show that if $r = 0$ then $\gcd(a, b) = b$. Suppose $r = 0$, then $a = qb$, so $b \in D(a)$. We know that $b \in D(b)$, so $b \in D(a) \cap D(b)$. Let x be an arbitrary element of $D(a) \cap D(b)$. We know that b is the largest element of $D(b)$ and $x \in D(b)$, so $x \leq b$.

Now we show that if $r > 0$ then $\gcd(a, b) = \gcd(b, r)$. Suppose $r > 0$. We will show that $D(a) \cap D(b) = D(b) \cap D(r)$.

($\rightarrow$) Suppose first that $d \in D(a) \cap D(b)$. Then $d \mid a$ and $d \mid b$, so there are integers j and k such that $a = jd$ and $b = kd$. From the equation $a = qb + r$ we get $r = a - qb$ and then $r = jd - qkd = (j - qk)d$, so $d \mid r$ and $d \in D(r)$. As such we have $d \in D(b) \cap D(r)$.

($\leftarrow$) Suppose now that $d \in D(b) \cap D(r)$. Then $d \mid b$ and $d \mid r$, so there are integers j and k such that $b = jd$ and $r = kd$. From the equation $a = qb + r$ we get $a = (qjd + kd) = (qj + k)d$, so $d \mid a$ and $d \in D(a)$. So we have $d \in D(a) \cap D(b)$, and $D(a) \cap D(b) = D(b) \cap D(r)$. By the definition of the greatest common divisor, it follows that $\gcd(a, b) = \gcd(b, r)$. □

6.1.1 Computation and Euclidean algorithm

Computation

Given two positive integers a and b, we can come up with a good way to determine their gcd. Suppose we want to find $\gcd(672, 161)$. We begin by dividing $a = 672$ by $b = 161$, which gives us a quotient $q = 4$ and remainder $r = 28$:

$$672 = 4 \cdot 161 + 28$$

By the previous theorem, we can conclude that $\gcd(672, 161) = \gcd(a, b) = \gcd(b, r) = \gcd(161, 28)$. So we try to compute $\gcd(161, 28)$, which seems like an easier problem.

We can apply the same method again. We divide 161 by 28, to get a quotient of 5 and a remainder of 21:

$$161 = 5 \cdot 28 + 21$$

Since $\gcd(161, 28) = \gcd(28, 21)$, we now try to compute $\gcd(28, 21)$:

$$28 = 1 \cdot 21 + 7$$

Finally, $\gcd(28, 21) = \gcd(21, 7)$, but $21 = 3 \cdot 7 + 0$, so $7 \mid 21$ and therefore $\gcd(21, 7) = 7$. This is the answer to our original problem:

$$\gcd(672, 161) = \gcd(161, 28) = \gcd(28, 21) = \gcd(21, 7) = 7$$

(next page)

Cont.
Now we generalise. Suppose we want to find $\gcd(a, b)$ for positive integers a and b where $a \geq b$. We define a sequence of natural numbers $r_0, r_1, r_2, \ldots$ recursively as follows. To start off the sequence, we let $r_0 = a$ and $r_1 = b$ (notice that $r_0 \geq r_1$). Then we let q_2 and r_2 be the quotient and remainder when we divide r_0 by r_1:

$$r_0 = q_2 \cdot r_1 + r_2, \quad 0 \leq r_2 < r_1$$

If $r_2 \neq 0$, we divide r_1 by r_2 to get a quotient q_3 and remainder r_3. In general, having computed $r_0, r_1, \ldots, r_n$, if $r_n \neq 0$ then we divide r_{n-1} by r_n to produce a quotient and remainder q_{n+1} and r_{n+1}:

$$r_{n-1} = q_{n+1} \cdot r_n + r_{n+1}, \quad 0 \leq r_{n+1} < r_n$$

The calculation stops when we reach a remainder of 0.

Note that we can be sure that we will eventually have a remainder of 0. If we didn't, then the sequence of divisions would go on forever and we would end up with an infinite sequence of positive integers $r_0, r_1, r_2, \ldots$ with $r_0 \geq r_1 > r_2 \cdots$. $\{r_0, r_1, r_2 \ldots\}$ would be a nonempty set of natural numbers with no smallest element, contradicting the well-ordering principle.

Suppose m is the largest index for which $r_m \neq 0$. Then $r_{m+1} = 0$ and there are m divisions, which can be summarised as

$$\begin{aligned} r_0 &= q_2 \cdot r_1 + r_2 \\ r_1 &= q_3 \cdot r_2 + r_3 \\ &\vdots \\ r_{m-1} &= q_{m+1} \cdot r_m + 0 \end{aligned}$$

As per the previous theorem, we can conclude that

$$\gcd(a, b) = \gcd(r_0, r_1) = \gcd(r_1, r_2) = \cdots = \gcd(r_{m-1}, r_m) = r_m$$

$\gcd(a, b)$ is the last nonzero value in the sequence $r_0, r_1, r_2, \ldots$. This method of computing the gcd of two positive integers is called the *Euclidean algorithm*, named for Euclid.

6.1.2 Linear combinations—Extended Euclidean algorithm

Intuition
Consider finding the greatest common divisor of 444 and 1392. We apply the euclidean algorithm with $a = 1392$ and $b = 444$:

n	q_n	r_n	Division
0		1392	
1		444	$1392 = 3 \cdot 444 + 60$
2	3	60	$444 = 7 \cdot 60 + 24$
3	7	24	$60 = 2 \cdot 24 + 12$
4	2	12	$24 = 2 \cdot 12 + 0$
5	2	0	

It is instructive to see how the remainders were computed with respect to the initial inputs a and b. Rearranging the equations, see that

$$r_2 = 60 = 1392 - 3 \cdot 444 = a - 3b$$

From the second equation we get

$$r_3 = 24 = 444 - 7 \cdot 60 = b - 7r_2 = b - 7(a - 3b) = -7a + 22b$$

and the third equation gives

$$r_4 = 12 = 60 - 2 \cdot 24 = r_2 = 2r3 = (a - 3b) - 2(-7a + 22b) = 15a - 47b$$

See that each remainder can be written in the form $sa + tb$ for some integers s and t. We say that each remaindr is a *linear combination* of a and b. But the last nonzero remainder is the greatest common divisor of a and b, so we conclude that $\gcd(a, b) = r_4 = 15a - 47b$.
(next page)

Theorem. *For all positive integers a and b there are integers s and t such that* $\gcd(a, b) = sa + tb$.

Proof. As usual we may assume $a \geq b$; if not we can simply reverse the values of a and b. Let $r_0, r_1, \ldots, r_{m+1}$ be the sequence of numbers produced by the euclidean algorithm (this is guaranteed by the division algorithm), where $r_m \neq 0$ and $r_{m+1} = 0$. We will show that for every natural number $n \leq m$, r_n is a linear combination of a and b. In other words, for every natural number n, if $n \leq m$ then there are integers s_n and t_n such that $r_n = s_n a + t_n b$. We proceed by strong induction.

Suppose n is a natural number and $n \leq m$, and suppose also that for all $k < n$, r_k is a linear combination of a and b. We consider three cases:

Case 1: $n = 0$. Then $r_n = r_0 = a = s_0 a + t_0 b$, where $s_0 = 1$ and $t_0 = 0$.

Case 2: $n = 1$. Then $r_n = r_1 = b = s_1 a + t_1 b$, where $s_1 = 0$ and $t_1 = 1$.

Case 3: $n \geq 2$. Then r_n is the remainder when r_{n-2} is divided by r_{n-1}:

$$r_{n-2} = q_n \cdot r_{n-1} + r_n$$

By the inductive hypothesis, there are integers $s_{n-1}, s_{n-2}, t_{n-1}$, and t_{n-2} such that

$$r_{n-1} = s_{n-1} a + t_{n-1} b, \quad r_{n-2} = s_{n-2} a + t_{n-2} b$$

Therefore

$$\begin{aligned} r_n &= r_{n-2} - q_n \cdot r_{n-1} = (s_{n-2} a + t_{n-2} b) - q_n (s_{n-1} a + t_{n-1} b) \\ &= (s_{n-2} - q_n s_{n-1}) a + (t_{n-2} - q_n t_{n-1}) b \end{aligned}$$

so $r_n = s_n a + t_n b$, where $s_n = s_{n-2} - q_n s_{n-1}$ and $t_n = t_{n-2} - q_n t_{n-1}$.

This completes the inductive proof that for every $n \leq m$, r_n is a linear combination of a and b. Applying this statement in the case $n = m$, we conclude that $\gcd(a, b) = r_m$ is a linear combination of a and b. □

One advantage of this proof is that it provides us with a method to find integers s and t such that $\gcd(a, b) = sa + tb$. While carrying out the euclidean algorithm, we can compute numbers s_n and t_n recursively by using the formulas

$$\begin{aligned} s_0 &= 1, & t_0 &= 0, \\ s_1 &= 0, & t_1 &= 1, \\ \text{for } n \geq 2,\ s_n &= s_{n-2} - q_n s_{n-1}, & t_n &= t_{n-2} - q_n t_{n-1}. \end{aligned}$$

If m is the largest index for which $r_m \neq 0$, then $\gcd(a, b) = r_m = s_m a + t_m b$. This version of the euclidean algorithm where we can keep track of these extra numbers s_n and t_n is called the *extended Euclidean Algorithm.*

(next page)

Example—Extended Euclidean algorithm
We use the Extended Euclidean algorithm to find $\gcd(574, 168)$, and express it as a linear combination of 574 and 168.

n	q_n	r_n	s_n	t_n	Division
0		574	1	0	
1		168	0	1	$574 = 3 \cdot 168 + 70$
2	3	70	$1 - 3 \cdot 0 = 1$	$0 - 3 \cdot 1 = -3$	$168 = 2 \cdot 70 + 28$
3	2	28	$0 - 2 \cdot 1 = -2$	$1 - 2 \cdot (-3) = 7$	$70 = 2 \cdot 28 + 14$
4	2	14	$1 - 2 \cdot (-2) = 5$	$-3 - 2 \cdot 7 = -17$	$28 = 2 \cdot 14 + 0$
5	2	0			

We conclude that $\gcd(574, 168) = 14 = 5 \cdot 574 - 17 \cdot 168$.

Theorem. *For all positive integers a, b and d, if $d \mid a$ and $d \mid b$ then $d \mid \gcd(a, b)$.*

Proof. Let a, b, d be arbitrary positive integers and suppose that $d \mid a$ and $d \mid b$. Then there are integers j and k such that $a = jd$ and $b = kd$. Now by the previous theorem we can declare s and t to be integers such that $\gcd(a, b) = sa + tb$. Then

$$\gcd(a, b) = sa + tb = sjd + tkd = (sj + tk)d$$

so $d \mid \gcd(a, b)$. □

6.2 Prime factorization

Recall that saying some number $n \in \mathbb{Z}^+$ is prime means $\neg\exists a \in \mathbb{Z}^+ \exists b \in \mathbb{Z}^+ (n = ab \wedge a < n \wedge b < n)$. Recall from the section on induction that

Theorem. *Every integer $n > 1$ is either prime or a product of two or more primes.*

We mean this to say that every integer $n > 1$ has a *prime factorisation.*

6.2.1 Relatively prime and more

Definition. If a and b are positive integers and $\gcd(a, b) = 1$, then we say that a and b are *relatively prime.*

That is, a and b are relatively prime if their only common divisor is 1. For example, $D(50) = \{1, 2, 5, 10, 25, 50\}$ and $D(63) = \{1, 3, 7, 9, 21, 63\}$. $\gcd(50, 63) = 1$, so 50 and 63 are relatively prime.

Theorem. *For all positive integers $a, b,$ and c, if $c \mid ab$ and $\gcd(a, c) = 1$ then $c \mid b$.*

Proof. Suppose $c \mid ab$ and $\gcd(a, c) = 1$. Then there is some integer j such that $ab = jc$, and by a previous theorem, we can declare integers s and t such that $sa + tc = 1$. Therefore

$$b = b \cdot 1 = b \cdot (sa + tc) = sab + tbc = sjc + tbc = (sj + tb)c$$

so $c \mid b$. □

Theorem. *Assuming that if p is a prime number then $D(p) = \{1, p\}$, if $p \nmid a$ then $\gcd(a, p) = 1$.*

Proof. We know $D(p) = \{1, p\}$. Suppose $p \nmid a$. Then $p \notin D(a)$ and $p \notin D(p) \cap D(a)$. So only $1 \in D(p) \cap D(a)$ and $\gcd(a, p) = 1$. □

We will intuitively assume that if p is a prime number then $D(p) = \{1, p\}$.

Theorem. *For all positive integers a, b and p, if p is prime and $p \mid ab$ then either $p \mid a$ or $p \mid b$.*

Proof. Suppose p is prime and $p \mid ab$. From the previous theorem, if $p \nmid a$ then a and p are relatively prime. By the theorem before that, $p \mid b$. Thus either $p \mid a$ or $p \mid b$. □

We use a disjunctive syllogism here.
(next page)

Theorem. *Suppose p is a prime number. For positive integers $a_1, a_2, \ldots, a_k$, if $p \mid (a_1 a_2 \cdots a_k)$, then for some $i \in \{1, 2, \ldots, k\}$, $p \mid a_i$.*

Scratch work

$$\forall k \geq 1(\forall a_1 \in \mathbb{Z}^+ \cdots \forall a_k \in \mathbb{Z}^+(p \mid (a_1 \cdots a_k) \to \exists i \in \{1 \ldots k\}(p \mid a_i)))$$

Proof. We prove this theorem by induction on k.

Base case: $k = 1$. Then if $p \mid a_1$ for arbitrary $a_1 \in \mathbb{Z}^+$, this is exactly the goal, taking $i = 1$.

Induction step: Now assuming that the statement holds for any k integers, let $a_1, a_2, \ldots, a_{k+1}$ be positive integers such that $p \mid (a_1 a_2 \cdots a_k a_{k+1})$. Since $a_1 a_2 \cdots a_k a_{k+1} = (a_1 a_2 \cdots a_k) a_{k+1}$, by the previous theorem, either $p \mid (a_1 a_2 \cdots a_k)$ or $p \mid a_{k+1}$. In the first case, by the inductive hypothesis we have $p \mid a_i$ for some $i \in \{1, 2, \ldots, k\}$, and in the second we have $p \mid a_i$ where $i = k + 1$. □

Theorem. *Suppose that $p_1, p_2, \ldots p_k$ and $q_1, q_2, \ldots q_m$ are prime numbers, $p_1 \leq p_2 \leq \cdots \leq p_k$, $q_1 \leq q_2 \leq \cdots \leq q_m$, and $p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_m$. Then $k = m$ and for all $i \in \{1, \ldots, k\}$, $p_i = q_i$.*

Scratch work

$$\begin{aligned}\forall k \geq 1 \forall m \geq 1([\forall p_1 \ldots \forall p_k \forall q_1 \ldots \forall q_m (&p_1 \ldots p_k \text{ are prime} \wedge q_1 \ldots q_m \text{ are prime} \\ &\wedge p_1 \leq p_2 \leq \ldots p_k \wedge q_1 \leq q_2 \leq \ldots q_m \wedge p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_m] \\ &\to k = m \wedge \forall i \in \{1 \cdots k\} p_i = q_i))\end{aligned}$$

Proof. The proof will be by induction on k. We are essentially showing that for all $k \geq 1$, if the product of some *nondecreasing list* of k prime numbers is equal to the product of another nondecreasing list of prime numbers, then the two lists must be the same.

Base case: When $k = 1$, we have $p_1 = q_1 q_2 \cdots q_m$. If $m > 1$ then this contradicts the fact that p_1 is prime. Therefore $m = 1$ and $p_1 = q_1$.

Induction step: Suppose the statement is true for products of nondecreasing lists of k prime numbers, let $m \geq 1$ be arbitrary and suppose that $p_1, p_2, \ldots, p_{k+1}$ and $q_1, q_2, \ldots q_m$ are prime numbers, $p_1 \leq p_2 \cdots \leq p_{k+1}$, $q_1 \leq q_2 \leq \cdots \leq q_m$, and $p_1 p_2 \cdots p_{k+1} = q_1 q_2 \cdots q_m$. First notice that if $k + 1 \geq 1$, so if $m = 1$ then $p_1 p_2 \cdots p_{k+1} = q_1$, and as in the base case this contradicts the fact that q_1 is prime, so $m > 1$ (the point of this statement will be apparent by the end of the proof).

Clearly $p_{k+1} \mid (p_1 p_2 \cdots p_{k+1})$, so $p_{k+1} \mid (q_1 q_2 \cdots q_m)$; by the previous theorem it follows that $p_{k+1} \mid q_i$ for some $i \in \{1, \ldots, m\}$. Therefore $p_{k+1} \leq q_i \leq q_m$. A similar argument shows that $q_m \mid p_j$ for some $j \in \{1, \ldots, k+1\}$, so $q_m \leq p_j \leq p_{k+1}$. Since $p_{k+1} \leq q_m$ and $q_m \leq p_{k+1}$, we conclude that $p_{k+1} = q_m$. Cancelling these factors from the equation $p_1 p_2 \cdots p_{k+1} = q_1 q_2 \cdots q_m$ we get $p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_{m-1}$, and now the inductive hypothesis tells us that the remaining factors on both sides of the equation are equal, as required. (the statement confirming $m > 1$ is necessary since we use q_{m-1}). □

(next page)

Cont.
See that the previous theorem used the intuitive assumption that $a \mid b \to b \leq a$. I now use this to show that for a prime number p, $D(p) = \{1, p\}$.

Theorem. *For a prime number p, $D(p) = \{1, p\}$.*

Scratch work
Saying some number $n \in \mathbb{Z}^+$ is prime means $\neg\exists a \in \mathbb{Z}^+ \exists b \in \mathbb{Z}^+ (n = ab \wedge a < n \wedge b < n)$. Our goal is to show

$$\neg\exists x \in \mathbb{Z}^+ (x \neq 1 \wedge x \neq p \wedge x \mid p)$$

Proof. Suppose there exists some $x \in \mathbb{Z}^+$ in $D(p)$ that isn't 1 or p. Then we can declare some $k \in \mathbb{Z}$ such that $kx = p$. Since $x \mid p$ we can say that $x \leq p$, and since $x \neq p$, $x < p$. Similarly since $kx = p$ for positive integers x and p, and k is an integer, we can say that $k \mid p$, so $k \leq p$. But since $x \neq 1$ and $kx = p$, $k \neq p$; so $k < p$. Since we have $kx = p$, $x < p$, and $k < p$ for positive integers x and k, this contradicts the fact that p is prime. □

6.2.2 Existence and uniqueness of prime factorisations

Theorem. *(Fundamental theorem of arithmetic) For every integer $n > 1$ there are unique prime numbers $p_1, p_2, \ldots p_k$ such that $p_1 \leq p_2 \leq \cdots \leq p_k$ and $n = p_1 p_2 \cdots p_k$.*

Proof. By an earlier theorem, every integer greater than 1 is either prime or a product of primes. Listing the primes from smallest to largest gives us the required nondecreasing prime factorisation. Uniqueness of the factorisation follows from the second last theorem in the previous subsection. □

If we write the product of a list of prime numbers $p_1, p_2, \ldots, p_k$ in the form $1 \cdot p_1, p_2, \ldots, p_k$, then it is natural to introduce the convention that the product of the empty list is 1. With this convention we can extend the fundamental theorem of arithmetic to say that every positive integer has a unique prime factorisation, where the factorisation of the number 1 is the product of the empty list of prime numbers.
(next page)

Computation
The most straightforward way to find the prime factorisation of a positive integer is to search for its smallest prime divisor, factor it out, and repeat untill all factors are prime. For an example, the prime factorisations of 275, 276, and 277 are:

$$\begin{aligned} 275 &= 5 \cdot 55 = 5 \cdot 5 \cdot 11 \\ 276 &= 2 \cdot 138 = 2 \cdot 2 \cdot 69 = 2 \cdot 2 \cdot 3 \cdot 23 \\ 277 &= 277 \end{aligned}$$

When there are repeated primes in the prime factorisation of an integer, we often use exponent notation; in this case $275 = 5^2 \cdot 11$ and $276 = 2^2 \cdot 3 \cdot 23$.

More generally, we can write the prime factorisation of a positive integer n in the form $n = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$, where $p_1, p_2, \ldots, p_k$ are prime numbers, $p_1 < p_2 < \cdots < p_k$, and $e_1, e_2, \ldots, e_k$ are positive integers. Again, by the fundmaental theorem of arithmetic, this representation of n is unique.

Divisors
The theorem can provide insight into a number of concepts in number theory. Suppose n and d are positive integers and $d \mid n$. Then there is some positive integer c such that $cd = n$. Now let the prime factorisations of c and d be $c = p_1 p_2 \cdots p_k$ and $d = q_1 q_2 \cdots q_m$. Then $n = cd = p_1 p_2 \cdots p_k q_1 q_2 \cdots q_m$. If we rearrange the primes into nondecreasing order, then this must be the unique prime factorisation of n.

Rephrasing this using exponent notation, suppose the prime factorisation of n is $n = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$. Then the divisors of n are precisely the numbers of the form $p_1^{f_1} p_2^{f_2} \cdots p_k^{f_k}$, where for all $i \in \{1, 2, \ldots, k\}$, $0 \leq f_i \leq e_i$.

This can be applied to greatest common divisors. Suppose a and b are positive integers. Let $p_1, p_2, \cdots, p_k$ be a list of all primes that occur in the prime factorisation of *either* a or b. We can write a and b in the form

$$a = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}, \quad b = p_1^{f_1} p_2^{f_2} \cdots p_k^{f_k}$$

Where some of the exponnenets e_i and f_i might be 0. The common divisors of a and b are all numbers of the form $p_1^{g_1} p_2^{g_2} \cdots p_k^{g_k}$, where for every $i \in \{1, \ldots, k\}$, $g_i \leq e_i$ and $g_i \leq f_i$ (if either condition is breached the number is no longer a common divisor). The greatest common divisor can be found by letting each g_i have the largest possible value without violating the two conditions. This is $\min(e_i, f_i)$, which is the minimum of e_i and f_i. In other words

$$\gcd(a, b) = p_1^{\min(e_1, f_1)} p_2^{\min(e_2, f_2)} \cdots p_k^{\min(e_k, f_k)}$$

(next page)

Cont.
For example, we could calculate $\gcd(1392, 444) = 12$. Their prime factors can be written as

$$1392 = 2^4 \cdot 3 \cdot 29 = 2^4 \cdot 3^1 \cdot 29^1 \cdot 37^0$$
$$444 = 2^2 \cdot 3 \cdot 37 = 2^2 \cdot 3^1 \cdot 29^0 \cdot 37^1$$

so the gcd is

$$\begin{aligned}\gcd(1392, 444) &= 2^{\min(4,2)} \cdot 3^{\min(1,1)} \cdot 29^{\min(1,0)} \cdot 37^{\min(0,1)} \\ &= 2^2 \cdot 3^1 \cdot 29^0 \cdot 37^0 = 12\end{aligned}$$

Usually the Euclidean algorithm is a more efficient way to find the gcd compared to prime factorisation, but if one already knows the prime factorisation, the gcd can be computed easily.

Least common multiple
For any positive integers a and b, the *least common multiple* of a and b, denoted $\operatorname{lcm}(a, b)$, is the smallest possible integer m such that $a \mid m$ and $b \mid m$. Suppose that as before,

$$a = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}, \quad b = p_1^{f_1} p_2^{f_2} \cdots p_k^{f_k}$$

For each $i \in \{1, \ldots, k\}$, any common multiple of a and b must include a factor $p_i^{g_i}$ in its prime factorisation, where $g_i \geq e_i$ and $g_i \geq f_i$ (the common multiple must contain the prime factorisations of both a and b). The smallest possible value of each g_i is the maximum of e_i and f_i, which we will denote $\max(e_i, f_i)$, so

$$\operatorname{lcm}(a, b) = p_1^{\max(e_1,f_1)} p_2^{\max(e_2,f_2)} \cdots p_k^{\max(e_k,f_k)}$$

It is not hard to show that for any numbers e and f, $\min(e, f)+\max(e, f) = e+f$ (by cases). So

$$\begin{aligned}&\gcd(a, b) \cdot \operatorname{lcm}(a, b) \\ &= \left(p_1^{\min(e_1,f_1)} p_2^{\min(e_2,f_2)} \cdots p_k^{\min(e_k,f_k)}\right) \cdot \left(p_1^{\max(e_1,f_1)} p_2^{\max(e_2,f_2)} \cdots p_k^{\max(e_k,f_k)}\right) \\ &= p_1^{\min(e_1,f_1)+\max(e_1,f_1)} \cdots p_k^{\min(e_k,f_k)+\max(e_k,f_k)} \\ &= p_1^{e_1+f_1} \cdots p_k^{e_k+f_k} \\ &= (p_1^{e_1} \cdots p_k^{e_k}) \cdot (p_1^{f_1} \cdots p_k^{f_k}) = ab\end{aligned}$$

This gives us another way to compute $\operatorname{lcm}(a, b)$:

$$\operatorname{lcm}(a, b) = \frac{ab}{\gcd(a, b)}$$

(next page)

Example
We can compute $\text{lcm}(1392, 444)$ two ways:

$$\begin{aligned}\text{lcm}(1392, 444) &= 2^{\max(4,2)} \cdot 3^{\max(1,1)} \cdot 29^{\max(1,0)} \cdot 37^{\max(0,1)} \\ &= 2^4 \cdot 3^1 \cdot 29^1 \cdot 37^1 = 51504\end{aligned}$$

and

$$\text{lcm}(1392, 444) = \frac{1392 \cdot 444}{\gcd(1392, 444)} = \frac{618048}{12} = 51504$$

6.3 Modular Arithmetic

Definition. Suppose m is a positive integer. For any integers x and y, we will say that x is *congruent to y modulo m* if $\exists k \in \mathbb{Z}(x - y = km)$. In other words, x is congruent to y modulo m iff $m \mid (x - y)$. We will use the notation $x \equiv y \pmod{m}$ or $\equiv_m$ to mean that x is congruent to y modulo m.

First consider an extension of the earlier proof of the division algorithm.

Theorem. *Suppose n and m are integers and $m > 0$. Prove that there are integers q and r such that $n = qm + r$ and $0 \leq r < m$.*

Proof. We want to show

$$\forall n \in \mathbb{Z} \forall m \in \mathbb{Z}(m > 0 \rightarrow \exists q \in \mathbb{Z} \exists r \in \mathbb{Z}(n = qm + r \wedge 0 \leq r < m))$$

Let n be some arbitrary integer.

Case 1: $n \geq 0$. We proceed by strong induction. Suppose that for all integers k where $0 \leq k < n$, the division algorithm is satisfied. Let m be some positive integer.

Case 1a: $n < m$. This case is trivial. Letting $q' = 0$ and $r' = n < m$, $n = q'm + r'$.

Case 1b: $n \geq m$. Since $m > 0$, $0 \leq n - m < n$. By the inductive hypothesis $n - m$ satisfies the division algorithm, meaning that we can declare integers q and r such that $n - m = qm + r$ and $0 \leq r < m$. We can simplify the equality to $n = (q + 1)m + r$. Taking $q' = q + 1$ and $r' = r$, $n = q'm + r'$ for integers q' and r' and $0 \leq r' < m$. We've now shown that the division algorithm holds for $n \geq 0$.

Case 2: $n < 0$. Then $-n > 0$ and since n is an integer, $-n - 1 \geq 0$. Let m be some positive integer. Then by the previous case we can declare integers q and r such that $-n - 1 = qm + r$, where $0 \leq r < m$. The equality is equivalent to $n = -qm - r - 1 = -qm - m + m - r - 1 = (-q - 1)m + (m - r - 1)$. Since $r \geq 0$, $m - r \leq m$; and since $r < m$, $0 < m - r$. Together we get $0 < m - r \leq m$ and $0 \leq m - r - 1 < m$. Taking $q' = -q - 1$ and $r' = m - r - 1$, we now have integers q' and r' where $n = q'm + r'$ and $0 \leq r' < m$. This completes the proof. □

(next page)

Intuition
Consider the earlier proven theorems for reference:

Theorem. *For every positive integer m, $\equiv_m$ is an equivalence relation on $\mathbb{Z}$.*

Definition. Suppose A is a set and $\mathcal{F} \subseteq \mathscr{P}(A)$. We will say that $\mathcal{F}$ is *pairwise disjoint* if every pair of distinct elements of $\mathcal{F}$ are disjoint, or in other words $\forall X \in \mathcal{F} \forall Y \in \mathcal{F}(X \neq Y \to X \cap Y = \emptyset)$. $\mathcal{F}$ is called a *partition* of A if it has the following properties:

1. $\bigcup \mathcal{F} = A$
2. $\mathcal{F}$ is pairwise disjoint
3. $\forall X \in \mathcal{F}(X \neq \emptyset)$

Theorem. *Suppose R is an equivalence relation on set A. Then A/R is a partition of A.*

We've proven earlier that $\equiv_m$ is an equivalence relation on $\mathbb{Z}$. For any integer a, let $[a]_m$ be the equivalence class of a with respect to the equivalence relation $\equiv_m$. The set of all these equivalence classes is denoted $\mathbb{Z}/\equiv_m$:

$$[a]_m = \{b \in \mathbb{Z} \mid b \equiv_m a\}, \quad \mathbb{Z}/\equiv_m = \{[a]_m \mid a \in \mathbb{Z}\}$$

Since $\equiv_m$ is an equivalence relation on $\mathbb{Z}$, $\mathbb{Z}/\equiv_m$ is a partition of $\mathbb{Z}$.

Consider the case $m = 3$. Three equivalence classes partition the integers:

$$[0]_3 = \{b \in \mathbb{Z} \mid b \equiv 0 \pmod 3\} = \{0, 3, 6, 9, \ldots, -3, -6, -9\}$$
$$[1]_3 = \{b \in \mathbb{Z} \mid b \equiv 1 \pmod 3\} = \{1, 4, 7, 10, \ldots, -2, -5, -8\}$$
$$[2]_3 = \{b \in \mathbb{Z} \mid b \equiv 2 \pmod 3\} = \{2, 5, 8, 11, \ldots, -1, -4, -7\}$$

Every integer is congruent modulo 3 to exactly one of the numbers 0, 1, and 2. This is an instance of the following general theorem.
(next page)

Theorem. *Suppose m is a positive integer. Then for every integer a there is exactly one integer r such that $0 \leq r < m$ and $a \equiv r \pmod{m}$.*

Proof. Let a be an arbitrary integer. Let q and r be the quotient and remainder when a is divided by m (guaranteed by the previous theorem). So $a = qm + r$ and $0 \leq r < m$. Then $a - r = qm$, and $a \equiv_m r$. This proves existence.

To prove uniqueness, suppose r_1 and r_2 are integers such that $0 \leq r_1 < m$, $0 \leq r_2 < m$, $a \equiv_m r_1$ and $a \equiv_m r_2$. Then by symmetry and transitivity of the equivalence relation $\equiv_m$, $r_1 \equiv_m r_2$. So we can declare integer d such that $r_1 - r_2 = dm$. Now since $0 \leq r_1 < m$ and $-m < -r_2 \leq 0$, we have $r_1 - r_2 \leq r_1 < m$ and $r_1 - r_2 \geq 0 - r_2 > -m$. Together we have $-m < r_1 - r_2 < m$. So $-m < dm < m$ and $-1 < d < 1$. Since d is an integer, we must have $d = 0$. Therefore $r_1 - r_2 = dm = 0$, and $r_1 = r_2$. □

This is an existence and uniqueness proof. We prove the two separately.

The theorem says that every integer is modulo m to exactly one element of the set $\{0, 1, \ldots, m-1\}$. We say this set is a *complete residue system modulo m*. Note that since $\equiv_m$ is an equivalence relation; as derived earlier, this means

$$a \equiv_m r \quad \text{iff} \quad a \in [r]_m \quad \text{iff} \quad [a]_m = [r]_m$$

and that every equivalence class in $\mathbb{Z}/\equiv_m$ is equal to exactly one of the equivalence classes in the list $[0]_m, [1], \ldots, [m-1]_m$. Thus these m equivalent classes are distinct, and $\mathbb{Z}/\equiv_m = \{[0]_m, [1]_m, \ldots, [m-1]_m\}$.

6.3.1 Sums and products between equivalence classes

Lemma. *Suppose m is a positive integer. Then for all integers a, a', b, and b', if $a' \equiv_m a$ and $b' \equiv_m b$, then $a' + b' \equiv_m a + b$ and $a'b' \equiv_m ab$.*

Proof. Suppose $a' \equiv_m a$ and $b' \equiv_m b$. Then we can declare integers k_0 and k_1 such that $a' - a = k_0 m$ and $b' - b = k_1 m$. Thus we have $a' = k_0 m + a$ and $b' = k_1 m + b$.

We have $a'+b' = (k_0+k_1)m+(a+b)$, and then $(a'+b')-(a+b) = (k_0+k_1)m$. So $a' + b' \equiv_m a + b$.

We also have $a'b' = (k_0 m + a)(k_1 m + b) = k_0 k_1 m^2 + k_0 bm + k_1 am + ab = (k_0 k_1 m + k_0 b + k_1 a)m + ab$. So

$$a'b' - ab = (k_0 k_1 m + k_0 b + k_1 a)m$$

and $a'b' \equiv_m ab$. □

Consider any two equivalence classes X and Y in $\mathbb{Z}/\equiv_m$. Something suprising happens if we add or multiply elements of X and Y. It turns out that all sums of the form $x+y$, where $x \in X$ and $y \in Y$, belong to the same equivalence class, and all products xy also belong to the same equivalence class. In other words:

Theorem. *Suppose m is a positive integer and X and Y are elements of $\mathbb{Z}/\equiv_m$. Then:*

1. *There is a unique $S \in \mathbb{Z}/\equiv_m$ such that $\forall x \in X \forall y \in Y (x + y \in S)$.*
2. *There is a unique $P \in \mathbb{Z}/\equiv_m$ such that $\forall x \in X \forall y \in Y (xy \in P)$.*

Proof. Since X and Y are elements of $\mathbb{Z}/\equiv_m$, we can let a and b be integers such that $X = [a]_m$ and $Y = [b]_m$.

We start with part 1. Since $a + b$ is an integer, we can let $S = [a + b]_m$. Now let $x \in X$ and $y \in Y$ be arbitrary, then $x \in [a]_m$ and $y \in [b]_m$, so $x \equiv_m a$ and $y \equiv_m b$. By the previous lemma, it follows that $x + y \equiv_m a + b$, and $x + y \in [a + b]_m = S$. Since x and y were arbitrary, we conclude that $\forall x \in X \forall y \in Y (x + y \in S)$.

Now suppose S' is another equivalence class such that $\forall x \in X \forall y \in Y (x+y \in S')$. Since $a \in X$ and $b \in Y$, we have $a+b \in S$ and $a+b \in S'$. Therefore S and S' are not disjoint, and since $\mathbb{Z}/\equiv_m$ is a partition, this implies that $S = S'$.
(next page)

Now for part 2. We know that ab is an integer, so we can let $p = [ab]_m$. Let $x \in X$ and $y \in Y$ be arbitrary. Then $x \in [a]_m$ and $y \in [b]_m$ so $x \equiv_m a$ and $y \equiv_m b$. By the previous lemma it follows then that $xy \equiv_m ab$, and $xy \in [ab]_m$. Since x and y were arbitrary, we conclude that $\forall x \in X \forall y \in Y(xy \in [ab]_m)$.

Now suppose P' is some arbitrary set such that $\forall x \in X \forall y \in Y(xy \in P')$. Since $a \in X$ and $b \in Y$ we have $ab \in P'$; but we also have $ab \in [ab]_m$, so P' and $[ab]_m$ are not disjoint. Since $\mathbb{Z}/\equiv_m$ is a partition, we then must have $P' = [ab]_m$. We are done. □

Definition. Suppose X and Y are elements of $\mathbb{Z}/\equiv_m$. Then we define the sum and product of X and Y, denoted by $X + Y$ and $X \cdot Y$, as follows:

$$X + Y = \text{the unique } S \in \mathbb{Z}/\equiv_m \text{ such that } \forall x \in X \forall y \in Y(x + y \in S)$$
$$X \cdot Y = \text{the unique } P \in \mathbb{Z}/\equiv_m \text{ such that } \forall x \in X \forall y \in Y(xy \in P)$$

The proof shows that if $X = [a]_m$ and $Y = [b]_m$, then the sum of X and Y is the equivalence class $S = [a + b]_m$ and the product is $P = [ab]_m$. As such we have the following theorem.

Theorem. *For any positive integer m and any integers a and b,*

$$[a]_m + [b]_m = [a + b]_m \quad \textit{and} \quad [a]_m \cdot [b]_m = [ab]_m$$

To demonstrate these ideas, consider the case $m = 5$. We know that every element of $\mathbb{Z}/\equiv_5$ is equal to either $[0]_5, [1]_5, [2]_5, [3]_5$, or $[4]_5$. We will often choose to write equivalence classes in one of these forms. For example, $[2]_5 + [4]_5 = [6]_5$, but since $6 \equiv_5 1$, $[6]_5 = [1]_5$, and we can say that $[2]_5 + [4]_5 = [1]_5$. Similarly $[2]_5 \cdot [4]_5 = [8]_5 = [3]_5$. The following shows the addition and multiplication tables for $\mathbb{Z}/\equiv_5$.

+	$[0]_5$	$[1]_5$	$[2]_5$	$[3]_5$	$[4]_5$
$[0]_5$	$[0]_5$	$[1]_5$	$[2]_5$	$[3]_5$	$[4]_5$
$[1]_5$	$[1]_5$	$[2]_5$	$[3]_5$	$[4]_5$	$[0]_5$
$[2]_5$	$[2]_5$	$[3]_5$	$[4]_5$	$[0]_5$	$[1]_5$
$[3]_5$	$[3]_5$	$[4]_5$	$[0]_5$	$[1]_5$	$[2]_5$
$[4]_5$	$[4]_5$	$[0]_5$	$[1]_5$	$[2]_5$	$[3]_5$

·	$[0]_5$	$[1]_5$	$[2]_5$	$[3]_5$	$[4]_5$
$[0]_5$	$[0]_5$	$[0]_5$	$[0]_5$	$[0]_5$	$[0]_5$
$[1]_5$	$[0]_5$	$[1]_5$	$[2]_5$	$[3]_5$	$[4]_5$
$[2]_5$	$[0]_5$	$[2]_5$	$[4]_5$	$[1]_5$	$[3]_5$
$[3]_5$	$[0]_5$	$[3]_5$	$[1]_5$	$[4]_5$	$[2]_5$
$[4]_5$	$[0]_5$	$[4]_5$	$[3]_5$	$[2]_5$	$[1]_5$

6.3.2 Properties of operations between equivalence classes

How do addition and multiplication in $\mathbb{Z}/\equiv_m$ compare to addition and multiplication in $\mathbb{Z}$? Many properties of addition and multiplication in $\mathbb{Z}$ carry over easily to $\mathbb{Z}/\equiv_m$.

Theorem. *Suppose m is a positive integer. Then for all equivalence classes X, Y, and Z in $\mathbb{Z}/\equiv_m$:*

1. $X + Y = Y + X$ *(Addition is commutative).*
2. $(X + Y) + Z = X + (Y + Z)$ *(Addition is associative).*
3. $X + [0]_m = X$ *($[0]_m$ is an identity element for addition).*
4. *There is some $X' \in \mathbb{Z}/\equiv_m$ such that $X + X' = [0]_m$ (X has an additive inverse).*
5. $X \cdot Y = Y \cdot X$ *(Multiplication is commutative).*
6. $(X \cdot Y) \cdot Z = X \cdot (Y \cdot Z)$ *(Multiplication is associative).*
7. $X \cdot [1]_m = X$ *($[1]_m$ is an identity element for multiplication).*
8. $X \cdot [0]_m = [0]_m$.
9. $X \cdot (Y + Z) = (X \cdot Y) + (X \cdot Z)$ *(Multiplication distributes over addition).*

Proof. Since $X, Y, Z \in \mathbb{Z}/\equiv_m$, there are integers a, b, and c such that $X = [a]_m$, $Y = [b]_m$, and $Z = [c]_m$.

1. We have

$$X + Y = [a]_m + [b]_m = [a + b]_m = [b + a]_m = [b]_m + [a]_m = Y + X$$

   Where we have used the commutativity of addition of integers.

2. We have

$$\begin{aligned}(X + Y) + Z &= ([a]_m + [b]_m) + [c]_m \\ &= [a + b]_m + [c]_m \\ &= [(a + b) + c]_m \\ &= [a + (b + c)]_m \\ &= [a]_m + [b + c]_m \\ &= [a]_m + ([b]_m + [c]_m) = X + (Y + Z)\end{aligned}$$

3. We have

$$X + [0]_m = [a]_m + [0]_m = [a + 0]_m = [a]_m = X$$

(next page)

4. Letting $X' = [-a]_m$,

$$X + X' = [a]_m + [-a]_m = [a + (-a)]_m = [0]_m$$

5. We have

$$X \cdot Y = [a]_m \cdot [b]_m = [ab]_m = [ba]_m = [b]_m \cdot [a]_m = Y \cdot X$$

6. We have

$$\begin{aligned}(X \cdot Y) \cdot Z &= ([a]_m \cdot [b]_m) \cdot [c]_m = [ab]_m \cdot [c]_m = [(ab)c]_m = [a(bc)]_m \\ &= [a]_m \cdot [bc]_m = [a]_m \cdot ([b]_m \cdot [c]_m) = X \cdot (Y \cdot Z)\end{aligned}$$

   Where we have used the associativity of multiplication of integers.

7. We have

$$X \cdot [1]_m = [a]_m \cdot [1]_m = [a \cdot 1]_m = [a]_m = X$$

8. We have

$$X \cdot [0]_m = [a]_m \cdot [0]_m = [a \cdot 0]_m = [0]_m$$

9. We have

$$\begin{aligned}X \cdot (Y + Z) &= [a]_m \cdot ([b]_m + [c]_m) = [a]_m \cdot [b + c]_m = [a(b + c)]_m \\ &= [ab + ac]_m = [ab]_m + [ac]_m = ([a]_m \cdot [b]_m) + ([a]_m \cdot [c]_m) \\ &= (X \cdot Y) + (X \cdot Z)\end{aligned}$$ □

Theorem. *Suppose m is a positive integer. The additive identity in $Z/\equiv_m$ is unique.*

Proof. We have already demonstrated existence. We now show uniqueness. That is:

$$\begin{aligned}\forall Z_1 \in \mathbb{Z}/\equiv_m \forall Z_2 \in \mathbb{Z}/\equiv_m ([\forall X \in \mathbb{Z}/\equiv_m (X + Z_1 = X) \\ \wedge \forall X \in \mathbb{Z}/\equiv_m (X + Z_2 = X)] \rightarrow Z_1 = Z_2)\end{aligned}$$

Let Z_1 and Z_2 be arbitrary elements of $\mathbb{Z}/\equiv_m$. Suppose that they are both additive identities. Then $Z_2 + Z_1 = Z_2$ and $Z_1 + Z_2 = Z_1$, meaning

$$Z_1 = Z_1 + Z_2 = Z_2 + Z_1 = Z_2$$

Using commutativity of addition between equivalence classes. □

(next page)

Theorem. *Suppose m is a positive integer. Any element $X \in \mathbb{Z}/\equiv_m$ has a unique additive inverse.*

Proof. Let X be some arbitrary element in $Z/\equiv_m$. We have already demonstrated existence. Now we show uniqueness. That is

$$\forall Z_1 \in \mathbb{Z}/\equiv_m \forall Z_2 \in \mathbb{Z}/\equiv_m ([X + Z_1 = [0]_m \wedge X + Z_2 = [0]_m] \\ \rightarrow Z_1 = Z_2)$$

Suppose Z_1 and Z_2 are arbitrary elements of $\mathbb{Z}/\equiv_m$ such that $X + Z_1 = [0]_m$ and $X + Z_2 = [0]_m$. Then $Z_1 + [0]_m = Z_1 + X + Z_2$ and $[0] + Z_2 = Z_1 + X + Z_2$ (using commutativity). So since $[0]_m$ is the additive identity,

$$Z_1 = Z_1 + [0]_m = Z_1 + X + Z_2 = [0] + Z_2 = Z_2 \qquad \square$$

Theorem. *Suppose m is a positive integer. The multiplicative identity in $\mathbb{Z}/\equiv_m$ is unique.*

Proof. We have already demonstrated existence. Now we show uniqueness. That is:

$$\forall Z_1 \in /\equiv_m \forall Z_2 \in \mathbb{Z}/\equiv_m ([\forall X \in \mathbb{Z}/\equiv_m (X \cdot Z_1 = X) \\ \forall X \in \mathbb{Z}/\equiv_m (X \cdot Z_2 = X)] \rightarrow Z_1 = Z_2)$$

Let Z_1 and Z_2 be arbitrary elements of $\mathbb{Z}/\equiv_m$ and suppose they are both multiplicative identities. Then $Z_1 \cdot Z_2 = Z_1$ and $Z_2 \cdot Z_1 = Z_2$, so

$$Z_1 = Z_1 \cdot Z_2 = Z_2 \cdot Z_1 = Z_2$$

Using commutativity of products of equivalence classes. $\square$

We've shown that the identity elements for addition ($[0]_m$) and multiplication ($[1]_m$) are unique. We've also shown that the additive identity always exists and is unique. We will denote the additive inverse of X by $-X$. For example, $[4]_5 + [1]_5 = [0]_5$, so $-[4]_m = [1]_5$.

What about multiplicative inverses? If $X \in \mathbb{Z}/\equiv_m$, $X' \in \mathbb{Z}/\equiv_m$, and $X \cdot X' = [1]_m$, then we say that X' is a multiplicative inverse of X. For instance recall the multiplication table for $\mathbb{Z}/\equiv_5$; every element except $[0]_5$ has a multiplicative inverse. We can show that multiplicative inverses, when they exist, are unique. (next page)

Theorem. *Suppose m is a positive integer. For any element X in $\mathbb{Z}/\equiv_m$, assuming X has a multiplicative inverse, this inverse is unique.*

Proof. Assuming that a multiplicative inverse exists, we show

$$\forall Z_1 \in \mathbb{Z}/\equiv_m \forall Z_2 \in \mathbb{Z}/\equiv_m ([X \cdot Z_1 = [1]_m \wedge X \cdot Z_2 = [1]_m] \to Z_1 = Z_2)$$

Suppose Z_1 and Z_2 are arbitrary elements of $\mathbb{Z}/\equiv_m$ such that they are multiplicative inverses of X. Then $Z_2 \cdot [1]_m = Z_2 \cdot (X \cdot Z_1)$ and $[1] \cdot Z_1 = (Z_2 \cdot X) \cdot Z_1$. So since $[1]_m$ is the multiplicative identity, using commutativity and associativity:

$$Z_1 = [1]_m \cdot Z_1 = (Z_2 \cdot X) \cdot Z_1 = Z_2 \cdot (X \cdot Z_1) = Z_2 \cdot [1]_m = Z_2 \qquad \square$$

In general, if $X \in \mathbb{Z}/\equiv_m$, then the multiplicative inverse, when it exists, is denoted X^{-1}. When does an equivalence class have a multiplicative inverse? First recall the theorem.

Theorem. *For all positive integers a and b there are integers s and t such that* $\gcd(a, b) = sa + tb$.

Theorem. *Suppose that a and m are positive integers. Then $[a]_m$ has a multiplicative inverse iff m and a are relatively prime.*

Proof. Suppose first that $[a]_m$ has a multiplicative inverse, which we can write as $[a']_m$ for some $a' \in \mathbb{Z}$, then $[a]_m \cdot [a']_m = [aa']_m = [1]_m$. So we have $aa' \equiv_m 1$; this means we can declare some $k \in \mathbb{Z}$ such that $aa' - 1 = km$, or equivalently $1 = -km + aa'$. Thus 1 is a linear combination of m and a, and it follows that m and a are relatively prime.

For the other direction, assume that m and a are relatively prime. Then there are positive integers s and t such that $sm + ta = 1$. Therefore $ta - 1 = -sm$, and $ta \equiv_m 1$. We conclude that $[a]_m \cdot [t]_m = [at]_m = [1]_m$, so $[t]_m$ is the multiplicative inverse of $[a]_m$. $\square$

This theorem shows that for any positive integers m and a, we can use the extended euclidean algorithm to find $[a]_m^{-1}$; if the algorithm shows that $\gcd(m, a) \neq 1$, then $[a]^{-1}$ doesn't exist; and if we find that $\gcd(m, a) = 1 = sm + ta$ then $[a]_m^{-1} = [t]_m$.

6.3.3 Application

Example
A class has 25 students. The teacher bought several cartons of eggs, each containing a dozen eggs, and then distributed the eggs among the students. After giving an equal number of eggs to each student, she had 7 eggs left over. What is the smallest number of cartons of eggs she could have bought?

We must find the smallest positive integer x satisfying $25 \mid (12x - 7)$, which means $12x \equiv_{25} 7$. We can turn this into an equation by working with equivalence classes. Our congruence is equivalent to the equation $[12]_{25} \cdot [x]_{25} = [7]_{25}$. First we try to find the multiplicative inverse of $[12]_{25}$. Applying the extended euclidean algorithm, we find that $\gcd(25, 12) = 1 = 1 \cdot 25 - 2 \cdot 12$ (25 and 12 are relatively prime so the inverse exists!), so $[12]_{25}^{-1} = [-2]_{25} = [23]_{25}$. See now that

$$\begin{aligned}
[12]_{25} \cdot [x]_{25} &= [7]_{25} \\
[12]_{25}^{-1} \cdot ([12]_{25} \cdot [x]_{25}) &= [12]_{25}^{-1} \cdot [7]_{25} \\
([12]_{25}^{-1} \cdot [12]_{25}) \cdot [x]_{25} &= [23]_{25} \cdot [7]_{25} \\
[1]_{25} \cdot [x]_{25} &= [161]_{25} = [11]_{25} \\
[x]_{25} &= [11]_{25}
\end{aligned}$$

We are trying to solve $[12]_{25} \cdot [x]_{25} = [7]_{25}$. As such any element of $[11]_25$ are postitive integer solutions, and the smallest positive integer solution would then be $x = 11$.

See that we were lucky in this example that 25 and 12 were relatively prime, since the multiplicative inverse was necessary in our manipulation. The next subsection elaborates a bit more on ways around this.

6.3.4 Alternative approaches

Theorem. *Suppose m and a are positive integers, and let $d = \gcd(m, a)$. Then for every integer b, if $d \nmid b$ then thre is no integer x such that $ax \equiv_m b$.*

Proof. Since $d = \gcd(m, a)$, we can declare integers k_0, k_1 such that $m = k_0 d$ and $a = k_1 d$. Let b be some arbitrary integer, and suppose $d \nmid b$. Now, for contradiction, suppose there exists some integer x such that $ax \equiv_m b$. Then we can declare the integer k_2 such that $ax - b = k_2 m$. So we have $b = ax - k_2 m = xk_1 d - k_2 k_0 d = (xk_1 - k_2 k_0)d$; this would contradict the assumption that $d \nmid b$, so we are done. □

Theorem. *Suppose n and m are positive integers. Then for all integers a and b,*

$$na \equiv_{nm} nb \leftrightarrow a \equiv_m b$$

Proof. We have the string of equivalences:

$$\begin{aligned} na \equiv_{nm} nb &\leftrightarrow \exists k \in \mathbb{Z}(na - nb = knm) \\ &\leftrightarrow \exists k \in \mathbb{Z}(a - b = km) \leftrightarrow a \equiv_m b \end{aligned}$$

□

Example
Say we want to solve the following congruences:

$$77x \equiv_{374} 120, \quad 77x \equiv_{374} 121$$

We begin by computing $\gcd(374, 77) = 11$ (so $[77]_{374}$ has no multiplicative inverse and our earlier strategy doesn't work). Since $11 \nmid 120$, the first theorem tells us that the first congruence $77x \equiv_{374} 120$ has no solution.

To solve the second congruence (see that the first theorem doesn't apply in this case), we first rewrite it as $11 \cdot 7x \equiv_{374} 11 \cdot 11$. By the second theorem this is equivalent to $7x \equiv_{34} 11$. Now we have $\gcd(34, 7) = 1 = -1 \cdot 34 + 5 \cdot 7$, so $[7]_{34}^{-1} = [5]_{34}$ and the earlier strategy can be applied:

$$\begin{aligned} 7x \equiv_{34} 11 &\leftrightarrow [7]_{34} \cdot [x]_{34} = [11]_{34} \\ &\leftrightarrow [x]_{34} = [7]_{34}^{-1} \cdot [11]_{34} = [5]_{34} \cdot [11]_{34} = [55]_{34} = [21]_{34} \\ &\leftrightarrow x \in [21]_{34} \end{aligned}$$

Thus the solutions to the second congruence are the elements of $[21]_{34}$.

6.4 Euler's theorem

For some positive integer m, we saw that some elements of $\mathbb{Z}/\equiv_m$ have multiplicative inverses and some don't. Here we focus on the ones that do. We let $(\mathbb{Z}/\equiv_m)^*$ denote the set of elements of $\mathbb{Z}/\equiv_m$ that have multiplicative inverses. That is

$$(\mathbb{Z}/\equiv_m)^* = \{X \in \mathbb{Z}/\equiv_m | \ \exists X' \in \mathbb{Z}/\equiv_m \ (X \cdot X' = [1]_m)\}$$

The number of elements of $(\mathbb{Z}/\equiv_m)^*$ is denoted $\varphi(m)$. The function φ is called *Euler's phi function*, or *Euler's totient function.*

For every positive integer m, $(\mathbb{Z}/\equiv_m)^* \subseteq \mathbb{Z}/\equiv_m$ and $\mathbb{Z}/\equiv_m$ has m elements, so $\varphi(m) \leq m$. Also $[1]_m \cdot [1]_m = [1]_m$, so $[1]_m \in (\mathbb{Z}/\equiv_m)^*$ and therefore $\varphi(m) \geq 1$. For example,

$$(\mathbb{Z}/\equiv_{10})^* = \{[1]_{10}, [3]_{10}, [7]_{10}, [9]_{10}\}$$

so $\varphi(10) = 4$. Two important properties of $(\mathbb{Z}/\equiv_m)^*$ are closure under inverses and multiplication.

Theorem. *Suppose m is a positive integer.*

1. *For every X in $(\mathbb{Z}/\equiv_m)^*$, $X^{-1} \in (\mathbb{Z}/\equiv_m)^*$.*
2. *For every X and Y in $(\mathbb{Z}/\equiv_m)^*$, $X \cdot Y \in (\mathbb{Z}/\equiv_m)^*$.*

Proof.

1. Suppose $X \in (\mathbb{Z}/\equiv_m)^*$. Then X has a multiplicative inverse X^{-1} (that has been proven to be unique) and $X \cdot X^{-1} = [1]_m$. But this equation also tells us that X is the multiplicative inverse of X^{-1}; that is $(X^{-1})^{-1} = X$. Therefore $X^{-1} \in (\mathbb{Z}/\equiv_m)^*$.

2. Suppose $X \in (\mathbb{Z}/\equiv_m)^*$ and $Y \in (\mathbb{Z}/\equiv_m)^*$. Then X and Y have multiplicative inverses X^{-1} and Y^{-1}. Therefore

   $$(X \cdot Y) \cdot (X^{-1} \cdot Y^{-1}) = (X \cdot X^{-1}) \cdot (Y \cdot Y^{-1}) = [1]_m \cdot [1]_m = [1]_m$$

   This means that $X^{-1} \cdot Y^{-1}$ is the multiplicative inverse of $X \cdot Y$ and $X \cdot Y \in (\mathbb{Z}/\equiv_m)^*$. □

(next page)

Cont.
Suppose $X \in (\mathbb{Z}/\equiv_m)^*$. By the second part of the previous theorem, for every $Y \in (\mathbb{Z}/\equiv_m)^*$, $X \cdot Y \in (\mathbb{Z}/\equiv_m)^*$. We can define a function $f_X : (\mathbb{Z}/\equiv_m)^* \to (\mathbb{Z}/\equiv_m)^*$ by the formula $f_X(Y) = X \cdot Y$.

Theorem. *Defining* $f_X : (\mathbb{Z}/\equiv_m)^* \to (\mathbb{Z}/\equiv_m)^*$ *as*

$$f_X = \{(A, B) \in (\mathbb{Z}/\equiv_m)^* \times (\mathbb{Z}/\equiv_m)^* \mid B = X \cdot A\}$$

f_X *is one-to-one and onto.*

Proof. To prove f_X is one-to-one, we must show

$$\forall Y_1 \in (\mathbb{Z}/\equiv_m)^* \forall Y_2 \in (\mathbb{Z}/\equiv_m)^* (f_X(Y_1) = f_X(Y_2) \to Y_1 = Y_2)$$

Suppose Y_1 and Y_2 are elements of $(\mathbb{Z}/\equiv_m)^*$ such that $f_X(Y_1) = f_X(Y_2)$. Then $X \cdot Y_1 = X \cdot Y_2$, and therefore

$$Y_1 = [1]_m \cdot Y_1 = X^{-1} \cdot X \cdot Y_1 = X^{-1} \cdot X \cdot Y_2 = [1]_m \cdot Y_2 = Y_2$$

This proves that f_X is one-to-one.

Next we show f_X is onto, meaning

$$\forall Z \in (\mathbb{Z}/\equiv_m)^* \exists Y \in (\mathbb{Z}/\equiv_m)^* (Z = X \cdot Y)$$

Suppose $Z \in (\mathbb{Z}/\equiv_m)^*$. Then since $(\mathbb{Z}/\equiv_m)^*$ is closed under inverses and multiplication and $X \in (\mathbb{Z}/\equiv_m)^*$, $X^{-1} \cdot Z \in (\mathbb{Z}/\equiv_m)^*$. As such

$$f_X(X^{-1} \cdot Z) = X \cdot X^{-1} \cdot Z = Z$$

Since Z was arbitrary, f_X is onto. □

For example, consider the case $m = 10$ and $[X] = [3]_{10}$. Applying f_X to the four elements of $(\mathbb{Z}/\equiv_{10})^*$ gives the following

Y	$f_X(Y)$
$[1]_{10}$	$[3]_{10} \cdot [1]_{10} = [3]_{10}$
$[3]_{10}$	$[3]_{10} \cdot [3]_{10} = [9]_{10}$
$[7]_{10}$	$[3]_{10} \cdot [7]_{10} = [1]_{10}$
$[9]_{10}$	$[3]_{10} \cdot [9]_{10} = [7]_{10}$

Notice that since f_X is one-to-one and onto, each of the four elements of $(\mathbb{Z}/\equiv_{10})^*$ appears exactly once in the column under $f_X(Y)$; the entries in the second column are exactly the same as the entries in the first column, but listed in a different order.

6.4.1 Generalisation of commutativity and associativity

The commutative law for multiplication says that for any numbers a and b, $ab = ba$. The associative law says that for any numbers a, b and c, $(ab)c = a(bc)$. Here we show that these laws can be used to justify reordering and regrouping terms in a product of any list of numbers in any way.

Let us say that the *left-grouped product* of a list of numbers $a_1, a_2, \ldots, a_n$ is the product in which the terms are grouped as follows:

$$(\cdots(((a_1a_2)a_3)a_4)\cdots a_{n-1})a_n$$

More precisely, we have the recursive definition

$$\begin{aligned} &p(1) = a_1; \\ &\text{for all integers } n > 1\text{: } p(n) = p(n-1)a_n \end{aligned}$$

Theorem. *Any product of a list of numbers $a_1, a_2, \ldots, a_n$ (with the terms in that order, but with parentheses inserted to group the terms in any way) is equal to the left-grouped product.*

Proof. To simplify notation, we will assume that any product is the left grouped product unless parentheses are used to indicate otherwise. We use strong induction on n.

Suppose that the statement is true for any products of fewer than n terms, and consider the any product of $a_1, a_2, \ldots, a_n$. If $n = 1$, then the only product is the left-grouped product, so there is nothing to prove. Now suppose $n > 1$. Then our product has the form pq, where p is a product of $a_1, \ldots, a_{k-1}$ and q is a product of $a_k, \ldots, a_n$ for some k where $2 \leq k \leq n$. By the inductive hypothesis, $p = a_1 \cdots a_{k-1}$ and $q = a_k \cdots a_n$, where by our convention these two products are left-grouped. It will suffice then to prove that $(a_1 \cdots a_{k-1})(a_k \cdots a_n) = a_1 \cdots a_n$ since (any product of $a_1, a_2, \ldots a_n$)$= pq$. If $k = n$, then the left-hand side of the equation is already left grouped. If $k < n$, then

$$\begin{aligned} &(a_1 \cdots a_{k-1})(a_k \cdots a_n) \\ &\quad = (a_1 \cdots a_{k-1})((a_k \cdots a_{n-1})a_n) && \text{(Definition of left-grouped)} \\ &\quad = ((a_1 \cdots a_{k-1})(a_k \cdots a_{n-1}))a_n && \text{(Associativity)} \\ &\quad = (a_1 \cdots a_{n-1})a_n && \text{(Inductive hypothesis)} \\ &\quad = a_1 \cdots a_n && \text{(Definition of left-grouped)} \end{aligned}$$

and we are done. □

(next page)

Theorem. *Any two products of the numbers $a_1, a_2, \ldots, a_n$, with the terms in any order and grouped in any way, are equal.*

Proof. Say we have two products $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$, where the second product is some reordering of the first sequence. Using the previous theorem, both products are equal to their left-grouped products; we wil show that their left-grouped products are equal. As in the previous proof, we will assume that products are written as left-grouped unless specified otherwise.

We use induction on n. If $n = 1$ then the products are clearly equal since $a_1 = b_1$. Now suppose that the statement is true for products of length n, and suppose that $b_1 \cdots b_{n+1}$ is a reordering of $a_1 \cdots a_{n+1}$. Then b_{n+1} is one of $a_1, \ldots, a_{n+1}$.

If $b_{n+1} = a_{n+1}$ then

$$\begin{aligned} b_1 \cdots b_{n+1} &= (b_1 \cdots b_n)a_{n+1} && \text{(Definition of left-grouped)} \\ &= (a_1 \cdots a_n)a_{n+1} && \text{(Inductive hypothesis)} \\ &= a_1 \cdots a_{n+1} && \text{(Definition of left-grouped)} \end{aligned}$$

Now suppose $b_{n+1} = a_k$ for some $k \leq n$. We will write $a_1 \cdots \hat{a_k} \cdots a_n$ for the (left-grouped) product of the numbers $a_1, \ldots, a_n$ with the factor a_k left out. Then

$$\begin{aligned} b_1 \cdots b_{n+1} &= (b_1 \cdots b_n)a_k && \text{(Definition of left-grouped)} \\ &= (a_1 \cdots \hat{a_k} \cdots a_{n+1})a_k && \text{(Inductive hypothesis)} \\ &= ((a_1 \cdots \hat{a_k} \cdots a_n)a_{n+1})a_k && \text{(Definition of left-grouped)} \\ &= (a_1 \cdots \hat{a_k} \cdots a_n)(a_{n+1}a_k) && \text{(Associativity)} \\ &= (a_1 \cdots \hat{a_k} \cdots a_n)(a_k a_{n+1}) && \text{(Commutativity)} \\ &= ((a_1 \cdots \hat{a_k} \cdots a_n)a_k)a_{n+1}) && \text{(Associativity)} \\ &= (a_1 \cdots a_n)a_{n+1} && \text{(Inductive hypothesis)} \\ &= a_1 \cdots a_{n+1} && \text{(Definition of left-grouped)} \end{aligned}$$

Since the left-grouped products are equal, the reordered products are equal. □

6.4.2 The theorem

Theorem. *Suppose m is a positive integer and $X \in (\mathbb{Z}/\equiv_m)^*$. Then $X^{\varphi(m)} = [1]_m$.*

Proof. By he definition of Euler's phi function, there are $\varphi(m)$ elements in $(\mathbb{Z}/\equiv_m)^*$. Let $Y_1, Y_2, \ldots, Y_{\varphi(m)}$ be a list of these elements. Then since f_X is one-to-one and onto, each of these elements occurs exactly once in the list $f_X(Y_1), f_X(Y_2), \ldots, f_X(Y_{\varphi(m)})$. In other words, the two lists $Y_1, Y_2, \ldots, Y_{\varphi(m)}$ and $f_X(Y_1), f_X(Y_2), \ldots, f_X(Y_{\varphi(m)})$ contain exactly the same entries, but listed in different orders. It follows then, by the commutative and associative laws for multiplication, that if we multiply all of the entries in the two lists, the products will be the same:

$$\begin{aligned} Y_1 \cdot Y_2 \cdots Y_{\varphi(m)} &= f_X(Y_1) \cdot f_X(Y_2) \cdots f_X(Y_{\varphi(m)}) \\ &= (X \cdot Y_1) \cdot (X \cdot Y_2) \cdots (X \cdot Y_{\varphi(m)}) \\ &= X^{\varphi(m)} \cdot (Y_1 \cdot Y_2 \cdots Y_{\varphi(m)}) \end{aligned}$$

Where $X^{\varphi(m)}$ denotes X multiplied by itself $\varphi(m)$ times. Let $Z = Y_1 \cdot Y_2 \cdots Y_{\varphi(m)}$. Then the equation says $Z = X^{\varphi(m)} \cdot Z$. Since $(\mathbb{Z}/\equiv_m)^*$ is closed under multiplication, $Z \in (\mathbb{Z}/\equiv_m)^*$, so it has an inverse and

$$[1]_m = Z \cdot Z^{-1} = X^{\varphi(m)} \cdot Z \cdot Z^{-1} = X^{\varphi(m)} \cdot [1]_m = X^{\varphi(m)} \qquad \square$$

Theorem. *(Euler's theorem) Suppose m is a positive integer. Then for every positive integer a, if $\gcd(m, a) = 1$ then $a^{\varphi(m)} \equiv_m 1$.*

Proof. Suppose a is a positive integer and $\gcd(m, a) = 1$. Then by an earlier theorem $[a]_m$ has an inverse and therefore $[a]_m \in (\mathbb{Z}/\equiv_m)^*$. By the previous theorem, $[a]_m^{\varphi(m)} = [1]_m$. But then

$$[a]_m^{\varphi(m)} = \underbrace{[a]_m \cdot [a]_m \cdots [a]_m}_{\varphi(m) \text{ terms}} = [\underbrace{a \cdot a \cdots a}_{\varphi(m) \text{ terms}}]_m = [a^{\varphi(m)}]_m$$

Thus $[a^{\varphi(m)}]_m = [a]_m^{\varphi(m)} = [1]_m$, and therefore $a^{\varphi(m)} \equiv_m 1$. $\square$

6.4.3 Phi function computation

To apply Euler's theorem, we need to be able to compute $\varphi(m)$. One way to do this would be to check all the elements of $\mathbb{Z}/\equiv_m$ one-by-one and count how many have multiplicative inverses. This is impractical for large m. We want to find more efficient ways to compute $\varphi(m)$.

We begin by rephrasing the definition of $\varphi(m)$. We know that $\{0, 1, \ldots, m-1\}$ is a complete residue system modulo m. But since $0 \equiv_m m$, $[0]_m = [m]_m$ and we can also say that $\{1, 2, \ldots, m\}$ is a complete residue system. Thus $\mathbb{Z}/\equiv_m = \{[1]_m, [2]_m, \ldots, [m-1]_m, [m]_m\} = \{[a]_m \mid 1 \leq a \leq m\}$, where each element of $\mathbb{Z}/\equiv_m$ appears exactly once in this list of elements.

We proved that for any positive integer a, $[a]_m$ has a multiplicative inverse iff m and a are relatively prime. So

$$(\mathbb{Z}/\equiv_m)^* = \{[a]_m \mid 1 \leq a \leq m \text{ and } \gcd(m, a) = 1\}$$

Another approach for calculating Euler's phi function would then be

$\varphi(m) =$ the number of elements in the set$\{[a]_m \mid 1 \leq a \leq m$ and $\gcd(m, a) = 1\}$

Recall the theorem:

Theorem. *For a prime number p, $D(p) = \{1, p\}$.*

Using this approach, it is easy to compute $\varphi(p)$ when p is prime: if $1 \leq a \leq p-1$ then $\gcd(p, a) = 1$; but $\gcd(p, p) = p > 1$ (1 isn't prime). Therefore

$$\{a \mid 1 \leq a \leq p \text{ and } \gcd(p, a) = 1\} = \{1, 2, \ldots, p-1\}$$

So $\varphi(p) = p - 1$.

It is almost as easy to compute $\varphi(p^k)$ for any positive integer k. If a is a positive integer and $p \mid a$ then $\gcd(p^k, a) \geq p > 1$. If $p \nmid a$ then the only common divisor of p^k and a is 1 (the prime factorisation of p^k for prime p is just p^k, and since $p \nmid a$, p is not a prime factor of a, so their greatest common divisor is $p^0 = 1$.) so $\gcd(p^k, a) = 1$.

The elements of the set $\{a \mid 1 \leq a \leq p^k\}$ that are *not* relatively prime to p^k are precisely the ones that are divisible by p. Those elements are $p, 2p, 3p, \ldots, p^k$. In other words,

$$\{a \mid 1 \leq a \leq p^k \text{ and } \gcd(p^k, a) = 1\} = \{1, 2, \ldots, p^k\} \backslash \{p, 2p, \ldots, p^{k-1}p\}$$

the number of elements in this set is $p^k - p^{k-1} = p^{k-1}(p-1)$. Thus $\varphi(p^k) = p^{k-1}(p-1)$.
(next page)

Lemma.

Appendix A

Exercises

A.1 Ch 3

A.1.1 $\exists x(P(x) \to Q(x))$ is equivalent to $\forall x P(x) \to \exists x Q(x)$

Using logical equivalences:

$$\begin{aligned}
\exists x(P(x) \to Q(x)) &= \exists x[\neg(P(x) \wedge \neg Q(x))] && \text{(Definition)} \\
&= \neg\forall x[P(x) \wedge \neg Q(x)] && \text{(Quantifier negation)} \\
&= \neg[\forall x P(x) \wedge \forall x \neg Q(x)] && \text{(Distributivity over conjuction)} \\
&= \neg[\forall x P(x) \wedge \neg\exists x Q(x)] && \text{(Quantifier negation)} \\
&= \forall x P(x) \to \exists x Q(x) && \text{(Definition)}
\end{aligned}$$

A.1.2 If $\exists x(P(x) \to Q(x))$, then $\forall x P(x) \to \exists x Q(x)$

Proof. Suppose $\exists x(P(x) \to Q(x))$. Then we can choose some x_0 such that $P(x_0) \to Q(x_0)$. Now suppose that $\forall x P(x)$. Then in particular, $P(x_0)$, and since $P(x_0) \to Q(x_0)$, it follows that $Q(x_0)$. Since we have found a particular value of x for which $Q(x)$ holds, we can conclude that $\exists x Q(x)$. Thus $\forall x P(x) \to \exists x Q(x)$. □

A.1.3