Chapter 1

Logic and Proofs

Mathematics admits no "absolute truth". Instead, most mathematicians work within the axiom system known as Zermelo-Fraenkel with choice, or ZFC for short. ZFC formalizes the concept of a *set*, an abstraction of a collection of objects, called *elements*. For now, the details of ZFC are unimportant. This chapter describes the basic rules of logic. Chapter 2 provides an informal introduction to ZFC.

ZFC is believed to be logically consistent, and the "correctness" of mathematical statements is evaluated according to "provability" and "logical consistency" with respect to ZFC. Theorems proved in ZFC are said colloquially to be "true". Strictly speaking, however, mathematicians do not find metaphysical truths, but instead deduce logical *conclusions* starting from assumptions called *hypotheses*.

1.1. Statements, Negation, and Connectives

A *statement* is a sentence having a *truth value*, T (True) or F (False). Contact with the external world can be made via experience, but in mathematics *true* and *false* may be viewed as undefined terms.

As noted earlier, the basic objects of ZFC are sets, collections of elements. The examples below refer to the set of *integers*, or whole numbers: 0, 1, -1, 2, -2, and so forth.

Example 1.1. -4 is an even integer.

The decimal expansion of π contains the string '999999'. (True)

$2 + 2 = 5$. (False)

Example 1.2. Sentences that are *not* statements include "n is an even integer" (whose truth value depends on n) "10^{1000} is a large number" ("large" has not been given mathematical meaning), and the self-referential examples, "This sentence is true" (whose truth value must be specified as an axiom) and “This sentence is false” (which cannot be consistently assigned a truth value).

Conventionally, abstract statements are denoted P and Q.

Not. The *negation* of a statement P is its logical opposite $\neg P$. You may regard the negation as P preceded by the clause “It is not the case that…”, but usually a more pleasant wording can be found.

Example 1.3. P: $2 + 2 = 4$. $\neg P$: $2 + 2 \neq 4$.

Let P and Q be statements. New statements can be constructed using the “logical connectives” *and*, *or*, and *implies*.

And. The statement “P and Q” has its ordinary meaning: The compound statement is true provided both P and Q are true, and is false otherwise.

Example 1.4. $2 + 2 = 4$ and $0 < 1$. (True)

$2 + 2 = 5$ and $0 < 1$. (False)

$2 + 2 = 5$ and $1 < 0$. (False)

Or. The statement “P or Q” always has the “inclusive” meaning in mathematics: P is true, or Q is true, *or both*.

Example 1.5. $2 + 2 = 4$ or $0 < 1$. (True)

$2 + 2 = 5$ or $0 < 1$. (True)

$2 + 2 = 5$ or $1 < 0$. (False)

Remark 1.6. In colloquial English, “or” is frequently used in the “exclusive” sense. The sentence “You will earn a 70% on the final exam or you will not pass the course” is conventionally interpreted to mean “If you earn a 70% on the final exam, then you will pass the course, and if you do not earn a 70%, then you will not pass.”

Mathematicians and computer scientists denote “exclusive or” by “xor” to distinguish it from “or”. The statement “P xor Q” means P is true, or Q is true, *but not both*. When needed, “P xor Q” can be expressed as “(P or Q) and $\neg$(P and Q)”. To a mathematician, the New Hampshire state motto reads *Live free xor die.* In this book, xor does not systematically appear again.

Implies. A statement of the form “If P then Q”, which also reads “P *implies* Q”, is called a *logical implication* and plays a central role in mathematics. P is called the *hypothesis* of the implication and Q is the *conclusion*.

By definition, a logical implication is *valid* provided Q is true whenever P is true. In other words, “P implies Q” is a valid deduction unless P is true and Q is false.

Example 1.7. If $1 \neq 0$, then $1^2 \neq 0$. (True)

If $1 \neq 0$, then $1^2 = 0$. (False)

If $1 = 0$, then $0 = 0$. (True)

If $1 = 0$, then $1^2 = 0$. (True)

If "P implies Q" is valid, we think of Q as being *deduced* or *derived* from P. The definition of "valid implication" ensures that by starting with true hypotheses and making valid deductions, we obtain only true conclusions, not falsehoods. There are two noteworthy and potentially confusing consequences of this convention, however.

First, it is valid (not logically erroneous) to deduce an arbitrary conclusion from a false hypothesis. An implication with false hypothesis is said to be *vacuously true*. Humorous examples abound: "If $1 = 0$, then money grows on trees."

In particular, the third and fourth implications of the preceding example are vacuously true. It may be helpful to point out that in each case, we can give a proof. If $1 = 0$, then subtracting this equation from itself gives $0 = 0$, which proves the third statement, while squaring gives $1^2 = 0^2 = 0$, proving the fourth statement.

Second, a valid implication need not connect causally related statements. The implication "If $0 = 0$, then 2 is an even integer" is valid because both the hypothesis and conclusion are true, but is effectively a *non sequitur*; the conclusion does not "follow" from the hypothesis in any obvious sense. A valid implication does not, of itself, constitute a proof. In the example at hand, we know the implication is valid only because there exists a separate proof, consisting of implications whose validity can be checked directly.

In these two senses, mathematicians are liberal in deeming an implication to be valid. Again, "validity" is the weakest criterion that excludes the act of drawing a false conclusion from a true hypothesis.

Remark 1.8. If, in some axiom system, some statement P and its negation $\neg P$ are both true, then *every* statement Q is provable, since either "P implies Q" or "$\neg P$ implies Q" is vacuously true. The pair $\{P, \neg P\}$ is called a *logical contradiction*. An axiom system is *inconsistent* if a contradiction can be derived from it.

Work of K. Gödel in the 1930s showed ZFC cannot be proved consistent without using some other ("more powerful") axiom system whose consistency is unknown. However, if there is a contradiction in ZFC, there is a contradiction in ordinary arithmetic.

Belief in the consistency of ZFC is about as close as mathematics gets to an "article of faith".

In this book, and throughout mathematics in practice, valid deductions do actually link causally related statements. Most implications involve classes of objects, and assert that every object satisfying some condition must also satisfy some other condition.

Negation and Conjunctions. If P and Q are statements, then the statement "P and Q" is false if *at least one* of P and Q is false. If someone assures you two statements are both true, only one has to be false for the assurance to be unfounded. Formally, the compound statements

$$\neg(P \text{ and } Q), \qquad (\neg P) \text{ or } (\neg Q)$$

express the same logical condition.

Analogously, if someone assures you at least one statement of two is true, then both must be false for the assurance to be unfounded. Formally, the compound statements

$$\neg(P \text{ or } Q), \qquad (\neg P) \text{ and } (\neg Q)$$

express the same logical condition.

Together, the two preceding relationships are known as *De Morgan's laws*, after the 19th Century English logician A. De Morgan. Loosely, the conjunctions "and" and "or" are interchanged by negation, perhaps contrary to first impression.

Consequently, the order of negation and a connective matters:

Example 1.9. The integers 1 and 0 are **not both** zero. (True)

The integers 1 and 0 are **both not** zero. (False)

Remark 1.10. All too frequently, one sees humorous ambiguities of the type "While driving, teens should not use cell phones and obey traffic laws". To avoid confusion, this sentence should be phrased "While driving, teens should obey traffic laws and not use cell phones" (placing the negation where it clearly applies only to one clause) or "While driving, teens should not use cell phones, and should obey traffic laws" (explicitly delimiting the negation).

In formal logic, "$\neg P$ and Q" means "$(\neg P)$ and Q".

1.2. Quantification

To accommodate classes of objects in the framework of statements, we allow statements to contain *variables* standing for elements of a set, so long as each variable is "quantified", accompanied by the phrase "for every" or "there exists". The quantifiers are crucial; pay close attention to them while reading, and *do not omit them when thinking and writing.*

Example 1.11. For every integer n, $n^2 - n$ is an even integer. (True)

For every integer n, $n^2 \geq 0$. (True)

For every integer n, $n^2 = 1$. (False)

"For every" statements are said to involve *universal quantification*. Each statement encapsulates multiple statements. For example, the first statement of the preceding example encapsulates an infinite collection of statements, one for each integer: $0^2 - 0$ is an even integer; $1^2 - 1$ is an even integer; $(-1)^2 - (-1)$ is an even integer; and so forth.

Example 1.12. There exists an integer n such that $n^2 = 1$. (True)

There exists an integer n such that $n^2 = 2$. (False)

There exists an n such that both n and $n + 1$ are even. (False)

"There exists" statements are said to involve *existential quantification*. Again, each encapsulates multiple statements. For example, the third expresses that at least one truism is found among the statements: 0 and 1 are both even; 1 and 2 are both even; -1 and 0 are both even; and so forth. The compound statement is false because *every* individual statement is false.

Remark 1.13. The statements of the preceding examples contain only "bound" (i.e., quantified) variables.

Sentences containing "free" or "unbound" variables (such as "n is an even integer" or "$x^2 + x - 2 = 0$") are not statements. However, sentences containing unbound variables play the useful role of *conditions* in mathematics, selecting objects (perhaps integers n or real numbers x) for which the resulting statement is true.

Many mathematical theorems take the universally quantified form "For every x satisfying $P(x)$, condition $Q(x)$ is true". For stylistic variety, such statements may be worded as implications involving "arbitrary" values of variables.

Example 1.14. If x is a real number such that $x^2 + x - 2 = 0$, then $x = 1$ or $x = -2$. (True)

If n is an integer, then there exist unique integers q and r such that $n = 4q + r$ and $0 \leq r < 4$. (True)

If a, b, and c are positive integers, then $a^3 + b^3 \neq c^3$. (True)

Quantifiers and Negation. The universal quantifier "for every" may be viewed as an enhancement of the "and" conjunction: "For every integer n, the condition $P(n)$ is true" means that the infinitely many statements $P(0)$, $P(1)$, $P(-1)$, and so forth, are *all* true.

The existential quantifier "there exists" may be viewed similarly as an enhancement of "or": "There exists an integer n such that the condition $P(n)$ is true" means that among the infinitely many statements $P(0)$, $P(1)$, $P(-1)$, ..., *at least one* is true.

Example 1.15. Logical negation "converts" a "for every" statement into a "there exists" statement of negations, and converts a "there exists" statement into a "for every" statement of negations:

P: For every integer n, $n^2 \geq 0$.
$\neg P$: There exists an integer n such that $n^2 < 0$.

P: There exist integers m and n such that $m^2 + n^2 = 8$.
$\neg P$: For all integers m and n, $m^2 + n^2 \neq 8$.

Remark 1.16. This type of linguistic transformation needs to become second nature. Particularly, a positive assertion regarding a class of objects can be disproved by finding a counterexample, but cannot be proved by finding an example.

Example 1.17. Three logicians walk into a bar. The bartender asks, "Do all of you want a beer?"

The first logician replies, "I don't know." The second adds, "I also don't know." The third says, "Yes."

Remark 1.18. When the hypothesis of a logical implication contains a variable but no quantifier is explicitly present, the convention is to read "for every". For example, "If $x > 0$ then $x^2 > 0$" should be read "For every real number x, if $x > 0$ then $x^2 > 0$" (assuming the context dictates real numbers as opposed to, say, integers).

If an implicitly quantified statement is negated, the existential quantifier must be added explicitly: "There exists a real number $x > 0$ such that $x^2 \leq 0$".

To avoid confusion, including your own, include logical quantifiers explicitly. This book makes a special effort to set a good example.

Implications, and Multiple Quantifiers. Among the most subtle conditions in mathematics are those containing multiple quantifiers. Elementary algebra seldom ventures into this territory, but analysis, the mathematics underlying and extending differential and integral calculus, is suffused with definitions and theorems of this type. When you encounter multiply quantified statements, slow down and read several times to ensure you thoroughly understand the dependencies implicit in the ordering.

Example 1.19. For every integer n, there exists an integer M such that $n \leq M$. (True; every integer n is smaller than some other integer M.)

There exists an integer M such that for every integer n, $n \leq M$. (False; there is no largest integer M, i.e., no integer that is greater than every other integer n.)

Each of these statements can be interpreted usefully as a strategy in an adversarial game; see Exercises 1.15 and 1.16.

1.3. Truth Tables and Applications

The logical operators ("not", "and", "or", and "implies") introduced above are neatly summarized by *truth tables*:

P	Q	$\neg P$	P and Q	P or Q	P implies Q
T	T	F	T	T	T
T	F	F	F	T	F
F	T	T	F	T	T
F	F	T	F	F	T

Truth tables furnish a useful tool for studying sentences built of other statements and logical connectives. This section gives a few applications.

Logical Equivalence. Two statements P and Q are *logically equivalent* if each implies the other: P implies Q and Q implies P. For brevity, we may write P iff Q, "iff" being short for “if and only if”. A truth table calculation shows P and Q are equivalent precisely when they have the same truth value:

P	Q	P implies Q	Q implies P	P iff Q
T	T	T	T	T
T	F	F	T	F
F	T	T	F	F
F	F	T	T	T

The Converse. The implications “P implies Q” and “Q implies P” are said to be *converse* to each other. The preceding table shows these implications are not equivalent.

The Contrapositive. The implications “P implies Q” and “$\neg Q$ implies $\neg P$” are said to be *contrapositive* to each other. An implication and its contrapositive are logically equivalent:

P	Q	P implies Q	$\neg Q$	$\neg P$	$\neg Q$ implies $\neg P$
T	T	T	F	F	T
T	F	F	T	F	F
F	T	T	F	T	T
F	F	T	T	T	T

This fact of logic should become second nature to you. Implications with a large number of hypotheses are generally easier to understand and prove in contrapositive form.

Example 1.20. In each statement, x stands for a real number. Let P be the statement “$x^2 - 1 \neq 0$” and Q be the statement “$x \neq 1$”.

The implication P implies Q is true, but may require a few seconds’ thought to see.

The converse implication, “If $x \neq 1$, then $x^2 - 1 \neq 0$” is an invalid deduction. The number $x = -1$ is a *counterexample*: It satisfies the converse hypothesis Q, but not the converse conclusion P.

The contrapositive reads, “If $x = 1$, then $x^2 - 1 = 0$.” This implication is obviously true, and on general grounds its truth implies the truth of P implies Q.

Example 1.21. In each statement below, n is a positive integer. A positive integer n is said to be *prime* if $n > 1$, and if n has no positive divisors other than 1 and n.

Direct implication: If n is a prime, then $n = 2$ or n is odd. (True)

Converse: If $n = 2$ or n is odd, then n is a prime. (False: $n = 1$ and $n = 9$ are the two smallest of infinitely many counterexamples.)

Contrapositive: If $n \neq 2$ and n is not odd, then n is not prime. (True. Every such integer has the form $n = 2k$ for some integer $k > 1$.)

One final example, drawn from analysis rather than from algebra, will illustrate the power of the contrapositive.

Example 1.22. In each statement, $x \geq 0$ is a real number and n is a positive integer.

Direct implication: If $x < 1/n$ for every n, then $x = 0$.

Converse: If $x = 0$, then $x < 1/n$ for every n.

Contrapositive: If $x > 0$, then there exists an n such that $1/n \leq x$.

It turns out that all three statements are true. The second is easily seen, even though the conclusion consists of infinitely many statements: $0 < 1$, $0 < 1/2$, $0 < 1/3$, etc.

The third statement is true and not difficult to see; informally, $1/k \to 0$ as $k \to \infty$, so if $x > 0$, there is some positive integer n such that $1/n \leq x$.

The direct implication is therefore true, since its contrapositive is true. However, the direct implication exhibits a new phenomenon: The hypothesis consists of infinitely many statements, $x < 1$, $x < 1/2$, $x < 1/3$, etc., but *no finite number of these statements implies the conclusion*. Indeed, if we assume only finitely many inequalities of the form $x < 1/n$, there is a largest denominator, say N, and our collection of inequalities is equivalent to the single inequality $x < 1/N$, which does not imply $x = 0$.

Exercises

Exercise 1.1. Many types of logical errors exist, including conflating an implication with its converse or inverse, assuming the conclusion, over-generalization, using undefined terms or using the same term for different things, and logical disconnect (where a true statement is followed by "therefore the desired conclusion is true").

Analyze the following syllogisms. (Recall that an integer $p \geq 2$ is *prime* if the only positive divisors of p are 1 and p.) Determine whether the hypotheses and conclusions of each are true, and whether one of the errors of the preceding paragraph has been made, or if the hypotheses justify the conclusions. (This is a stronger requirement than "implies". For example, "$1 + 1 = 2$ implies $0 \cdot 1 = 0$" is true, but the conclusion is not justified by the hypothesis: Truth of the hypothesis makes the conclusion no easier to see.)

(a) 3 is prime, 5 is prime, 7 is prime. Each is an odd integer greater than 1. Therefore every odd integer greater than 1 is prime.

(b) Every even integer greater than 2 is not prime. Therefore every odd integer greater than 2 is prime.

(c) Every even integer greater than 2 is not prime. Therefore every prime greater than 2 is not even.

(d) It is repugnant to the nature of a prime to be even. Therefore every prime greater than 2 is odd.

Exercise 1.2. The following arguments are similar to those of the preceding exercise, but contain non-mathematical assertions. As such, they may contain errors of pure logic, but in addition may be questionable due to imprecisely defined terms. Discuss and clarify vague terms if necessary, and analyze the logic of each as in the preceding exercise.

(a) Successful students spend long hours in the library. Therefore, spending long hours in the library makes a successful student.

(b) People called Galileo a crank during his life. In his lifetime, Galileo was an unrecognized genius. People call me a crank. Therefore I am an unrecognized genius.

(c) All healthy dogs have four legs. All healthy dogs are animals. Therefore all animals have four legs.

(d) All healthy dogs are carbon-based life forms. All healthy dogs are earth animals. Therefore all earth animals are carbon-based life forms.

(e) I am alive. Therefore I will live forever. (Adapted from the bumper sticker, "I intend to live forever. So far, so good.")

(f) All other men die. I am not like other men. Therefore I'll not die. (Adapted from Vladimir Nabokov.)

Exercise 1.3. A store sign reads, "Everything on this table discounted up to 50%, or even more."

(a) What is logically conveyed by this sign?

(b) What implicit promises do you take from the sign's truth?

Exercise 1.4. A promotional flyer proclaims: You are the guaranteed recipient of at least two of the following.

- Hotel Resort Platinum Getaway!
- $2,500.00 Instant Scratch Ticket!
- Home Theater System (retail value $500)!
- $1,000.00 Instant Scratch Ticket!
- $10,000 in cash!

Assuming these statements are true, what is the maximum value of the guaranteed prizes? What is the minimum value of the guaranteed prizes? If you cannot answer one or both questions, explain why not and what additional information you would need.

Exercise 1.5. The human brain has evolved to detect "cheating"—behavior violating established rules. These rules have logical content, but the "cheating" interpretation can be remarkably easier to "see". For best results, work out parts (a) and (b) completely before reading part (c).

(a) Each card in a deck is printed with a letter "D" or "N" on one side and a number between 16 and 70 on the other. Your job is to assess whether or not cards satisfy the criterion: "Every 'D' card has a number greater than or equal to 21 printed on the reverse." You are also to separate cards that satisfy this criterion from those that do not.

Write the criterion as an “If…, then…” statement, and determine which of the following cards satisfy the criterion:

$\boxed{20}$	$\boxed{46}$	$\boxed{16}$	$\boxed{25}$
$\boxed{\text{D}}$	$\boxed{\text{D}}$	$\boxed{\text{N}}$	$\boxed{\text{N}}$
(i)	(ii)	(iii)	(iv)

(b) You are shown four cards:

$\boxed{18}$	$\boxed{35}$	$\boxed{\text{D}}$	$\boxed{\text{N}}$
(i)	(ii)	(iii)	(iv)

Which cards must be turned over to determine whether or not they satisfy the criterion of part (a)? (The question continues; complete these parts before proceeding.)

(c) The legal drinking age in a certain state is 21. Your job at a gathering is to ensure that no one under 21 years of age is drinking alcohol, and to report those that are. A group of four people consists of a 20 year old who is drinking, a 46 year old who is drinking, a 16 year old who is not drinking, and a 25 year old who is not drinking. Which of these people is/are violating the law?

After reporting this incident, you find four people at the bar: An 18 year old and a 35 year old with their backs to you, and two people of unknown age, one of whom is drinking. From which people do you need further information to see whether or not they are violating the law?

(d) Explain why the card question is logically equivalent to the drinking question. Which did you find easier to answer correctly? (This exercise is adapted from the *Wason selection task* in social psychology.)

Exercise 1.6. (a) List (with justification) all sets of three non-zero digits (i.e., integers between 1 and 9 inclusive) whose sum is 15.

(b) Consider a two-player adversarial game with the following rules: (i) Players alternately pick integers between 1 and 9; each number can be picked at most once. (ii) A player wins if they pick some set of three numbers that add to 15.

Does each digit appear equally often in a “winning triple”? If not, which number(s) is/are most or least frequent?

(c) Show that the digits between 1 and 9 can be placed uniquely into a 3×3 array so that (i) The top row is "2, 9, x" for some digit x; (ii) The digits in each row sum to 15, the digits in each column sum to 15, and the digits on the diagonals sum to 15. Conclude that a set of three numbers adds to 15 if and only if the positions of those digits win the game of tic-tac-toe.

Exercise 1.7. The President, a law-abiding citizen who always tells the truth, has time for one more Yes/No question at a press conference. In an attempt to embarrass the President, a reporter asks, "Have you stopped offering illegal drugs to visiting Heads of State?"

(a) Which answer ("Yes" or "No") is logically truthful?

(b) Suppose the President answers "Yes". Can the public conclude that the President has offered illegal drugs to visiting Heads of State? What if the answer is "No"?

(c) Explain why both answers are embarrassing.

This rhetorical technique is the bread and butter of "push polls", propaganda or smear campaigns often conducted by telephone, disguised as attempts to gauge public opinion. If the President were a Zen Buddhist she might reply "mu" (pronounced "moo"), meaning "Your question is too flawed in its hypotheses to answer meaningfully."

Exercise 1.8. One domino covers two neighboring squares of a chess board. In each part, assume dominoes are non-overlapping, aligned with board squares, and rest entirely on the board.

(a) Can a 7×7 board be covered by dominoes? A 6×6 board? 8×8?

(b) Suppose two opposite corner squares are removed from a 6×6 board. Prove that the remaining squares cannot be covered by dominoes. What if the board is 7×7?

(c) Suppose two neighboring corner squares are removed from a 6×6 board. Can the remaining squares be evenly covered by dominoes? What if the board is 7×7?

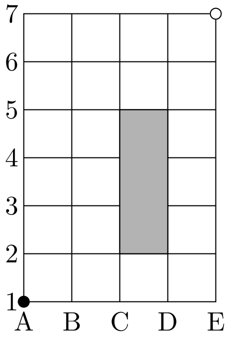
(d) Prove that if two squares of opposite color are removed from an 8×8 chess board, the remaining squares can be covered by dominoes. Hint: First show that each square can be visited precisely once by starting in one corner and successively stepping to neighboring squares.

Exercise 1.9. In the game Minesweeper, each square in an array is either empty or holds a mine. You task is to flag all the squares containing a mine, and only those squares. When a square is cleared (correctly marked as not containing a mine), the number of mines in adjacent squares is shown.

The 8 × 8 grid below (left) contains 20 uncleared squares (gray) and 44 cleared squares (white). Flag the mines (and only those squares), and explain your reasoning for each square.

(For your convenience, the squares are labeled using algebraic notation for chess. For instance, a1 is the lower left corner, and f5 is the left end of the peninsula in the middle.)

	a	b	c	d	e	f	g	h
8	▨	▨	▨	▨	▨		▨	▨
7	1	2	2	1		1	3	▨
6							3	▨
5						▨	▨	▨
4					1		1	
3								
2		1	2	1	2	2	2	
1	▨	▨	▨	▨	▨	▨	▨	▨



Exercise 1.10. Professor Calculus works at the corner of Avenue A and 1st Street (the black spot, above right), and lives at the corner of Avenue E and 7th Street (the open circle). Each day, Professor Calculus walks home along the city streets, traveling only northward or eastward, and not passing through the park (gray rectangle). How many distinct routes home are there?

Hint: How many distinct routes can Professor Calculus take to B1? To E1?

Exercise 1.11. In each pair P, Q of conditions, n represents an integer. (i) Give the negations of P and Q, and (ii) Form the implication P implies Q, its converse, and its contrapositive, and determine whether each is true.

(a) P: $n^2 - 4 = 0$. Q: $n = 2$.

(b) P: n is even. Q: n is an integer multiple of 4.

(c) P: n is even. Q: n is the square of an even integer.

Exercise 1.12. Let P and Q be arbitrary statements.

(a) Use a truth table to prove that "P implies Q" is logically equivalent to "$\neg P$ or Q".

(b) Use part (a) to re-show that an implication and its contrapositive are logically equivalent.

Exercise 1.13. Let x stand for a real number, and consider the conditions:

$$P : \ x^3 - 3x^2 + 2x \neq 0, \qquad Q : \ x \neq 1.$$

(a) Write out the direct implication P implies Q and the contrapositive. Are these statements true or false? Which is easier to decide?

(b) Write out the converse implication Q implies P and the inverse $\neg P$ implies $\neg Q$. Are these statements true or false? Which is easier to decide?

Exercise 1.14. Let P, Q, and R be arbitrary statements. Use a truth table to prove that the following pairs of statements are logically equivalent:

(a) "$\neg(P$ or $Q)$" and "$\neg P$ and $\neg Q$".

(b) "$\neg(P$ and $Q)$" and "$\neg P$ or $\neg Q$".

(c) "$(P$ or $Q)$ and R" and "$(P$ and $R)$ or $(Q$ and $R)$".

(d) "$(P$ and $Q)$ or R" and "$(P$ or $R)$ and $(Q$ or $R)$".

Exercise 1.15. (a) Consider the statement, "For every integer n, there exists an integer m such that $n < m$." Is this statement true or false? Explain.

(b) Consider the statement, "There exists an integer m such that for every integer n, $n < m$." Is this statement true or false? Explain.

(c) Consider an adversarial game ("Who can pick the larger number?") with the following rules: The first player picks an integer n and tells it to the second player. The second player picks an integer $m \neq n$. The player with the larger integer wins. If your goal is to win, would you rather play first, or second? Explain carefully how the statements in parts (a) and (b) are related to winning strategies in the game.

Exercise 1.16. (a) Consider the statement, "For every integer $n \leq 0$, there exists an integer $m \leq 0$ such that $n < m$." Is this statement true or false? Explain.

(b) Consider the statement, "There exists an integer $m \leq 0$ such that for every integer $n \leq 0$ distinct from m, we have $n < m$." Is this statement true or false? Explain.

(c) Consider an adversarial game ("Who can pick the larger non-positive number?") with the following rules: The first player picks an integer $n \leq 0$ and tells it to the second player. The second player picks an integer $m \leq 0$, $m \neq n$. The player with the larger integer wins. If your goal is to win, would you rather play first, or second? Explain carefully how the statements in parts (a) and (b) are related to winning strategies in the game.

(d) Explain carefully what property of the non-positive integers makes the answers to this question differ from those of the preceding question.

Chapter 2

An Introduction to Sets

Modern mathematics is built on the concept of a "set", a collection of "elements". These primitive notions will serve in lieu of definitions. This chapter informally introduces the set of complex numbers, connects sets with the basics of logic, and gives advice on constructing and writing mathematical proofs.

2.1. Specifying Sets

Example 2.1. The collection of all integers (whole numbers) is a set. Its elements are 0, 1, −1, 2, −2, and so forth. The set of integers is denoted **Z**, from the German *Zahl* (number). Formal axioms for the integers are given in Chapter 3.

Example 2.2. The collection of "prime numbers", integers p greater than 1 that have no divisors other than 1 and p, is a set. The numbers 2, 5, and $2^{13466917}-1$ are elements, while 4 and $2^{13466917} = 2 \cdot 2^{13466916}$ are not.

Example 2.3. The set of periodic table entries in the year 1960 has 102 elements. "Hydrogen", "promethium", and "astatine" are elements of this set, while "Massachusetts", "ammonia", and "surprise" are not.

Abstract sets will be denoted with capital letters, such as A or B. Elements are normally denoted with lowercase letters, such as a and b. We write "$a \in A$" as shorthand for the statement "a is an element of (the set) A", and "$b \notin A$" for the logical negation "b is not an element of A". For example, $0 \in \mathbf{Z}$, $-7 \in \mathbf{Z}$, and $\frac{1}{2} \notin \mathbf{Z}$.

Definition 2.4. Let A and B be sets. We say A is a *subset* of B, and write "$A \subseteq B$", if $x \in A$ implies $x \in B$, that is, if every element of A is an element of B. Two sets A and B are *equal* if $A \subseteq B$ and $B \subseteq A$, namely if they have exactly the same elements: $x \in A$ if and only if $x \in B$.

The most basic and explicit way of describing a set is to list its elements. Curly braces are used to denote a list of elements comprising a set. Sets do not "keep track of" what order the elements are listed, or whether their elements are multiply listed.

Example 2.5. Each set $A = \{-1, 0, 1\}$, $B = \{0, 1, -1\}$, and $C = \{0, 1, 0, -1, 1\}$ contains three elements, and in fact $A = B = C$.

Example 2.6. Let A be a set. For each element a in A, there is a *singleton* set $\{a\}$ contained in A. Take care to distinguish a and $\{a\}$; a is an object, while $\{a\}$ is a "package" that contains exactly one object.

Example 2.7. There exists an *empty set* $\varnothing$ containing *no* elements. For all x, the statement $x \in \varnothing$ is false. In particular, for every set A the logical implication "$x \in \varnothing$ implies $x \in A$" is vacuous (has a false hypothesis). Consequently, $\varnothing \subseteq A$ is true for all A.

Remark 2.8. The empty set is unique: If $\varnothing$ and $\varnothing'$ are sets having no elements, then $\varnothing \subseteq \varnothing'$ and $\varnothing' \subseteq \varnothing$ are both true, so $\varnothing = \varnothing'$.

In mathematics, we always restrict our attention to sets contained in a fixed set $\mathcal{U}$, called a *universe*. Specific subsets of $\mathcal{U}$ are conveniently described using *set-builder notation*, in which elements are selected according to logical conditions formally known as a *predicates*. The expression $\{x \text{ in } \mathcal{U} : P(x)\}$ is read "the set of all x in $\mathcal{U}$ such that $P(x)$".

Example 2.9. The expression $\{x \text{ in } \mathbf{Z} : x > 0\}$, read as "the set of all x in $\mathbf{Z}$ such that $x > 0$", specifies the set $\mathbf{Z}^+$ of *positive integers*.

To personify, if $\mathcal{U}$ is a population whose elements are individuals, then a subset A of $\mathcal{U}$ is a club or organization, and the predicate defining A is a membership card. We screen individuals x for membership in A by checking whether or not x carries the membership card for A, namely whether or not $P(x)$ is true.

Example 2.10. There can exist no "set $\mathcal{U}$ of all sets". If there were, the set $R = \{x \text{ in } \mathcal{U} : x \notin x\}$, comprising all sets that are not elements of themselves, would have the property that $R \in R$ if and only if $R \notin R$. This contradiction is known as *Russell's paradox*, after the English logician B. Russell.

Example 2.11. The expression $\{x \text{ in } \mathbf{Z} : x = 2n \text{ for some } n \text{ in } \mathbf{Z}\}$ is the set of *even integers*. We often denote this set $2\mathbf{Z}$, the idea being that the general even integer arises from multiplying some integer by 2.

Similarly, the set of *odd integers* could be expressed as

$$2\mathbf{Z} + 1 = \{x \text{ in } \mathbf{Z} : x = 2n + 1 \text{ for some } n \text{ in } \mathbf{Z}\}.$$

Remark 2.12. For brevity, we sometimes write, e.g., the set of even integers as $\{2n : n \in \mathbf{Z}\}$, read as "the set of $2n$ such that n is an element of $\mathbf{Z}$". This way of writing a set is convenient, and the meaning is generally clear, but it isn't technically proper; compare Example 2.10. To define a set formally, first give the universe, then specify the predicate.

Remark 2.13. The elements of a set may be other sets. For example, the set $A = \{2\mathbf{Z}, 2\mathbf{Z}+1\}$ has two elements, $2\mathbf{Z}$ and $2\mathbf{Z}+1$. Note carefully that A is not a subset of $\mathbf{Z}$: The elements of A are not themselves integers, but sets of integers.

2.2. Complex Numbers

Points in the Cartesian plane may be viewed as numerical entities in a way that extends the familiar real number line. The resulting "complex number system" illustrates many of the algebraic and geometric concepts introduced later.

Definition 2.14. A *complex number* is an expression $\alpha = a + ib$ in which a and b are real numbers and i is a symbol satisfying $i^2 = -1$. The real numbers a and b are, respectively, the *real part* and *imaginary part* of α. We say α is *real* if $b = 0$, *non-real* if $b \neq 0$, and *imaginary* if $a = 0$.

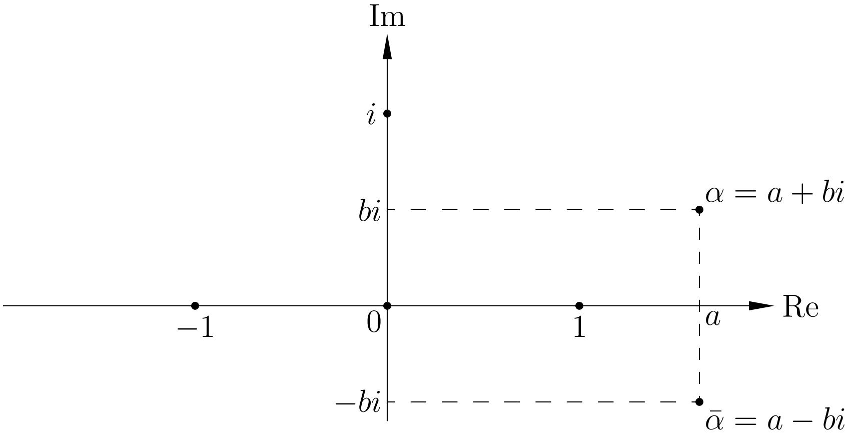


Figure 2.1. The complex plane.

Viewing the real and imaginary parts of a complex number $\alpha = a + bi$ as Cartesian coordinates, we identify α with the point (a, b), Figure 2.1.

Definition 2.15. The set of complex numbers is the *complex plane* $\mathbf{C}$. Each real number a is identified with the complex number $a + 0 \cdot i$. The set of all such points is the *real axis*. The set of all imaginary numbers $0 + b \cdot i$ is the *imaginary axis*. The *conjugate* of α is the complex number $\bar{\alpha} = a - bi$ obtained by reflecting α across the real axis.

Remark 2.16. Imaginary numbers may seem tainted with suspicion, as if they don't really exist but it's mathematically convenient to pretend they do. This sentiment traces back to the Ancient Greeks, who viewed numbers as lengths, what we now call "real numbers". Indeed, no real number has square equal to -1.

As noted above, however, i has a perfectly concrete existence as the point $(0, 1)$ in the Cartesian plane. Even the mysterious equation $i^2 = -1$ turns out to have a natural interpretation: Multiplication by i corresponds to a counterclockwise quarter-turn of the complex plane about the origin. Performing this operation twice, namely squaring i, amounts to a half-turn, which multiplies each complex number by -1.

From a modern perspective, the complex numbers earn their status as "numbers" by admitting operations of addition, subtraction, multiplication, and division that generalize the familiar algebraic properties of real numbers. We turn next to the algebraic and geometric descriptions of these operations.

Definition 2.17. Let $\alpha_1 = a_1 + ib_1$ and $\alpha_2 = a_2 + ib_2$ be complex numbers. Their *sum* is defined by the formula

$$\alpha_1 + \alpha_2 = (a_1 + ib_1) + (a_2 + ib_2) = (a_1 + a_2) + i(b_1 + b_2).$$

The formula for subtraction is similar and is left to you to work out. Adding two complex numbers corresponds to the parallelogram law for vector addition in the plane; see Figure 2.2.

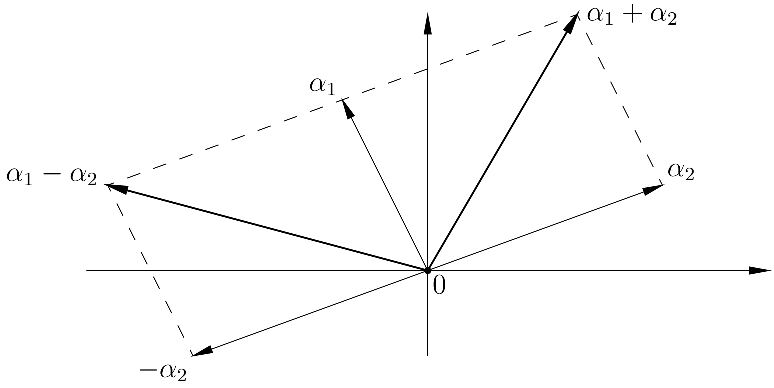


Figure 2.2. Adding and subtracting complex numbers.

Definition 2.18. A set A contained in **C** is *closed under addition* if for all α_1 and α_2 in A, the sum $\alpha_1 + \alpha_2$ is in A.

Example 2.19. The set $\{0\}$ is closed under addition, since $0 + 0 = 0$.

Example 2.20. Suppose A is closed under addition and $1 \in A$. Of necessity, $2 = 1 + 1$, $3 = 2 + 1$, $4 = 3 + 1$, and so forth, are in A. That is, A contains the set of positive integers. Since the set of positive integers is closed under addition, our hypotheses imply nothing further.

Similarly, if A is closed under addition and $\alpha \neq 0$ is an element of A, then every positive integer multiple of α is an element of A. Since these multiples are distinct, the set A must be infinite.

If A is closed under addition, it does not follow that A is "generated" by one element as in the previous examples.

Example 2.21. The set $\mathbf{Z}$ of integers is closed under addition in $\mathbf{C}$, as are the set $\mathbf{Q}$ of rational numbers (ratios of integers) and the set $\mathbf{R}$ of real numbers. None of these sets is obtained by adding a single element to itself repeatedly.

Example 2.22. The set $\mathbf{Z} + i\mathbf{Z} = \{m + in : m, n \in \mathbf{Z}\}$ of *Gaussian integers*, Figure 2.3, is closed under addition: If $\alpha_1 = m_1 + in_1$ and $\alpha_2 = m_2 + in_2$ are Gaussian integers, the addition formula for complex numbers gives $\alpha_1 + \alpha_2 = (m_1 + m_2) + i(n_1 + n_2)$. Since a sum of integers is an integer, the real and imaginary parts of $\alpha_1 + \alpha_2$ are integers. That is, $\alpha_1 + \alpha_2 \in \mathbf{Z} + i\mathbf{Z}$. Since α_1 and α_2 were arbitrary, $\mathbf{Z} + i\mathbf{Z}$ is closed under addition.

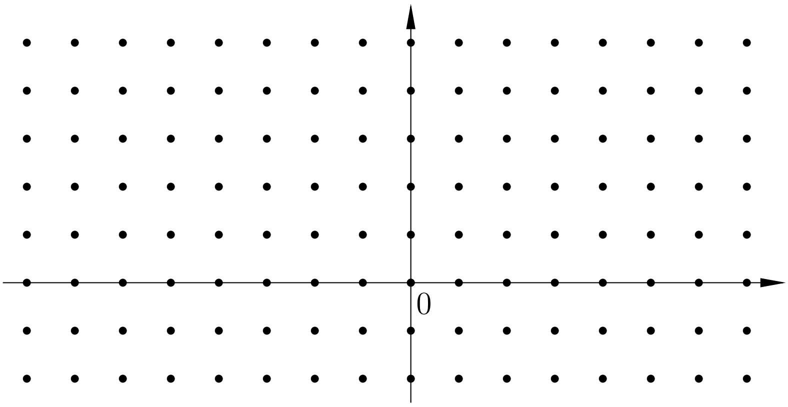


Figure 2.3. The Gaussian integers.

Example 2.23. The set A of complex numbers that are either real or imaginary, i.e., the union of the real and imaginary axes, is *not* closed under addition. Since "closed under addition" is a "for every" condition, its negation is a "there exists" condition; that is, it suffices to find a *single counterexample*. For instance, $1 \in A$ (since 1 is real) and $i \in A$ (since i is imaginary) but $1 + i \notin A$ (the sum is neither real nor imaginary), so A is not closed under addition.

To define multiplication of complex numbers, we treat i as a symbol distributing over the addition of real numbers, commuting with the multiplication of real numbers, and satisfying $i^2 = -1$. A short calculation using familiar laws of algebra leads us to

$$\begin{aligned}(a_1 + ib_1)(a_2 + ib_2) &= a_1a_2 + i(a_1b_2 + a_2b_1) + i^2b_1b_2 \\ &= (a_1a_2 - b_1b_2) + i(a_1b_2 + a_2b_1).\end{aligned}$$

Definition 2.24. Let $\alpha_1 = a_1 + ib_1$ and $\alpha_2 = a_2 + ib_2$ be complex numbers. Their *product* is defined by the formula

$$\alpha_1\alpha_2 = (a_1a_2 - b_1b_2) + i(a_1b_2 + a_2b_1).$$

Example 2.25. If $\alpha = a + bi$, then $i\alpha = i(a + bi) = -b + ai$. As expected, the vector $(-b, a)$ is obtained by rotating the vector (a, b) through a quarter-turn.

As a consistency check, $i(i\alpha) = i(-b + ai) = -a - bi = -\alpha$.

Example 2.26. If $\alpha = a + bi$, then

$$\alpha\bar{\alpha} = (a + bi)(a - bi) = a^2 - (bi)^2 = a^2 + b^2.$$

By the Pythagorean theorem, $\alpha\bar{\alpha} = (\text{distance from } 0 \text{ to } \alpha)^2$.

Complex multiplication is *commutative*: For all complex numbers α_1 and α_2, we have $\alpha_2\alpha_1 = \alpha_1\alpha_2$. We may therefore attempt to define division by declaring $\beta = \alpha_1/\alpha_2$ if and only if $\beta\alpha_2 = \alpha_1 = \alpha_2\beta$.

Remark 2.27. If multiplication were not commutative, the equations $\alpha_1 = \beta\alpha_2$ and $\alpha_1 = \alpha_2\beta$ might well be incompatible conditions for α_1.

To define complex division, let α_1 and α_2 be complex numbers with $\alpha_2 \neq 0$. We wish to write $\alpha_1/\alpha_2 = c_1 + ic_2$, namely, to find formulas for c_1 and c_2 in terms of the real and imaginary parts of the numerator and denominator.

The trick is analogous to rationalizing the denominator in high school algebra: Here we "realify" the denominator, multiplying the top and bottom by the conjugate number $\bar{\alpha}_2 = a_2 - ib_2$:

$$\frac{a_1 + ib_1}{a_2 + ib_2} = \frac{a_1 + ib_1}{a_2 + ib_2} \cdot \frac{a_2 - ib_2}{a_2 - ib_2} = \frac{(a_1a_2 + b_1b_2) + i(-a_1b_2 + a_2b_1)}{a_2^2 + b_2^2}.$$

Example 2.28. To divide $\alpha_1 = 2 - i$ by $\alpha_2 = 4 + 3i$, calculate as follows:

$$\begin{aligned} \frac{2 - i}{4 + 3i} &= \frac{2 - i}{4 + 3i} \cdot \frac{4 - 3i}{4 - 3i} = \frac{(8 - 3) + (-6 - 4)i}{4^2 + 3^2} \\ &= \frac{5 - 10i}{25} = \frac{1 - 2i}{5}. \end{aligned}$$

In practice, direct calculation is easier than memorizing the formula.

Example 2.29. If $\alpha = a + bi \neq 0$, then

$$\frac{1}{\alpha} = \frac{1}{a + ib} = \frac{a - ib}{a^2 + b^2} = \frac{a}{a^2 + b^2} - i\,\frac{b}{a^2 + b^2}.$$

That is, every non-zero complex number has a reciprocal.

The arithmetic operations on complex numbers satisfy familiar rules of algebra. For later reference, we collect several of these here. The proofs are straightforward calculations based on corresponding properties for real numbers and are left as exercises.

Proposition 2.30. *For all complex numbers α, β, and γ:*

(i) $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$ *and* $\beta + \alpha = \alpha + \beta$.

(ii) $(\alpha\beta)\gamma = \alpha(\beta\gamma)$ *and* $\beta\alpha = \alpha\beta$.

(iii) $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$ *and* $(\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma$.

Remark 2.31. In words, (i) complex addition is associative and commutative; (ii) complex multiplication is associative and commutative; (iii) multiplication distributes (on the left and on the right) over addition.

Example 2.32. For all *complex* α and β, the *difference of squares* identity holds: $\alpha^2 - \beta^2 = (\alpha+\beta)(\alpha-\beta)$.

Complex multiplication has a beautiful and useful geometric interpretation, most easily expressed in terms of *polar coordinates*. Recall that every point (a, b) in the plane can be written $(r\cos\theta, r\sin\theta)$ for some radius $r \geq 0$ and some angle θ, measured counterclockwise from the positive x axis and unique up to an added integer multiple of 2π.

Definition 2.33. Let $\alpha = a + bi = r\cos\theta + ir\sin\theta$ be a complex number. The radius r is called the *magnitude* of α, and the polar angle is the *argument* of α. If $-\pi < \theta < \pi$, we say θ is the *principal argument* of α.

Remark 2.34. The magnitude of $\alpha = a + ib$, denoted $|\alpha|$, is given by

$$|\alpha| = r = \sqrt{a^2+b^2} = \sqrt{\alpha\bar{\alpha}}.$$

Example 2.35. Since $i = 0 + 1 \cdot i = \cos\frac{\pi}{2} + i\sin\frac{\pi}{2}$, the magnitude of i is 1 and the principal argument of i is $\frac{\pi}{2}$.

Example 2.36. Let θ be a real number. By *Euler's formula* (see the appendix), we have $\cos\theta + i\sin\theta = e^{i\theta}$. The magnitude of $e^{i\theta}$ is 1, and the argument is θ.

Generally, $\alpha = |\alpha|(\cos\theta + i\sin\theta) = |\alpha|e^{i\theta}$.

If $e^{i\theta_1} = (\cos\theta_1 + i\sin\theta_1)$ and $e^{i\theta_2} = (\cos\theta_2 + i\sin\theta_2)$ are complex numbers of unit magnitude, the sum formulas for the cosine and sine functions allow us to write their product as

$$\begin{aligned} e^{i\theta_1} \cdot e^{i\theta_2} &= (\cos\theta_1 + i\sin\theta_1)(\cos\theta_2 + i\sin\theta_2) \\ &= (\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2) + i(\cos\theta_1\sin\theta_2 + \cos\theta_2\sin\theta_1) \\ &= \cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2) = e^{i(\theta_1+\theta_2)}. \end{aligned}$$

That is, *the law of exponents holds for imaginary exponents.* Since every complex number has polar form $\alpha = |\alpha|\,e^{i\theta}$, complex multiplication satisfies

$$\alpha_1\alpha_2 = \big(|\alpha_1|\,e^{i\theta_1}\big)\big(|\alpha_2|\,e^{i\theta_2}\big) = \big(|\alpha_1|\,|\alpha_2|\big)\,e^{i(\theta_1+\theta_2)}.$$

Geometrically, we multiply two complex numbers by multiplying their magnitudes and adding their arguments (polar angles). See Figure 2.4.

Example 2.37. Since $i = \cos\frac{\pi}{2} + i\sin\frac{\pi}{2} = e^{i\frac{\pi}{2}}$, we have

$$i\alpha = i|\alpha|\,e^{i\theta} = |\alpha|\,e^{i(\theta+\frac{\pi}{2})};$$

again we see that multiplication by i rotates the plane about the origin by a quarter-turn counterclockwise.

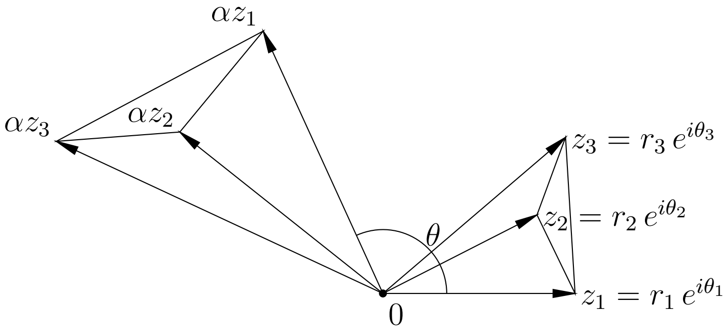

αz_1
αz_3
αz_2
$z_3 = r_3\, e^{i\theta_3}$
θ
$z_2 = r_2\, e^{i\theta_2}$
$z_1 = r_1\, e^{i\theta_1}$
0

Figure 2.4. Complex multiplication by $\alpha = |\alpha| e^{i\theta}$.

Example 2.38. Every non-zero complex number has precisely two complex square roots. This is particularly clear using polar form: Every non-zero complex number z may be written uniquely as $re^{i\theta}$ for some real $r > 0$ and real θ with $-\pi < \theta \leq \pi$. The numbers $\pm\sqrt{r}e^{i\theta/2}$ are the distinct square roots of z; see Exercise 2.9.

Example 2.39. If $\alpha x^2 + \beta x + \gamma = 0$ with α, β, and γ complex and $\alpha \neq 0$, then

$$x = \frac{-\beta \pm \sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha},$$

by the same completing-the-square proof you have seen for real coefficients. There are no "exceptional" cases; every quadratic has exactly two complex solutions, counting multiplicity.

Definition 2.40. A set A contained in $\mathbf{C}$ is *closed under multiplication* if, for all α_1 and α_2 in A, the product $\alpha_1 \cdot \alpha_2$ is an element of A.

Example 2.41. The set of complex numbers of magnitude 1 is the *unit circle*

$$U(1) = \{z \text{ in } \mathbf{C} : |z| = 1\} = \{z \text{ in } \mathbf{C} : z = e^{i\theta} \text{ for some real } \theta\}.$$

The set $U(1)$ is closed under multiplication: If $|\alpha_1| = 1$ and $|\alpha_2| = 1$, i.e., $\alpha_1, \alpha_2 \in U(1)$, then $|\alpha_1\alpha_2| = |\alpha_1||\alpha_2| = 1$, so $\alpha_1\alpha_2 \in U(1)$.

Example 2.42. The *finite* subsets $\{1\}$ and $\{-1, 1\}$ of $U(1)$ are also closed under multiplication. More generally, for each positive integer n there exists a subset U_n of $U(1)$ that contains exactly n elements and is closed under multiplication:

$$\begin{aligned} U_n &= \{1 = e^0,\ e^{i\,2\pi/n},\ e^{i\,4\pi/n},\ \dots,\ e^{i\,2\pi(n-1)/n}\} \\ &= \{e^{i\,2\pi k/n} : k = 0, \dots, n-1\}. \end{aligned}$$

The sets U_4 and U_6 are shown in Figures 2.5 and 2.6. Each is finite and closed under multiplication.

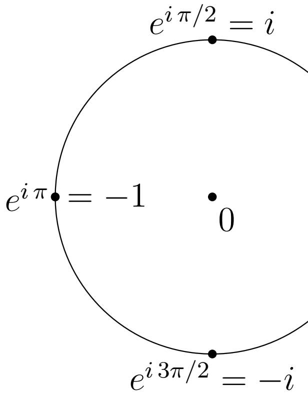


Figure 2.5. The set U_4.

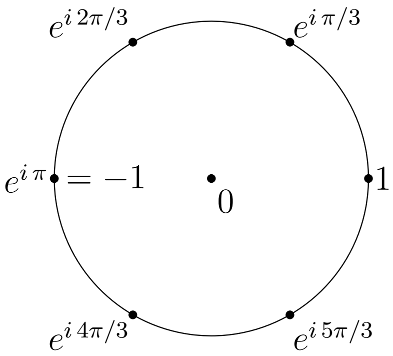


Figure 2.6. The set U_6.

The elements of U_n are precisely the complex numbers $\zeta = re^{i\theta}$ satisfying the equation $\zeta^n = 1$, namely the so-called *nth roots of unity*. To see why, note that $1 = \zeta^n = r^n e^{in\theta}$ precisely when $r = 1$ and $n\theta$ is an integer multiple of 2π. Assuming without loss of generality that $0 \leq \theta < 2\pi$, we have $0 \leq n\theta < 2n\pi$, so that $n\theta = 0$, $2\pi, 4\pi, \dots, 2(n-1)\pi$, or $n\theta = 2k\pi$ for some integer k with $0 \leq k < n$.

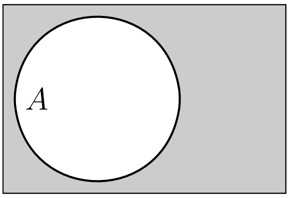
To see that the set of nth roots of unity is closed under multiplication, note that if $\zeta_1^n = 1$ and $\zeta_2^n = 1$, then $(\zeta_1\zeta_2)^n = \zeta_1^n\,\zeta_2^n = 1$, which means $\zeta_1\zeta_2$ is an nth root of unity.

2.3. Sets and Logic, Partitions

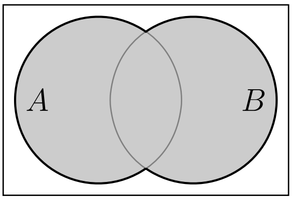
Let $\mathcal{U}$ be a universe, and let A and B be subsets of $\mathcal{U}$. The statements $x \in A$ and $x \in B$ may be viewed as predicates P and Q on elements of $\mathcal{U}$. By definition, the logical implication "$x \in A$ implies $x \in B$" corresponds to the set relation "$A \subseteq B$". Logical negation, disjunction (or), and conjunction (and) similarly have natural interpretations in terms of A and B.

The *complement* of A:	$A^c = \{x \text{ in } \mathcal{U} : x \notin A\}$.
The *union* of A and B:	$A \cup B = \{x \text{ in } \mathcal{U} : x \in A \text{ or } x \in B\}$.
The *intersection* of A and B:	$A \cap B = \{x \text{ in } \mathcal{U} : x \in A \text{ and } x \in B\}$.

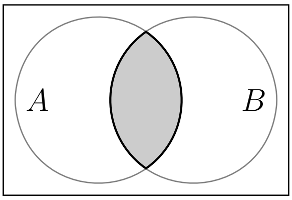
A *Venn diagram* pictorially represents subsets of a universe $\mathcal{U}$. The universe is depicted as a rectangle, and subsets are disks or, if necessary, more complicated shapes. The complement of A, or the union and intersection of two sets A and B, might be drawn as indicated:



A^c



$A \cup B$



$A \cap B$

Two sets A and B are *disjoint* if $A \cap B = \varnothing$, namely, if A and B have no elements in common. A Venn diagram of disjoint sets might be drawn as a pair of non-overlapping disks.

Example 2.43. The sets $2\mathbf{Z}$ and $2\mathbf{Z}+1$ of even and odd integers are disjoint: No integer is both even and odd. The sets $A = 2\mathbf{Z}$ and $B = \mathbf{Z}^+$ are *not* disjoint: For example, 2, 4, and 84 are elements of $A \cap B$, since each is both positive and a multiple of 2.

Definition 2.44. Let A be a set. The *power set* of A, $\mathcal{P}(A)$, is the set of all subsets of A.

Example 2.45. If $A = \{0, 1\}$ has two elements, the power set $\mathcal{P}(A)$ has four elements:

$$\mathcal{P}(A) = \big\{\varnothing, \{0\}, \{1\}, A\big\}.$$

The empty set and A itself are always subsets of A, so a power set is never empty. Indeed, $\mathcal{P}(\varnothing) = \{\varnothing\}$ has a single element.

Definition 2.46. Let A be a set, and I a set of indices. A family of non-empty subsets $\{A_i\}_{i\in I}$ of A constitutes a *partition* of A if each element of A is an element of *exactly one* of the sets A_i.

Remark 2.47. In other words, $\{A_i\}_{i\in I}$ is a partition of A if $A_i \neq \varnothing$ for all i (each set is non-empty), $A_i \cap A_j = \varnothing$ for $i \neq j$ (each pair of sets is disjoint), and A is the union of the sets A_i.

Example 2.48. The sets $A_0 = 2\mathbf{Z}$ and $A_1 = 2\mathbf{Z} + 1$ are a partition of $A = \mathbf{Z}$; every integer is either even or odd, and no integer is both. Here the index set is $I = \{0, 1\}$.

The sets $A_0 = 3\mathbf{Z}$, $A_1 = 3\mathbf{Z} + 1$, $A_2 = 3\mathbf{Z} + 2$ are another partition of $\mathbf{Z}$, since every integer leaves a unique remainder of 0, 1, or 2 upon division by 3:

$\mathbf{Z}$	$\cdots$	-4	-3	-2	-1	0	1	2	3	4	5	6	$\cdots$
A_0	$\cdots$		-3			0			3			6	$\cdots$
A_1	$\cdots$			-2			1			4			$\cdots$
A_2	$\cdots$	-4			-1			2			5		$\cdots$

Example 2.49. We will prove in Chapter 3 (Theorem 3.16) that if $n > 1$ is an integer, there is a partition of $\mathbf{Z}$ into n subsets, $A_k = n\mathbf{Z} + k$ with $k = 0, \dots, n-1$ an integer. An integer x is an element of A_k if and only if x leaves a remainder of k upon division by n.

In Chapter 8, we will write $[k]_n = n\mathbf{Z} + k$ and form a set $\mathbf{Z}_n$ having n elements: $\mathbf{Z}_n = \{[0]_n, [1]_n, \dots, [n-1]_n\}$. Note that $\mathbf{Z}_n \subseteq \mathcal{P}(\mathbf{Z})$: The *elements* of $\mathbf{Z}_n$ are *subsets* of $\mathbf{Z}$.

Advice on Writing Proofs. Discovering and writing proofs are nearly opposite activities. You'll find that most of the writing you do in discovering mathematics does not need to be written up; it's just "scaffolding".

Example 2.50. Assume $\alpha \in \mathbf{C}$. Prove: $|\alpha| = |\bar{\alpha}|$.

(Preliminary Work). When proving an identity such as this we have an obvious strategy: Express each side in terms of simpler information and see if the answers agree. Here, set $\alpha = a + ib$ with a and b real. Then $\bar{\alpha} = a - ib$, so we have

$$|\alpha| = \sqrt{a^2 + b^2}, \qquad |\bar{\alpha}| = \sqrt{a^2 + (-b)^2} = \sqrt{a^2 + b^2}.$$

These are indeed equal.

(The Written Solution). Assume $\alpha \in \mathbf{C}$. Prove: $|\alpha| = |\bar{\alpha}|$.

Proof: Let α be an arbitrary complex number, and write $\alpha = a + ib$ with a and b real. We have $\bar{\alpha} = a - ib$, and therefore

$$|\bar{\alpha}| = \sqrt{a^2 + (-b)^2} = \sqrt{a^2 + b^2} = |\alpha|,$$

as was to be shown.

Remark 2.51. When writing up a formal proof of an algebraic identity Q, the preferred style is to build a chain of equalities from one side to the other. *Do not* write down the desired conclusion Q and then manipulate each side until you have an identity P. At best, this "two-column" argument establishes the converse, Q implies P, which is not equivalent to P implies Q, and does not even imply the truth of Q. See Exercises 2.23 and 2.24 for pitfalls of the "two-column" style of proof.

Example 2.52. Prove or disprove: $2\mathbf{Z} + 1 = 2\mathbf{Z} - 1$.

(Preliminary Work). By the definition of equality of sets, we are to determine whether each set is a subset of the other. Some initial formalization can be performed mechanically. Give each set a name, write down its definition, and express the question in terms of this framework.

Here, we have two sets of integers,

$$A = 2\mathbf{Z} + 1 = \{x \text{ in } \mathbf{Z} : x = 2u + 1 \text{ for some } u \text{ in } \mathbf{Z}\},$$
$$B = 2\mathbf{Z} - 1 = \{y \text{ in } \mathbf{Z} : y = 2v - 1 \text{ for some } v \text{ in } \mathbf{Z}\}.$$

We wish to show either that $A \subseteq B$ and $B \subseteq A$ (which by definition means $A = B$ as sets), or that at least one of these inclusions is false.

Next, try to determine intuitively whether or not the statement is false (which can be shown by exhibiting a counterexample, an element of one set that is not an element of the other set) or true. To get an element of $2\mathbf{Z}+1$, add 1 to an even integer: $1 = 0+1$, $3 = 2+1$, $5 = 4+1$, $-1 = -2+1$, and so forth, are elements. Similarly, subtracting 1 from an even integer gives an element of $2\mathbf{Z}-1$: $-1 = 0-1$, $1 = 2-1$, $3 = 4-1$, $-3 = -2-1$, and so forth, are elements.

This evidence doesn't merely suggest the two sets *are* equal, it even points to a strategy of proof: Any integer one greater than an even integer is one less than the next largest even integer. We'll sketch out an informal proof to settle notation and iron out any unforeseen logical wrinkles.

The statement "$A \subseteq B$" may be phrased "if $x \in A$, then $x \in B$". If $x \in A$, then by the definition of A there exists an integer u such that $x = 2u+1 = 2(u+1)-1$. Setting $v = u+1$ (an integer because u is), we see x has the form $2v-1$ for some integer v, which by definition means $x \in B$. This shows $A \subseteq B$.

The inclusion $B \subseteq A$ is entirely similar, so at this stage we can write up a formal proof. The considerations above that led to the proof are customarily omitted from the formal write-up. Note, however, that the proof involves choices not easily known ahead of time; the scratch work is important!

(The Written Solution). Show $2\mathbf{Z}+1 = 2\mathbf{Z}-1$.

Proof: By definition, $A = \{x \text{ in } \mathbf{Z} : x = 2u+1 \text{ for some } u \text{ in } \mathbf{Z}\}$ and $B = \{y \text{ in } \mathbf{Z} : y = 2v-1 \text{ for some } v \text{ in } \mathbf{Z}\}$. Assume $x \in A$. By hypothesis, there exists an integer u such that $x = 2u+1$. Let $v = u+1$, so $u = v-1$, and note v is an integer. Since

$$x = 2u+1 = 2(v-1)+1 = 2v-2+1 = 2v-1,$$

$x \in B$. Since x was arbitrary (i.e., x could have been any element of A), we have shown $A \subseteq B$.

Conversely, suppose $y = 2v-1 \in B$ for some integer v. Let $u = v-1$, so that $v = u+1$. Then

$$y = 2v-1 = 2(u+1)-1 = 2u+1,$$

so $y \in A$. Since y was arbitrary, we have shown $B \subseteq A$.

Since $A \subseteq B$ and $B \subseteq A$, we have $A = B$.

Writing proofs requires practice. The final result should be a coherent, logical, step-by-step argument starting with the given hypotheses and leading to the conclusion.

Example 2.53. Let A and B be subsets of $\mathcal{U}$. Find the most general conditions on A and B under which $A \cap B = A$.

(Examples). If you're comfortable with sets and operations, go for the frontal assault ("reducing to the definitions", below). Otherwise, proceed by writing out examples on scratch paper or a blackboard. If Venn diagrams are more natural, use those. If concrete sets are easier to think about, use those. At this stage it's all right to let $\mathcal{U} = \mathbf{Z}$, the set of integers, but in the final proof, do not make any assumptions on the nature of $\mathcal{U}$, A, or B.

(Simpler cases). Since the target condition involves two sets, we can reduce to a simpler question by “fixing” one set and letting the other set vary.

If $A = \varnothing$, then $A \cap B = \varnothing \cap B = \varnothing = A$ regardless of B. If $A = \mathcal{U}$, then $A \cap B = \mathcal{U} \cap B = B$, which is not equal to A unless $B = \mathcal{U}$.

These examples show the condition $A \cap B = A$ *can* be true, but is *not always* true. The guiding task is to discover what common aspect these examples possess. If you’re still not sure, draw a Venn diagram with a circle representing A, and ask: What condition on B guarantees that $A \subseteq A \cap B$? Draw circles that are disjoint from A, that are contained in A, that partially overlap A, or that contain A. The evidence of this “experiment” should point toward the desired condition.

(Reducing to the definitions). The condition $A \cap B = A$ encapsulates two set inclusions, $A \cap B \subseteq A$ and $A \subseteq A \cap B$. The first inclusion is true for all pairs of sets: If $a \in A \cap B$, then $a \in A$ and $a \in B$, so perforce $a \in A$. Since a is an arbitrary element of $A \cap B$, this argument shows $A \cap B \subseteq A$.

We are therefore seeking the most general conditions under which $A \subseteq A \cap B$, namely, “$a \in A$ implies ‘$a \in A$ and $a \in B$’ ”. Clearly, this is equivalent to “$a \in A$ implies $a \in B$”, which may be rephrased as $A \subseteq B$, our putative answer.

As a consistency check, recall that $A = \varnothing$ and $A = \mathcal{U} = B$ satisfied the condition. In each case, $A \subseteq B$ holds. If the purported abstract condition is violated by examples, it’s definitely wrong.

(Putative conclusion). As the result of considerations above, we claim that $A \cap B = A$ if and only if $A \subseteq B$. To *prove* this formally, it suffices to establish two logical implications:

$$A \cap B = A \text{ implies } A \subseteq B, \qquad A \subseteq B \text{ implies } A \cap B = A.$$

Here, approximately, is what you’d normally write up:

(The Written Solution). $A \cap B = A$ if and only if $A \subseteq B$.

Proof: ($A \cap B = A$ implies $A \subseteq B$) Assume $A \cap B = A$, namely $A \cap B \subseteq A$ and $A \subseteq A \cap B$. Since the first inclusion holds for all sets, our initial hypothesis is equivalent to $A \subseteq A \cap B$.

Let a be an arbitrary element of A. Since $A \subseteq A \cap B$ by hypothesis, $a \in A \cap B$, so $a \in A$ and $a \in B$. In particular, $a \in B$. We have shown that if $a \in A$, then $a \in B$; this means that $A \subseteq B$, as was to be shown.

($A \subseteq B$ implies $A \cap B = A$) By hypothesis, if $a \in A$, then $a \in B$, so if $a \in A$, then $a \in A$ and $a \in B$. Since a is arbitrary we have $A \subseteq A \cap B$. The reverse inclusion $A \cap B \subseteq A$ holds for all sets A and B. We have shown that if $A \subseteq B$, then $A \cap B = A$. This completes the proof.

Find your own writing style. *Do write accurately and precisely*, but don't be pedantic or excessively wordy.

Avoid pronouns, especially "it". In the middle of even a simple proof, two or three objects tend to be under consideration, and "it" can often refer to any of them. If you're unable to decide exactly what "it" refers to, you've located something you don't fully understand.

Exercises

Exercise 2.1. Let A be a set and assume $a \in A$. Determine whether each statement is always true, sometimes true, or never true. If the statement is sometimes true, give examples of A and/or a for which the statement is true or is false.
(a) $a \in \{a\}$, (b) $a \subseteq A$, (c) $\{a\} \subseteq \varnothing$, (d) $\varnothing \in A$, (e) $\{a\} \in A$.

Exercise 2.2. On graph paper, carefully sketch the indicated sets of complex numbers. Determine whether each pair of sets is disjoint; if not, describe the intersection.

(a) $A = \{\alpha \in \mathbf{C} : |\alpha| \le 1\}$;
$B = \{\alpha \in \mathbf{C} : |\alpha - 2| \le 1\}$;
$C = \{\alpha \in \mathbf{C} : |\alpha - 3i| \le 1\}$.

(b) $A = \{\alpha \in \mathbf{C} : |\alpha| \le 1/2\}$;
$B = \{\alpha \in \mathbf{C} : 1 \le |\alpha| \le 2\}$;
$C = \{\alpha \in \mathbf{C} : 3 \le |\alpha - i| \le 4\}$.

(c) $A = \{\alpha \in \mathbf{C} : |\alpha + 1| \le 1\}$;
$B = \{\alpha \in \mathbf{C} : |\alpha - 1| \le 1\}$;
$C = \{\alpha \in \mathbf{C} : |\alpha - i| \le 3\}$.

(d) $A = \{\alpha \in \mathbf{C} : 0 \le \operatorname{Re}\alpha\}$;
$B = \{\alpha \in \mathbf{C} : \operatorname{Re}\alpha \le 1\}$;
$C = \{\alpha \in \mathbf{C} : 0 \le \operatorname{Im}\alpha\}$.

Exercise 2.3. If A and B are sets of complex numbers, we define their *sum* to be the set

$$\begin{aligned} A + B &= \{c \in \mathbf{C} : c = a + b \text{ for some } a \text{ in } A \text{ and } b \text{ in } B\} \\ &= \{a + b : a \in A, b \in B\}. \end{aligned}$$

If A and B are sets of complex numbers, prove that

$$A + B = \bigcup_{a \in A} \big(\{a\} + B\big).$$

Exercise 2.4. For the following pairs of sets, sketch A, B, and $A + B$. (See Exercise 2.3 for the definition of $A + B$.)

(a) $A = \{-2, -1, 0, 1, 2\}$, $B = \{-i, 0, 2i\}$.
(b) $A = \{-2, -1, 0, 1, 2\}$, $B = \{z \text{ in } \mathbf{C} : |z| \le 1/2\}$.
(c) $A = \{-2, -1, 0, 1, 2\}$, $B = \{z \text{ in } \mathbf{C} : |z| \le 1\}$.
(d) $A = [-1, 1]$ (the closed real interval), $B = \{z \text{ in } \mathbf{C} : |z| \le 1\}$.

Exercise 2.5. If n is an integer, let $A_n = [n, n + 1)$ be the half-open interval consisting of all real x such that $n \le x < n + 1$. In each part, sketch a few of the indicated sets and establish the stated property.

(a) The collection $\{A_n\}_{n\in\mathbf{Z}}$ is a partition of $\mathbf{R}$.

(b) If $B_n = \{\alpha \text{ in } \mathbf{C} : \operatorname{Re}\alpha \in A_n\}$, then $\{B_n\}_{n\in\mathbf{Z}}$ is a partition of $\mathbf{C}$.

(c) If $C_n = \{\alpha \text{ in } \mathbf{C} : \operatorname{Im}\alpha \in A_n\}$, then $\{C_n\}_{n\in\mathbf{Z}}$ is a partition of $\mathbf{C}$.

(d) If $D_n = \{\alpha \text{ in } \mathbf{C} : |\alpha| \in A_n\}$, then $\{D_n\}_{n\in\mathbf{Z}}$ is a partition of $\mathbf{C}$. (Note: The D_n are not all non-empty.)

Exercise 2.6. If A and B are sets of complex numbers, give a definition of their *product* analogous to Exercise 2.3, and then state and prove the corresponding property for products of sets.

Exercise 2.7. For the following pairs of sets, sketch A, B, and AB. (See Exercise 2.6 for the definition of AB.)

(a) $A = \{-1, 1, i, 2\}$, $B = \{-i, 0, 2i\}$.

(b) $A = \{1, 2, 3\}$, $B = \{z \text{ in } \mathbf{C} : |z| = 1\}$.

(c) $A = B = \{z \text{ in } \mathbf{C} : |z| = 1\}$.

(d) $A = \{1, i, -1, -i\} = \{e^{k\pi i/2} : k \in \mathbf{Z}\}$,
$B = \{1, \frac{1}{2}(-1 + i\sqrt{3}), \frac{1}{2}(-1 - i\sqrt{3})\} = \{e^{2k\pi i/3} : k \in \mathbf{Z}\}$.

Exercise 2.8. Let $A = 2\mathbf{Z}$ and $B = 3\mathbf{Z}$.

(a) Find $A \cap B$; that is, determine which integers are in $A \cap B$.

(b) List the elements of $A \cup B$ between -12 and 12.

(c) Show that $A+B$ is closed under addition and closed under taking negatives. (That is, if $n \in A+B$, then $-n \in A+B$.) Show that $1 \in A+B$ and argue that consequently $A + B = \mathbf{Z}$.

Exercise 2.9. Let r and θ be real numbers such that $r > 0$ and $-\pi < \theta \le \pi$.

(a) Show that $\sqrt{r}e^{i\theta/2}$ and $-\sqrt{r}e^{i\theta/2}$ are distinct square roots of $z = re^{i\theta}$.

(b) Prove that z has at most two distinct square roots. Hint: Use Example 2.32.

(c) Suppose that x and y are real and that $z = x + iy$ is non-zero. If $(u + iv)^2 = z$, find algebraic formulas for u and v in terms of x and y, and show that u, v are real.

(d) Find and plot the square roots of i in polar and rectangular forms.

(e) Find and plot the square roots of $-\frac{1}{2}(1 + i\sqrt{3})$ in polar and rectangular forms.

Exercise 2.10. In each part, let $A = \mathbf{Z} + i\mathbf{Z}$.

(a) Let $B = \{z \text{ in } \mathbf{C} : |z| \le 2\}$. Sketch the set $A \cap B$ and list the elements.

(b) How many elements of A satisfy $|z| \le 5$?
Suggestion: Listing them all may be a bit tedious, but by using symmetry you can cut your work by a factor of four.

(c) Let $n > 0$ be an integer, and let C_n be the number of elements of A satisfying $|z| \le n$. Prove $C_n < (2n + 1)^2$.

(d) Modify the idea of part (c) to prove $(n + 1)^2 < C_n$.

Exercise 2.11. In each part, let $A = \mathbf{Z} + i\mathbf{Z}$.

(a) Show A is closed under multiplication.

(b) Which elements of A have a reciprocal (multiplicative inverse) in A?

Exercise 2.12. (a) Suppose $A \subseteq \mathbf{Z} + i\mathbf{Z}$ is closed under addition and closed under taking negatives. Prove that if $1 \in A$ and $i \in A$, then $A = \mathbf{Z} + i\mathbf{Z}$.

(b) Prove that $(1 + i)\mathbf{Z} + (3 + 2i)\mathbf{Z} = \mathbf{Z} + i\mathbf{Z}$.

Exercise 2.13. Each part refers to the set

$$\mathbf{Q}[\sqrt{2}] = \mathbf{Q} + \mathbf{Q}\sqrt{2} = \{m + n\sqrt{2} : m, n \in \mathbf{Q}\}.$$

(a) Show $\mathbf{Q}[\sqrt{2}]$ is closed under addition. Suggestion: Compare Example 2.22.

(b) Show $\mathbf{Q}[\sqrt{2}]$ is closed under multiplication.

(c) Show that if $\alpha = m + n\sqrt{n}$ is a non-zero element of $\mathbf{Q}[\sqrt{2}]$, there exists a unique $\alpha' = m' + n'\sqrt{2}$ in $\mathbf{Q}[\sqrt{2}]$ such that $\alpha\alpha' = 1$.

Exercise 2.14. Let $\zeta = e^{2\pi i/3} = \frac{1}{2}(-1 + \sqrt{3})$, and let $A = \mathbf{Z} + \zeta\mathbf{Z}$.

(a) Give a formal definition of the set A.

(b) Prove A is closed under addition.

(c) Prove A is closed under multiplication. (This depends on the specific value ζ.)

(d) Show that $U_6 = U(1) \cap A$, and illustrate with a sketch.

(e) Which elements of A have a reciprocal in A? Explain.

Exercise 2.15. Let A and B be subsets of $\mathcal{U}$.

(a) Prove $A \subseteq B$ if and only if $B^c \subseteq A^c$, and illustrate with a Venn diagram.

(b) How is part (a) related to contrapositives?

Exercise 2.16. Let A, B, and C be subsets of a universe $\mathcal{U}$, and let P, Q, and R be the predicates $x \in A$, $x \in B$, and $x \in C$. Use truth tables to prove:

(a) $(A \cup B) \cup C = A \cup (B \cup C)$.

(b) $(A \cap B) \cap C = A \cap (B \cap C)$.

Exercise 2.17. Let A, B, and C be subsets of a universe $\mathcal{U}$. As in Exercise 2.16, use truth tables to establish *De Morgan's laws* (a) and (b) and the *distributive laws* (c) and (d).

(a) $(A \cup B)^c = A^c \cap B^c$.

(b) $(A \cap B)^c = A^c \cup B^c$.

(c) $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$.

(d) $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$.

Exercise 2.18. Draw Venn diagrams illustrating each part of the preceding exercise, and compare with Exercise 1.14.

Exercise 2.19. (a) Let $A = \{a, b, c\}$ be a set with three distinct elements. List the elements of the power set $\mathcal{P}(A)$.

(b) How would your answer to part (a) differ if $A = \{0, 1, 2\}$?

(c) Describe how you could use your answer to part (a) to list the elements of the power set of $A' = \{a, b, c, d\}$. Suggestion: There are two types of subsets of A', those having d as an element and those not having d as an element.

Exercise 2.20. Let A and B be subsets of $\mathcal{U}$.

(a) Suppose $A \subseteq B$. Prove $\mathcal{P}(A) \subseteq \mathcal{P}(B)$ as subsets of $\mathcal{P}(\mathcal{U})$.

(b) Suppose that $\mathcal{P}(A) = \mathcal{P}(B)$ as subsets of $\mathcal{P}(\mathcal{U})$. Prove $A = B$.

Exercise 2.21. Let A and B be subsets of $\mathcal{U}$. Their *difference* is defined to be $A \setminus B = \{x \text{ in } A : x \notin B\}$.

(a) Prove $A \setminus B = A \cap B^c$, and illustrate with a Venn diagram.

(b) List the elements of $\mathbf{Z} \setminus \mathbf{Z}^+$ between -5 and 5.

(c) List the elements of $2\mathbf{Z} \setminus 3\mathbf{Z}$ between -12 and 12.

(d) List the elements of $3\mathbf{Z} \setminus 2\mathbf{Z}$ between -12 and 12.

Exercise 2.22. Let A and B be subsets of $\mathcal{U}$. Their *symmetric difference* is defined to be $A \triangle B = (A \setminus B) \cup (B \setminus A)$.

(a) Prove $A \triangle B = (A \cup B) \setminus (A \cap B)$ and illustrate with a Venn diagram.

(b) Prove $A \triangle B = \{x \text{ in } \mathcal{U} : x \in A \text{ or } x \in B \text{ but not both}\}$. This condition is called *exclusive or*, denoted “xor”.

Exercise 2.23. Explain in detail what is wrong with this two-column “proof” that $-1 = 1$.

$$
\begin{aligned}
-1 &= 1 && \text{to be shown,}\\
(-1)^2 &= 1^2 && \text{square both sides,}\\
1 &= 1 && \text{true statement.}
\end{aligned}
$$

Therefore $-1 = 1$.

Exercise 2.24. Let a and b denote real numbers, and assume $a = b$.

(a) What is wrong with the following “proof” that $2 = 1$?

$$
\begin{aligned}
b^2 &= ab && a = b,\\
b^2 - a^2 &= ab - a^2 && \text{subtract } a^2,\\
(b + a)(b - a) &= a(b - a) && \text{factor each side,}\\
(b + a) &= a && \text{cancel common factor,}\\
2a &= a && a = b,\\
2 &= 1 && \text{cancel common factor.}
\end{aligned}
$$

(b) If the proof is read from bottom to top, is each step valid?

Chapter 4

Mappings and Relations

As in calculus, a “function” or “mapping” is a rule associating a unique “output” to each “input”. While this intuitive description is adequate for informal work, rigorous mathematics requires more precision: The sets of allowable inputs and potential outputs must be made an intrinsic part of a function.

Example 4.1. Consider the familiar squaring function $f(x) = x^2$, where x ranges over the set of real numbers. If we set $y = f(x)$, we might wish to “solve” for x in terms of y. At first glance this is trivial: set $x = \sqrt{y}$. Unfortunately, closer inspection reveals two fatal flaws. First, if $y < 0$, there is no real x satisfying $x^2 = y$. In this context, the square root is *undefined*. Second, if $y > 0$, there exist *two* values of x with $x^2 = y$; the input x is not a function of the output y, so the square root is *not well-defined*. In either event, we have not associated a unique output to each input.

In high school, you learned to avoid complications with square roots by only considering non-negative numbers y, and agreeing that $\sqrt{y}$ always refers to the non-negative square root. Technically you are no longer inverting the function $f(x) = x^2$ with x real, but a *different function defined by the same formula*, for which the allowable inputs and potential outputs have been explicitly restricted.

Remark 4.2. The squaring function is arguably artificial in this respect, but for other familiar functions, such as the circular trig functions, the inability to invert causes genuine annoyances. Consider longitude (measured in degrees) as a function of position on the earth. Upon circumnavigating the earth to the east, longitude increases by 360°. But this cannot be the whole story; if it were, each geographic location would have multiple longitudes, any two differing by a whole multiple of 360°.

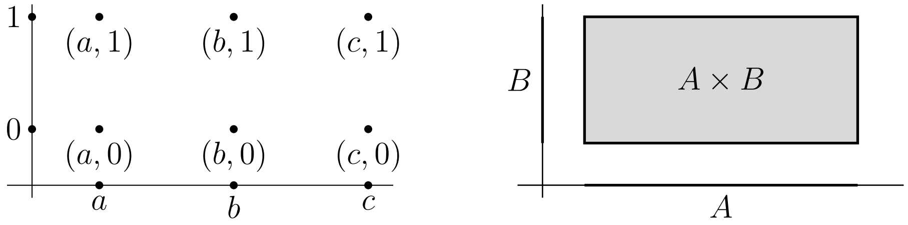


Figure 4.1. Cartesian products.

Instead, when you circumnavigate the globe in an eastward direction, you must cross a line where longitude "jumps down" by 360°. This discontinuity is a mathematical artifact of the impossibility of inverting sine and cosine to recover longitude continuously as a real-valued function of position on the earth.

The earth is approximately spherical and rotates with respect to the distant stars. A *sidereal day*, or 24 hours, is the time required for the earth to rotate 360° with respect to the stars. This *duration* is the same for all points on the earth, but the *starting time* (midnight) depends on one's longitude. By international treaty, the earth's surface is divided into twenty-four *time zones*, each a sector of longitude 15° wide (with substantial allowances for geographical and political boundaries). The times in neighboring zones differ by one hour.

The global discontinuity of longitude has a notable practical consequence: the existence of the International Date Line, an imaginary "cut" along the surface of the earth joining the south and north poles, along which local time "jumps" by 24 hours, affecting global travelers and international stock traders alike.

4.1. Mappings, Images, and Preimages

Before giving the formal definition of a mapping, we need to construct an appropriate set universe.

Definition 4.3. Let A and B be sets. Their *Cartesian product* $A{\times}B$ is the set of all "ordered pairs" from A and B,

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

Example 4.4. The Cartesian plane $\mathbf{R}{\times}\mathbf{R} = \mathbf{R}^2$ is the set of ordered pairs of real numbers. Similarly, $\mathbf{R}{\times}\mathbf{R}{\times}\mathbf{R}$, or $\mathbf{R}^3$, is Cartesian space, the set of ordered triples of real numbers.

Example 4.5. If $A = \{a,b,c\}$ and $B = \{0,1\}$, then the Cartesian product $A{\times}B$ is the six-element set $\{(a,0),(b,0),(c,0),(a,1),(b,1),(c,1)\}$ in the left-hand diagram in Figure 4.1.

For the same set B, $B{\times}B = \{(0,0),(1,0),(0,1),(1,1)\}$.

Example 4.6. If $A = \varnothing$ or $B = \varnothing$, then $A \times B = \varnothing$.

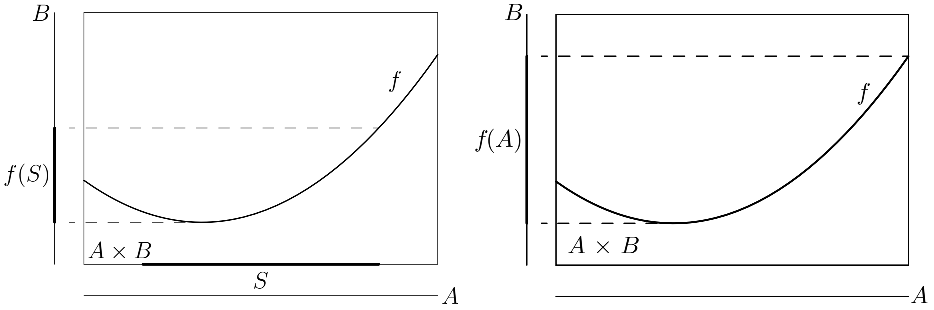


Figure 4.2. The image of a set under a mapping.

An abstract Cartesian product can be visualized conveniently by depicting the set A on a horizontal axis and the set B on a vertical axis, and taking the set of points lying above or below A *and* to the left or right of B. The right-hand diagram in Figure 4.1 shows the case where A and B are intervals.

Definition 4.7. A *mapping* f from A to B is a subset f of $A{\times}B$ satisfying the following condition:

For every a in A, there exists a unique b in B such that $(a, b) \in f$.

The set A is the *domain* of f, and B is the *codomain*. We write $b = f(a)$, and call b the *value* of f at a. We also say a is *mapped to* b by f, or that f *maps* a to b.

The notation $f : A \to B$, read "f from A to B", signifies that f is a mapping from A to B. A mapping $f : A \to B$ associates a unique value b in the codomain to each element a of the domain. If the Cartesian product $A{\times}B$ is viewed as a rectangle, a mapping is a "graph" in the sense of calculus, namely a subset intersecting each vertical line in the rectangle exactly once. The vertical line at horizontal position a intersects f (a.k.a. the graph of f) at location $b = f(a)$.

Remark 4.8. If A is non-empty, there exists no mapping $f : A \to \varnothing$.

If B is arbitrary (empty or not), there is a unique mapping $f : \varnothing \to B$.

Definition 4.9. Let $f : A \to B$ be a mapping. If $S \subseteq A$, we define the *image* of S under f to be the set

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

see Figure 4.2. In particular, the *image of* f is the set $f(A) \subseteq B$ of all values of f.

Definition 4.10. Let $f : A \to B$ be a mapping. If $T \subseteq B$, we define the *preimage* of T under f to be the set

$$f^{-1}(T) = \{a \text{ in } A : f(a) \in T\} \subseteq A$$

of elements of the domain mapped into T by f; see Figure 4.3.

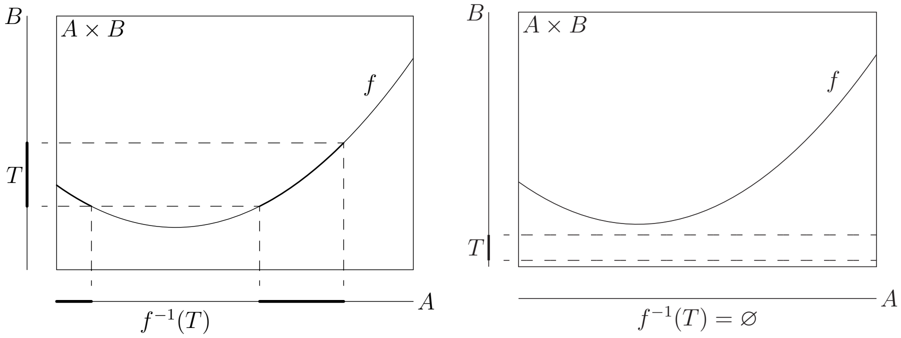


Figure 4.3. The preimage of a set.

Remark 4.11. A mapping $f : A \to B$ may be viewed as a "poll" taken of a population A, with responses in the set B. The image under f of a set $S \subseteq A$ is the set of responses from individuals in S. The preimage of a set $T \subseteq B$ is the set of individuals whose responses are in T.

Example 4.12. If A is a non-empty set, we define the *identity mapping* $I_A : A \to A$ by $I_A(a) = a$ for all a in A. Under the identity mapping, every set is its own image and its own preimage.

Example 4.13. Let A and B be non-empty sets. For each b in B, there is a *constant mapping* $c_b : A \to B$ defined by $c_b(a) = b$ for all a in A. The image of an arbitrary non-empty subset of A is the singleton $\{b\}$. The preimage of a set T is either the empty set (if $b \notin T$) or the entire domain A (if $b \in T$).

Proposition 4.14. *Let $f : A \to B$ be a mapping, S_1 and S_2 subsets of A, and T_1 and T_2 subsets of B. Then*

(i) $f(S_1 \cup S_2) = f(S_1) \cup f(S_2)$.

(ii) $f(S_1 \cap S_2) \subseteq f(S_1) \cap f(S_2)$.

(iii) $f^{-1}(T_1 \cup T_2) = f^{-1}(T_1) \cup f^{-1}(T_2)$.

(iv) $f^{-1}(T_1 \cap T_2) = f^{-1}(T_1) \cap f^{-1}(T_2)$.

Proof. To prove two sets are equal, we must establish inclusions in both directions. Assume S_1 and S_2 are subsets of A.

(The inclusion $f(S_1 \cup S_2) \subseteq f(S_1) \cup f(S_2)$). If $b \in f(S_1 \cup S_2)$, then by definition there exists an element a in $S_1 \cup S_2$ such that $f(a) = b$. Since either $a \in S_1$ or $a \in S_2$ by definition of the union of sets, either $b \in f(S_1)$ or $b \in f(S_2)$, which means $b \in f(S_1) \cup f(S_2)$. This proves $f(S_1 \cup S_2) \subseteq f(S_1) \cup f(S_2)$.

(The inclusion $f(S_1) \cup f(S_2) \subseteq f(S_1 \cup S_2)$). If $b \in f(S_1) \cup f(S_2)$, there exists an a in $S_1 \subseteq S_1 \cup S_2$ such that $f(a) = b$ or there exists an a in $S_2 \subseteq S_1 \cup S_2$ such that

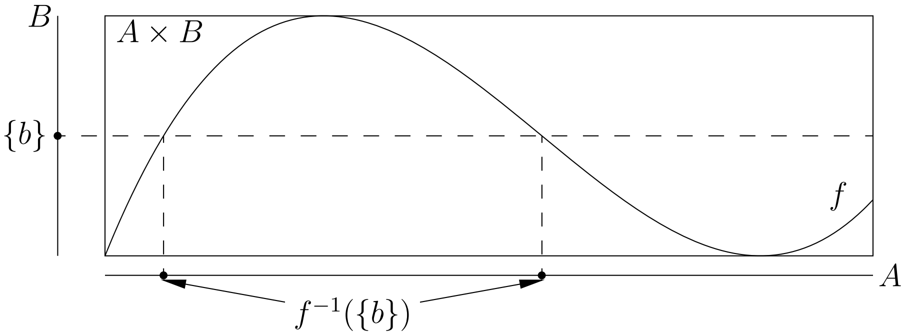


Figure 4.4. The preimage of a point under a surjective mapping.

$f(a) = b$. In either case, there exists an a in $S_1 \cup S_2$ such that $f(a) = b$, which means $b \in f(S_1 \cup S_2)$. This proves $f(S_1) \cup f(S_2) \subseteq f(S_1 \cup S_2)$.

The other parts are entirely similar and are left to you. □

4.2. Surjectivity and Injectivity

Our inability to invert the map $f : \mathbf{R} \to \mathbf{R}$, $f(x) = x^2$, had two aspects: When we wrote $y = f(x)$, some y were associated with no values of x, and some were associated with multiple values of x.

Definition 4.15. A mapping $f : A \to B$ is *surjective* if for every b in B, there exists an a in A such that $f(a) = b$.

Remark 4.16. A mapping $f : A \to B$ is surjective if, for every b in B, the equation $f(a) = b$ can be solved (perhaps non-uniquely) for a in A. In other words, f is surjective if every element of the codomain is a value of f.

In terms of sets, f is surjective if the preimage of $T = \{b\}$ is non-empty for each b in B, or the image of f is the entire codomain, $f(A) = B$; see Figure 4.4.

Geometrically, a mapping $f : \mathbf{R} \to \mathbf{R}$ is surjective if every horizontal line hits the graph of f at least once.

Definition 4.17. Let $f : A \to B$ be a mapping. Points a_1 and a_2 in A are *identified by f* if $f(a_1) = f(a_2)$, namely if a_1 and a_2 are mapped to the same value by f.

Definition 4.18. A mapping f is *injective* if $f(a_1) = f(a_2)$ implies $a_1 = a_2$. Contrapositively, if $a_1 \neq a_2$, then $f(a_1) \neq f(a_2)$.

Remark 4.19. A mapping f is injective if and only if f does not identify any pairs of distinct elements. Equivalently, the preimage of an arbitrary singleton $T = \{b\}$ contains at most one element.

Geometrically, $f : \mathbf{R} \to \mathbf{R}$ is injective if every horizontal line hits the graph of f at most once.

Remark 4.20. Continuing Remark 4.11, a mapping $f : A \to B$ is surjective if every allowable answer to the poll is given by at least one individual. Similarly, f is injective if no two people give the same response; knowledge of the response uniquely determines the individual who gave that response.

Definition 4.21. A mapping $f : A \to B$ is *bijective* if f is both surjective and injective.

Remark 4.22. A surjective mapping is sometimes called a *surjection*. An *injection* and a *bijection* are defined similarly.

Remark 4.23. If $f : A \to B$ is bijective, then each element a in A corresponds to exactly one element b in B. For this reason, many authors use the phrase "one to one correspondence" to connote a bijection. We avoid this name, since it can be easily confused with a "one to one mapping", an alternative name for an injection.

Remark 4.24. If $f : A \to B$ is bijective, the equation $f(a) = b$ can be solved uniquely for each b in B. Procedurally, f "relabels" elements of the set A using elements of B as names.

Example 4.25. Define $f_1 : \mathbf{R} \to [0, \infty)$ by $f_1(x) = x^2$. This mapping is surjective (every non-negative real y can be written as $x^2 = f_1(x)$ for at least one real x), but not injective (since $f_1(-1) = 1 = f_1(1)$, but $-1 \neq 1$).

Example 4.26. Define $f_2 : (0, \infty) \to \mathbf{R}$ by $f_2(x) = x^2$. This mapping is not surjective (there is no real x such that $x^2 = f_2(x) = -1$), but *is* injective. To establish injectivity, suppose $a_1^2 = f_2(a_1) = f_2(a_2) = a_2^2$. Subtracting and factoring, we find $0 = a_2^2 - a_1^2 = (a_2 - a_1)(a_2 + a_1)$, which implies $a_1 = a_2$ or $a_1 = -a_2$. The latter is impossible since a_1 and a_2 are positive by hypothesis.

We have shown that if $f_2(a_1) = f_2(a_2)$, then $a_1 = a_2$. Since a_1 and a_2 were arbitrary, f_2 is injective.

Note carefully that the mappings f_1 and f_2 in these examples are defined by the same formula, but have distinct domains and/or codomains.

Example 4.27. Let $\zeta = e^{2\pi i/3} = \frac{1}{2}(-1 + i\sqrt{3})$, and consider $A = \{1, \zeta, \zeta^2\} \subseteq \mathbf{C}^\times$. Since ζ is a cube root of unity, $(\zeta^2)^2 = \zeta^4 = \zeta$ and $(\zeta^2)^3 = 1$.

The mapping $f : A \to A$ defined by $f(z) = z^2$ is bijective: $1 = f(1)$, $\zeta = f(\zeta^2)$, and $\zeta^2 = f(\zeta)$.

The mapping $g : A \to A$ defined by $f(z) = z^3$ is neither injective nor surjective. Indeed, $f(z) = 1$ for every z in A.

Example 4.28. Define $f : \mathbf{Z} \to \mathbf{Z}$ by $f(a) = 1 - a$. Prove f is bijective.

(Injectivity). Let a_1 and a_2 be arbitrary integers, and assume that $f(a_1) = f(a_2)$. By the definition of f, $1 - a_1 = 1 - a_2$, so $a_1 = a_2$ by elementary algebra. Since a_1 and a_2 were arbitrary, f is injective.

(Surjectivity). Informally, we wish to solve $b = f(a) = 1 - a$ for a in terms of b. Rearrangement gives $a = 1 - b$.

Formally, let b be an arbitrary integer, and consider the integer $a = 1 - b$. Since $f(a) = f(1-b) = 1-(1-b) = b$, we have shown that for every integer b, there exists an integer a such that $f(a) = b$. This means f is surjective.

Example 4.29. Let $f : \mathbf{Z} \to \mathbf{Z}$ be defined by $f(a) = 1 - 2a$. Prove that f is injective (one-to-one) but not surjective (onto).

(Injectivity). Let a_1 and a_2 be integers, and assume $f(a_1) = f(a_2)$, i.e., that $1 - 2a_1 = 1 - 2a_2$. Subtracting the left side from the right, $0 = 2a_1 - 2a_2 = 2(a_1 - a_2)$. By Theorem 3.13 (ii), $a_1 - a_2 = 0$ as well. Since $f(a_1) = f(a_2)$ implies $a_1 = a_2$, f is injective.

(Non-surjectivity). To show that f is not surjective, it suffices to exhibit an integer not in the image of f. Let $b = 0$. The equation $f(a) = b$ becomes $1-2a = 0$, or $1 = 2a$. There exists no integer a satisfying this condition, which means 0 is not in the image of f.

Example 4.30. Define $f : \mathbf{Z}^+ \to \mathbf{Z}$ by

$$f(a) = \begin{cases} k = \dfrac{a-1}{2} & \text{if } a = 2k+1 \text{ is odd,} \\ -k = -\dfrac{a}{2} & \text{if } a = 2k \text{ is even.} \end{cases}$$

Prove f is bijective. (Informally, there are just as many positive integers as there are integers!)

(Initial exploration). To understand f intuitively, list its first several values. The inputs (elements of the domain) are $1, 2, 3, 4, \dots$. To find an output, determine whether the input is even or odd, and evaluate the corresponding formula. Thus $f(1) = 0$ (1 is odd), $f(2) = -1$ (2 is even), $f(3) = 1$, $f(4) = -2$, $f(5) = 2$, and so forth:

a	1	2	3	4	5	6	7	8	9
$f(a)$	0	-1	1	-2	2	-3	3	-4	4

In other words, f alternately “counts off” one negative, one positive. Using arrows to indicate successive values: Since the same value is never achieved twice, f is injective.

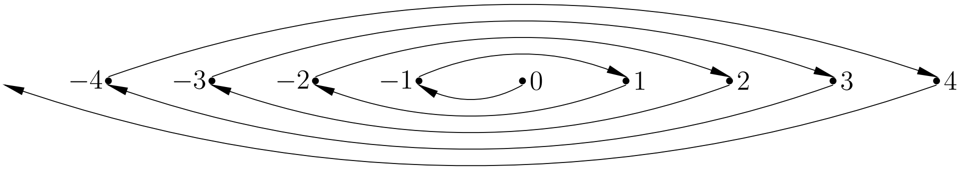


Figure 4.5. Counting the integers.

Since every integer value is achieved, f is surjective. We must convert this intuition into a formal proof.

(Injectivity). Let a_1 and a_2 be integers, and assume $f(a_1) = f(a_2)$. Because f is defined "piecewise", it's most convenient to consider three separate cases.

Case 1: a_1 and a_2 both odd. By hypothesis and the definition of f, $(a_1 - 1)/2 = (a_2 - 1)/2$, and elementary algebra implies $a_1 = a_2$.

Case 2: a_1 and a_2 both even. Here, $-a_1/2 = -a_2/2$, and again we find $a_1 = a_2$.

Case 3: a_1 and a_2 have opposite parity (one is odd, one is even). Without loss of generality, we may assume a_1 is odd and a_2 is even. (Otherwise, swap their names.) Since $f(a_2) < 0 \leq f(a_1)$, the hypothesis $f(a_1) = f(a_2)$ is false. Said contrapositively, if $f(a_1) = f(a_2)$, we are not in Case 3.

Since the conclusion $a_1 = a_2$ followed in each case, we have shown that f is injective.

(Surjectivity). Let b be an arbitrary integer, and consider two cases:

Case 1: $b < 0$. Let $a = -2b$. Since a is an even integer, we have $f(a) = -a/2 = b$; there exists an a such that $f(a) = b$ provided $b < 0$.

Case 2: $0 \leq b$. Let $a = 1 + 2b$. Since a is odd, $f(a) = (a - 1)/2 = 2b/2 = b$; there exists an a such that $f(a) = b$ provided $0 \leq b$.

Since every integer b is either negative or non-negative, we have handled all possibilities. In each case, there exists an integer a such that $f(a) = b$, so f is onto.

Example 4.31. Let A be an arbitrary set, and let $\mathcal{P}(A)$ be its power set. The following argument of G. Cantor shows there is no surjection $f : A \to \mathcal{P}(A)$.

Let $f : A \to \mathcal{P}(A)$ be an arbitrary mapping. For each a in A, the value $f(a)$ is a *subset* of A, so the statement $a \in f(a)$ is meaningful for each a. Let

$$T = \{a \text{ in } A : a \notin f(a)\}.$$

To prove f is not surjective, it suffices to show $f(t) \neq T$ for all t in A. We will prove that if $f(t) = T$ for some t, then set theory is logically inconsistent. Contrapositively, if set theory is logically consistent, then $f(t) \neq T$ for all t in A.

If $f(t) = T$, we may ask which alternative is true: $t \notin T$ or $t \in T$. By the definition of T, if $t \in f(t) = T$, then t fails to satisfy the criterion for membership in T, so $t \notin T$. However, if $t \notin f(t) = T$, then t satisfies the criterion of membership, so $t \in T$. In summary, the statement $t \in T$ is logically equivalent to its negation $t \notin T$. This completes the proof.

4.3. Composition and Inversion

Definition 4.32. Let $f : A \to B$ and $g : B \to C$ be mappings. Their *composition* is the mapping $g \circ f : A \to C$ defined by

$$(g \circ f)(a) = g\big(f(a)\big) \quad \text{for each } a \text{ in } A.$$

In this situation we say g *is composable with* f.

Remark 4.33. In other words, plug the output of f into g, obtaining $(g \circ f)(a)$.

When context clearly signifies composition, the operator symbol $\circ$ may be omitted, and the composition $g \circ f$ denoted gf.

Proposition 4.34 (Mapping composition is associative)**.** *If $f : A \to B$, $g : B \to C$, and $h : C \to D$ are composable, then $h(gf) = (hg)f$ as mappings from A to D.*

Proof. If a is an arbitrary element of A, then

$$\begin{aligned}\big[h \circ (g \circ f)\big](a) &= h\big[(g \circ f)(a)\big] = h\big[g\big(f(a)\big)\big] \\ &= (h \circ g)\big(f(a)\big) = \big[(h \circ g) \circ f\big](a).\end{aligned}$$ □

Surjectivity and injectivity of mappings f and g are related to whether or not the composition gf is surjective and/or injective. Think of two functions "cooperating", with g acting on the output of f. If f achieves every value in B and g achieves every value in C, then in tandem they achieve every value in C. Similarly, if neither g nor f identifies any pair of distinct points, then gf does not either. Before reading further, you should express these observations formally as logical implications and try to prove them.

Proposition 4.35. *Let $f : A \to B$ and $g : B \to C$ be mappings.*

(i) *If f and g are surjective, then gf is surjective.*

(ii) *If f and g are injective, then gf is injective.*

Proof. (i) Suppose $f : A \to B$ and $g : B \to C$ are surjective. Let c in C be arbitrary. Because g is surjective, there exists a b in B such that $g(b) = c$. Since f is surjective, there is an a in A such that $f(a) = b$. But $(gf)(a) = g\big(f(a)\big) = g(b) = c$. We have shown that for every c in C, there exists an a in A such that $(gf)(a) = c$, which by definition means gf is surjective.

(ii) Exercise 4.4 (a). □

Conversely, suppose we know that gf is surjective, or that gf is injective. What can we deduce about f and g?

In our cooperation metaphor, if gf achieves every value in C, then g itself must as well, since any value not achieved by g is certainly not achieved by gf. Thus, if gf is surjective, then g is surjective.

Similarly, if f identifies some pair of points, then gf identifies that pair as well, since g cannot split asunder what f has joined. Formally, if gf is injective, then f is injective, Exercise 4.4 (b).

The following examples show that nothing more can be deduced.

Example 4.36. Let $f : [-1, 1] \to \mathbf{R}$ and $g : \mathbf{R} \to [-1, 1]$ be defined by $f(x) = \arcsin x$, $g(x) = \sin x$. The mapping f is injective but not surjective (why?), g is surjective but not injective (why?), while $gf : [-1, 1] \to [-1, 1]$ is the identity map (which is bijective), and $fg : \mathbf{R} \to \mathbf{R}$ is neither injective nor surjective.

Example 4.37. An arbitrary mapping $f : A \to B$ can be "factored" into the composition of an injection followed by a surjection. Define $\gamma_f : A \to A{\times}B$ and $\pi_2 : A{\times}B \to B$ by

$$\gamma_f(a) = \big(a, f(a)\big), \qquad \pi_2(a,b) = b.$$

Geometrically, "γ_f lifts a to the graph of f" and "π_2 projects $A{\times}B$ onto the second factor". Clearly, $f = \pi_2 \circ \gamma_f : A \to B$, γ_f is injective, and π_2 is surjective.

Inversion of Mappings. In algebra, "inversion" generally refers to *undoing*. For addition, inversion means subtraction. For multiplication, inversion refers to division. For mappings, inversion refers to composition.

Definition 4.38. Let A and B be sets. A mapping $f : A \to B$ is *invertible* if there exists a mapping $g : B \to A$ that *inverts* f, i.e., such that $g \circ f$ is the identity map of A and $f \circ g$ is the identity map of B.

Remark 4.39. If $f : A \to B$ is invertible and the mapping $g : B \to A$ inverts f, then $(g \circ f)(a) = a$ for all a in A and $(f \circ g)(b) = b$ for all b in B. That is,

For all a in A and all b in B, $g(b) = a$ if and only if $f(a) = b$.

Proposition 4.40. *Let A and B be sets, and let $f : A \to B$ a mapping.*

(i) *f is invertible if and only if f is bijective.*

(ii) *If f is invertible, there exists a unique map inverting f.*

Remark 4.41. Both conclusions hold (with essentially vacuous proof) if either A or B is the empty set. It suffices to assume A, B are non-empty.

Proof. (i) Assume f is invertible, and let g be a mapping that inverts f, i.e., that satisfies $gf = I_A$ and $fg = I_B$. If $f(a_1) = f(a_2)$ for some a_1 and a_2 in A, applying g to both sides gives $a_1 = a_2$, so f is injective. If b is an arbitrary element of b, and if $a = g(b)$, then $f(a) = (fg)(b) = b$, so f is surjective.

Conversely, suppose f is bijective: For each b in B, there exists a unique a in A such that $b = f(a)$. Define $g(b) = a$. This prescription defines a mapping $g : B \to A$ that satisfies the condition in Remark 4.39, so f is invertible.

(ii) If $g_1, g_2 : B \to A$ invert f, then

$$g_1 = g_1 \circ I_B = g_1 \circ (f \circ g_2) = (g_1 \circ f) \circ g_2 = I_B \circ g_2 = g_2. \qquad \square$$

A mapping f is invertible if and only if f is injective and surjective. We now consider what happens if each condition holds individually.

Left Inverses. Assume f is one-to-one; not every element of B need be a value of f, but every value (every element of $f(A)$, the image of A under f) is achieved exactly once. We may define $h : f(A) \to A$ analogously to Remark 4.39: For all b in $f(A)$, $h(b) = a$ if and only if $f(a) = b$.

If we apply f to a in A, then apply h to $b = f(a)$, we find

$$(hf)(a) = h\big(f(a)\big) = h(b) = a \qquad \text{for all } a \text{ in } A.$$

That is, $hf = I_A$, the identity map on A; h is a *left inverse* of f.[1]

In order to obtain a mapping $g : B \to A$ satisfying $gf = I_A$, we must "enlarge" the domain of h; any convenient "rule" will do. For example, pick an element a_0 in A arbitrarily, and define, for b in B,

$$g(b) = \begin{cases} a & \text{if } b = f(a) \text{ for some } a \text{ in } A, \\ a_0 & \text{otherwise.} \end{cases}$$

The easy verification that $gf = I_A$ is left to you.

Example 4.42. Define $f : \mathbf{R} \to \mathbf{R}$ by $f(x) = e^x$; see Figure 4.6, left. For each $y > 0$ (namely, for each y in the image of f), we have $y = e^x$ if and only if $x = \ln y$. Define

$$g(y) = \begin{cases} \ln y & \text{if } y > 0, \\ 0 & \text{if } y \le 0; \end{cases}$$

see Figure 4.6, right. Then $(gf)(x) = g\big(f(x)\big) = x$ for all x; what about $f\big(g(y)\big)$?

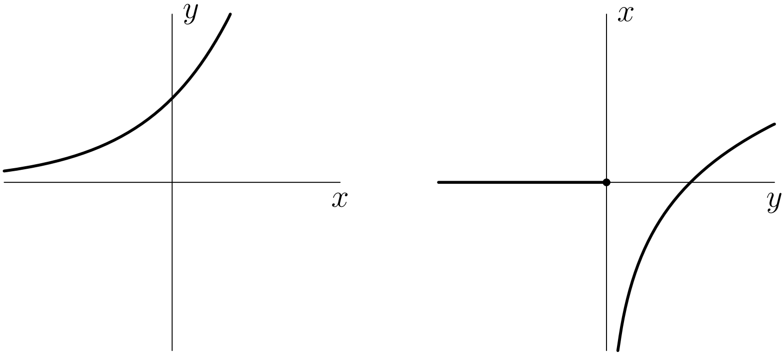


Figure 4.6. A left inverse of $f(x) = e^x$.

Right Inverses. Assume f is onto; every element of B is a value of f, but some values b may be achieved at distinct points: $f(a_1) = f(a_2)$ but $a_1 \ne a_2$. Define $g : B \to A$ by the following prescription: For each b in B, use the Axiom of Choice to pick an a in A such that $f(a) = b$, and define $g(b) = a$.[2]

It is straightforward to check that $fg = I_B$, the identity map on B.[3] Any particular g defined this way is called a *branch* of f^{-1}.

[1] In general, $fh \ne I_B$, the identity map on B, since (i) h is defined only on the image of f, and (ii) the image of fh, which is a subset of the image of f, may be a *proper* subset of B.

[2] The Axiom of Choice asserts that if $\{S_i\}_{i\in I}$ is a collection of non-empty sets indexed by a set I, it is possible to choose, for each i in I, an element x_i of S_i.

[3] In general, $gf \ne I_A$, the identity map on A, since if $f(a_1) = b = f(a_2)$ for $a_1 \ne a_2$, we cannot have both $g(b) = a_1$ and $g(b) = a_2$.

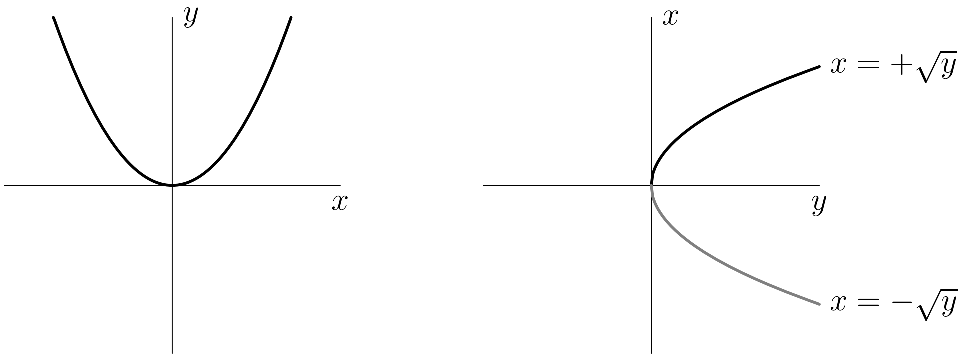


Figure 4.7. Right inverses of $f(x) = x^2$.

Example 4.43. Define $f : \mathbf{R} \to [0, \infty)$ by $f(x) = x^2$; see Figure 4.7, left. For each $y > 0$, there exist two real x such that $f(x) = y$, namely $x = \pm\sqrt{y}$. In particular, there are two "obvious" branches of f^{-1}, defined by $g_{\pm}(y) = \pm\sqrt{y}$ for $y \geq 0$; see Figure 4.7, right. (There are infinitely many other choices, though all are discontinuous.) For any such choice, $(fg)(y) = f\bigl(g(y)\bigr) = y$ for all $y \geq 0$. What about $g\bigl(f(x)\bigr)$?

4.4. Equivalence Relations

Definition 4.44. Let A be a non-empty set. A *relation* on A is a subset $R \subseteq A{\times}A$. Elements a and b of A are *R-related*, written aRb, if $(a, b) \in R$.

Remark 4.45. A binary relation is naturally associated to the *Boolean* (true/false-valued) function $\rho : A \times A \to \{T, F\}$ defined by $\rho(a, b) = aRb$.

Example 4.46. *Equality* on A is the *diagonal* $R = \Delta = \{(a, a) : a \in A\}$.

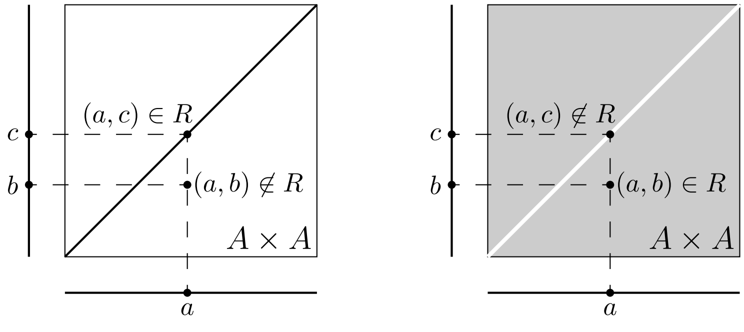


Figure 4.8. Equality and inequality: $a \neq b$ and $a = c$.

Example 4.47. The *inequality* relation $\neq$ is the complement of the equality relation, $A{\times}A \setminus \Delta = \{(a_1, a_2) : a_1 \neq a_2\}$, Figure 4.8, right.

Remark 4.48. Generally, if $R_2 = A \times A \setminus R_1$, the relations R_1 and R_2 are logical opposites: One relation holds for a pair of elements if and only if the other fails for the same pair.

Example 4.49. Let $A = \mathbf{Z}$ be the set of integers. The *less-than* relation is the set $R = \{(n_1, n_2) : n_1 < n_2\}$, Figure 4.9, left.

Example 4.50. Let A be the set $\mathbf{Z}$ of integers. The *parity* relation on $\mathbf{Z}$ is the set $R = \{(n_1, n_2) : n_2 - n_1 \text{ is even}\}$. Two integers are related if and only if they are both even or both odd, Figure 4.9, right.

Example 4.51. Let $f : A \to A$ be a mapping. Viewing f as a subset of $A \times A$ defines the *maps-to-under-f* relation on A: aRb if and only if $f(a) = b$, if and only if a maps to b under f.

Definition 4.52. Let R be a relation on a set A. We say R is

- *reflexive* if aRa for all a in A;
- *symmetric* if, for all a and b in A, aRb implies bRa;
- *transitive* if, for all a, b, and c in A, aRb and bRc imply aRc.

Example 4.53. Though not a formal example, the "friendship" relation may help you assimilate the conditions in the preceding definition. Let A be some set of people, and let aRb mean "b is a friend of a".

R is reflexive if and only if every person is their own friend; R is symmetric if and only if all friendships are mutual; R is transitive if and only if every friend-of-a-friend is a friend.

Definition 4.54. A reflexive, symmetric, and transitive relation is an *equivalence relation*.

If R is an equivalence on A and $a \in A$, the *equivalence class* of a is the set

$$[a] = \{x \text{ in } A : aRx\} \subseteq A$$

comprising all elements related to a.

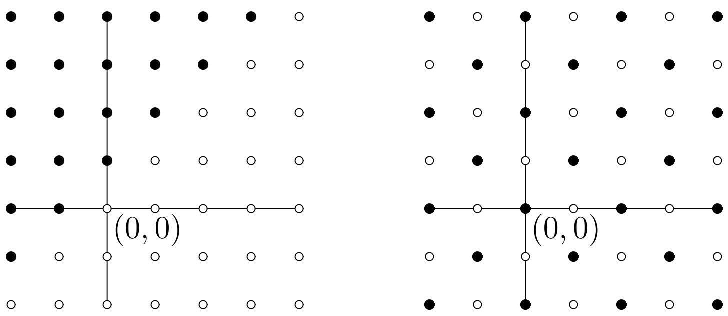


Figure 4.9. Less-than and parity on **Z**.

Example 4.55. Equality is an equivalence relation on an arbitrary set: For all a, b, and c, we have $a = a$ (reflexivity), $a = b$ implies $b = a$ (symmetry), and if $a = b$ and $b = c$, then $a = c$ (transitivity).

Inequality is symmetric, but neither reflexive nor transitive.

Example 4.56. Less-than is transitive (if $a < b$ and $b < c$, then $a < c$), but neither reflexive nor symmetric.

Example 4.57. The parity relation is an equivalence relation: For all integers a, b, and c, $a - a$ is even (reflexivity), if $b - a$ is even (aRb), then $a - b$ is even (bRa), and if $b - a$ and $c - b$ are even (aRb and bRc), then $c - a = (c - b) + (b - a)$ is even (aRc).

Equivalence Relations and Partitions. Let A be a non-empty set. Recall that a *partition* of A is a collection of non-empty, disjoint subsets whose union is A. Partitions and equivalence relations are two ways of viewing a single mathematical structure: Every equivalence relation gives rise to a partition, every partition gives rise to an equivalence relation, and these associations are inverse to each other.

Proposition 4.58. *Let R be an equivalence relation on A. The equivalence classes of R partition A.*

Proof. Since $a \in [a]$ for each a, every element of A lies in at least one equivalence class. It remains to prove that two arbitrary equivalence classes $[a]$ and $[b]$ are either disjoint or identical. To prove this it suffices to show that if $[a] \cap [b] \neq \varnothing$ (i.e., the classes are not disjoint), then $[a] = [b]$.

Let's first run through the argument using the friendship metaphor. If a and b have a friend in common, then a and b are themselves friends (transitivity). Consequently, every friend of a is a friend of b (transitivity again) and *vice versa*, so a and b have exactly the same set of friends.

Formally, if $[a]\cap[b] \neq \varnothing$, there exists a c in A such that $c \in [a]\cap[b]$. Consequently, aRc and bRc. By symmetry of R, aRc and cRb, and by transitivity aRb. This means $a \in [b]$ and $b \in [a]$.

It is now easy to prove $[a] \subseteq [b]$ and $[b] \subseteq [a]$: If $x \in [a]$, then xRa, and since aRb, transitivity guarantees xRb, meaning $x \in [b]$. Reversing the roles of a and b completes the argument.

We have shown that non-disjoint equivalence classes are identical, so the set of equivalence classes of R is indeed a partition of A. □

Remark 4.59. Conversely, if A is partitioned into subsets $\{A_i\}_{i\in I}$, there is an induced equivalence relation defined by aRb if and only if there exists an index i such that $a \in A_i$ and $b \in A_i$. Informally, aRb if and only if both elements lie in the same subset of the partition. Be sure to convince yourself that R is an equivalence relation, and that the partition induced by R is the original partition.

Example 4.60. The equivalence classes of the equality relation are the singletons, sets having one element: $[a] = \{a\}$ for each a in A.

Example 4.61. The parity relation on $\mathbf{Z}$ has two equivalence classes: $[0] = 2\mathbf{Z}$ and $[1] = 2\mathbf{Z} + 1$.

Partitions and Prejudice. Our minds organize the external world by categorizing, unconsciously identifying people, objects, or phenomena that share some attribute, or conversely, sharply distinguishing things that are nearly identical.

Example 4.62. A physicist, a statistician, and a mathematician saw a flock of 100 sheep, of which one was black. The physicist said, "We can deduce that one in 100 sheep is black." The statistician said, "No, only that *in this sample of* 100 *sheep*, one is black." The mathematician corrected, "No, we can only deduce that one sheep in this sample is black *on one side*."

Often we cope fluently with such hierarchies: a particular mandarin orange, mandarin oranges, oranges, citrus fruit, fruit.... At other times, prejudice deceives us into identifying individuals according to superficial characteristics (such as gender, ethnicity, religion, or scientific field) and incorrectly presuming "all such people are alike".

In mathematics, we can sometimes turn prejudice to good use. Perhaps we don't care which integer we're dealing with, but only if it's even or odd, or if it leaves a remainder of 5 on division by 12. Maybe we're dealing with pairs of points in the plane, but don't care where they're located, only that the second is located one unit to the right of the first. In such cases, an equivalence relation allows us to formalize our prejudice and discard irrelevant information.

Let A be a set, R an equivalence relation on A, and $\{A_i\}_{i\in I}$ the partition of A into equivalence classes. Each "index" i is associated with the non-empty set $A_i \subseteq A$, and the index set I is in bijective correspondence with the set of equivalence classes. We call the set of equivalence classes the *quotient* of A by R, denoted $I = A/R$ and read as "A modulo R" (or "A mod R" for short). *Elements* of A/R are *collections* of objects in A. The equivalence relation R is "unable to distinguish" elements of A_i, so when R "looks at" A it "sees" $I = A/R$.

Example 4.63. Two real numbers θ_1 and θ_2 determine the same longitude on the earth if and only if their difference is a multiple of one full turn, say 360°. To formalize this in the language of quotients, let $A = \mathbf{R}$ be the set of real numbers (a.k.a. the number line), and define the relation R by $\theta_1 R \theta_2$ if and only if $\theta_2 - \theta_1$ is an integer multiple of 360. By an argument entirely similar to that given for the parity relation in Example 4.55, R is an equivalence relation.

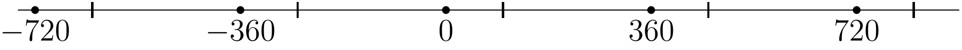


Figure 4.10. The number line, and the space of angles.

The set of equivalence classes is indexed by the half-open interval $[0, 360)$, since every angle is equivalent mod R to a unique number between 0 and 360 (excluding 360, which is equivalent to 0). We call this set the "space of angles", Figure 4.10.

Mappings and Equivalence Classes. Let A be non-empty, R an equivalence relation on A, and $f : A \to B$ a mapping. We will often be interested in trying to define an "induced" map $\bar{f}$ from the quotient set $\bar{A} = A/R$ to B.

Think of the elements of an equivalence class $[a]$ as a clique of friends who are polled by f, the question being "Which element of B do you map to?" If the clique responds unanimously ("We all map to b"), then by fiat $\bar{f}$ maps $[a]$ in $\bar{A}$ to b in B. If *every* clique reaches a unanimous decision, there is a mapping $\bar{f} : A/R \to B$ defined by $\bar{f}\big([a]\big) = f(a)$.

If the responses are mixed for some clique $[a]$, then $\bar{f}$ is undefined; a mapping must be single-valued for every input, but the members of $[a]$ do not decide unanimously where to be mapped by f.

Definition 4.64. Let $f : A \to B$ be a mapping, and R an equivalence relation on A. If aRa' implies $f(a) = f(a')$, we say f is *constant on equivalence classes* of R, or f is *well-defined* modulo R. If f is well-defined modulo R, we define the *induced mapping* $\bar{f} : A/R \to B$ by $\bar{f}\big([a]\big) = f(a)$ for each a in A.

Remark 4.65. If R is an equivalence relation on A, there is a "quotient mapping" $\Pi : A \to A/R$ defined by $\Pi(a) = [a]$. If $f : A \to B$ is well-defined modulo R and $\bar{f} : A/R \to B$ denotes the induced mapping, then $f = \bar{f} \circ \Pi$. We say "f factors through the quotient A/R".

Example 4.66. Let $A = \mathbf{Z}$ be the set of integers, R the parity relation, and $f : \mathbf{Z} \to \{1, -1\}$ the mapping defined by $f(a) = (-1)^a$. Under f, every even integer maps to 1 and every odd integer maps to -1, so f is well-defined modulo parity. Intuitively, to compute $(-1)^a$ for some integer a, we only need to know whether a is even or odd.

The quotient space $A/R = \{[0], [1]\} = \{2\mathbf{Z}, 2\mathbf{Z} + 1\}$ is a set having two elements, and the induced map $\bar{f} : A/R = \{2\mathbf{Z}, 2\mathbf{Z} + 1\} \to \{1, -1\}$, defined by

$$\bar{f}\big([0]\big) = (-1)^0 = 1, \qquad \bar{f}\big([1]\big) = (-1)^1 = -1,$$

is bijective.

Example 4.67. Let $A = \mathbf{Z}$, R the parity relation, and $f : \mathbf{Z} \to \mathbf{Z}$ defined by $f(a) = a^2$. The integers 0 and 2 are elements of $[0]$, but $f(0) = 0 \neq 4 = f(2)$. Thus f is not well-defined modulo parity.

This should be no surprise: To compute the square of an integer a, it is not enough to know whether a is even or odd.

Example 4.68. Let $A = \mathbf{R}$ be the set of real numbers, R the "longitude" relation, and define $f : \mathbf{R} \to \mathbf{R}^2$ by $f(t) = (\cos t, \sin t)$, the standard trigonometric parametrization of the circle, with trig functions in "degrees mode".

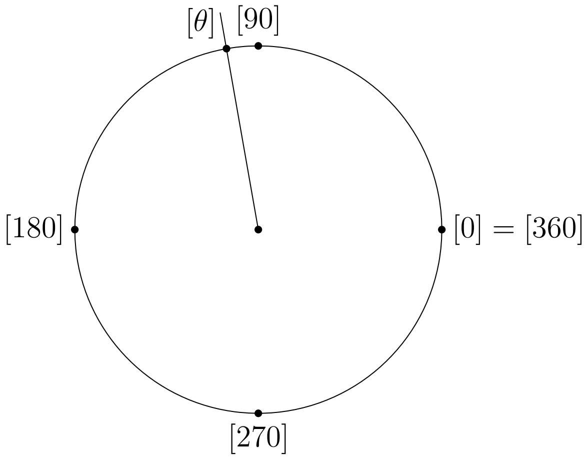


Figure 4.11. The space of angles as a circle.

If $\theta_2 - \theta_1$ is an integer multiple of 360, then $\cos\theta_1 = \cos\theta_2$ and $\sin\theta_1 = \sin\theta_2$, so $f(\theta_1) = f(\theta_2)$. Consequently, there is an induced mapping from the space of angles to the unit circle in the plane. In words, f factors through locations on the earth.

Since $\cos\theta_1 = \cos\theta_2$ and $\sin\theta_1 = \sin\theta_2$ if *and only if* $\theta_2 - \theta_1$ is an integer multiple of 360, the mapping $\bar{f}$ is bijective, so *the space of angles may be regarded as the unit circle*. Geometrically, each equivalence class $[\theta]$ corresponds to a unique point of the unit circle, Figure 4.11

Exercises

Exercise 4.1. Let $f : A \to B$ be a mapping, and let U and V be subsets of A. Prove the following:

(a) If $U \subseteq V$, then $f(U) \subseteq f(V)$.

(b) If f is injective and $f(U) \subseteq f(V)$, then $U \subseteq V$.

Exercise 4.2. Let $f : A \to B$ and $g : B \to C$ be mappings, and assume $S \subseteq A$, $T \subseteq C$.

(a) Prove $g\big(f(S)\big) = (gf)(S)$.

(b) Prove $f^{-1}\big(g^{-1}(T)\big) = (gf)^{-1}(T)$.

Exercise 4.3. Let $f : A \to B$ be a mapping.

(a) Assume $T \subseteq B$ is arbitrary. Prove $f\big(f^{-1}(T)\big) \subseteq T$, and that equality holds if f is surjective. Give an example of a mapping f and a set T for which the inclusion is proper.

(b) Assume $S \subseteq A$ is arbitrary. Prove $S \subseteq f^{-1}\big(f(S)\big)$, and that equality holds if f is injective. Give an example of a mapping f and a set S for which the inclusion is proper.

Exercise 4.4. (a) Let $f : A \to B$ and $g : B \to C$ be injective mappings. Prove gf is injective. (This is the second assertion of Proposition 4.35.)

(b) Suppose gf is injective. Prove f is injective.
Suggestion: Prove the contrapositive.

Exercise 4.5. Let m, n, and q be positive integers.

(a) Let A be a set containing m elements, B a set containing n elements, and assume $m > nq$. Prove that if $f : A \to B$ is a mapping, then there exists a b in B such that $f^{-1}(\{b\})$ contains at least $q + 1$ elements. (This result is known as the *Pigeonhole Principle*. If you distribute $m > nq$ pigeons among n holes, then some hole contains more than q pigeons.)
Suggestion: Write B as a union of singleton sets, and use Proposition 4.14. The contrapositive may be more natural to prove.

(b) With the same notation, let $f : A \to B$ be a mapping. Prove that if f is injective, then $m \leq n$, and that if f is surjective, then $m \geq n$. Show by example that both converse statements are false.

(c) With the same notation, assume $m = n$, and let $f : A \to B$ be a mapping. Prove that f is injective if and only if f is surjective. (Suggestion: Use part (b) to prove that f is injective if and only if $f(A)$ contains m elements, if and only if f is surjective.)

Exercise 4.6. Let $f : A \to B$ be a mapping. If $S \subseteq A$, define the *restriction* of f to S to be the mapping $f|_S : S \to B$ defined by $f|_S(a) = f(a)$ for all a in S.

(a) Prove that f is injective if and only if $f|_S$ is injective for *every* subset S of A.

(b) Assume f is a bijection. Prove that if S is a non-empty subset of A, then the restriction $f|_S$ is a bijection from S to $f(S)$ and the restriction $f|_{A\setminus S}$ is a bijection from $A \setminus S$ to $B \setminus f(S)$.

Exercise 4.7. Let m and b be integers, and define a mapping $f : \mathbf{Z} \to \mathbf{Z}$ by $f(x) = mx + b$.

(a) Prove f is injective if and only if $m \neq 0$.

(b) Find necessary and sufficient conditions on m and b for f to be surjective. If f is bijective, find a formula for the inverse mapping.

Exercise 4.8. Let $f : A \to B$ be an arbitrary mapping, and define mappings $\Gamma_f : A \to A \times B$ and $\Pi : A \times B \to B$ by

$$\Gamma_f(a) = \big(a, f(a)\big), \qquad \Pi(a, b) = b.$$

Prove that Γ_f is injective, Π is surjective, and $f = \Pi \circ \Gamma_f$. Illustrate with a sketch. (In other words, every mapping factors as an injection followed by a surjection.)

Exercise 4.9. Define $f : \mathbf{C} \to \mathbf{C}$ by $f(z) = z^2$.

(a) By writing $z = x + iy$ with x and y real, calculate the real and imaginary parts of $f(z)$.

(b) By writing $z = re^{i\theta}$ with $r \geq 0$ and θ real, re-calculate $f(z)$, and use your result to describe the geometric action of the mapping f.

(c) Find the preimages of the singletons $\{1\}$, $\{-1\}$, $\{i\}$, and $\{\rho e^{i\phi}\}$.

(d) Let $A = \{z \text{ in } \mathbf{C} : \operatorname{Re} z > 0\} \cup \{z \text{ in } \mathbf{C} : \operatorname{Re} z = 0, 0 \leq \operatorname{Im} z\}$. Show that f maps A bijectively to $\mathbf{C}$.

(e) By part (d), there exists a branch of inverse $g : \mathbf{C} \to A$ of f. Give a formula for $g(z)$ assuming $z = re^{i\theta}$ (specify necessary restrictions on θ), and show $h(z) = -g(z)$ is another branch of f^{-1}.

Exercise 4.10. Repeat parts (a)–(c) of the preceding exercise for the mapping $f : \mathbf{C} \to \mathbf{C}$ defined by $f(z) = z^3$.

Exercise 4.11. Let $n > 1$ be an integer, and define $f : \mathbf{C} \to \mathbf{C}$ by $f(z) = z^n$. By writing $z = re^{i\theta}$, describe the geometric action of f, and find the preimage of $\{\rho e^{i\phi}\}$. If $\rho > 0$, how many points are in the preimage, and how are these points situated geometrically in $\mathbf{C}$?

Exercise 4.12. Let $U_5 = \{e^{2\pi ki/5} : 0 \leq k \leq 4\}$ be the set of fifth roots of unity. Show that the formula $f(z) = z^2$ defines a mapping $f : U_5 \to U_5$, find the image of f, and determine whether or not f is bijective.

Exercise 4.13. Let $U_6 = \{e^{2\pi ki/6} : 0 \leq k \leq 5\}$ be the set of sixth roots of unity. Show that the formula $f(z) = z^2$ defines a mapping $f : U_6 \to U_6$, find the image of f, and determine whether or not f is bijective.

Exercise 4.14. Show that the mapping $f : [-1, 1] \to (-1, 1)$ defined by

$$f(x) = \begin{cases} \frac{x}{2} & x = \pm 2^{-n} \text{ for some integer } n \geq 0, \\ x & \text{otherwise} \end{cases}$$

is a bijection, and sketch the graph.

Exercise 4.15. Let $g : \mathbf{R} \to \mathbf{R}$ be a real-valued function of one real variable. We say that g is *even* if $g(-x) = g(x)$ for all x in $\mathbf{R}$, and that g is *odd* if $g(-x) = -g(x)$ for all x in $\mathbf{R}$. (Analogous formulas define the notions of "even" and "odd" functions whose domain and/or codomain is $\mathbf{Z}$ or any other set in which negatives are defined.)

(a) Find all functions that are *both* even and odd.

(b) Let $f : \mathbf{R} \to \mathbf{R}$ be an arbitrary function. Show that the functions

$$f_{\text{even}}(x) = \tfrac{1}{2}\big[f(x) + f(-x)\big], \qquad f_{\text{odd}}(x) = \tfrac{1}{2}\big[f(x) - f(-x)\big]$$

are even and odd, respectively.

(c) Suppose there exist an even function E and an odd function O such that $f(x) = E(x) + O(x)$ for all real x. Find formulas for E and O. Hint: Compute $f(-x)$.

(d) Prove that every function $f : \mathbf{R} \to \mathbf{R}$ can be written *uniquely* as the sum of an even function and an odd function. These functions are called the *even part* and *odd part* of f.

(e) Find the even and odd parts of $f(x) = x^3 - 2x^2 + x + 1$, $g(x) = e^x$, and $h(x) = \cos x$.

Exercise 4.16. Suppose $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ is a polynomial.

(a) Prove the following are equivalent: (i) p is even. (ii) $a_{2k+1} = 0$ for all k. (iii) There exists a polynomial q such that $p(x) = q(x^2)$.

(b) State and prove a similar characterization of odd polynomials.

Exercise 4.17. The *hyperbolic functions* cosh and sinh are defined by

$$\cosh x = \tfrac{1}{2}(e^x + e^{-x}), \quad \sinh x = \tfrac{1}{2}(e^x - e^{-x}), \qquad x \text{ real}.$$

(a) Show that $\cosh^2 - \sinh^2 = 1$. Carefully sketch the graphs of cosh and sinh on a single set of axes. Suggestion: First calculate $\cosh \pm \sinh$.

(b) Show that for all real x,

$$\cosh(2x) = \cosh^2 x + \sinh^2 x, \qquad \sinh(2x) = 2\cosh x \sinh x.$$

(c) Show that $\cosh' = \sinh$ and $\sinh' = \cosh$.

(d) Find algebraic formulas for $\sinh^{-1}$, and for two branches of $\cosh^{-1}$. Use algebra to show that the branches of $\cosh^{-1}$ differ by a sign.
Hint: Solve (e.g.) $y = \sinh x$ for x by multiplying through by e^x and rearranging to get a quadratic in e^x; then use the quadratic formula.

Exercise 4.18. The *hyperbolic tangent* and *hyperbolic secant* functions are

$$\tanh = \frac{\sinh}{\cosh}, \qquad \operatorname{sech} = \frac{1}{\cosh^2}.$$

(a) Carefully sketch the graphs of tanh and sech on a single set of axes. Show that $\tanh^2 = 1 + \operatorname{sech}^2$, and find formulas for $\tanh'$ and sech'.

(b) Find an algebraic formula for the inverse function $\tanh^{-1}$. Hint: Solve $y = \tanh x$ for x by cross-multiplying and rearranging.

Exercise 4.19. Let x and y be arbitrary real numbers. Show that

$$\cosh(x+y) = \cosh x \cosh y + \sinh x \sinh y,$$
$$\sinh(x+y) = \sinh x \cosh y + \cosh x \sinh y,$$
$$\tanh(x+y) = \frac{\tanh x + \tanh y}{1 + \tanh x \tanh y}.$$

Exercise 4.20. Let ϕ be a real number, and recall Euler's formula

$$e^{i\phi} = \cos\phi + i \sin\phi.$$

(a) Express $e^{-i\phi}$ in terms of $\cos\phi$ and $\sin\phi$.

(b) Show that

$$\cos\phi = \frac{e^{i\phi} + e^{-i\phi}}{2}, \qquad \sin\phi = \frac{e^{i\phi} - e^{-i\phi}}{2i}.$$

(c) Show that for all real ϕ,

$$\cosh(i\phi) = \cos\phi, \qquad \sinh(i\phi) = i\sin\phi.$$

(The hyperbolic functions are defined in Exercise 4.17.)

Exercise 4.21. Let A be a non-empty set, and let $R = \varnothing \subseteq A{\times}A$. Prove R is symmetric and transitive, but not reflexive.

Exercise 4.22. Let A be a set of people, and define a binary relation R by aRb if and only if a trusts b. What does it mean to say that R is reflexive? R is symmetric? R is transitive?

Exercise 4.23. Define a relation R on $\mathbf{Z}$ by aRb if and only if $|a| = |b|$.

(a) Prove R is an equivalence relation.

(b) Let $f : \mathbf{Z} \to \mathbf{Z}$ be defined by $f(a) = a^2$. Is f well-defined mod R?

(c) Let $g : \mathbf{Z} \to \mathbf{Z}$ be defined by $g(a) = 3a$. Is g well-defined mod R?

(d) Prove $f : \mathbf{Z} \to \mathbf{Z}$ is well-defined mod R if and only if f is an even function; see Exercise 4.15.

Exercise 4.24. Let $A = \mathbf{Z}$, and define a relation R by aRb if and only if $b - a$ is an integer multiple of 4.

(a) Prove R is an equivalence relation; describe the equivalence classes of R and the quotient $\mathbf{Z}/R$.

(b) Let $f : \mathbf{Z} \to \mathbf{C}$ be defined by $f(a) = (-1)^a$. Prove f is well-defined mod R. Is $\bar{f}$ injective?

(c) Let $g : \mathbf{Z} \to \mathbf{C}$ be defined by $g(a) = i^a$. Is g well-defined mod R? If so, is $\bar{g}$ injective?

Exercise 4.25. Fix an integer $n \geq 1$, and define a relation R on $\mathbf{Z}$ by aRb if and only if $b - a$ is an integer multiple of n.

(a) Show that R is an equivalence relation.

(b) Show that the equivalence classes of R are precisely the sets $A_k = [k]_n$ of Exercise 3.15.

(c) For each n with $3 \leq n \leq 7$, write out the integers from -10 to 10 inclusive, and using n colors or n symbols of your choosing, mark each integer in your list according to its equivalence class mod n.

Exercise 4.26. Let $f : A \to A$ be a mapping, and suppose the “maps-to” relation, aRb if and only if $b = f(a)$, is an equivalence relation. What can you say about f?

Exercise 4.27. If R_1 and R_2 are equivalence relations on a set A, we say R_1 *is finer than* R_2 if for all a and a' in A, aR_1a' implies aR_2a'.

(a) Show that R_1 is finer than R_2 if and only if $R_1 \subseteq R_2$, if and only if "R_1 is more discriminating than R_2".

(b) If R is an equivalence relation on A and $f : A \to B$ is a mapping, prove that f is well-defined mod R if and only if R is finer than the "identified by f" relation a_1Ra_2 if and only if $f(a_1) = f(a_2)$.

Exercise 4.28. Let R be an equivalence relation on A. If $f : A \to B$ is a mapping such that a_1Ra_2 if and only if $f(a_1) = f(a_2)$, i.e., whose level sets are precisely the equivalence classes of R, prove that the induced mapping $\bar{f} : A/R \to B$ is injective.

Exercise 4.29. Let $f : A \to B$ be a mapping, and define a relation on A by a_1Ra_2 if and only if $f(a_1) = f(a_2)$; see also Exercise 4.27.

(a) Prove R is an equivalence relation, and the equivalence classes of R are preimages of singletons, namely *level sets* of f: $f^{-1}(\{b\})$ for some b in B.

(b) Let $f : \mathbf{R} \to \mathbf{R}$ be defined by $f(x) = x^2$. Describe the equivalence classes of f.

(c) Let $f : \mathbf{R}^2 \to \mathbf{R}$ be defined by $f(x, y) = x^2 + y^2$. Describe the equivalence classes of f.

Exercise 4.30. Let $f : A \to B$ be a mapping. A mapping $g : A \to B$ is said to be *constant on the level sets of* f if $f(a_1) = f(a_2)$ implies $g(a_1) = g(a_2)$. (Compare the preceding three exercises.)

(a) Define $f : \mathbf{R}^2 \to \mathbf{R}$ by $f(x, y) = x^2 + y^2$. Which of the following are constant on the level sets of f?

$$g_1(x, y) = \left(1 - \sqrt{x^2 + y^2}\right)^2, \qquad g_2(x, y) = x^2 - y^2, \qquad g_3(x, y) = 1.$$

(b) For a general mapping $f : A \to B$, prove the following are equivalent: (i) g is constant on the level sets of f. (ii) There exists a mapping $\phi : B \to B$ such that $g = \phi \circ f$.

Exercise 4.31. Fix an integer $n \geq 2$, let R be congruence mod n, the relation of Exercise 4.25, let $\mathbf{Z}_n = \mathbf{Z}/R$ denote the set of equivalence classes of R, and let $\Pi : \mathbf{Z} \to \mathbf{Z}_n$ be the quotient map.

(a) If a is an integer, show the mapping $f_a : \mathbf{Z} \to \mathbf{Z}_n$ defined by $f_a(x) = \Pi(ax)$ is well-defined mod R. (In other words, if xRy, then $(ax)R(ay)$.)

(b) Show there is a mapping $\bar{f}_a : \mathbf{Z}_n \to \mathbf{Z}_n$ satisfying $\bar{f}_a\big(\Pi(x)\big) = f_a(x)$ for all integers x.

(c) If $n = 8$, tabulate the values $f_a(x)$ for $0 \leq x < 8$ and $1 \leq a \leq 5$.

Exercise 4.32. In each part, either show that the indicated mapping exists or explain why the mapping does not exist. (The main point to check is whether the given formula is *well-defined*, i.e., returns a single value for a single input. This is an issue because a single input may have multiple representations.)

(a) $f : \mathbf{Z}_6 \to \mathbf{Z}_3$ defined by $f([x]_6) = [x]_3$.

(b) $f : \mathbf{Z}_6 \to \mathbf{Z}_4$ defined by $f([x]_6) = [x]_4$.

(c) $f : \mathbf{Z}_6 \to \mathbf{Z}_4$ defined by $f([x]_6) = [2x]_4$.

(d) $f : \mathbf{Z}_3 \to \mathbf{Z}_6$ defined by $f([x]_3) = [x]_6$.

(e) $f : \mathbf{Z}_3 \to \mathbf{Z}_6$ defined by $f([x]_3) = [2x]_6$.

(e) $f : \mathbf{Z}_3 \to \mathbf{Z}_6$ defined by $f([x]_3) = [3x]_6$.

Exercise 4.33. Let I_1 and I_2 be intervals of real numbers. A mapping $f : I_1 \to I_2$ is *increasing* if, for all x and y in I_1, $x < y$ implies $f(x) < f(y)$.

(a) Prove that a composition of increasing mappings is increasing.

(b) Prove that if f is an increasing bijection, then the inverse f^{-1} is an increasing bijection.

Exercise 4.34. Let $\mathcal{P}(\mathbf{C})$ denote the set of *complex paths*, i.e., mappings from some closed, bounded real interval I to $\mathbf{C}$. We say two paths $\gamma_1 : I_1 \to \mathbf{C}$ and $\gamma_2 : I_2 \to \mathbf{C}$ are *reparametrizations*, and write $\gamma_1 \sim \gamma_2$, if there exists an increasing bijection $\tau : I_1 \to I_2$ such that $\gamma_1 = \gamma_2 \circ \tau$. Prove that $\sim$ is an equivalence relation on $\mathcal{P}(\mathbf{C})$, and that equivalent paths have the same image.

Exercise 4.35. Imagine a world where the natural numbers are known, but the integers are not. For example, $2 = 1+x$ "has a solution", but $1 = 2+x$ "has no solution". For a time, mathematicians cope by inventing symbols, such as "-1", to denote "negative" numbers. These fictitious entities turn out to be so useful that logical care demands they be placed on a firm logical foundation.

This exercise outlines an implementation, taking its cue from the equation $n_1 = n_2 + x$. The goal is to construct—using only natural numbers and operations of set theory—a larger collection of "numbers" that contains a copy of $\mathbf{N}$ and in which the equation $m = n + x$ has a solution when m and n are numbers *of the more general type*.

Intuitively, an integer will be an *ordered pair of natural numbers*. The pair $x = (n_1, n_2)$ corresponds to the solution of $n_1 = n_2 + x$. For example, the pair $(1, 2)$ corresponds to the solution of $1 = 2 + x$, namely to the integer $x = -1$.

Many different pairs represent the same number; $(4, 1)$, $(9, 6)$, and $(1968, 1965)$ all correspond to 3. Two pairs (n_1, n_2) and (m_1, m_2) represent the same number exactly when $n_1 - n_2 = m_1 - m_2$, that is, when $n_1 + m_2 = n_2 + m_1$. We are therefore led to define the relation

$$(m_1, m_2) \sim (n_1, n_2) \quad \text{if and only if } n_1 + m_2 = n_2 + m_1. \tag{4.1}$$

(a) Prove that (4.1) defines an equivalence relation on the set $X = \mathbf{N}\times\mathbf{N}$.

(An *integer* is an equivalence class of $\mathbf{N}\times\mathbf{N}$ with respect to (4.1).)

(b) Define the *sum* of two integers by adding representatives:

$$(m_1, m_2) \oplus (n_1, n_2) = (m_1 + n_1, m_2 + n_2).$$

Show that "addition" $\oplus$ is well-defined modulo (4.1). Explicitly, if $(m_1, m_2) \sim (m_1', m_2')$ and $(n_1, n_2) \sim (n_1', n_2')$, then

$$(m_1, m_2) \oplus (n_1, n_2) \sim (m_1', m_2') \oplus (n_1', n_2').$$

(c) Motivated by the idea that (m_1, m_2) and (n_1, n_2) represent $m_1 - m_2$ and $n_1 - n_2$, define the *product* of two integers by

$$(m_1, m_2) \odot (n_1, n_2) = (m_1 n_1 + m_2 n_2, m_1 n_2 + n_1 m_2).$$

Show that "multiplication" $\odot$ is well-defined modulo (4.1).

(d) Using commutativity and associativity of addition and multiplication *of natural numbers*, verify that $\oplus$ and $\odot$ are commutative and associative.

(e) Show that if m and n are natural numbers, then

$$(m, 0) \oplus (n, 0) = (m + n, 0),$$
$$(m, 0) \odot (n, 0) = (mn, 0).$$

That is, the equivalence class $\big[(n, 0)\big]$ corresponds to the natural number n, so we have succeeded in building a copy of **N** inside **Z**.

(f) Show that $m = n + x$, i.e., $(m_1, m_2) = (n_1, n_2) \oplus (x_1, x_2)$, has a solution for all integers m and n.