Discrete Mathematics

UNIVERSITY OF CAMBRIDGE

Computer Laboratory

Computer Science Tripos Part 1A
Mathematics Tripos Part 1A (CS Option)

Peter Robinson

Michaelmas 2003 & Lent 2004

William Gates Building
15 JJ Thomson Avenue
Cambridge
CB3 0FD

http://www.cl.cam.ac.uk/

Introduction

This course will develop the idea of formal proof by way of examples involving simple objects such as integers and sets. The material enables academic study of Computer Science and will be promoted with examples from the analysis of algorithms and cryptography.

Syllabus

Integers

- Proof. Deduction, contradiction. Integers, mathematical induction. [2 lectures]
- Factors. Division: highest common factors and least common multiples. Euclid's algorithm: solution in integers of $ax + by = c$, the complexity of Euclid's algorithm. Euclid's proof of the infinity of primes. Existence and uniqueness of prime factorisation. Irrationality of $\sqrt{p}$. [3 lectures]
- Modular arithmetic. Congruences. Units modulo m, Euler's totient function. Chinese remainder theorem. Wilson's theorem. The Fermat-Euler theorems, testing for primes. Public key cryptography, Diffie-Hellman, RSA. [3 lectures]

Sets, relations and functions

- Sets, subsets and Boolean operations. Partitions. Boolean logic. [2 lectures]
- Binary relations. Composition of relations. Equivalence relations and quotients of sets. Closures and Warshall's algorithm. Partial orders and total orders. Hasse diagrams. Well founded relations and well ordering. Well founded induction. [3 lectures]
- Functions; Injective, surjective and bijective functions. Numbers of such functions between sets. Sorting. The Schröder-Bernstein theorem. Countability. A countable union of countable sets is countable. The uncountability of $\mathbb{R}$. Existence of transcendental numbers. [3 lectures]

Objectives

On completing the course, students should be able to:

- Write a clear statement of a problem as a theorem in mathematical notation.
- Prove and disprove assertions using a variety of techniques.
- Describe, analyse and use Euclid's algorithm.
- Explain and apply prime factorisation.
- Perform calculations with modular arithmetic.
- Use number theory to explain public key cryptography.
- Analyse problems using set theory.
- Formulate statements using Boolean logic.
- Recognise relations and discuss their properties.
- Describe and analyse Warshall's algorithm.
- State, prove and apply the Schröder-Bernstein theorem.
- Distinguish countable and uncountable sets.

Appropriate books

The following books are relevant for the course:

- NL Biggs: *Discrete Mathematics,* Oxford University Press, 1989, ISBN 0-19-853427-2, £22.95.
- JH Conway & RK Guy: *The book of numbers,* Springer-Verlag, 1996, ISBN 0-387-97993-X, £21.95
  A beautiful book – deeply subtle mathematics presented in an accessible and exciting way.
- H Davenport: *The higher arithmetic* (6th edition), Cambridge University Press, 1992, ISBN 0-521-42227-2, £14.95.
- P Giblin: *Primes and programming,* Cambridge University Press, 1993, ISBN 0-521-40988-8, £15.95.
- RL Graham, DE Knuth & O Patashnik: *Concrete mathematics* (2nd edition), Addison Wesley, 1994, ISBN 0-201-55802-5, £26.00.
  The ultimate reference book.
- JF Humphreys & MY Prest: *Numbers, groups and codes,* Cambridge University Press, 1989, ISBN 0-521-35938-4, £14.95.
  Close to the approach in this course.
- HF Mattson: *Discrete Mathematics,* Wiley, 1993, ISBN 0-471-59966-2, £21.95.
- N Nissanke: *Introductory logic and sets for Computer Scientists*, Addison-Wesley, 1999, ISBN 0-201-17957-1, £20.95.
- G Pólya: *How to solve it,* Penguin, 1990, ISBN 0-14-012499-3, £8.99.
- KH Rosen: *Discrete mathematics and its applications* (4th edition), McGraw-Hill, 1999, IBN 0-07-116756-0, £23.99.
  An excellent book covering a wide range of topics and useful throughout the course.

These notes do not constitute a complete transcript of all the lectures and they are not a substitute for text books. They are intended to give a reasonable synopsis of the subjects discussed, but they give neither complete proofs of all the theorems nor all the background material.

Proof

What is a proof? If a theorem is a logical statement, the proof is meant to convince you that the statement is true. When faced with a proof you should convince yourself of three things:

- The arguments put forward are all true and the sequence follows logically from beginning to end.
- The arguments are sufficient to prove the theorem.
- The arguments are all necessary to prove the theorem.

A proof has to encompass all the possible cases permitted by the statement of the proof. Usually it will not be possible to work through all of these in turn, so some generality will be required. On the other hand, a single counter-example *is* sufficient to show that a theorem is false. Indeed, such a counter-example should be as simple as possible. Good mathematicians like to avoid effort.

This should not be confused with proof by contradiction. This is an elegant technique in which we prove a theorem by accepting the possibility that it is not true. If it is not true, there must be a counter-example. Examining this counter-example then gives rise to a logical inconsistency. If all the intermediate steps are correct, the only explanation is that the original assumption (accepting that the theorem was not true) was itself mistaken. In other words, the theorem *is* true.

Examples

- **Theorem**: $a^n + b^n = c^n$ has no solutions.

  Proof: Left as an exercise for the reader.

- **Theorem:** The whole numbers that can be expressed as the difference of two squares are precisely those that leave a remainder of 0, 1 or 3 when divided by 4.

  Proof: Work through a sequence of simpler problems.

  a) Any odd number can be expressed as the difference of two squares – consider $(\mathrm{n}+1)^2 - \mathrm{n}^2$.

  b) No even number can be expressed as the difference of two squares – false, consider $4 = 2^2 - 0^2$.

  c) Any exact multiple of 4 can be expressed as the difference of two squares – consider $(n+1)^2 - (n-1)^2$.

  d) No odd multiple of two can be expressed as the difference of two squares – assume true and find a contradiction by examining cases.

  Now combine these results. (d) shows that any difference of two squares leaves a remainder of 0, 1 or 3 when divided by 4. (a) shows that a number that leaves remainder 0 when divided by 4 can be expressed as the difference of two squares, and (c) shows that a number that leaves a remainder of 1 or 3 can.

- **Theorem:** $\sqrt{2}$ is *irrational*, that is, it can not be written as a fraction $\frac{x}{y}$ for whole numbers x and y.

  Proof: Assume that $\sqrt{2} = \frac{x}{y}$ for whole numbers x and y. Without loss of generality, we can assume that x and y are not both even and deduce a contradiction.

How to solve it

Pólya suggests the following four step plan for problem solving:

Understanding the problem

What is the unknown? What are the data? What is the condition?

Is it possible to satisfy the condition? Is the condition sufficient to determine the unknown? Or is it insufficient? Or redundant? Or contradictory?

Draw a figure. Introduce suitable notation.

Separate the various parts of the condition. Can you write them down?

Devising a plan

Find the connection between the data and the unknown. You may be obliged to consider auxiliary problems if an immediate connection cannot be found. You should obtain eventually a plan of the solution.

Have you seen it before? Or have you seen the same problem in a slightly different form?

Do you know a related problem? Do you know a theorem that could be useful?

Look at the unknown! And try to think of a familiar problem having the same or a similar unknown.

Here is a problem related to yours and solved before. Could you use it? Could you use its results? Could you use its method? Should you introduce some auxiliary element in order to make its use possible?

Could you restate the problem? Could you restate it still differently? Go back to definitions.

If you cannot solve the proposed problem try to solve first some related problem. Could you imagine a more accessible related problem? A more general problem? A more special problem? An analogous problem? Could you solve a part of the problem? Keep only a part of the condition, drop the other part; how far is the unknown then determined, how can it vary? Could you derive something useful from the data? Could you think of other data appropriate to determine the unknown? Could you change the unknown or data, or both if necessary, so that the new unknown and the new data are nearer to each other?

Did you use all the data? Did you use the whole condition? Have you taken into account all essential notions involved in the problem?

Carrying out the plan

Carrying out your plan of the solution, check each step. Can you see clearly that the step is correct? Can you prove that it is correct?

Looking back

Can you check the result? Can you check the argument?

Can you derive the result differently? Can you see it at a glance?

Can you use the result, or the method, for some other problem?

Integers

We start with the sets of natural numbers, $\mathbb{N} = \{1, 2, 3, \ldots\}$, the natural numbers augmented with 0, $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$, and integers, $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, and will refer to the rational numbers (fractions), $\mathbb{Q}$, and the real numbers $\mathbb{R}$. The curly brackets just wrap up enumerations of elements. The empty set is $\emptyset = \{\}$. We will discuss the notation for sets more formally in the second half of the course, but here is enough to get started.

A particular value, x, is an *element* of a set X if it is in it. We write this with a sort of Greek epsilon: $x \in X$. So $-3 \in \mathbb{Z}$ but $-3 \notin \mathbb{N}$.

One set, X, is a *subset* of another set, Y, if every element of X is also an element of Y. We write this with a rounded less-than-or-equal sign: $X \subseteq Y$. So $\mathbb{N} \subseteq \mathbb{N}_0 \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$.

We can also define sets by *predicates* or conditions: $\mathbb{N} = \{x \in \mathbb{Z} \mid x > 0\}$. This notation is a bit unfortunate because we will also use the vertical bar to indicate exact divisibility: $3 \mid 6$. So the set of even numbers might be defined as $E = \{x \in \mathbb{Z} \mid 2|x\}$, which is a bit confusing. Sorry. The vertical bar is also used in another way to count the number of elements $|X|$ in a (finite) set.

There are two particularly important properties of the natural numbers, which turn out to be equivalent: induction and well-ordering.

Mathematical induction

Let $P(n)$ be any mathematical assertion involving the natural number n which may be true or false. (Think of P as a function with n as an argument and returning a Boolean result.) The principle of mathematical induction states that, if

- $P(1)$ is true, and
- whenever $P(k)$ is true then $P(k+1)$ is true as well

then $P(n)$ is true for every natural number n.

The two conditions are known as the *base case* and the *inductive step*, and they give rise to the *conclusion*.

Examples

- $1 + 2 + 3 + \cdots + n = \frac{1}{2} n(n+1)$.

  Base case: $1 = ½ \times 1 \times 2$.

  Inductive step: Suppose $1 + 2 + 3 + \ldots + k = ½ \times k \times (k+1)$.
  Then $1 + 2 + 3 + \ldots + k + (k+1) = ½ \times k \times (k+1) + (k+1) = ½ \times (k+1) \times (k+2)$.

- Let $a_n = 2^{3n+1} + 3^{n+1}$. Then, for all positive integers n, a_n is exactly divisible by 5.

  Base case: $a_1 = 2^4 + 3^2 = 16 + 9 = 25$, which is divisible by 5.

  Inductive step: Suppose a_k is divisible by 5.
  Then $a_{k+1} = 2^{3(k+1)+1} + 3^{(k+1)+1} = 8 \times 2^{3k+1} + 3 \times 3^{k+1} = 5 \times 2^{3k+1} + 3 \times a_k$, which is divisible by 5.

- If n is a positive integer and x and y are any numbers, then
  $$(x+y)^n = \binom{n}{0}x^n + \binom{n}{1}x^{n-1}y^1 + \cdots + \binom{n}{i}x^{n-i}y^i + \cdots + \binom{n}{n}y^n$$
  where $\binom{n}{k}$ is the *binomial coefficient*, defined to be $\frac{n!}{k!(n-k)!}$ with $0! = 1$.

  Base case: $\binom{1}{0} = \binom{1}{1} = 1$, so $(x+y)^1 = x + y = \binom{1}{0}x^1 + \binom{1}{1}y^1$.

  Inductive step: Observe that $\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$ by direct algebra.

  Assume the expansion for $(x+y)^k$ and multiply it by $(x+y)$ to produce $(x+y)^{k+1}$ and group terms with the same powers of x and y in the sum.

An alternative statement of the principle (known as *course of values* induction) states that, whenever $\mathrm{P}(k)$ can be inferred from the truth of $\mathrm{P}(j)$ for all $j < k$, then $\mathrm{P}(n)$ is true for every natural number n. Note that $\mathrm{P}(1)$ is true since there are no natural numbers j with $j < 1$.

Well ordering

Any non-empty subset of $\mathbb{N}$ contains a smallest element.

That may seem obvious, but it is not true for the integers, rationals or reals. It is also an important property that will extend to other sets where each element does not have a natural successor and so ordinary induction can not be used. However, some sort of ordering relation $\leq$ is still necessary. We can use well ordering to prove results in a way similar to induction.

Fundamental theorem of arithmetic

A natural number p is *prime* if $p > 1$ and p is only divisible by 1 and itself. Every natural number greater than 1 can be expressed as a product of primes.

Proof: Use contradiction. Let $\mathrm{S} = \{n \in \mathbb{N} \mid n \text{ can not be expressed as a product of primes}\}$. S is not empty (or there would be no counter-examples). Let $s \in \mathrm{S}$ be its smallest element. s can not be prime, since it is in S. So $s = ab$ for some $a, b \in \mathbb{N}$ with $1 < a, b < s$. a and b are smaller than the least element of S, and so can not be in S. Write them as products of primes and combine them to give an expression for s.

We will be studying prime numbers in (much) more detail later.

Exercises

1. Prove that $1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{1}{6}n(n+1)(2n+1)$.

2. Find the sum of the first n cubes. Calculate the first few cases, formulate a general rule and confirm it by induction.

3. Evaluate the sum $\frac{1}{2!} + \frac{2}{3!} + \frac{3}{4!} + \cdots + \frac{n}{(n+1)!}$.

4. Show that 7 divides $2^{4n+2} + 3^{2n+1}$ and 13 divides $3^{n+1} + 4^{2n-1}$ for all natural numbers n.

5. The *Fibonacci* numbers are defined by $f_0 = 0, f_1 = 1$ and $f_n = f_{n-1} + f_{n-2}$ for $n > 1$.

   Show that $f_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right)$ for all $n \geq 0$.

   Hint: If using induction, you need to consider two base cases.

6. Prove that, for all $n \in \mathbb{N}_0$ and $x \in \mathbb{R}$ with $x \geq -1$, $(1+x)^n \geq 1 + n\,x$.

7. A *triomino* is an L-shaped pattern made from three square tiles. A $2^k \times 2^k$ chessboard, whose squares are the same size as the tiles, has an arbitrary square painted purple. Show that the chessboard can be covered with triominoes so that only the purple square is exposed.

8. A prison houses 100 inmates, one in each of 100 cells, guarded by a total of 100 warders. One evening, all the cells are locked and the keys left in the locks. As the first warder leaves, she turns every key, unlocking all the doors. The second warder turns every second key, re-locking every even numbered cell. The third warder turns every third key and so on. Finally the last warder turns the key in just the last cell. Which doors are left unlocked and why?

   Hint: This is a question about division.

9. Let $\mathrm{S} = \{1, 2, \ldots, n\}$. Write $\sum_{s \in S} f(s)$ for the sum $\sum_{s=1}^{n} f(s)$ and $\prod_{s \in S} f(s)$ for the product similarly, with the convention that the empty product $\prod_{s \in \phi} f(s) = 1$. For example, with $n = 2$, $\sum_{s \in S} f(s) = f(1) + f(2)$ and $\prod_{s \in S} f(s) = f(1) \times f(2)$.

   Use induction to prove that $\prod_{s \in S}(1 + x_s) = \sum_{T \subseteq S} \prod_{t \in T} x_t$, where all the $x_i \in \mathbb{R}$ and the sum is taken over all possible subsets $\mathrm{T} \subseteq \mathrm{S}$. Again, for $n = 2$, the left hand side is $(1+x_1)(1+x_2)$, and the possible values for S on the right hand side are Ø, $\{1\}$, $\{2\}$, and $\{1, 2\}$, giving corresponding products of 1, x_1, x_2, and x_1x_2.

   Deduce that $\prod_{s \in S}(1 - x_s) = \sum_{T \subseteq S} (-1)^{|T|} \prod_{t \in T} x_t$.

10. [Mathematical Tripos Part 1A 1988, Paper 6, Question 9]

    State the principle of mathematical induction. Prove your statement, assuming that every non-empty subset of the natural numbers contains a least element.

    Hint: Consider an assertion $\mathrm{P}(n)$ that satisfies the two conditions for mathematical induction. So $\mathrm{P}(1)$ is true and $\mathrm{P}(k)$ implies $\mathrm{P}(k+1)$. You need to show that $\mathrm{P}(n)$ is true for every natural number n. Use contradiction. Consider the set $\mathrm{S} = \{x \in \mathbb{N} \mid \mathrm{P}(x) \text{ is false}\}$. Show that S can not be empty and so has a least element. Call it s. Show that $s \neq 1$ and consider $\mathrm{P}(s-1)$.

    The Master of Regents' College and his wife invite n Fellows and their spouses to a party. After the party the Master asks everyone (including his own wife) how many people they shook hands with, and receives $2n + 1$ different answers. Of course, no woman shook hands with her own husband. Show that the person who shook the most hands was not the Master's wife.

    How many hands did the Master shake?

    Hint: Consider the largest and smallest numbers of people with whom a guest could shake hands. What does this tell you about the answers that the Master received? What does this tell you about the relationship between the person who shook most hands and the person who shook least?

11. Use induction to prove that the natural numbers are well-ordered, thereby completing the proof that the two properties are equivalent.

    Hint: Use contradiction. Suppose that X is a set of natural numbers which contains no least element. You need to prove that X is empty. Let L be the set of natural numbers n such that n is not greater than or equal to any element in X. Show by induction that L is the set of natural numbers, so X is, indeed, empty.

12. [Not to be taken too seriously.] Comment on the following alleged proofs by induction (with acknowledgements to Professor JWS Cassels):

    - Let n be a natural number and a_j be real numbers for $1 \leq j \leq n$. Then $a_j = a_k$ for $1 \leq j \leq n$, $1 \leq k \leq n$.

      Proof Certainly true for $n = 1$. Assume the result is true for n and prove it for $n+1$. By case n of the result, we have $a_1 = a_2 = \cdots = a_n$. Applying this to the a_{j+1} instead of the a_j we have $a_2 = \cdots = a_n = a_{n+1}$. Hence $a_1 = a_2 = \cdots = a_n = a_{n+1}$, which is the result for $n+1$.

    - Every natural number n is interesting.

      Proof There certainly are some interesting natural numbers: 0 is the smallest, 1 is the only natural number whose reciprocal is a natural number, 2 is the smallest prime, 3 is the number of persons in the Trinity, and so on. So, if the statement were false, there would be a smallest natural number n which is not interesting. This is a contradiction, since n would be a very interesting number indeed.

    - Every odd integer > 1 is prime.

      Proof The economist's proof runs as follows. 3 is prime, 5 is prime, 7 is prime. Three cases in a row is surely enough.

      If, however, we imagine an idealised economist who would not be satisfied by this, then the rest of the proof would continue as follows: Look at the next odd integer, 9. Well, it is admittedly not a prime; there must be some unusual factor of some kind operating. Let's go on looking at the figures. 11 is prime, 13 is prime. Two more confirmations, so it must be true.

- Every prime is odd.

  Proof $3, 5, 7, 11, 13, 17, 19, \ldots$ are all odd. There only remains 2, which must be the oddest prime of all.

- $n^2 - n + 41$ is prime for all natural numbers n.

  Proof The physicist's proof runs as follows. Write a computer program to check successively that $n^2 - n + 41$ is prime for $n = 0, 1, 2, \ldots 40$. Since quite a number of cases have now been verified using very expensive equipment, the result must be true.

Factors

The operations of addition, multiplication and ordering on the integers have some useful properties.

Division

Given integers a and b, we say that a *divides* b or a is a *factor* of b (written $a \mid b$) if $b = qa$ for some integer q. Moreover, a is a *proper divisor* of b if $a \mid b$ and $a \neq \pm 1$ or $\pm b$.

Observations

- If $a \mid b$ and $b \mid c$ then $a \mid c$.

  Proof: If $a \mid b$ and $b \mid c$ then $b = q\,a$ and $c = r\,b$ for some q and r.
  So $c = (r\,q)\,a$ and $a \mid c$.

- If $d \mid a$ and $d \mid b$ then $d \mid (a\,x + b\,y)$ for any integers x and y.

  Proof: If $d \mid a$ and $d \mid b$ then $a = q\,d$ and $b = r\,d$ for some q and r.
  So $a\,x + b\,y = q\,x\,d + r\,y\,d = (q\,x + r\,y)\,d$ and $d \mid (a\,x + b\,y)$.

  $(a\,x + b\,y)$ is called a *linear combination* of a and b (or of x and y).

Division algorithm

Given $a \in \mathbb{Z}$ and $b \in \mathbb{N}$, there exist unique integers $q, r \in \mathbb{Z}$ with $a = bq + r$ and $0 \leq r < b$. q is called the *quotient* and r is the *remainder* after dividing a by b. The latter is written as $a \bmod b$ or, sometimes, as a % b. So $b \mid a$ if, and only if, r = 0, that is, $a \bmod b = 0$.

Proof: *Existence.* Consider $\mathrm{R} = \{a - bk \mid k \in \mathbb{Z} \text{ and } (a - bk) \geq 0\}$. $\mathrm{R} \subseteq \mathbb{N}_0$ and is not empty, so use well ordering to find its smallest element, r. $r \in \mathrm{R}$, so $r \geq 0$ and we can write $r = a - bq$. Now $r < b$ or $r - b$ would be a smaller element of R.

Uniqueness. Suppose $a = bq_1 + r_1$ and $a = bq_2 + r_2$ with $0 \leq r_1 < b$ and $0 \leq r_2 < b$. Then $b(q_1 - q_2) + (r_1 - r_2) = 0$, but $-b < (r_1 - r_2) < b$, so $r_1 = r_2$ and $q_1 = q_2$.

This is not actually an algorithm in the normal sense understood by computer scientists, but there are algorithms that implement division in hardware or software. The important mathematical result is the existence and uniqueness of quotients and remainders.

Highest common factors

Given $a, b \in \mathbb{N}$, the *highest common factor* (*HCF*) or *greatest common divisor* (*GCD*) of a and b, written as (a, b), is defined to be $d \in \mathbb{N}$ satisfying:

- $d \mid a$ and $d \mid b$, and
- if $e \mid a$ and $e \mid b$ then $e \mid d$.

The second condition implies that $e \leq d$, but is a more general expression that allows the proofs that follow to be extended easily into sets other than the integers.

Observations

- The HCF exists and is unique.

  Proof: *Existence*. Consider $\mathrm{D} = \{as + bt \mid s, t \in \mathbb{Z} \text{ and } (as + bt) > 0\}$.
  $a = a1 + b0 \in \mathrm{D}$ so $\mathrm{D} \neq \emptyset$. By well ordering D has a least element, d, and $d = as + bt$ for some s and t. Use the division algorithm to write $a = dq + r$ with $0 \leq r < d$.
  Now $r = a - dq = a - (as + bt)q = a(1 - sq) + b(-tq)$. If $r > 0$, then $r \in \mathrm{D}$. But $r < d$ and d is minimal in D, so $r \notin \mathrm{D}$ and $r \leq 0$. But $r \geq 0$ so $r = 0$ and $d \mid a$. $d \mid b$ similarly.

  Now suppose $e \mid a$ and $e \mid b$. Say $a = fe$ and $b = ge$.
  Then $d = as + bt = fes + get = e(fs + gt)$ and $e \mid d$.

  Uniqueness. Suppose d_1 and d_2 are both HCFs satisfying the two conditions.
  Then $d_1 \mid d_2$ and $d_2 \mid d_1$, so $d_1 = d_2$.

- There are integers x and y with $d = ax + by$. Moreover, x and y can be calculated efficiently.

  Proof: $x = s$ and $y = t$ in the above for existence. See below for an efficient algorithm.

- If $a, b \in \mathbb{N}$ and $a = bq + r$ for integers q and r with $0 \leq r < b$, then $(a, b) = (b, r)$. This will give rise to an efficient algorithm for HCFs.

  Proof: Suppose $d = (a, b)$ and $a = bq + r$ by the division algorithm.
  $d \mid a$ and $d \mid b$ so $d \mid (a - bq) = r$. Therefore $d \mid (b, r)$.
  But $(b, r) \mid b$ and $(b, r) \mid r$, so $(b, r) \mid a$. Therefore $(b, r) \mid (a, b) = d$, so $(b, r) = d$.

- If $a \mid bn$ and $(a, b) = 1$, then $a \mid n$.

  Proof: $a \mid bn$, so write $bn = aq$. $(a, b) = 1$, so find x and y with $ax + by = 1$.
  Now $n = nax + nby = nax + aqy = a(nx + qy)$, and so $a \mid n$.

- If $a \mid n$, $b \mid n$ and $(a, b) = 1$, then $ab \mid n$.

  Proof: $(a, b) = 1$, so write $n = nax + nby$ as before. $a \mid n$, so $ab \mid nb$ and $ab \mid nby$.
  $b \mid n$ so $ab \mid nax$ similarly. Hence $ab \mid n$.

- $\left(\frac{a}{(a,b)}, \frac{b}{(a,b)}\right) = 1$.

  Proof: Use contradiction. Suppose $\left(\frac{a}{(a,b)}, \frac{b}{(a,b)}\right) = k > 1$. Then $k\,(a, b) \mid a$ and $k\,(a, b) \mid b$, which contradicts (a, b) being the highest common factor.

We say that a and b are *co-prime* if $(a, b) = 1$.

The *least common multiple* of a and b is the smallest number m which is exactly divisible by both a and b. This is sometimes written as $[a, b]$ and is equal to $ab \div (a, b)$.

Euclid's algorithm

Given $a, b \in \mathbb{N}$, use the division algorithm to write:

$$\begin{aligned}
a &= q_1 b + r_1 && 0 \le r_1 < b \\
b &= q_2 r_1 + r_2 && 0 \le r_2 < r_1 \\
r_1 &= q_3 r_2 + r_3 && 0 \le r_3 < r_2 \\
&\ldots \\
r_{i-2} &= q_i r_{i-1} + r_i && 0 \le r_i < r_{i-1} \\
&\ldots \\
r_{n-2} &= q_n r_{n-1} && \text{with remainder } r_n = 0
\end{aligned}$$

Then $(a, b) = (b, r_1) = (r_1, r_2) = \ldots = (r_{n-2}, r_{n-1}) = r_{n-1}$.

Moreover, we can now work backwards through the algorithm to calculate the integers x and y with $(a, b) = ax + by$.

Alternatively, we can produce the same result working forwards by observing that line i is just the difference of line $i-2$ and q_i times line $i-1$. Write $r_{-1} = a$ and $r_0 = b$, so q_i is just the integer quotient of r_{i-2} divided by r_{i-1}. Now express $r_i = s_i a + t_i b$ so $s_{-1} = 1$, $t_{-1} = 0$, $s_0 = 0$ and $t_0 = 1$ and observe that $r_i = r_{i-2} - q_i r_{i-1}$, $s_i = s_{i-2} - q_i s_{i-1}$ and $t_i = t_{i-2} - q_i t_{i-1}$.

Here is a worked example:

i	q_i	r_i		s_i	t_i
		$a = 55$	$= 2.20 + 15$	1	0
		$b = 20$	$= 1.15 + 5$	0	1
1	2	15	$= 3.5 + 0$	1	–2
2	1	5		–1	3
3	3	0		4	–11

The last line tells us that $4.55 - 11.20 = 0$ so $4k.55 - 11k.20 = 0$. This is rather like the finding the complementary function that solves the homogeneous part of a differential equation.

The penultimate line tells us that $(55, 20) = 5 = -1.55 + 3.20$. This is rather like finding the particular solution for an inhomogeneous differential equation.

Observations

- The signs of s_i alternate $1\ 0\ 1 - + - \ldots$ and those of t_i alternate $0\ 1 - + - + \ldots$.

  Proof: a, b and all the remainders r_i are positive, so the quotients q_i will be as well.

- $s_{i-1} t_i - s_i t_{i-1} = (-1)^i$ for $i \ge 0$, so, in particular, s_i and t_i are co-prime.

  Proof: By induction.

  Corollary: We have a linear combination of s_i and t_i which is equal to 1, so $(s_i, t_i) = 1$.

- $|s_n| = \frac{b}{(a,b)}, |t_n| = \frac{a}{(a,b)}$.

  Proof: Note that $r_n = s_n a + t_n b = 0$ and divide through by (a, b) to show $|s_n|\frac{a}{(a,b)} = |t_n|\frac{b}{(a,b)}$. Remember $\left(\frac{b}{(a,b)}, \frac{a}{(a,b)}\right) = 1$, so $\frac{b}{(a,b)} \mid |s_n|$.

  But $(s_n, t_n) = 1$ by the above, so $|s_n| \mid \frac{b}{(a,b)}$. Hence $|s_n| = \frac{b}{(a,b)}$.

Applications

- Given $a, b, c \in \mathbb{Z}$ with a and b not both zero, the linear Diophantine equation $ax + by = c$ has a solution with $x, y \in \mathbb{Z}$ if, and only if, $(a, b) \mid c$.

  Proof: ($\Rightarrow$) $(a, b) \mid a$ and $(a, b) \mid b$, so $(a, b) \mid (ax + by) = c$.

  ($\Leftarrow$) Suppose $(a, b) \mid c$. Write $f = \frac{c}{(a,b)}$. Find s and t with $(a, b) = as + bt$ using Euclid. Now $afs + bft = (as + bt)f = (a,b)\frac{c}{(a,b)} = c$, as required.

- Moreover, any solution to $au + bv = c$ has $u = x - \frac{kb}{(a,b)}$ and $v = y + \frac{ka}{(a,b)}$ for some $k \in \mathbb{Z}$.

  Proof: Suppose $ax + by = c$ and $au + bv = c$. Then $a(x–u) + b(y–v) = 0$, so $a(x–u) = b(v–y)$. Divide by (a, b), so $\frac{a}{(a,b)}(x-u) = \frac{b}{(a,b)}(v-y)$.

  But $\left(\frac{a}{(a,b)}, \frac{b}{(a,b)}\right) = 1$, so $\frac{a}{(a,b)} \mid (v-y)$. Hence $v = y + \frac{ka}{(a,b)}$.

  Now $u = x - \frac{b(v-y)}{a} = x - \frac{kb}{(a,b)}$.

- The general solution is just the sum of the particular solution $u = x$ and $v = y$ with the complementary function $u = \frac{kb}{(a,b)}$ and $v = -\frac{ka}{(a,b)}$ where k is an arbitrary constant.

- $a \div b$ can be written as the continued fraction $q_1 + \cfrac{1}{q_2 + \cfrac{1}{q_3 + \cdots}} = q_1 + \frac{1}{q_2 +}\frac{1}{q_3 +}\cdots$.

  Proof: Write $a = q_1 b + r_1$ so $\frac{a}{b} = q_1 + \frac{r_1}{b} = q_1 + \frac{1}{b/r_1}$.

  But $b = q_2 r_1 + r_2$ so $\frac{b}{r_1} = q_2 + \frac{1}{r_1/r_2}$, and so on until $\frac{r_{n-2}}{r_{n-1}} = q_n + 0$.

Efficiency

Euclid's algorithm finds (a, b) in $O(\log a)$ steps.

Proof: $a = q_1 b + r_1 \geq b + r_1 > 2r_1 > 2^2 r_3 > 2^3 r_5 > \ldots > 2^k r_{2k-1}$. So $r_{2k-1} < a / 2^k$. In particular, $k > \log_2 a$ implies that $r_{2k-1} < 1$, so $r_{2k-1} = 0$ and the algorithm has finished. Hence Euclid's algorithm takes at most $2 \log_2 a$ steps.

In fact we can do better than this. If $a > b$ and b has d digits (to the base 10), then Euclid's algorithm will take at most $5d + 2$ steps to find (a, b).

It is actually rather hard to say how many steps will be required for any given pair of numbers. So we follow Pólya's advice and ask a different question. What is the smallest number that will require n steps? This will arise when $q_i = 1$ for $1 \leq i < n$ and $q_n = 2$.

Using the earlier notation, $|s_i| = |s_{i-1}| + |s_{i-2}|$ and $|t_i| = |t_{i-1}| + |t_{i-2}|$ so $|s_i| = f_i$ and $|t_i| = f_{i+1}$ where f_i is the i^{th} Fibonacci number. So, if $b < f_n$, $|s_n| < f_n$ and we need fewer than n steps.

However, if $n = 5d + 2$, then $f_n > 1.6^{n-2} = 1.6^{5d} > 10^d > b$, as required. Of course, this is still $O(\log a)$.

Primes

A natural number p is *prime* if $p > 1$ and p has no proper divisor.

Observations

- If p is a prime and $p \mid ab$ for $a, b \in \mathbb{N}$ but $p \nmid a$, then $p \mid b$.

  Proof: If $p \nmid a$, then $(p, a) = 1$ and so $p \mid b$.

- There are infinitely many primes.

  Proof: Use contradiction. Suppose that the only primes were $p_1, p_2, p_3, \ldots p_n$. Consider $N = p_1 p_2 p_3 \ldots p_n + 1$. The smallest number that divides exactly into N must be a prime, but each of $p_1, p_2, p_3, \ldots p_n$ leaves remainder 1 when divided into N. Hence N itself must be a prime, but it isn't in the list.

- If p is a prime then $\sqrt{p}$ is irrational; that is, it can not be expressed as a ratio of two natural numbers.

  Proof: Use contradiction. Suppose $\sqrt{p} = \frac{a}{b}$ for $a, b \in \mathbb{N}$ with $(a, b) = 1$. Then $p \mid pb^2 = a^2$, so $p \mid a$. Write $a = pc$ so $pb^2 = p^2 c^2$ and $p \mid b$. Hence $p \mid (a, b)$.

Digressions

- $2^{13466917} - 1$ is prime.
- The Mersenne number $2^n - 1$ is prime only when n is prime, but that is not sufficient.
- The Fermat number $2^n + 1$ is prime only when n is of the form 2^m, but that is not sufficient.
- If p is a Fermat prime, then it is possible to construct a regular p–gon using only pencil, ruler and compasses.
- Let $\Pi(x)$ be the number of primes $\leq x$. Then $\Pi(x) \approx x / \ln x$.
- *Prime pair conjecture:* There are infinitely many primes p with $p + 2$ also prime.

- *Goldbach conjecture:* Every even integer greater than 2 can be expressed as the sum of two primes.

Fundamental theorem of arithmetic

Every natural number greater than 1 can be expressed as a product of primes. Moreover, the expression is unique up to the order of the primes.

Proof

Existence. Use contradiction. Let $n \in \mathbb{N}$ be the smallest counter-example. If n is prime, then we are done. Otherwise $n = ab$ for some $a, b \in \mathbb{N}$ with $a, b < n$. Write a and b as products of primes and combine them to give an expression for n.

Uniqueness. Suppose $n = p_1 p_2 p_3 \ldots p_r = q_1 q_2 q_3 \ldots q_s$ with the p_i and q_j all prime. $p_1 \mid n$, so $p_1 \mid q_1 q_2 q_3 \ldots q_s$. Now, either $p_1 \mid q_1$ or $p_1 \mid q_2 q_3 \ldots q_s$. In the latter case, continue until $p_1 \mid q_j$ for some j. But q_j is prime, so $p_1 = q_j$. Renumber so $j = 1$. Now $p_1 p_2 p_3 \ldots p_r = p_1 q_2 q_3 \ldots q_s$ so $p_2 p_3 \ldots p_r = q_2 q_3 \ldots q_s$. Continue in this way until $p_r = q_s$ and $r = s$.

Observation

- If $m = p_1^{r_1} p_2^{r_2} \ldots p_k^{r_k}$ and $n = p_1^{s_1} p_2^{s_2} \ldots p_k^{s_k}$ then
  $(m,n) = p_1^{\min(r_1,s_1)} p_2^{\min(r_2,s_2)} \ldots p_k^{\min(r_k,s_k)}$ and
  $[m,n] = p_1^{\max(r_1,s_1)} p_2^{\max(r_2,s_2)} \ldots p_k^{\max(r_k,s_k)}$.

Exercises

1. Are the following statements true or false?
   - $(a, b)\,(c, d) = (ac, bd)$
   - $(a, b)\,(a, d) = (a^2, bd)$
   - $(a, b) = (a, d) = 1$ implies that $(a, bd) = 1$
2. Prove that, if x and y are integers such that $57x + 44y = 1$, then there is an integer k such that $x = 17 - 44k$ and $y = 57k - 22$.
3. Does the equation $1992x + 1752y = 12$ have a solution in integers? Find all the integer solutions to the equation $1992x + 2622y = 12$.
4. Find all sets of integers x, y and z such that $56x + 63y + 72z = 1$.

   Hint: Consider the values taken by $56x + 63y$ as x and y range through $\mathbb{Z}$.
5. A photocopier charges 7.2p for each copy. However, it only accepts 10p coins and gives no change, although unused credit is carried forward. What is the smallest number of copies that must be made if the user is not to forgo any change?
6. Define the *least common multiple* of a and b to be $m = [a, b] = ab \div (a, b)$. Show that:
   - $a \mid m$ and $b \mid m$, and
   - if $a \mid n$ and $b \mid n$ then $m \mid n$.
7. Show that there are infinitely many prime numbers of the form $4k + 3$. [*Hint:* Consider $N = 2^2.3.5.7.\ldots.p_n - 1$.]
8. Let $\mathbb{Z}[\sqrt{5}] = \{a + b\sqrt{5} \mid a, b \in \mathbb{Z}\}$.
   Show that $\alpha, \beta \in \mathbb{Z}[\sqrt{5}] \Rightarrow \alpha + \beta, \alpha - \beta, \alpha \times \beta \in \mathbb{Z}[\sqrt{5}]$.

   Given $\alpha = a + b\sqrt{5} \in \mathbb{Z}[\sqrt{5}]$, define the *conjugate* of α to be $\overline{\alpha} = a - b\sqrt{5}$, and the *norm* of α to be $\mathrm{N}(\alpha) = \left|\alpha \times \overline{\alpha}\right| = |a^2 - 5b^2|$. Show that $\mathrm{N}(\alpha \times \beta) = \mathrm{N}(\alpha) \times \mathrm{N}(\beta)$.

   Define a *unit* $\varepsilon \in \mathbb{Z}[\sqrt{5}]$ to be an element that divides exactly into 1. Show that ε is a unit if and only if $\mathrm{N}(\varepsilon) = 1$.

   By considering residues modulo 5 (see the next section), show that there is no $\alpha \in \mathbb{Z}[\sqrt{5}]$ with $\mathrm{N}(\alpha) = 2$.

   Factor 4 in $\mathbb{Z}[\sqrt{5}]$ in two different ways, say $4 = \alpha_1\beta_1 = \alpha_2\beta_2$, where α_1 has no factors in common with α_2 other than units.

   Deduce that there is no analogue in $\mathbb{Z}[\sqrt{5}]$ to the uniqueness of prime factorisation.

   [It makes sense to define $\pi \in \mathbb{Z}[\sqrt{5}]$ to be prime if $\alpha \mid \pi \Rightarrow \alpha \mid 1$ or $\pi = \varepsilon\,\alpha$ for some unit ε.]

9. A Pythagorean Triad is a triple (a, b, c) with $a, b, c \in \mathbb{N}$ such that $a^2 + b^2 = c^2$. For example, $(3, 4, 5)$ and $(5, 12, 13)$ are Pythagorean Triads. Check the following:
   - $(m(p^2–q^2), 2mpq, m(p^2+q^2))$ is a Pythagorean Triad for any $m, p, q \in \mathbb{N}$ with $p > q$.
   - If (a, b, c) is a Pythagorean Triad, then we can write $a = md$, $b = me$ and $c = mf$ where d, e and f are pairwise co-prime (that is, $(d, e) = (e, f) = (f, d) = 1$), and exactly one of d and e is even, say $e = 2g$. Moreover, $f + d = 2h$ and $f – d = 2i$ for $h, i \in \mathbb{N}$. Since $g^2 = hi$ and $(h, i) = 1$, it follows that $h = p^2$ and $i = q^2$ for $p, q \in \mathbb{N}$.
   - Hence every Pythagorean Triad is of the form $(m(p^2–q^2), 2mpq, m(p^2+q^2))$. Moreover, different values of m, p and q give rise to different values.
10. Recall the Fibonacci numbers $\{f_n\}$.
    - Show, by induction on k or otherwise, that $f_{n+k} = f_k f_{n+1} + f_{k–1} f_n$.
    - Deduce that $f_n \mid f_{ln}$ for all $l \geq 1$.
    - Show that $(f_n, f_{n–1}) = 1$.
    - Deduce also that $(f_m, f_n) = (f_{m-n}, f_n)$ and hence that $(f_m, f_n) = f_{(m, n)}$.
    - Show that $f_m f_n \mid f_{mn}$ if $(m, n) = 1$.

Programming

11. Write an ML function to factor an integer into a list of prime factors.
12. Write an ML function to implement Euclid's algorithm. Given two integers a and b, this should return a triple (x, y, z) such that $ax + by = z$ where z is the greatest common divisor of a and b.

Modular arithmetic

If $a, b \in \mathbb{Z}$ and $m \in \mathbb{N}$ then we say that a and b are *congruent modulo* m if $m \mid (a - b)$, and we write this as $a \equiv b \pmod{m}$.

This equivalent to saying that there is $q \in \mathbb{Z}$ such that $a = b + qm$.

Observations

- For all $a \in \mathbb{Z}$ and $m \in \mathbb{N}$ we have $a \equiv a \pmod{m}$.

  Proof: $m \mid 0 = (a - a)$.

- If $a \equiv b \pmod{m}$ then $b \equiv a \pmod{m}$.

  Proof: If $a \equiv b \pmod{m}$ then $m \mid (a - b)$, so $m \mid (b - a)$, and $b \equiv a \pmod{m}$.

- If $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m}$, then $a \equiv c \pmod{m}$.

  Proof: If $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m}$, then find $r, s \in \mathbb{Z}$ such that $a - b = rm$ and $b - c = sm$. Then $a - c = (r + s)\, m$ and $a \equiv c \pmod{m}$.

- If $a_1 \equiv b_1 \pmod{m}$ and $a_2 \equiv b_2 \pmod{m}$ then
  $a_1 + a_2 \equiv b_1 + b_2 \pmod{m}$, $a_1 - a_2 \equiv b_1 - b_2 \pmod{m}$ and $a_1 \times a_2 \equiv b_1 \times b_2 \pmod{m}$.

  Proof: Find $q_1, q_2 \in \mathbb{Z}$ such that $a_1 - b_1 = q_1 m$ and $a_2 - b_2 = q_2 m$. Then
  $(a_1 + a_2) - (b_1 + b_2) = (q_1 + q_2)\, m$, so $a_1 + a_2 \equiv b_1 + b_2 \pmod{m}$,
  $(a_1 - a_2) - (b_1 - b_2) = (q_1 - q_2)\, m$, so $a_1 - a_2 \equiv b_1 - b_2 \pmod{m}$, and
  $a_1 a_2 - b_1 b_2 = (b_1 + q_1 m)(b_2 + q_2 m) - b_1 b_2 = (b_1 q_2 + q_1 b_2 + q_1 q_2 m)\, m$, so
  $a_1\, a_2 \equiv b_1\, b_2 \pmod{m}$.

- However, $a_1 \equiv b_1 \pmod{m}$ and $a_2 \equiv b_2 \pmod{m}$ does *not* imply that $a_1^{b_1} \equiv a_2^{b_2} \pmod{m}$. For example, consider $a_1 = a_2 = 2$, $b_1 = 1$, $b_2 = 4$, and $m = 3$.

Examples

Here are the addition and multiplication tables modulo 4:

$+$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\times$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

and multiplication modulo 5:

×	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Applications

- No integer congruent to 3 modulo 4 can be expressed as the sum of two squares.

  Proof: All squares modulo 4 are congruent to either 0 or 1, so the sum of two squares will be congruent to 0, 1 or 2.

- No integer congruent to 7 modulo 8 can be expressed as the sum of three squares.

  Proof: All squares modulo 8 are congruent to 0, 1 or 4, so the sum of three squares will be congruent to 0, 1, 2, 3, 4, 5 or 6.

  It transpires that any integer can be expressed as the sum of four squares, but this is harder to prove.

- $5 \mid (2^{3n+1} + 3^{n+1})$

  Proof: Observe that $2^{3n+1} \equiv 2, 1, 3, 4, 2, 1, 3, 4, \ldots \pmod 5$ and $3^{n+1} \equiv 3, 4, 2, 1, 3, 4, 2, 1 \pmod 5$ for $n = 0, 1, 2, 3, \ldots$, so their sum will be congruent to $0 \pmod 5$.

- There is no integer solution to $x^3 - x^2 + x + 1 = 0$.

  Proof: Consider the equation modulo 2. $x \equiv 0$ could not be a solution, but $x \equiv 1$ might be. This tells us that any solution would have to be odd. However, considering the equation modulo 3 shows that none of $x \equiv 0$, $x \equiv 1$ or $x \equiv 2$ could be a solution and so there is no solution.

- $641 \mid (2^{2^5} + 1)$.

  Proof: Consider $p = 641$, so $p = 625 + 16 = 5^4 + 2^4$ and $2^4 \equiv -5^4 \pmod p$. Observe also that $p-1 = 640 = 5 \times 128 = 5 \times 2^7$ so $5 \times 2^7 \equiv -1 \pmod p$. Combine these to see that $2^{32} = 2^4 \times 2^{28} \equiv (-5^4) \times (2^7)^4 = -(5 \times 2^7)^4 \equiv -(-1)^4 = -1 \pmod p$. So $p \mid (2^{32}+1)$.

Congruences

The *residues modulo m* are $\mathbb{Z}_m = \{0, 1, 2, \ldots (m-1)\}$.

Addition, subtraction and multiplication all work for residues, but what about division?

Given $a, c \in \mathbb{Z}$ and $m \in \mathbb{N}$, the *congruence* $ax \equiv c \pmod m$ has a solution for x if, and only if, $(a, m) \mid c$.

Proof

$ax \equiv c \pmod m$ has a solution
$\Leftrightarrow$ we can find x with $m \mid (ax - c)$
$\Leftrightarrow$ we can find x and y with $ax - c = my$

$\Leftrightarrow$ we can find x and y with $ax - my = c$
$\Leftrightarrow (a, m) \mid c$ by the application of Euclid's algorithm to linear Diophantine equations.

Moreover, by considering the complementary function to the equation, the solution is unique modulo $m \div (a, m)$.

Units

In particular, we can calculate the reciprocal of a modulo m if, and only if, $(a, m) = 1$. Such values a are called *units* modulo m and we write $\mathrm{U}_m = \{a \in \mathbb{Z}_m \mid a \text{ is a unit}\}$.

Observations

- If $a, b \in \mathrm{U}_m$ then $ab \in \mathrm{U}_m$.

  Proof: If $a, b \in \mathrm{U}_m$ then we can find x, y so that $ax \equiv by \equiv 1 \pmod{m}$. So the product $(ab)(xy) = (ax)(by) \equiv 1 \pmod{m}$ and ab is a unit.

- The reciprocal of a modulo m is unique modulo m.

  Proof: Suppose $ax \equiv 1 \pmod{m}$ and $ay \equiv 1 \pmod{m}$. Then $m \mid (ax - 1)$ and $m \mid (ay - 1)$, so $m \mid ((ax - 1) - (ay - 1)) = a(x–y)$. But $(m, a) = 1$ so $m \mid (x - y)$ and $x \equiv y \pmod{m}$.

- We can calculate reciprocals of units by using the extended Euclid's algorithm to express $(a, m) = 1$ as a linear combination of a and m.

Euler's totient function

Define $\varphi(m)$ to be the number of natural numbers less than m and co-prime to m, so $\varphi(m)$ is the number of units modulo m.

Given a prime p, observe $\varphi(p) = (p - 1)$ and $\varphi(p^n) = p^n - p^{n-1}$.

Chinese Remainder Theorem

Given two natural numbers m and n with greatest common divisor 1, there is a simultaneous solution to the congruences $x \equiv a \pmod{m}$ and $x \equiv b \pmod{n}$ and this solution is unique $\pmod{mn}$.

Proof

Existence. Use Euclid's algorithm to find s and t such that $ms + nt = 1$. Let $c = bms + ant$. Now $nt \equiv 1 \pmod{m}$ so $c \equiv ant \equiv a \pmod{m}$. Similarly $c \equiv b \pmod{n}$.

Uniqueness. Suppose there is a further solution d. Observe that $c - d \equiv 0 \pmod{m}$ and $c - d \equiv 0 \pmod{n}$, so $c - d \equiv 0 \pmod{mn}$ as required.

Corollaries

- Euler's totient function is multiplicative: if $(m, n) = 1$ then $\varphi(mn) = \varphi(m)\varphi(n)$.

  Proof: Given $c \in \mathrm{U}_m$ and $d \in \mathrm{U}_n$ find $e \in \mathbb{Z}_{mn}$ with $c \equiv e \pmod{m}$ and $d \equiv e \pmod{n}$. Then $e \in \mathrm{U}_{mn}$ and each such pair (c, d) is linked to a unique e.

- $\varphi(m) = m \prod_{\text{prime } p \mid m} (1 - 1/p)$.

  Proof: Consider the unique expression of m as a product of primes.

Wilson's theorem

If p is a prime, then $(p-1)! \equiv -1 \pmod{p}$.

Proof

Associate each of the numbers $1, 2, \ldots, p-1$ with its reciprocal $\pmod{p}$. The reciprocal of a may be the same as a, but only if $a^2 \equiv 1 \pmod{p}$ which requires $a = 1$ or $p-1$. Apart from these, the numbers $2, 3, \ldots, p-2$ can be paired off so that the product of each pair is $1 \pmod{p}$. It follows that $2.3. \ldots .(p-2) \equiv 1 \pmod{p}$. Multiply by $p-1 \equiv -1 \pmod{p}$ to obtain the result.

This proof actually fails if $p = 2$ or 3, but these cases are easily verified independently.

Euler's theorem[1]

Given $m \geq 2$ and a with $(a, m) = 1$, then $a^{\varphi(m)} \equiv 1 \pmod{m}$.

Proof

Let $U_m = \{x \mid 0 < x < m$ and $(x, m) = 1\}$ be the set of units modulo m. Say $U_m = \{u_1, u_2, \ldots u_f\}$ where $f = \varphi(m)$.

Multiply each of these u_i by a modulo m. The resulting values are coprime to m, since u_i and a are. Moreover they are distinct, since a is a unit and can be divided, so $au_i \equiv au_j \pmod{m} \Rightarrow u_i \equiv u_j \pmod{m}$. So they are just a permutation of the f values in U_m.

Hence the product $(au_1)(au_2)\ldots(au_f) \equiv u_1u_2\ldots u_f \pmod{m}$. But $u_1, u_2, \ldots u_f$ are all units and so can be divided out, leaving $a^f \equiv 1 \pmod{m}$ as required.

Corollary (Fermat's little theorem)[2]

Given a prime p and a not divisible by p, then $a^{p-1} \equiv 1 \pmod{p}$.

Moreover, for any a, $a^p \equiv a \pmod{p}$.

Observation

This gives a test for primality. If a number p does *not* satisfy $a^{p-1} \equiv 1 \pmod{p}$ for any single value of a, then p can *not* be prime.

However, passing this test is not sufficient to prove primality. Composite numbers p that satisfy $a^{p-1} \equiv 1 \pmod{p}$.are called *pseudo-prime* with respect to the base a. *Carmichael numbers* are Fermat pseudo-primes for all bases a with $(a, p) = 1$. For example, $561 = 3 \times 11 \times 17$. Observe that $(3-1)|(561-1)$, $(11-1)|(561-1)$ and $(17-1)|(561-1)$, so $a^{561-1} \equiv 1 \pmod{3, 11 \text{ and } 17}$ for all a with $(a, p) = 1$, and so $a^{561-1} \equiv 1 \pmod{561}$ by the Chinese Remainder Theorem.

The Fermat-Euler test $a^{\frac{p-1}{2}} \equiv \pm 1 \pmod{p}$ is sharper but is still not sufficient. In particular, it reveals 561 to be composite since $2^{\frac{561-1}{2}} = 2^{280} \equiv 421 \pmod{561}$. However, it fails to catch 1729.

[1] Humphreys & Prest, p 58.
[2] Humphreys & Prest, p 54.

Public key cryptography

With the increasing use of computer networks and digital, electronic communications, it becomes important to ensure that messages can be sent securely with the meaning revealed only to the intended recipient and that they can be authenticated as having been sent by the real originator.

The general approach is to choose some large modulus m and encode blocks of a message as numbers in $\mathbb{Z}_m$.

Caesar's cypher encodes a message a as $a + e \pmod{m}$ for some encryption key, e. This is decoded by calculating $(a + e) - e \equiv a \pmod{m}$. Unfortunately, the code is also easily broken by frequency analysis.

Using larger blocks and changing e in some agreed sequence unknown to interceptors gives a *one-time pad*, which is secure but difficult to administer.

A further problem is the distribution of the keys. The key can be any secret shared by the two participants. How can one pass it safely to the other? The trick is to imagine a box with two locks and proceed as follows:

- The sender (conventionally called Alice) places the secret in the box, locks one of the locks with her key and sends the locked box to the recipient (conventionally called Bob).
- Bob locks the second lock with his key and returns the box to Alice.
- Alice unlocks the first lock and returns the box to Bob.
- Bob unlocks the second lock, opens the box and extracts the secret.

Note that Alice and Bob never have to share their private keys with anyone else but the box is always securely locked when in transit between them. The trick is to find an arithmetic equivalent of a box with two locks.

Modular addition is a possibility. Alice and Bob agree on a modular base m (which can be made public) and choose private values a and b. Alice now sends a shared secret s to Bob as follows:

- A $\rightarrow$ B: $m_1 \equiv s + a \pmod{m}$
- B $\rightarrow$ A: $m_2 \equiv m_1 + b = s + a + b \pmod{m}$
- A $\rightarrow$ B: $m_3 \equiv m_2 - \mathrm{a} = \mathrm{s} + b \pmod{m}$

Bob can now recover s. Unfortunately, anyone overhearing the conversation (traditionally called Eve) can recover $s \equiv m_1 - m_2 + m_3 \pmod{m}$.

Modular multiplication is another possibility. As long as a and b are co-prime to m, Alice and Bob can calculate multiplicative inverses and replace the subtractions by divisions in the above protocol. The same problem arises and Eve can recover s or, strictly speaking, $s \pmod{m/(m,s)}$ if $(m,s) > 1$.

However, modular exponentiation really does work.

Diffie-Hellman key exchange[3]

Choose a large prime modulus, p. Pick e with $(e, p-1) = 1$ and find d such that $de \equiv 1 \pmod{p-1}$ so $de = 1 + (p-1)t$ for some t.

Then $(a^e)^d = a^{ed} = a^{1+(p-1)t} = a(a^{p-1})^t \equiv a1^t \pmod{p} = a$.

We now have a protocol:

[3] Davenport, p 191.

- Alice chooses p and the value e and sends p and the message a^e to Bob.
- Bob picks another value f with inverse g and sends $(a^e)^f$ back to Alice.
- Alice works out $((a^e)^f)^d = ((a^e)^d)^f = a^f$ and sends it back to Bob.
- Bob now works out $(a^f)^g$ back to decode and find a.

Breaking this requires *discrete logarithms*, which is as hard as factoring a large integer.

The RSA code[4]

The Rivest, Shamir and Adleman (RSA) public key system[5] uses Euler's Theorem to provide secure communications and digital signatures.

Let p and q be two primes with product m so $\varphi(m) = (p-1)(q-1)$. Choose e (the *encryption exponent*) relatively prime to $\varphi(m)$ and use Euclid's algorithm to find d (the *decryption exponent*) and c such that $ed + \varphi(m)c = 1$ so $ed \equiv 1 \pmod{\varphi(m)}$.

Now, given $a < p, q$, $(a^e)^d = a^{ed} = a^{1-\varphi(m)c} = a(a^{\varphi(m)})^{-c} \equiv a1^{-c} = a \pmod{m}$ provided $(a, m) = 1$, which is ensured by $a < p, q$.

We now have a protocol:

- Alice picks two large primes and publishes their product m and the value e.
- Bob encodes a message a as $a^e \pmod{m}$ and sends it to Alice.
- Alice recovers a by raising the encoded message to the power $d \pmod{m}$.

Anyone intercepting the message knows m and e but not d which can only be calculated easily if $\varphi(m)$ is known. However, this is believed to be difficult, at least as difficult as factoring m.

Conversely, if d is known, then m can be factored as follows:

$de \equiv 1 \pmod{\varphi(m)}$, so suppose that $de - 1 = n\varphi(m)$. Observe $\varphi(m) = (p-1)(q-1) = pq - p - q + 1$, which is slightly smaller than $pq = m$. So n is slightly greater than $(de - 1)/m$. Calculating this fraction and rounding up will give n.

Once n is known, $\varphi(m) = (de - 1)/n$. Now $m + 1 - \varphi(m) = p + q$ and $m = pq$, so p and q are the roots of the quadratic equation $x^2 - (m + 1 - \varphi(m))x + m = 0$.

The encoding and decoding processes are symmetric and can be performed in either order. Thus Alice can prove her identity by taking a challenge a and returning $a^d \pmod{m}$ which anyone can then decode but only she could have encoded.

Coin-tossing by telephone[6]

Let p be a prime of the form $4k + 3$ and suppose $a \equiv x^2 \pmod{p}$. Now $x^{4k+2} = x^{p-1} \equiv 1 \pmod{p}$, so $(a^{k+1})^2 \equiv x^{4k+4} \equiv x^2 \equiv a \pmod{p}$ and $x = a^{k+1}$ is a solution to the original equation. So we can calculate square roots $\bmod p$.

Let p and q be two such primes with product n and suppose $a \equiv z^2 \pmod{n}$. Now a is also a square modulo both p and q, say $a \equiv x^2 \pmod{p}$ and $a \equiv y^2 \pmod{q}$. Use the Chinese Remainder Theorem to construct 4 solutions $z \equiv \pm s, \pm t \pmod{n}$.

[4] Humphreys & Prest, p 60.
[5] R Rivest, A Shamir & L Adleman: *A method for obtaining digital signatures and public-key cryptosystems,* Communications ACM 21(2), February 1978, pp 120-6.
[6] Giblin, p 145.

Observe that, if we know both s and t, it is possible to factor n. $s^2 \equiv t^2 \equiv a(\mathrm{mod}\, n)$, so $pq = n|(s^2 - t^2) = (s+t)(s-t)$. However, s and t are distinct so neither $(s+t)$ nor $(s-t)$ is divisible by n. Without loss of generality, $p|(s+t)$ and $q|(s-t)$, and we can use Euclid to find p and q as the HCFs of n and $(s+t)$ and $(s-t)$ respectively.

We now have a protocol:

- Alice picks two large primes and tells Bob their product n.
- Bob picks s co-prime to n and tells Alice $a \equiv s^2 (\mathrm{mod}\, n)$.
- Alice calculates the 4 roots, picks one at random and tells Bob.
- If this is $\pm s$, Bob concedes defeat. Otherwise it is $\pm t$ which allows Bob to factor n and, by so doing, win.

Practical remarks

These mathematical results are not sufficient by themselves to build secure encryption systems. Care must be taken over the actual choice of the prime numbers used and, even more importantly, over the systems procedures. The security course explores these issues further.

Exercises

1. Show that a number is divisible by 9 if, and only if, the sum of its digits is divisible by 9. (This is known as *casting out the 9s*.) For example, 23714 is not divisible by 9 as $2+3+7+1+4 = 17$ which is not divisible by 9.
2. Find a similar test for divisibility by 11.
3. Is it possible to form a sum of numbers using each of the digits 0 to 9 exactly once whose total is 100? (Tricks like exponentiation are not allowed.)
4. A $1\ 000\ 000$ digit number is exactly divisible by 99. A new number is formed by reversing the order of its digits. What is the probability that the new number is also exactly divisible by 99?
5. The International Standard Book Number (ISBN) found in the front of many books is a 10 digit code such as 0-521-35938-4 (where the hyphens can be ignored). In this case, the 0 indicates that the book was published in the UK and some other English speaking countries, 521 is the publisher (the Cambridge University Press), 35938 is the book number and 4 a check digit. The check digit is chosen so that if the ISBN is $d_1 d_2 \ldots d_{10}$ then $d_{10} = \sum_{i=1}^{9} i \cdot d_i \pmod{11}$. It may be that the last digit has to be 10, in which case X is written, as in 0-387-97993-X.

   Prove that $\sum_{i=1}^{10} i \cdot d_i \equiv 0 \pmod{11}$ and verify that the two given ISBNs satisfy the congruence. Prove that the check digit will show up common copying errors caused by interchanging two adjacent digits (so, for example, 67 becomes 76) or doubling the wrong one of a triple (so, for example, 667 becomes 677). Why do you think the modulus 11 was chosen instead of the more natural 10?
6. Show that the equation $x^5 - 3x^2 + 2x - 1 = 0$ has no solutions for $x \in \mathbb{Z}$.
7. Solve the following congruences:
   - $77x \equiv 11 \pmod{40}$
   - $12y \equiv 30 \pmod{54}$
   - $z \equiv 13 \pmod{21}$ and $3z \equiv 2 \pmod{17}$
8. A band of 15 pirates acquires a hoard of gold pieces. When they come to divide up the coins, they find that three are left over. Their discussion of what to do with these extra coins becomes animated and, by the time some semblance of order returns, there remain only seven pirates capable of making an effective claim on the hoard. However, when the hoard is divided between these seven, it is found that two pieces are left over. There ensues an unfortunate repetition of the earlier disagreement, but this does at least have the consequence that the four pirates who remain are able to divide the hoard evenly between themselves. What is the smallest number of gold pieces that could have been in the hoard?[7]
9. Calculate $20!\ 21^{20} \pmod{23}$.
10. Calculate $3^{1000000000} \pmod{257}$.
11. Show that $42 \mid (n^7 - n)$ for all positive integers n.
12. An unwise person publishes the RSA enciphering scheme $(m, e) = (3901, 1997)$ via which he wishes to receive messages. You intercept the transmission

    1099 1307 2477 3490 0506 0615 0952 2697 0016 3333 0601

[7] Humphreys & Prest, p 50.

Factor m and hence find the deciphering key d such that $de \equiv 1 \pmod{\varphi(m)}$. Assuming that each block of four digits encodes two letters under the map a-z, space, ?, !, 0-9 become 00-25, 26, 27, 28, 29-38, decipher the text. (You may need to write and use the programs below.)

13. The previous question uses code blocks that are larger than the two primes whose product forms the base. Verify that a particular code block which shares a factor with m still can be encoded and decoded correctly. Why does this work?

14. 11 is a prime of the form $4k + 3$ (with $k = 2$) so we can extract the square root of a by raising a to the power $k + 1 = 3$. For example, the square root of 5 is $5^3 = 125 \equiv 4 \pmod{11}$ and we can check that $4^2 = 16 \equiv 5 \pmod{11}$. However, the same approach fails to calculate the square root of 6. Explain.

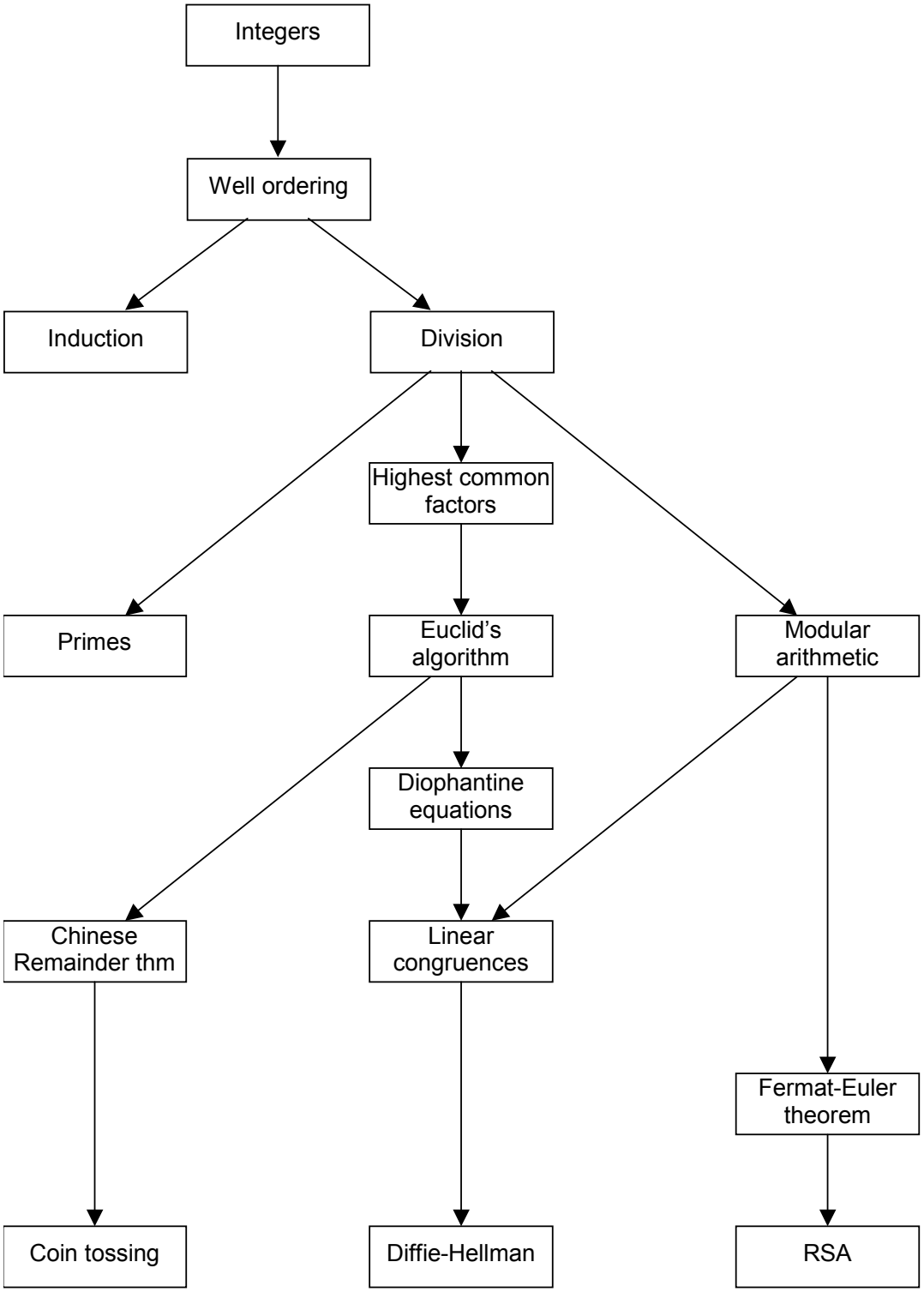
Programming

15. Write an ML function to calculate the reciprocal of a number to a given modular base. This may well use the function for Euclid's algorithm written earlier.

16. Write an ML function to calculate powers of numbers to a given modular base.

Revision guide

The following diagram shows the development of the key ideas presented in the first half of the course:



Sets

A set is just a collection of objects, or *elements*. We write $x \in \mathrm{A}$ when an element x is in the set A and $x \notin \mathrm{A}$ when it isn't.

Sets can be finite or infinite. (Indeed, there are many different infinite sizes.) If they are finite, you can define them explicitly by listing their elements, otherwise a pattern or restriction can be used:

$\mathrm{L} = \{\mathrm{a}, \mathrm{b}, \mathrm{c}\}$
$\mathrm{M} = \{\text{alpha}, \text{bravo}, \text{charlie}\}$
$\mathbb{N} = \{1, 2, 3, \ldots\}$
$\mathbb{Z} = \{\ldots -3, -2, -1, 0, 1, 2, 3, \ldots\}$
$\mathrm{P} = \{x \in \mathbb{N} \mid x > 1 \text{ and } 1 < y < x \Rightarrow (x, y) = 1\}$

Given a finite set, A, write $|\mathrm{A}|$ for the number of elements in A.

Write $\varnothing$ for the empty set, $\{\}$. So $|\varnothing| = 0$.

One set, A, is a *subset* of another set, B, if every element of A is also a member of B. We write this with a rounded less-than-or-equal sign: $\mathrm{A} \subseteq \mathrm{B}$. So $\mathbb{N} \subseteq \mathbb{Z}$. When the containment is strict (as in this case), we write $\mathbb{N} \subset \mathbb{Z}$ for a *proper* subset.

Two sets, A and B, are equal if they contain the same elements. This will often be proved by showing that each is a subset of the other: $\mathrm{A} \subseteq \mathrm{B}$ and $\mathrm{B} \subseteq \mathrm{A}$.

Sets can themselves be members of other sets. There is an important distinction between, for example, $\{\mathrm{a}, \mathrm{b}, \mathrm{c}\}$ and $\{\{\mathrm{a}, \mathrm{b}, \mathrm{c}\}\}$, or between $\varnothing$ and $\{\varnothing\}$.

Again, patterns can be used:

$$\mathrm{S} = \{\mathrm{X} \subseteq \mathrm{L} \mid \mathrm{a} \in \mathrm{X}\} = \{\{\mathrm{a}\}, \{\mathrm{a}, \mathrm{b}\}, \{\mathrm{a}, \mathrm{c}\}, \{\mathrm{a}, \mathrm{b}, \mathrm{c}\}\}$$

then $\mathrm{L} \in \mathrm{S}$ but $\mathrm{a} \notin \mathrm{S}$.

The power set, $\mathcal{P}(\mathrm{X})$, is the set of all subsets of X. So, for the set L above:

$$\mathcal{P}(\mathrm{L}) = \{\varnothing, \{\mathrm{a}\}, \{\mathrm{b}\}, \{\mathrm{c}\}, \{\mathrm{a}, \mathrm{b}\}, \{\mathrm{a}, \mathrm{c}\}, \{\mathrm{b}, \mathrm{c}\}, \{\mathrm{a}, \mathrm{b}, \mathrm{c}\}\}$$

It is important to specify the *universe of discourse* when discussing sets. This is the set of all possible elements that might be considered. It is often written as Ω.

Russell's paradox

Consider $\mathrm{R} = \{\mathrm{X} \mid \mathrm{X} \notin \mathrm{X}\}$. Then $\mathrm{L} \in \mathrm{R}$ and $\mathrm{S} \in \mathrm{R}$, but is $\mathrm{R} \in \mathrm{R}$?

Combining sets

There are several ways of combining existing sets to make new ones:

The *complement* of a set, A, is the collection, A^{c} or $\overline{\mathrm{A}}$, of elements (within the universe of discourse) that are not in A. $\mathrm{A}^{\mathrm{c}} = \{\mathrm{x} \in \Omega \mid \mathrm{x} \notin \mathrm{A}\}$. This is a case where it is necessary to be particularly clear about the universe.

Other operations include:

Union	$\mathrm{A} \cup \mathrm{B} = \{x \mid x \in \mathrm{A} \text{ OR } x \in \mathrm{B}\}$
Intersection	$\mathrm{A} \cap \mathrm{B} = \{x \mid x \in \mathrm{A} \text{ AND } x \in \mathrm{B}\}$

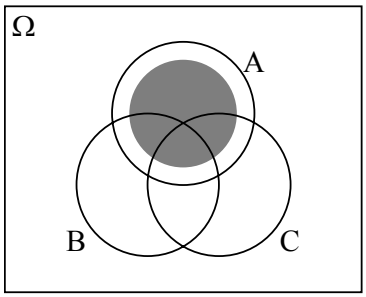
Difference	$A \setminus B$	$= \{x \mid x \in A \text{ AND } x \notin B\}$ $= A \cap B^c$
Symmetric difference	$A \Delta B$	$= (A \setminus B) \cup (B \setminus A)$

These operations satisfy various properties:

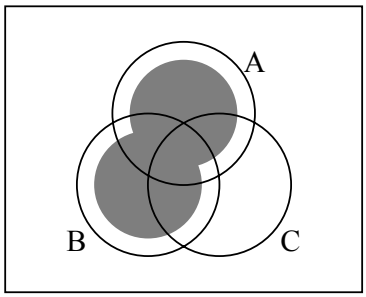
Idempotence	$A \cup A = A$	$A \cap A = A$
Complements	$A \cup A^c = \Omega$	$A \cap A^c = \varnothing$
	$(A^c)^c = A$	
Commutativity	$A \cup B = B \cup A$	$A \cap B = B \cap A$
Associativity	$A \cup (B \cup C) = (A \cup B) \cup C$	$A \cap (B \cap C) = (A \cap B) \cap C$
De Morgan's Laws	$(A \cup B)^c = A^c \cap B^c$	$(A \cap B)^c = A^c \cup B^c$
Distributivity	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
Empty set	$A \cup \varnothing = A$	$A \cap \varnothing = \varnothing$
Universal set	$A \cup \Omega = \Omega$	$A \cap \Omega = A$
Absorption	$A \cup (A \cap B) = A$	$A \cap (A \cup B) = A$

Venn diagrams

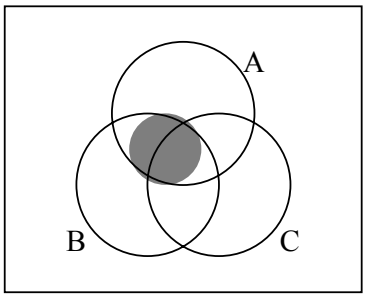
Venn diagrams provide a way of showing combinations of sets:



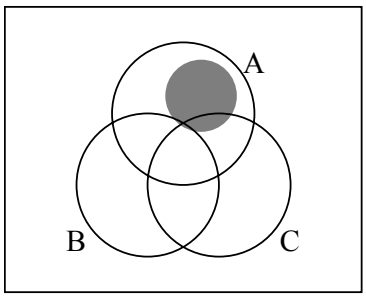
A



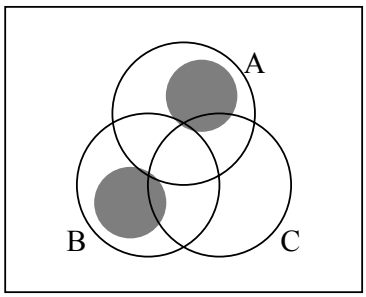
$A \cup B$



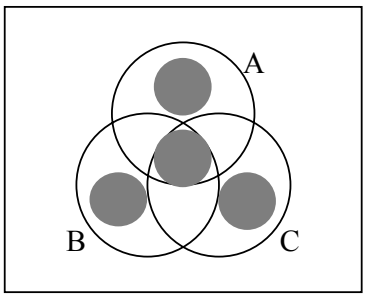
$A \cap B$



$A \setminus B$



$A \Delta B$



$A \Delta B \Delta C$

Normal form

Symmetric difference can be expressed as a union of differences and each difference can be expressed as an intersection of sets and complements of sets. De Morgan's Laws can be used to expand complemented expressions and distributivity can be used to expand intersections into

unions. Together these transformations allow any expression to be reduced to a union of terms each of which is the intersection of the underlying sets and their complements.

This expression is unique (up to the order of the terms and the factors within each term) so two expressions can be checked for equality by reducing them to this normal form.

Partitions

A *partition* of a set Ω is just a division of the whole of Ω into non-overlapping subsets.

Mathematically, a partition P of a set Ω is a subset $\mathrm{P} \subseteq \mathcal{P}(\Omega)$ with $\varnothing \notin \mathrm{P}$ such that

- $\bigcup_{S \in P} S = \Omega$ (P *covers* Ω) and
- If $\mathrm{S}, \mathrm{T} \in \mathrm{P}$ then $\mathrm{S} \cap \mathrm{T} \neq \varnothing$ implies that $\mathrm{S} = \mathrm{T}$ (the elements of P are *disjoint*).

Examples:

- $\{\{\mathrm{a}, \mathrm{b}, \mathrm{c}\}\}$ and $\{\{\mathrm{a}\}, \{\mathrm{b}, \mathrm{c}\}\}$ are both partitions of $\{\mathrm{a}, \mathrm{b}, \mathrm{c}\}$.
- Neither $\{\mathrm{a}, \mathrm{b}, \mathrm{c}\}$ nor $\{\{\mathrm{a}, \mathrm{b}\}, \{\mathrm{b}, \mathrm{c}\}\}$ are partitions of $\{\mathrm{a}, \mathrm{b}, \mathrm{c}\}$.

Product sets

The *product* of two sets A and B is the set of pairs of elements from A and B:

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

For example, $\{\mathrm{a}, \mathrm{b}\} \times \{\mathrm{b}, \mathrm{c}\} = \{(\mathrm{a}, \mathrm{b}), (\mathrm{a}, \mathrm{c}), (\mathrm{b}, \mathrm{b}), (\mathrm{b},\mathrm{c})\}$. Note that the order matters, so $\mathrm{A} \times \mathrm{B} \neq \mathrm{B} \times \mathrm{A}$

This can be extended to ordered n-tuples:

$$\mathrm{A}^1 = \mathrm{A}$$
$$\mathrm{A}^n = \mathrm{A} \times \mathrm{A}^{n-1} \text{ for } n > 1$$

For convenience, we write elements as $(\mathrm{a}, \mathrm{b}, \mathrm{c})$ rather than $(\mathrm{a}, (\mathrm{b}, \mathrm{c}))$. This gives the usual notation for Euclidean space, $\mathbb{R}^3$.

Disjoint sums

The *disjoint sum* of two sets A and B is $\mathrm{A} + \mathrm{B} = (\{0\} \times \mathrm{A}) \cup (\{1\} \times \mathrm{B})$.

So $\{\mathrm{a}, \mathrm{b}\} + \{\mathrm{b}, \mathrm{c}\} = \{(0, \mathrm{a}), (0, \mathrm{b}), (1, \mathrm{b}), (1, \mathrm{c})\}$ while $\{\mathrm{a}, \mathrm{b}\} \cup \{\mathrm{b}, \mathrm{c}\} = \{\mathrm{a}, \mathrm{b}, \mathrm{c}\}$.

Boolean logic

Propositions are statements that can be either true (T) or false (F). They will often include a symbol, x say, which can be thought of as an argument; the proposition $\mathrm{P}(x)$ will be true for some values of x and false for other values.

Propositions can be combined to make new ones:

P	Q	NOT P $\neg P, \sim P$	P AND Q $P \wedge Q$	P OR Q $P \vee Q$	P IMPLIES Q $P \Rightarrow Q$	P EQUIVALENT TO Q $P \Leftrightarrow Q$
F	F	T	F	F	T	T
F	T	T	F	T	T	F
T	F	F	F	T	F	F
T	T	F	T	T	T	T

Note that OR is *inclusive* - P OR Q is true if either or both of P and Q are true.

$P \Rightarrow Q$ means P *implies* Q. This is the same as saying, "If P is true then Q is true." If P is false, it says nothing about Q. It is actually equivalent to (NOT P) OR Q. It is also equivalent to $\neg Q \Rightarrow \neg P$.

Boolean logic enjoys a collection of properties that are similar to the ones shown above for sets. These can be used to prove statements by reducing them to a standard "sum of products" form.

Quantifiers

Given a proposition, $P(x)$, involving a variable, x, in some set, S, $P(x)$ may be true for some values of x and false for others. If it is true for *every* x, we write "$\forall x \in S \,.\, P(x)$" to mean "$P(x)$ is true for all x in S." If there is *at least one* x for which $P(x)$ is true, we write "$\exists x \in S \,.\, P(x)$" to mean "There exists an x in S such that $P(x)$ is true." $\forall$ is called the *universal quantifier* and $\exists$ the *existential quantifier*. The full stop in the middle can be read as a colon.

For example, we could write the definition of a highest common factor as:

$\forall a, b \in \mathbb{N} \,.\, \exists d \in \mathbb{N}$ s.t.

- $d \mid a \wedge d \mid b$, and
- $\forall e \in \mathbb{N} \,.\, (e \mid a \wedge e \mid b) \Rightarrow e \mid d.$

Note that the quantifier must precede the proposition to which it refers. It can not just be tagged on to the end of a proposition as might be the case when writing in English. This is rather like a scope rule in a programming language and 'variables' must be declared before they are referenced. In the example above, it would not make sense to refer to d before the values of a and b were known.

Exercises

1. Let $\mathrm{A} = \{1, 3, 5\}$ and $\mathrm{B} = \{2, 3\}$. Write down explicit sets for:
   - $\mathrm{A} \cup \mathrm{B}$ and $\mathrm{A} \cap \mathrm{B}$
   - $\mathrm{A} \setminus \mathrm{B}$ and $\mathrm{B} \setminus \mathrm{A}$
   - $(\mathrm{A} \cup \mathrm{B}) \setminus \mathrm{B}$ and $(\mathrm{A} \setminus \mathrm{B}) \cup \mathrm{B}$
   - $\mathrm{A} \Delta \mathrm{B}$ and $\mathrm{B} \Delta \mathrm{A}$
   - $\mathrm{A} \times \mathrm{B}$, $\mathrm{B} \times \mathrm{A}$ and $\mathrm{A} \times \varnothing$
   - $\mathrm{A} + \mathrm{B}$, $\mathrm{B} + \mathrm{A}$ and $\mathrm{A} + \varnothing$
2. Let A, B and C be sets. Prove or find counter-examples to:
   - $\mathrm{A} \cup (\mathrm{B} \cup \mathrm{C}) = (\mathrm{A} \cup \mathrm{B}) \cup \mathrm{C}$
   - $\mathrm{A} \cup (\mathrm{B} \cap \mathrm{C}) = (\mathrm{A} \cap \mathrm{B}) \cup \mathrm{C}$
   - $\mathrm{A} \cup (\mathrm{B} \cap \mathrm{C}) = (\mathrm{A} \cap \mathrm{B}) \cup (\mathrm{A} \cap \mathrm{C})$
   - $\mathrm{A} \setminus (\mathrm{B} \cup \mathrm{C}) = (\mathrm{A} \setminus \mathrm{B}) \cap (\mathrm{A} \setminus \mathrm{C})$
   - $\mathrm{A} \setminus (\mathrm{B} \Delta \mathrm{C}) = (\mathrm{A} \setminus \mathrm{B}) \Delta (\mathrm{A} \setminus \mathrm{C})$
   - $(\mathrm{A} \times \mathrm{C}) \cup (\mathrm{B} \times \mathrm{D}) = (\mathrm{A} \cup \mathrm{B}) \times (\mathrm{C} \cup \mathrm{D})$
   - $(\mathrm{A} \times \mathrm{C}) \cap (\mathrm{B} \times \mathrm{D}) = (\mathrm{A} \cap \mathrm{B}) \times (\mathrm{C} \cap \mathrm{D})$
3. What is the difference between $\forall\, x\, .\, (\exists\, y\, .\, \mathrm{P}(x, y))$ and $\exists\, y\, .\, (\forall\, x\, .\, \mathrm{P}(x, y))$? You might like to consider the universe to be the set of people and $\mathrm{P}(x, y)$ to mean "x loves y".
4. If $|\mathrm{A}| = m$ and $|\mathrm{B}| = n$, what are the sizes of $\mathrm{A} \times \mathrm{B}$ and $\mathrm{A} + \mathrm{B}$? How big is $\mathcal{P}(\mathrm{A})$? How many subsets of A are there of even size?
5. Of 100 students, 35 play football, 36 row and 24 play tiddlywinks. 13 play football and row, 2 play football and tiddlywinks but never row, 12 row and play tiddlywinks while 4 play every game in sight to avoid work of any form. How many students participate in none of these three vices?
6. Enumerate all the partitions of $\{1, 2, 3\}$.
7. Simplify the Boolean expression $\neg(\neg(\mathrm{a} \wedge \neg(\mathrm{a} \wedge \mathrm{b})) \wedge \neg(\neg(\mathrm{a} \wedge \mathrm{b}) \wedge \mathrm{b}))$.
8. Use Boolean simplification to show that $\{[(a \Rightarrow b) \vee (a \Rightarrow d)] \Rightarrow (b \vee d)\} = a \vee b \vee d$.
9. Consider the argument: "If Anna can cancan or Kant can't cant, then Greville will cavil vilely. If Greville will cavil vilely, Will will want. But Will won't want. Therefore Kant can cant." By rewriting the statement inside the double quotes as a single Boolean expression in terms of four variables and simplifying, show that it is true and hence that the argument is valid.

10. Given a set $\mathrm{A} \subseteq \Omega$, define the *indicator* or *characteristic* function for A for $\mathrm{x} \in \Omega$ by $I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$. Prove that:

    - $I_{\mathrm{A}\cup\mathrm{B}}(x) = \mathrm{MAX}\,(I_{\mathrm{A}}(x), I_{\mathrm{B}}(x))$
    - $I_{\mathrm{A}\cap\mathrm{B}}(x) = \mathrm{MIN}\,(I_{\mathrm{A}}(x), I_{\mathrm{B}}(x)) = I_{\mathrm{A}}(x)\, I_{\mathrm{B}}(x)$
    - $I_{\mathrm{A}^{\mathrm{c}}}(x) = 1 - I_{\mathrm{A}}(x)$
    - $\mathrm{A} = \{x \in \Omega \mid I_{\mathrm{A}}(x) = 1\}$
    - $|A| = \sum_{x\in\Omega} I_A(x)$

11. Observe that $|\mathrm{A} \cup \mathrm{B}| = |\mathrm{A}| + |\mathrm{B}| - |\mathrm{A} \cap \mathrm{B}|$ and
$|\mathrm{A} \cup \mathrm{B} \cup \mathrm{C}| = |\mathrm{A}| + |\mathrm{B}| + |\mathrm{C}| - |\mathrm{A} \cap \mathrm{B}| - |\mathrm{A} \cap \mathrm{C}| - |\mathrm{B} \cap \mathrm{C}| + |\mathrm{A} \cap \mathrm{B} \cap \mathrm{C}|$.

    This leads to the general *principle of inclusion and exclusion*. Given a collection, $\mathscr{C}$, of sets, say $\mathscr{C} = \{\mathrm{A}_s \mid s \in \mathrm{S}\}$ for some index set, $\mathrm{S} = \{1, 2, \ldots, n\}$, write

    $\bigcup_{s\in S} A_s = A_1 \cup A_2 \cup \cdots \cup A_n$ and then $\left|\bigcup_{s\in S} A_s\right| = -\sum_{\phi\neq T\subseteq S} (-1)^{|T|} \left|\bigcap_{t\in T} A_t\right|$, where the sum is taken with T ranging over all the non-empty subsets of S.

    Write this out in full for $\mathrm{S} = \{1, 2, 3\}$ and check that the result is the same as the second observation above.

12. There are n students at St Botolph's College, each with an individual pigeon hole in the Porters' Lodge. Because of the University policy on anonymous candidature, the porters are obliged to deliver examination results by posting n letters randomly, one in each of the n pigeon holes. Of course, there are $n!$ ways of doing this. Show that $n!\sum_{k=0}^{n} \frac{(-1)^k}{k!}$ of them result in no student receiving the correct letter.

    Hint: This is a question about inclusion and exclusion. Each delivery run gives a permutation of $\{1, 2, \ldots, n\}$. Let Ω be the set of all such permutations, and A_k be the subset of permutations resulting in the letter to student k being correctly delivered. We need to know $\left|\overline{\bigcup_{k=1}^{n} A_k}\right|$, which is just $|\Omega| - \left|\bigcup_{k=1}^{n} A_k\right|$.

    What is the limiting probability as $n \to \infty$?

Relations

A *relation*, R, between two sets, A and B, is just a subset $\mathrm{R} \subseteq \mathrm{A} \times \mathrm{B}$. A relation, R, on a single set, A, is just a subset $\mathrm{R} \subseteq \mathrm{A} \times \mathrm{A}$.

We write $a \ \mathrm{R} \ b$ as shorthand for $(a, b) \in \mathrm{R}$.

Write $\mathcal{R}(\mathrm{A}, \mathrm{B})$ for the collection of all relations between two sets, A and B. Obviously $\mathcal{R}(\mathrm{A}, \mathrm{B}) = \mathcal{P}(\mathrm{A} \times \mathrm{B})$.

Composition of relations

Suppose $\mathrm{R} \subseteq \mathrm{A} \times \mathrm{B}$ is a relation between A and B, and $\mathrm{S} \subseteq \mathrm{B} \times \mathrm{C}$ is a relation between B and C, then we define the *composition* of R and S to be the relation between A and C defined by $\mathrm{R} \circ \mathrm{S} = \{(a, c) \mid \exists\, b \in \mathrm{B} \,.\, (a, b) \in \mathrm{R} \wedge (b, c) \in \mathrm{S}\} \subseteq \mathrm{A} \times \mathrm{C}$. Sometimes this is written as $\mathrm{S} \circ \mathrm{R}$, which may seem confusing but is actually sensible for reasons that will become apparent later.

This can be extended to n-fold composition. Given a relation R on a set, A, write:

$\mathrm{R}^1 = \mathrm{R}$
$\mathrm{R}^n = \mathrm{R}^{n-1} \circ \mathrm{R}$ for $n > 1$.

We can also define the *inverse* of a relation: $\mathrm{R}^{-1} = \{(b, a) \mid (a, b) \in \mathrm{R}\}$, which is a relation between B and A. Observe that $(\mathrm{R} \circ \mathrm{S})^{-1} = \mathrm{S}^{-1} \circ \mathrm{R}^{-1}$.

Abstractions

There is an important pattern underlying the mathematics in the remainder of this course. We will introduce a series of abstract properties. From each of these properties we will prove some general theorems. We will then find specific examples that satisfy the abstract properties, and will be able to apply the general theorems to the specific examples.

The first abstract property applies to relations, and considers a special sort of relation.

Equivalence relations

An *equivalence relation* is a relation, R, on a set, A, satisfying three properties:

- Reflexive: $\forall\, a \in \mathrm{A} \,.\, (a, a) \in \mathrm{R}$
- Symmetric: $(a, b) \in \mathrm{R} \Rightarrow (b, a) \in \mathrm{R}$
- Transitive: $(a, b) \in \mathrm{R} \wedge (b, c) \in \mathrm{R} \Rightarrow (a, c) \in \mathrm{R}$

Intuitively, we can think of an equivalence relation as a sort of weak equality: $(a, b) \in \mathrm{R}$ means that a and b are indistinguishable within some framework.

Examples

- Given $n \in \mathbb{N}$, define R on $\mathbb{Z}$ by $(a, b) \in \mathrm{R} \Leftrightarrow n \mid (b - a)$.
- Define S on $\mathbb{Z} \times \mathbb{N}$ by $((z_1, n_1), (z_2, n_2)) \in \mathrm{S} \Leftrightarrow z_1 \, n_2 = z_2 \, n_1$.

Proving that these specific examples satisfy the abstract properties for an equivalence relations involves writing down the preconditions for each of the three properties and deducing that the corresponding conclusion is true. The beginning and ending of each part will almost be like applying a rubber stamp.

Equivalence classes

Given an equivalence relation R on a set, A, define the *equivalence class* of an element $a \in \mathrm{A}$ to be the set of elements of A related to a: $[a] = \{b \in \mathrm{A} \mid (a, b) \in \mathrm{R}\} = \{b \in \mathrm{A} \mid a \mathrm{R} b\}$.

The set of equivalence classes $\{[a] \mid a \in \mathrm{A}\}$ forms a partition of A, that is:

- The classes cover A: $\bigcup_{a \in \mathrm{A}} [a] = \mathrm{A}$.

  Proof: Given any $a \in \mathrm{A}$, $a \mathrm{R} a$ by reflexivity so $a \in [a]$ and $a \in \bigcup_{a \in \mathrm{A}} [a]$.

- They are disjoint (or equal): $\forall a, b \in \mathrm{A} . [a] \cap [b] \neq \varnothing \Rightarrow [a] = [b]$.

  Proof: Suppose $x \in [a] \cap [b]$, so $a \mathrm{R} x$ and $b \mathrm{R} x$. Then $x \mathrm{R} b$ by symmetry. Given any $v \in [b]$, observe $b \mathrm{R} v$. Now $a \mathrm{R} x$, $x \mathrm{R} b$ and $b \mathrm{R} v$, so $a \mathrm{R} v$ by transitivity. Therefore $v \in [a]$ and so $[b] \subseteq [a]$. But $[a] \subseteq [b]$ similarly, so $[a] = [b]$.

The set of equivalence classes is called the *quotient set*, $\mathrm{A}_{/\mathrm{R}}$.

In the two examples above, $\mathbb{Z}_{/\mathrm{R}}$ represents the integers *modulo n* and $(\mathbb{Z} \times \mathbb{N})_{/\mathrm{S}}$ represents the rational numbers. Equivalence classes are useful for analysing sets where there may be different representations of the same value. For example, $\frac{1}{2} = \frac{2}{4}$.

Closures

Given a set, Ω, a property, P, of subsets of Ω and a particular subset $\mathrm{S} \subseteq \Omega$, which may or may not satisfy P, we might ask the question, "What is the smallest subset of Ω containing S which does satisfy P?" That is, find $\mathrm{C} \subseteq \Omega$ such that:

- $\mathrm{S} \subseteq \mathrm{C}$.
- $\mathrm{P}(\mathrm{C})$ is true.
- If $\mathrm{D} \subseteq \Omega$ also satisfies $\mathrm{S} \subseteq \mathrm{D}$ and $\mathrm{P}(\mathrm{D})$, then $\mathrm{C} \subseteq \mathrm{D}$.

Such a set, C, is called the P-*closure* of S. Such a closure need not necessarily exist. However, there is one particular class of properties for which closures *will* always exist. These are the *intersection-closed* properties.

Let $\mathscr{C} = \{\mathrm{S} \subseteq \Omega \mid \mathrm{P}(\mathrm{S}) \text{ is true}\}$ be the collection of all subsets of Ω satisfying P. P is intersection-closed if, for any subset $\mathscr{B} \subseteq \mathscr{C}$, the intersection of all the subsets in $\mathscr{B}$ also satisfies P. That is, if $I = \bigcap_{S \in B} S$, then $\mathrm{P}(I)$ is true.

We can now calculate the P-closure of a given subset $\mathrm{S} \subseteq \Omega$ as $\bigcap \{B \subseteq \Omega \mid S \subseteq B \wedge P(B) \text{ is true}\}$ as long as $\mathrm{P}(\Omega)$ is true.

If we consider relations on a set A, which are just subsets of $\Omega = \mathrm{A} \times \mathrm{A}$, it turns out that reflexivity, symmetry and transitivity are all intersection-closed properties. Given a relation, $\mathrm{R} \subseteq \mathrm{A} \times \mathrm{A}$, this allows us to form the:

- Reflexive closure: $\mathrm{r}(\mathrm{R}) = \mathrm{R} \cup \mathrm{I}_{\mathrm{A}}$ where $\mathrm{I}_{\mathrm{A}} = \{(a, a) \mid a \in \mathrm{A}\}$
- Symmetric closure: $\mathrm{s}(\mathrm{R}) = \mathrm{R} \cup \mathrm{R}^{-1}$

- Transitive closure: $t(\mathrm{R}) = \mathrm{R}^+$ where $\mathrm{R}^+ = \mathrm{R} \cup \mathrm{R}^2 \cup \mathrm{R}^3 \cup \mathrm{R}^4 \cup \ldots$, and R^k is the k-fold composition of R with itself.

  Observe that $(a, b) \in t(\mathrm{R}) \Leftrightarrow \exists\, n \in \mathbb{N}$ and $x_0 = a, x_1, x_2, \ldots x_n = b$ with $(x_i, x_{i+1}) \in \mathrm{R}$ for $0 \le i < n$.

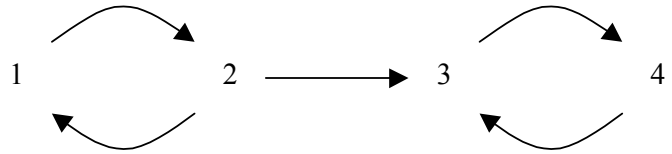
  If A is finite, the union can stop at R^n, where $n = |\mathrm{A}|$.

Warshall's algorithm

It is helpful to regard calculating the transitive closure as a route finding problem in a graph. Consider the elements of A to be locations identified by natural numbers, two of which are related by R if they are directly connected. Two locations are related by R^2 if they are connected via a path with two steps, by R^3 if they are linked via a path with three steps and so on. Two locations are related by the transitive closure of R if they are indirectly connected via an arbitrarily long sequence of intermediate steps.

Suppose that A is finite with $|\mathrm{A}| = n$. Then R can be represented as an $n \times n$ array of Boolean values. Calculating R^2 is rather like a matrix multiplication requiring n^2 values to be found, each of which is the sum of n products, which makes it an $O(n^3)$ operation. It turns out that, when $|\mathrm{A}| = n$, the union of powers of R in the transitive closure can stop at R^n (no path will require more than n intermediate steps), so forming the transitive closure naïvely is an $O(n^4)$ operation.

For example, consider the following graph:



$$m = \begin{bmatrix} \cdot & t & \cdot & \cdot \\ t & \cdot & t & \cdot \\ \cdot & \cdot & \cdot & t \\ \cdot & \cdot & t & \cdot \end{bmatrix},\ m^2 = \begin{bmatrix} t & \cdot & t & \cdot \\ \cdot & t & \cdot & t \\ \cdot & \cdot & t & \cdot \\ \cdot & \cdot & \cdot & t \end{bmatrix},\ m^3 = \begin{bmatrix} \cdot & t & \cdot & t \\ t & \cdot & t & \cdot \\ \cdot & \cdot & \cdot & t \\ \cdot & \cdot & t & \cdot \end{bmatrix} \text{ and } m^4 = \begin{bmatrix} t & \cdot & t & \cdot \\ \cdot & t & \cdot & t \\ \cdot & \cdot & t & \cdot \\ \cdot & \cdot & \cdot & t \end{bmatrix} \text{ so}$$

$t(\mathrm{R}) = m \vee m^2 \vee m^3 \vee m^4 = \begin{bmatrix} t & t & t & t \\ t & t & t & t \\ \cdot & \cdot & t & t \\ \cdot & \cdot & t & t \end{bmatrix}$. (In fact, the m^4 term is unnecessary for this example because the longest path is only three steps long.)

However, we can do better than this.

The outer loop of the naïve algorithm iterates over the length of the path linking two locations. Warshall's algorithm has a different structure in which the outer loop iterates over highest numbered intermediate point encountered along a path.

Suppose that $\mathrm{A} = \{1, 2, 3, \ldots, n\}$ and represent R by the Boolean matrix $[m(i,j)]$ where $m(i,j) = ((i,j) \in \mathrm{R})$. Now define $m_k(i,j)$ to be true if an only if there is a path from i to j using only intermediate locations numbered between 1 and k. So $m_0 = m$ representing direct connections that do not require any intermediate locations.

In order to get from i to j using only intermediate locations numbered between 1 and $k{+}1$, either we can do it using only locations between 1 and k or we must visit location $k{+}1$, but there will be at most one such visit. So $m_{k+1}(i,j) = m_k(i,j)) \vee (m_k(i, k{+}1) \wedge m_k(k{+}1, j))$. Finally m_n will be the transitive closure, allowing any intermediate locations.

For the above example, $m_0 = m = \begin{bmatrix} \cdot & t & \cdot & \cdot \\ t & \cdot & t & \cdot \\ \cdot & \cdot & \cdot & t \\ \cdot & \cdot & t & \cdot \end{bmatrix}$, $m_1 = \begin{bmatrix} \cdot & t & \cdot & \cdot \\ t & t & t & \cdot \\ \cdot & \cdot & \cdot & t \\ \cdot & \cdot & t & \cdot \end{bmatrix}$, $m_2 = \begin{bmatrix} t & t & t & \cdot \\ t & t & t & \cdot \\ \cdot & \cdot & \cdot & t \\ \cdot & \cdot & t & \cdot \end{bmatrix}$, $m_3 = \begin{bmatrix} t & t & t & t \\ t & t & t & t \\ \cdot & \cdot & \cdot & t \\ \cdot & \cdot & t & t \end{bmatrix}$ and $m_4 = \begin{bmatrix} t & t & t & t \\ t & t & t & t \\ \cdot & \cdot & t & t \\ \cdot & \cdot & t & t \end{bmatrix} = t(\mathrm{R})$. In this case, all the steps are necessary because the only route from 3 to 3 is via 4 which appears only at the last iteration.

It is necessary to iterate over *i, j* and k (in the right order), so this is an O(n^3) operation.

Partial orders

A *partial order* is a relation R on a set, A, satisfying three properties:

- Reflexive: $\forall\, a \in \mathrm{A}\,.\,(a, a) \in \mathrm{R}$
- Anti-symmetric: $(a, b) \in \mathrm{R} \wedge (b, a) \in \mathrm{R} \Rightarrow a = b.$
- Transitive: $(a, b) \in \mathrm{R} \wedge (b, c) \in \mathrm{R} \Rightarrow (a, c) \in \mathrm{R}$

Intuitively, we can think of $(a, b) \in \mathrm{R}$ as meaning $a \le b$. The notation $a\ \mathrm{R}\ b$ meaning $(a, b) \in \mathrm{R}$ may have seemed strange for arbitrary relations, but $a \le b$ reads more easily than $(a, b) \in \le$. The order is partial because it is possible to have pairs of elements that are not comparable - $(a, b) \notin \mathrm{R}$ and $(b, a) \notin \mathrm{R}$.

It is possible to have two different partial orders on a single set so it may be necessary to refer to the pair (A, R) or $(\mathrm{A}, \le)$ rather than just A, to avoid ambiguity.

Total order

A partially ordered set, $(\mathrm{A}, \le)$, is *totally ordered* if any pair of elements of A can be compared using $\le$. That is, $\forall\, a, b \in \mathrm{A}\,.\,(a \le b) \vee (b \le a)$.

Examples

- Conventional numerical order on $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ or $\mathbb{R}$, all of which are total.
- Division order on $\mathbb{N}$ or $\mathbb{N}_0$: $(a, b) \in \mathrm{D} \Leftrightarrow a \mid b$. This is partial, for example 2 and 3 can not be compared.
- However, the division order on $\mathbb{Z}$ is *not* a partial order.
- Division order on $\mathrm{D}_n = \{x \in \mathbb{N} \mid x \mid n\}$ for any $n \in \mathbb{N}$.
- For any set A, the power set of A ordered by subset inclusion, $(\mathbb{P}(\mathrm{A}), \subseteq)$. Again, this is partial.
- For any two partially ordered sets $(\mathrm{A}, \le_\mathrm{A})$ and $(\mathrm{B}, \le_\mathrm{B})$ there are two important orders on the product set $\mathrm{A} \times \mathrm{B}$.

- *Product* order: $(a_1, b_1) \leq_P (a_2, b_2) \Leftrightarrow (a_1 \leq_A a_2) \wedge (b_1 \leq_B b_2)$.
- *Lexicographic* order: $(a_1, b_1) \leq_L (a_2, b_2) \Leftrightarrow (a_1 <_A a_2) \vee (a_1 = a_2 \wedge b_1 \leq_B b_2)$.

If $(A, \leq_A)$ and $(B, \leq_B)$ are both total orders, then the lexicographic order on $A \times B$ will be total, but the product order will generally only be partial.

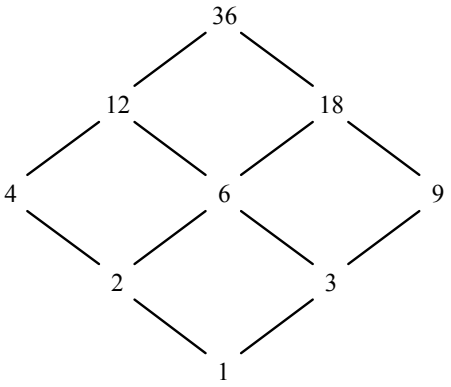
$(a_1, b_1) \leq_P (a_2, b_2) \Rightarrow (a_1, b_1) \leq_L (a_2, b_2)$, so the product order (considered as a subset of $(A \times B) \times (A \times B)$) is a subset of the lexicographic order.

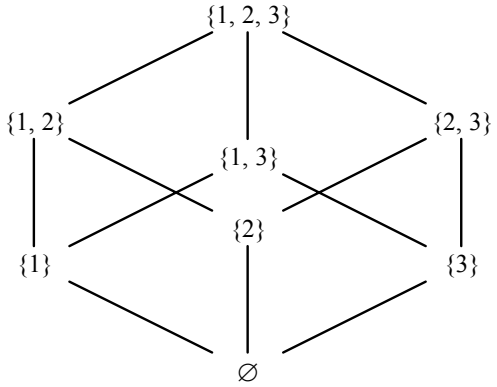
- For any totally ordered (finite) alphabet A, $A^* = \{\varepsilon\} \cup A \cup A^2 \cup A^3 \cup \ldots$ is the set of all strings made from that alphabet, where ε is the empty string. The *full lexicographic* order, $\leq_F$, on A^* is defined recursively as follows. Given two words $u, v \in A^*$, if $u = \varepsilon$ then $u \leq_F v$ and if $v = \varepsilon$ then $v \leq_F u$. Otherwise, both u and v are non-empty so we can write $u = u_1 x$ and $v = v_1 y$ where u_1 and v_1 are the first letters of u and v respectively. Now $u \leq_F v \Leftrightarrow (u = \varepsilon) \vee (u_1 <_A v_1) \vee (u_1 = v_1 \wedge x \leq_F y)$

Hasse diagrams

A *Hasse diagram* represents a partial order pictorially as a directed graph with nodes for the elements of the underlying set and arcs between pairs of elements related by the order but with no intermediate elements in the order. For simplicity, we omit the arcs for reflexivity and even omit the arrows on the arcs if they point up the page.

Here are the Hasse diagrams for $(D_{36}, |)$ and $(\mathcal{P}(\{1, 2, 3\}), \subseteq)$:





Well founded relations

Given a partial order, $(A, \leq)$, we can define a relation, $(A, <)$, as follows:

$$a_1 < a_2 \Leftrightarrow (a_1 \leq a_2) \wedge (a_1 \neq a_2)$$

Similarly, given a relation, $(A, <)$, we can write $\leq$ for the relation on A defined by:

$$a_1 \leq a_2 \Leftrightarrow (a_1 < a_2) \vee (a_1 = a_2)$$

$<$ is not reflexive and neither $<$ nor $\leq$ need be transitive. We can also write $>$ and $\geq$ for the inverse relations defined by:

$$a_1 > a_2 \Leftrightarrow a_2 < a_1$$
$$a_1 \geq a_2 \Leftrightarrow a_2 \leq a_1$$

A relation $(A, <)$ is *well founded* if $\forall S \subseteq A \,.\, S \neq \varnothing \Rightarrow \exists m \in S \,.\, \forall a \in A \,.\, a < m \Rightarrow a \notin S$

That is, the relation < is well founded if and only if every non-empty subset $S \subseteq A$ contains a < -minimal element. The condition for m to be minimal is expressed slightly strangely to accommodate the possibility of it not being possible to compare it with some elements of S.

In other words, if a set with a well-founded relation $<$ contains a counterexample to some property, then it contains a $<$ -minimal counterexample. (Just let S be the set of all counterexamples.) Note that the minimum need not be unique.

Note also that the definition is asymmetric: it says nothing about ascending sequences.

Proposition

Let $<$ be a relation on a set A. The relation $<$ is well founded if and only if every infinite sequence $a_1, a_2, a_3, \ldots$ of elements in A with $a_1 \geq a_2 \geq a_3 \geq \ldots$ is ultimately constant, so there is some $N \in \mathbb{N}$ such that $a_n = a_N$ for all $n \geq N$.

Proof: ($\Rightarrow$) Suppose $(A, <)$ is well founded and $a_1 \geq a_2 \geq a_3 \geq \ldots$. Let $S = \{a_1, a_2, a_3, \ldots\}$, so $S \subseteq A$ and $S \neq \varnothing$. By well-foundedness, there is a minimal element $m \in S$. Suppose that $m = a_N$. Now, for any $n \geq N$, $a_n \leq a_N = m$. But, if $a_n < m$, then $a_n \notin S$, which would be a contradiction. Hence $a_n = m = a_N$.

($\Leftarrow$) Suppose that every descending sequence is ultimately constant and $S \subseteq A$ with $S \neq \varnothing$. Pick $a_1 \in S$. Either a_1 is a minimal element in S, in which case we are done, or we can pick $a_2 \in S$ with $a_2 < a_1$. Proceed in this way to form $a_1 > a_2 > a_3 > \ldots$, which would be an infinite descending sequence, which is not possible. Hence there is a minimal element in S.

Examples

- Conventional numerical order $<$ on $\mathbb{N}$ but *not* on $\mathbb{Z}$ or $\mathbb{Q}$ or $\mathbb{R}$.
- If two relations $<_A$ on A and $<_B$ on B are well founded, then the lexicographic relation $<_L$ on $A \times B$ defined by:
  $$(a_1, b_1) <_L (a_2, b_2) \Leftrightarrow (a_1 <_A a_2) \vee (a_1 = a_2 \wedge b_1 <_B b_2)$$
  is also well founded.

  Proof: Suppose $S \subseteq A \times B$ with $S \neq \varnothing$. Let $U = \{a \in A \mid \exists\, b \in B \,.\, (a,b) \in S\}$. $U \subseteq A$ and $U \neq \varnothing$, so there is a minimal element $u \in U$. Now let $V = \{b \in B \mid (u,b) \in S\}$. $V \subseteq B$ and $V \neq \varnothing$, so there is a minimal element $v \in V$. Now $(u,v) \in S$ is a minimal element.

  Alternative proof: Suppose $(a_1, b_1) \geq_L (a_2, b_2) \geq_L (a_3, b_3) \geq_L \ldots$. Then $a_1 \geq_A a_2 \geq_A a_3 \geq_A \ldots$ by the definition of the lexicographic relation, so the sequence is ultimately constant because $(A, <_A)$ is well founded. So $\exists\, M \in \mathbb{N}$ such that $\forall\, m \geq M \,.\, a_m = a_M$. Now $b_M \geq_B b_{M+1} \geq_B b_{M+2} \geq_B \ldots$, so this sequence is also ultimately constant because $(B, <_B)$ is well founded. So $\exists\, N \in \mathbb{N}$ such that $\forall\, n \geq N \,.\, b_n = b_N$. But $N \geq M$, so $\forall\, n \geq N \,.\, (a_n, b_n) = (a_N, b_N)$ and the original sequence is ultimately constant.
- The product relation (defined in the obvious way) is well founded.

  Proof: $(a_1, b_1) \geq_P (a_2, b_2) \geq_P (a_3, b_3) \geq_P \ldots$ then $(a_1, b_1) \geq_L (a_2, b_2) \geq_L (a_3, b_3) \geq_L \ldots$, so again the sequence is ultimately constant.
- The full lexicographic relation on A^* is *not* well founded if A has more than one element. Consider the sequence b, ab, aab, aaab, ….
- Any irreflexive relation on a finite set is well founded.

- No reflexive relation on a set can be well founded. In particular, no partial order is well founded. However, we often refer to a partial order $\leq$ being well founded when we actually mean that the derived relation $<$ is well founded.

Well ordering

A total order $(\mathrm{A}, \leq)$ is a *well ordering* if $<$ is well founded. This is equivalent to saying that every subset of A contains a minimal element with respect to $\leq$. That is, $\forall\ \mathrm{S} \subseteq \mathrm{A}\ .\ \mathrm{S} \neq \varnothing \Rightarrow \exists\ m \in \mathrm{S}\ .\ \forall\ s \in \mathrm{S}\ .\ s \leq m \Rightarrow s = m$. The minimal element will be unique since $\leq$ is total. This is the characterisation of well ordering in the natural numbers that was used in the first half of this course.

A *chain* in a partially ordered set $(\mathrm{A}, \leq)$ is a subset $\mathrm{C} \subseteq \mathrm{A}$ that is totally ordered by $\leq$. An infinite descending chain is a sequence of elements $a_1 > a_2 > a_3 > \ldots$. $(\mathrm{A}, \leq)$ is well ordered precisely when there are no infinite descending chains of elements in A.

Topological sorting

Suppose that $(\mathrm{A}, \leq)$ is partially ordered but is not totally ordered. Can we find a *topological sort* of A, that is, a total order on A that respects $\leq$? If A is finite then $\leq$ is necessarily well-founded and we can develop an algorithm as follows.

A is itself a subset of A and so it contains a minimal element. Put this first in the total order. Now take the rest of A, find a minimal element and put it second. Continue in this way to build up a total order on the whole of A.

In fact there is a slight subtlety. At the each step there may be more than one minimal element. These can not be compared with each other, so it does not matter what order they have in the total order. Instead of putting just one of them into the total order we could include all of them in some arbitrary order before going on to the next step and finding the minimal elements in the remainder of A.

The former is a *depth first* algorithm, the latter a *breadth first* one, and they may well give rise to two different total orders, each of which respects the original partial order.

Complete ordering

A partially ordered set, $(\mathrm{A}, \leq)$, is *complete* if every (ascending) chain in A has a least upper bound in A. The least upper bound need not appear in the chain itself. This will prove useful in giving a mathematical meaning to the behaviour of programs.

Well founded induction

Given a well founded relation, $<$, on a set A and a Boolean proposition, P, involving elements of A, let M be the set of minimal elements in A (with respect to $<$). The *principle of well founded induction* states that, if

- $\forall\ m \in \mathrm{M}\ .\ \mathrm{P}(m)$ is true, and
- $\forall\ b \in \mathrm{A}\ .\ [(\forall\ c \in \mathrm{A}\ .\ c < b \Rightarrow \mathrm{P}(c)) \Rightarrow \mathrm{P}(b)]$,

then $\forall\ a \in \mathrm{A}\ .\ \mathrm{P}(a)$. These are rather like the base case and inductive step in mathematical induction. In fact we can even omit the first condition if we understand $c < b \Rightarrow \mathrm{P}(c)$ to be true if $b \in \mathrm{M}$ (so there is no $c < b$).

Examples

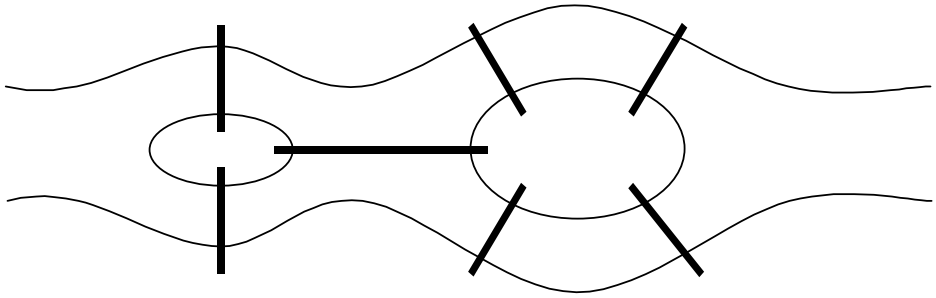
- Ackermann's function is defined by

```
fun ack (0, n) = n + 1
  | ack (m, 0) = ack (m-1, 1)
  | ack (m, n) = ack (m-1, ack (m, n-1));
```

Is `ack` well defined for all values of `m` and `n` ≥ 0? The answer is "yes" and the proof uses well founded induction on the arguments in $\mathbb{N}_0 \times \mathbb{N}_0$ under the lexicographic order.

Given an argument pair `(m, n)`, observe that the definition only uses applications of `ack` with argument pairs that come earlier in the lexicographic order, so `ack (m, n)` is well defined if they are. However, `ack` is well defined for the (unique) minimal argument pair `(0, 0)`, so it is well defined for all argument pairs.

- The town of Königsberg spans a river with two islands linked to the banks and each other by seven bridges:



Is it possible to set out from any point in the town and cross each bridge exactly once, returning to the starting point? In this case the answer is "no".

In general, this is the problem of finding an *Eulerian circuit* in a graph. A graph consists of a set of *nodes* or *vertices* (the two river banks and the two islands in Königsberg) linked by *arcs* or *edges* (the bridges). Formally, $G = (V, E)$ where V is the set of vertices, E the set of edges and $E \subseteq V \times V$ so E is just a relation on V. Often we will consider directed graphs, but in this case E is symmetric and the arcs are not directed. A connected graph has an Eulerian circuit if (and only if) every vertex has even degree, that is, it has an even number of edges connected to it.

Clearly this condition is necessary, and the proof that it is also sufficient uses induction on the set of graphs under the product order: $G_1 \leq G_2 \Leftrightarrow (V_1 \subseteq V_2) \wedge (E_1 \subseteq E_2)$, which will be well founded if V_1 and V_2 are finite.

Observe that the empty graph $G = (\varnothing, \varnothing)$ is uniquely minimal with respect to this ordering and satisfies the theorem.

Consider a connected graph G where every node has even degree. Pick a random vertex and set out on a circuit. Every vertex has even degree, so we only stop when we have returned to the start. Delete the edges in this circuit from G. The resulting graph will have a number of connected components, each of which precedes the original graph in the product order and so each has an Eulerian circuit (or is an isolated vertex). Link each of these into the original circuit to give an Eulerian circuit for the whole graph.

Exercises

1. Let $A = \{1, 2, 3, 4\}$, $B = \{a, b, c, d\}$ and $C = \{x, y, z\}$, and let $R = \{(1, a), (2, d), (3, a), (3, b), (3, d)\}$ and $S = \{(b, x), (b, z), (c, y), (d, z)\}$. What is the composition of R and S, $R \circ S$?

2. Let $A = \{1, 2, 3, 4\}$ and consider the relation $R = \{(1, 1), (2, 2), (2, 3), (3, 2), (4, 2), (4, 4)\}$ on A. Is R reflexive, symmetric, anti-symmetric or transitive? Find the reflexive, symmetric and transitive closures of R.

3. If $|A| = k$ and $|B| = m$, how many relations are there between A and B?

   If further $|C| = n$, how many ternary relations are there in $A \times B \times C$? [*Hint:* a binary relation is just a subset of $A \times B$, so a ternary relation is just a subset of $A \times B \times C$.]

4. Let $A = \{1, 2, 3\}$. How many relations are there on A? How many equivalence relations are there on A?

   Hint: Consider partitions of A.

5. Let R and S be relations between A and B. Prove that $(R \cap S)^{-1} = R^{-1} \cap S^{-1}$ and $(R \cup S)^{-1} = R^{-1} \cup S^{-1}$. Prove further that $R \subseteq S \Rightarrow R^{-1} \subseteq S^{-1}$

6. Give an example of a relation R on a set A with $|A| = n$ such that $t(R) \neq R^1 \cup R^2 \cup \ldots \cup R^{n-1}$. What must $t(R)$ be in this case?

7. Show that the smallest equivalence relation containing the two equivalence relations R and S on a set A is $t(R \cup S)$.

8. Define a relation R on $\mathbb{N}$ by $(x, y) \in R \Leftrightarrow \exists$ prime $p \,.\, y = px$, that is, y is a prime multiple of x. Describe in words the reflexive, symmetric and transitive closures of R.

9. Which of the following are true for the relation defined in the previous question:

   - $r(s(R)) = s(r(R))$
   - $r(t(R)) = t(r(R))$
   - $s(t(R)) = t(s(R))$

   Which of them hold for all relations on $\mathbb{N}$?

10. Express the smallest equivalence relation containing an arbitrary relation using the symmetric, reflexive and transitive closures.

    What is the smallest partial order containing the relation R above? Is it possible to find the smallest partial order containing an arbitrary relation?

11. Give two topological sorts of $\mathbb{N} \times \mathbb{N}$ that respect the product order. One should have the property that, given any point $(x, y) \in \mathbb{N} \times \mathbb{N}$, any infinite subset of $\mathbb{N} \times \mathbb{N}$ should include a point (x', y') with $(x, y) < (x', y')$ and the other should cause this property not to hold in general.

Functions

A (*total*) *function* f from a set A to a set B, written $f\colon \mathrm{A} \to \mathrm{B}$, is a relation $f \subseteq \mathrm{A} \times \mathrm{B}$ that satisfies:

- Uniquely defined: $(a, b_1) \in f \wedge (a, b_2) \in f \Rightarrow b_1 = b_2$
- Everywhere defined: $\forall\, a \in \mathrm{A}\ \exists\, b \in \mathrm{B}\ .\ (a, b) \in f$

We write $f(a)$ for the unique element $b \in \mathrm{B}$ with $(a, b) \in f$ to give the usual notation.

If only the first of these two properties holds, then f is a *partial function* from A to B which is undefined for certain elements of A. It is sometimes convenient to refer to this undefined value explicitly as $\perp$ (pronounced *bottom*). A partial function from A to B is the same as a total function from A to $(\mathrm{B} + \{\perp\})$.

A is called the *domain* of f, and B is called its *range*.

The set $f(\mathrm{A}) = \{b \in \mathrm{B} \mid \exists\, a \in \mathrm{A}\ .\ f(a) = b\}$ is the *image* of A under f.

Some authors call B the target or co-domain, and use the word range for the image $f(\mathrm{A})$.

Given two functions $f\colon \mathrm{A} \to \mathrm{B}$ and $g\colon \mathrm{B} \to \mathrm{C}$, the *composition* of f and g is the function $h\colon \mathrm{A} \to \mathrm{C}$ defined by $h(a) = g(f(a))$. This is just $f \circ g$ the composition of f and g as relations, and explains why $g \circ f$ is a sensible notation for that composition.

It is sometimes convenient to write $\mathrm{A} \to \mathrm{B}$ for the set of all functions from A to B.

Counting functions

If A and B are finite sets with $|\mathrm{A}| = m$ and $|\mathrm{B}| = n$, then $|\mathrm{A} \to \mathrm{B}| = n^m$.

$\mathrm{A} \to \mathrm{B}$ is sometimes written as B^{A}, so $|\mathrm{B}^{\mathrm{A}}| = |\mathrm{B}|^{|\mathrm{A}|}$.

Classifications of functions

A function $f\colon \mathrm{A} \to \mathrm{B}$ is *injective* (also described as *one-to-one* or *1-1*) if $\forall\, a_1, a_2 \in \mathrm{A}\ .\ f(a_1) = f(a_2) \Rightarrow a_1 = a_2$.

Example: $f\colon \mathbb{N} \to \mathbb{N}$ defined by $f(x) = x^2$ is injective, but $f\colon \mathbb{Z} \to \mathbb{Z}$ defined the same way would not be.

A function $f\colon \mathrm{A} \to \mathrm{B}$ is *surjective* (also described as *onto*) if $\forall\, b \in \mathrm{B}\ .\ \exists\, a \in \mathrm{A}\ .\ f(a) = b$.

Example: $f\colon \mathbb{Z} \to \mathbb{Z}$ defined by $f(x) = x + 1$ is surjective, but $f\colon \mathbb{N} \to \mathbb{N}$ defined the same way would not be.

A function $f\colon \mathrm{A} \to \mathrm{B}$ is *bijective* (also described as a *one-to-one correspondence*) if it is both injective and surjective. A bijection from a set to itself is a *permutation*.

If a function $f\colon \mathrm{A} \to \mathrm{B}$ is injective, then its inverse as a relation $f^{-1} \subseteq \mathrm{B} \times \mathrm{A}$ satisfies the uniquely defined criterion for a function. If f is surjective, then f^{-1} satisfies the everywhere defined criterion. So, given a bijection $f\colon \mathrm{A} \to \mathrm{B}$, its inverse as a relation is also a function $f^{-1}\colon \mathrm{B} \to \mathrm{A}$. In fact f^{-1} is also a bijection.

Given a universe Ω, define a relation on $\mathcal{P}(\Omega)$ by $\mathrm{A} \approx \mathrm{B}$ if and only if there is a bijection from A to B. This is an equivalence relation and two sets are said to have the same *cardinality* if they are related by it. For finite sets this means that they have the same number of elements and it is reasonable to extend the definition to infinite sets.

Observe that $\mathbb{N}_0 \approx \mathbb{N}$ (map $n \to n + 1$). Indeed, $\mathbb{Z} \approx \mathbb{N}$ (map $z \to 2z + 1$ if $z \geq 0$ and $-2z$ otherwise) and $\mathbb{N} \times \mathbb{N} \approx \mathbb{N}$ (map $(m, n) \to \frac{1}{2}(m+n-1)(m+n-2) + n$). In fact, $\mathbb{Q} \approx \mathbb{N}$ as well, but proving that requires a little preparation.

Sorting

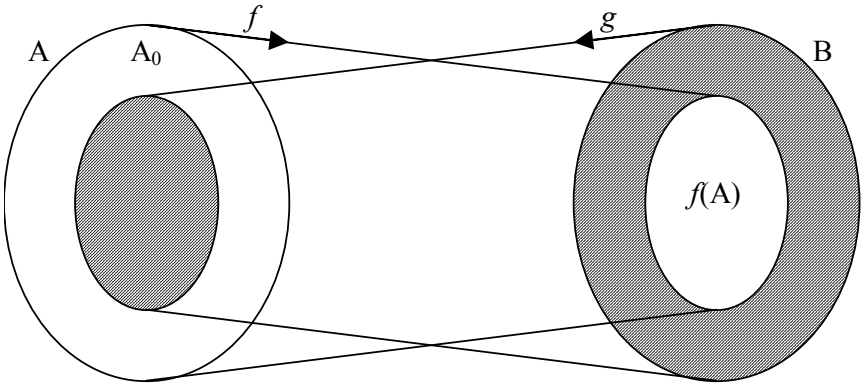
If A is a finite, totally ordered set with $|\mathrm{A}| = m$, there are $m!$ permutations of A. Sorting A involves choosing the single permutation from these $m!$ that makes a chain in A. Encoding this in binary would require $\log_2(m!) \approx m \log_2(m)$ bits of information. Any algorithm to sort A would yield one bit of information for each comparison of two elements and so we should not expect to do better than $\mathrm{O}(m \log_2(m))$ comparisons.

Schröder-Bernstein theorem

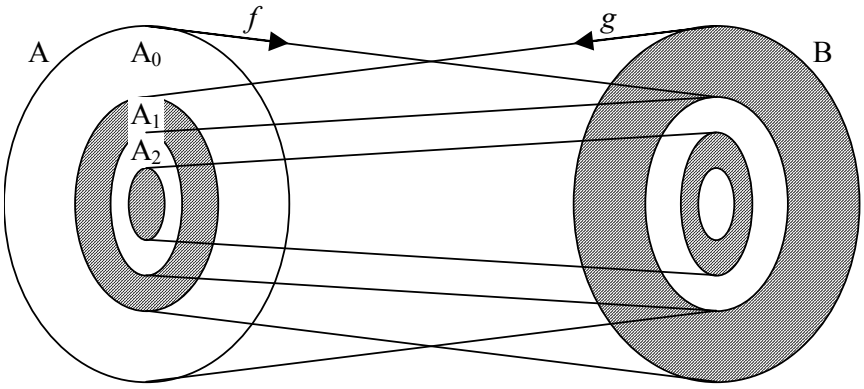
Suppose $f\colon \mathrm{A} \to \mathrm{B}$ and $g\colon \mathrm{B} \to \mathrm{A}$ are injections. Then there is a bijection from A to B.

Proof

Let $\mathrm{B}_0 = \mathrm{B} \setminus f(\mathrm{A})$ and $\mathrm{A}_0 = \mathrm{A} \setminus g(\mathrm{B})$.



Define $\mathrm{B}_n = f(\mathrm{A}_{n-1})$ and $\mathrm{A}_n = g(\mathrm{B}_{n-1})$ for $n > 0$. Observe that $f\colon \mathrm{A}_{n-1} \to \mathrm{B}_n$ and $g\colon \mathrm{B}_{n-1} \to \mathrm{A}_n$ are bijections.



Now define $\mathrm{A}_{\text{even}} = \mathrm{A}_0 \cup \mathrm{A}_2 \cup \mathrm{A}_4 \cup \cdots = \bigcup_{\text{even } n} \mathrm{A}_n$ and define A_{odd}, B_{even} and B_{odd} similarly.

Let $\mathrm{A}_\infty = \mathrm{A} \setminus (\mathrm{A}_{\text{even}} \cup \mathrm{A}_{\text{odd}})$ so $\mathrm{A} = \mathrm{A}_{\text{even}} \cup \mathrm{A}_{\text{odd}} \cup \mathrm{A}_\infty$ and these three sets are disjoint. Proceed similarly for B.

Define h: $\mathrm{A} \to \mathrm{B}$ to be equal to f on A_{even}, to g^{-1} on A_{odd} and to either on A_{∞}, which gives the desired bijection.

Countability

Recall that two sets have the same cardinality if there is a bijection between them. A set is *countably infinite* if it has the same cardinality as $\mathbb{N}$. This cardinality is known as $\aleph_0$ (the Hebrew letter Aleph with a subscript of 0). A *countable* set is either finite or countably infinite.

The Schröder-Bernstein theorem shows that any set, A, is countable if, and only if, there is an injection $\mathrm{A} \to \mathbb{N}$ or, equivalently, there is a surjection $\mathbb{N} \to \mathrm{A}$.

Example: $\mathbb{Q}$ is countable. We can construct injections $\mathbb{Q} \to \mathbb{Z} \times \mathbb{N}$, $\mathbb{Z} \times \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ and $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$, which can be composed to form an injection $\mathbb{Q} \to \mathbb{N}$ and so $\mathbb{Q}$ is countable.

Countable union of countable sets

Suppose that $\{\mathrm{A}_i \mid i \in \mathrm{I}\}$ is a countable collection of countable sets. That is, the index set I is countable and for each $i \in \mathrm{I}$ the set A_i is countable. Then $\bigcup_{i \in I} A_i$ is countable.

Proof: I is countable, so there is an injection f: $\mathrm{I} \to \mathbb{N}$. For each $i \in \mathrm{I}$ the set A_i is countable, so there is an injection g_i: $\mathrm{A}_i \to \mathbb{N}$. Define h: $\cup \mathrm{A}_i \to \mathbb{N}$ as follows. For any $x \in \cup\, \mathrm{A}_i$, let m be the minimal element of $\{ f(i) \mid x \in \mathrm{A}_i \}$ and let j be $f^{-1}(m)$. Now let $h(x) = p_m^{g_j(x)}$ where p_m is the m^{th} prime. Then h is an injection and so $\cup \mathrm{A}_i$ is countable.

Uncountability of $\mathcal{P}(\mathbb{N})$

Suppose that $\mathcal{P}(\mathbb{N})$ is countable, so there is a bijection f: $\mathbb{N} \to \mathcal{P}(\mathbb{N})$.

Let $\mathrm{A} = \{n \in \mathbb{N} \mid n \notin f(n)\} \subseteq \mathbb{N}$ so $\mathrm{A} \in \mathcal{P}(\mathbb{N})$, and let $a \in \mathbb{N}$ be such that $f(a) = \mathrm{A}$.

Now ask whether or not $a \in \mathrm{A}$? Suppose so, then $a \notin f(a) = \mathrm{A}$, a contradiction. Suppose not, then $a \in f(a) = \mathrm{A}$, another contradiction. Hence f could not exist and so $\mathcal{P}(\mathbb{N})$ is not countable.

Uncountability of $\mathbb{R}$

Suppose that $\mathbb{R}$ is countable, so there is a bijection g: $\mathbb{R} \to \mathbb{N}$. Define h: $\mathcal{P}(\mathbb{N}) \to \mathbb{R}$ as follows. Given $\mathrm{A} \subseteq \mathbb{N}$, let $h(\mathrm{A}) = \sum_{a \in \mathrm{A}} 10^{-a}$, giving a decimal number between 0 and 1 with ones in digit positions corresponding to members of A and zeroes elsewhere. h is an injection, so the composition $h \circ g$: $\mathcal{P}(\mathbb{N}) \to \mathbb{N}$ is also an injection. But this would imply that $\mathcal{P}(\mathbb{N})$ were countable which gives a contradiction, so $\mathbb{R}$ is not countable.

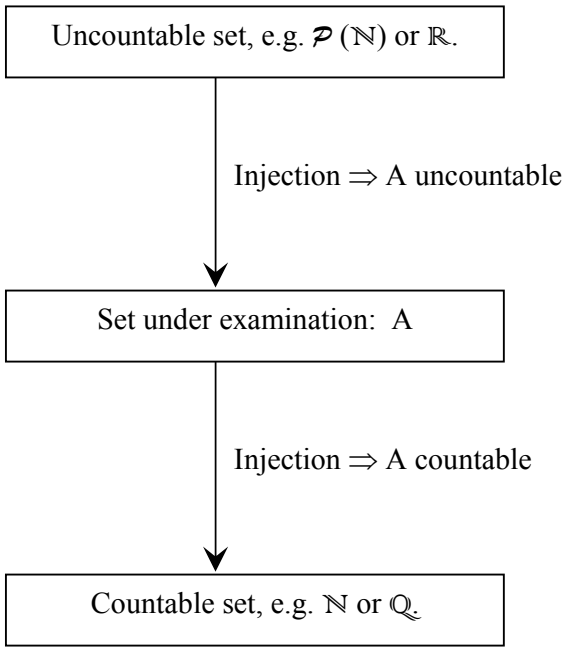
Investigating cardinality

We have now seen two ways to investigate the cardinality of a set:

- To prove that the set A is countable, we must construct an injection from A into a set that is known to be countable. For example, $\mathbb{Q}$ was shown to be countable by constructing an injection into $\mathbb{N}$.

- To prove that the set A is uncountable, we must construct an injection from a set that is known to be uncountable into A. For example, $\mathbb{R}$ was shown to be uncountable by construction an injection from $\mathcal{P}(\mathbb{N})$.

We can picture this as follows:



Algebraic and transcendental numbers

An *algebraic* number is a real number, x, that is the root of a polynomial with integer coefficients: $a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots + a_n x^n = 0$ with $a_i \in \mathbb{Z}$ and for some $n \in \mathbb{N}$.

There are only countably many such equations and each has only a finite number of roots, so there are only countably many algebraic numbers. However, there are uncountably many real numbers. Therefore there exist *transcendental* numbers which are not algebraic. Indeed, most (in some sense) numbers are transcendental. π and e (the base of natural logarithms) are both examples, but proving that they (or any other numbers) are transcendental is harder…

Exercises

1. Let $A_2 = \{1, 2\}$ and $A_3 = \{a, b, c\}$. List the elements of the four sets $A_i \to A_j$ for $i, j \in \{2, 3\}$. Annotate those elements which are injections, surjections and bijections.
2. Let B be a fixed subset of the set A. Define a relation R on the subsets of A in $\mathcal{P}(A)$ by $(X, Y) \in R \Leftrightarrow X \cap B = Y \cap B$. Show that R is an equivalence relation and describe a bijection between $\mathcal{P}(A)_{/R}$ and $\mathcal{P}(B)$.
3. Suppose that $f: A \to B$ and $g: B \to C$ are both injective. Show that their composition is also injective. Suppose instead that they are both surjective; show that their composition is surjective too. What can be deduced if f and g are both bijections?
4. If possible, find explicit bijections between the following pairs of sets. If this is not possible in general, explain why and say if any special cases can have bijections.
   - $A \times (B \times C) \leftrightarrow (A \times B) \times C$
   - $A \times A \leftrightarrow A$
   - $A \times A \leftrightarrow A + A$
   - $[(A \times B) \to C] \leftrightarrow [A \to (B \to C)]$
   - $[(A \to B) \to C] \leftrightarrow [A \to (B \to C)]$
   - $[(A + B) \to C] \leftrightarrow (A \to C) \times (B \to C)$
5. Let R and S be equivalence relations on A and B respectively with p and q the natural mappings of A and B into $A_{/R}$ and $B_{/S}$. Suppose that $f: A \to B$ is an arbitrary function. Show that the following two statements are equivalent:
   - $\exists\, g: A_{/R} \to B_{/S}$ with $p \circ g = f \circ q$.
   - $\forall\, a_1, a_2 \in A\ .\ (a_1, a_2) \in R \Rightarrow (f(a_1), f(a_2)) \in S$.
6. A function $f: A \to B$ between two partially ordered sets is *monotonic* if it respects the ordering in A and B, that is, $a_1 \leq_A a_2 \Rightarrow f(a_1) \leq_B f(a_2)$. Two partially ordered sets, A and B, are *isomorphic* if there is a monotonic bijection $f: A \to B$ whose inverse f^{-1} is also monotonic. Show that $(\mathcal{P}(\{a, b, c\}), \subseteq)$ and $(\{0, 1\}^3, \leq_P)$ are isomorphic.
7. If A and B are finite sets with $|A| = m$ and $|B| = n$, how many partial functions are there $A \to B$?
8. Show that the collection of all finite subsets of $\mathbb{N}$ is countable but that the collection of all subsets of $\mathbb{N}$ is not countable.
9. By considering indicator functions or otherwise, find a bijection from the set of functions $\{f: \mathbb{N} \to \{0, 1\}\}$ to $\mathcal{P}(\mathbb{N})$ and so deduce that the former is uncountable.

10. Which of the following sets are finite, which are countably infinite and which are uncountable?

    - $\{f\colon \mathbb{N} \to \{0, 1\} \mid \forall\, n \in \mathbb{N} \,.\, f(n) \leq f(n+1)\}$
    - $\{f\colon \mathbb{N} \to \{0, 1\} \mid \forall\, n \in \mathbb{N} \,.\, f(2n) \neq f(2n+1)\}$
    - $\{f\colon \mathbb{N} \to \{0, 1\} \mid \forall\, n \in \mathbb{N} \,.\, f(n) \neq f(n+1)\}$
    - $\{f\colon \mathbb{N} \to \mathbb{N} \mid \forall\, n \in \mathbb{N} \,.\, f(n) \leq f(n+1)\}$
    - $\{f\colon \mathbb{N} \to \mathbb{N} \mid \forall\, n \in \mathbb{N} \,.\, f(n) \geq f(n+1)\}$

11. Show that $\mathbb{Q} \times \mathbb{Q}$ is countable and deduce that any collection of disjoint discs (that is, circular areas) in the plane $\mathbb{R}^2$ is countable. Is the same true if "discs" is replaced by "circles" (that is, just the perimeters of the circles)?

Revision guide

The following diagram shows the development of the key ideas presented in the second half of the course: