Discrete Mathematics I

UNIVERSITY OF
CAMBRIDGE
Computer Laboratory

Computer Science Tripos Part IA (50%)

Peter Robinson

Michaelmas 2006

William Gates Building
15 JJ Thomson Avenue
Cambridge
CB3 0FD

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

Introduction

This course will develop the algebraic formulation of problems and formal proof by way of examples involving the integers. The material enables academic study of Computer Science and will be illustrated with examples from cryptography and the analysis of algorithms.

These notes only cover the first half of the course. A further 12 lectures will be given in the Lent Term, concentrating on sets, relations and functions.

Syllabus

- Proof. Deduction, contradiction. Integers, mathematical induction. [3 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 $\sqrt{p}$.
  [4 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. [5 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.

Appropriate books

The following books are relevant for the course:

- JA Anderson: Discrete mathematics with combinatorics, Prentice-Hall, 2001, ISBN 0-13-086998-8, £41.99.
- 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, but using different notation.
- G Pólya: *How to solve it,* Penguin, 1990, ISBN 0-14-012499-3, £8.99.
- KH Rosen: *Discrete mathematics and its applications* (5th edition), McGraw-Hill, 2002, ISBN 0-07-119881-4, £42.99.
  An excellent book covering a wide range of topics and useful throughout the whole of Computer Science.

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.

Further support material is available on-line at http://www.cl.cam.ac.uk/Teaching/current/DiscMathI/. In particular, there is an on-line help system in the form of a set of frequently asked questions which are revised as new questions are asked.

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 $(n+1)^2 - n^2$.

  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. (c) shows that a number that leaves remainder 0 when divided by 4 can be expressed as the difference of two squares, and (a) 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 $\dfrac{x}{y}$ for whole numbers x and y.

  Proof: Assume that $\sqrt{2} = \dfrac{x}{y}$ 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, $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ (rather like the type **int** in ML), and will refer to the rational numbers (fractions), $\mathbb{Q}$(which includes the type **real** in ML), and the real numbers $\mathbb{R}$ (which includes significantly more values). 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: $x \in \mathrm{X}$. 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: $\mathrm{X} \subseteq \mathrm{Y}$. 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 $\mathrm{E} = \{x \in \mathbb{Z} \mid 2|x\}$, which is a bit confusing. Sorry. The vertical bar is also used in another way to count the number of elements $|\mathrm{X}|$ 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 $\mathrm{P}(n)$ 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

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

then $\mathrm{P}(n)$ 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

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

  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
$$\left(x+y\right)^n = \binom{n}{0}x^n + \binom{n}{1}x^{n-1}y^1 + \cdots + \binom{n}{i}x^{n-i}y^i + \cdots + \binom{n}{n}y^n$$
where $\binom{n}{k}$ is the *binomial coefficient*, defined to be $\dfrac{n!}{k!(n-k)!}$ 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.

Course of values induction

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 $j < 1$.

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: Let $\mathrm{P}(n)$ be the proposition "n can be expressed as a product of primes." For any integer k, either k is prime and so $\mathrm{P}(k)$ is true, or $k = ab$ with $1 < a, b < k$. In this case, both $\mathrm{P}(a)$ and $\mathrm{P}(b)$ are true, so we can express k as the product of the expressions for a and b. Hence $\mathrm{P}(k)$ is true.

If there are no values a and b with $1 < a, b < k$, then $k = 2$ and $\mathrm{P}(k)$ is true.

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

We can now use well ordering to prove that every natural number greater than 1 can be expressed as a product of primes in a different way.

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.

This proves that n can be expressed as a product of primes, but gives us no help in showing how to do it. Factoring a large integer into primes is computationally hard, which assures us of the security of some of the codes we will be considering.

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 $3^{n+1} + 4^{2n\text{-}1}$ 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.

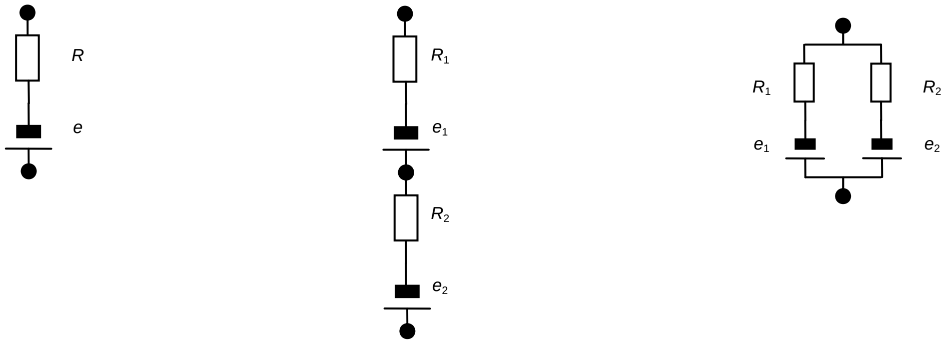
6. Prove that, for all $n \in \mathbb{N}$ and $x \in \mathbb{R}$ with $x \geq -1$, $(1 + x)^n \geq 1 + n\,x$.

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.

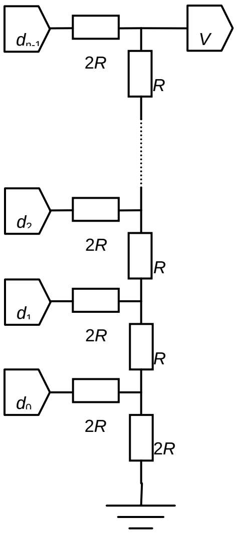
9. Thevenin's theorem states that any two-terminal network consisting of voltage sources and resistors is equivalent to a single voltage source and resistor in series. The simplest case is a single voltage source (*e*) and a single resistor (*R*). Two possible ways of connecting networks are series and parallel composition:



The series circuit is equivalent to a voltage source e = e1 + e2 and a resistance R = R1 + R2.

Show that the that the parallel circuit is equivalent to a voltage source $e = \frac{R_1 e_2 + R_2 e_1}{R_1 + R_2}$ and a resistance $R = \frac{R_1 R_2}{R_1 + R_2}$.

The following circuit shows a digital to analogue converter built from a ladder of resistors:



where *R* is an arbitrary, fixed resistance and where d_i are *n* input bits (either 0 or 1). Use induction to prove that the output voltage is given by $V = \dfrac{\sum_{i=0}^{n-1} d_i 2^i}{2^n}$.

10. Let $S = \{1, 2, \ldots, n\}$. 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. For example, with $n = 2$, $S = \{1, 2\}$ so $\sum_{s \in S} f(s) = f(1) + f(2)$ and $\prod_{s \in S} f(s) = f(1) \times f(2)$. By convention that the empty sum $\sum_{s \in \phi} f(s) = 0$ and the empty product $\prod_{s \in \phi} f(s) = 1$.

    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 $T \subseteq S$. Again, for $n = 2$, the left hand side is $(1+x_1)(1+x_2)$, and the possible values for T on the right hand side are $\emptyset$, $\{1\}$, $\{2\}$, and $\{1, 2\}$, giving corresponding products of 1, x_1, x_2, and $x_1 x_2$, so the sum is $1 + x_1 + x_2 + x_1 x_2$ (which is correct!).

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

11. Prove Pythagoras' theorem.

    This has nothing to do with Discrete Mathematics, but you ought to know a proof!

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

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?

13. [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 $b \in \mathbb{Z}$, we say that a *divides* b or a is a *factor* of b (written $a \mid b$) if $b = q\,a$ for some integer $q \in \mathbb{Z}$. 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 $b \mid a$ then $a = \pm b$.

  Proof: If $a \mid b$ and $b \mid a$ then $b = q\,a$ and $a = r\,b$ for some q and r.
  So $a = (r\,q)\,a$ and $r\,q = 1$. Now r, $q \in \mathbb{Z}$, so $r = q = \pm 1$, and $a = \pm b$.

- 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 $\mathrm{R} = \{a - bk \mid k \in \mathbb{Z}$ 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.

A further complication arises if we consider $b \in \mathbb{Z}$ rather than $b \in \mathbb{N}$. You might like to think about the best way to complete the following table:

		Quotient	Remainder
a	b	a div b	$a \bmod b$
7	3	2	1
~7	3		
7	~3		
~7	~3		

Highest common factors

Given a, $b \in \mathbb{Z}$ (not both 0), 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 $\mathrm{D} = \{as + bt \mid s, t \in \mathbb{Z}$ and $(as + bt) > 0\}$.
  If $a > 0$, then $a = a1 + b0 \in \mathrm{D}$. If $a < 0$, then $a = a(-1) + b0 \in \mathrm{D}$. If $a = 0$, then repeat the argument with b. In all cases, $\mathrm{D} \neq \emptyset$. 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 $(a, b) = ax + by$. 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.

- Let $\mathrm{L} = \{as + bt \mid s, t \in \mathbb{Z}\}$ be the set of linear combinations of a and b, and $\mathrm{M} = \{n\,(a,b) \mid n \in \mathbb{Z}\}$ be the set of multiples of their highest common factor. Then $\mathrm{L} = \mathrm{M}$.

  Proof: We need to show that $L \subseteq M$ and $M \subseteq L$.

- 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 $ab \mid n$.

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

  Proof: Use contradiction. Suppose $\left(\dfrac{a}{(a,b)}, \dfrac{b}{(a,b)}\right) = k > 1$. 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

It turns out that it is important to calculate highest common factors quickly. Simply working through the positive integers and trying all possible common factors would give an $\mathrm{O}(n)$ algorithm. We can do much better. The approach relies on the following result:

- If $a, b \in \mathbb{N}$ and $a = bq + r$ for integers q and r with $0 \le r < b$, then $(a, b) = (b, r)$.

  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 $(b, r) = d$.

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

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

Then $(a, b) = (b, r_1) = (r_1, r_2) = \dots = (r_{n-2}, r_{n-1}) = (r_{n-1}, 0) = r_{n-1}$.

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
$n =$	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.

- $s_{i-1}t_i - s_it_{i-1} = (-1)^i$ for $i \geq 0$, 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 $(s_i, t_i) = 1$.

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

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

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

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: This must be true since the set of linear combinations of two integers is equal to the set of multiples of their HCF. However, it is helpful to find the actual values of x and y.

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

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

  Now $afs + bft = (as+bt)f = (a,b)\dfrac{c}{(a,b)} = c$. So $x_0 = fs$ and $y_0 = ft$ is a solution.

- Moreover, any solution to $ax + by = c$ has $x = x_0 - \dfrac{kb}{(a,b)}$ and $y = y_0 + \dfrac{ka}{(a,b)}$ for some arbitrary $k \in \mathbb{Z}$.

  Proof: Suppose $ax_0 + by_0 = c$ and $ax + by = c$. Then $a(x_0-x) + b(y_0-y) = 0$, so $a(x_0-x) = b(y-y_0)$. Divide by (a, b), so $\dfrac{a}{(a,b)}(x_0 - x) = \dfrac{b}{(a,b)}(y - y_0)$.

  But $\left(\dfrac{a}{(a,b)}, \dfrac{b}{(a,b)}\right) = 1$, so $\dfrac{a}{(a,b)} \mid (y - y_0)$. Hence $y = y_0 + \dfrac{ka}{(a,b)}$.

  Now $x = x_0 - \dfrac{b(y-y_0)}{a} = x_0 - \dfrac{kb}{(a,b)}$.

- The general solution is just the sum of the particular solution $x = x_0$ and $y = y_0$ with the complementary function $x = -\dfrac{kb}{(a,b)}$ and $y = \dfrac{ka}{(a,b)}$ where $k \in \mathbb{Z}$ 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 $\mathrm{O}(\log a)$ 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.

Recall that $2^{10} = 1024 \approx 1000 = 10^3$. So, if a has d digits, then $a < 10^d < (2^{10})^{d/3} = 2^{10d/3}$, and the algorithm terminates in at most $20d/3$ steps, or less than $7d$ 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 $\mathrm{O}(\log a)$.

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 $pb^2 = p^2c^2$ and $p \mid b$. Hence $p \mid (a, b)$.

Digressions

- $2^{32\,582\,657} - 1$ is prime.
- The Mersenne number $\mathrm{M}_n = 2^n - 1$ is prime only when n is prime, but that is not sufficient. For example, 11 is prime, but $2^{11} - 1 = 2047 = 23 \times 89$.
- 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 $\le x$. 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 (again…). 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. 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. Complete the details of the following proof:
   - $(m(p^2-q^2), 2mpq, m(p^2+q^2))$ 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 as long as $(p,q) = 1$.

9. 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 $(m, n) = 1$.

10. Let $\mathbb{Z}[\surd 5] = \{a + b\surd 5 \mid a, b \in \mathbb{Z}\}$.

    - Show that $\alpha, \beta \in \mathbb{Z}[\surd 5] \Rightarrow \alpha + \beta, \alpha - \beta, \alpha \times \beta \in \mathbb{Z}[\surd 5]$.
    - Given $\alpha = a + b\surd 5 \in \mathbb{Z}[\surd 5]$, define the *conjugate* of α to be $\overline{\alpha} = a - b\surd 5$, 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* $\varepsilon \in \mathbb{Z}[\surd 5]$ to be an element that divides exactly into 1. Show that ε is a unit if and only if $\mathrm{N}(\varepsilon) = 1$. Find a unit that is *not* ± 1.
    - By considering residues modulo 5 (see the next section), show that there is no $\alpha \in \mathbb{Z}[\surd 5]$ with $\mathrm{N}(\alpha) = 2$.
    - Factor 4 in $\mathbb{Z}[\surd 5]$ 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 $\mathbb{Z}[\surd 5]$ to the uniqueness of prime factorisation.

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

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
  $a_1\, a_2 \equiv b_1\, b_2 \pmod{m}$.

  Mathematicians will recognise that this is a ring.

- However, $a \equiv b \pmod{m}$ does *not* imply that $c^a \equiv c^b \pmod{m}$. For example, consider $a = 1$, $b = 4$, $c = 2$, 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

×	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 $p-1 = 640 = 5\times128 = 5\times2^7$ so $5\times2^7 \equiv -1 \pmod p$. Combine these to see that $2^{32} = 2^4 \times 2^{28} \equiv (-5^4)\times(2^7)^4 = -(5\times2^7)^4 \equiv -(-1)^4 = -1 \pmod p$. 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 $(\mathrm{mod}\ mn)$.

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 $e \equiv c \pmod{m}$ and $e \equiv d \pmod{n}$. Then $e \in \mathrm{U}_{mn}$ and each such pair (c, d) is linked to a unique e.

- $\varphi(m) = m \prod_{\text{prime } p|m} (1 - {}^{1}\!/_{p})$.

  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 $(\text{mod } p)$. 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 $\mathrm{U}_m = \{x \mid 0 < x < m$ and $(x, m) = 1\}$ be the set of units modulo m. Say $\mathrm{U}_m = \{u_1, u_2, \ldots u_f\}$ 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.

This is often referred to as the Fermat-Euler Theorem, but Fermat's contribution was a special case:

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, $561 = 3\times11\times17$. Observe that $(3–1)|(561–1)$, $(11–1)|(561–1)$ and $(17–1)|(561–1)$, so $a^{561–1} \equiv 1 \pmod{3, 11}$ and $17)$ for all a with $(a, p) = 1$, and so $a^{561–1} \equiv 1 \pmod{561}$ by the Chinese Remainder Theorem.

[1] Humphreys & Prest, p 58.

[2] Humphreys & Prest, p 54.

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

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 $a_1 = a + e\ (\bmod m)$ for some encryption key, e. This is decoded by calculating $a_1 - e = (a + e) - e \equiv a\ (\bmod m)$. 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 e and f. Alice now sends a message a to Bob as follows:

- A → B: $a_1 \equiv a + e\ (\bmod m)$
- B → A: $a_2 \equiv a_1 + f = a + e + f\ (\bmod m)$
- A → B: $a_3 \equiv a_2 - e = a + f\ (\bmod m)$

Bob can now recover $a = a_3 - f\ (\bmod m)$. Unfortunately, anyone overhearing the conversation (traditionally called Eve) can recover $a \equiv a_1 - a_2 + a_3\ (\bmod m)$.

Modular multiplication is another possibility. As long as e and f 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 a or, strictly speaking, $a\ (\bmod m/(m,a))$ if $(m,a) > 1$.

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.

Given a message encoded as a natural number $a < p$, observe that p can not be a factor of a. Now use Fermat's Little Theorem: $(a^e)^d = a^{ed} = a^{1+(p-1)t} = a(a^{p-1})^t \equiv a1^t \pmod{p} = a$.

This gives a protocol:

- Alice chooses p and the value e, encodes a message a as $a_1 \equiv a^e \pmod{p}$ and sends it with p to Bob.
- Bob picks another value f with inverse g and sends $a_2 = a_1^f \equiv (a^e)^f \pmod{p}$ back to Alice.
- Alice works out $a_3 = a_2^d \equiv ((a^e)^f)^d = ((a^e)^d)^f \equiv a^f \pmod{p}$ and sends it back to Bob.
- Bob now works out $a_3^g = (a^f)^g \equiv a$ to recover the original message.

Breaking this from intercepting the intermediate messages requires *discrete logarithms*, which is as hard as factoring a large integer. Note that Alice does not know f and g, and Bob does not know e and d. However, three messages have to be transmitted to pass the single value a.

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 with only a single message transmission.

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 $ed \equiv 1 \pmod{\varphi(m)}$.

Given a message encoded as a natural number a less than both p and q, observe that neither p nor q can be a factor of a, so $(a, m) = 1$. Now use Euler's Theorem:
$(a^e)^d = a^{ed} = a^{1-\varphi(m)c} = a(a^{\varphi(m)})^{-c} \equiv a1^{-c} = a \pmod{m}$.

This gives a protocol:

- Alice picks two large primes and publishes their product m and the value e while keeping d secret.
- Bob encodes a message a as $a_1 = a^e \pmod{m}$ and sends it to Alice.
- Alice recovers a by calculating $a_1^d = (a^e)^d \equiv a \pmod{m}$.

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 $x^2 – (m + 1 – \varphi(m))x + m = 0$.

[3] Davenport, p 191.

[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.

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 $z \equiv \pm s, \pm t \pmod n$.

Observe that, if we know both s and t, it is possible to factor n. $s^2 \equiv t^2 \equiv a \pmod n$, 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 $a \equiv s^2 \pmod n$.
- 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.

[6] Giblin, p 145.

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 $d_1d_2\ldots d_{10}$ then $d_{10} = \sum_{i=1}^{9} i \cdot d_i \, (\text{mod} 11)$. 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 $\sum_{i=1}^{10} i \cdot d_i \equiv 0(\text{mod} 11)$ 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 $\mathrm{x}^5 - 3\mathrm{x}^2 + 2\mathrm{x} - 1 = 0$ has no solutions for $x \in \mathbb{Z}$.
7. Solve the following congruences:
   - $77x \equiv 11 \ (\text{mod } 40)$
   - $12y \equiv 30 \ (\text{mod } 54)$
   - $z \equiv 13 \ (\text{mod } 21)$ and $3z \equiv 2 \ (\text{mod } 17)$
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 $20!\ 21^{20}$ (mod 23).
10. Calculate $3^{1000000000}$ (mod 257).
11. Show that $42 \mid (n^7 - n)$ for all positive integers n.

[7] Humphreys & Prest, p 50.

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

    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.

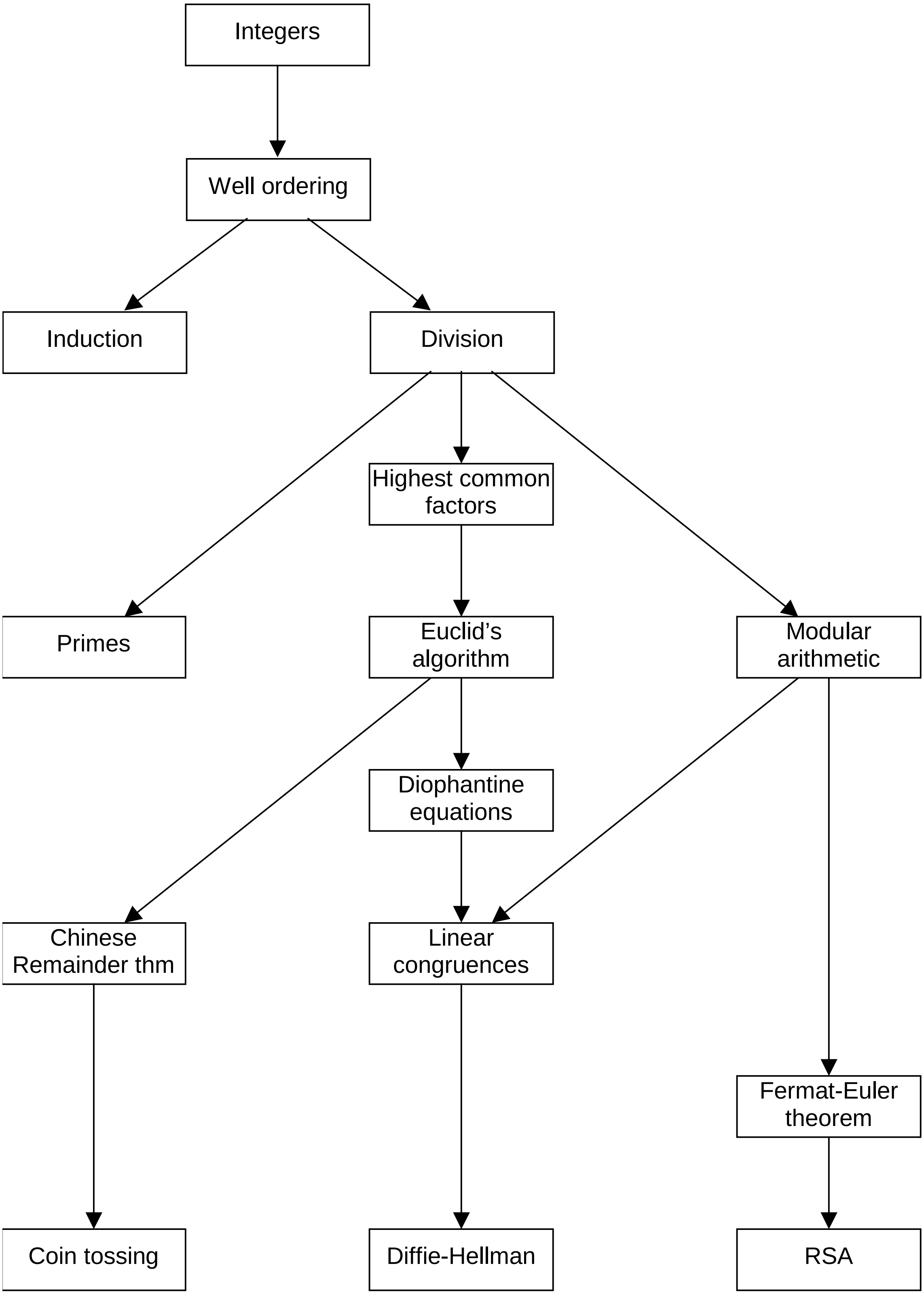
Programming

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 this part of the course:


Integers
Well ordering
Induction
Division
Highest common factors
Primes
Euclid's algorithm
Modular arithmetic
Diophantine equations
Chinese Remainder thm
Linear congruences
Fermat-Euler theorem
Coin tossing
Diffie-Hellman
RSA