Math 2301

Number Theory & Abstract & Linear Algebra

1st Semester 2010

The course is split into three parts:

- Number Theory
- Abstract Algebra: groups, rings & fields.
- Linear Algebra

The Number Theory and Abstract Algebra notes follow.

The relationship between these topics is as follows: *Number theory* studies properties of the integers $\mathbb{Z}$, and the integers mod n, $\mathbb{Z}/n\mathbb{Z}$. Both these sets have an addition operation $+$. *Group theory* studies sets with an addition (or multiplication) operation quite generally. *Rings* are sets with both addition and multiplication, and *fields* are special rings. A *vector space* over a field k consists of a group V, but also a way to multiply elements of V by elements in k, called scalar multiplication.

(The previous paragraph should make more sense at the end of the course.)

You will be expected to be able to write proofs on assignments and on the exam. None of the proofs given in this course require great ingenuity or memorization, so all proofs given could be examined. Sometimes a result is stated but not proved. In this case the word “omitted” will appear below the statement of the result. The proofs of these results will not be examined, but you should be familiar with the statement of these theorems.

The notes contain many exercises. Although you are encouraged to attempt these exercises, please note that they do not currently have solutions available, and many of these exercises are more difficult than the assignment and exam questions. Depending on demand, I may provide exercises with solutions in addition to the assignments and the worked examples in the notes.

CHAPTER 0

General Remarks

Throughout this course

log means natural log.

0.1. Sets

Recall that a *set* is a well defined collection of mathematical objects. We write $\{a_1, \ldots, a_n\}$ for the set whose members are $a_1, \ldots, a_n$. For example $\{1, 2, 3, 4\}$ is the set containing 1, 2, 3 and 4.

Any element of a set is considered only to occur once in the set. There is no notion of repetition. Thus $\{1, 2, 1, 3, 4, 2, 3\}$ is equal to $\{1, 2, 3, 4\}$.

We write $x \in S$ to indicate that x is an element of the set S, and $x \notin S$ to indicate that x is not a member of S. The symbol $\in$ is read as "is an element of" or "is a member of".

If S is finite $|S|$ denotes the number of elements in S, called the *cardinality* of S.[1]

The *natural numbers* are[2] the numbers $1, 2, 3, \ldots$ The set of natural numbers is denoted $\mathbb{N}$:

$$\mathbb{N} = \{1, 2, 3, \ldots\}.$$

The *integers* are the positive and negative whole numbers, and 0. The set of integers is denoted $\mathbb{Z}$:

$$\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}.$$

Some well known sets:

0.1.1 Example

- ★ $\mathbb{N}$, $\mathbb{Z}$.
- ★ $\mathbb{Q}$ the set of rational numbers (positive and negative fractions, and 0).
- ★ $\mathbb{R}$ the set of all real numbers.
- ★ $\mathbb{R}^+$ the set of all positive real numbers.
- ★ The set of all $n \times n$ matrices with real entries, denoted $M_n(\mathbb{R})$.

Two sets are *equal* if they have exactly the same elements. An important set is the set with no elements, denoted $\emptyset$ and called the *empty set* .

0.1.2 Definition Let A and B be sets. We write $A \subseteq B$ and say that A is a *subset* of B if every element is an element of B.

0.1.3 Example

- ★ $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$.
- ★ $A \subseteq A$ for every set A.
- ★ $\emptyset \subseteq A$ for every set A. (Because $\emptyset$ has no elements, so all of its elements occur in A, vacuously.)

---

[1]Some authors write $\#S$.

[2]Some authors also include 0.

To prove that two sets A and B are equal, we often show that $A \subseteq B$ and $B \subseteq A$. Then A and B have exactly the same elements, hence must be equal. (Compare: to prove that two real numbers a and b are equal, we sometimes show $a \leq b$ and $b \leq a$.) Such a proof is structured as follows: let $a \in A$ be arbitrary. Then (somehow) show that $a \in B$. Since a was *any* element of A, this proves $A \subseteq B$. Next let $b \in B$ and show $b \in A$. This proves $B \subseteq A$, so $A = B$.

If $P(x)$ is some property of x we use the notation

$$\{x \in S \mid P(x)\}$$

to mean the set of all x in S satisfying the condition $P(x)$. The vertical line is read "such that".[3]

If A and B are sets, the difference $A \setminus B$ is the set[4] consisting of all elements in A but not in B:

$$A \setminus B = \{a \in A \mid a \notin B\}.$$

Note that $A \setminus B$ is not the same set as $B \setminus A$.

0.1.4 Example

★ $\mathbb{R} \setminus \{0\}$. This is the set of all real numbers, except for those occurring in $\{0\}$. Thus it is the set of non-zero real numbers. We cannot write $\mathbb{R} \setminus 0$, because 0 is not a set.

★ $\mathbb{R} \setminus \mathbb{Q}$. This is the set of irrational numbers. It contains elements like $\sqrt{2}$, π.

0.2. Algorithms

At some places in the number theory notes we give algorithms for performing certain calculations. Usually these algorithms are described in English with examples. For the computationally minded, pseudo-code is given for some of these algorithms (without justification or discussion). This material is set off with the heading "Algorithm". It is not necessary to memorize this code. You simply need to be able to perform the algorithms by hand as described in the main text and in the examples.

0.3. Logic and Proofs

We briefly review some of the proof methods needed in this course. Let P and Q be any mathematical statements.

0.3.1. Or. (P or Q) means P is true, or Q is true, or both. This is different than the meaning of or in English, which is sometimes exclusive: one or the other (but not both).

0.3.2. Implication. $P \implies Q$ means: if P is true, then Q is true.

To prove $P \implies Q$ we must show that every time P is true, Q is also true. To disprove $P \implies Q$ it is enough to find a single example where P is true but Q is false.

0.3.3. If and Only If Proofs. We write

$$P \iff Q$$

if $P \implies Q$ and $Q \implies P$. This means: if P is true then Q is true, and if Q is true then P is true. We say that P holds *if and only if* Q holds. We abbreviate "if and only if" by "iff".

To prove P iff Q, it is usually necessary to split the proof into two parts: P implies Q (the "if" part) and Q implies P (the "only if") part. In these notes[5] we mark the beginning of the $P \implies Q$ proof

[3] Some authors use a colon instead of a vertical line.

[4] Some authors denote this set $A - B$.

[5] Some authors call the $\implies$ direction *sufficiency* (P true suffices to prove Q true) and the $\impliedby$ direction *necessity* (if Q is true, it is necessarily the case that P is true). For example, democracy implies elections (elections $\impliedby$ democracy), but elections do not imply democracy. So elections are necessary but not sufficient conditions for democracy.

by a forwards arrow $\implies$ and the beginning of the $Q \implies P$ proof by a backwards arrow $\impliedby$ (meaning $P \impliedby Q$).

0.3.1 Example To prove that a square matrix A is invertible iff $\det(A) \neq 0$ we must show:

$\implies$ if A is invertible then $\det(A) \neq 0$, and

$\impliedby$ if $\det(A) \neq 0$ then A is invertible.

0.3.4. Proof by contradiction. Suppose we want to prove that a statement P is true. One approach is to assume that P is false, and then show that this leads to an absurdity or contradiction (such as $1 < 0$ or $0 = 1$ etc). Since the statement "P is false" is thus untenable, P must be true.

This is an indirect method of proof, but it has the advantage that it gives us something to work with: namely the assumption that P is false.

0.3.5. Contrapositive. $P \implies Q$ means: every time P is true, Q is also true. Rephrasing: there is no instance when Q is false but P is true. That is: if (not Q) then not P. Thus $P \implies Q$ is equivalent to (not Q) $\implies$ (not P). This statement is called the *contrapositive* of $P \implies Q$.

Thus one strategy to prove $P \implies Q$ is to assume Q is false, and show that P is false.

0.3.6. Definitions. By convention, definitions in mathematics are usually stated in the form: X is a *(something)* if X satisfies (some property). This actually means: X is a *(something)* if and only if X satisfies (some property).

For example, we define x to be *even* if x is divisible by 2. This really means: x is even iff x is divisible by 2.

0.3.7. Existence and Uniqueness Proofs. Many statements in mathematics assert that there exists a unique object X with some property P. To prove such a statement usually requires two arguments:

- We must show that some such object X exists: for example by exhibiting one explicitly, or by proof by contradiction etc.
- We must prove X is unique. To do so, we assume that Y is any object with the given property P. Then we show that Y must be equal to X. Since Y was *any* object with property P, *every* object with property P must be equal to X. That is, X is the *only* such object.

0.3.8. Induction. Induction can be used to prove a statement $P(n)$ holds for all natural numbers as follows:

- Prove that $P(1)$ holds. This is the *base case* .
- Assuming $P(k)$ holds (the *inductive hypothesis*), prove that $P(k + 1)$ holds. This is the *inductive step* .

The base case shows $P(1)$ holds, then the inductive step shows $P(1) \implies P(2)$ so $P(2)$ holds, then the inductive step shows $P(2) \implies P(3)$, so $P(3)$ holds etc. Thus $P(n)$ holds for every n.

An alternate form of the inductive step is: assuming $P(1), P(2), \ldots P(k)$ *all* hold, prove that $P(k+1)$ holds. This method of proof is sometimes called *strong induction* .

Part 1

Number Theory

CHAPTER 1

Number Theory

1.1. Introduction

Number Theory is the study of properties of the integers. As such, it has been studied for millennia. To give a flavour of the subject, we state a few problems (of varying difficulty) that have been historically important.

1.1.1 Example [Pythagorean triples] The ancient Greeks were interested in finding solutions of

$$a^2 + b^2 = c^2$$

with $a, b, c \in \mathbb{N}$ (Pythagorean triples). Some solutions are:

a	b	c
3	4	5
5	12	13
$\vdots$	$\vdots$	$\vdots$
441	1960	2009
1206	1608	2010

Are there infinitely many solutions? How can we find them all?

What about more complicated equations?

1.1.2 Example [Fermat's Last Theorem] The French mathematician Fermat conjectured that the equation

$$x^n + y^n = z^n$$

has no solutions with x, y, $z \in \mathbb{N}$ once $n \geq 3$. This statement became known as *Fermat's Last Theorem* and inspired centuries of research. It was finally proved in about 1995.

Other longstanding problems in number theory concern the *prime numbers* 2, 3, 5, 7, 11, ...

1.1.3 Definition A natural number $n \geq 2$ is *prime* if the only divisors of n are 1 and n. [The number 1 is not considered to be prime.]

Primes are of great interest, because:

1.1.4 Theorem [Fundamental Theorem of Arithmetic] Every natural number can be factored into primes in a unique way.

Proof Later (see theorem 1.8.3). □

(Of course the order that factors occur in is not unique: $12 = 2^2 \cdot 3 = 3 \cdot 2^2$.)

This raises further questions such as:

- Are there infinitely many prime numbers?
- How many primes are there up to a given bound? Etc.

1.1.5 Example [Goldbach's conjecture] Goldbach conjectured that every even number $n \geq 4$ can be written as the sum of 2 primes $n = p + q$. For example, $4 = 2 + 2$, $6 = 3 + 3$, $8 = 3 + 5$, $10 = 3 + 7$, $\ldots$, $2010 = 829 + 1181$, $\ldots$

1.1.6 Example [Twin Prime Conjecture] It is conjectured that there exist infinitely many primes p such that $p + 2$ is also prime. Eg $(3, 5)$, $(5, 7)$, $(11, 13)$, $(1997, 1999)$, $(2027, 2029)$, $\ldots$

Today number theory is critical in many applications:

- Securely encrypting data. Eg credit card numbers sent over the web, medical records etc. Almost all encryption schemes rely on number theoretic ideas. See §2.12.
- Recovering corrupted digital data "error correcting". Eg on an audio CD, about 1/3 of the information is error correction information.

Why study number theory?

- It is a very old and important branch of mathematics. It has links to many areas: algebra, analysis, algebraic geometry etc.
- It contains some of the simplest to state but hardest to prove problems in mathematics. Results in number theory such as the proof of Fermat's Last Theorem represent some of humanity's highest intellectual achievements.
- It serves as a basis for generalization to abstract algebra.
- It is critical for many modern security applications.

1.2. Divisibility

Our first goal is to prove that every integer factors uniquely into primes, theorem 1.1.4. This is surprisingly(?) difficult to prove. We first need to develop some background material.

1.2.1 Definition Let a, b be integers. We say a *divides* b and write $a \mid b$ if there exists an integer c with

$$b = ac.$$

If no such c exists we write $a \nmid b$.

For example $4 \mid 12$:

$$\begin{array}{ccc} 4 & \mid & 12 \\ a & & b \end{array} \qquad 12 = 4 \cdot 3, \qquad \text{so } c = 3$$

1.2.2 Example

★ $6 \mid 12$.

★ $8 \nmid 100$.

Some people find this notation backwards, but it is standard. When writing by hand, do not confuse $\nmid$ with plus, $+$.

We state the basic properties of divisibility for natural numbers.

1.2.3 Theorem Let a, b, $c \in \mathbb{N}$. Then

(a) $1 \mid a, \quad a \mid a, \quad a \mid 0.$
(b) If $a \mid b$ then $a \leq b$.
(c) $a \mid 1$ iff $a = 1$.
(d) If $a \mid b$ and $b \mid a$ then $a = b$.
(e) If $a \mid b$ and $b \mid c$ then $a \mid c$
(f) If $a \mid c$ and $b \mid d$ then $ab \mid cd$.

(g) If $a \mid b$ and $a \mid c$ then a divides any linear combination of b and c: $a \mid (bx + cy)$ for any $x, y \in \mathbb{N}$.

Proof

(a) $a = a \cdot 1$ so $1 \mid a$ and $a \mid a$. $0 = a \cdot 0$ so $a \mid 0$.

(b) If $a \mid b$ then $b = ac$ for some $c \in \mathbb{Z}$. But $a, b > 0$ so $c > 0$ also, so $c \geq 1$. Thus $0 \leq a(c-1) = ac - a = b - a$ so $a \leq b$.

(c) $\implies$ If $a \mid 1$ then $a \leq 1$ by (b) but $a \in \mathbb{N}$ so $a \geq 1$. Hence a=1.

$\impliedby$ If $a = 1$ then $a \mid 1$ by (a).

(d) Follows from (b).

(e) If $a \mid b$ then $b = ad$ for some integer d. If $b \mid c$ then $c = be$ for some e. So $c = be = a(de)$ so $a \mid c$.

(f), (g) Exercises. □

We state an obvious result that we will often use:

1.2.4 Theorem Let a, b, $c \in \mathbb{N}$. If $ac = bc$ then $a = b$.

Proof If $a > b$ then $ac > bc$. If $b > a$ then $bc > ac$. Hence $a = b$. □

1.3. Greatest Common Divisor

Since any number dividing both a and b is $\leq a$, b there will be a largest common divisor.

1.3.1 Definition Let a, $b \in \mathbb{N}$. The *greatest common divisor (gcd)* of a and b is the largest integer d dividing both a and b. This is denoted[1] $d = \gcd(a, b)$.

1.3.2 Example

★ The positive divisors of 12 are 1, 2, 3, 4, 6, 12. The positive divisors of 42 are 1, 2, 3, 6, 7, 14, 21, 42 so

$$\gcd(12, 42) = 6.$$

★ The positive divisors of 35 are 1, 5, 7, 35. So $\gcd(35, 42) = 7$, $\gcd(12, 35) = 1$.

1.3.3 Definition Let a, $b \in \mathbb{N}$. We say that a and b are *relatively prime* or *coprime* if $\gcd(a, b) = 1$.

Note: do not confuse relatively prime with prime.

1.3.4 Example 12 and 35 are relatively prime. Note that neither is a prime number.

1 Exercise [Optional. For Computer Scientists] Write a program to calculate the probability that 2 numbers < 100 are relatively prime. What about two numbers < 1000? Make a conjecture about the probability that two randomly chosen numbers are relatively prime. Hint: the probability is not 60%.

We shall soon develop an efficient method of finding gcd's, called *Euclid's algorithm.* This is an extremely important algorithm in computational mathematics.

2 Exercise

(a) Define the gcd of n positive integers $a_1, \ldots, a_n$ to be the largest natural number dividing all the a_i. Prove that

$$\gcd(a, b, c) = \gcd(a, \gcd(b, c)).$$

[1] Some authors write $d = (a, b)$ for the gcd, but this notation may be confused with ordered pairs, vectors etc.

(b) Suppose a, b, $c \in \mathbb{N}$ with $\gcd(a,b) = \gcd(a,c) = \gcd(b,c) = 1$. Prove that $\gcd(a,b,c) = 1$.
(c) Find a, b, $c \in \mathbb{N}$ with $\gcd(a,b,c) = 1$ but $\gcd(a,b)$, $\gcd(a,c)$, $\gcd(b,c)$ all > 1.

1.4. The Division Algorithm

What happens if $b \nmid a$? In this case we cannot write $a = bc$. Instead there will be a remainder.

1.4.1 Example $3 \nmid 16$. Instead we can write $16 = 5 \cdot 3 + 1$. We say the quotient is 5 and the remainder 1.

In the previous example, $16 = 4 \cdot 3 + 4 = 3 \cdot 3 + 7 = \cdots$ so the quotient and remainder are not unique. Normally we pick the smallest non-negative remainder possible. The fact that we can always pick such a quotient and remainder is called the "Division Algorithm", although it is not really an algorithm in the modern sense of the word.

1.4.2 Theorem [Division Algorithm] Let a, $b \in \mathbb{N}$. Then there exist unique integers q, r such that

$$a = qb + r \qquad \text{with } 0 \leq r < b.$$

Proof We must prove (1) Existence and (2) Uniqueness.

(1) Existence: Consider the set (arithmetic progression)

$$\{\ldots, a - 3b, a - 2b, a - b, a, a + b, a + 2b, a + 3b, \ldots\}$$

Let r be the smallest non-negative member of this set, say

$$r = a - qb$$

for some $q \in \mathbb{N}$. Then $a = qb + r$ and $r \geq 0$ by definition of r. We have to show $r < b$. The proof is by contradiction.

Suppose $r \geq b$. Then $0 \leq r - b = (a - qb) - b = a - (q + 1)b$. But now $a - (q + 1)b$ is a non-negative member of our set and smaller than r. This contradicts the definition of r as the smallest non-negative member. This proves $r < b$.

(2) Uniqueness: Suppose $a = q_1 b + r_1 = q_2 b + r_2$ with $0 \leq r_1 < b$ and $0 \leq r_2 < b$. Then

$$q_1 b + r_1 = q_2 b + r_2.$$

If $r_1 < r_2$ then $0 < r_2 - r_1 \leq r_2 < b$ but $r_2 - r_1 = (q_1 - q_2)b$ so we have an integer multiple of b in the interval $(0, b)$. This is a contradiction.

A similar proof works if $r_2 < r_1$ (exercise). Thus we must have $r_2 = r_1$. Then $a = q_1 b + r_1 = q_2 b + r_2$ so $q_1 b = q_2 b$, but $b \neq 0$, so $q_1 = q_2$ by theorem 1.2.4. □

1.4.3 Example

- ★ Any natural number is of the form $2k$ or $2k + 1$.

  Proof Let $n \in \mathbb{N}$. Using the division algorithm, divide n by 2. The remainder must be either 0 or 1. Thus $n = 2k$ or $n = 2k + 1$.

- ★ Any integer is of the form $4k$ or $4k + 1$ or $4k + 2$ or $4k + 3$, similarly. Etc.

1.4.4 Example The square of any integer is of the form $4m$ or $4m + 1$. Eg no member of the following sequence is a square: $11, 111, 1111, 11111, \ldots$.

Proof Let $n \in \mathbb{N}$. If $n = 2k$ then $n^2 = 4k^2$ is of the form $4m$. If $n = 2k + 1$ then $n^2 = (2k + 1)^2 = 4(k^2 + k) + 1$ is of the form $4m + 1$. The second statement follows since $11 = 4 \cdot 2 + 3$, $111 = 4 \cdot 27 + 3$ and so on. In general $111 \cdots 111 = 111 \cdots 11108 + 3 = 4k + 3$. □

3 Exercise

(a) Show that $n^2 - n$ is always even, that $n^3 - n$ is always divisible by 6 and that $n^5 - n$ is always divisible by 30.
(b) Show that if n is odd then $n^2 - 1$ is divisible by 8.
(c) Show that no integers exist satisfying $x + y = 100$, $\gcd(x, y) = 3$.

1.5. The Euclidean GCD Algorithm

The naive method for finding the gcd of two integers a and b would be to list all the positive divisors[2] of a and b and pick the largest. This is extremely inefficient, because in order to find the divisors we must factorize a and b, and this is extremely time consuming.Using the fastest methods known it could take literally billions of years to factorize a 500 digit number on a supercomputer.

Luckily there is extremely efficient process for finding the gcd of two integers a and b, that does not rely on factorizing.

The heart of the algorithm is the following result:

1.5.1 Theorem If $a = qb + r$ then $\gcd(a, b) = \gcd(b, r)$.

Proof Let $d = \gcd(a, b)$, and $c = \gcd(b, r)$. Since $d \mid a$ and $d \mid b$, we have $d \mid (a - qb) = r$ by theorem 1.2.3. So $d \mid b$, $d \mid r$. Hence $d \leq \gcd(b, r) = c$.

Now $c \mid b$ and $c \mid r$. Thus $c \mid (qb + r) = a$ by theorem 1.2.3. But now $c \mid b$, $c \mid a$, so $c \leq \gcd(a, b) = d$. Hence $c = d$. □

Let a, $b \in \mathbb{N}$. Since $\gcd(a, b) = \gcd(b, a)$, we may assume $a \geq b$. Use the division algorithm repeatedly to write

$$\begin{array}{ccccccc}
a & = & q_1 b & + & r_1 & & 0 \leq r_1 < b \\
b & = & q_2 r_1 & + & r_2 & & 0 \leq r_2 < r_1 \\
r_1 & = & q_3 r_2 & + & r_3 & & 0 \leq r_3 < r_2 \\
r_2 & = & q_4 r_3 & + & r_4 & & 0 \leq r_4 < r_3 \\
\vdots & & & \vdots & & & \\
r_{n-2} & = & q_n r_{n-1} & + & \boxed{r_n} & & 0 \leq r_n < r_{n-1} \\
r_{n-1} & = & q_{n+1} r_n & + & 0 & &
\end{array}$$

The remainders slide one place left on each line.

Since $r_1 > r_2 > \cdots > r_{n-1}$, we eventually reach a step where the remainder is 0. The previous theorem implies

$$\gcd(a, b) = \gcd(b, r_1) = \gcd(r_1, r_2) = \cdots = \gcd(r_{n-1}, r_n) = r_n$$

since $r_n \mid r_{n-1}$.

The gcd is the last non-zero remainder obtained.

[2] Assuming results to come about unique factorization.

1.5.2 Algorithm [Euclid's Algorithm, Simple Form] For integers a, b with $a \geq b \geq 0$ and $a > 0$ this algorithm returns $\gcd(a, b)$

(a) While $b > 0$ do $(a, b) = (b, a \pmod{b})$
(b) Return a

1.5.3 Example Find $\gcd(98, 36)$ by Euclid's algorithm.

$$
\begin{array}{lcllc}
98 & = & 2 \cdot 36 & + & 26 \\
36 & = & 1 \cdot 26 & + & 10 \\
26 & = & 2 \cdot 10 & + & 6 \\
10 & = & 1 \cdot 6 & + & 4 \\
6 & = & 1 \cdot 4 & + & \boxed{2} \\
4 & = & 2 \cdot 2 & + & 0
\end{array}
$$

So $\gcd(98, 36) = 2$.

1.5.4 Example Find $\gcd(42823, 6409)$.

$$
\begin{array}{lcllc}
42823 & = & 6 \cdot 6409 & + & 4369 \\
6409 & = & 1 \cdot 4369 & + & 2040 \\
4369 & = & 2 \cdot 2040 & + & 289 \\
2040 & = & 7 \cdot 289 & + & \boxed{17} \\
289 & = & 17 \cdot 17 & + & 0
\end{array}
$$

So $\gcd(42823, 6409) = 17$.

We can set this out more compactly. There is no need to write each remainder 3 times. Below we just write each r once. We can also think of a and b as the first two remainders, say $r_0 = b$ and $r_{-1} = a$.

$$
\begin{array}{ll}
q & r \\
 & 42823 \\
6 & 6409 \\
1 & 4369 \\
2 & 2040 \\
7 & 289 \\
17 & \boxed{17} \\
 & 0
\end{array}
$$

So $\gcd(42823, 6409) = 17$.

4 Exercise Find $\gcd(1596725, 113256)$ by Euclid's algorithm.

The algorithm is fast, because the size of the remainders decreases very rapidly. It can be shown that there exists a constant $c > 0$ such that if $1 \leq a, b \leq N$ the Euclidean algorithm will always terminate in $\leq c \log N$ steps. In fact the average number of steps is $\simeq 0.85 \log N$. See [2, Theorem 2.1.3]. The total number of bit operations to perform the algorithm is bounded by $c_1(\log N)^2$ for some constant c_1. As a rough rule of thumb: it does not take much longer to calculate $\gcd(a, b)$ than it does to calculate the product ab.

1.6. The Extended GCD Algorithm

We can squeeze a bit more from Euclid's algorithm. With a little more calculation we can write the gcd as an integer combination of the original numbers a and b:

$$\gcd(a,b) = ax + by, \qquad \text{for some integers } x,\ y.$$

In fact, we write each of the remainders r_i in turn as a linear combination of a and b:

$$r_i = ax_i + by_i, \qquad \text{for some integers } x_i,\ y_i.$$

Since r_n is the gcd, the desired x and y are x_n and y_n.

The running time of this algorithm is only a constant multiple of the simple gcd algorithm.

1.6.1 Example Write the gcd of 42823 and 6409 in the form $42823x + 6409y$ for integers x, y.

Solution: Recall Example 1.5.4:

$$\begin{array}{lcccl} 42823 & = & 6\cdot 6409 & + & 4369 \\ 6409 & = & 1\cdot 4369 & + & 2040 \\ 4369 & = & 2\cdot 2040 & + & 289 \\ 2040 & = & 7\cdot 289 & + & \boxed{17} \\ 289 & = & 17\cdot 17 & + & 0 \end{array}$$

Let the quotients be labelled $q_0 = 6$, $q_1 = 1$, $q_2 = 2$ etc and the remainders be $r_1 = 4369$, $r_2 = 2040$ etc. It is convenient[3] to let $r_0 = 6409$ and $r_{-1} = 42823$. For $i = -1, 0, 1, \ldots$ we now solve

$$r_i = 42823x_i + 6409y_i, \qquad i \geq -1.$$

The cases $i = -1$ and $i = 0$ are easy:

$$42823 = 42823\cdot 1 + 6409\cdot 0 \qquad x_{-1} = 1, \quad y_{-1} = 0 \tag{1.6.1}$$

$$6409 = 42823\cdot 0 + 6409\cdot 1, \qquad x_0 = 0, \quad y_0 = 1 \tag{1.6.2}$$

Now multiply equation (1.6.2) by $q_0 = 6$ and subtract from equation (1.6.1). We use the gcd calculation above to rewrite the lhs:

$$4369 = 42823\cdot 1 + 6409\cdot(-6), \qquad x_1 = 1, \quad y_1 = -6 \tag{1.6.3}$$

We keep repeating this process. We multiply equation (1.6.3) by $q_1 = 1$ and subtract from equation (1.6.2):

$$2040 = 42823\cdot(-1) + 6409\cdot 7, \qquad x_2 = -1, \quad y_2 = 7.$$

Multiply by $q_2 = 2$ and subtract:

$$289 = 42823\cdot 3 + 6409\cdot(-20), \qquad x_3 = 3, \quad y_3 = -20.$$

Finally multiply by $q_3 = 7$ and subtract:

$$17 = 42823\cdot(-22) + 6409\cdot 147, \qquad x_3 = -22, \quad y_3 = 147.$$

Thus the required x and y are $x = -22$, $y = 147$.

1.6.2 Example Find integers x and y such that $267x + 118y = 1$.

Solution:

[3] It may have been neater to start labelling at $i = 0$ instead of $i = -1$ but these subscripts are conventional.

$$\begin{aligned} 267 &= 2 \cdot 118 + 31 \\ 118 &= 3 \cdot 31 + 25 \\ 31 &= 1 \cdot 25 + 6 \\ 25 &= 4 \cdot 6 + \boxed{1} \\ 6 &= 6 \cdot 1 + 0 \end{aligned}$$

So

$$\begin{array}{rcccccccl} 267 &=& 1 & \cdot & 267 & + & 0 & \cdot & 118 & \\ 118 &=& 0 & \cdot & 267 & + & 1 & \cdot & 118, & \text{Multiply by -2} \\ 31 &=& 1 & \cdot & 267 & + & (-2) & \cdot & 118, & \text{Multiply by -3} \\ 25 &=& (-3) & \cdot & 267 & + & 7 & \cdot & 118, & \text{Multiply by -1} \\ 6 &=& 4 & \cdot & 267 & + & (-9) & \cdot & 118, & \text{Multiply by -4} \\ 1 &=& (-19) & \cdot & 267 & + & 43 & \cdot & 118 & \end{array}$$

Hence $x = -19$, $y = 43$ is a solution.

Let us write down systematically the calculations in the general case. Consecutive lines of the gcd calculation look like:

$$r_{i-2} = q_{i-1} \cdot r_{i-1} + r_i = ax_{i-2} + by_{i-2} \tag{1.6.4}$$

$$r_{i-1} = q_i \cdot r_i + r_{i+1} = ax_{i-1} + by_{i-1} \tag{1.6.5}$$

Multiplying equation (1.6.5) by q_{i-1} and subtracting from the equation (1.6.4):

$$\begin{aligned} r_i &\overset{1.6.4}{=} r_{i-2} - q_{i-1} r_{i-1} \\ &= (ax_{i-2} + by_{i-2}) - q_{i-1}(ax_{i-1} + by_{i-1}) \\ &= a(x_{i-2} - q_{i-1}x_{i-1}) + b(y_{i-2} - q_{i-1}y_{i-1}) \\ &= ax_i + by_i. \end{aligned}$$

So

$$\left\{ \begin{aligned} r_i &= r_{i-2} - q_{i-1} r_{i-1} \\ x_i &= x_{i-2} - q_{i-1} x_{i-1} \\ y_i &= y_{i-2} - q_{i-1} y_{i-1} \end{aligned} \right. \tag{1.6.6}$$

These are all *recurrence relations*, where the new term is obtained from the two previous ones. In fact they are exactly the same recurrence relations, just with different starting conditions:

$$\begin{aligned} r_{-1} &= a, & r_0 &= b \\ x_{-1} &= 1, & x_0 &= 0 \\ y_{-1} &= 0, & y_0 &= 1 \end{aligned}$$

We can set out the computation in a compact form. We write the q_i on each line, and obtain r_i, x_i and y_i from the previous two lines using equations (1.6.6).

1.6.3 Example Find integers x and y with $42823x + 6409y = 17$.

Solution: This is Example 1.5.4 again. We solve it this time using the recurrence relations. Set out the r_i, x_i and y_i in rows. To obtain each new row, take the second last row and subtract q times the last row from it. This is just like a row reduction step in linear algebra.

This is the initial set up:

i	q_i	r_i	x_i	y_i
-1		42823	1	0
0	6	6409	0	1

The only calculation we have done is to find q_0, which is the greatest integer less than or equal to $42823/6409 \simeq 6.68$. Now we keep doing "row reduction" type steps. Subtract 6 times the last row ($i = 0$) from the previous row ($i = -1$):

i	q_i	r_i	x_i	y_i
-1		42823	1	0
0	6	6409	0	1
1	1	4369	1	-6

We also found q_1 which is the integer part of $6409/4369 \simeq 1.47$, so $q_1 = 1$. So next subtract 1 times the last row ($i = 1$) from the previous row ($i = 0$):

i	q_i	r_i	x_i	y_i
-1		42823	1	0
0	6	6409	0	1
1	1	4369	1	-6
2	2	2040	-1	7

Repeat until $r_i = 0$. The required x and y are the entries from the last row with non-zero r_i value:

i	q_i	r_i	x_i	y_i
-1		42823	1	0
0	6	6409	0	1
1	1	4369	1	-6
2	2	2040	-1	7
3	7	289	3	-20
4	17	17	-22	147
		0		

Thus $x = -22$, $y = 147$.

If only the gcd is required, the x and y columns can be omitted.

1.6.4 Algorithm [Euclid's Algorithm, Extended Form] Let $a \geq b \geq 0$ with $a > 0$. The algorithm returns an integer triple (x, y, g) where $g = \gcd(a, b)$ and $ax + by = g$.

(a) $(x, y, g, u, v, r) = (1, 0, a, 0, 1, b)$
(b) While $r > 0$ do
(c) Set q equal to the largest integer $\leq g/r$
(d) Set $(x, y, g, u, v, r) = (u, v, r, x - qu, y - qv, g - qr)$
(e) Return (x, y, g)

1.6.5 Theorem Let a, b, $c \in \mathbb{N}$. The equation

$$ax + by = c$$

has a solution in integers x, y iff c is a multiple of the gcd of a and b.

Proof Let $d = \gcd(a, b)$.

$\Longrightarrow$ Suppose $ax + by = c$. Then $d \mid a$, $d \mid b$ so $d \mid (ax + by) = c$, by theorem 1.2.3.

$\Longleftarrow$ Suppose $d \mid c$. Let $c = de$. Using the Extended Euclidean algorithm we can find integers X and Y such that $aX + bY = d$, as discussed above. Multiplying through by e,

$$a(Xe) + b(Ye) = de = c$$

so take $x = Xe$, $y = Ye$. □

5 Exercise In each case, evaluate $\gcd(a, b)$ and write this as a linear combination of a and b.

(a) $a = 503$, $b = 238$.

(b) $a = 13487$, $b = 8747$.
(c) $a = 5784025$, $b = 146927$.

1.7. Consequences of Euclid's Algorithm

Theorem 1.6.5 has several very useful consequences. We shall need these results to prove the Fundamental Theorem of Arithmetic, theorem 1.1.4.

1.7.1 Theorem If $\gcd(a, b) = 1$ and $a \mid bc$ then $a \mid c$.

Proof By theorem 1.6.5 there exist integers x and y with $ax + by = 1$. Then $acx + bcy = c$. Now $a \mid a$ and $a \mid bc$ so a divides the lhs by theorem 1.2.3. Thus $a \mid c$. □

1.7.2 Corollary [Euclid] Let p be a prime number. If $p \mid bc$ then $p \mid b$ or $p \mid c$.

Proof Suppose $p \nmid b$. We must show $p \mid c$. The only divisors of p are 1 and p, and $p \nmid b$, so $\gcd(p, b) = 1$. The result follows on letting $a = p$ in the previous theorem. □

This result is false if p is not prime.

1.7.3 Example $6 \mid (4 \cdot 9) = 36$ but $6 \nmid 4$, $6 \nmid 9$. The problem is that $\gcd(6, 4) = 2$, $\gcd(6, 9) = 3$.

1.7.4 Corollary If p is prime and $p \mid a_1 a_2 \cdots a_n$ then $p \mid a_i$ for some i.

Proof Induction on n, using Corollary 1.7.2. □

1.7.5 Corollary Suppose $m_1 \mid a$ and $m_2 \mid a$ with $\gcd(m_1, m_2) = 1$. Then the product $m_1 \cdots m_n$ divides a.

Proof Since $m_1 \mid a$ we may write $a = m_1 n_1$. Then $m_2 \mid m_1 n_1$ and $\gcd(m_2, m_1) = 1$, so $m_2 \mid n_1$ by theorem 1.7.1. Thus $m_1 m_2 \mid m_1 n_1 = a$. □

By induction we obtain:

1.7.6 Corollary Suppose $m_i \mid a$ with $1 \le i \le n$, and suppose the m_i are pairwise relatively prime. Then the product $m_1 \cdots m_n$ divides a.

6 Exercise Let $f = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0$ with $a_i \in \mathbb{Z}$. Suppose there exists a rational number x_0 with $f(x_0) = 0$. Show that x_0 must be an integer. Conclude that $\sqrt[n]{2}$ is irrational for every $n \ge 2$.

Finally, we give another characterization of the gcd. Recall our definition:

Let a, b, $d \in \mathbb{N}$. Then $d = \gcd(a, b)$ iff:

(1) $d \mid a$, $d \mid b$ (d is a common divisor)
(2) If e is any natural number such that $e \mid a$, $e \mid b$ then $e \le d$.

There is a slightly different characterization:

1.7.7 Theorem Let a, b, d. Then $d = \gcd(a, b)$ iff

(1) $d \mid a$, $d \mid b$ (d is a common divisor)
(2′) If e is any natural number such that $e \mid a$, $e \mid b$ then $e \mid d$.

Proof $\Longrightarrow$ Suppose $d = \gcd(a, b)$. Then (1), (2) above hold by definition. We have to show (2′). Use theorem 1.6.5 to write $d = ax + by$. Now if $e \mid a$, $e \mid b$ then $e \mid (ax + by)$, so $e \mid d$.

$\Longleftarrow$ Suppose (1) and (2′) hold. We must show (2) holds. Let $e \mid a$, $e \mid b$. Then $e \mid d$ by (2′) so $e \le d$ by theorem 1.2.3. □

Most authors *define* the gcd by (1) and (2′), because this definition generalizes to other systems.

7 Exercise What happens if we define the gcd of two integers (instead of natural numbers) using (1) and (2′)? Explain why the gcd is no longer unique. Is this a problem?

What happens if we define the gcd of two polynomials with rational coefficients using (1) and (2′)?

8 Exercise Recall that a regular polygon is a polygon in which all the sides are the same length and all the angles are the same (equilateral triangle, square etc).

Describe all regular polygons which may be fitted around a common vertex. For example, 4 squares, or 3 hexagons.

Hint: The interior angle of a regular n-gon is $(n-2)\pi/n$. What equation arises if you try to fit m different n-gons around a single point? Now split into two cases: n even or odd.

1.8. The Fundamental Theorem of Arithmetic

We now prove the so-called Fundamental Theorem of Arithmetic: that every natural numbers factors into primes in a unique way. That is, we may write a natural number n uniquely as a product:

$$n = p_1^{a_1} \cdots p_s^{a_s}$$

where the p_i are distinct primes, and $a_i \geq 1$. (Of course the order that we write the factors is not uniquely determined, but we often write the primes in increasing order.)

1.8.1 Example

- ★ $2007 = 3^2 \cdot 223$.
- ★ $2008 = 2^3 \cdot 251$.
- ★ $2009 = 7^2 \cdot 41$.
- ★ $2010 = 2 \cdot 3 \cdot 5 \cdot 67$.
- ★ $2011 = 2011$.

Unique factorization is far from obvious.

1.8.2 Example Let $\mathbb{E}$ denote the set of even natural numbers

$$\mathbb{E} = \{2, 4, 6, 8, \ldots\}$$

Let us say $a \mid_{\mathbb{E}} b$ if there exists $c \in \mathbb{E}$ with $b = ac$. For example $4 \mid_{\mathbb{E}} 8$ since $8 = 4 \cdot 2$. But $2 \nmid_{\mathbb{E}} 6$ since $6 \neq 2c$ for any $c \in \mathbb{E}$.

Now 2, 6, 18 are "$\mathbb{E}$-primes" (have no proper factors in $\mathbb{E}$). And

$$36 = 2 \cdot 18 = 6 \cdot 6$$

so 36 has different factorizations into "$\mathbb{E}$-primes".

There are also counterexamples in systems such as $\{a + b\sqrt{-5} \mid a, b \in \mathbb{Z}\}$—see Math 4304. (This is the reason Fermat's Last Theorem is so hard to prove.)

1.8.3 Theorem [Fundamental Theorem of Arithmetic] Every natural number factors into primes in a unique way.

Proof We must prove (1) Existence and (2) Uniqueness.

(1) We must show that every natural number can be written as a product of primes.[4] If this is not the case, there is a smallest counterexample N that cannot be written as such a product. Then N cannot be prime (or it would be a product of one prime), so it must have some proper divisor d,

[4] The integer 1 is considered to be a product of zero primes.

$d \neq 1$, N. Let $N = dM$. Then d, $M < N$ so by minimality of N, both d and M can be written as a product of primes. But then so can $N = dM$.

(2) Suppose

$$n = p_1 p_2 \cdots p_r = q_1 q_2 \cdots q_s$$

where p_i, q_j are primes. Without loss of generality we may assume $r \leq s$. Now $p_1 \mid q_1 q_2 \cdots q_s$, so by Corollary 1.7.4, $p_1 \mid q_j$ for some j. Since q_j is prime, $p_1 = q_j$. We may relabel the indices and thus assume $j = 1$ so $p_1 = q_1$. By theorem 1.2.4, we may then cancel the common p_1. This leaves

$$p_2 \cdots p_r = q_2 \cdots q_s.$$

Now repeat with p_2. If $r < s$ we eventually get

$$1 = q_{r+1} \cdots q_s$$

which is impossible (theorem 1.2.3). Thus $r = s$, and (after relabelling indices), $p_1 = q_1$, $p_2 = q_2, \ldots, p_r = q_r$. □

1.8.4 Theorem If $\gcd(a, b_1) = \cdots = \gcd(a, b_n) = 1$ then $\gcd(a, b_1 \cdots b_n) = 1$.

9 Exercise Prove theorem 1.8.4.

10 Exercise Prove that n is a square iff the exponent of every prime occurring in the factorization of n is even.

11 Exercise If $n = p_1^{a_1} \cdots p_k^{a_k}$ and $m = p_1^{b_1} \cdots p_k^{b_k}$ where we allow a_i, $b_i \geq 0$ (so that the primes occurring are the same in a and b), what is the prime factorization of $\gcd(m, n)$? Prove your claim.

This gives an incredibly *inefficient* method of finding gcd's.

12 Exercise Prove that if $\gcd(a, b) = 1$ then $\gcd(a^n, b^m) = 1$ for any m, $n \in \mathbb{N}$.

1.9. Distribution of Primes

Another famous result of Euclid's is the following. This proof is more than 2300 years old.

1.9.1 Theorem There are infinitely many prime numbers.

Proof The proof is by contradiction. Suppose there are only finitely many primes. Let the complete list be $p_1, p_2, \ldots, p_n$. Let $N = p_1 p_2 \cdots p_n + 1$. According to the Fundamental Theorem of Arithmetic, N must be divisible by some prime. This must be one of the primes in our list. Say $p_k \mid N$. But $p_k \mid p_1 \cdots p_n$, so $p_k \mid (N - p_1 \cdots p_n) = 1$ which is absurd (theorem 1.2.3). □

Note that we do not know that the N constructed is prime. We only prove it has *some* prime factor not in the list p_1, p_2, ..., p_n. For example, given primes 2, 3, 5, 7, 11, 13 we form $N = 2 \cdot 3 \cdots 13 + 1$. This is not prime, but factors as $N = 59 \cdot 509$. So we have discovered a new prime 59 (and also 509) not in our original list.

13 Exercise Show that there are arbitrarily large gaps between consecutive prime numbers. Hint: For $n \geq 2$ consider the numbers $n! + 2$, $n! + 3$, ...

Understanding exactly how the primes numbers are distributed is one of the great challenges in number theory. We introduce a counting function.

1.9.2 Definition Let $\pi(x)$ be the number of primes $\leq x$.

Here is a table of $\pi(x)$ (here log denotes natural log):

x	$\pi(x)$	$\pi(x)/x$	$x/\pi(x)$	$x/\log(x)$	$\frac{\pi(x)}{x/\log(x)}$
10^3	168	0.168	6.0	145	1.159
10^4	1229	0.123	8.1	1086	1.132
10^5	9592	0.0959	10.4	8686	1.104
10^6	78498	0.0785	12.7	72382	1.084
10^7	664579	0.0665	15.0	620420	1.071
10^8	5761455	0.0576	17.4	5428661	1.061

Here $\pi(x)/x$ represents the probability that a number up to x is prime, and $x/\pi(x)$ is the reciprocal. So $1000/\pi(1000) \simeq 6.0$ means there is about a 1/6 chance that a number up to 1000 is prime. There is only about a 1/8.1 chance that a number up to 10000 is prime etc.

Notice that as x increases by a factor of 10, $x/\pi(x)$ seems to increase by a difference of about 2.3. Then recall $\log 10 \simeq 2.3025$. So if we let $f(x) = x/\pi(x)$ then we have observed that

$$f(10x) \simeq f(x) + \log(10)$$

Note that the log function itself has this property: $\log(10x) = \log(x) + \log(10)$. So perhaps $f(x) \simeq \log(x)$ ie perhaps $\pi(x) \simeq x/\log(x)$. The fifth column gives $x/\log(x)$, which is roughly the same size as $\pi(x)$ as you can see.

This is all hand-waving, but it actually gives the correct insight:

1.9.3 Theorem [Prime Number Theorem, Hadamard, de la Vallée Poussin 1896]

$$\pi(x) \sim x/\log(x).$$

Here log is natural log and the $\sim$ means that the ratio of the lhs to the rhs has limit 1 as $x \to \infty$. Better approximations are known, but the problem remains of great interest.

14 Exercise Using the Prime Number Theorem, calculate the probability that a randomly chosen odd number with 100 decimal digits is prime. (You may assume that the Prime Number Theorem gives a very accurate estimate for numbers of this size). How many random choices of numbers of this size would be required until there was better than 99% chance that one of them would be prime?