2 The Euclidean algorithm

Do you understand the number 5? 6? 7? At some point our level of comfort with individual numbers goes down as the numbers get large. For some it may be at 43, for others, 4. In any case, the most basic way of understanding a natural number is knowing its prime factorization.

Ironically, there is no "easy" way to determine if a number is prime or not, or what its prime factors are. For instance, say you had a random 5 digit number: 72193. There is no simple way to determine if it is prime or not, or what its divisors (other than 1 obviously) are, short of trying to divide it by all possible smaller numbers: 2, 3, 4, 5, 6, etc. (Of course it suffices to try to divide it by all primes smaller than it, but to put this in to practice, you already need to know which smaller numbers are primes and which are not.)

On the other hand, it turns out to be easy to determine the gcd (greatest common divisor) of two numbers. This may seem somewhat paradoxical at first, but having two numbers instead of just one lets you compare them. In fact, if we think about what it should mean to understand something (a number), one good interpretation is to understand how it is similar and how it is different from other things (other numbers). And, from a multiplicative point of view, the gcd tells us precisely what two numbers have in common. Furthermore, the gcd will help us with the problems of understanding primes and prime factorization.

2.1 The gcd by subtraction

Let $a, b, d \in \mathbb{N}$.

First note that if d is a *common divisor* of a and b, i.e.,

$$a = a'd, \;\; b = b'd,$$

for some $a', b' \in \mathbb{N}$, then

$$a - b = a'd - b'd = (a' - b')d$$

so d is a divisor of $a - b$. Similarly, if d is a common divisor of $a - b$ and b, then it is also a divisor of $a = (a - b) - b$. Hence the common divisors of a and b are the same as the common divisors of $a - b$ and b. In particular,

$$\gcd(a, b) = \gcd(b, a - b)$$

Euclid used this idea to make an efficient algorithm to determine $\gcd(a, b)$.

The Euclidean algorithm goes as follows. Set

$$a_1 = \max\{a, b\}, \;\; b_1 = \min\{a, b\}.$$

Then we inductively compute

$$a_{i+1} = \max\{b_i, a_i - b_i\}, \;\; b_{i+1} = \min\{b_i, a_i - b_i\},$$

stopping only when we have

$$a_k = b_k.$$

(This procedure produces smaller and smaller pairs of natural numbers so must eventually terminate by descent. The max/min business is to ensure we always have $a_i \geq b_i$ so that the $a_i - b_i$ appearing in the next step is positive.)

Example. *Do* $a = 15$, $b = 6$.

The reason this works is as follows. Since $\gcd(a, b) = \gcd(b, a - b)$, we have

$$\gcd(a, b) = \gcd(a_1, b_1) = \gcd(a_2, b_2) = \cdots = \gcd(a_k, b_k) = \gcd(a_k, a_k) = a_k.$$

Example. *Do* $a = 18$, $b = 5$.

If $\gcd(a, b) = 1$, we say a and b are *relatively prime.*

Exercise 2.1. *Exercises 2.1.1, 2.1.3, 2.1.5.*

Exercise 2.2. *Compute* $\gcd(84, 63)$ *using the above method. Write out each step.*

2.2 The gcd by division with remainder

A more efficient version of the Euclidean algorithm is as follows. Set

$$a_1 = \max\{a, b\}, \quad b_1 = \min\{a, b\},$$

$$a_{i+1} = b_i, \quad b_{i+1} = \text{remainder in } a_i/b_i,$$

halting when we have a pair

$$a_k, \quad b_k \text{ with } b_k | a_k.$$

Then

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

This algorithm is essentially the same as the subtraction version, but the division can do several steps of subtraction at once.

Example. *Do* $a = 18$, $b = 5$.

Write a and b in binary. Suppose $a > b$ and a is n bits (binary digits) long. Then the remainder in a/b has at most $n - 1$ bits, so this algorithm will terminate at most n steps. In other words, if $\max a, b < 2^{n+1}$, then we can determine $\gcd(a, b)$ in at most n steps (called *logarithmic time.*) This is as efficient as one could hope for. A computer could handle numbers thousands of digits long in a fraction of a second. On the other hand, even with very advanced algorithms, a modern computer might take up to a year to factor a 200 digit number.

Another advantage of the division version is it can deal with other number systems. E.g., if you want to compute $\gcd(17, 4 + i)$ in $\mathbb{Z}[i]$, you can divide 17 by $4 + i$ (and get $4 - i$ exactly), but subtraction gives you nothing.

Exercise 2.3. *Compute* $\gcd(42, 8)$ *using the division method. Write out each step.*

2.3 Linear representation of the gcd

If we go back to the subtraction version of the Euclidean algorithm, it is clear that at each step a_i and b_i are (integral) linear combinations of a and b. Hence

$$\gcd(a, b) = a_k = ma + nb \tag{1}$$

for some $m, n \in \mathbb{Z}$.

Often it will be helpful to determine not just $\gcd(a, b)$ but also the above m and n. This can be done through a variety of equivalent methods, sometimes called the *extended Euclidean algorithm*. We will present the *tableau method*, which is more efficient than the one in the text.

Consider the example from the text: $a = 34$, $b = 19$. The idea is to use a little linear algebra, and is similar to matrix row reduction, but we build a table, starting with the following two rows. For clarification I will write the underlying equation on the right, though in practice you will omit this.

m	n	x	$\longleftrightarrow$	$ma + nb = x$
1	0	34		$1 \cdot a + 0 \cdot a = 34$
0	1	19		$0 \cdot a + 1 \cdot b = 19$

The entries running down the x column will just be the successive numbers $a_1, b_1, b_2, \ldots, b_k$ from the division algorithm. The m and n entries for the b_i row will just be the coefficients needed for $ma + nb = b_i$. For example, here $b_2 = a_1 - b_1$, so the next row will just be obtained by subtracting the second from the first (do this to each column) to get

m	n	x	$\longleftrightarrow$	$ma + nb = x$
1	0	34		$1 \cdot a + 0 \cdot a = 34$
0	1	19		$0 \cdot a + 1 \cdot b = 19$
1	1−	15		$1 \cdot a - 1 \cdot b = 15$

We do this again to get

m	n	x	$\longleftrightarrow$	$ma + nb = x$
1	0	34		$1 \cdot a + 0 \cdot a = 34$
0	1	19		$0 \cdot a + 1 \cdot b = 19$
1	−1	15		$1 \cdot a - 1 \cdot b = 15$
−1	2	4		$-1 \cdot a + 2 \cdot b = 4$

Now 4 goes into 15 3 times, so we should subtract 3 times the last row from the previous row to get

m	n	x	$\longleftrightarrow$	$ma + nb = x$
1	0	34		$1 \cdot a + 0 \cdot a = 34$
0	1	19		$0 \cdot a + 1 \cdot b = 19$
1	−1	15		$1 \cdot a - 1 \cdot b = 15$
−1	2	4		$-1 \cdot a + 2 \cdot b = 4$
4	−7	3		$4 \cdot a - 7 \cdot b = 3$

With one more step we are done:

$$\begin{array}{cccc} m & n & x & \longleftrightarrow & ma+nb=x \\ 1 & 0 & 34 & & 1\cdot a+0\cdot a=34 \\ 0 & 1 & 19 & & 0\cdot a+1\cdot b=19 \\ 1 & -1 & 15 & & 1\cdot a-1\cdot b=15 \\ -1 & 2 & 4 & & -1\cdot a+2\cdot b=4 \\ 4 & -7 & 3 & & 4\cdot a-7\cdot b=3 \\ -5 & 9 & 1 & & -5\cdot a+9\cdot b=1 \end{array}$$

We know we are done now because the last b_j ($x=1$) divides the previous b_j ($x=3$). Hence the tableau method has shown two things:

$$\gcd(a,b)=1$$

and

$$\gcd(a,b)=-5a+9b.$$

Exercise 2.4. *Exercise 2.3.2. (You may use either the tableau method, the method in the text, or any other equivalent method you like. Just write down each step and explain your method if it is not one of two mentioned above.)*

2.4 Primes and factorization

Proposition 2.1. *Let $n\in\mathbb{N}$, $n>1$. Then has a prime factorization*

$$n=p_1p_2\cdots p_k$$

where each p_i is prime. (Here the p_i's are not necessarily distinct.)

Proof. By definition, n is either prime or $n=ab$ for some $a,b>1$. If a and b are prime, we are done. If not, then we repeat this with a and b, until we have reduced all factors to products of primes. This process terminates by descent, and we are done. □

Note the *existence* of a prime factorization relies only on the definition of prime and Fermat's descent, and not any any arithmetic of $\mathbb{N}$. Hence this will hold for any number system in which descent holds (such as $\mathbb{N}$ or $\mathbb{Z}$, but not $\mathbb{Q}$ or $\mathbb{R}$). Contrast that to the following.

Proposition 2.2. (Prime divisor property) *Let p be prime. If $p|ab$ then $p|a$ or $p|b$.*

Proof. Suppose $p\nmid a$. Then $\gcd(a,p)=1$ since p is prime. From Section 2.3, we have

$$1=ma+np$$

for some $m,n\in\mathbb{Z}$. Hence

$$b=mab+npb.$$

Since $p|mab$ (by assumption) and $p|npb$ (trivially), we have $p|b$, which proves the proposition. □

This seemingly evident fact relies on the arithmetic of $\mathbb{N}$ (or $\mathbb{Z}$ if you will)—in particular the fact that $\gcd(a, b)$ is a linear combination of a and b (which we proved with via the Euclidean algorithm). We will see that in other number systems, such as $\mathbb{Z}[\sqrt{-5}] = \left\{a + b\sqrt{-5} : a, b \in \mathbb{Z}\right\}$, this prime divisor property is not true.

The prime divisor property allows us to prove another basic result which we normally take for granted. (This result and the previous are in fact equivalent.)

Proposition 2.3. (Fundamental theorem of arithmetic; aka. Unique prime factorization; aka. Unique factorization) *Let $n \in \mathbb{N}$, $n > 1$. Then the prime factorization of n is unique, up to reordering the primes.*

Proof. (By contradiction.) Suppose n has two distinct factorizations

$$p_1 \cdots p_k = q_1 \cdots q_\ell.$$

By cancelling out any common factors on each side, we may assume none of the p_i's equal any of the q_j's.

Now p_1 divides the product on the left, hence $p_1 | q_1 \cdots q_\ell$. So by the prime divisor property,

$$p_1 | q_1 \text{ or } p_1 | q_2 q_3 \cdots q_\ell.$$

But $p_1 \nmid q_1$ because $p_1 \neq q_1$ and q_1 is prime. Hence

$$p_1 | q_2 \cdots q_\ell \implies p_1 | q_2 \text{ or } p_1 | q_3 \cdots q_\ell.$$

Similarly $p_1 \nmid q_2$, and continuing this argument we eventually get

$$p_1 | q_{\ell-1} q_\ell \implies p_1 | q_{\ell-1} \text{ or } p_1 | q_\ell.$$

But both of these are impossible, contradicting our supposition. □

We will often be working with integers and not just natural numbers, so it may be helpful to restate it in terms of integers. It is not explicitly stated this way in the text at this point.

Fundamental theorem of arithmetic (for integers). Let $n \in \mathbb{Z}$, $n \neq 0$. Then n can be expressed in a unique way (up to reordering) as

$$n = u p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$$

where u is a *unit* (± 1), $p_i \in \mathbb{N}$ is prime, $e_i \in \mathbb{N}$ and $k \geq 0$.

(Note we allow the possibility that no primes appear ($k = 0$) in the factorization to include the units ± 1 in the statement.)

This result is called the fundamental theorem of arithmetic (though not by the text), because it plays such an important role in number theory (which is sometimes also called arithmetic). Its importance was first pointed out by Gauss (because he noticed that it doesn't hold for $\mathbb{Z}[\sqrt{-5}]$, which makes $\mathbb{Z}[\sqrt{-5}]$ much harder to work with). However, it will hold in some other number systems, such as the Gaussian integers $\mathbb{Z}[i]$ (the statement will be essentially identical to our second version).

2.5 Consequences of unique prime factorization

Here we will explore some simple consequences of unique factorization.

First we will make the following observation:
n is a square $\iff n = p_1^{2e_1} \cdots p_k^{2e_k}$

Proof. ($\Rightarrow$) If $n = m^2$, write $m = p_1^{e_1} \cdots p_k^{e_k}$.
($\Leftarrow$) Clearly $n = (p_1^{e_1} \cdots p_k^{e_k})^2$. □

This observation does not require unique prime factorization, but the next proposition does. (Pay attention how. Unfortunately the text is not careful in pointing out where unique factorization is used and how it is required here.)

Proposition 2.4. *If ab is a square and $\gcd(a, b) = 1$, then a and b are squares.*

Proof. If ab is square, write

$$ab = p_1^{2e_1} \cdots p_k^{2e_k}.$$

Now by unique factorization, the prime factorizations of a and b must be of the form

$$a = p_1^{f_1} \cdots p_k^{f_k}, \quad b = p_1^{g_1} \cdots p_k^{g_k}$$

where $f_i, g_i \geq 0$ and $f_i + g_i = 2e_i$. If a and b share no common factors, then by reordering our primes, we can write

$$a = p_1^{2e_1} \cdots p_j^{2e_j}, \quad b = p_{j+1}^{2e_{j+1}} \cdots p_k^{2e_k}.$$

Hence a and b are squares. □

If you don't have unique factorization, then it could happen that there are distinct primes p, q and r such that $p^2 = qr$, so that the proposition would be false.

Proposition 2.5. *If N is nonsquare, then $\sqrt{N}$ is irrational.*

Proof. We prove the contrapositive. Suppose $\sqrt{N} = \frac{a}{b}$ is rational. Write

$$a = p_1^{m_1} \cdots p_k^{m_k}$$

and

$$b = p_1^{n_1} \cdots p_k^{n_k}$$

where $m_i, n_i \geq 0$. (Note this is not the unique prime factorization, but by allowing 0's in the exponents, we can write a and b as products of the same primes.) Then

$$N = \frac{a^2}{b^2} = p_1^{2(m_1 - n_1)} \cdots p_k^{2(m_k - n_k)}.$$

Hence N is a square. □

Exercise 2.5. *Does Proposition 2.5 (the irrational square roots result in Section 2.5) require that the prime factorization is unique? Explain.*

Exercise 2.6. *Prove the following (slighly generalized) assertion from Section 2.5: Unique prime factorization implies that each prime power divisor of a natural number n (i.e., a prime power that divides n) actually appears in the prime factorization of n. (Hint: it's easy, but you need to write down what it means to be a divisor.)*

The previous exercise implies that a common prime power divisor of a and b appears in the prime factorization of each. Hence, $\gcd(a, b)$ *is the product of the largest common prime powers in the prime factorizations of* a *and* b. By largest common prime power we mean for each prime occuring in both a and b, the largest power of it which divides both. (The text does not talk about prime powers here, making its statements somewhat vague.)

Example. *Do* $a = 2^3 \cdot 5^2 \cdot 7$, $b = 2^5 \cdot 3^4 \cdot 7^2$.

Thus it is easy to compute the gcd if you know the prime factorizations. We can rewrite this mathematically as follows. Write

$$a = p_1^{m_1} \cdots p_k^{m_k}, \quad b = p_1^{n_1} \cdots p_k^{n_k}$$

where $m_i, n_i \geq 0$ (again not the unique factorization as exponents of 0 are allowed) in terms of common primes. Then

$$\gcd(a, b) = p_1^{\min(m_1, n_1)} \cdots p_k^{\min(m_k, n_k)}.$$

Similarly one has

$$\operatorname{lcm}(a, b) = p_1^{\max(m_1, n_1)} \cdots p_k^{\max(m_k, n_k)}.$$

Exercise 2.7. *Exercises 2.5.1 and 2.5.2.*

2.6 Linear Diophantine equations

Note: This section could have been done right after Section 2.3, and might be more natural there.

The simplest Diophantine equations (remember Diophantine equations? the alleged subject of number theory?) are the binary (2-variable) linear ones:

$$ax + by = c, \quad a, b, c \in \mathbb{Z}. \tag{2}$$

For fixed a, b, c the first thing to ask is, is there a solution (in integers)? If so, determine all solutions.

Graphically, this is a line in the plane with rational slope $-\frac{a}{b}$ and rational y-intercept $\frac{c}{b}$:

$$y = -\frac{a}{b}x + \frac{c}{b}$$

(assuming $b \neq 0$). But we will not solve it graphically—it is easier to use the divisor theory we developed.

Proposition 2.6. *Equation (2) has a solution (in integers) if and only if* $\gcd(a, b) | c$.

Proof. ($\Rightarrow$) If there is a solution, then

$$\gcd(a, b) | ax \text{ and } \gcd(a, b) | by \implies \gcd(a, b) | c.$$

($\Leftarrow$) If $\gcd(a, b) | c$, we can write $c = \gcd(a, b)d$ and by Section 2.3, for some $m, n \in \mathbb{Z}$

$$\gcd(a, b) = am + bn \implies c = \gcd(a, b)d = amd + bnd$$

exhibiting a solution of $x = md$ and $y = nd$ where m and n are found via the extended Euclidean algorithm (Section 2.3). □

Note the homogeneous equation

$$ax + by = 0 \tag{3}$$

has infinitely many solutions. Specifically, let $d = \gcd(a, b)$ and write $a = a'd$, $b = b'd$. Then all integer solutions are given by

$$\left\{(x, y) = (kb', -ka') : k \in \mathbb{Z}\right\}.$$

Proposition 2.7. *Suppose (2) has a solution (x_0, y_0) (which we obtained in the proof of the proposition above.) All solutions to (2) are given by $(x_0, y_0) + (x, y)$ where (x, y) is a solution to (3).*

Proof. You should have seen this proof in linear algebra already.

($\Rightarrow$) Suppose (x_1, y_1) is another solution to (2). We want to show it is of the desired form. Then

$$(ax_1 + by_1) - (ax_0 + by_0) = c - c = 0.$$

Hence $(x, y) = (x_1 - x_0, y_1 - y_0)$ is a solution to (3).

($\Leftarrow$) Suppose (x, y) is a solution to (3). Then

$$a(x_0 + x) + b(y_0 + y) = (ax_0 + by_0) + (ax + by) = c + 0 = c$$

so $(x_0, y_0) + (x, y)$ is a solution to (2). □

Exercise 2.8. *Exercises 2.6.1, 2.6.2, 2.6.3, 2.6.4.*

2.7 *The vector Euclidean algorithm

2.8 *The map of relatively prime pairs

We will skip these sections. I do not find them particularly interesting, however they are used in Conway's graphical theory of quadratic forms (see Chapter 5), which I do find interesting, but I do not plan to cover.

2.9 Discussion

Read it if you want it. I made all the comments I wanted to already.