Numbers, Groups and Cryptography

Gordan Savin

Contents

CHAPTER 1

Euclidean Algorithm

1. Euclidean Algorithm

Euclid was a Greek mathematician who lived in Alexandria around 300 B.C. He devised a clever and fast algorithm to find the greatest common divisor of two integers. In modern language, the Euclidean Algorithm is simply division with remainder. Recall that dividing two positive integers a, b means finding a positive integer q, called a quotient, such that

$$a = qb + r$$

with $0 \leq r < b$. The number r is called a remainder. If $r = 0$ then we say that b divides a, and write $b|a$. For example,

$$3 \mid 6 \text{ whereas } 4 \nmid 6.$$

Dividing two integers is probably the most difficult of the four standard binary operations. However, some of the most fundamental properties, such as the uniqueness of factorization, are based on the Euclidean Algorithm.

But how is this related to finding the *Greatest Common Divisor* (or simply gcd) of two integers m and n? If factorizations of m and n into prime factors are known, then it is easy to figure out what the gcd is. Consider, for example, $m = 756$ and $n = 360$. Then

$$756 = 2^2 \cdot 3^3 \cdot 7 \text{ and } 360 = 2^3 \cdot 3^2 \cdot 5.$$

We see that 2 and 3 are the only primes appearing in factorizations of both numbers. Moreover, 2^2 and 3^2 are the greatest powers of 2 and 3, respectively, which divide both 756 and 360. It follows that

$$\gcd(756, 360) = 2^2 \cdot 3^2 = 36.$$

There are two issues here, however. First, we have secretly assumed the uniqueness of factorization. The second issue, also very important, is that factoring into primes is a very difficult process, in general. A much better way was discovered by Euclid. Assume, for example, that we want to find the greatest common divisor of 60 and 22. Subtract $60 - 22 = 38$. Notice that any number dividing 60 and 22 also divides 22 and 38. Conversely, any number that divides 22 and 38 also divides $22 + 38 = 60$ and 22. Thus, instead of looking for common divisors of 60 and 22, we can look at common divisors of 22 and 38 instead. Since the later pair of numbers is smaller, the problem of finding the greatest common divisor of 60 and 22 has just become easier, and we have accomplished that without ever dividing or multiplying two numbers. In fact, we can do even better. Instead of subtracting 22 from 60 once, we can subtract it twice, to get $60 - 2 \times 22 = 16$. Thus we get that

$$\gcd(60, 22) = \gcd(22, 16).$$

Of course, we do not stop here. Since $22 - 16 = 6$, it follows that

$$\gcd(22, 16) = \gcd(16, 6)$$

and so on. At every step we replace a pair (a, b) by a pair (b, r) where r is the remainder of the division of a by b. The whole division process in this case is given here:

$$\begin{aligned} 60 &= 2 \cdot 22 + 16 \\ 22 &= 1 \cdot 16 + 6 \\ 16 &= 2 \cdot 6 + 4 \\ 6 &= 1 \cdot 4 + 2 \\ 4 &= 2 \cdot \boxed{2} + 0 \end{aligned}$$

This shows that

$$\gcd(60, 22) = \gcd(22, 16) = \cdots = \gcd(4, 2) = 2.$$

The last statement is obvious since 2 divides 4. In general, starting with a pair of numbers a and b we use the division algorithm to generate a sequence of numbers $(b > r_1 > r_2 > \ldots)$ as follows. First, we divide a by b:

$$a = q_1 b + r_1,$$

then divide b by r_1 , and so on...

$$\begin{aligned} b &= q_2 r_1 + r_2 \\ r_1 &= q_3 r_2 + r_3 \\ &\vdots \\ r_{n-2} &= q_n r_{n-1} + r_n \\ r_{n-1} &= q_{n+1} r_n + 0. \end{aligned}$$

This process stops when the remainder is 0. Since $b > r_1 > r_2 \ldots$ it stops in less than b steps. We claim that the last non-zero remainder r_n is equal to the greatest common divisor of a and b. In order to verify this, notice that the first equation $a = qb + r_1$ implies that any common divisor of b and r_1 also divides a. Likewise, if we rewrite the first equation as $a - qb = r_1$, it is clear that any common divisor of a and b also divides r_1. This shows that $\gcd(a, b) = \gcd(b, r_1)$. Arguing in this fashion, we see that

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

where, of course, the last statement is obvious since r_n divides r_{n-1}. We can summarize what we have discovered:

Euclidean Algorithm gives an effective way to compute the greatest common divisor of two integers. Moreover, the algorithm does not rely on the uniqueness of factorization.

The Euclidean Algorithm can be viewed as a special case of the *Continued Fraction Algorithm.* We need some notation before proceeding to the definition of the algorithm. If α is a real number then let $[\alpha]$ denote the greatest integer less than or equal to α. For example, since $\sqrt{2} = 1.4\ldots$ the greatest integer less than or equal to $\sqrt{2}$ is 1:

$$[\sqrt{2}] = 1.$$

Note that $[\alpha] = \alpha$ if α is an integer. The Continued Fraction Algorithm is defined as follows:

(1) Let $\alpha > 1$, and put $\beta = \alpha - [\alpha]$.
(2) If $\beta = 0$ stop, else put $\alpha_1 = 1/\beta$ and go to (1).

In order to illustrate the relation between the two algorithms, let us work out the case when $\alpha = 60/22$. Then the Continued Fraction Algorithm generates the following numerical data.

i	$[\alpha_i]$	β_i	α_{i+1}
0	2	16/22	22/16
1	1	6/16	16/6
2	2	4/6	6/4
3	1	2/4	4/2
4	2	0	STOP

Here, of course, we put $\alpha_0 = \alpha$ and $\beta_0 = \beta$. As it can be seen from the table, the Continued Fraction Algorithm generates the same numbers as the Euclidean Algorithm starting with the pair $(60, 22)$. More precisely, if $\alpha = a/b$ and $b > r_1 > r_2\ldots$ are non-negative integers generated by the Euclidean Algorithm applied to the pair (a, b), then

$$\beta = \frac{r_1}{b},\ \beta_1 = \frac{r_2}{r_1},\ \beta_2 = \frac{r_3}{r_2}\ldots$$

and

$$[\alpha] = q_1,\ [\alpha_1] = q_2,\ [\alpha_2] = q_3\ldots$$

Since $r_{n+1} = 0$ for some n, $\beta_n = 0$ and the algorithm stops for every rational number (fraction). The name (continued fraction) comes from the fact the process can be recorded as

$$\frac{60}{22} = 2 + \frac{16}{22} = 2 + \frac{1}{1 + \frac{6}{22}} = \ldots$$

$$\ldots = 2 + \cfrac{1}{1 + \frac{1}{2+\frac{1}{1+\frac{1}{2}}}}.$$

The continued fraction, in turn, gives a series of - so called - partial convergents:

$$2,\ 2 + \frac{1}{1},\ 2 + \cfrac{1}{1 + \frac{1}{2}}, 2 + \cfrac{1}{1 + \frac{1}{2+\frac{1}{1}}} \text{ and } 2 + \cfrac{1}{1 + \frac{1}{2+\frac{1}{1+\frac{1}{2}}}}.$$

An easy calculation gives that these five partial convergents are equal to, respectively,

$$2, 3, \frac{8}{3}, \frac{11}{4} \text{ and } \frac{30}{11} = \frac{60}{22}.$$

A rather different phenomenon occurs when we apply the algorithm to $\alpha = \sqrt{2}$. Then going through the loop for the first time yields

(1) $\alpha = \sqrt{2},\ \beta = \sqrt{2} - [\sqrt{2}] = \sqrt{2} - 1.$
(2)
$$\alpha_1 = \frac{1}{\beta} = \frac{1}{\sqrt{2} - 1} = \sqrt{2} + 1.$$

Going through the loop for the second time yields

(1) $\alpha_1 = \sqrt{2} + 1,\ \beta_1 = \sqrt{2} + 1 - [\sqrt{2} + 1] = \sqrt{2} - 1.$
(2)
$$\alpha_2 = \frac{1}{\beta_1} = \frac{1}{\sqrt{2} - 1} = \sqrt{2} + 1.$$

Thus the second output $\alpha_2 = \sqrt{2} + 1$ is the same as the first output of the algorithm. Of course, the third output will be the same, and so on:

$$\alpha_1 = \alpha_2 = \alpha_3 = \ldots = \sqrt{2} + 1.$$

We see that the Continued Fraction Algorithm is stuck in the loop in this case. In particular, this shows that $\sqrt{2}$ *is not a fraction.* We can write $\sqrt{2}$ as a *continued fraction*

$$\sqrt{2} = 1 + \cfrac{1}{2 + \frac{1}{2+\frac{1}{2+\frac{1}{\ddots}}}}.$$

This is a rather confusing expression, however. A proper way to view this expression is in terms of partial convergents. A partial convergent is obtained by stopping the continued fraction expression at a point. For example, the third convergent for $\sqrt{2}$ would be

$$\sqrt{2} = 1 + \frac{1}{2 + \frac{1}{2}} = \frac{7}{5}.$$

We mention, without a proof, that the partial convergents form a sequence of rational approximations of $\sqrt{2}$. For example, the first four convergents for $\sqrt{2}$ are

$$1,\ \frac{3}{2}=1.5,\ \frac{7}{5}=1.4,\ \frac{17}{12}\approx 1.41\ldots$$

Exercises

1) Use the Euclidean Algorithm to find the greatest common divisor of
a) 1084 and 412.
b) 1979 and 531.
c) 305 and 185.

2) Use calculations from the previous exercise to express in a continued fraction form:
a)
$$\frac{1084}{412}.$$
b)
$$\frac{1979}{531}.$$
c)
$$\frac{305}{185}.$$

3) If $a \mid b$ and $b \mid c$ show that $a \mid c$.

4) Use the Continued Fraction Algorithm to show that $\sqrt{3}$ is not rational.

5) Compute the fourth convergent of the continued fraction expansion of $\sqrt{3}$. Note how well it approximates $\sqrt{3}$.

6) Use the Continued Fraction Algorithm to show that $\sqrt{7}$ is not rational.

7) Compute the fourth convergent of the continued fraction expansion of $\sqrt{7}$. Note how well it approximates $\sqrt{7}$.

8) Find the number α whose continued fraction expansion is

$$\alpha = 1+\cfrac{1}{1+\frac{1}{1+\frac{1}{1+\frac{1}{\ddots}}}}.$$

2. Fundamental Theorem of Arithmetic

Pick three non-zero integers a, b and c. A purpose of this section is to find integer solutions of the following equation:

$$ax+by=c.$$

This is crucial to proving uniqueness of factorization, which will be done in the next section.

We shall first derive necessary conditions. Consider, for example, the equation

$$9x + 12y = 5.$$

It does not have any integer solutions because 3 divides the left hand side (since 3 divides 9 and 12) but 3 does not divide the right hand side, which is 5. Thus, the equation $ax + by = c$ does not have solutions unless the greatest common divisor of a and b divides c. This is a *necessary* condition. We shall see that this is also a *sufficient* solution. For example, if we replace 5 by 3 in the above equation, then

$$9x + 12y = 3$$

does have a solution, since $9(-1) + 12(1) = 3$. Before we state the main result of this section we make the following reduction. Let d denote the greatest common divisor of a and b. We assume now that d divides c, so we can write $c = kd$ for some integer k. If (x_0, y_0) is a solution of $ax + by = d$, hence

$$ax_0 + by_0 = d$$

then, after multiplying this equation by k, we have

$$a(kx_0) + b(ky_0) = kd = c,$$

making (kx_0, ky_0) a solution of the original equation $ax + by = c$. Thus we have reduced to solving:

$$ax + by = \gcd(a, b).$$

Theorem 1. *(Fundamental Theorem of Arithmetic) If a,b are two positive integers, then there exist integers x,y such that* $ax + by = \gcd(a, b)$.

Proof. We shall construct an explicit solution by applying the Euclidean Algorithm to a and b. Assume, for simplicity, that the algorithm terminates in three steps. Then r_3 is the last non-zero remainder and equal to $\gcd(a, b)$.

$$\begin{aligned} a &= q_1 b + r_1 \\ b &= q_2 r_1 + r_2 \\ r_1 &= q_3 r_2 + r_3. \end{aligned}$$

We can consider this as a system of three linear equations in five variables: a, b r_1, r_2 and r_3. We shall reduce this system to one equation in three variables a, b and r_3 as follows. First, rewrite the system as

$$\begin{aligned} r_1 &= a - q_1 b \\ r_2 &= b - q_2 r_1 \\ r_3 &= r_1 - q_3 r_2. \end{aligned}$$

Next, eliminate r_2 and r_1 using the second and the first equation, respectively. More precisely, first substitute $r_2 = b - q_2 r_1$ into the last equation, which then becomes

$$r_3 = -q_3 b + (1 + q_2 q_3) r_1.$$

Next, substitute $r_1 = a - q_1 b$ to obtain

$$r_3 = (1 + q_2 q_3) a - (q_1 + q_1 q_2 q_3 + q_3) b.$$

Since $r_3 = \gcd(a, b)$, we see that $(x, y) = (1 + q_2 q_3, -(q_1 + q_1 q_2 q_3 + q_3))$ is a solution of the equation $ax + by = \gcd(a, b)$. □

For example, if $a = 123$ and $b = 36$, then the Euclidean Algorithm gives a sequence of equations

$$\begin{array}{rcl} a & = & 3 \cdot b + 15 \\ 36 & = & 2 \cdot 15 + 6 \\ 15 & = & 2 \cdot 6 + 3 \\ 6 & = & 2 \cdot \boxed{3} + 0. \end{array}$$

In particular, $\gcd(123, 36) = 3$. As in the proof of Theorem 1, we rewrite these equations as

$$\begin{array}{rcl} 15 & = & a - 3 \cdot b \\ 6 & = & b - 2 \cdot 15 \\ 3 & = & 15 - 2 \cdot 6 \end{array}$$

and, after two substitutions, the last equation becomes $3 = 5a - 17b$ from which we identify a solution:

$$(x, y) = (5, -17).$$

The substitutions, however, are often rather cumbersome. In fact, this process could be made easier using the continued fraction expansion

$$\frac{123}{36} = 3 + \frac{1}{2 + \frac{1}{2 + \frac{1}{2}}}.$$

The partial convergents are

$$\frac{2}{1}, \frac{7}{2}, \frac{17}{5} \text{ and } \frac{41}{12} = \frac{123}{36}.$$

The convergent before the last one, in this case $\frac{17}{5}$, gives a solution of the equation $123x + 36y = 3$. As this example illustrates, the Continued Fraction Algorithm gives a solution (x_0, y_0) of the equation $ax + by = d$ ($d = \gcd(a, b)$) such that

$$|x_0| \leq \frac{b}{d} \text{ and } |y_0| \leq \frac{a}{d}.$$

We now address the question of finding all solutions. For example, if x_0, y_0 is a solution of $ax + by = d$, and ℓ is any integer, then

$$a(x_0 + \ell b) + b(y_0 - \ell a) = a x_0 + b y_0 = d$$

thus $(x_0 - \ell b, y_0 + \ell a)$ is another solution. In fact we have the following:

THEOREM 2. *If x_0, y_0 is a solution of $ax + by = d$, where $d = \gcd(a, b)$, then every solution of this equation is of the form:*

$$x = x_0 + k\frac{b}{d},\ y = y_0 - k\frac{a}{d}$$

where k is any integer.

PROOF. The proof is a simple manipulation of equations. Let (x_1, y_1) and (x_2, y_2) be two solutions meaning,

$$\begin{cases} ax_1 + by_1 = d \\ ax_2 + by_2 = d. \end{cases}$$

Then, if we multiply the first equation by y_2 and x_2, and the second equation by y_1 and x_1 we get two columns of equations

$$ax_1y_2 + by_1y_2 = dy_2 \qquad ax_1x_2 + by_1x_2 = dx_2$$

$$ax_2y_1 + by_2y_1 = dy_1 \qquad ax_1x_2 + by_2x_1 = dx_1$$

and, after subtracting the second equation from the first in both columns, we get two equations

$$a(x_1y_2 - x_2y_1) = d(y_2 - y_1) \qquad b(x_1y_2 - x_2y_1) = d(x_1 - x_2).$$

Since d divides a and b, we can divide both equations by d, hence

$$\frac{b}{d}(x_1y_2 - x_2y_1) = x_1 - x_2$$

$$\frac{a}{d}(x_1y_2 - x_2y_1) = y_2 - y_1$$

and this is the same as

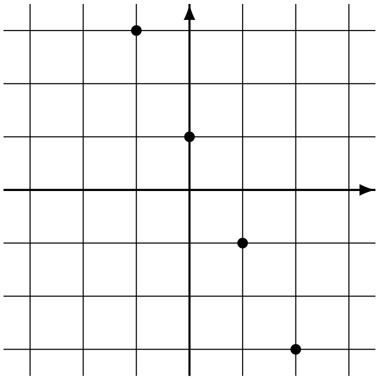
$$x_1 = x_2 + k\frac{b}{d}$$

$$y_1 = y_2 - k\frac{a}{d}$$

where $k = x_1y_2 - x_2y_1$. This proves the theorem. □

Example: If $2x + y = 1$ then one solution is $(1, -1)$. All solutions form a sequence (see the figure)

$$\ldots(-1, 3), (0, 1), (1, -1), (2, -3), \ldots$$



Exercises

1) Find all integer solutions of $13853x + 6951y = \gcd(13853, 6951)$.

2) Find all integer solutions of $15750x + 9150y = \gcd(15750, 9150)$.

3) Show that $427x + 259y = 13$ has no integer solutions.

Use the Fundamental Theorem of Arithmetic in the following two exercises:

4) Let a and b be two integers. Show that any common divisor of a and b divides the greatest common divisor of a and b.

5) Let a and b be two relatively prime integers ($\gcd(a, b) = 1$). Show that if a and b divide c, then ab divides c.

6) Show that

$$\gcd(ad, bd) = \gcd(a, b)d.$$

7) Let a and b be two positive integers. The lowest common multiple of a and b, denoted by $\operatorname{lcm}(a, b)$, is the smallest positive integer divisible by *both* a and b. Show that $\operatorname{lcm}(a, b)$ divides any common multiple of a and b. Hint: If m is a common multiple, write

$$m = q \cdot \operatorname{lcm}(a, b) + r$$

with $0 \leq r < \operatorname{lcm}(a, b)$.

8) Show that

$$\operatorname{lcm}(a, b) \gcd(a, b) = ab.$$

Hint: Do the case $\gcd(a, b) = 1$ first.

9) Use the Euclidean Algorithm and the formula in the exercise 8) to find
a) $\operatorname{lcm}(13853, 6951)$
b) $\operatorname{lcm}(15750, 9150)$.

3. Uniqueness of Factorization

A positive integer $p > 1$ is called *prime* if it is divisible only by 1 and itself. Positive integers greater than 1 which are not primes are called composite. For example, 5 is a prime, while $12 = 3 \cdot 4$ is a composite number. We note that 1 is considered neither a prime nor a composite integer. Every positive integer greater than 1 can be factored into prime factors. This is not entirely obvious, if you think about it. The argument is based on the following:

Well ordered axiom: Every non-empty subset S of positive integers has a smallest element.

Thus, if there are positive integers that cannot be factored into primes, then all such integers form a non empty set S. By the axiom there exists a smallest element n in S. Now n cannot be a prime, since prime numbers can be factored into primes (obviously!). Thus $n = a \cdot b$ for two integers less than n. Since a and b are less than n they do not belong to S and, therefore, can be factored into primes. It follows that $n = a \cdot b$ can also be factored into primes. This is a contradiction. Thus S must be empty.

In practice this argument can be illustrated as follows. If we can factor into primes all integers less than 12 then we can factor 12 as well. Indeed, 12 is not a prime since $12 = 3 \cdot 4$. Since 3 is a prime and $4 = 2^2$ we get that $12 = 2^2 \cdot 3$.

We can find (some) prime numbers using a classical method: The Sieve of Eratosthenes. First, we list integers as many as we can or wish, starting with 2. Then we remove all numbers greater than 2 and divisible by 2, for they are obviously not prime. The first remaining integer is 3. Thus it must be a prime. Then we remove all integers greater than 3 and divisible by 3, and so on. What remains, in the end, are prime numbers. For numbers up to 19 this process can be illustrated by the following table:

2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2	3		5		7		9		11		13		15		17		19
2	3		5		7				11		13				17		19

The first row contains all numbers from 2 to 19. The second row has multiples of 2, except 2, removed. The third row has all multiples of 3, except 3, removed. The fourth row, not pictured here, would have all multiples of 5, except 5, removed and so on. In the end we can conclude that 2, 3, 5, 7, 11, 13, 17 and 19 are primes less than 20.

One of the most important results in number theory is the uniqueness of factorization into primes. This is not at all obvious. Take, for example, 42. We can write $42 = 2 \cdot 21$ and, since $21 = 3 \cdot 7$, we have $42 = 2 \cdot 3 \cdot 7$. This process can not be continued since all the factors 2, 3 and 7 are prime. However, factoring of 42 could have been achieved by first writing $42 = 3 \cdot 14$ and then using $14 = 2 \cdot 7$. In both cases we arrive to the same answer $42 = 2 \cdot 3 \cdot 7$, up to a permutation of factors.

THEOREM 3. *Every positive integer can be uniquely, up to ordering of factors, factored as a product of primes.*

For example, $6 = 2 \cdot 3$ and $6 = 3 \cdot 2$ are the only possible prime factorizations of 6. Proof of uniqueness of factorization is based on the following statement.

LEMMA 4. *Let p be prime, and a and b be two integers such that p divides ab. Then p divides a or b. For example,* 3 *divides* $36 = 4 \cdot 9$ *and* 3 *divides* 9*, one of the factors.*

PROOF. To prove the lemma, we need to show that if p does not divide one of the factors, say a, then p must divide the other factor b. To that end, if p does not divide a then the only common divisor of p and a is 1. Thus, by the fundamental theorem of arithmetic, there exist integers x and y such that

$$ax + py = 1.$$

Multiplying this equation by b we have $abx + pby = b$. Since p divides ab, it divides $abx + pby$ and, therefore, b. □

The statement can be generalized to arbitrary number of factors. If $p|a_1, a_2, \ldots, a_p$ then $p|a_i$ for some i s.t. $1 \leq i \leq r$. Indeed, either $p|a_1$ or $p|(a_2, \ldots, a_r)$, then either $p|a_2$ or $p|(a_3, \ldots, a_r)$, and so on ...

We are now ready to show uniqueness of factorization. Let n be a positive integer and $n = p_1, \ldots, p_r$ and $n = q_1, \ldots, q_s$ be two factorizations into primes. Then

$$p_1 \cdot \ldots \cdot p_r = q_1 \cdot \ldots \cdot q_s.$$

So, $p_1|q_1 \cdot \ldots \cdot q_s$ and by Lemma $p_1|q_i$ (for some i s.t. $1 \leq i \leq s$) and since q_i is a prime, $p_1 = q_i$. After rearranging q_i's we can assume that $p_1 = q_1$, and after canceling $p_1 = q_1$ on both sides we get

$$p_2 \cdot \ldots \cdot p_r = q_2 \cdot \ldots \cdot q_s.$$

The same argument gives $p_2 = q_2$, $p_3 = q_3$ and so on. The theorem is proved.

Now that we have shown that every positive integer can be factored uniquely into primes, a natural question is how to factor a given integer into primes n? For example, if $n = p \cdot q$ is a product of two large, unfamiliar primes, there is no easy way to factor n, and this fact lies at the heart of many modern cryptography schemes. However, if the primes p and q are relatively close to each other, there is a neat trick which is based on the following identity:

$$n = \left(\frac{p+q}{2}\right)^2 - \left(\frac{p-q}{2}\right)^2.$$

As an example we shall factor 826277, which is a product of two close primes, which means that $p-q$ is small. Since p and q are odd, the difference $p-q$ is even. Thus the first possibility is $p-q=2$, which implies that

$$\left(\frac{p+q}{2}\right)^2 = 826277 + 1^2 = 826278.$$

Since 826278 is not a square, our guess $p-q=2$ is incorrect. So let's try the next possibility $p-q=4$. Then

$$\left(\frac{p+q}{2}\right)^2 = 826277 + 2^2 = 826281.$$

Since $826281 = 909^2$ we deduce that $p+q=1818$ which, combined with $p-q=4$, gives that $p=911$ and $q=907$ or

$$826277 = 911 \cdot 907.$$

Exercises

1) Use the Sieve of Erathostenes to determine all primes less than 100.

2) Let a and b two positive integers such that $a+b$ is a prime number. Prove that the greatest common divisor of a and b is 1.

3) The numbers 3992003 and 1340939 are each products of two close primes. Find the primes.

4. Efficiency of the Euclidean Algorithm

The goal of this section is to understand in how many steps the Euclidean Algorithm terminates. An answer to this problem introduces, somewhat surprisingly, Fibonacci numbers. We start with two examples. In the first, we take $a=144$ and $b=71$. Then

$$\begin{aligned}
&144 = 2 \cdot 71 + 2\\
&71 = 35 \cdot 2 + 1\\
&2 = 2 \cdot 1 + 0,
\end{aligned}$$

and the algorithm terminates in 3 steps. On the other had, if we take $a = 144, b = 89$, then the Euclidean Algorithm has 10 steps:

$$\begin{aligned}
&144 = 1 \cdot 89 + 55\\
&89 = 1 \cdot 55 + 34\\
&55 = 1 \cdot 34 + 21\\
&34 = 1 \cdot 21 + 13\\
&21 = 1 \cdot 13 + 8\\
&13 = 1 \cdot 8 + 5\\
&8 = 1 \cdot 5 + 3\\
&5 = 1 \cdot 3 + 2\\
&3 = 1 \cdot 2 + 1\\
&2 = 2 \cdot 1 + 0.
\end{aligned}$$

The main result of this section is that the Euclidean Algorithm terminates in the number of steps which is not greater than 5 times the number of digits of b. If we take this as granted, we see that the second example illustrates a worst case scenario since $b = 89$ has 2 digits.

We start by expressing the number of digits of a number b in terms of $\log_{10} b$. To that end, note that every number between 10^k and 10^{k+1} (but not equal to 10^{k+1}) has precisely $k+1$ digits. In particular, since $b = 10^{\log_{10} b}$ and

$$10^{[\log_{10} b]} \leq b < 10^{[\log_{10} b]+1}$$

the number of digits of b is

$$[\log_{10} b] + 1 \approx \log_{10} b.$$

Proposition 5. *Let a and b be two positive integers such that $a \geq b$. The Euclidean Algorithm for the pair (a, b) terminates in the number of steps not greater than 5 times the number of digits of b.*

Proof. Assume that the Euclidean algorithm terminates after n steps. Let

$$b = r_n > r_{n-1} > \cdots > r_1 > r_0 = 0$$

be all remainders, which we have indexed in the opposite order than in the previous occasions. In particular, we have

$$\begin{array}{rcl}
a &=& q_{n-1}b + r_{n-1}\\
b &=& q_{n-2}r_{n-1} + r_{n-2}\\
&\vdots&\\
r_3 &=& q_1 r_2 + r_1\\
r_2 &=& q_0 r_1.
\end{array}$$

Since $q_i \geq 1$ the sequence of equalities can be replaced by a sequence of inequalities

$$r_{i+1} \geq r_i + r_{i-1}.$$

Next, note that $r_1 \geq 1$ and, $r_2 \geq 2$, since $r_2 > r_1$. The inequalities imply that

$$r_3 \geq r_2 + r_1 \geq 2 + 1 = 3$$
$$r_4 \geq r_3 + r_2 \geq 3 + 2 = 5$$
$$r_5 \geq r_4 + r_3 \geq 5 + 3 = 8.$$

The numbers $1, 2, 3, 5, 8 \ldots$ belong to the the Fibonacci sequence F_i. This sequence is defined by $F_0 = 1$, $F_1 = 1$ and the recursive relation

$$F_{i+1} = F_i + F_{i-1}.$$

The first 12 terms of the Fibonacci Sequence are

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89 \text{ and } 144$$

which have already appeared in the second example above. In particular, it follows that

$$r_i \geq F_i$$

for all i. Thus $b \geq F_n$ and the number of digits of b is greater than or equal to the number of digits of F_n.

In order to estimate the number of digits of F_n we need a closed formula for F_n, which is

$$F_n = \frac{(1+\sqrt{5})^{n+1} - (1-\sqrt{5})^{n+1}}{2^{n+1}\sqrt{5}}.$$

(A derivation of this formula is given as an exercise at the end of this section.) Since $(1-\sqrt{5})/2 \approx -0.6$ it follows that

$$\left(\frac{1-\sqrt{5}}{2}\right)^{n+1} \to 0$$

as $n \to \infty$. Thus the second term in the formula for F_n can be ignored for purposes of estimating the number of digits of F_n. Therefore, using

$$F_n \approx \frac{(1+\sqrt{5})^{n+1}}{2^{n+1}\sqrt{5}},$$

we obtain that

$$\log_{10}(F_n) \approx (n+1)\log_{10}\left(\frac{1+\sqrt{5}}{2}\right) - \log_{10}(\sqrt{5}) \approx \frac{n}{5},$$

where we used that $\log_{10}\left(\frac{1+\sqrt{5}}{2}\right) \approx 1/5$. We have shown that the number of digits of b is greater than or equal to the number of steps (n) divided by 5 or, equivalently, the number of steps is not more than 5 times the number of digits of b. The proposition is proved. □

Exercises

1) Let

$$f_n = \frac{(1+\sqrt{5})^n - (1-\sqrt{5})^n}{2^n\sqrt{5}}.$$

Show that $f_{n+2} = f_{n+1} + f_n$. In words, show that f_n satisfy the same recursion relation as the Fibonacci numbers. What else do you need to check in order to verify that f_n are indeed the Fibonacci numbers?

The purpose of the next two exercises is to derive a weaker estimate on the efficiency of the Euclidean Algorithm without using the Fibonacci numbers.

2) Observe that the remainders in the Euclidean Algorithm satisfy the inequality

$$r_{i+1} \geq r_i + r_{i-1} > 2r_{i-1}.$$

Use this to show that the number of steps in the Euclidean Algorithm is less than or equal to $2\log_2(b)$.

3) Show that $2\log_2(b)$ is greater than six times the number of digits of b. Hint: notice that $\log_2 10 > 3$.

CHAPTER 2

Groups and Arithmetic

1. Groups

Addition and multiplication of numbers are examples of binary operations in mathematics. More generally, binary operation on a set G is a procedure that, from any two elements a and b in G, produces an element in G denoted by $a \cdot b$. The set G is a *group* if the binary operation $\cdot$ satisfies the following properties (axioms):

(1) (associativity) For any three elements a, b and c in G

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c.$$

(2) (existence of identity) There exists an element e in G, called identity, such that for all a in G

$$a \cdot e = e \cdot a = a.$$

(3) (existence of inverse) For every a in G there exists b, called an inverse of a, such that

$$a \cdot b = b \cdot a = e.$$

The inverse of a is usually denoted by a^{-1}.

The group is commutative if $a \cdot b = b \cdot a$ for all pairs of elements in G. An example of a finite non-commutative group will be given in the next section. As a first example, consider the set of integers

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

and addition as the binary operation. As it is customary, we will use use $+$ and not $\cdot$ to denote this binary operation. The addition of integers is associative

$$a + (b + c) = (a + b) + c$$

so the first axiom is satisfied. Next, we need to check that there exists an integer e such that $a + e = e + a = a$ for every integer a. Clearly $e = 0$ satisfies this axiom. Finally, $-a$ is an inverse of a, for any integer a. Thus the set of integers is a group for addition. The same holds for rational numbers $\mathbb{Q}$, real numbers $\mathbb{R}$ and complex numbers $\mathbb{C}$. Each is a group for the usual addition of numbers with 0 as identity.

As the next example, consider $\mathbb{Q}^\times$, the set of non-zero rational numbers, and multiplication as the operation. Again, as you well know, the multiplication is associative, 1 is an identity and any non-zero fraction n/m has an inverse. The inverse is, of course, the fraction m/n. It follows that rational numbers form a group for multiplication with 1 as identity. The same holds for non-zero real numbers $\mathbb{R}^\times$ and non-zero complex numbers $\mathbb{C}^\times$.

Since a group is not commutative in general, we need to be careful with certain "common" formulas. For example, one may be tempted to write $(a \cdot b)^{-1} = a^{-1} \cdot b^{-1}$, but this is not right. In fact, the correct version is

$$(a \cdot b)^{-1} = b^{-1} \cdot a^{-1},$$

in words, the inverse of a product is the product of inverses, but in the reverse order. Indeed,

$$(a \cdot b) \cdot (b^{-1} \cdot a^{-1}) = a \cdot (b \cdot b^{-1}) \cdot a^{-1} = a \cdot a^{-1} = e.$$

In fact, the formula $(a \cdot b)^{-1} = b^{-1} \cdot a^{-1}$ is more intuitive. For example, if you build a brick wall, you start with the first row, then second and so on, from the bottom to the top. But to take down the wall you go in the opposite order, form the top to the bottom. Thus, we have the following general formula:

$$(a_1 \cdot a_2 \cdot \ldots \cdot a_n)^{-1} = a_n^{-1} \cdot \ldots \cdot a_2^{-1} \cdot a_1^{-1}.$$

An important property of any group law is the cancellation property. More precisely,

Proposition 6. *Let G be a group and a, x and y three elements of G. If $a \cdot x = a \cdot y$ then $x = y$.*

Proof. Let b be the inverse of a. Then $a \cdot x = a \cdot y$ implies that

$$b \cdot (a \cdot x) = b \cdot (a \cdot y).$$

Using associativity this can be rewritten as

$$(b \cdot a) \cdot x = (b \cdot a) \cdot y.$$

Since $b \cdot a = e$, the identity element in the group, the last equation reduces to $x = y$, as claimed. □

As an example where the cancellation property does not hold, consider the set of all 2×2 matrices with integer coefficients. The matrix multiplication is defined by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} aa' + bc' & ab' + bd' \\ ca' + dc' & cb' + dd' \end{pmatrix}.$$

It is associative, and

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

is an identity, but the cancellation property does not hold as the following example shows. First, a simple calculation shows that

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

and

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

This implies that

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}.$$

However, we cannot cancel

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

since

$$\begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}.$$

In any group we can define integer powers g^n of any element g. If n is positive, then

$$g^n = \underbrace{g \cdot \ldots \cdot g}_{n-\text{times}}.$$

Next, $g^0 = e$ and finally, still assuming that n is positive, we define

$$g^{-n} = \underbrace{g^{-1} \cdot \ldots \cdot g^{-1}}_{n-\text{times}}.$$

The powers satisfy the usual elementary-school rules:

$$g^n \cdot g^m = g^{n+m}$$

and

$$(g^n)^m = g^{nm}.$$

A nonempty subset H of a group G is a subgroup of G if the following two axioms hold:

(1) The product of any two elements in H is contained in H.
(2) The inverse of any element in H is also contained in H.

It follows from the axioms that the identity element e of G is contained in H. Indeed, pick an element h in H. Then h^{-1} is in H, by the second axiom, and $e = h \cdot h^{-1}$ is in H, by the first axiom. Note that H is also a group, with respect to the same binary operation.

Examples:

(1) Let G be a group and e its identity element. Then $H = \{e\}$ is a subgroup. This group is called the *trivial* subgroup of G.

(2) Let $G = \mathbb{Z}$, the group of integers with respect to the usual addition. Let H be the subset of all even integers. This is a subgroup of G, since the sum of two even integers is even, and the inverse of an even integer is even.

(3) Let $G = \mathbb{Z}$, the group of integers with respect to the usual addition. Let H be the subset of all positive integers. This subset is closed under addition, but not under taking inverse. Thus H is not a subgroup.

Here is a very general and important example of a subgroup. Let G be a group and g an element in G. Let $\langle g \rangle$ be the set of all integral powers of g:

$$\langle g \rangle = \{\ldots, g^{-2}, g^{-1}, e, g, g^2, \ldots\}.$$

Note that $\langle g \rangle$ is a subgroup. Indeed, since $g^n \cdot g^m = g^{n+m}$ the first axiom holds, and since $(g^n)^{-1} = g^{-n}$ the second axiom holds.

If $G = \langle g \rangle$ for some element g in G, that is, if any element in G is a power of g, we say that G is a *cyclic* group.

Exercises

1) Let G be a group and e and e' two identity elements. Show that $e = e'$. Hint: consider $e \cdot e'$ and calculate it using first that e is an identity and then using that e' is an identity.

2) Let G be a group and a an element in G. Show that the inverse of a is unique.

3) Let G be a group with identity e and such that $a^2 = e$ for every element a in G. Show that G is commutative, $a \cdot b = b \cdot a$ for every two elements a and b in G. Hint: consider $(a \cdot b)^2$.

4) Fix n, a positive integer. Let

$$n\mathbb{Z} = \{\ldots - 2n, -n, 0, n, 2n, \ldots\}$$

be the set of all integer multiples of n. Show that $n\mathbb{Z}$ is a subgroup of $\mathbb{Z}$ with respect to addition as the binary operation.

5) Let H be a non trivial subgroup of $\mathbb{Z}$. Show that $H = n\mathbb{Z}$ for some positive integer n. Hint: n is the smallest positive integer in H.

Let $SL_2(\mathbb{Z})$ be the set of all integer valued 2×2 matrices of determinant one. The purpose of the following six exercises is to show that $SL_2(\mathbb{Z})$ is a group with respect to the (usual) multiplication of matrices. The group $SL_2(\mathbb{Z})$ is easily the most interesting, important and difficult group in mathematics.

6) Before we proceed with checking the group axioms, you first need to verify that $SL_2(\mathbb{Z})$ is *closed with respect to multiplication.* Show that the

product of any two elements A and B in $SL_2(\mathbb{Z})$ is also in $SL_2(\mathbb{Z})$. (We need to verify that the result of multiplication is back in the proposed group $SL_2(\mathbb{Z})$, otherwise the question - whether $SL_2(\mathbb{Z})$ is a group - does not make any sense.)

7) Show that the matrix multiplication is associative. That is, show that for any three matrices A, B and C in $SL_2(\mathbb{Z})$ we have

$$A(BC) = (AB)C.$$

8) Show that the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

is an identity for the multiplication.

9) Show that the inverse is given by the formula:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

10) Compute (and verify) the inverse of

$$\begin{pmatrix} 3 & 4 \\ 2 & 3 \end{pmatrix}.$$

11) Show that $(AB)^{-1} \neq A^{-1}B^{-1}$ in $SL_2(\mathbb{Z})$ where

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

2. Congruences

Fix a positive integer n. We say that two integers a and b are congruent modulo n if the difference $a - b$ is divisible by n. This is classically denoted by

$$a \equiv b \pmod{n}.$$

For example, if $n = 12$, then 14 is congruent to 2 modulo 12, since the difference of 14 and 2 is divisible by 12. We write

$$14 \equiv 2 \pmod{12}.$$

We can add and multiply integers modulo n. For example, the usual addition of integers gives $10 + 4 = 14$. Since $14 \equiv 2 \pmod{12}$ we say that adding 10 and 4 modulo 12 gives 2. We write this as

$$10 + 4 \equiv 2 \pmod{12}.$$

Similarly, the usual multiplication of integers gives $3 \cdot 7 = 21$. Since $21 \equiv 9 \pmod{12}$ we say that multiplying 3 and 7 modulo 12 gives 9. We write this as

$$3 \cdot 7 \equiv 9 \pmod{12}.$$

This may come as a surprise, but modular arithmetic is used in everyday life: The addition modulo 12 is simply the clock arithmetic. For example, if it is 10 o'clock now, in 4 hours it will be 2 o'clock.

Modular addition is well defined in the sense that the result of addition does not depend on choices of integers modulo n. If in the above example we replace 10 and 4 by 22 and 28, then the sum is 50 which is again congruent to 2 modulo 12. In general, we have to show that

$$a \equiv b \pmod{n} \text{ and } a' \equiv b' \pmod{n}$$

imply that $a + a' \equiv b + b' \pmod{n}$. Indeed, if $a = b + kn$ and $a' = b' + k'n$, then

$$a + a' = b + b' + (k + k')n \equiv b + b' \pmod{n}.$$

Similarly, in order to check that the modular multiplication is well defined, we have to show that

$$a \equiv b \pmod{n} \text{ and } a' \equiv b' \pmod{n}$$

imply that $aa' \equiv bb' \pmod{n}$. Indeed, if $a = b + kn$ and $a' = b' + k'n$, then

$$aa' = (b + kn)(b' + k'n) = bb' + (k + k' + n)n \equiv bb' \pmod{n}.$$

In order to illustrate the power of congruences we prove the following criterion which describes, in simple terms, whether a positive integer is divisible by 9.

PROPOSITION 7. *A positive integer m is divisible by 9 if and only if the sum of digits of m, in base 10, is divisible by 9.*

PROOF. Write $m = a_n a_{n-1} \cdots a_1 a_0$ where $a_n, a_{n-1}, \ldots, a_1, a_0$ are the digits of m. In other words,

$$m = a_n 10^n + a_{n-1} 10^{n-1} + \cdots + a_1 10^1 + a_0 10^0.$$

Next, notice that $10 \equiv 1 \pmod{9}$. If we raise this equation to the k-th power, we get $10^k \equiv 1 \pmod{9}$ for every k. It follows that

$$m \equiv a_n + a_{n-1} + \cdots + a_1 + a_0 \pmod{9},$$

as claimed. □

For example, 111111111 is divisible by 9 since the sum of its digits is 9.

A new and sophisticated application of congruences is given by the International Standard Book Number (ISBN). This number, usually printed on the back cover of the book, is intended to uniquely identify the book.

In addition, the number is designed so that it includes a *check* digit. The purpose of this digit, as it will be explained in a moment, is to detect one of the following two errors, which can likely occur in the process of typing an ISBN number:

- Entering incorrectly one digit of an ISBN number.
- Switching two digits of an ISBN number.

An ISBN number is a 10 digit number $x_1 \dots x_{10}$ divided into four parts of variable length, for example,

0-12-732961-7.

The first group identifies where the book was published. It is 0 in this example, which indicates that the book in question was published in US, Canada, UK, Australia or New Zealand. The second group identifies a publisher. The third group identifies a particular title and edition. The last number, 7 in this example, is the check digit x_{10}. It is computed, modulo 11, using the first 9 digits:

$$x_{10} \equiv \sum_{i=1}^{9} ix_i \pmod{11}.$$

In particular, x_{10} can take values $0, 1, \dots, 9$ and 10. Since 10 has two digits, it is entered as X.

Proposition 8. *Assume that a ten digit number $y_1 \dots y_{10}$ is obtained by entering a ten digit ISBN number $x_1 \dots x_{10}$ so that either one digit was entered incorrectly or two digits were switched. Then*

$$y_{10} \not\equiv \sum_{i=1}^{9} iy_i \pmod{11}.$$

Proof. Notice that, by adding $10x_{10}$ to both sides, the congruence

$$\sum_{i=1}^{9} ix_i \equiv x_{10} \pmod{11}.$$

is equivalent to

$$\sum_{i=1}^{10} ix_i \equiv 0 \pmod{11}.$$

Now assume that the number $y_1 \dots y_{10}$ differs from the valid ISBN number $x_1 \dots x_{10}$ in only one digit y_j. Then

$$\sum_{i=1}^{10} iy_i \equiv \sum_{i=1}^{10} iy_i - \sum_{i=1}^{10} ix_i = j(y_j - x_j) \not\equiv 0 \pmod{11}$$

since 11 does not divide $j(y_j - x_j)$. Thus $y_1 \dots y_{10}$ is not a valid ISBN number. Next, assume that the number $y_1 \dots y_{10}$ is obtained from the valid

ISBN number $x_1 \ldots x_{10}$ by switching two digits, $x_j \neq x_k$. That is, $y_j = x_k$ and $y_k = x_j$. Then

$$\sum_{i=1}^{10} iy_i \equiv \sum_{i=1}^{10} iy_i - \sum_{i=1}^{10} ix_i = (k-j)(y_k - y_j) \not\equiv 0 \pmod{11}$$

since 11 does not divide $(k-j)(y_k - y_j)$. □

In some cases it is even possible to recover a valid ISBN number. Assume that a number $y_1 \ldots y_{10}$ is obtained from a valid ISBN number $x_1 \ldots x_{10}$ by switching consecutive digits, x_i and x_{i+1}. Then, in order to find the valid ISBN number, we need to find and switch two consecutive digits y_i and y_{i+1} such that

$$\sum_{i=1}^{10} i \cdot y_i \equiv y_{i+1} - y_i \pmod{11}$$

For example, consider the number 0-354-18834-6. Then

$$1 \cdot 0 + 2 \cdot 3 + 3 \cdot 5 + 4 \cdot 4 + 5 \cdot 1 + 6 \cdot 8 + 7 \cdot 8 + 8 \cdot 3 + 9 \cdot 4 - 6 \equiv 2 \pmod{11}.$$

Thus, we are looking for two consecutive digits in 0-354-18834-6 with difference 2. These digits are either 3 and 5 or 4 and 6. It follows that the original ISBN number is either 0-534-18834-6 or 0-354-18836-4.

Exercises

1) Show that a positive integer m is divisible by 11 if and only if the alternating sum of its digits is divisible by 11. Hint: notice that $10 \equiv -1 \pmod{11}$.

2) Is 212121212121212121212121 divisible by 9? Is it divisible by 11?

3) The number 3-540-97285-9 is obtained from a valid ISBN number by switching two consecutive digits. Find the ISBN number.

4) The number 0-31-030369-0 is obtained from a valid ISBN number by switching two consecutive digits. Find the ISBN number.

5) Let n be a positive integer. If $x \equiv y \pmod{n}$ and $y \equiv z \pmod{n}$ prove that $x \equiv z \pmod{n}$ for any three integers x, y and z.

3. Modular arithmetic

Fix a positive integer n. If a is any integer, then the set of all integers congruent to a is called the *class* of a modulo n. The set of classes of integers modulo n is denoted by $\mathbb{Z}/n\mathbb{Z}$. Since every number a can be written as $a = qn + r$ where $0 \leq r < n$, every class modulo n is uniquely represented by an element in the set

$$\{0, 1, \ldots, n-1\}.$$

The set of congruence classes is a group with respect to modular addition, as all group axioms are inherited from $\mathbb{Z}$. The class of 0 (all multiples of n) is the identity element. The inverse of the class of a is the class of $-a$. For example, consider $n = 5$. Then we have the following 5 classes:

$$\begin{gathered}\{\ldots, -5, 0, 5, 10, \ldots\}\\ \{\ldots, -4, 1, 6, 11, \ldots\}\\ \{\ldots, -3, 2, 7, 12, \ldots\}\\ \{\ldots, -2, 3, 8, 13, \ldots\}\\ \{\ldots, -1, 4, 9, 14, \ldots\}\end{gathered}$$

A complete set of representatives is $\{0, 1, 2, 3, 4\}$, so we write

$$\mathbb{Z}/5\mathbb{Z} = \{0, 1, 2, 3, 4\}.$$

The addition modulo 5 in terms of these representatives is given by the following table:

+	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

The situation with modular multiplication is more delicate. The multiplication is associative, this is inherited from integers, and the class of 1 is an identity element. But what about inverses? Integers, except -1 and 1, do not have an inverse with respect to multiplication. In particular, any group structure with respect to modular multiplication is not simply inherited from $\mathbb{Z}$. In order to understand the situation, we start with an example. The multiplication table modulo 5 is:

$\cdot$	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

We see from the table that every non-zero number modulo 5 has a multiplicative inverse, and $(\mathbb{Z}/5\mathbb{Z})^\times = \{1, 2, 3, 4\}$ is a group with respect to multiplication. It turns out that the same holds for reduction modulo p for every prime p:

PROPOSITION 9. *Let p be a prime. Then $(\mathbb{Z}/p\mathbb{Z})^\times = \{1, 2, 3, \ldots p - 1\}$ is a group for modular multiplication.*

Proof. First of all, note that the set of all invertible elements in $\mathbb{Z}/p\mathbb{Z}$ is closed under multiplication. Indeed, if a and b are invertible, then ab is also invertible since

$$(ab)^{-1} = b^{-1}a^{-1}.$$

Since 0 is clearly not invertible, it remains to show that every a, $1 \le a \le p-1$, has an inverse. To that end, we must find an integer x such that $ax \equiv 1 \pmod{p}$. Since p is a prime and $a < p$ it follows that $\gcd(a, p) = 1$. Thus, there exist two integers x and y such that

$$ax + py = 1$$

which clearly implies that $ax \equiv 1 \pmod{p}$, as desired. □

As an example consider the case $p = 31$ and $a = 7$. In order to find the inverse of 7 we need to solve the equation

$$7x + 31y = 1.$$

As we know this is accomplished by applying the Euclidean Algorithm:

$$31 = 4 \cdot 7 + 3$$

$$7 = 2 \cdot 3 + 1.$$

The first equation gives $3 = 31 - 4 \cdot 7$ which we then substitute for 3 in the second equation yielding

$$7 = 2 \cdot (31 - 4 \cdot 7) + 1 \text{ or } 9 \cdot 7 - 2 \cdot 31 = 1.$$

Thus, we have obtained that 9 is the multiplicative inverse of 7 modulo 31.

Once we know that 9 is the inverse of 7 modulo 31, we can solve the linear equation

$$7x \equiv 11 \pmod{31}$$

simply by multiplying both sides of the congruence by 9:

$$9 \cdot (7x) \equiv 9 \cdot 11 \pmod{31}.$$

Since $9 \cdot 7 \equiv 1 \pmod{31}$ and $9 \cdot 11 \equiv 6 \pmod{31}$, it follows that 6 is a solution of the original congruence.

Exercises

1) Write down the addition table for $\mathbb{Z}/5\mathbb{Z}$ using $\{5, 6, 7, 8, 9\}$ as representatives of classes. Verify, by inspection, that the entries of the resulting table are congruent to those obtained using $\{0, 1, 2, 3, 4\}$ as representatives of classes.

2) Write down the multiplication table for $\mathbb{Z}/5\mathbb{Z}$ using $\{5, 6, 7, 8, 9\}$ as representatives of classes. Verify, by inspection, that the entries of the resulting table are congruent to those obtained using $\{0, 1, 2, 3, 4\}$ as representatives of classes.

3) Write down the multiplication table for $\mathbb{Z}/11\mathbb{Z}$, the set of integers modulo 11. The subset of invertible integers modulo 11 is denoted by $(\mathbb{Z}/11\mathbb{Z})^\times$. Extract the multiplication table for $(\mathbb{Z}/11\mathbb{Z})^\times$.

4) Write down the multiplication table for $\mathbb{Z}/10\mathbb{Z}$, the set of integers modulo 10. The subset of invertible integers modulo 10 is denoted by $(\mathbb{Z}/10\mathbb{Z})^\times$. Extract the multiplication table for $(\mathbb{Z}/10\mathbb{Z})^\times$. How many invertible elements do we have here?

5) Write down the multiplication table for $\mathbb{Z}/12\mathbb{Z}$, the set of integers modulo 12. The subset of invertible integers modulo 12 is denoted by $(\mathbb{Z}/12\mathbb{Z})^\times$. Extract the multiplication table for $(\mathbb{Z}/12\mathbb{Z})^\times$. How many invertible elements do we have here?

6) Use the Euclidean Algorithm to compute the multiplicative inverse of 131 modulo 1979.

7) Use the previous problem to solve

$$131x \equiv 11 \pmod{1979}.$$

8) Use the Euclidean Algorithm to compute the multiplicative inverse of 127 modulo 1091.

9) Use the previous problem to solve

$$127x \equiv 11 \pmod{1091}.$$

4. Theorem of Lagrange

Fix a group G with an identity element e. Let g be any element in G. For every positive integer n, define $g^n = g \cdots g$ where we have n factors. The order of an element g is the smallest positive integer k such that

$$g^k = e.$$

The case $k = \infty$ is allowed. In that case we say that g has infinite order. The order of g is denoted by $o(g)$. If G is finite, then the number of elements in G is called the order of G. Moreover, every element in a finite group has a finite order. This can be easily seen as follows. Consider (an infinite) sequence

$$g, g^2, g^3, \ldots$$

in G. Since G is finite, the elements in the sequence cannot be all different. Thus $g^n = g^m$ for some $n > m$. After canceling g^m on both sides we get $g^{n-m} = e$ which shows that g has finite order.

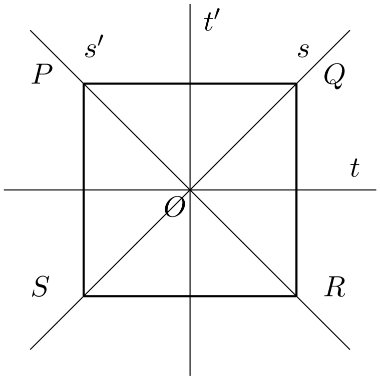
Examples:

(1) $G = \mathbb{Q}^\times$ and $g = 2$. Then the powers of 2 are $2, 2^2, 2^3, \cdots \neq 1$, so 2 has infinite order. Only two rational numbers have a finite order. They are -1 and 1.

(2) $G = (\mathbb{Z}/7Z)^\times$ and $g = 2$. Then, modulo 7, the powers of 2 are $1, 2, 2^2 = 4, 2^3 = 1$. Hence the order of 2 is 3. One can easily tabulate orders of all elements in G, and they are given by the following table:

g	$o(g)$
1	1
2	3
3	6
4	3
5	6
6	2

(3) Let G be the group of all isometries of a square. It contains 4 rotations, with angles 0, 90, 180 and 270 degrees. For sake of definiteness, we assume that the rotations are in counter-clockwise direction, and r_φ will denote the rotation for angle φ. In addition, we have four axial symmetries as seen on the picture. We have eight isometries in all.



The group law is the composition of isometries $\circ$. As an example, we shall calculate $s \circ t$. The idea is to pick a triangle, for example $\Delta(P, O, Q)$, and then act on its vertices by t and s. This gives:

$$s \circ t(P) = s(t(P)) = s(S) = S$$

and

$$s \circ t(Q) = s(t(Q)) = s(R) = P.$$

Since $s \circ t(O) = O$, we conclude that $s \circ t$ moves the triangle $\Delta(P, O, Q)$ to the triangle $\Delta(S, O, P)$. Since any isometry is completely determined by its action on any triangle, it follows that $s \circ t = r_{90}$. A similar calculation shows that $t \circ s = r_{270}$. In particular, the group here is not commutative. The orders of elements are tabulated in the following table:

g	$o(g)$
r_0	1
r_{90}	4
r_{180}	2
r_{270}	4
s	2
s'	2
t	2
t'	2

Next, observe from the tables, that the order of the group (6 and 8 in the last two examples, respectively) is divisible by the order of any group element. This is not an accident. In fact, we have the following important result:

THEOREM 10. *(Theorem of Lagrange) Let G be a finite group. Then the order of any element g in G divides the order of G. In particular, if $|G|$ denotes the order of G, then $g^{|G|} = e$.*

PROOF. Assume that the order of g is k. In order to prove that k divides the order of G we shall arrange all elements of the group in a rectangle with k columns as follows. As the first row, we write down all possible (different) powers of g:

$$e \quad g \quad g^2 \quad \dots \quad g^{k-1}$$

If this row contains all elements of G, then $|G| = k$ and we are done. If not then there exists an element x in G which is not in the first row. Then we can write the second row of elements in G by multiplying x by all possible powers of g. This gives a rectangular table

$$\begin{array}{ccccc} e & g & g^2 & \dots & g^{k-1} \\ x & xg & xg^2 & \dots & xg^{k-1} \end{array}$$

Since $xg^i \neq xg^j$ if $g^i \neq g^j$, the elements of the second row are all different. Also, the two rows are disjoint. If not, then $xg^i = g^j$, for some integers i and j, which implies that $x = g^{i-j}$. But this is impossible since x was picked so that it was not on the first row. If the two rows account for all elements of G, then $|G| = 2k$ and we are done. Otherwise we pick an element y not on the first two rows, form the third row by

$$y \quad yg \quad yg^2 \quad \dots \quad yg^{k-1}$$

and argue as before to show that the third row contains k distinct elements and that the third row is disjoint from the first two. Since G is finite, this process eventually stops after, say, r steps. Thus, we can arrange all group elements in an $r \times k$ rectangle. It follows that $|G| = rk$. □

We give two applications of the Theorem of Lagrange. Consider the group $(\mathbb{Z}/p\mathbb{Z})^\times$ where p is a prime. The order of this group is $p-1$ thus, Theorem 10 implies that $a^{p-1} = e$ for every a in $(\mathbb{Z}/p\mathbb{Z})^\times$. In terms of integers and congruences, we have

$$a^{p-1} \equiv 1 \pmod{p},$$

for every integer a not divisible by p. This statement is known as the Fermat Little Theorem (FLT). Thus, the FLT, no matter how famous, is just a special case of the theorem of Lagrange!

Let G be any group of order p, where p is a prime. Let g be an element in G different from the identity element e. Then the order of g divides p and is not 1. It follows that the order of g is p, and the powers

$$e, g, g^2, \ldots, g^{p-1}$$

account for all elements in the group. Since multiplying the powers of g amounts to adding exponents modulo p, the group G is essentially the same as the group $\mathbb{Z}/p\mathbb{Z}$.

Exercises

1) Let G be a group and g an element in G of order n. Let m be a positive integer such that $g^m = e$. Show that n divides m. Hint: write $m = qn + r$ with $0 \leq r < n$.

2) Repeat the argument of the Theorem of Lagrange with $G = (\mathbb{Z}/13\mathbb{Z})^\times$ and $g = 5$.

3) Consider the group of symmetries of a square, as in the Example 3 above. Compute the action of all group elements on the four points P, Q, R and S. For example, the action of the group on P is given by

g	r_0	r_{90}	r_{180}	r_{270}	s	s'	t	t'
$g(P)$	P	S	R	Q	R	P	S	Q

4) Use the information obtained in the previous exercise to write down the multiplication table for the group of isometries of a square.

5) Let G be a group and g an element in G of order 9. What is the order of g^3? What is the order of g^2?

6) Let g be a group element such that $g^9 = e$ and $g^{16} = e$ where e is the identity element. Show that $g = e$.

5. Chinese Remainder Theorem

Let m be a positive integer bigger than one, and consider multiplication modulo m. Let $(\mathbb{Z}/m\mathbb{Z})^\times$ be the group of invertible elements, with respect to modular multiplication, in $\mathbb{Z}/m\mathbb{Z}$. One of the goals of this section is to determine the order of this group. Consider, for example, $m = 8$. Then the full multiplication table is

*	1	2	3	4	5	6	7
1	1	2	3	4	5	6	7
2	2	4	6	0	2	4	6
3	3	6	1	4	7	2	5
4	4	0	4	0	4	0	4
5	5	2	7	4	1	6	3
6	6	4	2	0	6	4	2
7	7	6	5	4	3	2	1

Notice that every row (and column) contains 1 or 0, but not both. Thus every number is either invertible or a *zero divisor*. For example, 4 is a zero divisor since

$$2 \cdot 4 \equiv 0 \pmod{8}.$$

Even without looking at the table it is easily seen that zero divisors cannot be invertible. For example, if

$$4x \equiv 1 \pmod{8}$$

for some integer x then, after multiplying both sides of this equation by 2, we get

$$0 \equiv 2 \pmod{8},$$

a contradiction. More generally, if $\gcd(a, m) > 1$ then a is a zero divisor modulo m. On the other hand, if $\gcd(a, m) = 1$, then there exist two integers x and y such that

$$ax + my = 1$$

which shows that a is invertible modulo m. Summarizing, for every m, the set $(\mathbb{Z}/m\mathbb{Z})^\times$ of invertible elements for multiplication in $\mathbb{Z}/m\mathbb{Z}$ is

$$(\mathbb{Z}/m\mathbb{Z})^\times = \{0 \leq a \leq m-1 \mid \gcd(a, m) = 1\}.$$

The order of $(\mathbb{Z}/m\mathbb{Z})^\times$ is denoted by $\varphi(m)$ and is called the "Euler function". It is equal to the number of classes of integers modulo m, relatively prime to m. If $m = 8$ then an integer a is prime modulo 8 if and only if it is odd. There are 4 classes of integers modulo 8 represented by odd numbers. In particular $\varphi(8) = 4$.

The theorem of Lagrange implies the following generalization of the Fermat Little Theorem due to Euler.

Theorem 11. *If a is relatively prime to m, then*

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

Take, for example, $m = 8$ and $a = 3$. Then

$$3^{\varphi(8)} = 3^4 = 81 \equiv 1 \pmod{8}.$$

Our next task is to figure out how to calculate $\varphi(m)$. If $m = p$ is prime, then $\varphi(p) = p - 1$. To compute $\varphi(m)$ in general we will prove the following two properties of φ:

(1) If m and n are relatively prime, then $\varphi(mn) = \varphi(m)\varphi(n)$.
(2) If p is a prime, then $\varphi(p^n) = p^n - p^{n-1} = p^{n-1}(p-1)$.

Since every integer can be factored into a product of primes, these two properties suffice to calculate $\varphi(m)$ for any integer m. Consider, for example, $m = 12$. Then, by the first property,

$$\varphi(12) = \varphi(4)\varphi(3).$$

The second property implies that $\varphi(3) = 3 - 1 = 2$ and $\varphi(4) = 2^2 - 2 = 2$. Thus, $\varphi(12) = 4$. This is correct, since $1, 5, 7$ and 11 are the four integers less than 12 and prime to 12.

To prove (1) we will use the *Chinese Remainder Theorem* (CRT) which says the following. Let m, n be two relatively prime integers. Then for any two integers a, b, the system:

$$\begin{cases} x \equiv a \pmod{m} \\ x \equiv b \pmod{n} \end{cases}$$

has a unique solution x modulo mn or, a unique integer solution such that $0 \leq x < m \cdot n$

The Chinese Reminder Theorem is an unusual type of statement since there is one unknown and two equations. However, note that there are m and n choices for a and b, respectively, which means that there are mn possible systems in all. But mn is also the number of classes modulo mn, which is the number of possible choices for x. Thus, in order to prove the CRT, it is natural to consider proving all possible systems at once. This is done as follows. Consider a mapping

$$i : \mathbb{Z}/mn\mathbb{Z} \to (\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z})$$

defined by $i(x) = (x, x)$ where, in (x, x), the first x is considered modulo m while the second x is considered modulo n.

As a working example, consider the case $m = 4, n = 3$ and the system

$$\begin{cases} x \equiv 2 \pmod{4} \\ x \equiv 1 \pmod{3} \end{cases}$$

So, in this case the map i is defined by reducing twelve elements of $\mathbb{Z}/12\mathbb{Z} = \{0, 1, 2, ..., 11\}$ modulo 4 and 3, respectively. For example $i(7) = (3, 1)$. The complete list is given by the following table:

$$\begin{array}{rcl}
0 & \mapsto & (0,0) \\
1 & \mapsto & (1,1) \\
2 & \mapsto & (2,2) \\
3 & \mapsto & (3,0) \\
4 & \mapsto & (0,1) \\
5 & \mapsto & (1,2) \\
6 & \mapsto & (2,0) \\
7 & \mapsto & (3,1) \\
8 & \mapsto & (0,2) \\
9 & \mapsto & (1,0) \\
10 & \mapsto & (2,1) \\
11 & \mapsto & (3,2) \\
12 & \mapsto & (0,0)
\end{array}$$

In particular, we see that $x = 10$ is a unique solution to our pair of congruences. Thus, in order to prove CRT, it suffices to show that the map i is a bijection (i.e. it is one-to-one and onto) between $\mathbb{Z}/mn\mathbb{Z}$ and $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$.

We shall first show that i is one to one. If $i(x) = i(y)$ then

$$x \equiv y \pmod{m} \text{ and } x \equiv y \pmod{n}$$

or,

$$m|x - y \text{ and } n|x - y.$$

Since m and n are relatively prime, it follows that $mn|(x - y)$ or $x \equiv y \pmod{mn}$. In words, the map i is one to one, as claimed

But, if the map i is one-to-one, it has to be onto since the two sets have the same cardinalities. Thus, as x runs through all integers modulo mn, the pairs (x, x) fill the set $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$.

Although we have just shown that the congruence system has a solution, a natural question is how to explicitly find the solution. For example, consider the system

$$x \equiv 3 \pmod{8}$$

$$x \equiv 5 \pmod{9}.$$

We can find a solution x modulo 72 of this system as follows. Since $x \equiv 3 \pmod{8}$, x must be one of the following numbers:

$$3, 11, 19, 27, 35, 43, 51, 59, 67.$$

Indeed, these are all numbers less then 72 and congruent to 3 modulo 8. These numbers are obtained by adding (multiples of) 8 to 3. By inspection we find that 59 is the only number here congruent to 5 modulo 9.

More abstractly the system can be solved using the following two steps. First, any number of the form $x = 3 + 8y$ solves the first equation. Next, substitute $x = 3 + 8y$ into the second equation. This gives

$$3 + 8y \equiv 5 \pmod 9$$

and

$$8y \equiv 2 \pmod 9.$$

Since 8 is relatively prime to 9 we can solve this equation by multiplying both sides of the equation by the inverse of 8 modulo 9. The inverse is 8, thus $y \equiv 8 \cdot 2 \equiv 7 \pmod 9$, and

$$x = 3 + 8y = 3 + 8 \cdot 7 = 59.$$

Using the map i we can prove that $\varphi(m, n) = \varphi(m)\varphi(n)$ for relatively prime integers m and n. To that end, note that $\varphi(mn)$ is the number of elements in $\mathbb{Z}/mn\mathbb{Z}$ prime to x, while $\varphi(m)\varphi(n)$ is equal to the number of pairs (a, b) in $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ such that a is prime to m and b is prime to n. Since any pair (a, b) is given by $i(x) = (x, x)$ for some x in $\mathbb{Z}/mn\mathbb{Z}$, it is clear that x is prime to mn if and only if $a \equiv x$ is prime to m and $b \equiv x$ is prime to n. It follows that i gives a bijection

$$i : (\mathbb{Z}/mn\mathbb{Z})^\times \to (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times.$$

This completes the proof of the first property of Euler's function:

$$\varphi(mn) = \varphi(m)\varphi(n).$$

The proof of the second property of φ is is simple. Indeed, if $0 \le a \le p^n - 1$ is prime to p^n then p does not divide a. Thus, in order to count a relatively prime to p^n, we need to throw away those that are multiples of p. But the multiples of p are

$$0, p, 2p, \ldots, p^n - p,$$

which is p^{n-1} elements in all. Thus $\varphi(p^n) = p^n - p^{n-1}$, as claimed.

As an illustration of what we have done so far we show how to determine the last two digits of a large power. For example, the number 3^{83} is too large for any calculator, but we can determine its last two digits as follows. Note that any number is congruent modulo 100 to the number given by the last two digits, for example, $523412 \equiv 12 \pmod{100}$. Thus we need to compute 3^{83} modulo 100. Since $100 = 2^2 \cdot 5^2$, it follows that $\varphi(100) = \varphi(2^2)\varphi(5^2) = (2^2 - 2)(5^2 - 5) = 40$, and $a^{40} \equiv 1 \pmod{100}$ for any integer a relatively prime to 100. Thus

$$3^{83} = 3^{80} \cdot 3^3 \equiv 3^3 \equiv 27 \pmod{100}.$$

Exercises

1) Let m and n be two relatively prime integers. If m and n divide a show that mn divides a.

2) Put $\varphi(1) = 1$. Compute

$$\sum_{d|1000} \varphi(d)$$

where the sum is taken over all divisors d of 1000 including 1 and 1000.

3) Solve the system of congruences

$$x \equiv 5 \pmod{11}$$
$$x \equiv 7 \pmod{13}$$

4) Solve the system of congruences

$$x \equiv 11 \pmod{16}$$
$$x \equiv 16 \pmod{27}$$

6) Find the last two digits of 3^{9999}.

7) Find the last two digits of 2^{9999}. (Note that 2 is not relatively prime to 100.) Hint: Compute 2^{9999} modulo 25. Why is this enough?

8) Compute 3^{25} modulo 45.

CHAPTER 3

Rings and Fields

1. Fields and Wilson's Theorem

A ring is a set R with two binary operations, addition and multiplication (traditionally denoted by $+$ and $\cdot$, respectively) satisfying the following conditions:

(1) The set R is a group for addition. The unique identity element for the addition operation is denoted by 0.
(2) The addition operation is commutative.
(3) The multiplication operation is associative. There exists an identity element denoted by 1. We assume that $1 \neq 0$.
(4) The two operations are related by the distributive law:

$$a \cdot (b + c) = a \cdot b + a \cdot c,$$

$$(b + c) \cdot a = b \cdot a + c \cdot a.$$

If, in addition, the multiplication is commutative then we say that the ring R is commutative. Examples of rings include the rings of integers $\mathbb{Z}$, integers modulo m, rational numbers $\mathbb{Q}$, real numbers $\mathbb{R}$ and complex numbers $\mathbb{C}$. These rings are all commutative. An example of a non-commutative ring is the ring of all 2×2 matrices with integer coefficients.

The set of elements in R invertible with respect to multiplication is denoted by $R^\times$. If a and b are two invertible elements then the product $a \cdot b$ is also invertible since

$$(a \cdot b)^{-1} = b^{-1} \cdot a^{-1}.$$

The set of invertible elements $R^\times$ is a group with respect to multiplication. For example, $\mathbb{Z}^\times = \{-1, 1\}$, $(\mathbb{Z}/m\mathbb{Z})^\times$ consist of all classes of integers modulo m that are relatively prime to m and $\mathbb{Q}^\times$ consists of all non-zero rational numbers.

We note that 0 is not invertible in any ring R. This can be easily seen as follows. If r is any element in R then, by the distributive property,

$$r \cdot 0 = r \cdot (0 + 0) = r \cdot 0 + r \cdot 0.$$

Since we can cancel $r \cdot 0$ from both sides, we conclude that $r \cdot 0 = 0$. In words, multiplying 0 by any element in R gives always 0. Since 0 is different from 1, 0 cannot be invertible.

A commutative ring R is, furthermore, a *field* if every element different from 0 has a multiplicative inverse. Examples of fields include

$$\mathbb{Z}/p\mathbb{Z},\ \mathbb{Q},\ \mathbb{R},\ \mathbb{C},$$

where p is a prime. We point out that the finite field $\mathbb{Z}/p\mathbb{Z}$ is often denoted by $\mathbb{F}_p$. We shall use both notations.

If R is a commutative ring, then we can manipulate expressions involving multiplication and addition in the usual fashion. For example, the identity

$$(a+b)^2 = a^2 + 2ab + b^2$$

holds, and this can be verified using the ring axioms as follows:

$$\begin{aligned}
&(a+b)(a+b) = && \\
&= (a+b)a + (a+b)b && \text{distributive law} \\
&= (a^2+ba) + (ab+b^2) && \text{distributive law} \\
&= a^2 + (ba+ab) + b^2 && \text{associativity for } + \\
&= a^2 + 2ab + b^2 && \text{commutativity for } \cdot
\end{aligned}$$

Starting from a commutative ring R we can construct a new ring: the ring of polynomials with coefficients in R. More precisely, let $R[x]$ denote the set of all polynomials

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

with coefficients $a_n, a_{n-1}, \ldots, a_1, a_0$ in R. Then $R[x]$ is a ring with respect to the usual addition and multiplication of polynomials. If $a_n \neq 0$ then the degree of the polynomial f is n. Finding roots of polynomials is a central theme in number theory. Finding roots of a degree one polynomial amounts to solving a linear equation $ax = b$ where a and b are elements in R. This is done by "dividing" by a or, more precisely, multiplying both sides of the equation by an inverse of a with respect to multiplication, provided that such an inverse exists. Consider, for example, the ring $R = \mathbb{Z}/34\mathbb{Z}$ and there linear equations

$$7x \equiv 4 \pmod{34},$$

$$6x \equiv 7 \pmod{34},$$

$$6x \equiv 4 \pmod{34}.$$

Since 5 is the multiplicative inverse of 7 modulo 34, we can multiply both sides of the first equation by 5 to obtain

$$x \equiv 5 \cdot 4 \pmod{34}.$$

This shows that $5 \cdot 4 = 20$ is a (unique) solution of the first equation. Since 6 is not invertible modulo 34 ($\gcd(6, 34) = 2$) the approach used for the first equation will not work for the other two equations. We claim that the first equation has no solutions, while the third has two. Indeed, If x is a solution of the second equation, then $6x - 7 = 34k$ for some integer k. Since 2 divides both 6 and 34, it follows that 2 divides 7 and this is a contradiction. Thus the second equation has no solution. The third equation

$$6x \equiv 4 \pmod{34}$$

can be rewritten as

$$2(3x-2) \equiv 0 \pmod{34}.$$

This equation has two solutions, x_1 and x_2, given by (unique) solutions of the following two congruences

$$\begin{cases} 3x_1 - 2 \equiv 0 \pmod{34} \\ 3x_2 - 2 \equiv 17 \pmod{34}. \end{cases}$$

Notice that the identity $2 \cdot 17 = 0$ in $\mathbb{Z}/34\mathbb{Z}$ is responsible for the fact that we have two solutions of a linear (degree one) equation. More generally, assume that in a ring R we have two two non-zero elements a and b such that

$$ab = 0.$$

Such numbers are called *zero divisors*. Then the equation $ax = 0$ clearly has at least two solutions: $x = 0$ and b.

PROPOSITION 12. *Fields have no zero divisors, that is, if $a \cdot b = 0$ and $a \neq 0$ then $b = 0$.*

PROOF. Since $a \neq 0$, there exists a multiplicative inverse a^{-1}. Then

$$a \cdot b = 0 \Rightarrow a^{-1}(a \cdot b) = a^{-1} \cdot 0 \Rightarrow (a^{-1} \cdot a)b = 0 \Rightarrow b = 0.$$

□

As the next example, consider the quadratic equation

$$x^2 - 1 = (x+1)(x-1) = 0.$$

If the ring R has no zero divisors, which is true if R is a field, then the product $(x+1)(x-1)$ is zero if and only if one of the two factors is 0. This implies that $x = 1$ or -1. Otherwise, this equation can have more than one solution as the example $R = \mathbb{Z}/8\mathbb{Z}$ shows. The solutions of

$$x^2 - 1 = 0$$

in $\mathbb{Z}/8\mathbb{Z}$ are 3 and -3, in addition to obvious 1 and -1. Indeed, if $x = 3$, then

$$(x+1)(x-1) = 4 \cdot 2 = 8 \equiv 0 \pmod{8}.$$

Again, the presence of zero divisors in $\mathbb{Z}/8\mathbb{Z}$ is responsible for additional solutions of the equation $x^2 - 1 = 0$.

As an application of our study of the equation $x^2 - 1 = 0$, we shall now derive Wilson's theorem which says that

$$(p-1)! \equiv -1 \pmod{p}$$

for every odd prime p. Another way to state this theorem is that $(p-1)! = -1$ holds in $\mathbb{Z}/p\mathbb{Z}$. Here, as usual, $(p-1)! = 1 \cdot 2 \cdot \dots \cdot (p-1)$. The proof of this fact relies on the fact that $\mathbb{Z}/p\mathbb{Z}$ is a field. In particular, $x^2 - 1$ has only two roots in $\mathbb{Z}/p\mathbb{Z}$: 1 and -1.

Before proceeding to the general case consider the case of $p = 7$. The multiplication table in $(\mathbb{Z}/7\mathbb{Z})^\times$ is shown below:

*	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	1	3	5
3	3	6	2	5	1	4
4	4	1	5	2	6	3
5	5	3	1	6	4	2
6	6	5	4	3	2	1

The table clearly shows the multiplicative inverse of each element: 2 is the inverse of 4 and 3 is the inverse of 5. Thus

$$1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 = (1 \cdot 6) \cdot (2 \cdot 4) \cdot (3 \cdot 5) = (-1) \cdot (1) \cdot (1) = -1$$

where we have used that $6 \equiv -1 \pmod 7$.

This example is quite illustrative since it shows that cancellations in $(p-1)!$ come from pairing elements with their inverses. However, as it can also be seen in the example, some elements are their own inverses. Such elements satisfy the equation $x^2 = 1$. It follows that

$$(p-1)! = \prod_{x^2=1} x.$$

Since $\mathbb{Z}/p\mathbb{Z}$ is a field, the equation $x^2 = 1$ has only two solutions: $x = 1$ and $x = -1$. It follows that $(p-1)! = -1$, as desired.

Exercises

1) Solve $x^{62} - 16 = 0$ in $\mathbb{Z}/31\mathbb{Z}$. Hint: use the Fermat Little Theorem to reduce the exponent.

2) Solve $19x - 11 = 0$ in $\mathbb{Z}/31\mathbb{Z}$.

3) Solve $13x - 11 = 0$ in $\mathbb{Z}/31\mathbb{Z}$.

4) Solve the following three equations in the ring $\mathbb{Z}/30\mathbb{Z}$.

$$21x - 24 = 0,$$

$$24x - 11 = 0,$$

$$11x - 24 = 0.$$

5) Let R be a ring and let -1 denote the inverse of 1 for addition. Show that, for every element r,

$$(-1) \cdot r = -r$$

where $-r$ is the inverse of r for addition. Hint: use $0 \cdot r = 0$.

2. Field Characteristic and Frobenius

Let F be a field. It is a set with two operations $+$ (addition) and $\cdot$ (multiplication) which satisfy the set of axioms given in the previous lecture. In particular, the field F has at least two elements 0 and 1. Every positive integer n can be identified with an element $\underline{n}$ of F defined by

$$\underline{n} = \underbrace{1 + \cdots + 1}_{n-\text{times}}.$$

Note that

$$\underline{n} + \underline{m} = \underline{n + m}.$$

Moreover, it follows from the distributive property in F that the numbers $\underline{n}$ and $\underline{m}$ multiply by the formula

$$\underline{n} \cdot \underline{m} = (\underbrace{1 + \cdots + 1}_{n-\text{times}})(\underbrace{1 + \cdots + 1}_{m-\text{times}}) = \underbrace{1 + \cdots + 1}_{nm-\text{times}} = \underline{nm}.$$

This shows that the elements $\underline{n}$ add and multiply in the same way as integers. In particular, it is safe to write n instead of $\underline{n}$, and we shall do so when it causes no confusion.

We have now two possibilities. Either $\underline{n} \neq 0$ for every positive integer n, or there exists a positive integer n such that $\underline{n} = 0$ in F. In the first case we say that the field F has *characteristic 0*. Examples of such fields are $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$. In the second case we say that the field F has a positive characteristic or, more precisely, *characteristic* p where p is the smallest positive integer such that $\underline{p} = 0$. For example, $F = \mathbb{Z}/3\mathbb{Z}$ has the field characteristic 3 since

$$1 + 1 + 1 = 0$$

in $\mathbb{Z}/3\mathbb{Z}$. More generally, if p is a prime, then the finite field $\mathbb{Z}/p\mathbb{Z}$ has the characteristic p. We have the following important observation:

The field characteristic is either 0 *or a positive prime number* p.

This is really easy. Assume, for example, that $6 = 1{+}1{+}1{+}1{+}1{+}1 = 0$ in a field F. This equation can be rewritten as

$$0 = 1 + 1 + 1 + 1 + 1 + 1 = (1 + 1 + 1)(1 + 1).$$

Since the field F does not have zero divisors we must have either $1{+}1{+}1 = 0$ or $1{+}1 = 0$ in F, that is, the characteristic of F is 2 or 3. Of course, the same argument can be repeated for every composite n such that $\underline{n} = 1{+}\cdots{+}1 = 0$. This shows that the field characteristic must be 0 or a prime number.

Another important observation of this discussion is that if the characteristic of the field F is 0, then the ring of integers $\mathbb{Z}$ can be viewed as a subring of F. If the characteristic of F is p then F contains $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ as a sub-field.

Proposition 13. *(5-th grader's dream) Let F be a field of characteristic p. Then for any two elements a and b in F we have*

$$(a+b)^p = a^p + b^p.$$

Proof. The p-th power of $a+b$ can be expressed in terms of binomial coefficients:

$$(a+b)^p = a^p + \left(\begin{array}{c} p \\ 1 \end{array}\right) a^{p-1}b + \left(\begin{array}{c} p \\ 2 \end{array}\right) a^{p-2}b^2 + \cdots + \left(\begin{array}{c} p \\ p-1 \end{array}\right) ab^{p-1} + b^p.$$

The binomial coefficients are computed using the Pascal triangle. In the characteristic p, the coefficients are computed modulo p. For example, if $p = 7$, then the first eight rows of the Pascal triangle calculated modulo 7 are.

$$\begin{array}{ccccccccccccccc}
&&&&&&&1&&&&&&& \\
&&&&&&1&&1&&&&&& \\
&&&&&1&&2&&1&&&&& \\
&&&&1&&3&&3&&1&&&& \\
&&&1&&4&&6&&4&&1&&& \\
&&1&&5&&3&&3&&5&&1&& \\
&1&&6&&1&&6&&1&&6&&1& \\
1&&0&&0&&0&&0&&0&&0&&1
\end{array}$$

Here the first 5 rows coincide with the usual Pascal triangle. The first difference appears in the sixth row, where instead of 10 (twice) we have $3 \equiv 10 \pmod 7$. The vanishing of coefficients modulo 7 in the 8-th row or, more generally, vanishing of coefficients modulo p in the $p+1$-st row can be easily explained. Recall that the binomial coefficients are *integers* and given by the following formula:

$$\left(\begin{array}{c} p \\ k \end{array}\right) = \frac{p!}{n!(p-n)!}.$$

Since p is prime, both factors in the denominator do not contain p as a factor, as they are products of numbers less than p. On the other hand, the numerator is divisible by p. Thus, the quotient of the numerator and the denominator is still divisible by p. It follows that the binomial coefficients are 0, when considered as elements of the field F. This completes the proof. □

The above proposition shows that the map $\mathrm{Fr}(x) = x^p$, also called the "Frobenius map", is rather interesting, since

$$\mathrm{Fr}(ab) = \mathrm{Fr}(a) \cdot \mathrm{Fr}(b)$$

and

$$\mathrm{Fr}(x+y) = \mathrm{Fr}(x) + \mathrm{Fr}(x)$$

which implies that Fr is a *homomorphism* for both group structures at the same time. In this sense it is similar to the conjugation of complex numbers. But that is not all. Recall that if x is a complex number such that $x = \bar{x}$, then x is a real number. We have a similar statement for fields of characteristic p.

Proposition 14. *Let F be a field of characteristic p. Let x be an element in F. Then* $\mathrm{Fr}(x) = x$ *if and only if x is in the subfield* $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ *of F.*

Proof. The equation $\mathrm{Fr}(x) = x$ can be written as

$$x^p - x = 0.$$

Thus, an element of the field F satisfies $\mathrm{Fr}(x) = x$ if and only if it is a root of the polynomial $x^p - x$. By the Little Fermat Theorem, all elements of $\mathbb{F}_p$ are roots of this polynomial. In this way we have already accounted for p roots of $x^p - x$. Since, over any field, a polynomial of degree p cannot have more then p roots, the elements of $\mathbb{F}_p$ are precisely all roots of $x^p - x$. □

We shall now construct some additional examples of finite fields. Complex numbers $z = x + iy$ such that x and y are integers are called Gaussian integers. Here, of course, $i^2 = -1$. Gaussian integers also admit modular arithmetic. If n is a positive integer, two Gaussian integers are said to be congruent modulo n

$$a + bi \equiv c + di \pmod{n}$$

if $a \equiv c \pmod{n}$ and $b \equiv d \pmod{n}$. Addition and multiplication are performed just as with the complex numbers, except the coefficients x and y are always considered modulo n. For example, if $n = 11$ then

$$(2 + 5i)(5 + 4i) = -10 + 33i \equiv 1 + 0i \pmod{11}.$$

If p is a prime such that $p \equiv 3 \pmod{4}$ then the set of Gaussian integers modulo p is a field of characteristic p which is usually denoted by $\mathbb{F}_{p^2}$. Since x and y are integers modulo p, there are p choices for both, x and y. In all, we have p^2 elements in $\mathbb{F}_{p^2}$. This counting principle can be illustrated as follows. There are 100 two-digit numbers: $00, 01, \ldots, 99$. There are 10 choices for the first digit and 10 choices for the second digits. In all, we can write down

$$10 \times 10 = 100$$

numbers using precisely two digits. If $a + bi$ is a Gaussian integer then, modulo p, we have

$$(a + bi)^p \equiv a^p + b^p i^p \pmod{p}.$$

By Fermat's Little Theorem $a^p \equiv a \pmod{p}$ and $b^p \equiv b \pmod{p}$. It remains to compute i^p. Since $p \equiv 3 \pmod{4}$, we can write $p = 4k + 3$. Since $i^4 = 1$ and $i^3 = -i$ it follows that

$$i^p = i^{4k+3} = i^{4k} \cdot i^3 = -i.$$

Summarizing, we have shown that

$$(a+bi)^p \equiv a-bi=\overline{a+bi} \pmod p$$

making the analogy of between the Frobenius map and the complex conjugation rather convincing.

Exercises

1) Build the first 12 rows of the Pascal triangle modulo 11.

2) Prove that

$$\left(\begin{array}{c} n+1 \\ k \end{array}\right)=\left(\begin{array}{c} n \\ k \end{array}\right)+\left(\begin{array}{c} n \\ k-1 \end{array}\right).$$

3. Quadratic Numbers

Quadratic integers are solutions of quadratic equations $x^2+px+q=0$ such that p and q are integers. For example, $2+\sqrt{3}$ is a solution of

$$x^2-4x+1=0.$$

Thus it is a quadratic integer. Solutions of quadratic equations such that p and q to are rational numbers are called *quadratic rationals.*

Quadratic integers and rationals can be grouped together to form rings and fields. For example, let $\mathbb{Z}[\sqrt{3}]$ be the set of all real numbers $a+b\sqrt{3}$ such that a and b are integers. Similarly, let $\mathbb{Q}[\sqrt{3}]$ be the set of all real numbers $a+b\sqrt{3}$ such a and b are rational numbers. Both $\mathbb{Z}[\sqrt{3}]$ and $\mathbb{Q}[\sqrt{3}]$ are closed under addition (clearly) and under multiplication, since

$$(a+b\sqrt{3})(c+d\sqrt{3})=(ac+3bd)+(ad+bc)\sqrt{3}.$$

The elements of $\mathbb{Z}[\sqrt{3}]$ and $\mathbb{Q}[\sqrt{3}]$ are (usual) real numbers, so the ring axioms are inherited. More exciting, however, is to verify that $\mathbb{Q}[\sqrt{3}]$ is a field. This can be checked as follows. Consider, for example $\alpha=7+3\sqrt{3}$. Then

$$\alpha^{-1}=\frac{1}{7+3\sqrt{3}}=\frac{1}{7+3\sqrt{3}}\cdot\frac{7-3\sqrt{3}}{7-3\sqrt{3}}=\frac{7}{22}-\frac{3}{22}\sqrt{3}.$$

The real issue here was whether the inverse of α, which clearly exists since α is a non-zero real number, is contained in $\mathbb{Q}[\sqrt{3}]$. The trick here was provided by so-called rationalization of the denominator, which we all learn in grade school. Of course this works for general $\alpha=a+b\sqrt{3}$. Let $\bar{\alpha}=a-b\sqrt{3}$ be the *conjugate* of α. Then

$$\alpha^{-1}=\frac{1}{\alpha}=\frac{1}{\alpha}\cdot\frac{\bar{\alpha}}{\bar{\alpha}}=\frac{a-b\sqrt{3}}{a^2-3b^2}.$$

The number $\alpha\bar{\alpha} = a^2 - 3b^2$ is abbreviated by $N(\alpha)$ and called the *norm* of α. Note that $N(\alpha)$ is not zero unless α is 0. Indeed, $a^2 - 3b^2 = 0$ implies that

$$\sqrt{3} = \frac{a}{b} \text{ or } \sqrt{3} = -\frac{a}{b}$$

which is not possible since $\sqrt{3}$ is not a rational number. This shows that $\mathbb{Q}[\sqrt{3}]$ is a field.

The ring $R = \mathbb{Z}[\sqrt{3}]$ admits modular arithmetic just as the ordinary integers. More precisely, we say that

$$a + b\sqrt{3} \equiv c + d\sqrt{3} \pmod{p}$$

if

$$\begin{cases} a \equiv c \pmod{p} \\ b \equiv d \pmod{p}. \end{cases}$$

Clearly, modulo p, there are p^2 classes of elements of $\mathbb{Z}[\sqrt{3}]$. This set will be denoted by R/pR, by analogy with $\mathbb{Z}/p\mathbb{Z}$. Again, we have a question whether an element $a+b\sqrt{3}$ in R/pR has a multiplicative inverse. The same formula for inverse works here: the inverse of $a+b\sqrt{3}$ is given by dividing the conjugate $a - b\sqrt{3}$ by the norm $a^2 - 3b^2$. In particular, $a + b\sqrt{3}$ is invertible if and only if $a^2 - 3b^2$ is an integer invertible modulo p.

For example, consider $5 + 3\sqrt{3}$ modulo 7. The norm of $5 + 3\sqrt{3}$ is -2. The multiplicative inverse of -2 modulo 7, is -4. Thus the inverse of $5+3\sqrt{3}$ is

$$(5 - 3\sqrt{3})(-4) = -20 + 12\sqrt{3} \equiv 1 + 5\sqrt{3} \pmod{7}.$$

Indeed,

$$(5 + 3\sqrt{3})(1 + 5\sqrt{3}) = 50 + 28\sqrt{3} \equiv 1 \pmod{7}.$$

As the next example, consider $4 + 3\sqrt{3}$ modulo 11. The norm of $4 + 3\sqrt{3}$ is $4^2 - 3 \cdot 3^2 = -11 \equiv 0 \pmod{11}$. In particular, the norm is not invertible modulo 11. Moreover,

$$(4 + 3\sqrt{3})(4 - 3\sqrt{3}) \equiv 0 \pmod{11}$$

and this shows that $4+3\sqrt{3}$ is a zero divisor. Zero divisors are not invertible, thus $4 + 3\sqrt{3}$ does not have a multiplicative inverse in $R/11R$.

The two examples illustrate the following dichotomy for quadratic integers modulo a prime p: Either the norm is relatively prime to p, and the integer is invertible, or the norm is divisible by p, and the integer is a zero divisor. This observation is useful in determining the order of the group of invertible elements of R/pR which is, as usual, denoted by $(R/pR)^{\times}$. The answer depends on whether 3 is a square modulo p.

Theorem 15. *Let $R = \mathbb{Z}[\sqrt{3}]$ and p an odd prime different from 3. Then:*

(1) *If 3 is not a square modulo p then the order of $(R/pR)^{\times}$ is $p^2 - 1$.*

(2) *If* 3 *is a square modulo p then the order of $(R/pR)^\times$ is $(p-1)^2$. In particular, if* 3 *is not a square modulo p then R/pR is a field with p^2 elements.*

Proof. The proof of this theorem is not difficult at all! If $x^2 - 3y^2 \equiv 0 \pmod{p}$, for some $x + y\sqrt{3}$ in R/pR then

$$3 \equiv (xy^{-1})^2 \pmod{p}$$

which implies that 3 is a square modulo p. Thus, if 3 is not a square modulo p then the norm is invertible modulo p for every non-zero α. It follows that α^{-1} exists for every non-zero α. This takes care of the first case, when 3 is not a square modulo p. In the second case we need to count all elements with norm congruent to 0 modulo p. Since 3 is a square modulo p, then

$$3 \equiv s^2 \pmod{p}$$

for some integer s, and the norm $x^2 - 3y^2$ can be factored as

$$x^2 - 3y^2 \equiv (x - sy)(x + sy) \pmod{p}.$$

Since $\mathbb{Z}/p\mathbb{Z}$ has no zero divisors, the norm is 0 modulo p if and only if

$$x \equiv sy \pmod{p} \text{ or } x \equiv -sy \pmod{p}.$$

These are linear equations and are easy to solve. Each has p solutions. (x is determined once we pick y, and there are p choices for y.) The two equations have one solution in common, namely $(x, y) = (0, 0)$. Thus we have $2p - 1$ non-invertible elements, and $p^2 - (2p - 1) = (p - 1)^2$ invertible elements. □

Consider $p = 7$, for example. Then 3 is not a square modulo 7 as the following table shows:

x	1	2	3	4	5	6
x^2	1	4	2	2	4	1

It follows that $R/7R$ is a field. If $p = 11$, then $5^2 \equiv 3 \pmod{11}$, and 3 is a square modulo 11. Thus $x + y\sqrt{3}$ is not invertible modulo 11 if $x = 5y$ or $-5y$. Here is the list of all non-invertible elements in $R/11R$, twenty one in all, since 0 appears twice on the list.

y	0	1	2	3	4	5	6	7	8	9	10
x	0	5	10	4	9	3	8	2	7	1	6
x	0	6	1	7	2	8	3	9	4	10	5

Of course, the number 3 in the previous discussion can be replaced by any non-square integer D, positive or negative. For example, if $D = -1$, then $\mathbb{Z}[\sqrt{D}] = \mathbb{Z}[i]$, is the ring of Gauss's integers. We record:

Theorem 16. *Let $R = \mathbb{Z}[\sqrt{D}]$, where D is a non-square integer. Let p be an odd prime not dividing D.*

(1) *If D is not a square modulo p then the order of $(R/pR)^{\times}$ is $p^2 - 1$.*
(2) *If D is a square modulo p then the order of $(R/pR)^{\times}$ is $(p-1)^2$.*

In particular, if D is not a square modulo p then R/pR is a field with p^2 elements.

Exercises

1) Find a multiplicative inverse of $7 - 3\sqrt{5}$ modulo 11 and then modulo 17. In each case verify that the answer is correct.

2) List all non-invertible modulo 11 quadratic integers of the type $a + b\sqrt{5}$.

3) Find the inverse of $2 + 5i$ modulo 31. Is there an inverse of $2 + 5i$ modulo 29?

4) Is $\frac{1+\sqrt{5}}{2}$ (the golden mean) a quadratic integer?

5) Describe all quadratic integers of the type $a + b\sqrt{5}$.

CHAPTER 4

Primes

1. Infinitude of primes

In this and several successive lectures we focus our attention on primes. We start by giving two proofs that there exist infinitely many primes. The first and elementary proof involves a method of constructing a new prime, from a given list of primes. More precisely, let $S = \{p_1, p_2, \dots, p_n\}$ be a set of primes and consider the number

$$m = p_1 \cdot p_2 \cdot \ldots \cdot p_n + 1.$$

Since $m \equiv 1 \pmod{p_i}$ for any p_i in S, the number m is not divisible by p_i. Thus any prime divisor q of m is not in S. For example, if $S = \{2, 3, 5\}$, then

$$n = 2 \cdot 3 \cdot 5 + 1 = 31,$$

which is a prime already, different from 2, 3 and 5. In particular, no matter how many primes we write down, we can always construct more. Thus, there are infinitely many primes.

This proof is elementary and cute, but it has a drawback. It can not be generalized easily to primes in progressions. We shall give another proof that there are infinitely many primes using analysis! The central role in arguments here is played by the Riemann zeta function

$$\zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \ldots$$

Note that $\zeta(1)$ is the (divergent) harmonic series while $\zeta(2), \zeta(3) \ldots$ are its convergent analogues that you have studied in Calculus II. These (convergence) properties can be established by comparing $\zeta(s)$ with the (improper) integral of the function $1/x^s$ over $[1, \infty)$. Indeed, forming lower and upper sums corresponding to the subdivision $[0, \infty) = [1, 2] \cup [2, 3] \cup \ldots$ gives

$$\zeta(s) - 1 \leq \int_1^\infty \frac{1}{x^s} \, dx \leq \zeta(s).$$

Since the integral can be easily shown to be equal to $1/(s-1)$ we see that $\zeta(s)$ converges for $s > 1$ and

$$\zeta(1) = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} + \ldots = \infty.$$

We are now ready to give the second proof that there are infinitely many primes. Suppose we have only finitely many primes, say $2, 3$ and 5 only, then every integer n could be factored as $n = 2^s 3^t 5^u$, and the harmonic series could be rewritten as a finite product

$$(1 + \frac{1}{2} + \frac{1}{2^2} + \ldots)(1 + \frac{1}{3} + \frac{1}{3^2} + \ldots)(1 + \frac{1}{5} + \frac{1}{5^2} + \ldots).$$

Each of the three factors is a convergent geometric series that can be easily summed up by the formula

$$1 + \frac{1}{p} + \frac{1}{p^2} + \ldots = \frac{1}{1 - \frac{1}{p}} = \frac{p}{p-1}.$$

In particular, if there were no other primes but only 2 3 and 5, then the harmonic series would converge and be equal to

$$1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} + \ldots = \frac{2}{1} \cdot \frac{3}{2} \cdot \frac{5}{4}.$$

This is a contradiction, since the harmonic series diverges. Clearly, the same argument shows that no finite list of primes is complete.

We now show two identities, which will be useful in the next section, when we study primes in progressions:

$$\prod_p \frac{p}{p-1} = \infty,$$

and

$$\prod_p \frac{p}{p+1} = 0,$$

Note that these products involve infinitely many factors, so some explanation is on order. Just as an infinite sum (a series) is defined as a limit of the sequence of partial sums, an infinite product is also defined to be the limit of partial, finite products. For example, the first infinite product is defined to be

$$\lim_{N\to\infty} \prod_{p \le N} \frac{p}{p-1}.$$

We now show that this limit is infinite. Using the formula for the geometric series, write

$$\prod_{p \le N} \frac{p}{p-1} = \prod_{p \le N} (1 + \frac{1}{p} + \frac{1}{p^2} + \ldots).$$

After multiplying out the product of the geometric series on the right, we get a sum of $\frac{1}{n}$ over all integers n whose factorization into primes involves only primes $p \le N$. In particular, this sum includes all $n \le N$ and we have the following inequality

$$1 + \frac{1}{2} + \ldots + \frac{1}{N} \le \prod_{p \le N} \frac{p}{p-1}.$$

Since the harmonic series diverges, it follows that the product $\prod_p \frac{p}{p-1}$ diverges, as claimed. In order to check that the second infinite product converges to 0, the trick is to multiply the two products

$$\prod_{p\leq N} \frac{p}{p-1} \cdot \prod_{p\leq N} \frac{p}{p+1} = \prod_{p\leq N} \frac{p^2}{p^2-1}.$$

Again, using the formula for the geometric series, write

$$\prod_{p\leq N} \frac{p^2}{p^2-1} = \prod_{p\leq N} (1 + \frac{1}{p^2} + \frac{1}{p^4} + \ldots).$$

After multiplying out the product of the geometric series on the right, we get a sum of $\frac{1}{n^2}$ over all integers n whose factorization into primes involves only primes $p \leq N$. In particular, this sum includes all $n \leq N$ and we have the following inequalities

$$1 + \frac{1}{2^2} + \ldots + \frac{1}{N^2} \leq \prod_{p\leq N} \frac{p^2}{p-1} \leq 1 + \frac{1}{2^2} + \ldots + \frac{1}{n^2} + \ldots = \zeta(2).$$

The limit, as $N \to \infty$, of the sum on the left is also $\zeta(2)$. It follows that

$$\lim_{N\to\infty} \prod_{p\leq N} \frac{p^2}{p^2-1} = \zeta(2) = \frac{\pi^2}{6}$$

and

$$\lim_{N\to\infty} \prod_{p\leq N} \frac{p}{p+1} = 0$$

since this last limit is equal to the quotient $\frac{\zeta(2)}{\infty}$ of the two already computed limits.

Now that we know that there are infinitely many primes, we can ask how often they appear. A quite good answer to this question is given by the Prime Number Theorem, which says that

$$\pi(x) \sim \frac{x}{\ln x}$$

where $\pi(x)$ is the number of primes less than or equal to x. The proof of this results uses techniques of complex analysis applied to the Riemann zeta function.

Exercises

1) In this exercise you will prove that there are infinitely many primes p congruent to 2 modulo 3 using an analogue of the first proof of infinitude of primes. Let $S = \{p_1, \ldots p_m\}$ be any list of odd primes, and consider the number

$$m = 3 \cdot p_1 \cdot \ldots \cdot p_n + 2.$$

Note that m is odd and not divisible by $3, p_1, \ldots, p_n$. Thus, not one of the prime factors of m is in S (or equal to 2 or 3). Show that at least one of the prime factors of m is congruent to 2 modulo 3. Hint: if all prime factors are congruent to 1 modulo 3, what would this imply for m?

2) Show that

$$1 + \frac{1}{2^2} + \frac{1}{3^2} + \ldots + \frac{1}{n^2} + \ldots = \frac{\pi^2}{6}$$

by computing the integral

$$\int_{-\pi}^{\pi} (\frac{x}{2})^2 \; dx$$

in two ways. First directly, and then substituting the Fourier series for $x/2$

$$\frac{x}{2} = \sin x - \frac{\sin 2x}{2} + \frac{\sin 3x}{3} - \ldots.$$

Hint: your calculation can be simplified considerably by knowing that, if $n \neq m$,

$$\int_{-\pi}^{\pi} \sin(nx) \cdot \sin(mx) \; dx = 0.$$

3) Compute the improper integral

$$\int_{1}^{\infty} \frac{1}{x^s} \; dx.$$

2. Primes in progression

If a and d are two relatively prime integers, then a theorem of Dirichlet says that the arithmetic progression $a, a+d, a+2d, \ldots$ contains infinitely many primes. For example, if we take $d = 3$, then the possible a are 1 and 2, and Dirichlet's theorem says that the two sequences of integers

$$1, 4, \mathbf{7}, 10, \mathbf{13}, 16, \mathbf{19}, \ldots$$

and

$$\mathbf{2}, \mathbf{5}, 8, \mathbf{11}, 14, \mathbf{17}, 20 \ldots$$

congruent to 1 and -1 modulo 3 respectively, contain infinitely many primes each.

The theorem of Dirichlet is proved in a way similar to the second proof of the infinitude of primes, given in the previous section. We shall do this for $d = 3$. To that end, for every integer n not divisible by 3, let $\varepsilon(n)$ be 1 or -1 so that

$$n \equiv \varepsilon(n) \pmod 3.$$

The function ε is called a Dirichlet character because it is multiplicative, meaning that

$$\varepsilon(nm) = \varepsilon(n)\varepsilon(m).$$

The proof of this statement is easy. Indeed, multiplying the two congruences

$$\begin{cases} n \equiv \varepsilon(n) \pmod 3 \\ n \equiv \varepsilon(n) \pmod 3 \end{cases}$$

gives $nm \equiv \varepsilon(n)\varepsilon(m) \pmod 3$. On the other hand, from the definition of ε,

$$nm \equiv \varepsilon(nm) \pmod 3.$$

It follows that $\varepsilon(nm)$ is congruent to $\varepsilon(n)\varepsilon(m)$. Since each is equal to 1 or -1, they have to be equal, as integers.

Just as we have used the harmonic series to show that there are infinitely many primes we shall use the Dirichlet L-series

$$L = 1 - \frac{1}{2} + \frac{1}{4} - \ldots + \frac{\epsilon(n)}{n} + \ldots$$

There a couple of key observations to be made here. First, this series is an alternating series. In particular, it converges to a number sandwiched between any two consecutive partial sums, for example

$$1 - \frac{1}{2} < L < 1.$$

Second, since ϵ is multiplicative the series can be factored, similarly as the harmonic series,

$$L = (1 - \frac{1}{2} + \frac{1}{2^2} - \ldots)(1 - \frac{1}{5} + \frac{1}{5^2} - \ldots)(1 + \frac{1}{7} + \frac{1}{7^2} + \ldots) \ldots.$$

Here we have disregarded convergence issues, but for a moment. We will return to it shortly. In the meantime, note that the individual factors are again geometric series. The series with alternating signs appears for primes $p \equiv -1 \pmod 3$ and it is equal to

$$1 - \frac{1}{p} + \frac{1}{p^2} - \ldots = \frac{1}{1 + \frac{1}{p}} = \frac{p}{p+1}.$$

The series with all positive signs appears for primes $p \equiv 1 \pmod 3$ and it is equal to

$$1 + \frac{1}{p} + \frac{1}{p^2} + \ldots = \frac{1}{1 - \frac{1}{p}} = \frac{p}{p-1}.$$

It follows that the Dirichlet L-series is equal to the infinite product

$$L = \frac{2}{3} \cdot \frac{5}{6} \cdot \frac{7}{6} \cdot \frac{11}{12} \cdot \frac{13}{12} \ldots,$$

where for every prime p the factor is either $p/(p-1) > 1$ or $p/(p+1) < 1$, depending whether p is congruent to 1 or -1 modulo 3. Since $0 < L < \infty$, these two types of factors have to balance each other out, and this will imply that both sequences contain infinitely many primes. More precisely, let S^- be the set of all primes congruent to -1 modulo 3. If S^- is finite then, using

$$\frac{p}{p+1} = \frac{p-1}{p+1} \cdot \frac{p}{p-1}$$

for every p in S^-, we can rewrite L as

$$L = (\prod_{p \in S^-} \frac{p-1}{p+1}) \cdot (\prod_{p \neq 3} \frac{p}{p-1}).$$

Since (from the previous section)

$$(\prod_p \frac{p}{p-1}) = \infty$$

it follows that $L = \infty$, a contradiction. Similarly, let S^+ be the set of all primes congruent to $+1$ modulo 3. If S^+ is finite then, using

$$\frac{p}{p-1} = \frac{p+1}{p-1} \cdot \frac{p}{p+1}$$

for every p in S^+, we can rewrite L as

$$L = (\prod_{p \in S^+} \frac{p+1}{p-1}) \cdot (\prod_{p \neq 3} \frac{p}{p+1})$$

Since (from the previous section)

$$(\prod_p \frac{p}{p+1}) = 0$$

it follows that $L = 0$, a contradiction. Thus both sequences contain infinitely many primes, as claimed.

However, since the L-series is not absolutely convergent, the above discussion is problematic. We now briefly sketch how this can be fixed. Consider the Dirichlet L-function

$$L(s) = 1 - \frac{1}{2^s} + \frac{1}{4^s} - \ldots + \frac{\epsilon(n)}{n^s} + \ldots$$

This series is absolutely convergent if $s > 1$. If there are only finitely many primes congruent to -1 modulo 3 then

$$L(s) = *\, \zeta(s),$$

where $*$ stands for a finite product that we leave to the reader to figure out. Similarly, if there are only finitely many primes congruent to 1 modulo 3 then

$$L(s) = *\, \frac{\zeta(2s)}{\zeta(s)}.$$

Passing to the limit $s \to 1$ we get $L(1) = \infty$ in and $L(1) = 0$ in two cases, respectively, again contradicting the fact that $1/2 < L(1) < 1$.

Exercises

1) Write down the L series necessary to show that there are infinitely many primes congruent to 1 modulo 4, and infinitely many primes congruent to -1 modulo 4.

2) Show that

$$\prod_{p\equiv 1(3)} \frac{p}{p-1} = \infty \text{ and } \prod_{p\equiv 2(3)} \frac{p}{p-1} = \infty$$

Hint: if one of the two products is finite, what would that imply for the value of the Dirichlet L-series?

3) For every integer n relatively prime to 5 define

$$\epsilon(n) = \begin{cases} 1 \text{ if } n \equiv 1, 4 \pmod 5 \\ -1 \text{ if } n \equiv 2, 3 \pmod 5. \end{cases}$$

This is a Dirichlet character modulo 5. Show that ϵ is multiplicative, that is, $\epsilon(nm) = \epsilon(n)\epsilon(m)$ for any pair of integers n and m relatively prime to 5. Hint: $\epsilon(n)$ depends only on what n is equal to modulo 5, so this is a case by case verification based on the multiplication table modulo 5:

·	1	2	3	4
1	1	2	3	4
2	2	4	1	3
3	3	1	4	2
4	4	3	2	1

4) Use the Dirichlet character from the previous exercise to define an appropriate L-series to show that there are infinitely many primes congruent to 1 or 4 modulo 5, and that there are infinitely many primes congruent to 2 or 3 modulo 5. Hint: the individual terms in the L-series decrease, and the signs are $+ - - + + - - + + \cdots$ so the value of the series can be as easily estimated as for any alternating series.

3. Perfect Numbers and Mersenne Primes

A number n is "perfect" if it is equal to the sum of its proper divisors (i.e. divisors $d < n$). For example, if $n = 6$ then its divisors are $1, 2, 3$ and

$$1 + 2 + 3 = 6$$

so 6 is perfect. On the other hand, $n = 12$ is not perfect, since its divisors are $1, 2, 3, 4, 6$ and

$$1 + 2 + 3 + 4 + 6 = 16 \neq 12.$$

Perfect numbers were interesting to the ancient Greeks, and they devised a formula for them which involves Mersenne primes. The Greeks knew that if $2^\ell - 1$ is a prime then $2^{\ell-1}(2^\ell - 1)$ is a perfect number. For example, if $\ell = 3$ then $2^3 - 1 = 7$ is prime and $n = 4 \cdot 8 = 28$ is another perfect number. Indeed, the proper divisors of 28 are $1, 2, 4, 7, 14$ and

$$1 + 2 + 4 + 7 + 14 = 28.$$

Much later Euler showed that all even perfect numbers are of this shape:

Proposition 17. *An even number n is perfect if and only if it can be written as $n = 2^{\ell-1}(2^{\ell} - 1)$ so that $2^{\ell} - 1$ is a prime.*

This proposition is proved by means of the sigma function $\sigma(n)$ which is defined as the sum of all divisors of n, including 1 and n. Of course, the number n is perfect if and only if

$$\sigma(n) = 2n.$$

This equation, by itself, is of no use unless we can exploit somehow the function σ. To that end, we shall first establish some properties of the function σ which are quite analogous to those of the Euler Function φ. First of all, if p is prime then it is easy to calculate $\sigma(p^k)$. The divisors of p^k are $1, p^1, p^2, \ldots, p^k$, so

$$\sigma(p^k) = 1 + p + \cdots + p^k = \frac{p^{k+1} - 1}{p - 1}.$$

Next, if m and n are relatively prime then

$$\sigma(mn) = \sigma(m) \cdot \sigma(n).$$

This is not difficult to see, using the uniqueness of factorization. For example if $m = p^k$ and $n = q^l$ for two different primes, then

$$\sigma(m) = 1 + p + \ldots + p^k$$

$$\sigma(n) = 1 + q + \ldots + q^l$$

The product $\sigma(m)\sigma(n)$ is a sum of terms $p^i q^j$ for all $0 \le i \le k$ and $0 \le j \le l$, which are precisely all divisors of $mn = p^k q^l$. This shows that $\sigma(mn) = \sigma(m)\sigma(n)$. For example:

$$\sigma(28) = \sigma(4 \cdot 7) = \sigma(4)\sigma(7) = \frac{2^3 - 1}{2 - 1} \cdot \frac{7^2 - 1}{7 - 1} = 7 \cdot 8 = 2 \cdot 28.$$

With these properties in hand, it is easy to check that the numbers $n = 2^{\ell-1}(2^{\ell} - 1)$ are perfect if $2^{\ell} - 1$ is prime. Indeed, since $2^{\ell-1}$ and $2^{\ell} - 1$ are relatively prime,

$$\sigma(2^{\ell-1}(2^{\ell} - 1)) = \sigma(2^{\ell-1})\sigma(2^{\ell} - 1).$$

Since $2^{\ell} - 1$ is prime,

$$\sigma(2^{\ell} - 1) = 1 + (2^{\ell} - 1) = 2^{\ell}.$$

Since

$$\sigma(2^{\ell-1}) = 1 + 2 + \ldots + 2^{\ell-1} = 2^{\ell} - 1$$

it follows that $\sigma(n) = 2n$, as desired. Euler's converse, showing that all even perfect numbers are of this form, is also not too difficult to check. Write an even perfect number as $n = 2^{\ell-1} \cdot r$ with r odd. Since σ is multiplicative we have the following:

$$\sigma(n) = \sigma(2^{\ell-1}) \cdot \sigma(r) = \frac{2^\ell - 1}{2 - 1} \cdot \sigma(r) = (2^\ell - 1) \cdot \sigma(r).$$

Using this expression for $\sigma(n)$, the equation $\sigma(n) = 2n$ can be rewritten as

$$(2^\ell - 1) \cdot \sigma(r) = 2^\ell r.$$

Since $2^\ell - 1$ and 2^ℓ are relatively prime, $2^\ell - 1$ divides r, so $r = (2^\ell - 1)s$ for some integer s. Therefore, the equation $\sigma(n) = 2n$ can again be rewritten as

$$n = 2^{\ell-1} \cdot r = 2^k(2^\ell - 1)s,$$

which is quite close to what we want to prove. It remains to check that

(1) $s = 1$ and
(2) $2^\ell - 1$ is prime.

Let us prove (1) and (2) at the same time. Since n is even, $\ell > 1$ and hence $r > s$. If $s \neq 1$ then 1 and s are two different and proper divisors of r. Hence

$$\sigma(r) \geq 1 + s + r.$$

On the other hand, $(2^\ell - 1)\sigma(r) = 2^\ell(2^\ell - 1)s$ implies that

$$\sigma(r) = 2^\ell s = r + s.$$

This contradicts the inequality $\sigma(r) \geq 1 + s + r$. Thus $s = 1$ and

$$\sigma(r) = 1 + r$$

which is possible only if r is prime!

Of course, Euler's theorem tells us nothing about odd perfect numbers. There are no known examples of odd perfect numbers. It is still an open problem either to find an odd perfect number or to show that they do not exist. In one of the exercises bellow, you will show that odd perfect numbers cannot be divisible by an odd power of 3.

As we have seen, a prominent role of the description of even perfect numbers is played by Mersenne numbers:

$$M_\ell = 2^\ell - 1.$$

It is easy to see that M_ℓ is composite if ℓ is composite. For example, if $\ell = 6 = 2 \cdot 3$ then $2^6 - 1$ can be factored as

$$2^6 - 1 = (2^2)^3 - 1 = (2^2 - 1)(2^4 + 2^2 + 1) = 3 \cdot 31.$$

On the other hand, if ℓ is a prime, then M_ℓ may or may not be prime. The list of known Mersenne primes is quite short - and, therefore, the list of known perfect numbers is also short - and contains only about forty numbers. The first ten Mersenne primes are the corresponding M_ℓ for

$$\ell = 2, 3, 5, 7, 13, 17, 19, 31, 61, 89, 107, \text{ and } 127.$$

The numbers are named after Marin Mersenne (17th century, France) who gave a list of primes M_ℓ for ℓ up to 257. That list was not correct since it included composite numbers M_{67} and M_{257} and it did not include M_{61}, M_{89} and M_{107}. It is hard to believe that Mersenne was able to prove that M_{127} is prime since M_{127} is a huge number. Indeed, the number of digits of M_{127} is equal to

$$\log_{10}(2^{127} - 1) \approx 127 \cdot \log_{10}(2) \approx 38.$$

In order to factor M_{127} into prime factors, we can start dividing it by primes: 3, 5, 7, and so on. If we fail to find a prime factor less than $\sqrt{M_{127}} \approx 10^{19}$ then we can conclude that M_{127} is prime. By the Prime Number Theorem the total number of primes less than 10^{19} is

$$\pi(10^{19}) = \frac{10^{19}}{\ln(10^{19})} \approx 0.2 \times (\text{billion})^2.$$

It seems that we are destined to perform about billion-billions of increasingly complicated divisions to check that M_{127} is prime! Even if Mersenne could do about thousand divisions a day, and he started right after the BIG BANG (5 billion years ago) he would not be (yet) through one-tenth of one-percent of the verification.

The first rigorous proof of the primality of M_{127} was given by Lucas in 1878. The secret lies in a simple test, discovered by Lucas and simplified by Lehmer, which requires only one division. This test is the topic of our next section.

Exercises

1) Show that odd numbers divisible by an odd power of 3 are not perfect. Hint: write $n = 3^{2k+1}r$ with r prime to 3, and show that $\sigma(3^{2k+1})$ is divisible by 4.

2) The 11-th Mersenne number $2^{11} - 1 = 2047$ is not prime. Find its prime factors.

3) As of September 2006, the largest known and 44-th Mersenne prime is

$$M_{32,582,657}.$$

Assuming that a page (of a book) has typically 40 lines, and each line fits about 40 characters, estimate the number of pages needed just to print that number.

4. The Lucas Lehmer test

A prominent theme in this and some other sections will be the problem of determining whether an integer is prime. We shall introduce this topic through Mersenne numbers which, recall, are the numbers of the type

$$M_\ell = 2^\ell - 1.$$

As we have seen in the previous section, if ℓ is composite then M_ℓ is also composite. On the other hand, if ℓ is a prime, then M_ℓ may or may not be prime. The list of known Mersenne primes is quite short, and contains only about forty numbers. The primality of

$$M_{127} = 170141183460469231731687303715884105727$$

was not proved until 1876 when Lucas devised a simple test to accomplish this task. The number $2^{127} - 1$ remained the largest known prime until mid-twentieth century, when the introduction of modern computing machines made possible using the Lucas-Lehmer test for large numbers. Even today, the ongoing computer-based effort to determine larger and larger Mersenne primes is based on this remarkable test:

THEOREM 18. *(Lucas-Lehmer) Define recursively a sequence s_n of integers by $s_1 = 4$ and $s_{n+1} = s_n^2 - 2$. Let ℓ be an odd prime. Then $M_\ell = 2^\ell - 1$ is prime if and only if*

$$s_{\ell-1} \equiv 0 \pmod{M_\ell}.$$

A remarkable feature of the test is that the test involves only one division to determine whether M_ℓ is prime or not. However, notice that the numbers s_n become quickly very large. Indeed,

$$\begin{aligned}
s_1 &= 4\\
s_2 &= 14\\
s_3 &= 194\\
s_4 &= 37634\\
s_5 &= 1416317954\\
s_6 &= 2005....6114
\end{aligned}$$

Thus, it is a good idea to calculate the numbers s_n modulo M_ℓ. In practice, this means calculating again the numbers s_n up to $s_{\ell-1}$. For example, if $M_7 = 127$, then we take $s_1 = 4$ and $s_2 = 14$, but replace $s_3 = 194$ by $s_3 \equiv -60 \pmod{127}$. Continuing in this fashion,

$$s_4 \equiv (-60)^2 - 2 \equiv 42 \pmod{127}$$

$$s_5 \equiv 42^2 - 2 \equiv -16 \pmod{127}$$

$$s_6 \equiv (-16)^2 - 2 \equiv 0 \pmod{127}$$

which shows that 127 is indeed a prime number.

The proof of the Lucas Lehmer test uses a clever interplay of quadratic integers and elementary group theory. The main tool in the test is the ring

$R = \mathbb{Z}[\sqrt{3}]$ consisting of quadratic numbers $a + b\sqrt{3}$ where a and b are integers. Recall that

$$a + b\sqrt{3} \equiv c + d\sqrt{3} \pmod{p}$$

if $a \equiv c \pmod{p}$ and $b \equiv d \pmod{p}$. The set of classes modulo p is denoted by R/pR. It is also a ring with respect to modular addition and multiplication. Clearly, the order of R/pR is p^2. In particular, the order of the group of invertible elements $(R/pR)^\times$ is strictly less than p^2.

Let $\alpha = 2 + \sqrt{3}$ and $\beta = 2 - \sqrt{3}$. Then $\alpha\beta = 1$, so α is an invertible element in the ring R. In particular, α is invertible when considered as an element of R/pR for any prime p. Define a sequence of numbers by

$$t_n = \alpha^{2^{n-1}} + \beta^{2^{n-1}}.$$

We claim that $s_n = t_n$. First of all, $t_1 = \alpha + \beta = 4 = s_1$. Thus, in order to show that $t_n = s_n$, we need to show that t_n satisfy the same recursive relation as s_n. This is not difficult at all. Since $\alpha\beta = 1$,

$$t_n^2 - 2 = (\alpha^{2^{n-1}} + \beta^{2^{n-1}})^2 - 2 = \alpha^{2^n} + \beta^{2^n} = t_{n+1},$$

as desired.

We can now give a proof of the test in one direction: if $s_{\ell-1}$ is divisible by M_ℓ then M_ℓ is prime. The other direction, as well as a more conceptual approach to the Lucas - Lehmer test, will be given in the chapter on Quadratic Reciprocity. The proof is by contradiction. Assume that M_ℓ divides $s_{\ell-1}$ but M_ℓ is not prime. Then there exists a factor p of M_ℓ such that $p \leq \sqrt{M_\ell}$. Since p divides M_ℓ and M_ℓ divides $s_{\ell-1}$, it follows that p divides $s_{\ell-1}$:

$$\alpha^{2^{\ell-2}} + \beta^{2^{\ell-2}} \equiv 0 \pmod{p}.$$

We shall manipulate this congruence as follows. First, subtract $\beta^{2^{\ell-2}}$ from both sides to obtain

$$\alpha^{2^{\ell-2}} \equiv -\beta^{2^{\ell-2}} \pmod{p}.$$

Second, multiply both sides by $\alpha^{2^{\ell-2}}$ and use $\alpha\beta = 1$ to obtain

$$\alpha^{2^{\ell-1}} \equiv -1 \pmod{p}.$$

Finally, as the last step, square both sides to obtain

$$\alpha^{2^\ell} \equiv 1 \pmod{p}.$$

The last two congruences imply (as it will be explained in a moment) that the order of α in $(R/pR)^\times$ is precisely $2^\ell = M_\ell + 1$. However, the order of $(R/pR)^\times$ is less than p^2. Since $p^2 \leq M_\ell$, by our choice of p, it follows that the order of α is greater than the order of the group $(R/pR)^\times$. This is a contradiction. Hence M_ℓ is prime.

So why is the order of α precisely 2^ℓ? To answer this question we need the following general principle: If g is an element of order k in a group G

then $g^n = 1$ precisely when k divides n. This is rather easy to see, since $n = qk + r$ with $0 \le r < k$, and

$$1 = g^n = g^{qk+r} = (g^k)^q \cdot g^r = g^r.$$

Since the order of g is k, $g^r = 1$ is possible only if $r = 0$, which means that k divides n. This principle can be illustrated using the clock arithmetic. It takes a minimum of 12 hours for clock handles to return to the same position. Moreover, if the handles return to the same position after n hours, then n is a multiple of 12.

Therefore, if the order of α is strictly less than 2^ℓ, then it would be a proper divisor of 2^ℓ, and therefore a divisor of $2^{\ell-1}$. This would imply that

$$\alpha^{2^{\ell-1}} \equiv 1 \pmod{p},$$

a contradiction, since $\alpha^{2^{\ell-1}} \equiv -1 \pmod{p}$ and $1 \not\equiv -1 \pmod{p}$ since p is odd.

Exercises

1) Note that the number s_6 has 19 digits, so it does not fit on the display of probably any calculator. So how do we calculate it without any sophisticated software application? Here is a trick. Write s_5 as

$$1416317954 = 14163 \times 10^5 + 17954.$$

Thus, the square of s_5 is

$$14163^2 \times 10^{10} + 2 \cdot 14163 \cdot 17954 \times 10^5 + 17954^2$$

which can be calculated on any calculator with 10 digits. Use this trick (twice) to calculate s_7.

2) Use the Lucas-Lehmer test to show that the following Mersenne numbers are prime:

a) M_{19}
b) M_{31}

3) Use the Lucas-Lehmer test to show that the following Mersenne numbers are not prime:

a) M_{11}.
b) M_{23}.

CHAPTER 5

Roots

1. Roots

Let m be a positive integer. In this section we discuss solutions of the equation

$$x^k \equiv a \pmod{m}$$

under the assumption that a is relatively prime to m. In other words, we would like to compute $\sqrt[k]{a}$, the k-th root of a modulo m. We start with an example

$$x^3 \equiv 2 \pmod{m}$$

and $m = 5$ or 7. Since 5 and 7 are small we can simply list all cubes modulo 5 and modulo 7.

Cubes modulo 5:

x	1	2	3	4
x^3	1	3	2	4

Cubes modulo 7:

x	1	2	3	4	5	6
x^3	1	1	6	1	6	6

The first table shows that 3 is a cube root of 2 modulo 5. The second table shows that a cube root of 2 does not exist modulo 7. The tables show that every non-zero element modulo 5 has a unique cube root, while modulo 7 only two (1 and 6) out of six elements have a cube root. Moreover, the equations

$$x^3 \equiv 1 \pmod{7} \text{ and } x^3 \equiv 6 \pmod{7}$$

have three solutions, each. In other words, the mapping $x \mapsto x^3$ is one to one modulo 5 and three to one modulo 7.

Now let us move to the general case. If the map $x \mapsto x^k$ from $(\mathbb{Z}/m\mathbb{Z})^\times$ to $(\mathbb{Z}/m\mathbb{Z})^\times$ is one to one, then $x^k \equiv a \pmod{m}$ has a unique solution for every non-zero a. A criterion, when this happens, is given by the following proposition.

PROPOSITION 19. *Let G be a finite group of order n. Let k be a positive integer relatively prime to n. Recall that, by the fundamental theorem of arithmetic, there exist integers u and v such that $ku + nv = 1$. Then*

$$x = a^u$$

is the unique solution of the equation $x^k = a$, for every a in G. In particular, the map $x \mapsto x^k$ from G to G is one to one and onto.

PROOF. If we can show that the map $x \mapsto x^k$ is onto then it is automatically one to one, since G is finite. In other words, if we can show that for every a in G the equation $x^k = a$ has at least one solution then the solution is necessarily unique. Therefore, it remains to check that $x = a^u$ is a solution. That is easy,

$$x^k = (a^u)^k = a^{ku} = a^{1-nv} = a \cdot a^{-nv} = a \cdot (a^n)^{-v} = a.$$

In the last step we used that $a^n = e$, by the theorem of Lagrange. The proposition is proved. □

This proposition can be at once applied to the group $G = (\mathbb{Z}/m\mathbb{Z})^\times$. The order of this group is given by the Euler's function $\varphi(m)$. For example, if $m = 5$ then $\varphi(5) = 4$, which is relatively prime to $k = 3$, and the map $x \mapsto x^3$ is one-to-one. In general, if k is *relatively prime* to $\varphi(m)$ then

$$x^k \equiv a \pmod{m}$$

has a unique solution for every a relatively prime to m. Moreover, the solution of the equation is given by

$$x \equiv a^u \pmod{m}$$

where u is an integer such that $uk + v\varphi(m) = 1$. In other words, $u = 1/k$ modulo $\varphi(m)$, so our solution can be thought of as $a^{\frac{1}{k}}$. Going back to our example, the solution of

$$x^3 \equiv 2 \pmod{5}$$

is 2^u where u is the multiplicative inverse of 3 modulo $\varphi(5) = 4$. Since $3 \cdot 3 \equiv 1 \pmod{4}$, it follows that $2^3 \equiv 3 \pmod{5}$ is the solution of original equation, as we originally observed by writing down all cubes modulo 5.

Of course, having developed a method of calculating roots, we can be more ambitious and solve some more complicated equations. Consider, for example,

$$x^7 \equiv 7 \pmod{23}.$$

Here 23 is prime, so $\varphi(23) = 22$. Since $\gcd(22, 7) = 1$ there is precisely one solution. The multiplicative inverse of 7 modulo 22 is 19 since

$$7 \cdot 19 - 6 \cdot 22 = 1$$

thus the solution of the equation is 7^{19} which can be efficiently computed using the method of successive squaring: We first calculate (modulo 23) the successive squares

$$7^2,\ (7^2)^2 = 7^4,\ ((7^2)^2)^2 = 7^8, \ldots$$

and then combine these powers to obtain 7^{19}. This step uses a binary expansion of 19. Recall that the binary notation for 19 is 10011, which means that

$$19 = 1 \cdot 2^4 + 0 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0 = 16 + 2 + 1.$$

This implies that

$$7^{19} = 7^{16} \cdot 7^2 \cdot 7.$$

Summarizing, in order to calculate 7^{19} modulo 23 we first square successively 7 modulo 23 up to 7^{2^4} modulo 23:

$$\begin{array}{rcrl} 7 & \equiv & 7 & \pmod{23} \\ 7^2 & \equiv & 3 & \pmod{23} \\ 7^4 & \equiv & 9 & \pmod{23} \\ 7^8 & \equiv & 12 & \pmod{23} \\ 7^{16} & \equiv & 6 & \pmod{23} \end{array}$$

and then multiply

$$7^{19} = 7^{16} \cdot 7^2 \cdot 7 \equiv 6 \cdot 3 \cdot 7 \equiv 11 \pmod{23}.$$

This completes the calculation. The 7-th root of 7 modulo 23 is 11.

The method of successive squares is a very efficient method to calculate powers. In our example, it took four consecutive squaring and then two additional multiplications to compute 7^{19}, instead of 18 multiplications $7 \cdot 7 \cdot \ldots \cdot 7$. A general estimate of this method can be derived as follows. Since

$$n = 2^{\log_2(n)},$$

we need first to calculate approximately $\log_2(n)$ consecutive squares x^2, $x^4 = (x^2)^2, \ldots$ and then multiply some of them according to the binary expansion of n. This gives less than $2\log_2(n)$ operations in all as opposed to $n - 1$ multiplications $x \cdot x \cdot \ldots \cdot x$.

Of course, so far, we have completely ignored the case when k is not relatively prime to $\varphi(m)$. This case is much more difficult. For example, consider the case when $m = p$ is an odd prime number and $k = 2$. Then the map $x \mapsto x^2$ is two-to-one, since $x^2 = (-x)^2$, so the equation

$$x^2 \equiv a \pmod{p}$$

will have two solutions for $(p-1)/2$ possible a and no solution for the other $(p-1)/2$ choices for a. The answer to the question whether, for a given a, the equation has a solution or not is given by the *Quadratic Reciprocity.*

Exercises

1) Using the method of successive squaring compute 5^{143} modulo 1979.

2) Using the method of successive squaring compute 2^{143} modulo 1979.

3) Solve
a) $x^{11} \equiv 13 \pmod{35}$.
b) $x^7 \equiv 11 \pmod{63}$.
c) $x^5 \equiv 3 \pmod{64}$.

2. Property of the Euler Function

Recall that the Euler function $\varphi(n)$ is defined as the number of integers k, $1 \leq k \leq n$, relatively prime to n. In this section we want to establish the following formula:

$$\sum_{d|n} \varphi(d) = n$$

where the sum is over all divisors d of n. For example, if $n = 15$ then the divisors are $1, 3, 5$ and 15 and

$$\varphi(1) + \varphi(3) + \varphi(5) + \varphi(15) = 1 + 2 + 4 + 8 = 15.$$

This equality is not difficult to prove. If k is an integer between 1 and n then $\gcd(k, n)$ is a divisor of n. We can partition all integers k between 1 and n into disjoint subsets as follows. For every divisor d of n let S_d be the set of all integers k between 1 and n such that $\gcd(k, n) = d$. For example, if $n = 15$, then we have 4 divisors, 1, 3, 5, 15, and four subsets given by the following table:

$\gcd(k, 15)$	k
1	$1, 2, 4, 7, 8, 11, 13, 14$
3	$3, 6, 9, 12$
5	$5, 10$
15	15

Since the union of all S_d contains all integers from 1 to n, we have

$$\sum_{d|n} \#\{1 \leq k \leq n \mid \gcd(k, n) = d\} = n.$$

Our next step is to understand the cardinality of S_d. For example, let $n = 60$ and $d = 4$. If $\gcd(k, 60) = 4$ then k is a multiple of 4. The multiples of 4 between 1 and 60 are $4 \cdot l$ where $1 \leq l \leq 15$:

$$\{4,\ 4 \cdot 2,\ 4 \cdot 3,\ 4 \cdot 4,\ 4 \cdot 5,\ 4 \cdot 6,\ 4 \cdot 7,\ 4 \cdot 8,\ 4 \cdot 9,\ 4 \cdot 10,\ 4 \cdot 11,\ 4 \cdot 12,\ 4 \cdot 13,\ 4 \cdot 14\}.$$

Next, $\gcd(4l, 60) = 4$ if and only if $\gcd(l, \frac{60}{4}) = \gcd(l, 15) = 1$. It follows that

$$S_4 = \{4,\ 4 \cdot 2,\ 4 \cdot 4,\ 4 \cdot 7,\ 4 \cdot 8,\ 4 \cdot 11,\ 4 \cdot 13,\ 4 \cdot 14\},$$

that is, the set S_4 consists of integers $4 \cdot l$ where $1 \leq l \leq 15$ and l is relatively prime to 15. Thus, there are $\varphi(15)$ elements in S_4. This works in general. If d is a divisor of n, then the set S_d consists of integers $d \cdot l$ where

$$1 \leq l \leq \frac{n}{d}$$

and l is relatively prime to $\frac{n}{d}$. It follows that the cardinality of S_d is $\varphi(n/d)$.

Summarizing, we have partitioned the integers between 1 and n into the subsets S_d parameterized by the divisors d of n. Each subset S_d has $\varphi(n/d)$ elements. Adding the cardinalities of all S_d gives n:

$$\Sigma_{d|n} \varphi\left(\frac{n}{d}\right) = n.$$

This is what we wanted to prove. Indeed, if

$$d_1, d_2, \ldots, d_m$$

are all divisors of n, then

$$\frac{n}{d_1}, \frac{n}{d_2}, \ldots, \frac{n}{d_m}$$

are again all divisors of n. In particular,

$$\Sigma_{d|n} \varphi\left(\frac{n}{d}\right) = \Sigma_{d|n} \varphi(d)$$

and, therefore, for every integer n

$$n = \Sigma_{d|n} \varphi(d).$$

Exercises

1) Let p be a prime and $n = p^k$. By a direct calculation (compute all terms) check the formula

$$\sum_{d|n} \varphi(d) = n.$$

3. Primitive roots

Let F be any field. An element ζ in F is called an n-th root of 1 if $\zeta^n = 1$. The set of all n-th roots of 1 in F is denoted by $\mu_n(F)$ or simply μ_n. It is a subgroup of $F^{\times}$. Note that the order of $\zeta \in \mu_n$ divides n, but is not necessarily equal to n. For example, 1 is an n-th root of 1 for every n, but the order of 1 is one.

As the first result in this section, we show that the number of n-th roots of 1 in the field F is always less than or equal to n. To that end, note that n-th roots of 1 are precisely the roots of the polynomial equation

$$x^n - 1 = 0.$$

As you probably know, a polynomial $P(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_0$ of degree n with coefficients in F cannot have more than n roots. Indeed, if $z_1, \ldots, z_n$ are n roots of $P(x)$ then $P(x)$ can be factored as

$$P(x) = (x - z_1) \cdot \ldots \cdot (x - z_n).$$

Let ζ be a root of P. Then

$$P(\zeta) = (\zeta - z_1) \cdot \ldots \cdot (\zeta - z_n) = 0.$$

Since F is a field and has no zero divisors, one factor $\zeta - z_i$ must be equal to 0. This implies that $\zeta = z_i$ i.e. any root is equal to one of the n roots $z_1, \ldots, z_n$. In particular, the polynomial $x^n - 1$ cannot have more than n roots.

For example, if $F = \mathbb{C}$ and $n = 4$ then $x^4 - 1$ can be factored out as follows:

$$x^4 - 1 = (x^2 - 1)(x^2 + 1) = (x - 1)(x + 1)(x - i)(x + i).$$

This shows that $1, -1, i, -i$ are the four, 4-th roots of 1 in $\mathbb{C}$. The order of -1 is 2, while the orders of i and $-i$ are 4. An n-th root of 1 of order n is called a *primitive* n-th root of one. For example, i and $-i$ are primitive 4-th roots of 1. Every root of 1 is primitive for some n. Here is a table for small degree n together with the corresponding primitive roots:

n	primitive roots
1	1
2	-1
4	$i, -i$

An importance of primitive roots is in the following. If ζ is a primitive n-th root of 1 then the powers

$$1, \zeta, \zeta^2, \ldots, \zeta^{n-1}$$

are all distinct and also n-th roots of 1. In particular, any n-th root of 1 is a power of ζ. In other words, the group μ_n is cyclic of order n. Take, for example, i. Then the first four powers yield all complex 4-th roots of 1:

$$i^0 = 1,\ i^1 = i,\ i^2 = -1 \text{ and } i^3 = -1,$$

This brings us to the following question. Fix a field F and a positive integer n. Are there any primitive n-th roots of 1 in F, and how may primitive roots of order n can be in a field? This is answered nicely as follows.

Proposition 20. *Let n be a positive integer. Assume that the equation $x^n - 1 = 0$ has n different solutions in F. Then $\varphi(n)$ of them are primitive n-th roots of one, where φ is the Euler function. In particular, μ_n is a cyclic group of order n.*

Proof. The proof of this statement is by *mathematical induction.* If $n = 1$ then 1 is the unique solution of $x - 1 = 0$. It is a primitive root. Since $\varphi(1) = 1$, the proposition holds for $n = 1$. We shall derive the statement for a general n assuming that it holds for all $d < n$. Let ζ be an n-th root of 1. If ζ is not primitive then it satisfies an equation $x^d - 1 = 0$ for a proper divisor d of n. Write $n = d \cdot l$. Then

$$x^n - 1 = (x^d - 1)(x^{d(l-1)} + x^{d(l-2)} + \ldots + x^d + 1).$$

It follows that precisely d of the n solutions of $x^n - 1 = 0$ are solutions of $x^d - 1 = 0$. By the induction assumption, since $d < n$, $\varphi(d)$ of them are primitive d-th roots of 1. Thus, the total number of non-primitive n-th roots is

$$\sum_{d|n, d \neq n} \varphi(d),$$

where the sum is taken over all divisors d of n different from n. It follows that the number of primitive n-th roots of 1 is

$$n - \sum_{d|n, d \neq n} \varphi(d).$$

Since $n = \sum_{d|n} \varphi(d)$, by the formula for Euler's function, it follows that the number of primitive n-th roots is $\varphi(n)$. The proposition is proved. □

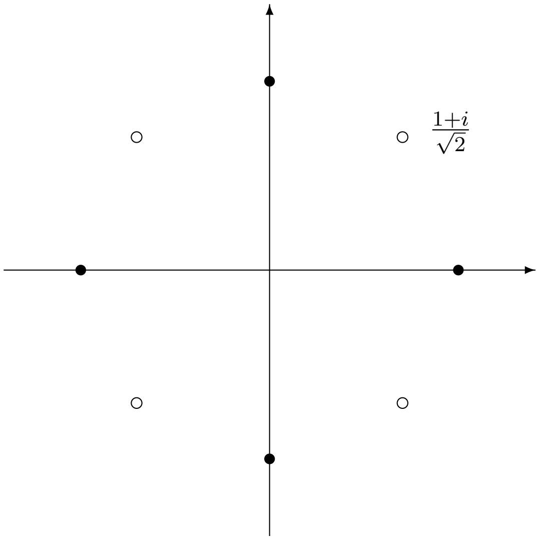
Example: If $F = \mathbb{C}$, the field of complex numbers, then n-th roots of 1 are

$$e^{2\pi ik/n} = \cos\frac{2\pi k}{n} + i\sin\frac{2\pi k}{n} = (e^{2\pi i/n})^k$$

for $k = 1, 2, \ldots, n$. They form a regular n-gon with vertices on the unit circle. The root $e^{2\pi ik/n}$ is primitive if and only if k is relatively prime to n. Thus we have (obviously) $\varphi(n)$ primitive roots here. For example, if $n = 8$, then

$$e^{2\pi i/8} = \cos\frac{\pi}{4} + i\sin\frac{\pi}{4} = \frac{1+i}{\sqrt{2}}.$$

The following figure represents complex 8-th roots of 1. The primitive are represented by blank circles.



We now move to finite fields. By Fermat's Little Theorem $x^{p-1} = 1$ for every non-zero element x in $\mathbb{Z}/p\mathbb{Z}$. In other words, all non-zero elements are roots of one! Thus we have exactly $p - 1$ solutions of the equation $x^{p-1} - 1 = 0$. Our results imply that $\varphi(p-1)$ elements in $\mathbb{Z}/p\mathbb{Z}$ are primitive or, in the group language, $\varphi(p - 1)$ elements in the multiplicative group $(\mathbb{Z}/p\mathbb{Z})^\times$ have order $p - 1$. Let us see this on the example $p = 11$. We list all non-zero elements and their orders:

x	1	2	3	4	5	6	7	8	9	10
$o(x)$	1	10	5	5	5	10	10	10	5	2

The table shows that there are $4 = \varphi(10)$ elements of order 10, $4 = \varphi(5)$ of order 5, $1 = \varphi(2)$ element of order 2 and $1 = \varphi(1)$ element of order 1. As we now know, this is by no means an accident.

Exercises

1) Let F be a field. Show that the set of all n-th roots of 1 is a subgroup of $F^\times$.

2) Let ζ be a primitive n-th root of 1. Then $1, \zeta, \ldots \zeta^{n-1}$ are all n-th roots of 1. Let $S_n = 1 + \zeta + \cdots \zeta^{n-1}$ be the sum of all n-th roots of 1. Show that $S_n = 0$ using the factorization

$$x^n - 1 = (x-1)(x-\zeta) \cdot \ldots \cdot (x - \zeta^{n-1}).$$

Hint: S_n shows up as a coefficient of $x^n - 1$.

3) Consider the finite field $\mathbb{Z}/31\mathbb{Z}$. Check that 3 is a primitive root by working out all powers. Deduce from this, quickly, which elements are sixth roots of 1 (not necessarily primitive). Add them up modulo 31. What number should you get?

4. Discrete logarithm

The fact that $\mathbb{Z}/p\mathbb{Z}$ contains a primitive root of order $p-1$ can be used to define a *discrete* logarithm for finite fields. Take, for example, $p = 11$. Since 2 is a primitive root modulo 11, its powers

$$2, 2^2, \dots, 2^9, 2^{10} = 1$$

give all non-zero elements modulo 11. A quick calculation shows that

I	1	2	3	4	5	6	7	8	9	10
2^I	2	4	8	5	10	9	7	3	6	1

If $x = 2^I$ we say that I is a discrete logarithm of x modulo 11, with respect to the base 2. The number $I = I(x)$ is called also the index of x. Note that $I(x)$ is represented by an integer modulo $10 = \varphi(11)$, by Fermat's little theorem. More generally, if g is a primitive root modulo p then any element x in $(\mathbb{Z}/p\mathbb{Z})^\times$ can be written as

$$x = g^{I(x)}$$

for a unique element $I(x)$ is in $\mathbb{Z}/(p-1)\mathbb{Z}$. In other words, we have an identification

$$I : (\mathbb{Z}/p\mathbb{Z})^\times \to \mathbb{Z}/(p-1)\mathbb{Z}.$$

But this is not all. Just as the usual logarithmic function enjoys the property $\log(xy) = \log(x) + \log(y)$ we also have

$$I(xy) = I(x) + I(y)$$

in this case. This is easy to check. Indeed, given two elements $x = g^{I(x)}$ and $y = g^{I(y)}$ in $(\mathbb{Z}/p\mathbb{Z})^\times$, then

$$xy = g^{I(x)}g^{I(y)} = g^{I(x)+I(y)}.$$

Since xy is also equal to $g^{I(xy)}$, by the definition of I, it follows that $I(xy) = I(x) + I(y)$, as claimed.

The map I is an example of a *group homomorphism* since it transfers the group operation from one group (in this case $(\mathbb{Z}/p\mathbb{Z})^\times$) to another group (in this case $\mathbb{Z}/(p-1)\mathbb{Z}$). Here is a formal definition of the group homomorphism.

DEFINITION 21. *Let G_1 and G_2 be two groups. A map $h : G_1 \to G_2$ is called a (group) homomorphism if*

$$h(x \cdot y) = h(x) \cdot h(y)$$

for any two elements x and y in G_1. (In order to keep the notation simple, we use $\cdot$ for the group law in each group.) Moreover, the map h is called an isomorphism if it is one to one and onto.

Any two isomorphic group are indistinguishable, as far as their "group properties" are concerned. For example, an element g in G_1 has the order n if and only $h(g)$ in G_2 has the order n. Thus, the discrete logarithm can be used to replace the modular multiplication by the modular addition in problems. For example, assume that we want to solve a congruence

$$4x \equiv 7 \pmod{11}.$$

We can write $4 = 2^{I(4)}$, $x = 2^{I(x)}$ and $7 = 2^{I(7)}$. Substituting in the congruence gives

$$2^{I(4)}2^{I(x)} \equiv 2^{I(7)} \pmod{11}.$$

Since $2^{I(4)}2^{I(x)} = 2^{I(4)+I(x)}$, and the discrete logarithm is determined modulo $10 = \varphi(11)$, we can replace the original congruence with

$$I(4) + I(x) \equiv I(7) \pmod{10}.$$

This can be solved quickly since $I(4) = 2$ and $I(7) = 7$, from the table. It follows that $I(x) = 5$ and, therefore, $x = 10$.

We know that there are primitive roots modulo any prime p. But, given a prime p, which integers modulo p are primitive roots? There seems no easy way to answer this question. Similarly, we can ask for which primes is 2 a primitive root? Or, for which primes is 3 a primitive root? Again, these questions have no known answer. Due to the lack of any better ideas we simply list the first primitive root g modulo p for primes under 100:

g	p
2	$3, 5, 11, 13, 19, 29, 37, 53, 59, 61, 67, 83$
3	$7, 17, 31, 43, 79, 89$
5	$23, 47, 73, 97$
6	41
7	71

Exercises

1) Use the discrete logarithm modulo 11 with base 2 to solve the following congruence:

$$7x \equiv 5 \pmod{11}.$$

Verify that the solution is correct.

2) Solve, using the discrete logarithm with base 2,

$$4x^2 \equiv 9 \pmod{11}.$$

3) The number 2 is a primitive root modulo 19. Compute the powers 2^I for $I = 1, 2, \ldots, 18, 19$ to obtain the table for the discrete logarithm with base 2 for integers modulo 19. Then use the table to solve the equation

$$x^5 \equiv 7 \pmod{19}.$$

5. Cyclotomic Polynomials

Let F be a field, such as the field of complex numbers $\mathbb{C}$ or a finite field $\mathbb{Z}/p\mathbb{Z}$. A number ζ in F is a primitive n-th root of 1 if $\zeta^n = 1$ and

$$\zeta^d \neq 1$$

for all proper divisors d of n. For example, the complex number i is a primitive fourth root of 1 since $i^4 = 1$ and $i^2 \neq 1$. In the previous section we showed that if the equation $x^n - 1 = 0$ has n solutions in a field, then $\varphi(n)$ of them are primitive n-th roots of 1. For example, if $F = \mathbb{C}$ then the solutions of $x^n - 1 = 0$ are

$$\zeta^k = e^{2\pi ik/n} = \cos(2\pi k/n) + i\sin(2\pi k/n)$$

for $k = 1, \ldots, n$ and ζ^k is primitive if and only if k is relatively prime to n.

Using the complex n-th roots of 1, we define (so-called) n-th cyclotomic polynomial as a product

$$\Phi_n(x) = \prod_{\gcd(k,n)=1} (x - \zeta^k).$$

Clearly, by design, the degree of this polynomial is $\varphi(n)$. Its roots are precisely the primitive n-th roots of 1. Since any n-th root of 1 is primitive d-th root of 1 for some divisor d of n, we have a factorization

$$x^n - 1 = \prod_{d|n} \Phi_d(x).$$

A remarkable fact, which we shall verify in a moment, is that $\Phi_n(x)$ have integer coefficients. Before we discuss how to calculate the cyclotomic polynomials, here is a list of the first six:

n	Φ_n
1	$x - 1$
2	$x + 1$
3	$x^2 + x + 1$
4	$x^2 + 1$
5	$x^4 + x^3 + x^2 + x + 1$
6	$x^2 - x + 1$

It would be certainly very cumbersome to figure out the coefficients of the cyclotomic polynomials from our definition. However, the cyclotomic polynomial can be efficiently calculated in the following (inductive) fashion. If $\Phi_d(x)$ are known for all proper divisors d of n then $\Phi_n(x)$ can be computed by dividing $x^n - 1$ by the product

$$\prod_{d|n, d\neq n} \Phi_d(x).$$

In practice this product can be easily factored out using standard tricks of algebra. For example, assume we want to compute Φ_{10}. The proper divisors of 10 are 1, 2, and 5. The cyclotomic polynomials $\Phi_1(x)$, $\Phi_2(x)$ and $\Phi_5(x)$ are given in the above table. Since

$$x^{10} - 1 = (x^5 - 1)(x^5 + 1)$$

and

$$\begin{cases} x^5 - 1 = (x-1)(x^4 + x^3 + x^2 + x + 1) \\ x^5 + 1 = (x+1)(x^4 - x^3 + x^2 - x + 1) \end{cases}$$

we can factor

$$x^{10} - 1 = (x-1)(x^4 + x^3 + x^2 + x + 1)(x+1)(x^4 - x^3 + x^2 - x + 1).$$

We recognize the first three factors as $\Phi_1(x)$, $\Phi_5(x)$ and $\Phi_2(x)$. Therefore, the last factor must be

$$\Phi_{10}(x) = x^4 - x^3 + x^2 - x + 1.$$

The inductive procedure also explains why the coefficients of $\Phi_n(x)$ are integers. Recall that, if $f(x)$ and $g(x)$ are two polynomials with integer coefficients then there exist two polynomials $q(x)$ and $r(x)$ such that

$$f(x) = q(x)g(x) + r(x)$$

where $r = 0$ or the degree of r is strictly less than the degree of $g(x)$. The coefficients of $q(x)$ may be rational numbers. However, if the leading coefficient of $g(x)$ is 1 (such polynomial is called *monic*) then $q(x)$ has integer coefficients. You can easily convince yourself to this fact by working out an example of synthetic division. In any case, the quotient of two monic polynomials is again monic, and this shows that $\Phi_n(x)$ is monic since it is a quotient of $x^n - 1$ by a monic polynomial.

We finish with the following remark. Originally, we defined the cyclotomic polynomials using complex roots. However, the inductive procedure used to compute the polynomials starts with $\Phi_1(x) = x - 1$, then $x^2 - 1$ is divided by $\Phi_1(x)$ to get $\Phi_2(x) = x + 1$ and so on, never uses any complex numbers whatsoever. It follows that the cyclotomic polynomials are completely independent of the field. For example, consider the field $\mathbb{Z}/11\mathbb{Z}$. Then the orders of non-zero elements are

x	1	2	3	4	5	6	7	8	9	10
$o(x)$	1	10	5	5	5	10	10	10	5	2

In particular, 2, 6, 7, and 8 are the four ($4 = \varphi(10)$) primitive 10-th roots of 1. We can calculate $\Phi_{10}(x)$ using these roots. Multiplying out

$$(x-2)(x-6)(x-7)(x-8) = x^4 - 23x^3 + 188x^2 - 628x + 672$$

and then reducing modulo 11 gives

$$x^4 - 23x^3 + 188x^2 - 628x + 672 \equiv x^4 - x^3 + x^2 - x + 1 \pmod{11}$$

which is $\Phi_{10}(x)$, as worked out before.

Exercises

1) Find all complex roots of the polynomial $x^6 - 1$, by factoring it into a product of cyclotomic polynomials

$$x^6 - 1 = \Phi_1(x)\Phi_2(x)\Phi_3(x)\Phi_6(x)$$

and then using the quadratic formula to find the roots of $\Phi_3(x)$ and $\Phi_6(x)$. Graph the solutions.

2) All five complex 6-th roots of 1 form a regular hexagon in the plane of complex numbers. The following nice geometric argument shows that S_6, the sum of all 6-th roots of 1, is 0: The angle between any two consecutive 6-th roots is 60 degrees. Thus, if we rotate the hexagon by 60 degrees we get back the same hexagon. This shows that the number S_6 has to be invariant under the rotation by 60 degrees. But there are no such complex number except 0, so $S_6 = 0$.

In this exercise you will show that the same argument works for any field. Let ζ be a primitive n-th root of 1. Then all n-th roots of 1 are $1, \zeta, \zeta^2, \ldots, \zeta^{n-1}$. Put

$$S_n = 1 + \zeta + \zeta^2 + \ldots + \zeta^{n-1}.$$

Show that $\zeta \cdot S_n = S_n$. Deduce form this that $S_n = 0$. (Notice that, if ζ is a complex number of norm 1, the multiplication by ζ is the same as rotation by the argument of ζ.)

3) Calculate Φ_{12} in two ways:

a) Factor $x^{12} - 1$, then use $\Phi_d(x)$ for all proper divisors d of 12, given in the table above.

b) Compute primitive roots modulo 13, then expand $\Phi_{12}(x)$ modulo 13.

4) Factor $x^9 - 1 = \Phi_1(x)\Phi_3(x)\Phi_9(x)$.

5) Calculate $\Phi_8(x)$ in two ways:

a) Factor $x^8 - 1$ into a product $\Phi_1(x)\Phi_2(x)\Phi_4(x)\Phi_8(x)$.

b) Calculate Φ_8 in the field of $\mathbb{F}_{17}$. The primitive 8-th roots of 1 modulo 17 are $2, 8, 9$ and 15.

6) Let p be a prime. Find an explicit expression for $\Phi_{p^n}(x)$.

7) This exercise involves some work. Compute $\Phi_{15}(x)$, $\Phi_{21}(x)$ and $\Phi_{35}(x)$.

8) Compute $\Phi_{105}(x)$. This is the first cyclotomic polynomial with a coefficient different from $-1, 0$ or 1.

CHAPTER 6

Quadratic Reciprocity

1. Squares modulo p

It this section we study the important problem of deciding whether an integer a is a square modulo p or, in other words, we would like to determine if the equation

$$x^2 \equiv a \pmod{p}$$

has a solution. Here p is an odd prime and, therefore, -1 is not congruent to 1 modulo p.

Before attacking the problem, let us revisit what happens for the group of non-zero real numbers $\mathbb{R}^\times = \mathbb{R} \setminus \{0\}$, In this case the mapping $x \mapsto x^2$ is two-to-one, and the image is the set of all positive real numbers. Thus,

$$x^2 - a = 0$$
$$x^2 + a = 0$$

have 2 and 0 solutions, respectively, for every positive real number a. A similar phenomenon occurs with $(\mathbb{Z}/p\mathbb{Z})^\times$. Consider, for example, $p = 11$. All squares modulo 11 are tabulated in the following table:

a	1	2	3	4	5	6	7	8	9	10
a^2	1	4	9	5	3	3	5	9	4	1

Thus, If $p = 11$ then the squares are: 1,3,4,5,9. These numbers are also called *quadratic residues* since they are the residues modulo p of integer squares. The map $x \mapsto x^2$ is two to one, as in the case of non-zero real numbers. The same happens for any prime number p. Indeed, if a and b are two non-zero elements in $\mathbb{Z}/p\mathbb{Z}$ such that $a^2 = b^2$ then

$$0 = a^2 - b^2 = (a - b)(a + b).$$

Since there are non-zero divisors modulo p, one of the two factors has to be 0. It follows that

$$a = -b \text{ or } a = b.$$

This shows that for every odd prime p there are $(p-1)/2$ squares. Here is another example with $p = 13$.

a	1	2	3	4	5	6	7	8	9	10	11	12
a^2	1	4	9	3	12	10	10	12	3	9	4	1

So, if $p = 13$, the squares are: 1,3,4,9,10,12 which makes $\frac{13-1}{2} = 6$ elements. The point of these examples is that, unlike the case of real numbers, it is not clear at all which integers are squares modulo p. For example, if we are given a prime $p = 131$, is 71 is a square mod 131? An answer to this question is given by *Quadratic Reciprocity*, a deep discovery of Gauss. The main goal of this Chapter is to state and partially prove the Quadratic Reciprocity.

As the first step in understanding this problem we shall make use of primitive roots modulo p.

Recall that there exists an element g in $(\mathbb{Z}/p\mathbb{Z})^\times$ of order $p-1$ (primitive root). Then every $a \in (\mathbb{Z}/p\mathbb{Z})^\times = \{1, 2, \dots, p-1\}$ can be written as

$$a = g^I, \text{ for } I = 1, \dots, p-1.$$

The integer I is the index of a with respect to g. Again, recall the example of $p = 11$ and $g = 2$:

I	1	2	3	4	5	6	7	8	9	10
2^I	2	4	8	5	10	9	7	3	6	1

Note that the indices of squares $\{1, 4, 9, 3, 5\}$ are even numbers. This is true in general. Indeed, if I is even then

$$g^I = (g^{I/2})^2$$

so g^I is a square. Since there are $\frac{p-1}{2}$ even indices, which is also the number of all squares, these account for all squares. Thus, if p is odd prime and g a primitive root of order $p-1$,

$$\begin{gathered} 1, g^2, g^4, \dots, g^{p-3} \text{ are squares, and} \\ g, g^3, g^5, \dots, g^{p-2} \text{ are non- squares} \end{gathered}$$

Now is a good time to introduce the Legendre Symbol, which is a notation introduced by Adrien-Marie Legendre. The symbol is a function

$$\left(\frac{\cdot}{p}\right) : (\mathbb{Z}/p\mathbb{Z})^\times \to \{\pm 1\}$$

defined as follows. If n is an element in $(\mathbb{Z}/p\mathbb{Z})^\times$ or, simply, an integer relatively prime to p then

$$\left(\frac{n}{p}\right) = \begin{cases} 1 & \text{if } n \text{ is a square modulo } p \text{ and} \\ -1 & \text{if } n \text{ is not a square modulo } p. \end{cases}$$

PROPOSITION 22. *Euler's criterion. The Legendre symbol can be computed using the following criterion*

$$a^{\frac{p-1}{2}} \equiv \left(\frac{a}{p}\right) \pmod{p}.$$

PROOF. Let g be a primitive root modulo p. Then a is congruent to g^I for some integer I, and

$$a^{\frac{p-1}{2}} \equiv (g^I)^{\frac{(p-1)}{2}} = (g^{\frac{p-1}{2}})^I \pmod{p}.$$

Since the order of g is $p-1$, the order of $g^{\frac{p-1}{2}}$ is precisely two. This implies that $g^{\frac{p-1}{2}} = -1$. Thus

$$a^{\frac{p-1}{2}} \equiv (-1)^I \pmod{p}.$$

If a is a square then I is even, and $(-1)^I = 1$. If a is not a square, then I is odd and $(-1)^I = -1$. The proposition is proved. □

We now derive two important consequences of Euler's criterion. We start with a special case of the Quadratic Reciprocity.

PROPOSITION 23. *Let p be an odd prime. Then -1 is a square modulo p if and only if $p \equiv 1 \pmod{4}$.*

PROOF. By the Euler's criterion,

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}.$$

If $p \equiv 1 \pmod{4}$ or, equivalently, $p = 4k+1$ then

$$(-1)^{\frac{p-1}{2}} = (-1)^{2k} = 1.$$

Thus -1 is a square if $p \equiv 1 \pmod{4}$. On the other hand, if $p \equiv 3 \pmod{4}$ or, equivalently, $p = 4k+3$ then

$$(-1)^{\frac{p-1}{2}} = (-1)^{2k+1} = -1.$$

Thus -1 is not a square if $p \equiv 3 \pmod{4}$. □

Another importance of the Legendre symbol lies in the fact that it is *multiplicative*:

PROPOSITION 24. *For any two $a, b \in (\mathbb{Z}/p\mathbb{Z})^\times$*

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right).$$

PROOF. This is easy to verify using the Euler's criterion criterion. Indeed,

$$\left(\frac{ab}{p}\right) = (ab)^{\frac{p-1}{2}} = a^{\frac{p-1}{2}} b^{\frac{p-1}{2}} = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right).$$

□

The multiplicativity of the Legendre symbol implies that a product of two squares in $(\mathbb{Z}/p\mathbb{Z})^\times$ is a square (obvious), a product of a square and a non-square is a non-square and, finally, a product of two non-squares is a square, as tabulated below:

·	S	N
S	S	N
N	N	S

The Legendre symbol is analogous to the sign function for non-zero real numbers:

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{if } x > 0 \iff x \text{ is a square} \\ -1 & \text{if } x < 0 \iff x \text{ is not a square} \end{cases}$$

A well known property of the sign function is

$$\operatorname{sgn}(xy) = \operatorname{sgn}(x) \cdot \operatorname{sgn}(y)$$

which states that the sign of a product of two numbers is the product of signs, as we have learned in the elementary school.

The Legendre symbol and the sign function are group homomorphisms. The Legendre symbol is a group homomorphism from $(\mathbb{Z}/p\mathbb{Z})^\times$ to $\{\pm 1\}$, and the sign function is a group homomorphism from $\mathbb{R}^\times$ to $\{\pm 1\}$. In both cases, of course, $\{\pm 1\}$ is considered as a group with the multiplication table

·	1	-1
1	1	-1
-1	-1	1

Exercises

1) 6 is a primitive root modulo 41. Using the method of consecutive squaring, calculate

$$6^{\frac{41-1}{2}} = 6^{20}$$

modulo 41. The answer should be congruent to -1 modulo 41.

2) Use Euler's criterion to determine if the following are squares:
a) 2 modulo 31.
b) 3 modulo 31.
c) 7 modulo 29.

3) Let n be a positive integer. Let p be a prime divisor of $n^2 + 1$. Prove that $p \equiv 1 \pmod 4$. Hint: Use

$$\left(\frac{n^2}{p}\right) = \left(\frac{-1}{p}\right).$$

4) Use the previous exercise to prove that there are infinitely many primes $p \equiv 1 \pmod 4$.

5) Let $GL_2(\mathbb{C})$ be the group of 2×2 invertible matrices with complex coefficients. A well known property of determinant tells us that the determinant is a homomorphism from $GL_2(\mathbb{C})$ to $\mathbb{C}^\times$. State that property.

2. Fields of order p^2

Let p be a prime. In this section we show that there is only one field of order p^2. This is a consequence of existence of primitive roots, Proposition 20.

Recall that the characteristic of a finite field F is the smallest number ℓ (necesarily a prime) such that $1+1+\cdots+1=0$ where 1 is added ℓ-times. In other words, ℓ is the order of 1 in F, considered as a group with respect to the addition. By the theorem of Lagrange, the order of F must be divisible by ℓ. In particular, if the order of F is p^n, then the characteristic of F is p. The field $\mathbb{F}_p$ is a subfield of F where $a \in \mathbb{F}_p$, an integer modulo p, is identified with

$$1+1+\cdots+1 \in F$$

where 1 is added a times. The Frobenius map $\mathrm{Fr}(x) = x^p$, rising every element x in F to its p-th power, has some special properties. It is additive,

$$\mathrm{Fr}(x+y) = (x+y)^p = x^p + y^p = \mathrm{Fr}(x) + \mathrm{Fr}(y)$$

for all x and y in F, and $\mathrm{Fr}(x) = x$ if and only if x is in the subfield $\mathbb{F}_p \subseteq F$. Indeed, elements $x \in \mathbb{F}_p$ satisfy $x^p = x$, by the Fermat's Little Theorem, and thus account for all p solutions of the equation $x^p = x$.

Existence of fields of order p^2. Assume first that $p \neq 2$. We previously constructed fields of order p^2. We briefly review the construction. The map $x \mapsto x^2$ from $\mathbb{F}_p^\times$ to $\mathbb{F}_p^\times$ is 2 to 1. In particular, there is d in $\mathbb{F}_p^\times$ which is not a square. Let

$$\mathbb{F}_p[\sqrt{d}] = \{x + y\sqrt{d} \mid x, y \in \mathbb{F}_p\}.$$

The addition in $\mathbb{F}_p[\sqrt{d}]$ is defined by

$$(x + y\sqrt{d}) + (u + v\sqrt{d}) = (x+u) + (y+v)\sqrt{d}$$

and the multiplication in $\mathbb{F}_p[\sqrt{d}]$ is defined by

$$(x + y\sqrt{d}) \cdot (u + v\sqrt{d}) = (xu + yvd) + (xv + yu)\sqrt{d}.$$

One shows that $\mathbb{F}_p[\sqrt{d}]$ is a field where $0+0\sqrt{d}$ is the identity with respect to the addition and $1+0\sqrt{d}$ is the identity with respect to the multiplication. Most of the field axioms are easy to verify. The hardest part is to show that every $z \neq 0 + 0\sqrt{d}$ has a multiplicative inverse. The multiplicative inverse of $z = x + y\sqrt{d}$ is

$$\frac{x}{x^2 - dy^2} - \frac{y}{x^2 - dy^2}\sqrt{d}.$$

Note that the inverse is well defined as long as $x^2 - dy^2 \neq 0$. If $x^2 - dy^2 = 0$ and $y \neq 0$ then $d = (x/y)^2$, and this is a contradiction since d is not a square. Thus, if $x^2 - dy^2 = 0$ then $y = 0$. Substituting $y = 0$ in $x^2 - dy^2 = 0$ gives $x = 0$. We have shown that $x^2 - dy^2 = 0$ implies that $z = 0 + 0\sqrt{d}$. In other words, z has a multiplicative inverse if $z \neq 0 + 0\sqrt{d}$.

Action of the Frobenius map on $\mathbb{F}_p[\sqrt{d}]$ can be made very explicit.

Proposition 25. *Realize the field $\mathbb{F}_{p^2}$ as the set of numbers $a + b\sqrt{d}$ where a and b are integers considered modulo p. Then* $\mathrm{Fr}(a+b\sqrt{d}) = a - b\sqrt{d}$ *or, in terms of congruences,*

$$(a + b\sqrt{d})^p \equiv a - b\sqrt{d} \pmod{p}.$$

Proof. This congruence is not too difficult to check. Since $(x + y)^p = x^p + y^p$ in characteristic p, we have

$$(a + b\sqrt{d})^p \equiv a^p + b^p(\sqrt{d})^p \pmod{p}.$$

By Fermat's Little Theorem $a^p \equiv a \pmod{p}$ and $b^p \equiv b \pmod{p}$, so the congruence can be further rewritten as

$$a^p + b^p(\sqrt{d})^p \equiv a + b(\sqrt{d})^p \pmod{p}.$$

It remains to deal with $(\sqrt{d})^p$. Since d is not a square modulo p,

$$d^{\frac{p-1}{2}} \equiv -1 \pmod{p}$$

by Euler's criterion. Thus

$$(\sqrt{d})^p = d^{\frac{p-1}{2}} \cdot \sqrt{d} \equiv -\sqrt{d} \pmod{p}$$

as desired. □

The explicit action of the Frobenius map on $\mathbb{F}_p[\sqrt{d}]$ gives another proof that the Frobenius fixes only elements in the subfield $\mathbb{F}_p$. Indeed, if $\mathrm{Fr}(z) = z$, for $z = x + y\sqrt{d}$ in $\mathbb{F}_p[\sqrt{d}]$, then the proposition implies that $y = 0$, i.e. z is in $\mathbb{F}_p$.

Uniqueness of fields of order p^2. Assume that F is a finite field of order p^2 with p odd. We shall show that F can be identified with the field $\mathbb{F}_p[\sqrt{d}]$. To that end, consider the multiplicative group $F^\times$. The order of $F^\times$ is $p^2 - 1$. By the theorem of Lagrange $x^{p^2-1} = 1$ for every element x in $F^\times$. By Proposition 20 (existence of primitive roots) $\varphi(p^2 - 1)$ of elements in $F^\times$ must be primitive, that is, of order exactly $p^2 - 1$. Let g be a primitive root, and define $h = g^{p+1}$. Since $p^2 - 1 = (p+1)(p-1)$, the order of h is precisely

$p-1$. It follows, by the observation above, that h is a primitive root in $\mathbb{F}_p^\times$. In particular, there exists an integer I such that $d = h^I$. Put

$$s = g^{\frac{p+1}{2}I}.$$

This is a well defined element in F since $p+1$ is even. Note that $s^2 = d$. In particular, s is not contained in the subfield $\mathbb{F}_p$ since d is not a square in $\mathbb{F}_p$ by the assumption. Consider a subset of F given by

$$F' = \{x + ys \mid x, y \in \mathbb{F}_p\}.$$

We want to show that $F' = F$. To that end it suffices to show that F' has p^2 elements. Assume that $x + ys = u + vs$. If $y \neq v$ then we can solve for s:

$$s = \frac{x-u}{v-y}.$$

This is a contradiction since s is not in $\mathbb{F}_p$. It follows that $x + ys = u + vs$ implies first that $y = v$ and then $x = u$ by canceling $ys = vs$ from both sides of $x + ys = u + vs$. Thus different choices for x and y give different elements in F' so F' has p^2 elements as desired.

We now know that any element in F can be uniquely written as $x + ys$ for some x and y in $\mathbb{F}_p$. Multiplying two elements in this expression gives

$$(x+ys)(u+vs) = xu + (xv+yu)s + yvs^2 = (xu + yvd) + (xv + yu)s$$

since $s^2 = d$. Note that this formula is identical to the one for the field $\mathbb{F}_p[\sqrt{d}]$. This shows that we can identify $\mathbb{F}_p[\sqrt{d}]$ and F by sending $x + y\sqrt{d}$ in $\mathbb{F}_p[\sqrt{d}]$ to $x + ys$ in F.

Summarizing, for any odd integer p we can construct a finite field of order p^2 as $\mathbb{F}_p[\sqrt{d}]$ by picking a non-square element d in $\mathbb{F}_p^\times$. Any other field of order p^2 can be identified with this field. In other words, there is only one field of order p^2, although there are different ways to write it down.

Now assume that F has 4 elements. Then elements in $F^\times$ are cube roots of 1. Thus, if ρ is a cube root different from 1 then $F^\times = \{1, \rho, \rho^2\}$ and multiplication table is clearly

$\cdot$	1	ρ	ρ^2
1	1	ρ	ρ^2
ρ	ρ	ρ^2	1
ρ^2	ρ^2	1	ρ

It remains to show that there is only one way to define addition on $F = \{0, 1, \rho, \rho^2\}$. Since

$$0 = \rho^3 - 1 = (\rho - 1)(\rho^2 + \rho + 1)$$

and $\rho \neq 1$ it follows that ρ satisfies the quadratic equation $\rho^2 + \rho + 1 = 0$. Since $-1 = 1$ in characteristic 2 this equation can be rewritten as $1 + \rho = \rho^2$,

$\rho + \rho^2 = 1$ and $\rho^2 + 1 = \rho$. In other words, the fact that ρ is a cube root of 1 forces the following addition table:

$+$	0	1	ρ	ρ^2
0	0	1	ρ	ρ^2
1	1	0	ρ^2	ρ
ρ	ρ	ρ^2	0	1
ρ^2	ρ^2	1	ρ	0

The diagonal terms are all 0 since $x + x = 2 \cdot x = 0 \cdot x = 0$ in any field of characteristic 2. We have shown that any field of order 4 can be identified with a field consisting of elements $\{0, 1, \rho, \rho^2\}$ with the two operations given by the two tables above.

The unique field of order p^2 is denoted by $\mathbb{F}_{p^2}$.

Exercises

1) Since $-1 = 2$ is not a square modulo 3, we can write $\mathbb{F}_9$ as $\mathbb{F}_3[i]$ where $i^2 = -1$. Write down the multiplication table for 8 non-zero elements in $\mathbb{F}_3[i]$. (For convenience write these elements as $a+bi$ where a, b are $-1, 0, 1$.)

2) Since the order of $\mathbb{F}_9^\times$ is 8 and $\varphi(8) = 4$, four elements in $\mathbb{F}_3[i]^\times$ should be primitive, that is, of order 8. Find them.

3. When is 2 a square modulo p?

We continue studying the problem of characterizing quadratic residues modulo p. In Proposition 23 it was shown that -1 is a square modulo p if and only if $p \equiv 1 \pmod 4$. We give another proof of this fact. Assume that -1 is a square modulo p, that is,

$$-1 \equiv a^2 \pmod p$$

for some a. This implies that the order of a in the group $(\mathbb{Z}/p\mathbb{Z})^\times$ is 4. Since the order of $(\mathbb{Z}/p\mathbb{Z})^\times$ is $p-1$ it follows, from the theorem of Lagrange, that 4 divides $p-1$ or $p \equiv 1 \pmod 4$. Conversely, assume that $p \equiv 1 \pmod 4$. Let g be a primitive root in $(\mathbb{Z}/p\mathbb{Z})^\times$. Then

$$-1 = g^{\frac{p-1}{2}} = (g^{\frac{p-1}{4}})^2,$$

hence -1 is a square. Notice the structure of this argument. First, the theorem of Lagrange is used to show that -1 could be a square only if $p \equiv 1 \pmod 4$. Second, if $p \equiv 1 \pmod 4$ then an *explicit* construction of a square root of -1 is given using the primitive root. This sort of thinking is useful in proving the following case of the Quadratic Reciprocity.

Proposition 26. *Let p be an odd prime. Then:*

(1) 2 *is a square modulo p if and only if* $p \equiv 1, 7 \pmod 8$.
(2) 2 *is not a square modulo p if and only if* $p \equiv 3, 5 \pmod 8$.

The proof of this result is based on an observation concerning complex 8-th roots of one. They are powers $1, \zeta, \dots, \zeta^7$ where

$$\zeta = \exp(\frac{\pi i}{4}) = \frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}}.$$

Notice the appearance of $\sqrt{2}$ here. The root ζ is expressed in terms of $\sqrt{2}$. Conversely, it is possible to write $\sqrt{2}$ in terms of ζ. Indeed, since

$$\zeta^7 = \exp(\frac{7\pi i}{4}) = \frac{1}{\sqrt{2}} - \frac{i}{\sqrt{2}}.$$

it follows that

$$\zeta + \zeta^7 = 2 \cdot \frac{1}{\sqrt{2}} = \sqrt{2}.$$

Now let's move to finite fields. Consider, for example, $\mathbb{Z}/17\mathbb{Z}$. Since $17 \equiv 1 \pmod 8$, 2 should be a square modulo 17. A square root of 2 modulo 17 can be constructed by following how the real $\sqrt{2}$ is expressed in terms of the complex 8-th roots of 1. To that end, recall that 3 is a primitive root modulo 17. In particular, the order of 3 modulo 17 is 16. It follows that $9 = 3^2$ is a primitive 8-th root modulo 17. By analogy with complex numbers we consider

$$9 + 9^7 \equiv 11 \pmod{17}.$$

Then $11^2 = 121 \equiv 2 \pmod{17}$, thus our construction of a square root of 2 works in this case. The same argument works for all primes p such that $p \equiv 1 \pmod 8$. However, in order to give a uniform proof of the theorem in all cases, we must work with the field $\mathbb{F}_{p^2}$. If p is an odd prime - in fact if p is any odd number - then

$$p \equiv 1, 3, 5 \text{ or } 7 \pmod 8.$$

Thus, since $1^2 = 1, 3^2 = 9, 5^2 = 25$ and $7^2 = 49$ are all congruent to 1 modulo 8,

$$p^2 \equiv 1 \pmod 8.$$

It follows that $p^2 - 1$, the order of the multiplicative group of the quadratic extension $\mathbb{F}_{p^2}$, is divisible by 8. Next, by the theorem of Lagrange, every element x in $\mathbb{F}_{p^2}^\times$ satisfies the equation

$$x^{p^2-1} = 1.$$

Thus, all non-zero elements of $\mathbb{F}_{p^2}$ are roots of 1 and, by Proposition 20 (existence of primitive roots), there is a primitive root g (of order $p^2 - 1$). In particular,

$$\zeta = g^{\frac{k(p^2-1)}{8}}$$

is a primitive 8-th roots of 1, and by analogy with complex numbers,

$$s = \zeta + \zeta^7$$

should be a square root of 2. This is not difficult to check. Indeed, squaring s and using that $\zeta^8 = 1$ gives

$$s^2 = \zeta^2 + 2\zeta^8 + \zeta^{14} = \zeta^2 + 2 + \zeta^6.$$

Since ζ is a primitive 8-th root of 1, ζ^4 is a primitive square root of 1. Hence $\zeta^4 = -1$. Since $\zeta^6 = \zeta^4 \cdot \zeta^2 = -\zeta^2$, the first and the last summand in the above expression for s^2 cancel each other out. This shows that $s^2 = 2$ or, in words, 2 has a square root in the field $\mathbb{F}_{p^2}$. The question now is whether or not s belongs to the base field $\mathbb{F}_p$. This is checked by applying the Frobenius, $\mathrm{Fr}(s) = s^p$. Recall that, by Proposition 14, an element s in $\mathbb{F}_{p^2}$ belongs to $\mathbb{F}_p$ if and only if $\mathrm{Fr}(s) = s$.

Let $k = 1, 3, 5, 7$ be such that $p \equiv k \pmod 8$. Since raising an eight root of 1 to the p-th power is the same as raising it to the k-th power,

$$\mathrm{Fr}(s) = \mathrm{Fr}\left(\zeta\right) + \mathrm{Fr}\left(\zeta^7\right) = \zeta^k + \zeta^{7k}.$$

The second summand depends only on what $7k$ is modulo 8. Now if $k = 1$ then $\mathrm{Fr}(s) = s$, and 2 is a square modulo p. If $k = 7$ then $7 \cdot 7 \equiv 1 \pmod 8$ and $\mathrm{Fr}(s) = s$ again, as Frobenius simply switches the two terms. Thus 2 is a square if $p \equiv 1, 7 \pmod 8$. If $k = 5$ then $7 \cdot 5 \equiv 11 \pmod 8$ and

$$\mathrm{Fr}(s) = \zeta^5 + \zeta^{11} = \zeta^4(\zeta + \zeta^7) = -(\zeta + \zeta^7) = -s.$$

A similar argument shows that $\mathrm{Fr}(s) = -s$ if $k = 3$. Thus 2 is a not square if $p \equiv 3, 5 \pmod 8$. The proposition is proved.

Exercises

1) Notice that $41 \equiv 1 \pmod 8$. Use 6, as a primitive root modulo 41, to construct a square root of 2 modulo 41.

2) Notice that $73 \equiv 1 \pmod 8$. Use 5, as a primitive root modulo 73, to construct a square root of 2 modulo 73.

3) Recall that

$$\frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2}$$

is a complex primitive 8-th root of 1. Notice how it is constructed from square roots of 2 and -1. Construct a primitive 8-th root of 1 modulo 17 using 6 and 4 which are square roots of 2 and -1 modulo 17.

4) Does the equation $x^2 - 6x + 11 = 0$ have a solution modulo 131? Hint: complete to a square, then use quadratic reciprocity.

4. Quadratic Reciprocity

THEOREM 27. *Quadratic Reciprocity. Let p and q be two different odd primes. Then*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}.$$

Note that the sign $(-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}$ is -1 only when both primes are congruent to 3 modulo 4. Thus, it is convenient to note that the quadratic reciprocity can be restated as

$$\left(\frac{p}{q}\right) = \left(\frac{q}{p}\right) \text{ if } p \equiv 1 \pmod 4,$$

$$\left(\frac{-p}{q}\right) = \left(\frac{q}{p}\right) \text{ if } p \equiv 3 \pmod 4.$$

The quadratic reciprocity, combined with the multiplicative property of the Legendre symbol, can be used to figure out quickly whether an integer is a quadratic residue. For example, say we want to figure out if 71 is a square modulo 131. Since 71 is a prime then, using thequadratic reciprocity,

$$\left(\frac{71}{131}\right) = \left(\frac{131}{71}\right) = \left(\frac{60}{71}\right)$$

where, for the second equality, we used that $131 \equiv 60 \pmod{71}$. Since $60 = 4 \cdot 3 \cdot 5$, by the multiplicativity property of the Legendre's symbol, we further have

$$\left(\frac{71}{131}\right) = \left(\frac{4}{71}\right)\left(\frac{3}{71}\right)\left(\frac{5}{71}\right).$$

Since $4 = 2^2$, the first symbol on the right side is 1. After applying the quadratic reciprocity to the second and third symbol,

$$\left(\frac{71}{131}\right) = -\left(\frac{71}{3}\right)\left(\frac{71}{5}\right).$$

Since $71 \equiv 2 \pmod 3$ is not a square and $71 \equiv 4 \pmod 5$ is a square, it follows that 71 is a square modulo 131.

We shall now give a proof of the quadratic reciprocity in a special case when one of the primes is 3. In fact, it will be more convenient to work with -3. The quadratic reciprocity (the second formulation) says that

$$\left(\frac{-3}{p}\right) = \left(\frac{p}{3}\right).$$

In words, -3 is a square modulo p if and only if p is a square modulo 3. Since $p \equiv 1 \pmod 3$ or $p \equiv 2 \pmod 3$, and 1 is a square modulo 3, while 2 is not, the quadratic reciprocity states that -3 is a square modulo p if and only if $p \equiv 1 \mod 3$. We shall now give a proof of this statement using cube roots of 1. Again, we use complex roots as a guide. Recall that

$$\zeta = \exp(\frac{2\pi i}{3}) = -\frac{1}{2} + \frac{\sqrt{-3}}{2}$$

is a primitive complex cube root of 1. Then

$$\zeta^2 = \exp(\frac{4\pi i}{3}) = -\frac{1}{2} - \frac{\sqrt{-3}}{2}.$$

Hence

$$\zeta - \zeta^2 = \sqrt{-3}.$$

Thus, the idea is to construct a square root of -3 modulo p using cube roots of 1. Consider, for example, $\mathbb{Z}/7\mathbb{Z}$. Since $2^3 \equiv 1 \pmod 7$, 2 is a primitive cube root of 1 and, by the analogy with complex numbers,

$$2 - 2^2 \equiv 5 \pmod 7$$

should be a square root of -3 modulo 7. Indeed $5^2 \equiv -3 \pmod 7$. If $p \equiv 2 \pmod 3$ then $\mathbb{F}_p$ does not contain cube roots of 1 and we need to use the larger field $\mathbb{F}_{p^2}$. Note that

$$p^2 \equiv 1 \pmod 3,$$

for any odd prime p. Let g be a primitive root (of order $p^2 - 1$) in $\mathbb{F}_{p^2}$. Then

$$\zeta = g^{\frac{p^2-1}{3}}$$

is a primitive cube root of 1 and

$$t = \zeta - \zeta^2$$

should be a square root of -3. Let us check this. Squaring t and using $\zeta^3 = 1$ gives

$$t^2 = \zeta^2 - 2\zeta^3 + \zeta^4 = \zeta^2 - 2 + \zeta.$$

Since the sum of all three cube roots is 0 $(1 + \zeta + \zeta^2 = 0)$ it follows that $t^2 = -3$, as desired. The remaining question is whether t is in $\mathbb{F}_p$. Again, we use the Frobenius. Let $k = 1, 2$ be such that $p \equiv k \pmod 3$. Since raising a cube root of 1 to the p-th power is the same as raising it to the k-th power,

$$\mathrm{Fr}(t) = \mathrm{Fr}\left(\zeta\right) - \mathrm{Fr}\left(\zeta^2\right) = \zeta^k - \zeta^{2k}.$$

If $k = 1$ then $\mathrm{Fr}(t) = t$, hence t is in $\mathbb{F}_p$. If $k = 2$ then, using $\zeta^4 = \zeta$, the Frobenius switches the two summands of t

$$\mathrm{Fr}(t) = \zeta^2 - \zeta^4 = \zeta^2 - \zeta = -(\zeta - \zeta^2) = -t.$$

Hence t is not in $\mathbb{F}_p$.

A general proof of the quadratic reciprocity proceeds along similar lines. We give a brief sketch. Let q be an odd prime, and let

$$q^* = \begin{cases} q \text{ if } q \equiv 1 \pmod 4 \\ -q \text{ if } q \equiv 3 \pmod 4). \end{cases}$$

Let ζ be a primitive root of 1 of order q. Gauss discovered that $\sqrt{q^*}$ can be computed using ζ by the following formula (Gauss sum)

$$\sqrt{q^*} = \sum_{x \in S} \zeta^x - \sum_{x \in N} \zeta^x$$

where S is the set of squares in $(\mathbb{Z}/q\mathbb{Z})^\times$ and N is the set of non-squares in $(\mathbb{Z}/q\mathbb{Z})^\times$. If p is a square modulo q the Frobenius fixes the Gauss sum. If p is not a square modulo q the Frobenius switches the two summands in the Gauss sum and, therefore, changes the sign of the Gauss sum. Hence q^* is a square modulo p if and only p is a square modulo q. This is the second formulation of the quadratic reciprocity.

Exercises

1) Use 3 as a primitive root modulo 43 to write down a square root of -3 using cube roots modulo 43.

2) Use 5 as a primitive root modulo 73 to write down a square root of -3 using cube roots modulo 73.

3) Use quadratic reciprocity to determine odd primes p such that 5 a square.

4) Using the quadratic reciprocity determine whether 66 and 80 are squares modulo 127.

5) Does the equation $x^2 - 6x + 28 = 0$ have a solution modulo 131? Hint: complete to a square, then use quadratic reciprocity.

6) Let $\zeta = e^{\frac{2\pi i}{5}}$ be a complex primitive 5-th root of 1. Verify Gauss's formula

$$\sqrt{5} = \zeta + \zeta^4 - \zeta^2 - \zeta^3.$$

Let p be a prime such that $p \equiv 1 \pmod 5$. Let g be a primitive root in $\mathbb{Z}/p\mathbb{Z}$. Then $\zeta = g^{\frac{p-1}{5}}$ is a primitive 5-th root of 1 in $\mathbb{Z}/p\mathbb{Z}$. A square root of 5 is given by the Gauss sum

$$\zeta + \zeta^4 - \zeta^2 - \zeta^3.$$

Use this formula to construct a square root of 5 in the following two exercises.

7) Use the primitive root $g = 3$ modulo 31 to find a square root of 5 modulo 31.

8) Use the primitive root $g = 2$ modulo 61 to find a square root of 5 modulo 61.

9) Let n be a positive integer. Let p be a prime divisor of $n^2 + 3$. Prove that $p \equiv 1 \pmod 3$. Hint: Use

$$\left(\frac{n^2}{p}\right) = \left(\frac{-3}{p}\right).$$

10) Use the previous exercise to prove that there are infinitely many primes $p \equiv 1 \pmod 3$.

CHAPTER 7

Applications of Quadratic Reciprocity

1. Fermat primes

In this section we shall use quadratic reciprocity to study Fermat numbers

$$F_n = 2^{2^n} + 1.$$

A Fermat number may or may not be prime. It is not known if there are infinitely many Fermat primes. There are only 5 known Fermat primes. They are:

n	F_n
0	3
1	5
2	17
3	257
4	65537

It is interesting to note that the known Fermat primes are precisely the first 5 Fermat numbers. In fact, Fermat claimed that

$$F_5 = 2^{2^5} + 1 = 4294967297,$$

is also prime. However, about 100 years after Fermat, Euler found the following factorization

$$4294967297 = 641 \cdot 6700417.$$

By 2003 it was known that F_n is composite for $5 \leq n \leq 32$. This is by no means easy to verify since the numbers F_n quickly become huge. Indeed,

$$F_{n+1} = (F_n - 1)^2 + 1$$

so the number of digits of F_{n+1} is roughly 2 times the number of digits of F_n. For example, F_5 has 10 digits, F_6 has 20 digits and F_7 has 39 digits:

$$F_7 = 340282366920938463463374607431768211457.$$

A prime factorization of F_7 was obtained in 1975 by Morrison and Brillhart:

$$F_7 = 59649589127497217 \cdot 5704689200685129054721.$$

The presence of large primes in this factorization illustrates how difficult it is to decide whether a Fermat number is composite. However, there is a very efficient test (Pépin's test) to decide whether a Fermat number F_n is composite or not. We emphasize that the proof of the test exploits that, if F_n is prime, the order of the group $(\mathbb{Z}/F_n\mathbb{Z})^\times$ is a pure power of 2. Indeed,

$$|(\mathbb{Z}/F_n\mathbb{Z})^\times| = F_n - 1 = 2^{2^n}.$$

Theorem 28. *Let $n \geq 1$. The Fermat number F_n is prime if and only if*

$$3^{\frac{F_n-1}{2}} \equiv -1 \pmod{F_n}.$$

Proof. Assume first that F_n is prime. We need to show that the congruence holds. Recall that, by Euler's criterion,

$$3^{\frac{F_n-1}{2}} \equiv \left(\frac{3}{F_n}\right) \pmod{F_n}.$$

Thus, we need to show that 3 is not a square modulo F_n. Since $F_n \equiv 1 \pmod 4$ the quadratic reciprocity implies that

$$\left(\frac{3}{F_n}\right) = \left(\frac{F_n}{3}\right).$$

Furthermore, since $2 \equiv -1 \pmod 3$,

$$F_n = 2^{2^n} + 1 \equiv (-1)^{2^n} + 1 = 2 \pmod 3.$$

Thus, we have

$$\left(\frac{3}{F_n}\right) = \left(\frac{2}{3}\right) = -1$$

and the congruence holds. To prove the converse, we need the following lemma.

Lemma 29. *Let G be a group with identity e. Let g be an element in G such that $g^{2^{k-1}} \neq e$ and $g^{2^k} = e$. Then the order of g is 2^k.*

Proof. The order of g divides 2^k since $g^{2^k} = e$. Now note that all divisors of 2^k are

$$1, 2, 2^2, \dots, 2^{k-1} \text{ and } 2^k.$$

Thus, if the order of g is less then 2^k, then it divides 2^{k-1}. But then $g^{2^{k-1}} = e$. This is a contradiction. The lemma is proved. □

Now we can finish the proof of the Theorem. Let p be a prime factor of F_n. If

$$3^{\frac{F_n-1}{2}} \equiv -1 \pmod{F_n}$$

then, since p divides F_n,

$$3^{\frac{F_n-1}{2}} \equiv -1 \pmod{p}.$$

By squaring both sides,

$$3^{F_n-1} \equiv 1 \pmod{p}.$$

Since $F_n - 1$ is a power of two, Lemma 29 implies that the order of 3 in the group $(\mathbb{Z}/p\mathbb{Z})^\times$ is precisely $F_n - 1$. Since the order of a group element is less than or equal to the order of the group, it follows that

$$F_n - 1 \leq p - 1$$

or $F_n \leq p$. Hence $p = F_n$, that is, F_n is prime. □

Note, however, that the test does not give a factorization of F_n if F_n is shown to be composite. Thus, by the end of 2005, factorizations of F_{14}, F_{20}, F_{22} and F_{24} were not known.

Let us apply the Pepin test to $F_3 = 257$. We need to calculate 3^{128} modulo 257. This can be done efficiently using successive squaring:

$$\begin{array}{rcrl}
3^2 & \equiv & 9 & \pmod{257} \\
3^4 & \equiv & 81 & \pmod{257} \\
3^8 & \equiv & 136 & \pmod{257} \\
3^{16} & \equiv & -8 & \pmod{257} \\
3^{32} & \equiv & 64 & \pmod{257} \\
3^{64} & \equiv & -16 & \pmod{257} \\
3^{128} & \equiv & -1 & \pmod{257}
\end{array}$$

which confirms that $F_3 = 257$ is prime. Note that we needed 7 steps to run the test. (7 comes from $128 = 2^7$.) In general, the number of consecutive squaring needed to run the test is equal to

$$\log_2((F_n - 1)/2) = 2^n - 1.$$

For example, if $n = 7$, then the number of steps needed to determine that F_7 is composite is equal to 127. This number pales in comparison to the actual prime factors of F_7.

The idea behind the Pepin's these can be described in terms of the problem of determining the order of a group element g. If the order of g is n, this can be verified by multiplying the group element n times. However, if $n = 2^m$, then the order of g can be determined by consecutive squaring

$$g^2, g^4 = (g^2)^2, \ldots, g^{2^m} = (g^{2^{m-1}})^2,$$

a process consisting of $m = \log_2(n)$ steps. In particular, if the order of the ambient group is a power of 2, then the order of any element can be determined by consecutive squaring.

Exercises

1) Use Pepin's test to show that F_4 is prime.

2) Let $n \geq 2$. Show that $F_n \equiv 2 \pmod{5}$.

3) Let $n \geq 2$. Assume that F_n is prime. Show that $\left(\frac{5}{F_n}\right) = -1$.

4) Use the previous exercise to prove the following version of Pepin's test for Fermat's numbers F_n if $n \geq 2$: A Fermat number F_n is prime if and only if

$$5^{\frac{F_n-1}{2}} \equiv -1 \pmod{F_n}.$$

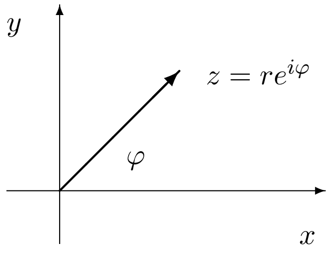
2. Quadratic Fields and the Circle Group

Pepin's test is based on the fact that the order of the multiplicative group $(\mathbb{Z}/F_n\mathbb{Z})^\times$ is a pure power of 2, if Fermat number $F_n = 2^{2^n} + 1$ is prime. A purpose of this section is to introduce a subgroup $T(p)$ of $\mathbb{F}_{p^2}^\times$ of order $p+1$. If $M_\ell = 2^\ell - 1$ is a Mersenne prime then the order of $T(M_\ell)$ is 2^ℓ, a pure power of 2. Using this observation group the Lucas Lehmer test can be formulated and proved in a way analogous to Pepin's test.

The group $T(p)$ is analogous to the multiplicative group of complex numbers of norm 1. Recall that the norm of a complex number $z = x + yi$ is

$$N(z) = z\bar{z} = (x+yi)(x-yi).$$

The complex number z can be written as $z = r \exp i\varphi$ where $r = \sqrt{N(z)}$ and φ the argument of z. This is the angle that z forms with respect to the x-axis.



Multiplication of two complex numbers in this form is given by

$$(r_1 e^{i\varphi_1})(r_2 e^{i\varphi_2}) = (r_1 r_2) e^{i(\varphi_1 + \varphi_2)}.$$

In particular, if $r_1 = r_2 = 1$ then the group law amounts to adding the arguments. This shows that the set of all complex numbers of norm 1 is a group with respect to multiplication, denoted by $\mathbb{T}$. Geometrically $\mathbb{T}$ is a circle in the plane of complex numbers.

Pick an integer d which is not a square (i.e. not a quadratic residue) modulo p. Recall that the finite field $\mathbb{F}_{p^2}$ can be realized as the set of numbers

$$z = x + y\sqrt{d}$$

where x and y are elements of the finite field $\mathbb{F}_p$, that is, integers considered modulo p. Let $\bar{z} = x - yi$ and define the norm $N(z)$ of z by

$$N(z) = z\bar{z} = (x + y\sqrt{d})(x - y\sqrt{d}) = x^2 - dy^2.$$

By analogy with complex numbers, define

$$T(p) = \{z = x + y\sqrt{d} \in \mathbb{F}_{p^2} \mid N(z) = 1\}.$$

Since $N(z_1z_2) = N(z_1)N(z_2)$, the product of two numbers with norm one has norm one. Therefore the set $T(p)$ is closed under multiplication. The inverse of z in $T(p)$ is simply $\bar{z}$. It follows that $T(p)$ is a subgroup of $\mathbb{F}_{p^2}^\times$. We shall now prove that $T(p)$ is a cyclic group of order $p+1$. The key is that the Frobenius map $\mathrm{Fr}(z) = z^p$ on $\mathbb{F}_{p^2}$ coincides with conjugation, as proved in Proposition 25:

$$(x + y\sqrt{d})^p \equiv x - y\sqrt{d} \pmod{p}.$$

Proposition 30. *The circle group $T(p)$ is a cyclic group of order $p+1$.*

Proof. Let γ be a primitive root in $\mathbb{F}_{p^2}$. Then the order of γ is $p^2 - 1 = (p-1)(p+1)$. Any element α in $\mathbb{F}_{p^2}^\times$ can be written as $\alpha = \gamma^I$ for a unique integer $1 \leq I \leq p^2 - 1$. If α is in $T(p)$ then

$$1 = N(\alpha) = \alpha \cdot \bar{\alpha} = \alpha \cdot \alpha^p = \alpha^{p+1} = \gamma^{I(p+1)}.$$

This shows that $p^2 - 1 = (p-1)(p+1)$, the order of γ, divides $I(p+1)$. It follows that $p-1$ divides I and the elements in $T(p)$ are

$$\gamma^{(p-1)}, \gamma^{2(p-1)}, \ldots, \gamma^{(p+1)(p-1)},$$

$p+1$ of them in all. Moreover, the order of γ^{p-1} is clearly $p+1$, completing the proof. □

Example: Let $p = 7$, and write $\mathbb{F}_{7^2}$ as the set of elements $x + iy$ where x and y are integers modulo 7. Then there are 8 norm one elements. They are

$$T(7) = \{\pm 1, \pm i, \pm(2+2i), \pm(2-2i)\}.$$

The order of $2+2i$ is 8. Indeed, squaring $2+2i$ gives

$$(2+2i)^2 = 0 + 8i \equiv i \pmod{7}$$

and, since the order of i is four, $2+2i$ has the order 8. In particular, the group $T(7)$ is cyclic.

In the proof of Lucas - Lehmer test it will be important to decide whether an element α in $T(p)$ is a square of another element in $T(p)$. (We are not discussing here whether α is a square in $\mathbb{F}_{p^2}^\times$.) Since -1 is contained in $T(p)$, the squaring map $\alpha \mapsto \alpha^2$ is two to one. In particular, $(p+1)/2$ elements in $T(p)$ are squares and the other $(p+1)/2$ elements in $T(p)$ are non-squares.

Let β be an element of $T(p)$ of order $p+1$. Then β^I for I even is clearly a square. Therefore β^I for

$$I = 2, 4, \ldots, p-1, p+1$$

give $(p+1)/2$ different squares in $T(p)$. These are, therefore, all squares in $T(p)$. It follows that β^I is a square or not depending whether I is even or odd. Finally, notice that $\beta^{\frac{p+1}{2}}$ is an element of order 2 in $T(p)$. There is only one element of order 2 in the whole multiplicative group $\mathbb{F}_{p^2}^{\times}$. That element is -1. Hence

$$\beta^{\frac{p+1}{2}} = -1.$$

We can now show the following analogue of Euler's criterion for the group $T(p)$.

Proposition 31. *Let p be an odd prime. For every α in the circle group $T(p)$*

$$\alpha^{\frac{p+1}{2}} = \begin{cases} 1 \text{ if } \alpha \text{ is a square in } T(p) \\ -1 \text{ if } \alpha \text{ is not a square .} \end{cases}$$

Proof. If α is square, we can write $\alpha = \beta^{2I}$ for some I. Then

$$\alpha^{\frac{p+1}{2}} = (\beta^{2I})^{\frac{p+1}{2}} = (\beta^{p+1})^I = 1.$$

If α is not a square, we can write $\alpha = \beta^{2I+1}$. Then

$$\alpha^{\frac{p+1}{2}} = (\beta^{(2I+1)})^{\frac{p+1}{2}} = (\beta^{p+1})^I \cdot \beta^{\frac{p+1}{2}} = 1 \cdot (-1) = -1.$$

□

Exercises

1) Let p be a prime congruent to 3 modulo 4. Then $\mathbb{F}_{p^2}$ can be realized as the set of elements $x + yi$ where x and y are integers considered modulo p and $i^2 = -1$. Show that i is a square in $T(p)$ if and only if $p \equiv 7 \pmod 8$.

2) The field $\mathbb{F}_{11^2}$ can be realized as the set of elements $x + yi$ where x and y are integers considered modulo p and $i^2 = -1$. List all 12 elements in the group $T(11)$. Find a generator of this group.

3. Lucas Lehmer test revisited

We are now ready to give a proof of the Lucas Lehmer test for Mersenne primes. The main tool is the group $T(p)$ introduced in the previous section.

Proposition 32. *Let ℓ be an odd prime. Then there exists a prime divisor p of the Mersenne number $M_\ell = 2^\ell - 1$ such that $\left(\frac{3}{p}\right) = -1$. In particular, if M_ℓ is a prime, then 3 is not a square modulo M_ℓ.*

PROOF. We start by noting that M_ℓ satisfies a couple of simple congruences. Since $2 \equiv -1 \pmod 3$ and ℓ is odd,

$$M_\ell \equiv (-1)^\ell - 1 = -2 \equiv 1 \pmod 3.$$

This implies M_ℓ is a square modulo 3. Next, since 4 divides 2^ℓ, we have $M_\ell \equiv -1 \pmod 4$. Let $M_\ell = p_1 \cdots p_n$ be a factorization of M_ℓ into primes. Since $M_\ell \equiv -1 \pmod 4$, an *odd* number of prime factors p_i in the factorization of M_ℓ satisfy $p_i \equiv -1 \pmod 4$. Using the mulitiplicativity property of the Legendre's symbol,

$$1 = \left(\frac{M_\ell}{3}\right) = \left(\frac{p_1}{3}\right) \cdot \ldots \cdot \left(\frac{p_1}{3}\right).$$

Next, using the quadratic reciprocity, the right hand side can be rewritten as

$$1 = (-1)^s \left(\frac{3}{p_1}\right) \cdot \ldots \cdot \left(\frac{3}{p_1}\right)$$

where s is the number of prime factors p_i such that $p_i \equiv -1 \pmod 4$. We know that s is odd. Hence $(-1)^s = -1$ and

$$\left(\frac{3}{p_1}\right) \cdot \ldots \cdot \left(\frac{3}{p_1}\right) = -1.$$

This implies that one of the Legendre's symbols is -1. The proposition is proved. □

Let p be an odd prime such that 3 is not a quadratic residue modulo p. Then the field $\mathbb{F}_{p^2}$ can be realized as

$$\mathbb{F}_{p^2} = \{x + y\sqrt{3} \mid x, y \in \mathbb{F}_{p^2}\}.$$

Let $\alpha = 2 + \sqrt{3}$, and let $\bar{\alpha} = 2 - \sqrt{3}$. Then, as real numbers,

$$\alpha \cdot \bar{\alpha} = (2 + \sqrt{3})(2 - \sqrt{3}) = 1.$$

In particular, α can be viewed as an element in $T(p)$ for all p such that 3 is not a square modulo p.

LEMMA 33. *Assume that $M_\ell = 2^\ell - 1$ is a Mersenne prime. Then $\alpha = 2 + \sqrt{3}$ is not a square of an element in $T(M_\ell)$.*

PROOF. Assume that $2+\sqrt{3}$ is a square of an element $x+y\sqrt{3}$ in $T(M_\ell)$, that is,

$$(x + y\sqrt{3})^2 = x^2 + 3y^2 + 2xy\sqrt{3} = 2 + \sqrt{3}.$$

This implies that $x^2 + 3y^2 = 2$. Since $x + y\sqrt{3}$ is in $T(M_\ell)$ we also have $x^2 - 3y^2 = 1$. Adding this two equations gives

$$2x^2 = 3.$$

Since $M_\ell \equiv 1 \pmod 8$, 2 is a square mod M_ℓ. It follows that 3 is a square. A contradiction, since 3 is not a square modulo a Mersenne prime. The lemma is proved. □

Now we are ready to prove a version (the first version) of the Lucas - Lehmer test for Mersenne primes. Note a similarity between this test and Pepin's test for Fermat's primes.

THEOREM 34. *Let ℓ be an odd prime. Let $\alpha = 2+\sqrt{3}$. Then $M_\ell = 2^\ell - 1$ is a prime if and only if*

$$\alpha^{2^{\ell-1}} \equiv -1 \pmod{M_\ell}$$

PROOF. Assume that the congruence is satisfied. We want to show that M_ℓ is a prime number. By Proposition 32 there exists a prime divisor p of M_ℓ such that 3 is not a square modulo p. We can view α as an element in $T(p)$. The congruence

$$\alpha^{2^{\ell-1}} \equiv -1 \pmod{p}$$

implies that the order of α in $T(p)$ is $2^\ell = M_\ell + 1$. On the other hand, α, as an element of $T(p)$, has the order less than or equal to the order of $T(p)$. Therefore

$$M_\ell + 1 \leq p + 1$$

and, after subtracting 1 from both sides, $M_\ell \leq p$. Hence $M_\ell = p$, that is, M_ℓ is a prime number.

Conversely, assume that M_ℓ is a prime. By Lemma 33, α is not a square in $T(M_\ell)$. By Proposition 31 (the analogue of Euler's criterion for the circle group)

$$\alpha^{2^{\ell-1}} \equiv -1 \pmod{M_\ell}.$$

The theorem is proved. □

Example: Consider $\ell = 5$, so the Mersenne number is 31. We need to compute $(2+\sqrt{3})^{16}$ modulo 31. This is done by consecutive squaring, four times:

$$\begin{array}{rcll}
(2+\sqrt{3})^2 & \equiv & 7+4\sqrt{3} & \pmod{31} \\
(7+4\sqrt{3})^2 & \equiv & 2-3\sqrt{3} & \pmod{31} \\
(2-3\sqrt{3})^2 & \equiv & 0-12\sqrt{3} & \pmod{31} \\
(0-12\sqrt{3})^2 & \equiv & -1+0\sqrt{3} & \pmod{31}
\end{array}$$

This shows that $(2+\sqrt{3})^{16}$ is congruent to -1 modulo 31. Thus 31 is prime by the test.

On the surface the test given by Theorem 34 appears unrelated to the Lucas - Lehmer test introduced in Section ???:

THEOREM 35. *(Lucas-Lehmer) Define recursively a sequence s_n of integers by $s_1 = 4$ and $s_{n+1} = s_n^2 - 2$. Let ℓ be an odd prime. Then $M_\ell = 2^\ell - 1$ is prime if and only if $s_{\ell-1} \equiv 0 \pmod{M_\ell}$.*

How do we relate the two tests? It turns out that the sequence s_n is defined using $\alpha = 2+\sqrt{3}$. Let $\beta = 2-\sqrt{3}$ and define a sequence t_n for

$n = 1, 2, \ldots$ by

$$t_n = \alpha^{2^{n-1}} + \beta^{2^{n-1}}.$$

We claim that $s_n = t_n$ for all integers n. Since

$$t_1 = \alpha + \beta = 4$$

we have that $t_1 = s_1$. In order to show that $t_n = s_n$ for all n, it suffices to show that t_n satisfy the same recursive relation as s_n. This is not difficult at all. Since $\alpha\beta = 1$,

$$t_n^2 - 2 = (\alpha^{2^{n-1}} + \beta^{2^{n-1}})^2 - 2 = \alpha^{2^n} + \beta^{2^n} = t_{n+1},$$

and the two sequences are the same.

We can now derive the classical Lucas - Lehmer test from Theorem 34. Assume that M_ℓ divides $s_{\ell-1}$. We want to show that M_ℓ is prime. The congruence

$$s_{\ell-1} = \alpha^{2^{\ell-2}} + \beta^{2^{\ell-2}} \equiv 0 \pmod{M_\ell}$$

implies that

$$\alpha^{2^{\ell-2}} \equiv -\beta^{2^{\ell-2}} \pmod{M_\ell}.$$

Multiply both sides of this congruence by $\alpha^{2^{\ell-2}}$. Since $\alpha^{2^{\ell-2}} \cdot \alpha^{2^{\ell-2}} = \alpha^{2^{\ell-1}}$ and $\alpha^{2^{\ell-2}} \cdot \beta^{2^{\ell-2}} = 1$, we obtain that

$$\alpha^{2^{\ell-1}} \equiv -1 \pmod{M_\ell}.$$

It follows that M_ℓ is prime by Theorem 34.

The converse is just as easy. Assume that M_ℓ is a prime. We want to show that M_ℓ divides $s_{\ell-1}$. By Theorem 34,

$$\alpha^{2^{\ell-1}} \equiv -1 \pmod{M_\ell}.$$

By changing the sign of both sides and factoring $\alpha^{2^{\ell-1}} = \alpha^{2^{\ell-2}} \cdot \alpha^{2^{\ell-2}}$ the congruence can be rewritten as

$$-\alpha^{2^{\ell-2}} \cdot \alpha^{2^\ell} \equiv 1 \pmod{M_\ell}.$$

This implies that $-\alpha^{2^{\ell-2}}$ is a multiplicative inverse of $\alpha^{2^{\ell-2}}$. But $\beta^{2^{\ell-2}}$ is also an inverse of $\alpha^{2^{\ell-2}}$. By uniqueness of the multiplicative inverse in a field,

$$\beta^{2^{\ell-2}} = -\alpha^{2^{\ell-2}}$$

and

$$\alpha^{2^{\ell-2}} + \beta^{2^{\ell-2}} \equiv 0 \pmod{M_\ell}.$$

This shows that $s_{\ell-1} = \alpha^{2^{\ell-2}} + \beta^{2^{\ell-2}}$ is divisible by M_ℓ, as desired.

Of course, there is noting terribly special about the number $2 + \sqrt{3}$. It is possible to develop many different yet similar versions of the Lucas-Lehmer test by replacing $2 + \sqrt{3}$. For example, we can take

$$\alpha = \frac{3}{2} + \frac{1}{2}\sqrt{5}.$$

Let p be an odd prime such that 5 is not a quadratic residue modulo p. Then the field $\mathbb{F}_{p^2}$ can be realized as

$$\mathbb{F}_{p^2} = \{x + y\sqrt{5} \mid x, y \in \mathbb{F}_{p^2}\}.$$

Since 2 is invertible modulo p, the number $\alpha = \frac{3}{2} + \frac{1}{2}\sqrt{5}$ can be viewed as an element of $\mathbb{F}_{p^2}$. Moreover, since

$$\alpha\bar{\alpha} = \left(\frac{3}{2} + \frac{1}{2}\sqrt{5}\right)\left(\frac{3}{2} - \frac{1}{2}\sqrt{5}\right) = 1,$$

α is an element of $T(p)$.

Now assume that ℓ is an odd prime such that M_ℓ is a Mersenne prime. If we can show that

- 5 is not a square modulo M_ℓ,
- α is not a square in $T(M_\ell)$,

then we can develop a version of the Lucas - Lehmer test using $\frac{3}{2} + \frac{1}{2}\sqrt{5}$ instead of $2 + \sqrt{3}$. It turns out that the two bullets hold if $\ell \equiv 3 \pmod 4$. In particular, we get a primality test for the Mersenne numbers M_ℓ such that $\ell \equiv 3 \pmod 4$:

THEOREM 36. *Define a recursive sequence u_n of integers by $u_1 = 3$ and $u_{n+1} = u_n^2 - 2$. If ℓ is a prime congruent to 3 modulo 4, then the Mersenne number M_ℓ is prime if and only if*

$$u_{\ell-1} \equiv 0 \pmod{M_\ell}.$$

A proof of this test is given as a sequence of exercises below. As an example, take $\ell = 7$. Then $M_7 = 2^7 - 1 = 127$. The first six values of u_n, modulo 127, are

$$\begin{array}{rcrl} u_1 & \equiv & 3 & \pmod{127} \\ u_2 & \equiv & 7 & \pmod{127} \\ u_3 & \equiv & 47 & \pmod{127} \\ u_4 & \equiv & 48 & \pmod{127} \\ u_5 & \equiv & 16 & \pmod{127} \\ u_6 & \equiv & 0 & \pmod{127} \end{array}$$

confirming that 127 is a Mersenne prime.

Exercises

1) Let ℓ be an odd prime such that $M_\ell = 2^\ell - 1$ is a Mersenne prime. Use the quadratic reciprocity to show that 5 is a square modulo M_ℓ if and only if $\ell \equiv 1 \pmod 4$.

2) Let $\ell \equiv 3 \pmod 4$ be an odd prime. Show that there exists a prime divisor p of M_ℓ such that $\left(\frac{5}{p}\right) = -1$.

3) Let $\ell \equiv 3 \pmod 4$ be an odd prime. Assume that $M_\ell = 2^\ell - 1$ is a Mersenne prime. Then $\alpha = \frac{3}{2} + \frac{1}{2}\sqrt{5}$ is not a square in $T(M_\ell)$.

4) Let $\ell \equiv 3 \pmod 4$ be an odd prime. Let $\alpha = \frac{3}{2} + \frac{1}{2}\sqrt{5}$. Show that $M_\ell = 2^\ell - 1$ is a Mersenne prime if and only if

$$\alpha^{2^{\ell-1}} \equiv -1 \pmod{M_\ell}.$$

5) Let $\ell \equiv 3 \pmod 4$ be an odd prime. Define a recursive sequence u_n by $u_1 = 3$ and $u_{n+1} = u_n^2 - 2$. Show that $M_\ell = 2^\ell - 1$ is a Mersenne prime if and only if

$$u_{\ell-1} \equiv 0 \pmod{M_\ell}.$$

6) Calculate u_{18} modulo $M_{19} = 524287$ to show that M_{19} is a Mersenne prime.

7) Let $M_\ell = 2^\ell - 1$ be a Mersenne number. Show that $M_\ell \equiv 1 \pmod 7$ if $\ell \equiv 1 \pmod 3$ and $M_\ell \equiv 3 \pmod 7$ if $\ell \equiv 2 \pmod 3$.

8) Let $M_\ell = 2^\ell - 1$ be a Mersenne number where $\ell \equiv 1 \pmod 3$. Show that there exists a prime divisor p of M_ℓ such that $\left(\frac{7}{p}\right) = -1$.

CHAPTER 8

Sums of two squares

1. Sums of two squares

In the next several section we shall describe all positive integers that can be written as a sum of two squares. As the first step, we shall solve this problem for prime numbers. Below we see a list of the first six odd primes indicating for each prime whether or not (Y or N) it can be written as a sum of two squares.

Prime	Y/N	Sum
3	N	-
5	Y	2^2+1^2
7	N	-
11	N	-
13	Y	3^2+2^2
17	Y	4^2+1^2

By this table we see that 3, 7, 11 cannot be written as a sum of two squares, while 5, 14, 17 can be written as a sum of two squares. In fact, the following is true.

$$p \equiv 1 \pmod 4 \Leftrightarrow \ p \text{ is a sum of two squares.}$$

The proof of this statement is based on the fact that -1 is a square modulo p if and only $p \equiv 1 \pmod 4$. Indeed, if $a^2+b^2=p$ then

$$a^2+b^2 \equiv 0 \pmod p.$$

Now note that a and b are less than p. It follows that b is relatively prime to p. In particular, there exists an inverse b^{-1} of b modulo p. Multiplying the congruence by b^{-2} gives

$$\left(\frac{a}{b}\right)^2+1 \equiv 0 \pmod p$$

or $(a/b)^2 \equiv -1 \pmod p$, which means that -1 is a square modulo p. Thus we have shown that, if p can can be written as a sum of two squares, then

$$p \equiv 1 \pmod 4.$$

But how do we prove that every prime $p \equiv 1 \pmod 4$ can be written as a sum of two squares? To begin, we know that the equation

$$x^2 \equiv -1 \pmod p$$

has a solution. This, in turn, gives a solution to

$$x^2 + y^2 \equiv 0 \pmod p.$$

(Just pick $y = 1$ and x such that $x^2 \equiv -1 \pmod p$.) For example, if $p = 13$ then

$$5^2 = 25 \equiv -1 \pmod{13}$$

hence

$$5^2 + 1^2 \equiv 0 \pmod{13}.$$

Without congruences, we have

$$5^2 + 1^1 = 2 \cdot 13.$$

In order to represent 13 as a sum of two squares we need to remove the factor 2 from $2 \cdot 13$. This will be accomplished by *Fermat's Method of Infinite Descent*. The aim here is to replace the equation

$$x^2 + y^2 = mp$$

by

$$x_1^2 + y_1^2 = rp$$

such that the integer r is smaller than the integer m. In fact, we will see that the method gives $r \leq m/2$ so, starting with $x^2+y^2 = mp$, we can find a solution to $x^2+y^2 = p$ in less than $\log_2 m$ steps. To explain how the *Method of Descent* works, we need first the following formula:

$$(x^2 + y^2)(u^2 + v^2) = (xu + yv)^2 + (xv - yu)^2.$$

This formula says that a product of two sums of two squares is again a sum of two squares. The best way to explain this formula is to use complex numbers. Recall, if $z = x + yi$ is a complex number (here $i^2 = -1$) then its conjugate is $\bar{z} = x - yi$. Then

$$z \cdot \bar{z} = (x + yi) \cdot (x - yi) = x^2 + y^2.$$

Thus, if we write $z = x + iy$ and $w = u - iv$, then

$$(x^2 + y^2)(u^2 + v^2) = (z \cdot \bar{z})(w \cdot \bar{w}) = (zw) \cdot (\bar{z}\bar{w}).$$

Since $zw = (xu + yv) + i(yu - xv)$ and $\bar{z} \cdot \bar{w} = \overline{zw}$ (the conjugate of the product is equal to the product of conjugates) we have

$$(x^2 + y^2)(u^2 + v^2) = (zw) \cdot (\overline{zw}) = (xu + yv)^2 + (xv - yu)^2$$

as claimed. Consider, for example, $2^2 + 1^2$ and $3^2 + 2^2$. Since

$$(2 + i)(3 + 2i) = 4 + 7i$$

then

$$(2^2 + 1^2)(3^2 + 2^2) = 4^2 + 7^2.$$

$\boxed{\text{METHOD OF DESCENT}}$

We start with an equation $x^2 + y^2 = mp$. Pick $u \equiv x \pmod{m}$ and $v \equiv y \pmod{m}$ such that

$$-\frac{m}{2} < u, v \leq \frac{m}{2}.$$

The choice of u and v implies that $u^2 + v^2$ is congruent to $x^2 + y^2$ modulo m, so we definitely can write

$$u^2 + v^2 = mr.$$

Using the formula for the product of two sums of two squares, we have

$$(xu + yv)^2 + (xv - yu)^2 = m^2 rp.$$

Since $v \equiv y \pmod{m}$ and $u \equiv x \pmod{m}$, it follows that

$$xv - yu \equiv xy - yx \equiv 0 \pmod{m}$$

and

$$xu + yv \equiv x^2 + y^2 \equiv 0 \pmod{m}.$$

Thus we can write

$$\left(\frac{xu + yv}{m}\right)^2 + \left(\frac{xv - yu}{m}\right)^2 = rp$$

where the two expressions in brackets are integers. Thus, our task is accomplished provided we can prove that r is smaller than m. But that is not difficult. Indeed,

$$mr = u^2 + v^2 \leq \left(\frac{m}{2}\right)^2 + \left(\frac{m}{2}\right)^2 = \frac{m^2}{2}$$

hence

$$r \leq \frac{m}{2}.$$

Summarizing, the whole algorithm can be simply written as follows:

(1) Pick $u \equiv x \pmod{m}$ and $v \equiv y \pmod{m}$ such that

$$-\frac{m}{2} < u, v \leq \frac{m}{2}.$$

(2) Put

$$x_1 = \frac{xu + yv}{m} \text{ and } y_1 = \frac{xv - yu}{m}$$

(3) Write

$$x_1^2 + y_1^2 = m_1 p.$$

If $m_1 = 1$ stop. Else go to (1).

Let us see how this algorithm gives a representation of 13 as a sum of two squares, starting with $5^2 + 1^2 = 2 \cdot 13$.

$5^2 + 1^2 = 2 \cdot 13$	$x^2 + y^2 = mp$
$1 \equiv 5 \pmod{2}$	$u \equiv x \pmod{m}$
$1 \equiv 1 \pmod{2}$	$v \equiv y \pmod{m}$
$1^2 + 1^2 = 2 \cdot 1$	$u^2 + v^2 = mr$
$3^2 + 2^2 = 1 \cdot 13$	$x_1^2 + y_1^2 = rp$

If we know a primitive root modulo p, the whole algorithm can be easily implemented on a computer. For example, 2 is a primitive root modulo 13. Thus

$$2^{\frac{13-1}{4}} \equiv 8 \pmod{13}$$

is a root of -1 modulo 13 and

$$8^2 + 1^2 = 65 = 5 \cdot 13$$

gives a starting input for the descent algorithm which will terminate - in this example - in at most two steps.

Exercises

1) Use $13 = 3^2 + 2^2$ and $17 = 4^2 + 1^2$ to write 221 as a sum of two squares.

2) Starting with the equation

$$8^2 + 1^2 = 5 \cdot 13$$

use the method of descent to represent 13 as a sum of two squares.

3) Represent 73 as a sum of two squares using the following two steps.

a) First, use the primitive root 5 to find x such that

$$x^2 + 1^2 = m \cdot 73.$$

b) Then use the method of descent to find a solution.

In the following three exercises we shall adopt the method of descent to find out which primes p can be written as a sum $x^2 + 2y^2 = p$.

4) Use the quadratic reciprocity to find a necessary condition such that a prime p can be written as $x^2 + 2y^2 = p$.

5) Prove the formula

$$(x^2 + 2y^2)(u^2 + 2v^2) = (xu + 2yv)^2 + 2(yu - xv)^2.$$

6) Note that -2 is a square modulo 11. Indeed, $8^2 + 2 = 6 \cdot 11$. Use this information, and the method of descent to construct a solution of $x^2 + 2y^2 = 11$.

2. Gaussian integers

Complex numbers $a+bi$ such that a and b are integers are called Gaussian integers. The ring $\mathbb{Z}[i]$ of Gaussian integers is in many ways similar to the ring of integers $\mathbb{Z}$. For example, there is an Euclidean-type algorithm for $\mathbb{Z}[i]$ which can be used to prove a uniqueness of factorization into primes. The key tool here is the norm $N(\alpha)$

$$N(\alpha) = \alpha \cdot \bar{\alpha} = a^2 + b^2.$$

Notice that $N(\alpha)$ is always a positive integer. It is 0 if and only if α is 0. The most important property of the norm is that it is multiplicative:

$$N(\alpha\beta) = (\alpha\beta) \cdot (\overline{\alpha\beta}) = (\alpha\bar{\alpha}) \cdot (\beta\bar{\beta}) = N(\alpha)N(\beta).$$

We start with the following simple proposition which describes the group $\mathbb{Z}[i]^\times$ of invertible Gaussian integers.

Proposition 37. *Let α be a Gaussian integer. The following three statements are equivalent:*

1. *α is invertible, that is, there exists a Gaussian integer β such that $\alpha \cdot \beta = 1$.*
2. *$N(\alpha) = 1$.*
3. *α is 1, -1, i or $-i$.*

Proof. If (1) holds then, after taking norms,

$$N(\alpha)N(\beta) = N(\alpha \cdot \beta) = N(1) = 1.$$

Since $N(\alpha)$ and $N(\beta)$ are positive integers, they have to be 1. Thus, we have shown that (1) implies (2). If $N(\alpha) = a^2 + b^2 = 1$ then α must be one of the given four Gaussian integers. In particular (2) implies (3). Finally, each of the four: 1, -1 , i and $-i$ is invertible, so (3) implies (1). □

We say that a Gaussian integer β divides α is there exists a Gaussian integer γ such that

$$\alpha = \gamma \cdot \beta.$$

For example, $2+i$ divides 5 since $5 = (2-i)(2+i)$. Notice that, if β divides α, then the norm $N(\beta)$ divides $N(\alpha)$. For example, the norm of $2+i$ is 5 while the norm of $3+2i$ is 13. Thus $2+i$ cannot possibly divide $3+2i$.

The Euclidean algorithm works as follows. If α and β are two integers (we shall omit "Gaussian" from now on) then there are two integers γ and ρ (that will be precisely defined in a moment) such that

$$\alpha = \gamma \cdot \beta + \rho \text{ and } N(\rho) < N(\beta).$$

The integer γ is defined to be an approximation of the complex fraction

$$\frac{\alpha}{\beta} \approx \gamma$$

obtained by rounding each coordinate to the nearest integer, in the usual sense. In particular, we shall round

$$0.5 \approx 1 \text{ and } -0.5 \approx 0.$$

Once we have γ then we put $\rho = \alpha - \gamma \cdot \beta$. As an example, let us calculate with $\alpha = 3 + 2i$ and $\beta = 3 - 2i$. Then

$$\frac{\alpha}{\beta} = \frac{3+2i}{3-2i} = \frac{5}{13} + \frac{12}{13} i \approx i = \gamma.$$

which implies that $\rho = (3+2i) - (3-2i)i = 1 - i$. Note that $N(1-i) = 2 < N(3+2i) = 13$.

It remains to check that $N(\rho) < N(\beta)$ holds in general. From the definition of γ, notice that

$$\frac{\alpha}{\beta} - \gamma = x + yi$$

with $|x|, |y| \leq 1/2$. Thus, it follows that

$$\frac{N(\rho)}{N(\beta)} = N\left(\frac{\alpha}{\beta} - \gamma\right) \leq \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{2}.$$

We have shown even more than it was needed: $N(\rho) \leq N(\beta)/2$. This property is important because it guarantees that the repeated application of the algorithm

$$\begin{array}{rcl} \alpha & = & \gamma_1 \cdot \beta + \rho_1 \\ \beta & = & \gamma_2 \cdot \rho_1 + \rho_2 \\ & \vdots & \end{array}$$

stops in a finite number of steps. In fact, since the reminder in each consecutive division is less than the half of the reminder in the previous step, it will not take more than $\log_2 N(\beta)$ steps in all.

If the last non-zero reminder is ρ_n then, arguing as in the case of usual integers, any common divisor δ of α and β divides ρ_n. Indeed, the first equation

$$\alpha = \gamma_1 \cdot \beta + \rho_1$$

shows that δ divides ρ_1, the second equation

$$\beta = \gamma_2 \cdot \rho_1 + \rho_2$$

shows that δ divides ρ_2 (since it divides β and ρ_1). Working in this manner from the top to the bottom, we see that δ divides ρ_n, the last non-zero reminder. (In particular, $N(\delta) \leq N(\rho_n)$.) In the other direction, the last equation

$$\rho_{n-1} = \gamma_{n+1}\rho_n + 0$$

implies that ρ_n divides ρ_{n-1}. Then the equation

$$\rho_{n-2} = \gamma_n \rho_{n-1} + \rho_n$$

implies that ρ_n divides ρ_{n-2}. Continuing in this fashion, we see that ρ_n is a common divisor of α and β. This shows that the reminder ρ_n can be considered the *greatest common divisor of α and β*. (Its norm is the greatest amongst all common divisors of α and β.)

Consider, for example, $\alpha = 3 + 2i$ and $\beta = 3 - 2i$. Then, as we have already computed, $\rho_1 = 1 - i$. The next step gives

$$\frac{\beta}{\rho_1} = \frac{3 - 2i}{1 - i} = \frac{5}{2} + \frac{1}{2}i \approx 3 + i = \gamma_2,$$

which implies that $\rho_2 = (3 - 2i) - (1 - i)(3 + i) = -1$, and then, finally, $\rho_3 = 0$. Thus -1 is the greatest common divisor of $3 + 2i$ and $3 - 2i$. Equipped with these tools, one can now show that Gaussian integers admit unique factorization into prime numbers. The proof is very similar as the one for usual integers. Instead of repeating similar arguments, we shall focus here on some differences.

First of all, instead of talking about prime numbers, it is more correct to talk about *indecomposable* numbers. A number α is indecomposable if

$$\alpha = \beta\gamma$$

then either β or γ is invertible. This concept works well in the case of usual integers. For example, the only way to factor 3 in $\mathbb{Z}$ is

$$3 = (-1) \cdot (-3).$$

Since -1 is invertible, 3 is indecomposable. Moreover, the factorization of (usual) integers is not quite unique, if we allow fudging with signs, since 6 can be factored in two ways:

$$6 = 2 \cdot 3 = (-2) \cdot (-3).$$

Of course, the two factorizations are not that much different, since 2 and -2 could be considered the same "prime number". This example should explain why, technically, a prime is not a number. Rather, a prime is a class of all indecomposable integers, such as

$$\{-2, 2\}$$

where the quotient of any two integers in the class is an invertible integer. (Note that the quotient of -2 and 2 is -1.) In the case of integers $\mathbb{Z}$ we have avoided this sort of intricacies by simply restricting the discussion to positive integers!

With this distinctions in mind, we are now ready to figure out primes in the ring of Gaussian integers.

PROPOSITION 38. *Let p be a prime such that $p \equiv 1 \pmod 4$. Write p as a sum of two squares, $p = a^2 + b^2$. Then $\alpha = a + bi$ is an indecomposable Gaussian integer.*

PROOF. Note that the norm of α is p. Now, if $\alpha = \beta\gamma$ then, after taking norms,

$$N(\alpha) = p = N(\beta)N(\gamma).$$

Since p is prime, it follows that $N(\beta) = 1$ or $N(\gamma) = 1$. In any case, one of the two, β or γ must be invertible Gaussian integer. The proposition follows. □

Consider, for example, $p = 5$. Then $5 = (2+i)(2-i)$ and both $2+i$ and $2-i$ are indecomposable. Furthermore, they represent different primes, since their respective classes include

$$2+i,\ -2-i,\ i\cdot(2+i) = -1+2i, \text{ and } -i\cdot(2+i) = 1-2i$$

and

$$2-i,\ -2+i,\ i\cdot(2-i) = 1+2i, \text{ and } -i\cdot(2-i) = -1-2i.$$

In words, $2+i$ and $2-i$ are two indecomposable gaussian integers representing different primes, since one is not obtained from the other by multiplying by 1, -1, i or $-i$. This happens for all $p \equiv 1 \pmod 4$. We can always fix a unique representative in the class of the form $a+bi$ or $a-bi$ with a and b positive and $a > b$.

Situation is somewhat different for $p = 2$. Since 2 is also a sum of two squares, we have

$$2 = 1^2 + 1^2 = (1+i)(1-i).$$

Here however, $1+i = i(1-i)$, so $1+i$ and $1-i$ represent the same prime.

Finally, it remains to discuss primes $p \equiv 3 \pmod 4$. These turn out to be indecomposable. Indeed, assume that we have a factorization

$$p = \alpha\cdot\beta$$

where α and β are two Gaussian integers. After taking norms,

$$N(p) = p^2 = N(\alpha)N(\beta).$$

If $N(\alpha) = p$, then $p = a^2 + b^2$ where $\alpha = a+bi$. This implies that p can be represented as a sum of two squares, a contradiction. Thus the only possibilities are that $N(\alpha) = 1$ and $N(\beta) = p^2$ or $N(\alpha) = 1$ and $N(\beta) = p^2$, i.e. either α or β is a unit. Hence p is indecomposable.

We are now ready to prove the following:

Theorem 39. *Every class of indecomposable Gaussian integers contains precisely one of the following:*

1. $1+i$.
2. *A usual prime p congruent to 3 modulo 4.*
3. $a+bi$ *or* $a-bi$ *where* $a > b$ *are two positive integers such that* a^2+b^2 *is a prime congruent to* 1 *modulo* 4.

In particular, every Gaussian integer α *can be written as a product*

$$\alpha = i^m\pi_1\cdot\ldots\cdot\pi_n$$

where m *is an integer modulo 4, and* π_k *are indecomposable integers from (1) - (3) above. This factorization is unique up to permutation of factors.*

Proof. Let π be an indecomposable gaussian integer. Consider the positive integer $N(\pi) = \pi\bar{\pi}$. If $N(\pi)$ is divisible by 2 then we have

$$\pi\bar{\pi} = N(\pi) = \ldots (1+i)(1-i)$$

from which we can conclude that $1+i$ or $1-i$ divides π. Thus π is in the class of $1+i$. Similarly, if $N(\pi)$ is divisible by a prime p congruent to 1 modulo 4 then

$$\pi\bar{\pi} = N(\pi) = \ldots (a+bi)(a-bi)$$

where $a^2+b^2=p$. Thus $a+bi$ or $a-bi$ divides π and π is in the class of $a+bi$ or $a-bi$. Finally, if $N(\pi)$ is divisible by a prime p congruent to 3 modulo 4 then π is the class of p. This proves the theorem, as the uniqueness of the representatives was already discussed. □

We can now give a complete characterization of numbers that can be written as a sum of two squares.

Corollary 40. *Let n be a positive integer. Factor n into primes: $n = p_1^{n_1} \cdots p_r^{n_r}$. Then n can be written as a sum of two squares if an only if n_i is even for every $p_i \equiv 3 \pmod 4$.*

Proof. If n satisfies the assumptions, then we can write $n = m^2 p_1 \cdots p_s$ where every p_i is either equal to 2 or congruent to 1 modulo 4. In particular, each can be written as a sum of two squares, $p_i = N(\pi_i)$. It follows that n is a sum of two squares

$$n = N(m\pi_1 \cdots \pi_s).$$

Conversely, assume that $n = a^2+b^2$. Factor $a+bi$ into a product of Gauss's primes:

$$a+bi = \pi_1 \cdots \pi_t.$$

Then $a - ib = \bar{\pi}_i \cdots \bar{\pi}_t$ and $n = (a+ib)(a-ib)$ is a product of $\pi_i\bar{\pi}_i$. Since $\pi_i\bar{\pi}_i = p$ for some $p \equiv 1 \pmod 4$ or 2, or $\pi_i\bar{\pi}_i = p^2$ for some $p \equiv 3 \pmod 4$, the exponents of prime factors of n are even for all $p \equiv 3 \pmod 4$, as desired. □

Exercises

1) Find the greatest common divisor of $11+7i$ and $5+3i$.

2) Factor

$$11+3i$$

into indecomposable gaussian integers. Hint: pretty much the only way to factor is to guess divisors. How do we guess here? Compute the norm!

3. Method of descent revisited

In the last section we have shown that a uniqueness of factorization holds for Gaussian integers and we have described all gaussian primes. Armed with this information, we can now gain a more conceptual understanding of Fermat's method of descent.

Let p be a prime congruent to 1 modulo 4. If we have a sum of two squares that is equal to a multiple of of p then the method of descent produces a sum of two squares equal to p. For example, if $p = 13$, then $5^2 + 1^2 = 2 \cdot 13$ and the method of descent produces $3^2 + 2^2 = 13$. The reader might have noticed that the method of descent is similar to the Euclidean algorithm for Gaussian integers. This is not a coincidence. Indeed, using Gaussian integers we can write

$$(5 + i)(5 - i) = 2 \cdot 13.$$

Writing 13 as a sum of two squares is equivalent to factoring $13 = (3 + 2i)(3 - 2i)$. Substituting this in the identity above,

$$(5 + i)(5 - i) = 2 \cdot (3 + 2i)(3 - 2i).$$

Since $3 + 2i$ is prime, uniqueness of factorization implies that $3 + 2i$ divides $5 + i$ or $5 - i$. Note that if $3 + 2i$ divides $5 + i$ then, by taking complex conjugates, $3 - 2i$ divides $5 - i$. In any case, 13 and $5 - i$ have a common divisor, either $3+2i$ or $3-2i$. Now note that $3-2i$ and $3+2i$ cannot, both, be divisors of $5-i$. Otherwise, since $3-2i$ and $3+2i$ are two different primes, their product $(3 - 2i)(3 + 2i) = 13$ would also divide $5 - i$, a contradiction. This shows that either $3 + 2i$ or $3 - 2i$ is a greatest common divisor of 13 and $5 - i$. The greatest common divisor of two Gaussian integers is, of course, the product of all common primes, with highest possible exponents. As for the ordinary integers, the greatest common divisor is computed by the Euclidean algorithm. Dividing 13 by $5 - i$ gives

$$\frac{13}{5-i} = \frac{5}{2} + \frac{1}{2}i = 2 + (\frac{1}{2} + \frac{1}{2}i)$$

and, after multiplying both sides by $5 - i$,

$$13 = (5 - i) \cdot 2 + (3 + 2i).$$

This shows that diving 13 with $5-i$ gives $3+2i$ as a reminder. As one easily checks that $5 - i$ is divisible by $3 + 2i$, it follows that $3 + 2i$ is the greatest common divisor of 13 and $5 - i$. More importantly, we have found a way to represent 13 as a sum of two squares:

$$3^2 + 2^2 = 13.$$

This method works in general. Indeed, assume that a multiple of p can be written as a sum of two squares

$$x^2 + y^2 = mp$$

such that x and y are not divisible by p. Recall that this can be always arranged, for example, with $y = 1$ since -1 is a square modulo p for p congruent to 1 modulo 4. Writing p as a sum of two squares is equivalent to factoring

$$p = \pi\bar{\pi}$$

as a product of two Gauss's primes. Then, by uniqueness of factorization, π divides $x + yi$ or $x - yi$. Assume, without any loss of generality, that π divides $x+yi$. Then $\bar{\pi}$ cannot divide $x+yi$. Otherwise $p = \pi\bar{\pi}$ would divide $x + yi$, a contradiction since p does not divide x and y. This shows that

$$\pi = \gcd(p, x + yi).$$

In particular, π can be obtained from p and $x + yi$ using the Euclidean algorithm. Dividing $x - yi$ by p gives

$$\frac{p}{x-yi} = \frac{p(x+yi)}{x^2+y^2} = \frac{x}{m} + \frac{y}{m}i.$$

Pick integers $u \equiv x \pmod{m}$ and $v \equiv y \pmod{m}$ such that

$$-\frac{m}{2} < u, v \leq \frac{m}{2}.$$

Then

$$\frac{p}{x-yi} = (\frac{x-u}{m} + \frac{y-v}{m}i) + (\frac{u}{m} + \frac{v}{m}i)$$

and, after multiplying both sides by $(x - yi)$,

$$p = (\frac{x-u}{m} + \frac{y-v}{m}i)(x-yi) + (\frac{xu+yv}{m} + \frac{xv-yu}{m}i).$$

It follows that dividing p by $x - yi$ gives a reminder

$$x_1 + y_1 i = \frac{xu+yv}{m} + \frac{xv-yu}{m}i.$$

Not that x_1 and y_1 are exactly the same as those produced by Fermat's descent. This shows that Fermat's method of descent can be recovered from the Euclidean algorithm for Gauss's integers.

Exercises

1) Compute the greatest common divisor of $9+7i$ and 13 to find $x+yi$ such that $x^2 + y^2 = 13$.

2) Compute the greatest common divisor of $9+8i$ and 17 to find $x+yi$ such that $x^2 + y^2 = 17$.

3) Compute the greatest common divisor of $7+3i$ and 29 to find $x+yi$ such that $x^2 + y^2 = 29$.

CHAPTER 9

Pell's Equations

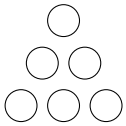
1. Shape numbers and Induction

Shape numbers count objects arranged in a special shape, such as a triangle or a square. The problem of finding and proving a formula for shape numbers will introduce us to one of the main topics of this section: mathematical induction. In the next section we shall compare certain shape numbers. This will lead us to the Pell equation, the main subject of this chapter.

The simplest example of shape numbers are triangular numbers T_n. The number T_n is a sum of the first n integers:

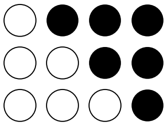
$$T_n = 1 + 2 + \ldots + n.$$

The number T_n is called a triangular number because it counts objects arranged in a triangular shape. For example $T_3 = 1 + 2 + 3$ counts the number of circles arranged in the following triangle:



T_3

In order to develop a closed formula for T_n (a formula that involves a definite number of operations) we apply the following trick. Combine two triangles with T_n circles to create an $(n + 1) \times n$-rectangle. For example, if $n = 3$ then the picture is:



This shows that $2T_n = (n+1)n$

$$T_n = \frac{(n+1)n}{2}.$$

Obviously, the trick we used here is very particular to this example. On the other hand, mathematical induction is a very general method that can be used to verify a statement which depends on a positive integer. The basic principle of induction is:

Every nonempty set of positive integers has a smallest element.

Suppose we have a statement that involves a positive integer n. We will denote it $S(n)$. To prove $S(n)$ for every $n = 1, 2, 3, \cdots$ it suffices to prove the following:

(1) $S(1)$ is true (this is called the *basis of induction*), and
(2) the truth of $S(n-1)$ (induction assumption) implies the truth of $S(n)$ (this is called the *inductive step*).

A way to think about this is the following. Consider the set of all n such that $S(n)$ fails. We want to show that this set is empty. Well, if not, then there is a minimal element of this set, say m. Now $m > 1$ by (1). But then the fact that $S(m-1)$ is true while $S(m)$ is false contradicts (2).

For example, let $S(n)$ be the statement that a closed formula for the n-th triangular number (or the sum of the first n integers) is given by

$$1 + 2 + \ldots + n = \frac{(n+1)n}{2}.$$

In order to verify this statement using the induction we need first to check it for $n = 1$ (basis of induction). Since (we substitute $n = 1$ in both sides of the above identity)

$$1 = \frac{(1+1)1}{2},$$

it follows that $S(1)$ is true.

The step of induction: We need to verify the statement $S(n)$ assuming that $S(n-1)$ is true. The statement $S(n-1)$ says that the sum of first

$n-1$ integers is equal to

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

Add n to both sides, to get

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

Since

$$\frac{(n-1)n}{2}+n=\frac{n}{2}[(n-1)+2]=\frac{n(n+1)}{2}$$

we see that $S(n)$ is true. This completes the induction in this example.

One can argue that we really did not need the induction to figure out the formula for T_n. However, the trick used to figure out T_n might not be available in some other situations. So here is an approach to a problem that makes induction indispensable. For example, assume we want to find a formula for the sum of the first n odd numbers. We can easily work out the sum of the first n odd integers for $n=1,2,3,4$ and 5. We see that the sums are

$$1,\ 4,\ 9,\ 16,\ \text{and } 25,$$

respectively. This leads us to conjecture that the sum of the first n odd integers is equal to n^2. This statement can be proved by mathematical induction. (See an exercise bellow.)

There are other versions of mathematical induction. For example, (2) can be replaced by:

(2') the truth of $S(1),S(2),\cdots,S(n-1)$ implies the truth of $S(n)$.

This sort of induction was used when we proved that cyclotomic polynomials $\Phi_n(x)$ have integral coefficients. Indeed, the base of induction is true since $\Phi_1(x)=x-1$. The step of induction is based on the factorization

$$x^n-1=\prod_{d|n}\Phi_d(x).$$

By induction assumption, $\Phi_d(x)$ for $d<n$ have integral coefficients. It follows that $\Phi_n(x)$ has integral coefficients. This conclusion is based on the fact that a quotient of two monic polynomials with integral coefficients is again a monic polynomial with integral coefficients.

Exercises

1) Use mathematical induction to prove that for every positive integer n

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

2) Use mathematical induction to prove that the sum of the first n odd integers is equal to n^2.

3) For every positive integer n, show that

$$\frac{(1+\sqrt{5})^n-(1-\sqrt{5})^n}{2^n\sqrt{5}}$$

is a positive integer. In fact, u_n is the n-th Fibonacci number.

2. Square-Triangular Numbers and Pell Equation

In the previous section we introduced so-called triangular numbers. The name comes from the fact that these numbers count objects arranged in a certain shape - a triangle in this case. Recall that the m-the triangular number is

$$T_m = 1 + 3 + ... + m = \frac{m(m+1)}{2}.$$

Similarly, we have square numbers $S_n = n^2$. In the following table we list first 8 triangular and square numbers

k	1	2	3	4	5	6	7	8
T_k	1	3	6	10	15	21	28	36
S_k	1	4	9	16	25	36	49	64

Notice that the numbers 1 and 36 appear on both lists. In other words, these two numbers are both, square and triangular numbers. Therefore, 1 and 36 are called *square-triangular numbers*. In the next couple of sections we shall completely solve the problem of finding all square - triangular numbers. Of course, the issue here is to solve the equation

$$T_m = S_n.$$

The smallest solution, as seen from the list, is $m = 1, n = 1$. An amazing fact is that all other solutions can be generated from this one. In order to explain, we need to rewrite the equation. First, we multiply each side of $T_m = S_n$ by 8 to get

$$4m(m+1) = 8n^2.$$

The left hand side can be manipulated by completing a square,

$$4m^2 + 4m = 4m^2 + 4m + 1 - 1 = (2m+1)^2 - 1.$$

The right hand side is equal to $2(2n)^2$. Thus, if we substitute

$$x = 2m + 1 \text{ and } y = 2n,$$

then the equation $T_m = S_n$ becomes

$$x^2 - 2y^2 = 1$$

and the first solution $(m, n) = (1, 1)$ of $T_m = S_n$ translates into a solution $(x, y) = (3, 2)$ of $x^2 - 2y^2 = 1$. The equation $x^2 - 2y^2 = 1$ is a special case of the general Pell equation

$$x^2 - Dy^2 = 1.$$

We have translated the problem of finding square-triangular number into the problem of finding integer solutions of a Pell equation. A remarkable fact about the Pell equation is that, starting with one solution, one can generate more solutions of the Pell equation. Here is how this is done. First, note that the left hand side of the equation can be factored as

$$x^2 - 2y^2 = (x + y\sqrt{2})(x - y\sqrt{2}).$$

Thus, the Pell equation can be rewritten as

$$(x + y\sqrt{2})(x - y\sqrt{2}) = 1.$$

Next, let us take the square of both sides to see what we get. Since

$$(x \pm y\sqrt{2})^2 = (x^2 \pm 2xy\sqrt{2} + 2y^2)$$

squaring of both sides of the Pell equation gives

$$(x^2 + 2y^2 + 2xy\sqrt{2})(x^2 + 2y^2 - 2xy\sqrt{2}) = 1$$

or, equivalently,

$$\left(x^2 + 2y^2\right)^2 - 2\,(2xy)^2 = 1.$$

Hence we have the following important observation. If (x, y) is a solution of $x^2 - 2y^2 = 1$ then so is (S, T) where

$$S + T\sqrt{2} = (x + y\sqrt{2})^2.$$

For example, since $(3, 2)$ is a solution of $x^2 - 2y^2 = 1$, squaring of $3 + 2\sqrt{2}$

$$(3 + 2\sqrt{2})^2 = 17 + 12\sqrt{2}$$

generates another solution $(x, y) = (17, 12)$ of $x^2 - 2y^2 = 1$. This solution corresponds to a solution $(m, n) = (8, 6)$ of the original equation $T_m = S_n$. More generally, if we have two solutions of the Pell equation, we can simply multiply them to get a third solution. For example,

$$(17 + 12\sqrt{2})(3 + 2\sqrt{2}) = 99 + 70\sqrt{2}$$

and we have generated yet another solution $(x, y) = (99, 70)$ of $x^2 - 2y^2 = 1$. A list of the three solutions and corresponding square-triangular numbers, that we have just found, is given in the following table:

x	y	m	n	$S_n = T_m$
3	2	1	1	1
17	12	8	6	36
99	70	49	35	1225

Thus 1225 is the third square-triangular number, so far discovered. But is there is square-triangular number between 36 and 1225? The answer is no. This follows from the following theorem, to be proved in the next several sections.

THEOREM 41. *Any integer solution of the equation* $x^2 - 2y^2 = 1$*, with positive integers* (x, y) *is obtained from the first solution* $(3, 2)$ *by taking a* k^{th} *power*

$$(3 + 2\sqrt{2})^k = x + y\sqrt{2}$$

for some positive integer k*.*

In fact, the same result holds for any equation of the type $x^2 - Dy^2 = 1$ if D is a non-square positive integer. If (x_1, y_1) is the solution with the smallest positive x and y coordinates, then every solution (x_k, y_k), with positive coordinates, is of the form:

$$x_k + y_k\sqrt{D} = (x_1 + y_1\sqrt{D})^k.$$

Note however that, for a general D, we have not yet established existence of a non-trivial solution. In fact, even for some small value of D the first solution can be rather large. For example, the first solution of $x^2 - 94y^2 = 1$ is

$$2143295 + 211064\sqrt{94}.$$

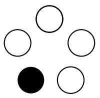
Our task, which will be accomplished in the next sections, consists of the following three steps.

(1) Show that $x^2 - Dy^2 = 1$ has a non-trivial solution (x, y) with x and y positive integers.
(2) Show that any positive integer solution is a power of the first solution.
(3) Find an effective way to construct the first solution.

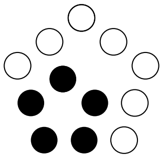
Exercises

1) By inspection, find the first positive solution of the Pell equation $x^2 - 6y^2 = 1$. Use the first solution to find the first three solutions of the equation.

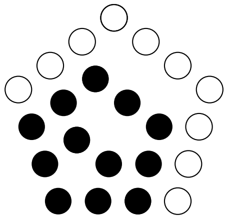
2) A number P_n is called pentagonal, if P_n pebbles can be arranged in the shape as pictured, with n pebbles along each edge. Write down a simple formula for P_n, n-th pentagonal number, following the two steps: First find two integers a and b such that $P_{n+1} - P_n = an + b$, and then use the formula $1 + 2 + \ldots + (n - 1) = n(n - 1)/2$.



$P_2 = 5$



$P_3 = 12$



$P_4 = 22$

3) In this problem we will classify square-pentagonal numbers. The equation $m^2 = P_n$, after substituting $n = (x+1)/6$, and $m = y/2$ becomes the Pell equation $x^2 - 6y^2 = 1$, which has $(5, 2)$ as a basic solution. Of course, the powers $(5 + 2\sqrt{6})^k$ will give all solutions to the Pell equation, but not all of these will give rise to square-pentagonal numbers. Calculate several low degree powers to see what is going on, and determine the first non-trivial square-pentagonal number.

4) Classify triangular-pentagonal numbers.

3. Dirichlet's approximation

Solving the Pell equation $x^2 - Dy^2 = 1$ is closely related to approximating $\sqrt{D}$ by fractions. Indeed, the equation $(x - y\sqrt{D})(x + y\sqrt{D}) = 1$ can be rewritten as

$$x - y\sqrt{D} = \frac{1}{x + y\sqrt{D}},$$

and, since $y \le x + y\sqrt{D}$, we get an inequality

$$|x - y\sqrt{D}| < \frac{1}{y} \Leftrightarrow \left|\frac{x}{y} - \sqrt{D}\right| < \frac{1}{y^2}.$$

For example, $(x, y) = (99, 70)$ is the third solution of the Pell equation $x^2 - 2y^2 = 1$ and

$$\sqrt{2} = 1.414213\ldots \approx \frac{99}{70} = 1.414228.$$

Thus, in order to show that the equation $x^2 - Dy^2 = 1$ has a non-trivial solution, we shall first develop a method of approximating irrational numbers by fractions and, along the way, show that the Pell equation always has a non-trivial solution. This method is due to Dirichlet and is based on the pigeon-hole or Dirichlet's principle. However, due to bird flu concerns, we shall leave pigeons out, and state the principle in terms of boxes. The principle says that if you put $n+1$ objects in n boxes then at least one box contains two objects.

We shall apply this principle in the following way. Suppose that we want to approximate, by fractions, a positive irrational number α. Let n be a positive integer and divide the interval $[0,1)$ into n subintervals of the length $1/n$:

$$\left[0,\frac{1}{n}\right),\left[\frac{1}{n},\frac{2}{n}\right),\ldots\left[\frac{n-1}{n},1\right).$$

This will be our boxes! We are going to put $n+1$ numbers into these boxes. For every integer n between 0 and n (inclusive 0 and n so there are $n+1$ integers here in all) let

$$z_j = j\cdot\alpha - [j\cdot\alpha]$$

where $[j\cdot\alpha]$ is the greatest integer part of $j\cdot\alpha$. Then $0\le z_j<1$ so we can put these numbers in the n intervals of length $1/n$. By the Dirichlet principle at least one of the n intervals will contain 2 of these numbers. Let us work out an example with $\alpha=\sqrt{3}$ and $n=5$. The numbers z_j are tabulated below.

j	$j\sqrt{3}$	z_j
0	0.000	.000
1	1.732	.732
2	3.464	.464
3	5.196	.196
4	6.928	.928
5	8.660	.660

We see that the five intervals $[0,0.2)$, $[0.2,0.4)$, $[0.4,0.6)$, $[0.6,0.8)$ and $[0.8,1)$ receive $2,0,1,2,1$ numbers z_j, respectively:

$$\{.000 \text{ and } .196\},\{\cdot\},\{.464\},\{.660 \text{ and } .732\},\{.928\}.$$

Now if z_j and z_i (assume that $j<i$) are two numbers in the same box (i.e. an interval of width $1/n$) then

$$|z_j - z_i| = |([i\alpha]-[j\alpha])-(i-j)\alpha| < \frac{1}{n}.$$

If we set $x=[j\alpha]-[i\alpha]$ and $y=i-j$ the inequality can be rewritten as

$$|x-y\alpha|<\frac{1}{n}.$$

Note that, by taking n larger and larger, the expression $|x-y\alpha|$ can be made arbitrarily small. In particular, we have infinitely many pairs (x,y) satisfying this inequality, for various values of n. Since $y=i-j<n$, the inequality implies the following weaker, yet more practical inequality

$$|x-y\alpha|<\frac{1}{y}.$$

In this way we have obtained the following theorem of Dirichlet on rational approximations of irrational numbers.

THEOREM 42. *Let α be a positive irrational number. Then there are infinitely many pairs of positive integers (x, y) such that*

$$\left|\frac{x}{y} - \alpha\right| < \frac{1}{y^2}.$$

Now that we have figured out a way to approximate irrational numbers by rational numbers, let us see if this helps us to find a solution of the Pell equation $x^2 - Dy^2 = 1$. As we have argued above, Dirichlet's principle implies that for every n there exist positive integers $x \leq n\sqrt{D}$ and $y \leq n$ such that

$$|x - y\sqrt{D}| < \frac{1}{n}.$$

Since $x + y\sqrt{D} \leq n\sqrt{D} + n\sqrt{D} = 2n\sqrt{D}$, it follows that

$$|x - y\sqrt{D}| \cdot |x + y\sqrt{D}| < \frac{2n\sqrt{D}}{n} = 2\sqrt{D}.$$

Thus, the Dirichlet approximation guarantees that there are infinitely many pairs (x, y) such that $x^2 - Dy^2$ is an integer between $-2\sqrt{D}$ and $2\sqrt{D}$. Since there are finitely many integers between these two numbers, there must be a number m (also between $-2\sqrt{D}$ and $2\sqrt{D}$) such that the equation

$$x^2 - Dy^2 = m$$

has infinitely many integer solutions. If (x_1, y_1) and (x_2, y_2) are two solutions, then the quotient

$$\frac{x_1 + y_1\sqrt{D}}{x_2 + y_2\sqrt{D}} = \frac{x_1x_2 - Dy_1y_2}{m} + \frac{x_2y_1 - x_1y_2}{m}\sqrt{D}$$

satisfies the equation $x^2 - Dy^2 = 1$. (As the norm of a quotient is equal to the quotient of the norms.) However, it is not clear that the coefficients of this numbers are integers. To assure that, we apply the Dirichlet principle once again! We use m^2 boxes $B_{a,b}$ where $1 \leq a, b \leq m$. If (x, y) is a solution of the equation $x^2 - dy^2 = m$ we put it in the box $B_{a,b}$ if

$$x \equiv a \pmod{m}$$

$$y \equiv b \pmod{m}$$

Since there are infinitely many solutions there is a box containing at least two solutions, say (x_1, y_1) and (x_2, y_2). Then

$$x_1x_2 - Dy_1y_2 \equiv x_1^2 - Dy_1^2 = m \pmod{m}$$

$$x_2y_1 - x_1y_2 \equiv x_1y_1 - x_1y_1 = 0 \pmod{m}$$

which guarantees that the quotient considered above has integer coefficients. This shows that the Pell equation always has a non-trivial solution.

Let us illustrate what we have just done with an example. The Dirichlet approximation assures us that the equation $x^2 - 2y^2 = m$ has infinitely may solutions for some m between $-2\sqrt{2} = -2.82\ldots$ and $2\sqrt{2} = 2.82\ldots$. Moreover, we can pick two solutions of $x^2 - 2y^2 = m$ which are congruent modulo m. For example, $(2,1)$ and $(10,7)$ are two solutions of $x^2 - 2y^2 = 2$ which are congruent modulo 2, since $10 \equiv 2 \pmod 2$ and $7 \equiv 1 \pmod 2$. Thus, the quotient of $10+7\sqrt{2}$ and $2+\sqrt{2}$ should give a solution to the Pell equation $x^2 - 2y^2 = 1$. Indeed,

$$\frac{10+7\sqrt{2}}{2+\sqrt{2}} = 3 + 2\sqrt{2}$$

and $(3,2)$ is the first non-trivial solution of $x^2 - 2y^2 = 1$.

Exercises

1) Use the Dirichlet principle with $n = 10$ to find a rational approximation of π.

4) Use the pigeonhole principle to show that there is a multiple of 2005 whose digits are 0's and 1's. Hint: consider the sequence 1, 11, 111, ...

4. Pell Equation: classifying solutions

Given a Pell Equation $x^2 - Dy^2 = 1$, we can order all solutions (x,y) such that x and y are positive integers, by the size of x or y coordinate. It does not matter whether we use x or y, since the two coordinates are tied up by the equation $x^2 = 1 + Dy^2$. The solution with the smallest positive x and y is called the *first* solution of the Pell equation.

Consider now the Pell equation $x^2 - 2y^2 = 1$. By inspection we find that $(3,2)$ has the smallest positive x and y coordinate. The main result of this section is to show that the first solution generates all other solutions of the Pell equation.

Theorem 43. *Any integer solution (u,v) of the equation $x^2 - 2y^2 = 1$, with u and v positive satisfies*

$$u + v\sqrt{2} = (3 + 2\sqrt{2})^k$$

for some positive integer k.

Proof. The idea is to keep dividing $u + v\sqrt{2}$ by $3 + 2\sqrt{2}$ until we get $3 + 2\sqrt{2}$. Of course, if $(u,v) = (3,2)$ there is nothing to prove, so assume that (u,v) is different form $(3,2)$, which means that $u > 3$. Dividing $u+v\sqrt{2}$ by $3 = 2\sqrt{2}$ gives

$$\frac{u+v\sqrt{2}}{3+2\sqrt{2}} = (u+v\sqrt{2})(3-2\sqrt{2}) = (3u-4v) + (3v-2u)\sqrt{2} = u_1 + v_1\sqrt{2}.$$

Notice that (u_1, v_1) is another solution of $x^2 - 2y^2 = 1$, since it is obtained by multiplying tow solutions: $u + v\sqrt{2}$ and $3 - 2\sqrt{2}$. We claim that it has the following properties:

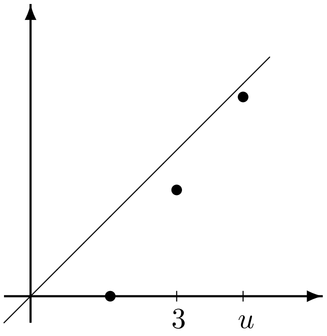
(1) $u_1 > 0$.
(2) $v_1 > 0$
(3) $u_1 < u$.

The first inequality is easy. Notice that if $x^2 - 2y^2 = 1$ then $x^2 > 2y^2$ or $x > \sqrt{2}y$. Thus

$$u > \sqrt{2} \cdot v \text{ and } 3 > \sqrt{2} \cdot 2.$$

Multiplying this two inequalities gives $3u > 4v$ or $3u - 4v > 0$ which is exactly what we want, since $u_1 = 3u - 4v$.

Next, we deal with the second inequality $v_1 > 0$. This is the trickiest part. However, we still have not used the fact that $(3, 2)$ is the solution of $x^2 - 2y^2 = 1$ with the smallest positive x and y coordinates. We use a bit of geometry to exploit this fact.



The solutions $(3, 2)$ and (u, v) sit on a branch of the hyperbola $x^2 - 2y^2 = 1$ contained in the first quadrant, which has the line

$$y = \frac{x}{\sqrt{2}}$$

as an asymptote. Since $u > 3$, the point (u, v) is closer to the asymptote then the point $(3, 2)$, as the picture indicates. It follows that the slope of the vector (u, v) is bigger then the slope of the vector $(3, 2)$:

$$\frac{v}{u} > \frac{2}{3}$$

which is equivalent to $2v - 3u > 0$. This shows that v_1 is positive, as claimed.

It remains to deal with the third inequality. This is easy. Since

$$u + v\sqrt{2} = (u_1 + v_1\sqrt{2})(3 + 2\sqrt{2}) = (3u_1 + 4v_1) + (3v_1 + 2u_1)\sqrt{2}$$

we see that $u = 3u_1 + 4v_1$. Since we have already shown that u_1 and v_1 are positive, u must be bigger than u_1.

Summarizing, starting with a positive solution (u, v) different from $(3, 2)$, we constructed another positive solution (u_1, v_1) with $u_1 < u$. If $u_1 > 3$, we can repeat this process and get another positive solution (u_2, v_2) such that $u_2 < u_1$. Continuing in this fashion, as long as we can, we get a sequence of positive solutions (u_i, v_i), $i = 1, 2, \ldots k - 1$ such that

$$u > u_1 > u_2 > \ldots > u_{k-1}$$

and $u_i > 3$ for all $i < k - 1$ and $u_{k-1} \leq 3$ (so we cannot continue with this construction). Since (u_{k-1}, v_{k-1}) is a positive solution with $u_{k-1} \leq 3$, it has to be $(3, 2)$. Thus

$$(u + v\sqrt{2})(3 + 2\sqrt{2})^{-k+1} = 3 + 2\sqrt{2}$$

or

$$u + v\sqrt{2} = (3 + 2\sqrt{2})^k.$$

□

Of course, Theorem 43 holds for any Pell equation. All solutions with positive coordinates are obtained from the first one. So, how does one find the first solution of $x^2 - Dy^2 = 1$? This might not be so easy. For example, the first solution of $x^2 - 94y^2 = 1$ is

$$2143295 + 211064\sqrt{94},$$

In particular, the first solution might not be easy to find by guessing even for small values of D. However, the first solution can be computed using the Continued Fractions algorithm, and this will be discussed in the next section.

5. Continued Fractions

In this section we shall describe, without a proof, an effective way to construct the first solution of the Pell equation $x^2 - Dy^2 = 1$. We assume some basic knowledge of determinants.

The method is based on the Continued Fractions algorithm which, we recall, is defined as follows:

(1) Let $\alpha > 1$. Put $\beta = \alpha - [\alpha]$ where $[\alpha]$ is the greatest integer part of α.
(2) If $\beta = 0$ stop, else put $\alpha_1 = 1/\beta$ and go to (1).

As we discussed in the first chapter, the algorithm does not stop unless α is a rational number. Thus, starting with an irrational number α, the the algorithm generates a sequence of irrational numbers

$$\alpha, \alpha_1, \alpha_2, \ldots$$

and the sequence of integers

$$[\alpha], [\alpha_1], [\alpha_2] \ldots.$$

The name continued fraction comes from the fact that we can write down α as the following sequence of expressions:

$$\alpha = [\alpha] + \frac{1}{\alpha_1} = [\alpha] + \frac{1}{[\alpha_1] + \frac{1}{\alpha_2}} = \ldots$$

A continued fraction expansion is called purely periodic with period p if $\alpha_p = \alpha$. Not that this implies that $\alpha_{p+1} = \alpha_1$, $\alpha_{p+2} = \alpha_2$ and so on. The key fact needed to find the first solution of the Pell equation is:

If D is not a square then the continued fraction of $\sqrt{D} + [D]$ is purely periodic.

A proof of this fact can be found in the book by Davenport (REFERENCE). We only give some examples here. First, take $D = 2$, then $\alpha = \sqrt{2} + [\sqrt{2}] = \sqrt{2} + 1$. The first loop of the continued fractions algorithm gives

$$\beta = \sqrt{2} + 1 - [\sqrt{2} + 1] = \sqrt{2} - 1$$

and

$$\alpha_1 = \frac{1}{\beta} = \frac{1}{\sqrt{2} - 1} \cdot \frac{\sqrt{2} + 1}{\sqrt{2} + 1} = \sqrt{2} + 1.$$

This implies that $\alpha_1 = \alpha$ and the fraction has the period 1! A similar calculation can be performed easily for $\sqrt{D} + [\sqrt{D}]$ with small D. We give a short list of answers here. The right column of the table gives the full period of integers

$$[\alpha], \ldots, [\alpha_{p-1}].$$

D	
2	2
3	2, 1
5	3
7	4, 1, 1, 1
11	3, 6
13	6, 1, 1, 1, 1
17	8
19	8, 2, 1, 3, 1, 2

We are now ready to explain how to obtain the first solution of the Pell equation using the continued fraction expansion. If p is the period, then we can expand

$$\alpha = [\alpha] + \cfrac{1}{\ddots + \frac{1}{[\alpha_{p-1}] + \frac{1}{\alpha}}}$$

For example, if $\alpha = \sqrt{3} + [\sqrt{3}]$, then $p = 2$, and we have

$$\sqrt{3} + 1 = 2 + \cfrac{1}{1 + \frac{1}{\sqrt{3}+1}}.$$

We claim that this identity can be rewritten in the form

$$\sqrt{3} = \frac{a\sqrt{3} + b}{c\sqrt{3} + d}$$

where a, b, c and d are integers. This is accomplished by working up from the last fraction on the bottom, as follows:

$$\sqrt{3} + 1 = 2 + \cfrac{1}{1 + \frac{1}{\sqrt{3}+1}} = 2 + \cfrac{1}{\frac{\sqrt{3}+2}{\sqrt{3}+1}} = 2 + \frac{\sqrt{3}+1}{\sqrt{3}+2} = 1 + \frac{2\sqrt{3}+3}{\sqrt{3}+2}$$

(note that we never "rationalize" or use $(\sqrt{3})^2 = 3$). After subtracting 1 from both sides, we get

$$\sqrt{3} = \frac{2\sqrt{3} + 3}{\sqrt{3} + 2}.$$

The coefficients a, b, c and d of the fraction on the right give a 2×2 matrix of determinant 1:

$$\begin{vmatrix} 2 & 3 \\ 1 & 2 \end{vmatrix} = \begin{vmatrix} 2 & 3 \cdot 1 \\ 1 & 2 \end{vmatrix} = 2^2 - 3 \cdot 1^2 = 1$$

We recognize here $(2, 1)$, the first solution of the Pell equation $x^2 - 3y^2 = 1$. Is this an accident? Let us work out another example, $D = 6$. The continued fraction of $\sqrt{6} + [\sqrt{6}] = \sqrt{6} + 2$ has the period of length 2:

$$\sqrt{6} + 2 = 4 + \cfrac{1}{2 + \frac{1}{\sqrt{6}+2}}.$$

Again, this equation can be rewritten as

$$\sqrt{6} = \frac{5\sqrt{6} + 12}{2\sqrt{6} + 5}$$

and the coefficients in the fraction on the right hand side give a 2×2 matrix with determinant 1:

$$\begin{vmatrix} 5 & 12 \\ 2 & 5 \end{vmatrix} = \begin{vmatrix} 5 & 6 \cdot 2 \\ 2 & 5 \end{vmatrix} = 5^2 - 6 \cdot 2^2 = 1$$

Again, we get the first solution $(5, 2)$ of the Pell equation $x^2 - 6y^2 = 1$.

More generally, using the period of the continued fraction for $\sqrt{D}+[\sqrt{D}]$ we can get an expression

$$\sqrt{D} = \frac{a\sqrt{D}+Dc}{c\sqrt{D}+a}$$

such that the determinant

$$\begin{vmatrix} a & Dc \\ c & a \end{vmatrix} = a^2 - D\cdot c^2$$

is 1 if the period of the continued fraction has even length and -1 if the period has odd length. In the later case, we get a solution (x_0, y_0) of the equation $x^2 - Dy^2 = -1$ from which, by squaring $x_0 + y_0\sqrt{D}$, we get the first solution of $x^2 - Dy^2 = 1$.

Exercises

1) Compute the period of the continued fraction of
   a) $\sqrt{15}+[\sqrt{15}]$.
   b) $\sqrt{19}+[\sqrt{19}]$.

2) Use the exercise 1) to find the first solution of
   a) $x^2 - 15y^2 = 1$.
   b) $x^2 - 19y^2 = 1$.

3) The continued fraction algorithm for a number $\alpha > 1$ is purely periodic of period 2 such that $[\alpha] = [\alpha_2] = \ldots = 1$ and $[\alpha_1] = [\alpha_3] = \ldots = 2$. Find the number α.

CHAPTER 10

Cryptography

1. Diffie-Hellman key exchange

A cypher is a method of making a message unreadable to the general public. There are two aspects of any cypher: The encryption procedure and the decryption procedure. One of the simplest cyphers is the shift (or Caesar) cypher. Here the encryption procedure consists of shifting the letters in a message by a fixed number of spots in the alphabetical order. Thus here the *key* is a positive integer k. For example, if $k = 5$ then, in order to encrypt a message we shift the letters of the alphabet

A	B	C	D	E	F	G	H	I	J	K	L	M
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

by five places in the usual ordering. Thus $A \mapsto F$, $B \mapsto G$, $C \mapsto H$ and so on. For example, the word OLYMPIC is encrypted as

TQDRUNH

On the other hand, if we receive an encrypted message

ITBSMNQQ

then, by shifting the letters by five places back we decrypt the message as

DOWNHILL

Since the encryption and decryption procedures are essentially the same the shift cypher is an example of a *symmetric cypher.* Those who know how to encrypt a message are also those that can decrypt the message.

Of course, this encryption procedure is very easy to break, since there exist only 26 possible sifts (keys). A slight improvement would be, for example, to break up the set of 26 letters into 2 groups of 14 letters, and then to apply two different shifts k_1 and k_2 for each of the two groups of letters. This is already more secure since there are $13 \times 13 = 196$ possible combinations. The two keys can be represented by the same number k as follows. If $a_n a_{n-1} \cdots a_2 a_1$ are the digits of k then we can build two numbers from k

by using the digits of k as follows:

$$\begin{cases} k_1 = \ldots a_3 a_1 \\ k_2 = \ldots a_4 a_2. \end{cases}$$

Thus the key in this case would also be given by one (secret) number k. Of course, we can do even better, by dividing 26 letters into 4 groups, say of 6, 7, 6 and 7 letters and perform an independent shift for each group. This gives $6 \times 7 \times 6 \times 7 = 1764$ possible combinations.

Still, if we were to use a variant of the shift key, we would be far from secure, especially in the days of fast computers. Indeed, the frequency of letters in three well known English language novels

	Alice in Wonderland	Hamlet	Treasure Island
space	19.75%	15.70%	18.61%
E	9.40%	9.04%	9.28%
T	7.43%	7.11%	6.96%
A	6.00%	5.87%	6.54%
O	5.69%	6.53%	6.03%

indicates that the letter E appears most often, followed by the letters T, A and O. Thus, a repeated use of any permutation cypher would almost certainly doom it.

A simple, but efficient idea, would be to change the key often. But how to accomplish an exchange of a new key if the present key has been broken? That is precisely what the idea of a *public key exchange* of Diffie and Hellman is about.

The Diffie-Hellman key exchange was developed around 1970 and is based on the discrete logarithm. Pick a prime p and a primitive root g modulo p. These numbers can be made public. In practice p has to be large, but here we shall take $p = 29$ and $g = 2$. Here is how Caesar and Cleopatra would perform a secure key exchange over unsecure channels:

(1) Caesar thinks of a (secret) number x, for example $x = 10$, and calculates $X = 2^x$ modulo 29:

$$2^{10} \equiv 9 \pmod{29}$$

He sends $X = 9$ to Cleopatra over an unsecured channel (the whole world can know this number).

(2) Cleopatra thinks of a (secret) number y, for example $y = 18$,1123 and calculates $Y = 2^y$ modulo 29:

$$2^{18} \equiv 13 \pmod{29}$$

She sends $Y = 13$ to Caesar over an unsecured channel (the whole world can know this number).

(3) Caesar receives the number Y from Cleopatra and calculates the *key* modulo 29:

$$k \equiv Y^x \equiv 7 \pmod{29}$$

(4) Cleopatra receives the number X from Caesar and calculates the *key* modulo 29:

$$k \equiv X^y \equiv 7 \pmod{29}$$

The key k is the same for Caesar and Cleopatra because

$$(2^x)^y \equiv (2^y)^x \pmod{29}.$$

With the key in hands, our heroes can exchange a message. Say, for example, that Cleopatra's message starts with MEET ME IN ... which she encrypts using the key $k = 7$ into

TLLA TL PU

At this point, in order to make the exchange of messages more secure, Cleopatra decides to exchange the key once more. She sends the number $Y = 7$ to Caesar and receives the number $X = 13$ from him. She then calculates the key and sends the rest of the message which, encrypted, reads

XIBUXKAOFX

In order to decipher the rest of the message we need to figure out the key. Despite the fact that we have the numbers X and Y, we cannot calculate the key k since we do not know x or y, the discrete logarithms of X and Y modulo 29 with respect to the base 2. While it is easy to calculate a power of 2 modulo any prime p, there is no known algorithm to perform the inverse operation! This is the strength of the Diffie-Hellman key exchange.

Of course, since $p = 29$ is quite small, it is possible to find out the discrete logarithm simply by listing all powers of 2:

I	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2^I	2	4	8	16	3	6	12	24	19	9	18	7	14	28

The list shows that Cleopatra has taken $y = 12$ in the second key exchange. With the number y in hand, we can calculate the key

$$k \equiv X^y \equiv 13^{12} \equiv 23 \pmod{29}.$$

Thus we have to shift the letters in XIBUXKAOFX by 23 places backward, which is the same as 3 places forward. Decrypting gives

ALEXANDRIA

Summarizing, the message MEET ME IN ALEXANDRIA has been encrypted into TLLA TL PU XIBUXKAOFX using the symmetric key exchange twice.

The Diffie-Hellman key exchange was published by Whitfield Diffie and Martin Hellman in 1976. Since the idea is based on concept developed by Ralph Merkle the key is sometimes called the Diffie-Hellman-Merkle key. In any case, all three are credited in the U. S. patent, which is now expired.

Exercises

1) A shift cypher key is exchanged using the Diffie-Hellman method with $g = 5$ and $p = 47$. The actual numbers exchanged were $X = 38$ and $Y = 3$. Find that key.

2) Using the key in the previous exercise decipher the message:

EQPITCVWNCVKQPU

2. RSA Code

The shift cypher, discussed in the previous lecture, is an example of a symmetric cypher. This means that the encryption and decryption procedures are essentially the same. In this lecture we introduce another type of cypher, the RSA cypher. It is an assymetric cypher, meaning that the cypher, once we have fixed a key, is used for a secure flow of information in only one direction. In a nutshell, the RSA cypher is based on the fact that it is very easy to multiply numbers but not easy to factor them.

In order to explain the RSA cypher, we shall restrict the alphabet to nine letters: E, A, O, H, L, M, N. K and I. This is done for practical reasons, only. It will allow us to work with small numbers and to do computations by hand. There are plenty of words that can be written using these letters. For example, the Hawaiian alphabet contains only twelve letters: these nine, U, P and W. The first step is to replace the letters by numbers:

E	A	O	H	L	M	N	K	I
1	2	3	4	5	6	7	8	9

This table itself is a symmetric cypher. Letters in every word can be replaced by the numbers (encrypting) and then words can be recovered by reverse substitution (decrypting). The table, of course, has to be kept secret, if there is any use of this. This code, however, is rather easy to break. We shall do much better in a moment.

King Kamehameha wants to create a secure channel to receive messages. The King takes two (large) prime numbers p and q which he multiplies to get a composite number $m = p \cdot q$. He also picks a number e relatively prime to $\varphi(m) = (p - 1)(q - 1)$. Then he announces to friends and foes alike that his *public* or encryption key is the number e with the modulus m. However, he keeps proprietary the factorization $m = p \cdot q$ on which his *private* or decryption key is based. Assume, for example, that the announced values of the public key are

$$m = 1517 \text{ and } e = 11.$$

Encryption procedure: Assume, for example, that a general wants to send to the King a message which starts with the word

KONA

The first step is to convert the letters to their number equivalents, given by the above table. This gives the following sequence of numbers

$$8372.$$

The next step is to break up the sequence 8372 into smaller subsequences such as:

$$83, 72 \text{ or } 837, 2 \text{ or } 8, 372, ...$$

The important thing here is that every subsequence consists of ≤ 3 digits because m has 4 digits. Assuming, for example, that the general picked 83 and 72, the next step is to replace 83 and 72 by their e-th power modulo m. With our choices, this is 83^{11} and 72^{11} modulo 1517:

$$83^{11} \equiv 821 \pmod{1517}$$

$$72^{11} \equiv 1097 \pmod{1517}$$

Of course, an efficient way to do this is by successive squaring:

$$\begin{array}{rcll} 83^1 & \equiv & 83 & \pmod{1517} \\ 83^2 & \equiv & 821 & \pmod{1517} \\ 83^4 & \equiv & 493 & \pmod{1517} \\ 83^8 & \equiv & 329 & \pmod{1517} \end{array}$$

which gives $83^{11} = 83^8 \cdot 83^2 \cdot 83 \equiv 329 \cdot 821 \cdot 83 \equiv 821 \pmod{1517}$. Similarly,

$$\begin{array}{rcll} 72^1 & \equiv & 72 & \pmod{1517} \\ 72^2 & \equiv & 633 & \pmod{1517} \\ 72^4 & \equiv & 201 & \pmod{1517} \\ 72^8 & \equiv & 959 & \pmod{1517} \end{array}$$

which gives $72^{11} = 72^8 \cdot 72^2 \cdot 72 \equiv 959 \cdot 633 \cdot 72 \equiv 1097 \pmod{1517}$.

In this way the general has encrypted the word KONA by a sequence of two numbers 821 and 1097. The second word of the message - which we have not revealed yet - is encrypted using the same procedure into a sequence 33, 108. Then the whole message

$$821, 1097, 33, 108$$

is sent to the King by an unsecured channel.

Decryption procedure: The King receives the encrypted message 821, 1097, 33 and 108. In order to recover the original message the King needs to solve the following four equations:

$$\begin{array}{rcll} x^{11} & \equiv & 821 & \pmod{1517} \\ x^{11} & \equiv & 1097 & \pmod{1517} \\ x^{11} & \equiv & 33 & \pmod{1517} \\ x^{11} & \equiv & 108 & \pmod{1517} \end{array}$$

The (unique) solutions to these four equations are given by 821^d, 1097^d 33^d and 108^d where d is the inverse of 11 modulo $\varphi(1517)$. This was explained earlier, but we repeat the argument here. If d is an inverse of 11 modulo $\varphi(m)$ then $11d = 1 + k\varphi(1517)$ for some integer k. Then

$$(x^{11})^d = x^{11d} = x \cdot x^{k\varphi(1517)} \equiv x \pmod{1517}$$

since $x^{\varphi(1517)} \equiv 1 \pmod{1517}$. Thus if we take the congruence $x^{11} \equiv 821 \pmod{1517}$ to the d-th power, we obtain

$$x \equiv 821^d \pmod{1517}$$

as claimed. But d cannot be calculated unless $\varphi(1517)$ or, equivalently, the factorization of 1517 is known. This information, however, is only known to the King since he has created the cypher:

$$1517 = 37 \cdot 41,$$

Thus, he knows that $\varphi(1517) = 36 \cdot 40 = 1440$ and the Euclidean algorithm applied to 1440 and 11 gives

$$1440 = 130 \cdot 11 + 10 \text{ and } 11 = 1 \cdot 10 + 1.$$

These two equations, combined, give $131 \cdot 11 - 1440 = 1$. It follows that $d = 131$ is the inverse of 11 modulo 1440. The number d is called the *decryption* or *private key*. Of course 821^{131} and 1097^{131} give back 83 and 72, so let's see what the rest of the message is. Calculating consecutive squares gives

$$\begin{array}{rcll} 33^1 & \equiv & 33 & \pmod{1517} \\ 33^2 & \equiv & 1089 & \pmod{1517} \\ 33^{2^2} & \equiv & 1144 & \pmod{1517} \\ 33^{2^3} & \equiv & 1082 & \pmod{1517} \\ 33^{2^4} & \equiv & 1117 & \pmod{1517} \\ 33^{2^5} & \equiv & 715 & \pmod{1517} \\ 33^{2^6} & \equiv & 1513 & \pmod{1517} \\ 33^{2^7} & \equiv & 16 & \pmod{1517} \end{array}$$

so

$$33^{131} = 33^{2^7} \cdot 33^2 \cdot 33 \equiv 16 \cdot 1089 \cdot 33 \equiv 49 \pmod{1517}.$$

Also,

$$\begin{array}{rcll} 108^1 & \equiv & 108 & \pmod{1517} \\ 108^2 & \equiv & 1045 & \pmod{1517} \\ 108^{2^2} & \equiv & 1302 & \pmod{1517} \\ 108^{2^3} & \equiv & 715 & \pmod{1517} \\ 108^{2^4} & \equiv & 1513 & \pmod{1517} \\ 108^{2^5} & \equiv & 16 & \pmod{1517} \\ 108^{2^6} & \equiv & 256 & \pmod{1517} \\ 108^{2^7} & \equiv & 305 & \pmod{1517} \end{array}$$

which gives

$$108^{131} = 108^{2^7} \cdot 108^2 \cdot 108 \equiv 305 \cdot 1045 \cdot 108 \equiv 53 \pmod{1517}.$$

Thus the second part of the message decrypted is 49, 53 which corresponds to HILO. (The King, unfortunately, will have to do all these calculations himself, if he is to keep the decryption key secret.)

Notice that the public key created by the King is used only for a secure transmission of messages to the King. If the King wants to send a message back to his general, or to anyone else for that matter, he can do so as long as the intended receiver has his or her own public key! This, truly a remarkable feature of RSA, has added a completely new dimension to cryptography.

Digital Signature. RSA has an additional security feature that allows identification of the sender. Assume, for example, that the King wants to send a message to the Queen Kea Lani. Of course, he can do that securely using Queen's public key. But how can she be sure that the message was sent by the King and not by an impostor? (That would be an example of a so called third party attack.) This problem is resoved by adding King's digital signature to the message. The digital signature is created as follows. First, the word KAMEHAMEHA is replaced by a sequence of numbers $826, 1426, 142$ using our substitution scheme. Then King's digital signature is the sequence

$$X = 826^{131}, Y = 1426^{131}, Z = 142^{131}$$

computed modulo 1517, where the exponent used is $d = 131$, King's private key. The King signs the message to the Queen by including his digital signature in the message, encrypted using Queen's public key. After the Queen decrypts the message she will recover King's digital signature X, Y, Z. She then uses King's public key $e = 11$ to compute X^{11}, Y^{11}, Z^{11} modulo 1517. These three numbers should be $826, 1426, 142$ if the King was the sender.

The RSA is named after Rivest, Shamir and Adelman. In 1977 they challenged mathematical community to decipher a message based on the modulus $n = pq$ that is 129 digits long. In 1994, after seventeen years, Atkins, Graff, Lenstra and Leyland finally succeeded in factoring n.

Exercises

1) With $m = 1517$ and $e = 11$, decipher the message:

$$1373 \qquad 1149 \qquad 108$$

Express the final answer in terms of our nine letter alphabet.

2) Encypt KO-NA using the RSA cypher with modulus $m = 1517$ and the encryption key $e = 7$.

3) An RSA cypher has modulus $m = 4189$ and the encryption key $e = 11$. Find the decryption key.

3. ElGamal Code

El Gamal is another public key cypher. It is based on the Diffie-Hellman key exchange. Again, assume that King Kamehameha wants to create a secure channel to receive messages. The King takes a large prime number p and an integer g of large order modulo p. Of course, the largest possible order modulo p is $p-1$ - for primitive roots - but we do not insist that g is a primitive root. Then the King picks a secret number x, and computes

$$X \equiv g^x \pmod{p}.$$

The triple (p, g, X) is made public, while the King keeps x secret. For example, assume that the numbers are $p = 131$, $g = 2$ and $x = 37$. Then

$$\begin{array}{rcll} 2^4 & \equiv & 16 & \pmod{131} \\ 2^8 & \equiv & -6 & \pmod{131} \\ 2^{16} & \equiv & 36 & \pmod{131} \\ 2^{32} & \equiv & 117 & \pmod{131} \end{array}$$

which gives $2^{37} = 2^{32} \cdot 2^4 \cdot 2 \equiv 117 \cdot 16 \cdot 2 \equiv 76 \pmod{131}$. Thus, the announced triple is

$$(p, g, X) = (131, 2, 76).$$

Encryption procedure: Unlike the RSA code, the encryption procedure here is not canonical. The encryptor needs to *pick* a (secret) number y which is used to calculate two numbers:

$$Y \equiv g^y \pmod{p}$$

and

$$k \equiv X^y \pmod{p}$$

(The number y should be large and relatively prime to $p-1$, if there is any use of this.) The number k is now the encryption key. Assume, for example, that $y = 19$. Then

$$Y \equiv 2^{19} \equiv 26 \pmod{131}$$

In order to calculate the key $k \equiv 76^{19} \pmod{131}$ we use the method of successive squares:

$$\begin{array}{rcrl} 76^2 & \equiv & 12 & \pmod{131} \\ 76^4 & \equiv & 13 & \pmod{131} \\ 76^8 & \equiv & 38 & \pmod{131} \\ 76^{16} & \equiv & 3 & \pmod{131} \end{array}$$

Thus $k \equiv 76^{19} \equiv 76^{16} \cdot 76^2 \cdot 76 \equiv 3 \cdot 12 \cdot 76 \equiv 116 \pmod{131}$.

Assume, for example, that we want to encrypt a nine letter word with the first four letters

HALE

The first step is to convert the letters to their number equivalents given by the table introduced in the previous sections:

E	A	O	H	L	M	N	K	I
1	2	3	4	5	6	7	8	9

In particular, HALE is converted into

4251

The next step is to break up the sequence 4251 into smaller subsequences each of which is less then $p = 131$. Assume, for example, that 4251 is broken into 42 and 51. Then the next step is to replace 42 and 51 by $42 \times k$ and $51 \times k$ modulo p. In our example, this would be

$$42 \times 116 \equiv 25 \pmod{131}$$

$$51 \times 116 \equiv 21 \pmod{131}$$

In this way HALE is encrypted by a sequence of two numbers 25, 21. The second part of the word - which we have not reveled yet - is encrypted using the same procedure into a sequence 101, 80, 6. Then the whole message, including $Y = 26$ as the header

$$Y; 25, 21, 101, 80, 6$$

is sent to the King by an unsecured channel.

Decryption procedure: The King receives the encrypted message

$$26; 25, 21, 101, 80, 6.$$

In order to recover the original message the King needs to compute the key k which is given by the formula

$$k \equiv Y^x \equiv 26^{37} \pmod{131}.$$

The key k is efficiently calculated using consecutive squares:

$$\begin{array}{rcll} 26^2 & \equiv & 21 & \pmod{131} \\ 26^4 & \equiv & 48 & \pmod{131} \\ 26^8 & \equiv & 77 & \pmod{131} \\ 26^{16} & \equiv & 34 & \pmod{131} \\ 26^{32} & \equiv & 108 & \pmod{131} \end{array}$$

This gives $k \equiv 26^{37} \equiv 26^{32} \cdot 26^4 \cdot 26 \equiv 108 \cdot 48 \cdot 26 \equiv 116 \pmod{131}$, exactly the same as the encryptor's k. The final step is now to divide the segments of the received message by 116. Since the multiplicative inverse of 116 modulo 131 is 96,

$$\begin{array}{rcll} 25 \times 96 & \equiv & 42 & \pmod{131} \\ 21 \times 96 & \equiv & 51 & \pmod{131} \\ 101 \times 96 & \equiv & 2 & \pmod{131} \\ 80 \times 96 & \equiv & 82 & \pmod{131} \\ 6 \times 96 & \equiv & 52 & \pmod{131} \end{array}$$

and the message is decrypted as 42, 51, 2, 49 and 53 that gives the word

HALEAKALA

The sender may want to change the choice of y often. If this is done by a computer we need a reliable *random number* generator. More generally, the following four natural problems appear and need to be considered in the cryptography:

(1) Computing discrete logarithm.
(2) Primality testing.
(3) Factoring attacks.
(4) Random number generation.

Primality testing and factoring are the main topics of the next chapter. We finish this section with an approach to computing discrete logarithms. Recall that ℓ is the discrete logarithm of a with respect to g, a primitive root modulo p, if

$$a \equiv g^\ell \pmod{p}.$$

Given a, one way to compute the discrete logarithm ℓ is to start listing all possible powers $g, g^2, g^3 \dots$ of g. The expected number of steps in this way is $(p-1)/2$. There is a better way, called the baby-step, giant-step method. It takes less than $2\sqrt{p}$ number of steps. To explain, let m be the smallest integer such that $p - 1 \leq m^2$. (Note that m is equal to $\sqrt{p-1}$ rounded up to the nearest integer.) Then ℓ, since it is less than p, can be written as

$$\ell = m \cdot i + j$$

for two integers i and j each of which is less than m. The first step in the algorithm is to list the list first m powers of g (These are the baby steps of the method.):

$$1, g^1, \dots g^{m-1}.$$

The next step is to start computing $a \cdot (g^{-m})^i$ for $i = 1, 2 \ldots$. (These are the giant steps of the method.) Since

$$g^j = a \cdot (g^{-m})^i$$

for some pair of integers i and j less than m, eventually, $a \cdot (g^{-m})^i$ will show up on the list.

Let's see how this works on an example. Take $p = 29$ and $g = 2$. Assume that we want to compute the discrete logarithm of $a = 11$. We can take here $m = 6$. Then the list of the first 6 powers of 2 modulo 29 is

j	0	1	2	3	4	5
2^j	1	2	4	8	16	3

Also $2^6 \equiv 6 \pmod{29}$ and 5 is the inverse of 6 modulo 29. Now we need to compute $11 \cdot 5$, $11 \cdot 5^2$ etc modulo 29 until we get a number on the first list:

i	0	1	2	3	4	5
$11 \cdot 5^i$	11	26	14	12	2	

We see that 2 is the first number on the second list which has already appeared on the first list. Thus

$$11 \cdot 5^4 \equiv 2^1 \pmod{29}$$

or

$$11 \equiv (2^6)^4 \cdot 2^1 \equiv 2^{25} \pmod{29}.$$

This shows that the discrete logarithm of 11 is 25.

Exercises

1) Let $p = 103$, $g = 2$ and $x = 47$. Compute

$$X \equiv 2^{47} \pmod{103}.$$

2) Using the $p = 103$, $g = 2$ and X from the previous lecture, use the ElGamal cypher with $y = 31$ to encrypt KO-NA. (Substitute KO-NA by 83-72.)

3) Use the baby-step giant-step algorithm to find ℓ (the discrete logarithm) such that

$$2^\ell \equiv 7 \pmod{53}.$$

4) A shift cypher k is exchanged using the Diffie-Hellman method with $p = 421$ and $g = 2$. The numbers exchanged over a public channel are $X = 229$ and $Y = 247$. Using the giant-step - baby-step method compute the discrete logarithm of X (or of Y) to find the two digit key $k = k_2k_1$ and decipher following the message which has been encrypted by shifting first 13 letters (A - M) by $k_1 + 2$ places and the second 13 letters (N - Z) by $k_2 + 3$ places:

JQFEYUOJX

Note: shifting first 13 letters by 2 places, for example, means that M shifts to B.

CHAPTER 11

Primality testing

1. Miller Rabin test

In this section we discuss a method to determine whether an odd integer n is composite or prime. To that end, pick an integer $a < n$. If n is prime then the Fermat Little Theorem implies that

$$a^n \equiv a \pmod{n}.$$

If this congruence is not satisfied, then n is composite, and the number a is called a *witness*. Let's how this works with $a = 2$ and several small (composite) odd numbers:

n	9	15	21	25	27	33
2^n	6	8	17	14	25	16

In each of these examples taking 2 to the power n modulo n is different from 2, confirming what we already knew - that n is composite. Unfortunately, while this method is very efficient to detect a composite number, it does not work in every case. Moreover, even if we replace 2 by another integer, the method may still not work since there are odd numbers n such that

$$a^n \equiv a \pmod{n}$$

for all numbers a. Positive integers with this property are called Carmichael numbers. The first Carmichael number is

$$561 = 3 \cdot 11 \cdot 17.$$

It is not difficult to see that $a^{561} \equiv a \pmod{561}$ for all integers a. Indeed, since a number is divisible by 561 if and only if it is divisible by the three factors 3, 11 and 17, the congruence $a^{561} \equiv a \pmod{561}$ is equivalent to the following three congruences:

$$\begin{aligned} a^{561} &\equiv a \pmod{3} \\ a^{561} &\equiv a \pmod{11} \\ a^{561} &\equiv a \pmod{17} \end{aligned}$$

As we shall see in a moment, these congruences hold since $\varphi(3) = 2$, $\varphi(11) = 10$ and $\varphi(17) = 16$ divide 560. Consider, for example, the third congruence. Since $560 = 16 \cdot 35$ we can rewrite $a^{561} - a$ as

$$a^{561} - a = a(a^{560} - 1) = a((a^{16})^{35} - 1).$$

If 17 divides a then 17 divides $a^{561} - a$. If 17 does not divide a then $a^{16} \equiv 1 \pmod{17}$, by the Fermat Little Theorem. Hence

$$(a^{16})^{35} \equiv 1^{35} \equiv 1 \pmod{17}$$

and 17 again divides $a(a^{560} - 1)$. This argument, of course, works for 3 and 11, since

$$560 = 2 \cdot 280 \text{ and } 560 = 10 \cdot 56.$$

This shows that 561 is a Carmichael number. The next four Carmichael numbers are

$$1105,\ 1729,\ 2465,\ \text{and } 2821.$$

It is interesting to note that until recently it was not known whether there are infinitely many Carmichael numbers. Then, in 1994, Alford, Granville and Pomerance showed that there are infinitely many Carmichael numbers. Carmichael numbers are characterized by the following (Korselt's) criterion:

Proposition 44. *A number $n > 2$ is a Carmichael number if and only if the following three conditions are true:*

1. *The number n is odd.*
2. *The number n is square free. This means that p^2 does not divide n for any prime p.*
3. *For every prime factor p of n, $p - 1$ divides $n - 1$.*

Proof. Obviously, any number that satisfies the two conditions is a Carmichael number. The verification is exactly the same as what we did for 561. So assume that n is a Carmichael number. Then $a^n \equiv a \pmod{a}$ for any integer a. To prove (1), take $a = -1$. Then

$$(-1)^n \equiv -1 \pmod{n}$$

which forces n to be odd since $n > 2$. This proves (1). To prove (2), let p^{e+1} is the largest power dividing n. We need to show that $e = 0$. Take $a = p^e$. Then

$$(p^e)^n \equiv p^e \pmod{n}.$$

This implies that n, and therefore p^{e+1}, divides $p^{en} - p^e = p^e(p^{e(n-1)} - 1)$. This implies that p divides $p^{e(n-1)} - 1$ and this is possible only if $e = 0$. This verifies (2). It remains to prove (3). Let a be a primitive root modulo p. Then

$$a^n \equiv a \pmod{p}$$

implies $a^{n-1} \equiv 1 \pmod{p}$. Now recall that if G is any group and g an element in G such that $g^m = 1$ then the order of g divides m. We can apply this to our situation, where $G = (\mathbb{Z}/p\mathbb{Z})^\times$, $g = a$ and $m = n - 1$. The order of a is $p-1$ since a is a primitive root. It follows that $p-1$ divides $n-1$. The proposition is proved. □

Although $a^n \equiv a \pmod{n}$ holds for all a for some composite integers n there is a way to go around this problem, as given by the Miller Rabin test. The idea behind the Miller Rabin test is very simple, and exploits the fact that $a^{\frac{n-1}{2}}$ must be 1 or -1 if n is a prime number and $a < n$. We are now ready to state the Rabin-Miller test.

Theorem 45. *Let n be an odd integer. Write $n-1 = 2^k \cdot q$ with q odd. If there exists an integer $a < n$ such that*

(1) $a^q \not\equiv 1 \pmod{n}$
(2) $a^{2^i q} \not\equiv -1 \pmod{n}$ *for all* $i = 0, 1, \ldots, k-1$,

then n is composite.

Proof. In order to prove this theorem, we need to show that if n is prime then one of the two conditions fails. If n is prime, then for any $a < n$ we have $a^{n-1} = a^{2^k \cdot q} \equiv 1 \pmod{n}$. Next, consider the sequence

$$a^q, a^{2q}, a^{4q}, \ldots, a^{2^{k-1}q}, a^{2^k q}.$$

Note that each number in the sequence is the square of the previous. Let $a^{2^i q}$ be the first, from the left, congruent to 1 modulo n. If this happens for $i = 0$ then the first condition fails. Otherwise, there are two consecutive numbers on the list such that

$$a^{2^{i-1}q} \not\equiv 1 \pmod{n}$$

and

$$a^{2^i q} \equiv 1 \pmod{n}$$

Since $\mathbb{Z}/n\mathbb{Z}$ is a field -1 is the only number different from 1 whose square is 1. This implies that

$$a^{2^{i-1}q} \equiv -1 \pmod{n}$$

which shows that the second condition fails. Summarizing, assuming that n is prime we have shown that one of the two condition fails for every a. This completes the proof. □

Let's see how this works with $n = 561$. Then $560 = 2^4 \cdot 35$. take $a = 2$. Then

$$2^{35} \equiv 263 \pmod{561}$$

which shows that the first condition is satisfied. Further, by successive squaring,

$$\begin{aligned} 263^2 &\equiv 166 \pmod{561} \\ 166^2 &\equiv 67 \pmod{561} \\ 67^2 &\equiv 1 \pmod{561} \end{aligned}$$

which shows that the second condition is satisfied as well. The test implies that 561 is composite.

Of course, we have not answered yet an important question. If n is a composite number does there exists an a which satisfies the conditions of the Miller-Rabin test? (Such number a is called a *strong witness* to compositeness of n.) After all, the Miller-Rabin test has been introduced precisely because of the lack of *witnesses* for some composite numbers. Luckily, the answer is yes. In fact, without giving a proof, we shall state that for any composite number n, more than 75% of integers a modulo n are strong witnesses. Thus if, for example, we pick randomly 5 integers modulo n, the chance that none of them is a strong witness is less than

$$\frac{1}{4^5} < \frac{1}{1,000}.$$

In other words, if the test fails to show that n is composite for 5 choices of a, then there is a good chance (99.9%) that n is prime.

Exercises

1) Calculate 3^n modulo n for the following 10 odd numbers: $n = 5, 7, 9, \dots 23$. Partition the 10 numbers into two groups: One consisting of n such that 3^n is congruent to 3 modulo n, and the other such that 3^n is not congruent to 3 modulo n.

2) The smallest Carmichael number is 561. Use Korselt's criterion to check that

$$1105,\ 1729,\ 2465,\ 2821$$

are Carmichael numbers. These four and 561 are the first five Carmichael numbers.

3) Use Korselt's criterion to determine which of the following numbers is a Carmichael number:
a) 3457
b) 5329
c) 6601
d) 8911
e) 9011

4) Use the Miller-Rabin test to show that following numbers are composite:
a) 899.
b) 3599.
c) 427.
d) 30227.

2. $p-1$ method

We now turn our attention to the problem of factoring a large composite number n which is a product of two large primes. Of course, there is no universal algorithm - otherwise the RSA cypher would not be secure - but there are efficient algorithms which work in certain situations. One such algorithm is the method introduced by J. Pollard in 1975. This method works well if the number n has at least one prime factor p such that $p-1$ is composed of *small* primes. If that is the case, then $p-1$ divides $B! = 1 \cdot 2 \cdot \ldots \cdot B$ for a relatively *small* integer B. For example, let $p = 17$. Then

$$17 - 1 = 16 = 2^4.$$

Thus, $17-1$ is a product of small primes, and we can take $B = 6$, as $6!$ is the smallest factorial divisible by $17 - 1 = 16$.

In any case, if $p-1$ divides $B!$ and a is a positive integer relatively prime to p then

$$a^{B!} \equiv 1 \pmod{p}$$

by Fermats' Little Theorem. This fact will ensure that the algorithm works, or at least stops in a small number of steps. A version of this algorithm is as follows:

Pick a small number a such as $a = 2$. If a is not relatively prime to n then we have found a factor of n. Otherwise, starting with $a_1 = a$ use the recursion formula

$$a_i \equiv a_{i-1}^i \pmod{n}$$

to construct a sequence of numbers $a_2, a_3, \ldots$ until

$$d = \gcd(a_i - 1, n) \neq 1.$$

There are two possible outcomes here. If $d \neq n$ then we have found a factor of n. If $d = n$ then the algorithm fails and we need to restart with a different a.

We remind the reader that, for every i, $\gcd(a_i - 1, n)$ is efficiently computed using the Euclidean algorithm. The first issue with this algorithm, as well as with any other, is whether it stops in finitely many steps. To see this, let p be a prime factor of n and B a positive integer such that $p-1$ divides B. We claim that the algorithm stops in less than B steps. Indeed, note that a_i is obtained by first squaring a, then cubing the square, and so on. Therefore a_i can be expressed by

$$a_i \equiv a^{i!} \pmod{n}.$$

Since $p-1$ divides $B!$, by Fermat's Little Theorem,

$$a_B \equiv a^{B!} \equiv 1 \pmod{p}.$$

This shows that p divides $\gcd(a_B - 1, n)$ and the algorithm stops in less than B steps. In particular, if B is small then the algorithm produces a non-trivial $\gcd(a_i - 1, n)$ in a small number of steps. Note that B could be small only if $p-1$ is not divisible by a large prime.

Of course, for a given a, it is possible that the algorithm stops after a small number of steps but it fails to factor n if the first non-trivial greatest common divisor of $a_i - 1$ and n is n. In order to understand better this issue, let us work out an example.

Example: We apply the algorithm to $n = 901$. We start with and $a = 2$. The first step. We compute

$$a_2 \equiv 2^2 \equiv 4 \pmod{901}.$$

Next, we compute the greatest common divisor of $a_2 - 1 = 3$ and 901. This is done using the Euclidean algorithm. Since

$$\begin{array}{rcl} 901 & = & 300 \cdot 3 + 1 \\ 3 & = & 3 \cdot 1 + 0 \end{array}$$

the greatest common divisor of 3 and 301 is 1. We go to the second step. We compute

$$a_3 \equiv 4^3 \equiv 64 \pmod{901}.$$

Since

$$\begin{array}{rcl} 901 & = & 14 \cdot 63 + 1 \\ 63 & = & 4 \cdot 19 + 7 \\ 19 & = & 2 \cdot 7 + 5 \\ 7 & = & 1 \cdot 5 + 2 \\ 5 & = & 2 \cdot 2 + 1 \\ 2 & = & 2 \cdot 1 + 0 \end{array}$$

the greatest common divisor of $a_3 - 1 = 63$ and 901 is 1. We go to the third step. We compute

$$a_4 \equiv 64^4 \equiv 596 \pmod{901}.$$

Since

$$\begin{array}{rcl} 901 & = & 1 \cdot 595 + 306 \\ 595 & = & 1 \cdot 306 + 289 \\ 306 & = & 1 \cdot 289 + 17 \\ 289 & = & 17 \cdot 17 + 0 \end{array}$$

the greatest common divisor of $a_4 - 1 = 595$ and 901 is 17. This produces a factorization

$$901 = 17 \cdot 53.$$

In fact, since for $p = 17$ we can take $B = 6$, the algorithm would have surely terminated in less then 6 steps. But why does the algorithm produces a proper factor of 901? The answer has to do with the multiplicative group

$$(\mathbb{Z}/53\mathbb{Z})^\times.$$

The order of this group is $52 = 53 - 1 = 4 \cdot 13$. In particular, by the theorem of Lagrange, the possible orders of elements in this group are divisors of 52:

$$1, 2, 4, 13, 26 \text{ and } 52.$$

If, by luck, a is picked so that the order of a is divisible by 13 then, since 13 does not divide 6!,

$$a^{6!} \not\equiv 1 \pmod{53}.$$

This and

$$a^{6!} \equiv 1 \pmod{17}$$

imply that $\gcd(a^{6!}-1, 901)$ is precisely 17. This is exactly what happened in our case since the order of 2 module 53 is dividible by 13.

This argument, of course, can be generalized provided that $q-1$ - the other factor of n - has a prime factor larger then B. We record, as a proposition, what we have just discovered:

Proposition 46. *Let $n = p \cdot q$ be a composite number. Let B be a positive integer such that $p-1$ divides $B!$. Assume that there exists a prime factor P of $q-1$ such that*

$$P > B.$$

Let a be a positive integer whose order, as an element of the group $(\mathbb{Z}/q\mathbb{Z})^\times$, is divisible by P. Then the $p-1$ algorithm, started with this a, will compute the prime factor p of n in at most B steps.

Of course, the success in factoring still depends on the choice of a. However, as we shall see in a moment, a random choice of a will almost certainly be the right choice. Consider, first, our case $q = 53$. Then $q - 1 = 4 \cdot 13$. If the order of a randomly picked integer a modulo 53 is divisible by 13, then the algorithm will produce the factor 17, certainly. Thus a bad choice of a happens only if the order of a divides 4. This means that a is a fourth root of 1. Since $\mathbb{Z}/53\mathbb{Z}$ is a field, the number of fourth roots of 1 in less than or equal to 4. Thus, the chance of picking a non-zero integer modulo 53 whose order is not divisible by 13 is less than or equal to

$$\frac{4}{52} = \frac{1}{13}$$

which is very small. In general, let P be a (large) prime dividing $q-1$ and write $q-1 = mP^e$ where P^e is the maximal power of P dividing $q-1$. If the order d of a non-zero integer a modulo q is not divisible by P then d divides m and a satisfies the congruence

$$a^m \equiv 1 \pmod{q}.$$

This shows that a is an m-th root of 1. Since $\mathbb{Z}/q\mathbb{Z}$ is a field, the number of m-th rots of 1 is less then or equal to m. Thus the chance of picking a non-zero integer modulo q whose order is not divisible by P - the algorithm may fail only for such a - is less than or equal to

$$\frac{m}{q-1} = \frac{1}{P^e}.$$

Of course, if one choice of a fails to produce the prime p, we simply pick a different a. The chance of not finding out the prime factor p with d different choices of a is less then

$$\left(\frac{1}{P^e}\right)^d$$

which quickly becomes very small, as d increases.

Exercises

1) Use the $p-1$ method to factor
   a) 9991.
   b) 30227.

3. $p+1$ method

Again, our task it to factor a composite number n. The $p+1$ method is similar to the $p-1$ method and it works well if $p+1$ is composed of *small* primes, as the name suggests. This method was introduced by H. C. Williams in 1982. The main tool is the circle group

$$T(p)$$

of norm one elements in $\mathbb{F}_{p^2}^\times$. This group has the order $p+1$. In a nutshell, the role of the group $(\mathbb{Z}/p\mathbb{Z})^\times$ in the $p-1$ method is played by the group $T(p)$ in the $p+1$ method.

Recall that the finite field $\mathbb{F}_{p^2}$ can be realized as follows. Pick an integer d which is not a square modulo p. Then the finite field $\mathbb{F}_{p^2}$ can be realized as the set of numbers

$$z = x + y\sqrt{d}$$

where x and y are integers considered modulo p. In analogy with complex numbers we shall call y the imaginary part of z and write

$$y = \Im(z).$$

Recall that the circle group $T(p)$ is the subgroup of $\mathbb{F}_{p^2}^\times$ consisting of all z such that

$$z \cdot \bar{z} = (x + y\sqrt{d})(x - y\sqrt{d}) = 1.$$

The following lemma is the key. It allows us to formulate a user-friendly version of the $p+1$ test, that is, a version that avoids an explicit use of the group $T(p)$.

Lemma 47. *Let p be an odd prime. Let d be an integer which is not a square modulo p. Let $z = x + y\sqrt{d} \neq 0$ where x and y are two integers. Let B be a positive integer such that $p+1$ divides $B!$. Then*

$$\Im(z^{B!}) \equiv 0 \pmod{p}.$$

Proof. Consider z as an element of the field $\mathbb{F}_{p^2}$. Let

$$a = \frac{z}{\bar{z}}.$$

Since

$$a \cdot \bar{a} = \left(\frac{z}{\bar{z}}\right)\left(\frac{\bar{z}}{z}\right) = 1$$

a is contained in $T(p)$. Since the order of $T(p)$ is $p+1$, and $p+1$ divides $B!$, the theorem of Lagrange implies that

$$a^{B!} \equiv 1 \pmod{p}.$$

Since $a = z/\bar{z}$, after multiplying both sides of the congruence by $\bar{z}^{B!}$, we obtain

$$z^{B!} \equiv \bar{z}^{B!} \pmod{p}.$$

This implies that the imaginary part of $z^{B!}$ is 0, as desired. The lemma is proved. □

Assume now that we want to factor a composite integer n. The following is a version of the $p+1$ test.

Pick a quadratic integer $z = x + y\sqrt{d}$. If $z\bar{z}$ is not prime to n, then we have found a factor of n. Otherwise, starting with $z_1 = z$, use the recursion formula

$$z_i \equiv z_{i-1}^i \pmod{n}$$

to construct a sequence of numbers $z_2, z_3, \ldots$. The algorithm stops for the first i such that

$$D = \gcd(\Im(z_i), n) \neq 1.$$

There are two possible outcomes here. If $D \neq n$ then we have found a factor of n. If $D = n$ then the algorithm fails and we need to restart with a different z.

The algorithm, regardless of success or failure, will stop in finitely many steps. Indeed, note that

$$z_i \equiv z^{i!} \pmod{n}.$$

Now let p be a prime factor of n and B a positive integer such that $p+1$ divides $B!$ Then, as we showed in the lemma above,

$$\Im(z_B) \equiv \Im(z^{B!}) \equiv 0 \pmod{p}.$$

This shows that p divides $\gcd(\Im(z_B), n)$ and the algorithm stops in less than B steps. Of course - if the algorithm is to be fast - the number B should be small, which is possible only if $p+1$ is a product of small primes.

A small drawback of this algorithm comes from the fact that the construction of the finite field $\mathbb{F}_{p^2}$ is based on the integer d which must not be a square modulo p. Since, a priori, we do not know the factorization $n = pq$, there is no way to know whether a randomly picked d is a square modulo p. However, we have a good chance (50%) that it is. Thus, if one

d does not work, we can try another. Take, for example, $d = -1$. Then $x + y\sqrt{-1} = x + yi$ is a Gaussian integer. Recall that -1 is not a square modulo p if $p \equiv 3 \pmod 4$. In particular, we can use Gaussian integers in the above algorithm to factor out primes p such that

(1) $p \equiv 3 \pmod 4$.
(2) $p + 1$ is a product of small primes.

Example: Let $n = 667$. We apply the $p+1$ method using $d = -1$ (Gaussian integers) and $z = 1 + 2i$. The first step. We calculate

$$z_2 \equiv (1 + 2i)^2 \equiv -3 + 4i \pmod{667}.$$

Since

$$\begin{aligned} 667 &= 161 \cdot 4 + 3 \\ 4 &= 1 \cdot 3 + 1 \\ 3 &= 3 \cdot 1 + 0 \end{aligned}$$

the greatest common divisor of 4 and 667 is 1. The second step. We calculate

$$z_3 \equiv (-3 + 4i)^3 \equiv 117 + 44i \pmod{667}.$$

Since

$$\begin{aligned} 667 &= 15 \cdot 44 + 7 \\ 44 &= 6 \cdot 7 + 2 \\ 7 &= 3 \cdot 2 + 1 \\ 2 &= 2 \cdot 1 + 0 \end{aligned}$$

the greatest common divisor of 44 and 667 is 1. The third step. We calculate

$$z_4 \equiv (117 + 44i)^4 \equiv 5 - 506i \pmod{667}.$$

Since

$$\begin{aligned} 667 &= 1 \cdot 506 + 161 \\ 506 &= 3 \cdot 161 + 23 \\ 161 &= 7 \cdot 23 + 0 \end{aligned}$$

the greatest common divisor of 506 and 667 is 23. In particular, we have obtained a factorization

$$667 = 23 \cdot 29.$$

Note that the algorithm found the factor $p = 23 \equiv 3 \pmod 4$ very quickly. This is because

$$p + 1 = 24 = 2^3 \cdot 3$$

which divides 4!. In particular, the algorithm for $n = 667$ is guaranteed to terminate in less than four steps for any initial value of z.

Exercises

1) Factor 5251 using the $p + 1$ method. Use
a) $z = 1 + 2i$
b) $z = 2 + i$

1) Factor 3953 using the $p + 1$ method. Use

a) $z = 1 + 2i$
b) $z = 2 + i$

4. Quadratic sieve

Quadratic sieve is a modern factorization method. It is based on the observation that for a composite number n we can have two positive integers x and y such that

$$x^2 \equiv y^2 \pmod{n}$$

yet

$$x \not\equiv \pm y \pmod{n}.$$

For example, 12 divides $10^2 - 4^2 = (10+4)(10-4)$, yet 12 does not divide either of the two factors $10 + 4 = 14$ and $10 - 4 = 6$. In particular, $\gcd(12, 14) = 2$ and $\gcd(12, 6) = 6$ are factors of 12. More generally, if we have been lucky to find such x and y for n, then $\gcd(n, x+y)$ and $\gcd(n, x-y)$ will be proper factors of n. Of course, the question is how to construct x and y such that $x^2 \equiv y^2 \pmod{n}$. This is what the quadratic sieve method is about.

A variant of this method goes as follows. Let $[\sqrt{n}]$ be the greatest integer lees than or equal to $\sqrt{n}$. The algorithm makes use of the integers $x_i = [\sqrt{n}] + i$, $i = 1, 2, \ldots$ for (relatively) small values of i. For every such i, the square x_i is greater than n (but not much greater) and when we reduce modulo n,

$$x_i^2 \equiv y_i \pmod{n},$$

then y_i should not be too large. Of course, if y_i is a square, then we have succeeded. However, even if y_i are not squares, it may still be possible to factor n as it will be explained in a moment.

Consider, for example, $n = 5917$. Then $[\sqrt{n}] = 76$. We compute y_i for the first several x_i. We tabulate the outcomes so that the maximal square dividing y_i is factored out:

$$\begin{array}{rcrl}
77^2 & \equiv & 3 \cdot 2^2 & \pmod{5917} \\
78^2 & \equiv & 167 & \pmod{5917} \\
79^2 & \equiv & 18^2 & \pmod{5917} \\
80^2 & \equiv & 3 \cdot 7 \cdot 23 & \pmod{5917} \\
81^2 & \equiv & 7 \cdot 23 \cdot 2^2 & \pmod{5917} \\
82^2 & \equiv & 3 \cdot 269 & \pmod{5917} \\
83^2 & \equiv & 3 \cdot 18^2 & \pmod{5917}
\end{array}$$

Note that the third congruence is $79^2 \equiv 18^2 \pmod{5917}$. It follows that 5917 divides

$$79^2 - 18^2 = (79 + 18)(79 - 18) = 97 \cdot 61.$$

Since $\gcd(5917, 97) = 97$ and $\gcd(5917, 61) = 61$ we have obtained a factorization

$$5917 = 61 \cdot 97.$$

However, there is another important point here. Even if we do not use the congruence $79^2 \equiv 18^2 \pmod{5917}$, we can still factor 5917. More precisely, note that the factor 3 in the first and the last congruence is responsible that $y_1 = 3 \cdot 2^2$ and $y_7 = 3 \cdot 18^2$ are not squares. Thus, if we multiply the first and the last congruence, then

$$(77 \cdot 83)^2 \equiv (3 \cdot 2 \cdot 18)^2 \pmod{5917}$$

which gives, again, a congruence $x^2 \equiv y^2 \pmod{n}$ with $x = 77 \cdot 83 = 6391$ and $y = 3 \cdot 2 \cdot 18 = 108$. It follows that 5917 divides

$$(6391 - 108)(6391 + 108) = 6283 \cdot 6499.$$

Since 5917 is too big to divide either of the two factors we have definitely succeeded in factoring 5917. Indeed, a short calculation using the Euclidean algorithm yields

$$\gcd(5917, 6283) = 61 \text{ and } \gcd(5917, 6499) = 97$$

and $5917 = 61 \cdot 97$.

Summarizing, even if none of the numbers y_i is a square, they are by construction relatively small, so there is a good chance that many of them can be expressed as a product of a square and some small primes. Then some of the congruences may be combined to get a difference of squares divisible by n. This is exactly what happened in the above example. Indeed, if we remove the two equations where relatively large primes 167 and 269 appear, we are left with five congruences

$$\begin{array}{rcrl}
77^2 & \equiv & 3 \cdot 2^2 & \pmod{5917} \\
79^2 & \equiv & 18^2 & \pmod{5917} \\
80^2 & \equiv & 3 \cdot 7 \cdot 23 & \pmod{5917} \\
81^2 & \equiv & 7 \cdot 23 \cdot 2^2 & \pmod{5917} \\
83^2 & \equiv & 3 \cdot 18^2 & \pmod{5917}
\end{array}$$

where only three primes 3, 7 and 23 are making up the square free part of each y_i. This fact, that there are more equations (five) than “obstructing” primes (three), guarantees that the congruences can be combined so that the product of corresponding y_i is a square.

The process can be formalized, using a bit of linear algebra, as follows. First of all, one picks a number B which depends on n and then we consider congruences $x_i^2 \equiv y_i \pmod{n}$ such that y_i can be expressed, up to a square, in terms of primes less than B. In practice there are recommendations what B should be (perhaps NSA knows). Asume, for example, that we have picked $B = 40$ for our $n = 5917$. This choice of B then eliminates the two equations involving primes 167 and 269. The remaining congruences can be tabulated as follows

x_i	3	7	23
77	1	0	0
79	0	0	0
80	1	1	1
81	0	1	1
83	1	0	0

This table is essentially a matrix, in this case a 5×3-matrix, with coefficients in the finite field $\mathbb{Z}/2\mathbb{Z}$. The rows are parameterized by x_i and columns by primes p less than B. The value of an entry, parameterized by the pair (x_i, p), is equal to the exponent of p as it appears in the square free part of y_i. Now, as soon as we have calculated enough y_i so that the number of columns is greater then the number of rows, we surely are going to have some linear relations between rows. In our example, if the rows are denoted by $v_1, v_2, \ldots$, we have the following relation (congruence):

$$v_1 + v_5 \equiv (0,0,0) \pmod 2.$$

Since the rows v_1 and v_5 correspond to $x_1 = 77$ and $x_7 = 83$, respectively, this congruence implies that $y_1 \cdot y_7$ is a square. We have used this already to factor 5917. Note, however, that there is another relation:

$$v_1 + v_3 + v_4 \equiv (0,0,0) \pmod 2.$$

Since the rows v_1, v_3 and v_4 correspond to $x_1 = 77$, $x_4 = 80$ and $x_5 = 81$, respectively, it follows that $y_1 \cdot y_3 \cdot y_4$ is a square and multiplying the congrunces $x_i^2 \equiv y_i \pmod{5917}$ for $i = 1, 3$ and 4 gives:

$$(77 \cdot 80 \cdot 81)^2 \equiv (3 \cdot 7 \cdot 23 \cdot 4)^2 \pmod{5917}.$$

Since $77 \cdot 80 \cdot 81 = 498960$ and $3 \cdot 7 \cdot 23 \cdot 4 = 1932$, it follows that 5917 divides

$$(498960 - 1932)(498960 + 1932) = 497028 \cdot 500892.$$

This, unfortunately, does not give a factorization of 5917 since 5917 divides the first factor and is relatively prime to the second. This example shows that a dependence relation between the rows of the matrix does not necessarily lead to a factorization of n.

Exercises

1) Factor 3837523 using

$$1964^2 \equiv 3^2 \cdot 13^3 \pmod{3837523}$$

and

$$14262^2 \equiv 5^2 \cdot 7^2 \cdot 13 \pmod{3837523}.$$

2) The quadratic sieve method finds quickly factors of the following three numbers, without combining the congruences. Find the factors.

a) $n = 10057$

b) $n = 26123$
c) $n = 64777$
d) $n = 17557$

3) Compute 255^2 and 317^2 modulo 64777. Combine these congruences to factor 64777.

4) Use the quadratic sieve method to factor 7097.

5) The composite number $n = 30227$ with the encryption exponent $e = 7$ is used to produce - using the RSA cypher - a secret message:

$$4110.$$

Use the quadratic sieve method to factor 30227. Use the factorization to compute the decryption exponent d and to decipher the message. Express the final answer using the replacement table:

A	O	H	L	M	N	K
2	3	4	5	6	7	8

CHAPTER 12

Elliptic curves

1. Cubic curves

In this chapter we shall study elliptic curves. The main interest in elliptic curves lies in the fact that the set of points of an elliptic curve is a group! For our purposes, an elliptic curve over a field F will be the set of solutions of a cubic equation of the type

$$y^2 = x^3 + ax^2 + bx + c,$$

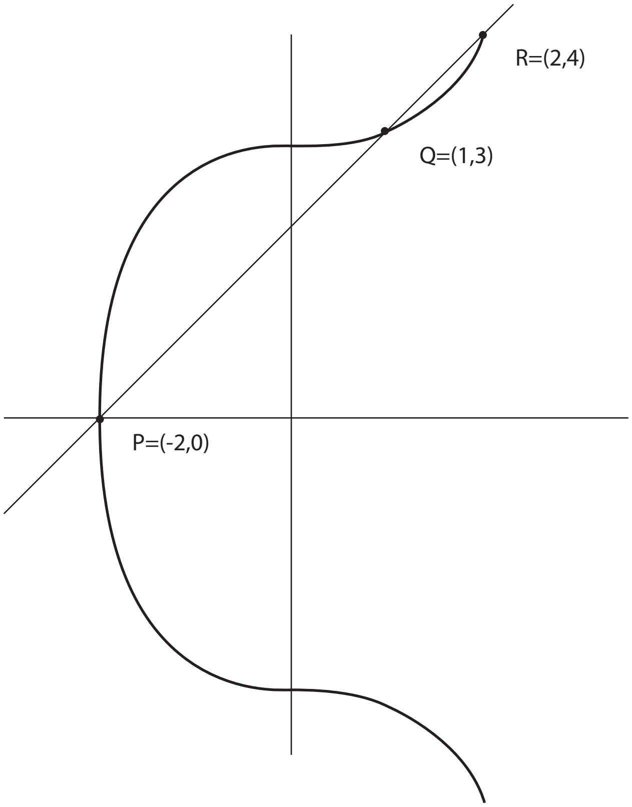
where a, b and c are in F.

Our first task is to get some understanding of elliptic curves by considering them over real numbers. In that case, the above equation cuts out a honest curve in the x, y-plane, the shape of which can be understood using calculus. Consider, for example the equation $y^2 = x^3 + 8$. The following five points on the curve are easily found:

x	y
-2	0
1	3
1	-3
2	4
2	-4

The graph given below is obtained by first graphing the five points, and then interpolating between these points. The interpolation - as given in the picture - can be justified by the following:

(1) Since $x^3 + 8 = y^2 \geq 0$, it follows that $x^3 \geq -8$ and $x \geq -2$. Thus the curve is situated right of $x = -2$. If $x = -2$ then $y = 0$ is the only possible value for y. If $x > -2$ then we have two possible values for y. We can say that the curve branches out of the point $P = (-2, 0)$ point. There is a a "positive" branch, consisting of points with $y > 0$, and a "negative" branch consisting of points with $y < 0$.



(2) The implicit differentiation gives

$$\frac{dy}{dx} = \frac{3x^2}{2y}.$$

Note that $\frac{dy}{dx} \geq 0$ if $y > 0$ and $\frac{dy}{dx} \leq 0$ if $y < 0$. It follows that the positive branch is increasing, while the negative branch is decreasing. Each branch has a unique inflection point at $x = 0$.

(3) For large x, the curve is similar to $y^2 = x^3$.

The group law is based on the fact that any line - well, almost any - intersects the curve $y^2 = x^3 + ax^2 + bx + c$ in 3 points. More precisely, let $P = (x_P, y_P)$ and $Q = (x_Q, y_Q)$ be two points on the curve $y^2 = x^3 + bx + c$.

The equation of the line through P and Q is

$$y = \frac{y_Q - y_P}{x_Q - y_P}(x - x_P) + y_P = Ax + B.$$

The intersection points of the line and the elliptic curve are computed by substituting $y = Ax + B$ in the cubic equation, which gives

$$(Ax + B)^2 = x^3 + ax^2 + bx + c$$

and, simplified,

$$x^3 + (a - A^2)x^2 + (b - 2A)x + c + B^2 = 0.$$

This is a cubic polynomial. As such, it has three roots. Two of the roots are known - x_P and x_Q - the coordinates of P and Q. The third root x_R is the x-coordinate of the third point $R = (x_R, y_R)$ of the intersection of the cubic curve with the line. The number x_R can be easily computed as follows. Recall that the sum of zeros of a cubic polynomial $x^3 - \alpha x^2 + \ldots$ is equal to the coefficient α. In our case $\alpha = A^2 - a$, thus $x_P + x_Q + x_R = A^2 - a$ which gives

$$x_R = -x_P - x_Q + A^2 - a.$$

Finally, the coordinate y_R of R is easily figured out since it is the (unique) point on the line through P and Q with x-coordinate equal to x_R. Thus,

$$y_R = A(x_R - x_P) + y_P.$$

Take, for example, $P = (-2, 0)$ and $Q = (1, 3)$ on the curve $y^2 = x^3 + 8$. The line through P and Q has the slope $A = 1$. Since $a = 0$, we have

$$x_R = -(-2) - 1 + 1^2 = 2 \text{ and } y_R = 1(2 - (-2)) + 0 = 4.$$

It follows that the line through P and Q intersects the curve $y^2 = x^3 + 8$ in the third point $R = (2, 4)$, as pictured.

This works fine except when the points P and Q have the same x-coordinate, $x_P = x_Q = C$ in which case the line through two points is vertical, given by an equation $x = C$. When combined with the equation $y^2 = x^3 + ax^2 + bx + c$ we get a quadratic, not cubic, equation

$$y^2 = C^3 + aC^2 + bC + c.$$

This quadratic equation has two solutions, the y-coordinates of P and Q. To overcome this imperfection, mathematicians have added points "at infinity" to the plane and build the so-called *projective* plane.

In order to explain the basic idea of projective geometry let's take a step back and consider a simpler problem of intersecting two lines in a plane. As you well know, two different lines in the plane either intersect in a point or are parallel. The projective plane is introduced precisely to overcome this imperfection. The projective plane is obtained by adding points to the usual plane, as follows. For every class of parallel lines in the plane you ad a point "at infinity" where all these parallel lines intersect. This is not too difficult to imagine. Indeed, if you stand in the middle of a long, straight road, two curbs of the road appear to meet in the distance. So you add that point “at infinity” to the plane. You also identify this point with the point where the curbs of the road “meet” as you turn about 180 degrees. (You do this so that the two parallel lines, given by the two road curbs, intersect in precisely one point.) Anyway, once you have added points at infinity, one for each class of parallel lines, you have built the projective plane. In the projective plane any two different lines intersect in precisely one point and, not surprisingly, a line will intersect a cubic curve always in three points. For example, the elliptic curve $y^2 = x^3 + ax^2 + bx + c$, when considered sitting in the projective plane, consists of points in the usual plane (solutions of $y^2 = x^3 + ax^2 + bx + c$) plus one point O at infinity, given as the intersection of all vertical lines. With this modification, a vertical line drawn through any two points of the cubic curve $y^2 = x^3 + ax^2 + bx + c$ will intersect the cubic curve in O as well.

We are now ready to introduce a group law an a cubic $y^2 = x^3 + ax^2 + bx + c$. The elements of the group are the points on the curve, including the point O. The group operation - addition of points - is denoted by $+$ and is specified by the following axioms:

(1) The point at infinity O is the identity element (zero).
(2) If a line intersects the curve in three points P Q and R then,
$$P + Q + R = O.$$
(3) The inverse of a point $P = (x, y)$ is $-P = (x, -y)$ - only the y-coordinate changes the sign.

For example, let $P = (-2, 0)$ and $Q = (1, 3)$ be the two points on the curve $y^2 = x^3 + 8$. The sum $P + Q$ is computed as follows. You draw the line through P and Q. It intersects the cubic curve in the third point $R = (2, 4)$, as we have shown before. Then
$$P + Q = -R = (2, -4).$$

Our definition, however, contains a gap. If $P = Q$ then there is no *unique* line through P and Q. In this case we take the tangent line at P in order to compute $2P$. This can be explained as follows. Assume that $P \neq Q$. If we move Q a little bit, then $P + Q$ also moves a little bit. Thus, in order to compute $2P$, it is natural to take Q close to P and then $P + Q$ approximates $2P$. As we take Q closer and closer to P, the line through P and Q will look more and more as the tangent line to the curve at P. Since the slope of the tangent line at P is
$$\lim_{Q \to P} \frac{y_Q - y_P}{x_Q - x_P} = \frac{dy}{dx}(P),$$
the x-coordinate of $R = -2P$ is given by the *doubling* formula
$$x_R = -2x_P + \left(\frac{dy}{dx}(P)\right)^2 - a.$$

Note that $dy/dx = \infty$ can happen. This happens for $P = (-2, 0)$ on the curve $y^2 = x^3 + 8$. The tangent line at P is a vertical line $x = -2$ which intersects the curve in P and in the point at infinity O. Thus, in this case, $2P = O$ and the point P has the order 2 in the group.

Of course, the formula for doubling is well defined only if the derivative dy/dx exists. Geometrically, this simply means that it is possible to draw the tangent line at every point of the cubic curve. If that is the case, the set of all points together with the point at infinity O, forms a commutative group. We state this as a fact without a proof. Otherwise, if there are some points where the derivative dy/dx does not exist, then the curve is called *singular* or *degenerate*. Degenerate curves are interesting in its own right and will be discussed in details in the next lecture.

Exercises

1) Complete the addition table for the points $P = (-1, 0)$, $Q = (0, 1)$, $-Q = (0, -1)$, $R = (2, 3)$, $-R = (2, -3)$, and the point at infinity O on the elliptic curve $y^2 = x^3 + 1$. (This shows that these six points form a subgroup of the elliptic curve group.)

$+$	O	P	Q	$-Q$	R	$-R$
O	O	P	Q	$-Q$	R	$-R$
P	P	O				
Q	Q			O		
$-Q$	$-Q$		O			
R	R					O
$-R$	$-R$				O	

2) Let $P = (1, 3)$ be a point on the elliptic curve $y^2 = x^3 + 8$. Compute $2P$, $4P$ and $8P$. (You should be able to compute $2P$ and $4P$ by hand. Some software application might be needed for $8P$.)

2. Degenerate curves

Doubling a point P on a cubic curve $y^2 = f(x)$ requires that the curve has a tangent line at P or, equivalently, that the derivative dy/dx exist at the point P. We allow $dy/dx = \infty$, which happens when the tangent line is vertical. Our task here is to identify points on the curve where dy/dx is not well defined. To this end, implicit differentiation of $y^2 = f(x)$ gives

$$\frac{dy}{dx} = \frac{f'(x)}{2y} = \pm\frac{f'(x)}{2\sqrt{f(x)}}.$$

In particular, dy/dx exists unless the last fraction is equal to $0/0$. This happens if and only if $x = \alpha$ is a double (or a triple) root of the cubic polynomial $f(x)$. If that is the case, then the point

$$S = (\alpha, 0)$$

on the curve $y^2 = f(x)$ is called a singular point and the curve is called a degenerate or singular curve. One can easily determine if the curve $y^2 = f(x)$ is singular as follows. Let x_1, x_2 and x_3 be the three roots of $f(x)$. Then the discriminant of the polynomial $f(x)$ is the number

$$\Delta = [(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)]^2.$$

Usefulness of this number lies in the fact that it can be described in terms of coefficients of the polynomial $f(x) = x^3 + ax^2 + bx + c$:

$$\Delta = a^2b^2 - 4a^3c - 4b^3 - 27c^2 + 18abc.$$

In particular, given a cubic curve $y^2 = f(x)$ one can easily check whether the curve is degenerate or not, by calculating the discriminant Δ. Singular cubic curves are not considered to be elliptic curves. Still, the the addition law - as defined by intersecting the cubic by lines - gives a group structure on the cubic curve with the singular point S removed.

We are now ready to describe the degenerate curves over real numbers in some detail. Let α be a double root of $f(x)$. By translating $x := x - \alpha$ we can assume, without loss of generality, that $\alpha = 0$. In particular, the equation of the curve is

$$y^2 = x^3 + dx^2.$$

Solutions of this equation can be easily determined by substituting $y = tx$. With this substitution the equation $y^2 = x^3 + dx^2$ can be rewritten as

$$(d - t^2)x^2 = x^3.$$

Now notice that the point $S = (0, 0)$ is the only point on the curve with $x = 0$. If $x \neq 0$ then we can divide both sides of $(d - t^2)x^2 = x^3$ by x^2. This gives parametric equations

$$\begin{cases} x = t^2 - d \\ y = tx = t^3 - td. \end{cases}$$

Thus any point on the curve - except perhaps S - is given by a picking a value for t and then calculating the coordinates x and y using these formulas.

Consider now the case when $d = 0$. As we shall verify in a moment the group law here amounts simply to adding ratios x/y. That is, if $P_1 = (x_1, y_1)$, $P_2 = (x_2, y_2)$, and $P_3 = (x_3, y_3)$ are three points on the intersection of a line and the curve $y^2 = x^3$ then

$$\frac{x_1}{y_1} + \frac{x_2}{y_2} + \frac{x_3}{y_3} = 0.$$

In order to verify this we shall use the parameterization $x = t^2$ and $y = t^3$ of the curve $y^2 = x^3$. Let t_1 and t_2 be non-zero numbers such that $P_1 = (t_1^2, t_1^3)$ and $P_2 = (t_2^2, t_2^3)$. The line through the points P_1 and P_2 has the slope

$$A = \frac{t_1^3 - t_2^3}{t_1^2 - t_2^2} = \frac{t_1^2 + t_1t_2 + t_2^2}{t_1 + t_2}.$$

Now, a short calculation shows that the coordinates x_3 and y_3 of the third point P_3 on the intersection of the line and the curve are

$$x_3 = -x_1 - x_2 + A^2 = \left(\frac{t_1t_2}{t_1 + t_2}\right)^2$$

and

$$y_3 = y_1 + A(x_3 - x_1) = -\left(\frac{t_1t_2}{t_1 + t_2}\right)^3.$$

Thus, the fraction x_3/y_3 is equal to

$$\frac{x_3}{y_3} = -\frac{t_1 + t_2}{t_1t_2} = -\frac{1}{t_1} - \frac{1}{t_2} = -\frac{x_1}{y_1} - \frac{x_2}{y_2}$$

which is exactly what we wanted to prove. Note that, as $t \to \infty$, then

$$\frac{x}{y} = \frac{1}{t} \to 0$$

so the point O “at infinity” naturally corresponds to the number 0. Thus, the points on the cubic curve $y^2 = x^3$ together with the point O, but with the singular point $(0, 0)$ removed, form a group where adding points corresponds to adding quotients x/y. In other words, the map

$$P = (x, y) \mapsto \frac{x}{y}$$

gives an identification of two groups, the group of the cubic curve $y^2 = x^3$ and the group of real numbers with respect to the usual addition. Of course, the curve $y^2 = x^3$ could be considered over any field F and the above conclusions are valid again.

Now we shall study the degenerate curves where $d \neq 0$. If $P = (x, y)$ is point on the curve $y^2 = x^3 + dx^2$ different form S, put

$$h(P) = \frac{y - x\sqrt{d}}{y + x\sqrt{d}},$$

and $h(O) = 1$. It turns out that the group law, in this case, amounts to multiplication of numbers $h(P)$:

$$h(P + Q) = h(P) \cdot h(Q).$$

A verification of this is straightforward, yet somewhat tedious, so it is omitted. A special case $h(2P) = h(P)^2$ is given as an exercise below. Note that $h(P)$ is a real if d is positive and it is a complex number on the unit circle

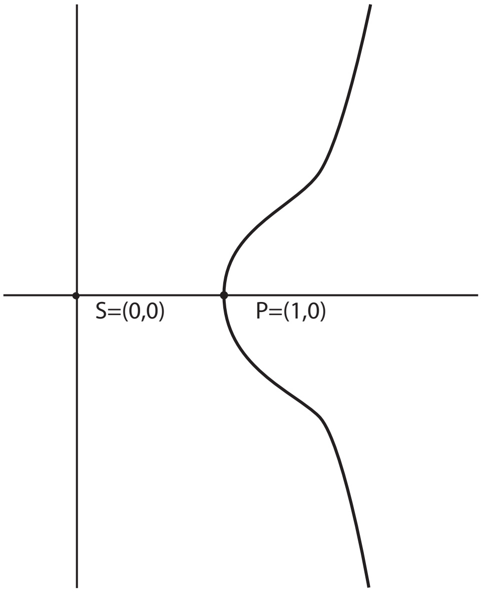


Figure 1

if d is negative. We shall illustrate the group structure by considering two special cases $d = -1$ and $d = 1$.

Consider first $d = -1$. The curve $y^2 = x^3 - x^2$ is pictured in Figure 1. The singular point is $S = (0, 0)$. It is isolated from the rest of the curve and, clearly, there is no tangent line to the curve at the point S. The set of points on the curve, including the point O but not the point S, forms a group. Let $i = \sqrt{-1}$. Then the map

$$h(P) = \frac{y - ix}{y + ix}$$

identifies this group with $\mathbb{T} = \{z \in \mathbb{C} \mid |z| = 1\}$. the group of complex numbers of norm 1. For example, consider the point $Q = (2, 2)$ on the curve $y^2 = x^3 - x^2$. Let us compute $P = 2Q$. The slope of the tangent line at Q is

$$A = \frac{dy}{dx}(Q) = \frac{3x^2 - 2x}{2y}(Q) = \frac{8}{4} = 2.$$

It follows that

$$x_P = -2x_Q + A^2 - a = -2(2) + 2^2 - (-1) = 1,$$

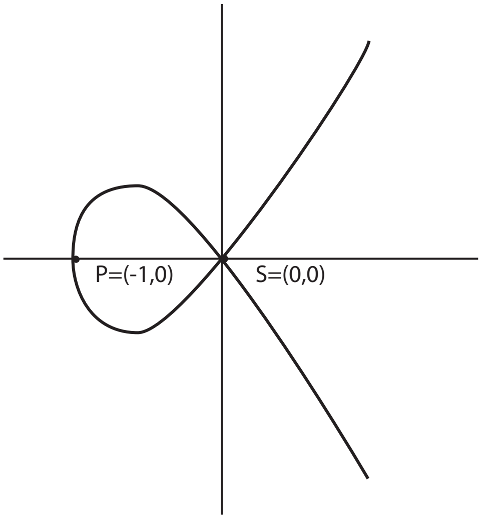


Figure 2

so $P = (1, 0)$, as given in the Figure 1. The order if P is 2, hence Q is of order 4. The map h, on the other hand, gives

$$h(Q) = \frac{2 - 2i}{2 + 2i} = i$$

which is of order 4 as an element of the group $\mathbb{T}$.

Consider now $d = 1$. The curve $y^2 = x^3 + x^2$ is pictured in Figure 2. This curve is self-intersecting at the singular point $S = (0, 0)$. Again, the set of points on the curve, together with the point O but with the point S removed, is a group. This time the map

$$h(P) = \frac{y - x}{y + x}$$

identifies the group of points with the multiplicative group of non-zero real numbers. Note that the expressions $y - x$ and $y + x$ define equations of the two tangent lines at the point S:

$$\begin{cases} y - x = 0 \\ y + x = 0 \end{cases}$$

These two lines do not intersect the cubic curve in any point different from S. This explains why

$$h(P) \neq 0, \infty$$

for any point $P \neq S$ on the curve. In particular, the map h is well defined.

Exercises

1) Let $P = (1, 1)$ be a point on the degenerate cubic curve $y^2 = x^3$. Compute nP for $n = 1, 2, \ldots, 5$. What do you think nP should be? If $nP = (x_n, y_n)$, what is x_n/y_n?

2) Let $P = (x, y)$ be a point on the degenerate curve $y^2 = x^3 + dx^2$. Show that $2P = (X, Y)$ where

$$Y - X\sqrt{d} = (y - x\sqrt{d})^2 = (y^2 - dx^2) - 2xy\sqrt{d}.$$

3. Curves modulo p

In this section we study cubic curves modulo an odd prime p or, in other words, congruences

$$y^2 \equiv x^3 + ax^2 + bx + c \pmod{p},$$

where a, b and c are integers. Consider, for example, the curve $y^2 = x^3 + 8$. Then, as one easily verifies,

$$0^2 \not\equiv 8^3 + 8 \pmod{11}$$

but

$$0^2 \equiv 8^3 + 8 \pmod{13}.$$

This shows that $(8, 0)$ is a not a solution of $y^2 = x^3 + 8$ modulo 11 but it is a solution modulo 13. The solution $(8, 0)$ is also called a point modulo 13 on the cubic curve. Note that $(1, 3)$ is an integral solution of $y^2 = x^3 + 8$ and it automatically gives a solution modulo p for any prime number p.

Points modulo p on a cubic curve can be added using the same group law defined by intersecting the curve with lines. In particular, solutions of the cubic equation modulo p, together with the point at infinity O, form a finite group denoted by $E(p)$. However, just as in the case of real curves, if

$$\Delta \equiv 0 \pmod{p},$$

for the discriminant Δ of the cubic curve E then the curve is degenerate modulo p and singular points need to be removed before to have a group structure. If p does not divide the discriminant, then we say that E has a good reduction modulo p.

The discriminant of $y^2 = x^3 + 8$ is $27 \cdot 8^2$. This shows that we do not need to worry about singular points for primes $p \neq 2, 3$. As the first example of computing on elliptic curves modulo a prime, we shall double the point $P = (1, 3)$ on $y^2 = x^3 + 8$ modulo $p = 11$. Recall that the coordinates of $-2P$ are given by

$$x_{-2P} = -2x_P + A^2 \text{ and } y_{-2P} = y_P + A(x_{-2P} - x_P)$$

where A is the slope of the tangent line at P. The slope is given by

$$A = \frac{dy}{dx}(P) = \frac{3x^2}{2y}(1, 3) = \frac{3}{6}.$$

Of course, in order to make sense of $3/6$ modulo 11 we need to find the multiplicative inverse of 6 modulo 11. Since $2 \cdot 6 - 5 \cdot 2 = 1$, the multiplicative inverse of 6 is 2, therefore $A = 3 \cdot 2 = 6$. We can now easily calculate

$$x_{-2P} = -2 \cdot 1 + 6^2 \equiv 1 \pmod{11}$$

and

$$y_{-2P} = 3 + 6 \cdot (1 - 1) \equiv 3 \pmod{11}.$$

It follows that $-2P = P$ or $3P = O$. Thus, we have shown that the order of P in $E(11)$ is 3.

As the next example, we shall double the point $P = (1, 3)$ modulo $p = 13$. The slope of the tangent at P is given by

$$A = \frac{dy}{dx}(P) = \frac{3x^2}{2y}(1, 3) = \frac{3}{6}.$$

Again, in order to make sense of $3/6$ modulo 13, we need to find the multiplicative inverse of 6 modulo 13. Since $11 \cdot 6 - 5 \cdot 13 = 1$, the multiplicative inverse of 6 is 11, therefore $A = 3 \cdot 11 \equiv 7 \pmod{13}$. We can now easily calculate

$$x_{-2P} = -2 \cdot 1 + 7^2 \equiv 8 \pmod{13}$$

and

$$y_{-2P} = 3 + 7 \cdot (8 - 1) \equiv 0 \pmod{13}.$$

It follows that $2P = (8, 0)$. Doubling further the point $2P = Q$ gives

$$A = \frac{dy}{dx}(Q) = \frac{3x^2}{2y}(8, 0) = \frac{10}{0}$$

which shows that the tangent line at Q is "vertical", and this means that $2Q = O$, the identity point. Thus, we have shown that the order of P in $E(13)$ is 4.

The theory of solutions of a cubic equation modulo p is rather rich. As an illustration we shall compute the number of solutions to $y^2 = x^3 + 8$ modulo 11 and modulo 13. Consider first $p = 11$. The first step is to calculate $f(x) = x^3 + 8$ modulo 11:

x	0	1	2	3	4	5	6	7	8	9	10
x^3	0	1	8	5	9	4	7	2	6	3	10
$f(x)$	8	9	5	2	6	1	4	10	3	0	7

Notice that $f(x)$ takes all possible values modulo 11. If $f(x) = 0$ then $y = 0$. If $f(x) \neq 0$ then $f(x)$ is a square for 5 different values of x and it is a non-square for 5 other values of x. Every time $f(x)$ is a non-zero square then there are two possible choices for y. Adding O, the point at infinity, we have calculated that the order of $E(11)$ is

$$|E(11)| = 1 + 2 \cdot 5 + 1 = 12.$$

A rather different calculation emerges for $p = 13$. Indeed, calculating $f(x) = x^3 + 8$ modulo 13 gives:

x	0	1	2	3	4	5	6	7	8	9	10	11	12
x^3	0	1	8	1	12	8	8	5	5	1	12	5	12
$f(x)$	8	9	3	9	7	3	3	0	0	9	7	0	7

In this case $f(x)$ does not take all possible values modulo 13. However, if $f(x) = 0$ then $y = 0$. This gives us 3 points on the curve modulo 13:

$$(7, 0), (8, 0) \text{ and } (11, 0).$$

If $f(x) \neq 0$, in order to find other solutions of $y^2 = f(x)$, we list all squares modulo 13:

$$1 \equiv (\pm 1)^2, 4 \equiv (\pm 2)^2, 9 \equiv (\pm 3)^2, 3 \equiv (\pm 4)^2, 12 \equiv (\pm 5)^2 \text{ and } 10 \equiv (\pm 6)^2.$$

If $f(x)$ is a non-zero square then there are two solutions for y. Since $f(x) \equiv 3 \pmod{13}$ for $x = 2, 5$ and 6, we have the following six points

$$(2, 4), (5, 4), (6, 4) \text{ and } (2, -4), (5, -4), (6, -4).$$

Since $f(x) \equiv 9 \pmod{13}$ for $x = 1, 3$ and 9, the following additional six points

$$(1, 3), (3, 3), (9, 3) \text{ and } (1, -3), (3, -3), (9, -3).$$

In all, counting the the point O, we find that

$$|E(13)| = 3 + 6 + 6 + 1 = 16.$$

The computation of the order of $E(11)$ can be easily generalized to the curve $y^2 = x^3 + c$ and any prime $p \equiv 2 \pmod{3}$. Note that the discriminant of $y^2 = x^3 + c$ is $-27c^2$. In particular, the curve has a good reduction for every odd prime not dividing $3c$.

Proposition 48. *Let c be a non-zero integer. Let E be the elliptic curve given by the equation $y^2 = x^3 + c$. Assume that p is a prime such that it does not divide c and such that $p \equiv 2 \pmod{3}$. Then*

$$|E(p)| = p + 1.$$

Proof. Let us look at the example of the curve $y^2 = x^3 + 8$ and $p = 11$, as worked out above. The key observation there is that the map $x \mapsto x^3$ is one to one. We claim that the map $x \mapsto x^3$, from $\mathbb{F}_p$ to $\mathbb{F}_p$, is one to one if $p \equiv 2 \pmod{3}$. First of all, $0 \mapsto 0$, so we need to show that the map $x \mapsto x^3$

is one to one when restricted to the multiplicative group $\mathbb{F}_p^\times$. If $x^3 = y^3$ then $(x/y)^3 = 1$. If $p \equiv 2 \pmod 3$ then $\mathbb{F}_p^\times$ has no elements of order 3. This implies that $x/y = 1$ and $x = y$. It follows that $x \mapsto x^3$ is a one to one map, as claimed.

Next, if $x \mapsto x^3$ is one to one, then so is $x \mapsto x^3 + c$. If $x^3 + c = 0$ then $y = 0$. This is one point on the curve. Next, $x^3 + c \neq 0$ is a square for $(p-1)/2$ values of x. For every such x we have two choices for y. Counting the point O, the order of $E(p)$ is

$$|E(p)| = 1 + 2\frac{(p-1)}{2} + 1 = p + 1,$$

as desired. □

Generally, if E is an elliptic curve, it is not easy to determine the order of $E(p)$. However, there is an estimate due to Hasse which says that, if we write the order of $E(p)$ as

$$|E(p)| = p - a_p + 1$$

then

$$|a_p| \leq 2\sqrt{p}.$$

For example, if E is the curve $y^2 = x^3 + 8$, this estimate says that the order of $E(13)$ is between 7 and 21. The actual order is 16, as we calculated. There are several families of elliptic curves for which it is possible to determine the order of $E(p)$ for all primes. One such family is $y^2 = x^3 + c$ and the other is $y^2 = x^3 + bx$. The family $y^2 = x^3 + bx$ is the subject of the next section.

Exercises

1) Let $P = (2,3)$ be a point on the elliptic curve $E := y^2 = x^3 - 10x + 21$ modulo the prime 557.

(1) Show that $189P = O$ while $63P \neq O$ and $27P \neq O$. Explain why this shows that the order of P is 189.

(2) Use the fact that P has the order 189 and Hasse's estimate to determine - precisely - the order of $E(557)$.

(The fastest way to compute mP is by consecutive squaring or doubling.)

2) Let $P = (-1,4)$ be a point on the curve $y^2 = x^3 + 17$. Then $Q = (21,3) \equiv nP \pmod{31}$ for some positive integer n. The number n is called the discrete logarithm of Q with base P, modulo 31. Calculate n using the giant step - baby step method. First, use Hasse's inequality to show that the order of $E(31)$ is less then or equal to 43. Since the root of 43 is less then 7, the number n can be written as $n = i + j7$ for some i and j less then 7. Now n is determined by following two steps:

(1) Compute and list iP modulo 31 for all $i = 1, 2, \ldots 7$.

(2) Compute $Q - j(7P)$ for $j = 1, 2, \dots$ until it is equal to iP for some i, $1 \leq i \leq 6$.

4. The curve $y^2 = x^3 + bx$

The family of elliptic curves $y^2 = x^3 + bx$ is arguably one of the most interesting. The discriminant of $y^2 = x^3 + bx$ is $-4b^3$. Thus $y^2 = x^3 + bx$ has a good reduction modulo an odd prime p as long as p does not divide b. In this section we determine the order of $E(p)$ for primes $p \equiv 3 \pmod 4$. The main tool is the observation that the function

$$f(x) = x^3 + bx$$

is odd, that is, $f(-x) = -f(x)$. Assume that $f(x) \neq 0$. If $p \equiv 3 \pmod 4$ then -1 is not a square. Thus, if $f(x)$ is not a square then $f(-x) = -f(x)$ is a square. On the other hand, if $f(x)$ is a square, then $f(-x) = -f(x)$ is not a square. It follows, if $f(x) \neq 0$, that $f(x)$ is a square for precisely one of the two elements $\{x, -x\}$. This plays a key role in the following proposition.

Proposition 49. *Let b be a non-zero integer. Let E be the elliptic curve given by the equation $y^2 = x^3 + bx$. Assume that p is a prime such that it does not divide b and such that $p \equiv 3 \pmod 4$. Then*

$$|E(p)| = p + 1.$$

Proof. We shall first count the points of order 2. They correspond to the solutions of $f(x) = 0$. Since $f(x) = x(x^2 + b)$ we see that $x = 0$ is one root of $f(x)$. We have two different cases:

Case 1: $-b$ is not a square modulo p. Then $f(x) = 0$ only for $x = 0$. In particular, $(0, 0)$ is the unique point of order 2. If $x \neq 0$ then $f(x) \neq 0$ and the equation $y^2 = f(x)$ has 2 or 0 solutions, depending whether $f(x)$ is a square or not. Furthermore, as we argued above, $f(x)$ is a square for precisely one of the two elements $\{x, -x\}$. Since we have $(p - 1)/2$ pairs $\{x, -x\}$ in all, and each pair contributes two points in $E(p)$, we can conclude that

$$|E(p)| = 2 + 2 \cdot \frac{p - 1}{2} = p + 1$$

where the first summand 2 accounts for $(0, 0)$ and the identity O.

Case 2: $-b$ is a square modulo p. Let $-b = s^2$. In this case $f(x) = 0$ for three values $x = 0, s$ and $-s$. If $f(x) \neq 0$ then, arguing in the same way as above, the pair $\{x, -x\}$ contributes two points on $E(p)$. Since the pair $\{s, -s\}$ also contributes two points on $E(p)$ (the points $(0, -s)$ and $(0, -s)$) the total number of points is again $p + 1$. □

It is possible to determine the order of $E(p)$ for the curve $y^2 = x^3 + bx$ even for $p \equiv 1 \pmod 4$. The order of $E(p)$ is, quite remarkably, related to expressing the prime p as a sum of two squares. Assume, for simplicity, that $b = 1$. Recall that an odd prime can be written as a sum of two squares

$$p = \alpha^2 + \beta^2$$

if and only if $p \equiv 1 \pmod 4$. Since p is odd, the numbers α and β must have different parity. Without any loss, we can assume that α is odd. Then the number of points on the curve $y^2 = x^3 + x$ modulo p is

$$|E(p)| = p \pm 2\alpha + 1$$

where the sign in front of 2α is negative if $\alpha \equiv 1 \pmod 4$, and positive if $\alpha \equiv 3 \pmod 4$. As an example, consider $p = 173$. Since $173 = 13^2 + 2^2$ and $13 \equiv 1 \pmod 4$ the number of points modulo 173 is

$$|E(173)| = 173 - 2 \cdot 13 + 1 = 148.$$

It is interesting to note that this number is on the lower end of the Hasse estimate for $p = 173$:

$$148 \leq |E(173)| \leq 200.$$

The doubling formula takes an interesting shape for the curve $y^2 = x^3 + bx$. Recall that

$$x_{-2P} = -2x_P + \frac{(3x_P^2 + b)^2}{4y_P^2}.$$

A simple calculation, using $4y_P^2$ as a common denominator, shows that the right hand side can be rewritten as a pure square:

$$x_{-2P} = \frac{(x_P^2 - b)^2}{4y_P^2}.$$

In particular, if $R = 2P$, then x_R is a square. This observation can be used to show that a point on the curve $y^2 = x^3 - bx$ is not obtained by doubling another point. Consider, for example, the point $R = (3, 3)$ on the curve $y^2 = x^3 - 6x$. Then, by the quadratic reciprocity, 3 is not a square modulo 31 and, therefore, the point R is not a double of another point on the curve modulo 31.

Exercises

1) Let $F_n = 2^{2^n} + 1$ be a Fermat prime. Show that the number of points on the curve $y^2 = x^3 + x$ modulo F_n is 2^{2^n}.

2) Show that the point $R = (-2, 4)$ on the curve $y^2 = x^3 - 12x$ is not obtained by doubling another point modulo 31.

CHAPTER 13

Factoring and testing using elliptic curves

1. Lenstra's factoring method

Factoring integers using elliptic curves was introduced by Lenstra in 1987. This method can be viewed as a vast generalization of $p-1$ and $p+1$ methods. In order to explain the idea, consider the point $P=(1,3)$ on the elliptic curve $y^2=x^3+8$. Working modulo $n=533$, we shall compute

$$2P,\ 4P,\ 8P,\ldots$$

To that end, recall that the formulas for doubling a point $P=(x_P,y_P)$ are

$$x_{-2P}=-2x_P+A^2$$

and

$$y_{-2P}=y_P+A(x_{-2P}-x_P),$$

where A is the slope of the tangent line at the point P. We apply this to $P=(1,3)$. The slope of the tangent line at the point P is

$$A=\frac{dy}{dx}(P)=\frac{3x^2}{2y}(1,3)=\frac{3}{6}.$$

Of course, in order to make sense of the fraction $3/6$ modulo 533 we need to invert 6 modulo 533. As usual, this is done using the Euclidean algorithm, applied to 533 and 6:

$$\begin{array}{rcl} 533 & = & 88\cdot 6+5 \\ 6 & = & 1\cdot 5+1 \\ 5 & = & 5\cdot 1+0 \end{array}$$

The first equation can be solved for 5 and then 5 can be eliminated from the second equation to obtain

$$89\cdot 6-533=1.$$

Thus the inverse of 6 modulo 533 is 89 so $A=3\cdot 89=267$. It follows that

$$x_{-2P}=-2\cdot 1+267^2\equiv 398 \pmod{533}$$

and

$$y_{-2P}=3+267\cdot 397\equiv 468 \pmod{533}.$$

So we get that

$$2P=(398,-468)\equiv(398,65) \pmod{533}.$$

The next step is to calculate $4P = 2(2P)$. The slope of the tangent line at $2P = (398, 65)$ is

$$A = \frac{dy}{dx}(2P) = \frac{3x^2}{2y}(398, 65) = \frac{3 \cdot 398^2}{130}.$$

This time we need to invert 130 modulo 533. The Euclidean algorithm gives:

$$\begin{aligned} 533 &= 4 \cdot 130 + 13 \\ 130 &= 10 \cdot 13 \end{aligned}$$

We cannot invert 130 modulo 533, since the Euclidean algorithm has computed 13 as a non-trivial greatest common divisor of 130 and 533. But that is not bad, since we have found that 13 is a non-trivial factor of 533. In particular, we have factored

$$533 = 13 \cdot 41.$$

So what happened here? Recall that we have shown, in Section 3 of Chapter 12, that $|E(13)| = 16$. Since $16 = 2^4$, consecutive doubling of any point modulo 13 has to give, eventually, the identity element O. In fact, we showed that the order of $P = (1, 3)$ modulo 13 is 4. Thus, the tangent line at $2P$ modulo 13 is ∞, meaning that the numerator of $A = 3x^2/2y$, evaluated at the point $2P$, must be divisible by 13. When we attempted to invert the the numerator modulo 533 we could not do it and found the factor 13 instead.

There is a bit more to say here. Similar to what can happen with Pollard's $p - 1$ test, the numerator could have been divisible by the whole 533 and we would not have discovered a proper factor of 533. But this could only happen if the order of P is 4 modulo 41, too. Since $41 \equiv 2 \pmod 3$ the order of the the elliptic curve $y^2 = x^3 + 8$ modulo 41 is

$$|E(41)| = 41 + 1 = 42,$$

by Section 3 of Chapter 12. Since 4 does not divide 42 there is no point on $E(41)$ of order 4. This explains why consecutive doubling of the point $P = (1, 3)$ on the curve $y^2 = x^3 + 8$ modulo 533 finds a non-trivial factor of 533.

We note that the group $E(41)$ is isomorphic to the group $\mathbb{Z}/42\mathbb{Z}$. (See the first exercise below.) The group $\mathbb{Z}/42\mathbb{Z}$ has only one element of order 2 and no elements of order 4. It follows that the order of a randomly picked point on $E(41)$ is very likely (with probability 40/42) not a power of 2. On the other hand, the order of any point in $E(13)$ is a power of 2. This shows that consecutive doubling of any point on the curve $y^2 = x^3 + 8$ modulo 533 would very likely yield a factorization of 533.

Of course, we do not need restrict ourselves to doubling points. We can use the above idea to factor a number $n = pq$ if we can find a a point P on an elliptic curve E such that P has small order modulo p and large order modulo q. Here is a more general approach modeled after Pollard's $p-1$ test:

Pick an Elliptic curve E and a point P on it. Compute $P_2 = 2P$, $P_3 = 3P_2 \ldots$ modulo n as long as you can.

Here is what is going on here: Assume that p is a prime factor of n. The order of P modulo p is finite, so there exists an integer B such that the order of P modulo p divides $B!$. Then

$$P_B = B! \cdot P$$

must be the identity element O on the curve $E(p)$. Put $Q = (B-1)! \cdot P$. The identity $B \cdot Q = O$ in the group $E(p)$ can be written as

$$(B-1)Q \equiv -Q \pmod{p}.$$

Thus, as you attempt to add $Q = (x_1, y_1)$ and $(B-1)Q = (x_2, y_2)$ modulo n, you have to compute the slope of the line through this two points. This involves inverting $x_2 - x_1$ modulo n. However, since $(B-1)Q$ is congruent to $-Q$ modulo p, $x_2 - x_1$ is divisible by p and the Euclidean algorithm produces a non-trivial common divisor of n and $x_2 - x_1$. If that divisor is equal to n, then the algorithm has failed, and we need to restart with a new point P or a new curve E. Of course, the number B has to be small if the algorithm is to run efficiently. This can happen only if the order of the curve E modulo p

$$|E(p)| = p - a_p + 1$$

is a product of small primes. If, with some luck, we have picked an elliptic curve such that $p - a_p + 1$ is divisible by small primes only, then we have a good chance to factor $n = pq$. Since a_p can take all kinds of values, the elliptic curve tests appear to have a considerable advantage over $p-1$ and $p+1$ methods.

As an example, we shall factor $n = 2501$ using the point $P = (1, 1)$ on the curve $y^2 = x^3 - 2x + 2$. We calculate $2P$ first. The slope of the tangent line at P is

$$A = \frac{dy}{dx} = \frac{3x^2 - 2}{2y}(P) = \frac{1}{2} \equiv 1251 \pmod{2501}$$

since the multiplicative inverse of 2 modulo 2501 is 1251. Now one quickly computes that

$$x_{-2P} = -2 \cdot 1 + 1251^2 \equiv 1874 \pmod{2501}$$

and

$$y_{-2P} = 1 + 1251 \cdot 1873 \equiv 2188 \pmod{2501}.$$

So we get that

$$Q = 2P = (1874, -2188) \equiv (1874, 313).$$

The next step is to compute $3Q = Q + 2Q$. The slope of the tangent at Q is

$$A = \frac{dy}{dx} = \frac{3x^2-2}{2y}(Q) = \frac{10535626}{626} = 1414 \cdot 835 = 218 \pmod{2501}.$$

since $10535626 \equiv 1414$ and the inverse of 626 is 835. From this we compute that

$$2Q = (1259, 1297).$$

Now we can add $Q + 2Q$. In order to compute the slope through Q and $2Q$ we need to invert

$$x_{2Q} - x_Q = 1259 - 1874 \equiv 1886 \pmod{2501}.$$

This is done using the Euclidean algorithm:

$$\begin{array}{rcl} 2501 &=& 1 \cdot 1886 + 615 \\ 1886 &=& 3 \cdot 615 + 41 \\ 615 &=& 15 \cdot 41 + 0 \end{array}$$

This shows that $\gcd(1886, 2501) = 41$. In particular we have found a factorization

$$2501 = 41 \cdot 61.$$

Since adding points on an elliptic curve is hard work, one may wonder if this approach to factoring has any merits. In order to get a better understanding of the efficiency of the test, assume that we want to factor a 60 digit composite number $n = pq$ where the prime factors p and q have around 30 digits each. The sieve method would require about

$$10^{30} = 1,000,000,000,000,000,000,000,000,000,000$$

steps. On the other hand, if there is an elliptic curve E such that

$$|E(p)| = p - a_p + 1 = 2^m$$

then the elliptic curve test could factor n by doubling a point on E in less then m steps. By Hasse's estimate, the number of digits of $E(p)$ is the same as the number of digits of p (add or take one), which is 30. It follows that

$$m = \log_2(|E(p)|) \approx \log_2(p) \approx \log_2(10^{30}) \approx 100.$$

Thus, the number of doublings is less than 100. Each doubling requires inverting of an integer modulo n. This is done using the Euclidean algorithm and it takes less then 5 times the number of digits of n. In all, about $100 \times 300 = 30,000$ numerical operations might be expected and this is, by a huge factor, less than 10^{30}.

Exercises

1) Compute the order of $P = (1, 3)$ on $y^2 = x^3 + 8$ modulo 41. Helpful hint: The order of the the group $E(41)$ is 42.

2) Let E be the elliptic curve $y^2 = x^3 + 17$. The prime factors $p < q$ of the composite number $7519 = pq$ satisfy property

$$|E(p)| = 64 = 2^6 \text{ and } |E(q)| = 111 = 3 \cdot 37.$$

Therefore, the order of any point P is a power of 2 modulo p and it is odd modulo q. In particular, doubling any point $P \neq O$ quickly gives the identity element modulo p, but nod modulo q. Double the point P, as many times as necessary, to factor 7519 where

a) $P = (-1, 4)$

b) $P = (2, 5)$.

3) Let E be the elliptic curve $y^2 = x^3 + 15$. Try to factor 7519 by consecutively doubling $P = (1, 4)$. What happens?

4) Factor 6077 using the point $P = (2, 5)$ on the curve $y^2 = x^3 + 17$. That is, compute $P_2 = 2P$, then $P_3 = 3P_2$ etc...

5) Reverse engineering. In this exercise we shall construct a composite number $n = p \cdot q$ which can be factored using the point $P = (1, 2)$ be on the curve $y^2 = x^3 - x + 4$. First, calculate $Q = 2P$ as a rational point. Next, calculate $2Q$. In order to calculate $3Q$ - do not do that - you need the slope through Q and $2Q$:

$$A = \frac{y_{2Q} - y_Q}{x_{2Q} - x_Q}.$$

Write $x_{2Q} - x_Q$ as a reduced fraction and let p be the biggest prime divisor of the numerator. Let q be the first prime number bigger then $2p$. This choice of q is not important, but it assures that the following number

$$n = pq$$

is well defined. Now use the point $P = (1, 2)$ on the curve $y^2 = x^3 - x + 4$ to factor n.

2. Degenerate curves modulo p

The factorization attack via cubic curves can be applied to degenerate curves, as well. Since degenerate curves are somewhat special and different from non-degenerate (i.e. elliptic) curves, it may be interesting to investigate them in more details. Recall that a curve $y^2 = f(x)$ is degenerate if $f(x)$ has a double root. By shifting the x-coordinate, if necessary, we can assume that the double root is 0, so the equation takes form

$$y^2 = x^3 + dx^2.$$

Consider first the case when $d = 0$. This is the case when $f(x)$ has a triple root. The equation is $y^2 = x^3$. Recall that the group law amounts to adding the fraction x/y. Thus, given a point $P = (x, y)$ on $y^2 = x^3$ and an integer m, then

$$mP = \left(\frac{x}{m^2}, \frac{y}{m^3}\right).$$

Indeed, if we put $x' = x/m^2$ and $y' = y/m^3$, then

$$\frac{x'}{y'} = m \cdot \frac{x}{y},$$

as required.

Now recall that factoring a composite number n using a point P on the curve $y^2 = x^3$ requires computing

$$P_2 = 2P, P_3 = 3P_2, P_4 = 4P_3 \ldots$$

modulo n. This amounts to dividing coordinates of $P, P_2, P_3 \ldots$ by (powers of) $2, 3, 4 \ldots$, modulo n. We can do that as long as $2, 3, 4 \ldots$ are relatively prime to n. Note that this algorithm stops once we have arrived at the smallest prime p dividing n. In essence, our algorithm is the oldest factorization method, the sieve of Eratosthenes.

The situation is more exciting if d is non-zero modulo p. We can easily find all solutions of the equation $y^2 = x^3 + dx^2$, as follows. If $x = 0$ then $y = 0$, and we have the singular point $S = (0, 0)$ on the curve. If $P = (x, y)$ is any other point on the curve, then $x \neq 0$ so $t = y/x$ is well defined. Substitute $y = tx$ in the cubic equation. This gives

$$x^2t^2 = x^3 + dx^2$$

which implies that $x = t^2 - d$. Since $y = xt$ we see that the point P is given by

$$P = (t^2 - d, t^3 - td).$$

However, if d is a square, $d = u^2$, then $P = S$ for $t = \pm u$. If d is not a square, then $P \neq S$ for any t. Thus, the number of solutions of $y^2 = x^3 + dx^2$ (including S) is $p - 1$ if d is a square, and $p + 1$ if d is not a square. The group $E_{ns}(p)$ is formed by excluding S and including the point at infinity O. Thus the order of the group $E_{ns}(p)$ is

$$|E_{ns}(p)| = \begin{cases} p - 1 \text{ if } d \text{ is a square, and} \\ p + 1 \text{ if } d \text{ is not a square.} \end{cases}$$

Without verification, we state here that the group law on $E_{ns}(p)$ amounts to multiplying numbers

$$h(P) = \frac{y - x\sqrt{d}}{y + x\sqrt{d}},$$

where $P = (x, y)$ is a point on E. (We also set $h(P) = 1$.) If d is a square modulo p then the above fraction is understood as an element of $\mathbb{F}_p^\times$. Otherwise, it is an element of the quadratic extension $\mathbb{F}_{p^2}$. Moreover, in this case, $h(P)$ lands in the circle group

$$T(p) = \{z = a + b\sqrt{d} \in \mathbb{F}_{p^2}^\times \mid z\bar{z} = 1\},$$

where $\bar{z} = a - b\sqrt{d}$. Indeed, $h(P) = z/\bar{z}$, where $z = y - x\sqrt{d}$, so

$$h(P) \cdot \overline{h(P)} = \frac{z}{\bar{z}} \cdot \frac{\bar{z}}{z} = 1.$$

We recall that $T(p)$ is a subgroup of $\mathbb{F}_{p^2}^\times$ analogous to the (unit) circle subgroup $\mathbb{T}$ of $\mathbb{C}^\times$. The order of $T(p)$ is $p+1$.

Summarizing, the map h is a homomorphism from $E_{ns}(p)$ to $\mathbb{F}_p^\times$ or $T(p)$ depending on whether d is a square or not, respectively.

Proposition 50. *The map h is one to one and onto. In other words, the map h is an isomorphism of groups.*

Proof. In order to verify this statement we need to show that the map h is one to one. Let $P = (x, y)$ such that $h(P) = 1$:

$$h(P) = \frac{y - x\sqrt{d}}{y + x\sqrt{d}} = 1.$$

This identity can be rewritten as

$$y - x\sqrt{d} = y + x\sqrt{d}$$

and $2x\sqrt{d} = 0$. It follows that $x = 0$. If $x = 0$, then $y^2 = x^3 + dx^2 = 0$ and $y = 0$, as well. This means that $P = S$. But P cannot be equal to S since S is excluded from the group. Thus we have shown that $h(P) = 1$ only for $P = O$. Now we can easily show that h is one to one. Indeed, if $h(P_1) = h(P_2)$ then, since h is a homomorphism,

$$h(P_1 - P_2) = 1$$

which means that $P_1 - P_2 = O$ or $P_1 = P_2$. Finally, the map has to be onto since it is a map between two finite sets with the same number of elements. □

The factorization method based on the curve $y^2 = x^3 + dx^2$ is, essentially, the $p-1$ method or $p+1$ method depending whether d is a square or not modulo p, respectively. The first of this two cases is quite believable, especially if we show the second, which we intend to do here. To that end assume, for simplicity, that $p \equiv 3 \pmod{4}$. Then we can chose $d = -1$. The field $\mathbb{F}_{p^2}$ can be realized as Gaussian integers modulo p. The factorization of a composite number $n = pq$ using the curve $y^2 = x^3 - x^2$ consists of picking an integer point $P = (a, b)$ and then calculating $P_i = i! \cdot P$ for $i = 2, 3, \ldots$ modulo n. Via the isomorphism h this amounts to calculating

$$\left(\frac{z}{\bar{z}}\right)^{i!}$$

where $z = b - ia$ and the quotient is understood in terms of Gaussian integers modulo n. The algorithm terminates when $P_B = O$ in $E_{ns}(p)$, which happens relatively quickly if $p+1$ is a product of small primes. Via the homomorphism h the identity $P_B = O$ is equivalent to

$$\left(\frac{z}{\bar{z}}\right)^{B!} \equiv 1 \pmod p$$

or, after multiplying both sides of the congruence by $\bar{z}^{B!}$,

$$z^{B!} \equiv \bar{z}^{B!} \pmod p.$$

But this means that imaginary part of $z^{B!}$ is trivial:

$$\Im(z^{B!}) \equiv 0 \pmod p.$$

This congruence is precisely what terminates the $p+1$ algorithm with z as the initial input. Summarizing, the factorization attack involving the cubic curve $y^2 = x^3 - x^2$ and a point $P = (a, b)$ is the same as the $p+1$ method starting with $z = b - ai$.

Through our emphasis on groups, we have seen that Lenstra's elliptic curve method is a natural generalization of Pollard's $p-1$ method. Although, superficially, Pollard's $p-1$ method and Lenstra's elliptic curve method appear different, in essence, they are quite similar. It is only the underlying group that is different. The following table summarizes which group is used in the four factorization methods.

Method:	Sieve	$p-1$	$p+1$	Lenstra
Group:	$\mathbb{Z}/p\mathbb{Z}$	$\mathbb{F}_p^\times$	$T(p)$	$E(p)$

3. Elliptic curve test for Mersenne primes

Recall that a Mersenne number is a number $M_\ell = 2^\ell - 1$ where where ℓ is an odd prime. The Lucas-Lehmer test for primality of M_ℓ makes use of the circle group $T(p)$. This group has the order $p+1$. In particular, if M_ℓ is a prime, then $T(M_\ell)$ is a *cyclic* group of order 2^ℓ. This observation is a crucial ingredient in the proof of the test.

Dick Gross has recently introduced an elliptic curve version of the Lucas-Lehmer test for Mersenne primes. In essence, Gross replaces the group $T(p)$ with the group $E(p)$ where E is a carefully chosen elliptic curve. Consider the elliptic curve E given by $y^2 = x^3 - 6x$. The curve E has two obvious integer points:

$$\begin{cases} Q = (0, 0) \\ P = (3, 3). \end{cases}$$

Note that Q is a point of order 2 on E. The discriminant of E is $\Delta(E) = 2^5 3^3$. In particular, E has a good reduction for all primes $p > 3$. The points P and Q can be considered as elements of the finite group $E(p)$ of points modulo p on the elliptic curve E. Recall that the order of $E(p)$ is $p+1$ for every prime $p \equiv 3 \pmod 4$ and $p \neq 3$. In particular, if M_ℓ is prime then the order of $E(M_\ell)$ is

$$|E(M_\ell)| = M_\ell + 1 = 2^\ell.$$

Moreover, as we shall see in a moment, the group $E(M_\ell)$ is also cyclic. These two observations are key ingredients in the test.

PROPOSITION 51. *Let $P = (3,3)$ and $Q = (0,0)$ be points on the elliptic curve $y^2 = x^3 - 6x$. Let ℓ be an odd prime. The number $M_\ell = 2^\ell - 1$ is prime if and only if the congruence*

$$2^{\ell-1}P \equiv Q \pmod{M_\ell}$$

holds for the curve $y^2 = x^3 - 6x$.

PROOF. Assume that the congruence holds. We want to show that M_ℓ is prime. Let p be a prime factor of M_ℓ. We can pick $p \equiv 3 \pmod 4$. Indeed, if all prime factors of M_ℓ are congruent to 1 modulo 4 then their product is also congruent to 1 modulo 4. But M_ℓ is congruent to 3 modulo 4 and this is a contradiction. Next, note that p is not 3. Indeed, since $2 \equiv -1 \pmod 3$, we have

$$M_\ell = 2^\ell - 1 \equiv (-1)^\ell - 1 \equiv -2 \equiv 1 \pmod 3.$$

Since the curve E has a good reduction modulo any prime $p > 3$, the congruence $2^{\ell-1}P \equiv Q \pmod{M_\ell}$ implies

$$2^{\ell-1}P \equiv Q \pmod p$$

for any prime p dividing M_ℓ. Since doubling Q gives O, the identity element in $E(p)$, we have

$$\begin{cases} 2^{\ell-1}P \not\equiv O \pmod p \\ 2^\ell P \equiv O \pmod p. \end{cases}$$

This shows that the order of P in $E(p)$ is exactly $2^\ell = M_\ell + 1$. (See the proof of the Lucas-Lehmer test for a detailed explanation why the order is 2^ℓ.) On the other hand, since we have picked $p \equiv 3 \pmod 4$,

$$|E(p)| = p + 1.$$

Since the order of P is less then or equal to the order of the group $E(p)$ it follows that $M_\ell + 1 \leq p + 1$ and $M_\ell \leq p$. This implies that $M_\ell = p$.

To prove the other direction, assume that M_ℓ is prime. If we can show that

(1) $E(M_\ell)$ is a cyclic group of order 2^ℓ
(2) P is an element of order 2^ℓ in $E(M_\ell)$

then we are done. Indeed, (2) implies that $2^{\ell-1}P$ has order 2. Since a cyclic group of order 2^ℓ has only one element of order 2 and the point $Q = (0,0)$ is an element of $E(M_\ell)$ of order 2, it follows that

$$2^{\ell-1}P \equiv Q \pmod{M_\ell}.$$

as desired. It remains to show (1) and (2) above. First of all, since $M_\ell \equiv 3 \pmod 4$ the order of $E(M_\ell)$ is $M_\ell + 1 = 2^\ell$. To show that $E(M_\ell)$ is cyclic we need the following simple lemma.

Lemma 52. *Let G be a commutative group of order 2^ℓ. Assume that G has only one element z of order 2. Then G is cyclic.*

Proof. Let G^2 be the set of squares of of all elements of G, let G^4 be the set of 4-th powers of all elements of G and so on. In this way we get a sequence of subgroups

$$G \supseteq G^2 \supseteq G^4 \supseteq \dots$$

in G. If G is not cyclic then the order of any element in G divides $2^{\ell-1}$. Then $g^{2^{\ell-1}} = 1$ for every element g in G. Then $G^{2^{\ell-1}}$ is the trivial subgroup of G. Thus, in order to show that G is cyclic it suffices to show that $G^{2^{\ell-1}}$ is not trivial. We will accomplish this as follows: We will show that the order of G^2 is half the order of G, the order of G^4 is half the order of G^2 and so on. In the end this shows that the order of $G^{2^{\ell-1}}$ is 2. In particular, $G^{2^{\ell-1}}$ is not trivial.

Consider the map $x \mapsto x^2$ from G onto G^2. If $x^2 = y^2$ then $(xy^{-1})^2 = 1$. This implies that $x = y$ or $x = yz$. It follows that squaring is a 2 to 1 map and, therefore, the order of G^2 is $2^{\ell-1}$. Since G^2 surely contains an element of order 2, it must contain z as it is a unique element of order 2 in G. The same argument now shows that the order of G^4 is $2^{\ell-2}$. Continuing in this fashion it follows that the order of $G^{2^{\ell-1}}$ is 2. The lemma is proved. □

Now let's go back to the curve $E(M_\ell)$. We already know that the order of $E(M_\ell)$ is 2^ℓ. In order to show that $E(M_\ell)$ is cyclic, by lemma, we need to show that $Q = (0,0)$ is the only element of order 2 in $E(M_\ell)$. Recall that elements of order 2 correspond to the roots of $x^3 - 6x = x(x^2 - 6)$. The quadratic reciprocity implies that 2 is a square modulo M_ℓ and 3 is not a square modulo M_ℓ. It follows that 6 is not a square modulo M_ℓ and 0 is the unique root of $x^3 - 6x$ modulo M_ℓ. This shows that $E(M_\ell)$ is cyclic.

It remains to show that $P = (3,3)$ has the order 2^ℓ in $E(M_\ell)$. We need another simple result for cyclic groups of order 2^ℓ:

Lemma 53. *Let G be a cyclic group of order 2^ℓ. Then elements of order 2^ℓ in G are precisely non-squares.*

Proof. This can be easily seen by writing down G as the group $\mathbb{Z}/2^\ell\mathbb{Z} = \{0, 1, \dots 2^\ell - 1\}$. The group operation is the modular addition. In particular, to square a number means to multiply the number by 2. It follows that even numbers $0, 2, \dots, 2^\ell - 2$ are squares while odd numbers $1, 3, \dots, 2^\ell - 1$ are non-squares. The order of any odd number is clearly 2^ℓ. □

If we can show that P is not obtained by doubling a point on the curve modulo m then the order of $P = (3,3)$ in $E(M_\ell)$ is 2^ℓ, by the lemma. If

$(3,3) \equiv 2R \pmod{M_\ell}$ for some $R = (x_R, y_R)$ in $E(M_\ell)$ then the doubling formula (worked out in Chapter **??** Section 4) shows that

$$3 \equiv \frac{(x_R+6)^2}{4y_R^2} \pmod{M_\ell}.$$

This is impossible since 3 is not a square modulo M_ℓ. Therefore the order of P in $E(M_\ell)$ is 2^ℓ, as desired. The proof of the test is now complete.

□

As an example, consider $31 = 2^5 - 1$. A calculation modulo 31 shows that

$$\begin{array}{rcll} 2P & \equiv & (14,5) & \pmod{31} \\ 4P & \equiv & (23,10) & \pmod{31} \\ 8P & \equiv & (16,21) & \pmod{31} \\ 16P & \equiv & (0,0) & \pmod{31} \end{array}$$

confirming that 31 is indeed a prime number.

The original test, as developed by Dick Gross, uses the curve $y^2 = x^3 - 12x$ and points $P = (-2,4)$ and $Q = (0,0)$. The test works the same: a Mersenne number $M_\ell = 2^\ell - 1$ is prime if and only if

$$2^{\ell-1} \equiv Q \pmod{M_\ell}$$

on E. The proof is completely analogous to the proof for the curve $y^2 = x^3 - 6x$. The main ingredients are given in the following exercises.

Exercises

1) Let E be the elliptic curve $y^2 = x^3 - 12x$. Then $P = (-2,4)$ and $Q = (0,0)$ are two points on E. Let ℓ be an odd prime such that $M_\ell = 2^\ell - 1$ is prime. Show that

(1) Q is the only point of order 2 in $E(M_\ell)$.
(2) P is not obtained by doubling any point on $E(M_\ell)$.

2) The previous problem shows that the curve $y^2 = x^3 - 12x$ can also be used to test primality of Mersenne numbers. Use the test to show that $31 = 2^5 - 1$ is prime.