Notes for Number Theory

Here are the notes I wrote up for a number theory course I taught. The notes cover elementary number theory but don't get into anything too advanced. My approach to things is fairly informal. I like to explain the ideas behind everything without getting too formal and also without getting too wordy.

If you see anything wrong (including typos), please send me a note at `heinold@msmary.edu`.

Last updated: March 11, 2024.

Contents

Chapter 1

Divisibility

1.1 Definition of divisibility

One of the most basic concepts in number theory is that of divisibility. The concept is familiar: 14 is divisible by 7, even numbers are divisible by 2, prime numbers are only divisible by themselves and 1, etc. Here is the formal definition:

Definition 1. *An integer d is a* divisor *of an integer n if there exists an integer k such that $n = dk$. We say that n is* divisible by *d, or that d* divides *n, and write $d \mid n$.*

For example, 20 is divisible by 4 because we can write $20 = 4 \cdot 5$; that is $n = dk$, with $n = 20$, $d = 4$, and $k = 5$. The equation $n = dk$ is the key part of the definition. It gives us a formula that we can associate with the concept of divisibility.

This formula is handy when it comes to proving things involving divisibility. If we are given that n is divisible by d, then we write that in equation form as $n = dk$ for some integer k. If we need to show that n is divisible by d, then we need to find some integer k such that $n = dk$.

Here are a few example proofs:

1. Suppose we want to prove the simple fact that if n is even, then n^2 is even as well.

   Proof. Even numbers are divisible by 2, so we can write $n = 2k$ for some integer k. Then $n^2 = (2k)^2 = 4k^2$, which we can write as $n^2 = 2(2k^2)$. We have written n^2 as 2 times some integer, so we see that n^2 is divisible by 2 (and hence even). □

2. Prove that if $a \mid b$ and $b \mid c$, then $a \mid c$

   Proof. Since $a \mid b$ and $b \mid c$ we can write $b = aj$ and $c = bk$ for some integers j and k.[2] Plug the first equation into the second to get $c = (aj)k$, which we can rewrite as $c = a(jk)$. So we see that $a \mid c$, since we have written c as a multiple of a. □

3. Prove that if $a \mid b$, then $ac \mid bc$, for any integer c.

   Proof. Since $a \mid b$, we can write $b = ak$ for some integer k. Multiply both sides of this equation by c to get $(ac)k = bc$. This equation tells us that $ac \mid bc$, which is what we want. □

[2] Note that we *must* use different integers here since the integer j that works for $a \mid b$ does not necessarily equal the integer k that works for $b \mid c$.

Here are a couple of other divisibility examples:

1. Disprove: If $a\,|(b+c)$, then $a\,|\,b$ or $a\,|\,c$.

   Solution. All we need is a single counterexample. Setting $a = 5$, $b = 3$, and $c = 7$ does the trick.

2. Find a divisor of 359951 besides 1 and itself.

   Solution. The only answers are 593 and 607. It would be tedious to find them by checking divisors starting with 2, 3, etc. A better way is to use the algebraic fact $x^2 - y^2 = (x-y)(x+y)$. This fact is very useful in number theory.

   With a little cleverness, we might notice that 359951 is $360000-49$, which is 600^2-7^2. We can factor this into $(600-7)(600+7)$, or 593×607.

1.2 The division algorithm

The *division algorithm*, despite its name, it is not really an algorithm. It states that when you divide two numbers, there is a unique quotient and remainder. Specifically, it says the following:

Theorem 1. *Let $a, b \in \mathbb{Z}$ with $b > 0$. Then there exist unique $q, r \in \mathbb{Z}$ such that $a = bq + r$ and $0 \leq r < b$.*

The integers q and r are called the *quotient* and *remainder*. For example, if $a = 27$ and $b = 7$, then $q = 3$ and $r = 6$. That is, $27 \div 7$ is 3 with a remainder of 6, or in equation form: $27 = 7 \cdot 3 + 6$. The proof of the theorem is not difficult and can be found in number theory textbooks.

One of the keys here is that the remainder is less than b. Here are some consequences of the theorem:

- Taking $b = 2$, tells us all integers are of the form $2k$ or $2k+1$ (i.e., every integer is either odd or even).[1]
- Taking $b = 3$, all integers are of the form $3k$, $3k+1$, or $3k+2$.
- Taking $b = 4$, all integers are of the form $4k$, $4k+1$, $4k+2$, or $4k+3$.
- For a general b, all integers are of the form bk, $bk+1$, …, or $bk+(b-1)$.

These are useful for breaking things up into cases. Here are a few examples:

1. Suppose we want to show that $n^3 - n$ is always divisible by 3. We can break things up into cases $n = 3k$, $n = 3k+1$, and $n = 3k+2$, like below:

$$\begin{aligned}
n = 3k: &\quad (3k)^3 - 3k = 27k^3 - 3k = 3(9k^3 - k),\\
n = 3k+1: &\quad (3k+1)^3 - (3k+1) = 27k^3 + 9k^2 + 3k + 1 - 3k - 1 = 3(9k^3 - 3k),\\
n = 3k+2: &\quad (3k+2)^3 - (3k+2) = 27k^3 + 18k^2 + 12k + 8 - 3k - 2 = 3(9k^3 + 6k^2 + 3k + 2).
\end{aligned}$$

   We see that in each case, $n^3 - n$ is divisible by 3. By the division algorithm, these are the only cases we need to check, since every integer must be of one of those three forms.

2. Prove that every perfect square is of the form $4k$ or $4k+1$.

   Proof. Every integer n is of the form $2k$ or $2k+1$.

   If $n = 2k$, then we have $n^2 = 4k^2$, which is of the form $4k$.[2]

   If $n = 2j+1$, then we have $n^2 = 4k^2 + 4k + 1 = 4(k^2+k) + 1$, which is of the form $4k+1$. □

[1] We use k instead of q here out of convention.

[2] We are being a bit informal here with the notation. When we say the number is of the form $4k$, the k is different from the k used in $n = 2k$. What we're really saying here is that the number is of the form 4 times some integer. A more rigorous way to approach this might be to let $n = 2j$, compute $n^2 = 4j^2$ and say $n^2 = 4k$, with $k = j^2$.

3. A prime number is a number greater than 1 whose only divisors are 1 and itself. Prove that every prime greater than 3 is of the form $6k+1$ or $6k+5$.

   Proof. Every integer is of the form $6k$, $6k+1$, $6k+2$, $6k+3$, $6k+4$, or $6k+5$. An integer of form $6k$ is divisible by 6. A integer of the form $6k+2$ is divisible by 2 as it can be written as $2(3k+1)$. Similarly, an integer of the form $6k+3$ is divisible by 3 and an integer of the form $6k+4$ is divisible by 2. None of these forms can be prime (except for the integers 2 and 3, which we exclude), so the only forms left that could be prime are $6k+1$ and $6k+5$. □

4. Prove that $16 \mid a^4+b^4-2$ for any odd integers a and b.

   Proof. We will start by writing a^4+b^4-2 as $(a^4-1)+(b^4-1)$. Let's take a look at the a^4-1 term. We can factor that into $(a^2-1)(a^2+1)$ and further into $(a-1)(a+1)(a^2+1)$. Since we know a is odd, we can write $a=2k+1$ and we have

   $$(a-1)(a+1)(a^2+1) = (2k)(2k+2)(4k^2+4k+1) = 8k(k+1)(2k^2+2k+1).$$

   Our goal is to show that this expression is divisible by 16, but so far we've only found a factor of 8. We get one additional factor of 2 from the $k(k+1)$ part of the expression. We know that one of k and $k+1$ is even, since k and $k+1$ are consecutive integers, and that gives us our factor of 16.

   So we have that a^4-1 is divisible by 16. A similar argument tells us that b^4-1 is divisible by 16, and from there we get that a^4+b^4-2 is divisible by 16, since it is the sum of two multiples of 16. □

5. One of the oldest and most famous proofs in math is that $\sqrt{2}$ is irrational. That is, $\sqrt{2}$ cannot be written as a ratio of integers. Here is the proof:

   Proof. First, note that the square of an even is even and the square of an odd is odd since $(2k)^2 = 2(2k^2)$ is even and $(2k+1)^2 = 2(2k^2+2k)+1$ is odd. In particular, if an integer a^2 is even, then a is even as well.

   Suppose $\sqrt{2} = p/q$, with p and q positive integers. By clearing common factors, we can assume the fraction is in lowest terms. Multiply both sides by q and square both sides to get $2q^2 = p^2$. This tells us that $2 \mid p^2$. Thus $2 \mid p$ by the statement above, and we can write $p = 2k$ for some integer k. So we have $2q^2 = (2k)^2$, which simplifies to $q^2 = 2k^2$. This tells us that $2 \mid q^2$ and hence $2 \mid q$. But this is a problem because p and q both have a factor of 2 and p/q is supposed to already be in lowest terms. So we have a contradiction, which shows that it must not be possible to write $\sqrt{2}$ as a ratio of integers.

   It is not to hard to extend this result to show that $\sqrt[n]{m}$ is irrational unless m is a perfect nth power.

Perfect squares

The second example above is sometimes useful when working with perfect squares. We record it as a theorem:

Theorem 2. *Every perfect square is of the form $4k$ or $4k+1$.*

For example, 3999 is not a perfect square since it is $4000-1$, which is of the form $4k-1$ (same as a $4k+3$ form). On the other hand, just because something is of the form $4k+1$ does not mean it is a perfect square. For instance, 41 is of the form $4k+1$, but isn't a perfect square.

As another example, $3n^2-1$ is not a perfect square for any integer n. To see this, break the problem into two cases: $n=2k$ and $n=2k+1$. If $n=2k$, then $3n^2-1 = 3(4k^2)-1 = 4(3k^2)-1$. If $n=2k+1$, then $3n^2-1 = 3(4k^2+4k+1)-1 = 4(3k^2+3k)+2$. Neither of these are of the form $4k$ or $4k+1$, so they are not perfect squares.

Similar results to the theorem above can be proved for other integers. For instance, every perfect square is of the form $5k$, $5k+1$, or $5k+4$.

More about remainders

Note that we can also take the remainders to be in other ranges besides from 0 to $b-1$. The most useful range is from $-b/2$ and $b/2$. For instance, with $b=3$, we can also write every integer as being of the form $3k-1$, $3k$, or $3k+1$. An integer of the form $3k-1$ is also of the form $3k+2$. As another example, we can write every integer in one of the forms $6k-2$, $6k-1$, $6k$, $6k+1$, $6k+2$, or $6k+3$.

1.3 The modulo operation

In grammar school, the remainder always seemed to me to be an afterthought, but in higher math, it is quite useful and important. It is built into most programming languages, usually with the symbol `mod` or `%`

Definition 2. *The remainder when an integer a is divided by b is denoted by $a \bmod b$. It is the integer r in the division algorithm expression $a = bq + r$, with $0 \leq r < b$. We have*

$$a \bmod b = a - \left\lfloor \frac{a}{b} \right\rfloor b.$$

For example, suppose we want to find $68 \bmod 7$. The definition above tells us to find the nearest multiple of 7 less than n and subtract. The closest multiple of 7 less than 68 is 63, and $68-63=5$. So $68 \bmod 7 = 5$.

This procedure applies to negatives as well. For instance, to compute $-31 \bmod 5$, the closest multiple of 5 less than or equal to -31 is -35, which is 4 away from -31, so $-31 \bmod 5 = 4$, or $-31 \equiv 4 \pmod 5$.

As one further example, suppose we want $179 \bmod 18$. 179 is one less than 180, a multiple of 18, so it leaves a remainder of 18-1 = 17. So $179 \bmod 18 = 17$.

A nice way to compute mods mentally or by hand is to use a streamlined version of the grade school long division algorithm. For example, suppose we want to compute $34529 \bmod 7$. Here is the procedure:

```
  34529
- 28
   65
 - 63
    22
   -21
     19
    -14
      5
```

The end result is $34529 \bmod 7 = 5$.

1.4 The greatest common divisor

Definition 3. *The* greatest common divisor*, or gcd, of two integers a and b is the largest integer that divides both a and b. We denote it by* $\gcd(a,b)$.[1]

For example, the gcd of 24 and 96 is 12, since 12 is the largest integer that divides both 24 and 96. As another example, the gcd of 18 and 25 is 1. Those numbers have no divisors in common besides 1.

[1] In some texts, the notation (a,b) is used instead of $\gcd(a,b)$.

The Euclidean algorithm

To find the gcd of two integers, the *Euclidean algorithm* is used. We'll start with an example, finding $\gcd(21, 78)$:

$$\begin{aligned}
&78 \bmod 21 = 15\\
&21 \bmod 15 = 6\\
&15 \bmod 6 = 3\\
&6 \bmod 3 = 0
\end{aligned}$$

The last nonzero remainder, 3, is the gcd.

In general, to find $\gcd(a, b)$, assume $a \leq b$, and compute

$$\begin{aligned}
&r_1 = b \bmod a\\
&r_2 = a \bmod r_1\\
&r_3 = r_1 \bmod r_2\\
&r_4 = r_2 \bmod r_1,\\
&\text{etc.}
\end{aligned}$$

until a remainder of 0 is obtained. The last nonzero remainder is the gcd. Here is how we might program it in Python[1]:

```
def gcd(a, b):
    while b != 0:
        b, a = a, b % a
    return a
```

Why the Euclidean algorithm works

The reason it works is that the common divisors of a and b are exactly the same as the common divisors as a and $b \bmod a$, so their gcds must be the same. Because of this, when we apply the Euclidean algorithm, the gcd of the two numbers on the left side stays constant all the way through the algorithm. For example, when we compute $\gcd(21, 78)$, we get the following:

$$\begin{aligned}
&78 \bmod 21 = 15\\
&21 \bmod 15 = 6\\
&15 \bmod 6 = 3\\
&6 \bmod 3 = 0.
\end{aligned}$$

The fact that $\gcd(a, b) = \gcd(a, b \bmod a)$ tells us that $\gcd(21, 78)$, $\gcd(15, 21)$, $\gcd(6, 15)$, and $\gcd(3, 6)$ are all the same, and since the last step gives us $6 \bmod 3 = 0$, we know that 3 is a divisor of 6, and hence

$$\gcd(21, 78) = \gcd(15, 21) = \gcd(6, 15) = \gcd(3, 6) = 3.$$

Also, since they are all the same, that means we can actually stop the process early. For instance, it's easy to see that $\gcd(15, 21) = 3$, so we could stop there.

It is worth showing why the common divisors of a and b are the same as the common divisors as a and $b \bmod a$. First, when we apply the division algorithm to a and b, we get $b = aq + r$, where $r = b \bmod a$. If a and b are both divisible by some common divisor d, then $r = b - aq$ will be as well, since we can factor d out of the right side. On the other hand, if a and r are both divisible by some common divisor d, then $b = r - aq$ will be as well, since we can factor d out of the right side.

[1] Note that the gcd is already built into Python's `fractions` module.

1.5 Gcds and linear combinations

In number theory, a *linear combination* of the integers a and b is an expression of the form $ax + by$ for some integers x and y. For example, a linear combination of $a = 6$ and $b = 15$ is an expression of the form $6x + 15y$, like $6(1) + 15(2) = 36$ or $6(4) + 15(-1) = 9$. Linear combinations are important in a variety of contexts.

Linear combinations have a close connection with the gcd. Suppose we want to know if it is possible to write 4 as a linear combination of 6 and 15. That is, can we find integers x and y such that $6x + 15y = 4$? The answer is no, since the left side is a multiple of 3 (namely $3(2x + 5y)$), but the right side is not a multiple of 3. By the same reasoning, in general, if c is not a multiple of $\gcd(a, b)$, then it is impossible to write c as a linear combination of a and b.

What about multiples of the gcd? Is it always possible to write $ax + by = c$ if c is a multiple of the gcd? The answer is yes. We just have to show how to write the gcd as a linear combination of a and b. Once we have this (see the proof below for how), we can then multiply through to get c. For instance, suppose we want to find x and y such that $6x + 15y = 21$. We have $\gcd(6, 15) = 3$ and it is not hard to find that $6(3) + 15(-1) = 3$. If we multiply through by 7, we get $6(21) + 15(-7) = 21$. So it just comes down to writing the gcd as a linear combination.

Here is a formal statement of the above along with a proof:

Theorem 3. *Let $a, b, c \in \mathbb{Z}$ with $d = \gcd(a, b)$. There exist integers x and y such that $ax + by = c$ if and only if $d \mid c$.*

Proof. First, since $d \mid a$ and $d \mid b$, we have $d \mid c$. Thus, we cannot write c as a linear combination of a and b if it is not divisible by d.

We now show that it is possible to write d as a linear combination of a and b. Start by letting e be the smallest positive linear combination of a and b. We need to show that $e = d$. By the division algorithm, we can write $a = eq + r$ for some integers q and r. Then we have

$$r = a - eq = a - (ax + by)q = a(1 - x) + b(qy).$$

So r is a linear combination of a and b. But, by the division algorithm, $0 \leq r < e$. Since e is the smallest positive linear combination of a and b, we must have $r = 0$. Thus, $a = eq + r$ with $r = 0$ tells us that $e \mid a$. A similar argument shows that $e \mid b$. So e is a common divisor of a and b. But, as we mentioned earlier, any linear combination of a and b is a multiple of d. So e is both a multiple of d and a common divisor of a and b, which means $e = d$.

Finally, if c is a multiple of d (say $c = dk$ for some integer k), and we have integers x and y such that $ax + by = d$, then we can multiply through by k to get $a(kx) + b(ky) = c$. □

In particular, we have the following important special case:

Corollary 4. *Let $a, b \in \mathbb{Z}$ with $d = \gcd(a, b)$. Then there exist integers x and y such that $ax + by = d$.*

This fact is useful when working with gcds because it gives us an equation to work with. On the other hand, be careful. Just because we can write $ax + by = c$, that does not mean $c = \gcd(a, b)$. All we are guaranteed is that c is a multiple of $\gcd(a, b)$.

1.6 The extended Euclidean algorithm

Theorem 3 tells us that the gcd is a linear combination, but it doesn't tell us how to find that linear combination. Being able to find that linear combination is important in a number of contexts. The trick is to use the Euclidean algorithm in a particular way.

In the earlier example when we found gcd(21, 78) using the Euclidean algorithm, we used the modulo operation. Written out fully using the division algorithm, the Euclidean algorithm on this example is as follows:

$$\begin{aligned}
78 &= 21 \cdot 3 + 15 \\
21 &= 15 \cdot 1 + 6 \\
15 &= 6 \cdot 2 + \boxed{3} \\
6 &= 3 \cdot 2 + 0.
\end{aligned}$$

At each step we shift the quotient and remainder diagonally down and left and repeat the process. We stop when we get a remainder of 0. The last nonzero remainder is the gcd, 3 in this case.

We can use this sequence of steps to find the linear combination by working backwards. Start with the second-to-last equation from the Euclidean algorithm and work back up in the following way:

$$\begin{aligned}
3 &= 15 - 6(2) \\
&= 15 - (21 - 15)(2) && \text{replacing 6 with } 21 - 15 \cdot 1 \\
&= 15(3) - (21)(2) && \text{writing in terms of 15 and 21} \\
&= (78 - 21 \cdot 3)(3) - 21(2) && \text{replacing 15 with } 78 - 21 \cdot 3 \\
&= 78(3) - 21(11) && \text{writing in terms of 21 and 78.}
\end{aligned}$$

So we have $3 = 78(3) + 21(-11)$. Here's what happens: In the Euclidean algorithm, we generate the sequence 78, 21, 15, 6, 3 of quotient/remainders. We start with the last equation from the Euclidean algorithm that contains those numbers and solve it for the gcd, 3, in terms of the next two terms of the sequence, 6 and 15 (namely, $3 = 15 - 6(2)$). At the next stage, solve the next equation up for the remainder ($6 = 21 - 15(2)$) and use it to eliminate 6. We then simplify to write things in terms of 15 and 21 and then solve the next equation for the remainder ($15 = 78 - 21(3)$). We use this to eliminate 15, simplify to write things in terms of 21 and 78, and then we are done because the equation is in terms of 21 and 78, which is what we want.

Note that at each step we can check our work by making sure the expression equals the gcd. For instance, in the third line above, $15(3) - 21(2) = 45 - 42 = 3$.

Here is another example. Suppose we want to find integers such that $11x + 41y = 1$. Start with the Euclidean algorithm:

$$\begin{aligned}
41 &= 11 \cdot 3 + 8 \\
11 &= 8 \cdot 1 + 3 \\
8 &= 3 \cdot 2 + 2 \\
3 &= 2 \cdot 1 + \boxed{1}
\end{aligned}$$

Then work backwards as follows:

$$\begin{aligned}
1 &= 3 - 2 \\
&= 3 - (8 - 3 \cdot 2) \\
&= 3(3) - 8(1) \\
&= (11 - 8 \cdot 1)(3) - 8(1) \\
&= 11(3) - 8(4) \\
&= 11(3) - (41 - 11 \cdot 3)(4) \\
&= 11(15) - 41(4)
\end{aligned}$$

Thus we have $x = 15$ and $y = -4$ that give us $11x + 41y = 1$.

This algorithm is called the *extended Euclidean algorithm*. It turns out to have a number of important uses, as we will see. Here is a short Python program implementing a version of it. This is a streamlined version of the algebra above, based on the algorithm given at http://en.wikipedia.org/wiki/Extended_Euclidean_algorithm.

```
def extended_euclid(a, b):
    s, old_s, t, old_t, r, old_r = 0, 1, 1, 0, b, a

    while r != 0:
        q = old_r // r
        old_r, r = r, old_r - q * r
        old_s, s = s, old_s - q * s
        old_t, t = t, old_t - q * t
    return (old_s, old_t)
```

1.7 A few example proofs

The keys to many proofs involving gcds are as follows:

1. Rewrite divisibility statements as equations. For instance, $a \mid b$ becomes $b = ak$ for some integer k.
2. Rewrite $\gcd(a,b) = d$ as a linear combination equation, like $ax + by = d$ for some integers x and y.
3. Algebraically manipulate the equations from (1) and (2).
4. If, at some point, you get a linear combination $au + bv = e$, you can conclude that $\gcd(a,b)$ divides e, but not necessarily that $\gcd(a,b) = e$.

Here are a few example proofs involving gcds:

1. If $d = \gcd(a,b)$, then $\gcd(a/d, b/d) = 1$.

   Proof. The basic idea here is intuitively clear: If we divide through by the gcd, then what's left should not have factors in common, since all the common factors should be in the gcd.

   Here is a more algebraic approach: Start by writing $ax + by = d$ for some integers x and y. We can then divide through by d to get $(a/d)x + (b/d)y = 1$. Theorem 3 tell us that $\gcd(a/d, b/d)$ divides any linear combination of a/d and b/d. So $\gcd(a/d, b/d)$ is a divisor of 1, meaning it must equal 1. □

2. If $d = \gcd(a,b)$, $a \mid c$, and $b \mid c$, then $ab \mid cd$.

   Proof. Write $c = aj$, $c = bk$, and $d = ax + by$ for some integers j, k, x, and y. Multiply the last equation through by c to get $cd = acx + bcy$. Then plug in $c = bk$ into the acx term and $c = aj$ into the bcy term to get $cd = abkx + bajy$. So we have $cd = ab(kx + jy)$, showing that $ab \mid cd$. □

3. If $k > 0$, then $\gcd(ka, kb) = k\gcd(a,b)$.

   Proof. Let $d = \gcd(a,b)$ and $d' = \gcd(ka,kb)$. We can write $d = ax + by$ for some integers x and y. Multiplying through by k gives $kd = kax + kby$. This is a linear combination of ka and kb, so we know that $d' \mid kd$. On the other hand, we can write $d' = kax' + kby'$ for some integers x' and y'. Divide through by k to get $d'/k = ax' + by'$. This is a linear combination of a and b, so $d \mid d'/k$, or $kd \mid d'$. Since $d' \mid kd$ and $kd \mid d'$, and $k > 0$, we must have $d' = kd$. □

4. If $c \mid (a-b)$, then $\gcd(a,c) = \gcd(b,c)$.

   Proof. Let $d_1 = \gcd(a,c)$ and $d_2 = \gcd(b,c)$. From these, we can write $ax_1 + cy_1 = d_1$ and $bx_2 + cy_2 = d_2$ for some integers x_1, x_2, y_1, and y_2. Also, since $c \mid (a-b)$ we can write $a - b = ck$ for some integer k, which we can solve to get $a = ck + b$ and $b = a - ck$. Plugging the former into the equation for d_1 gives $(ck+b)x_1 + cy_1 = d_1$, which we can write as $c(kx_1 + y_1) + bx_1 = d_1$. The left hand side is a linear combination of b and c, so it is a multiple of $d_2 = \gcd(b,c)$. Thus $d_2 \mid d_1$. Plugging $b = a - ck$ into the equation for d_2 and doing a similar computation gives $d_1 \mid d_2$. Thus $d_1 = d_2$. □

1.8 The least common multiple

A relative of the gcd is the *least common multiple*.

Definition 4. *The* least common multiple *(lcm) of two integers a and b, denoted* $\operatorname{lcm}(a,b)$*, is the smallest positive integer which is divisible by both a and b.*

The following gives an easy way to find the lcm:

Theorem 5. *Let a and b be integers. Then*

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

Here an example: If $a=14$ and $b=16$, then $\gcd(a,b)=2$, $ab=224$ and $\operatorname{lcm}(a,b)=224/2=112$. A simple way to think of this theorem is that ab is a multiple of both a and b, but it has some redundant factors in it. Dividing out by $\gcd(a,b)$ removes all of the redundancies, leaving the smallest possible common multiple. Here is a formal proof of the theorem:

Proof. Let $d=\gcd(a,b)$. We need to show that $ab/d=\operatorname{lcm}(a,b)$. We can do this by first showing that ab/d is a multiple of a and b and then showing that no other multiple is smaller than it.

First, since d is divisor of a and b, we can write $a=dk$ and $b=dj$. Then $ab/d=aj$ and $ab/d=bk$, which shows that ab/d is a multiple of both a and b.

Next, let m be any common multiple of a and b. So we have $m=as$ and $m=bt$ for some integers r and s, and we can write $1/a=s/m$ and $1/b=t/m$.

We can also write d as a linear combination $d=ax+by$ for some integers x and y. Dividing both sides of this equation by ab, we get $\frac{d}{ab}=\frac{x}{b}+\frac{y}{a}$. Plugging in our earlier equations for $1/a$ and $1/b$ gives $\frac{d}{ab}=\frac{xt}{m}+\frac{ys}{m}$, which we can rewrite as $m=(xt+ys)\frac{ab}{d}$. This tells us that ad/b is a divisor of m.

Since m is an arbitrary multiple of a and b, and ab/d divides it, that means that ab/d is the smallest possible multiple, which is what we wanted to prove. □

There are other ways to find $\operatorname{lcm}(a,b)$. One way would be to list all the multiples of a and all the multiples of b and find the first one they have in common. This, however, is slow unless a and b are small. Another way uses the prime factorization. See Section 2.1 for an example.

Here is a simple example where the lcm is useful. Suppose one thing happens every 28 days and other happens every 30 days, and that both things happened today. When will they both happen again? The answer is $\operatorname{lcm}(28,30)=420$ days from now.[1]

As another example, some people theorize that the timing of periodic cicadas has to do with the lcm. Some species of cicadas only emerge every 13 years, while others emerge every 17 years. People have noticed that these are values are both prime. This means that the lcm of these numbers with other numbers is relatively large. The things that eat cicadas are often on a boom-bust cycle, and it would be bad for cicadas to emerge in a year when there are a lot of predators. Suppose a certain predator is on a 4-year cycle. How often would a boom year coincide with a 13-year cicada emergence? The answer is every $\operatorname{lcm}(4,13)=52$ years. On the other hand, if cicadas were on say a 14-year cycle, then it would happen every 28 years. It would also be bad for both the 13-year and the 17-year cicadas to emerge at once, since they would be competing for resources. But this will only happen every $\operatorname{lcm}(13,17)=191$ years.

1.9 Relatively prime integers

Definition 5. *Integers a and b are called* relatively prime *(or* coprime*) if* $\gcd(a,b)=1$.

In other words, a and b are relatively prime provided they have no divisors in common besides 1 (or maybe -1 if they are negative). There are a number of facts in number theory that are only true for relatively prime integers. Here is one useful fact that follows quickly from Theorem 3:

Theorem 6. *Integers a and b are relatively prime if and only if $ax+by=1$ for some integers x and y.*

One of the most useful tools in number theory is the following result:

Theorem 7. *(Euclid's Lemma) If $c\mid ab$ with a and c relatively prime, then $c\mid b$.*

[1] A more general approach to handling these kinds of cyclical problems is covered in Section 3.10.

For example, if $c \mid (5 \times 12)$, and c has no factors in common with 5, then in order for it to divide $5 \times 12 = 60$, it must divide 12. On the other hand, if a and c do have factors in common besides 1, then the result might not hold. For instance, if $a = 10$, then $10 \mid (5 \times 12)$, but $10 \nmid 12$.

Proof. Since $\gcd(c, a) = 1$, we can write $cx + ay = 1$ for some integers x and y. We can solve this to get $ay = 1 - cx$. Further, since $c \mid ab$, we can write $ck = ab$ for some integer k. Multiply both sides by y and plug in $ay = 1 - cx$ to get $cky = (1 - cx)b$. We can rearrange this to get $c(ky + bx) = b$, which tells us that $c \mid b$, as desired. □

1.10 The gcd and lcm of more than two integers

The concept of gcd can be applied to more than two integers. Namely, $\gcd(a_1, a_2, \ldots, a_n)$ is the largest integer that divides each of the a_i. For instance, $\gcd(24, 36, 60) = 12$. The gcd can be computed from the Euclidean algorithm and the following fact:

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

We can repeatedly apply this rule if needed. For instance, to compute $\gcd(14, 28, 50, 77)$, we can compute $\gcd(14, 28) = 14$, then $\gcd(14, 50) = 2$ and finally, $\gcd(2, 77) = 1$. So $\gcd(14, 28, 50, 77) = 1$.

It is also possible to compute the gcd by extending the ideas of the Euclidean algorithm. We perform several modulos at each step, always modding by the smallest value. Here is some (slightly tricky) Python code implementing these ideas:

```python
def gcd(*A):
    while A[0] != 0:
        A = sorted([A[0]] + [x % A[0] for x in A[1:]])
    return A[-1]
```

One thing to be careful of is that it is possible to have $\gcd(a, b, c) = 1$, but not have $\gcd(a, b)$ and $\gcd(b, c)$ both equal to 1. For instance, $\gcd(2, 4, 5) = 1$ since 1 is the largest integer dividing 2, 4, and 5, but $\gcd(2, 4) \neq 1$.

For many theorems that require a bunch of integers, a_1, a_2, ..., a_n to not have any factors in common, instead of requiring $\gcd(a_1, a_2, \ldots, a_n) = 1$, often the following notion is used:

Definition 6. *Integers a_1, a_2, ..., a_n are said to be* pairwise relatively prime *if* $\gcd(a_i, a_j) = 1$ *for all $i, j = 1, 2, \ldots, n$ with $i \neq j$.*

The lcm of integers a_1, a_2, ..., a_n is the smallest integer that is a multiple of each of the a_i. Similarly to the gcd, the lcm can be computed by using the rule below to break things down.

$$\text{lcm}(a, b, c) = \text{lcm}(\text{lcm}(a, b), c).$$

Here is some Python code for the lcm:

```python
def lcm(*X):
    m = 1
    for a in X:
        m = a*m // gcd(a, m)
    return m
```

1.11 Some useful facts about divisibility and gcds

Here is a list of facts that might come in handy from time to time. It's not worth memorizing this list, but it may be useful to refer back to if you need a certain fact for something you are working on.

Most important facts

1. Euclid's lemma: If $a \mid bc$ with a and b relatively prime, then $a \mid c$.
2. If $d = \gcd(a, b)$, then $d = ax + by$ for some integers x and y.
3. Any linear combination of a and b is a multiple of $\gcd(a, b)$.
4. Integers a and b are relatively prime if and only $ax + by = 1$ for some integers x and y.

Divisibility

5. If $a \mid b$ and $b \mid c$, then $a \mid c$.
6. If $a \mid b$ and $c \mid d$, then $ac \mid bd$.
7. If $a \mid b$ and $a \mid c$, then $a \mid (bx + cy)$ for any integers x and y.
8. $a \mid b$ and $b \mid a$ if and only if $a = b$ or $a = -b$.

Gcds

9. If $c \mid a$ and $c \mid b$, then $c \mid \gcd(a, b)$.
10. If $d|a$ and $d|b$, then $d = \gcd(a, b)$ if and only if $\gcd(a/d, b/d) = 1$.
11. $\gcd(ka, kb) = |k| \gcd(a, b)$ for any integer $k \neq 0$.
12. Let $d = \gcd(a, b)$. If $a \mid c$ and $b \mid c$, then $ab \mid cd$.
13. $\gcd(a + bc, b) = \gcd(a, b)$ for any integer c.
14. If $\gcd(a, b) = 1$, then $\gcd(c, ab) = \gcd(c, a) \gcd(c, b)$.
15. If $a \mid bc$, then $a/\gcd(a, b)$ is a divisor of c.
16. $\gcd(a, a + n) \mid n$. In particular, $\gcd(a, a + 1) = 1$.
17. $\gcd(a, b) \operatorname{lcm}(a, b) = ab$.

Chapter 2

Primes

Prime numbers are one of the main focuses of number theory.

Definition 7. *An integer greater than 1 is called* prime *if its only divisors are 1 and itself. It is called* composite *otherwise.*

Notice that 1 is not considered to be prime. The reason is that primes are thought of as fundamental building blocks of numbers. As we will soon see, every number is a product of primes, each prime helping to build up the number. However, 1 doesn't do any building, as multiplying by 1 doesn't accomplish anything. There are other ways in which 1 behaves differently from prime numbers, and for these reasons 1 is not considered prime.

The first few primes are 2, 3, 5, 7, 11, 13, 17, 19. Much of number theory is concerned with the structure of the primes—how frequent they are, gaps between them, whether there is any sort of pattern to them, etc.

Euclid's lemma

Recall Euclid's lemma from Section 1.9. It states that if $c \mid ab$ and $\gcd(c, a) = 1$, then $c \mid b$. If a is prime, then $\gcd(c, a) = 1$, so Euclid's lemma holds whenever p is prime. Here is Euclid's lemma restated for primes:

Theorem 8. *(Euclid's lemma) If p is prime and $p \mid ab$, then $p \mid a$ or $p \mid b$.*

Euclid's lemma could be used as an alternate definition for prime numbers as it is not too hard to show that no other number besides 1 has this property. In fact, Euclid's lemma is used to define analogs of prime numbers (like prime ideals) in abstract algebra.

Using induction, Euclid's lemma can be extended as follows:

Corollary 9. *If p is prime and $p \mid a_1 a_2 \dots a_n$, then $p \mid a_i$ for some $i = 1, 2, \dots n$.*

Proof. We proceed by induction. The base case, $n = 2$, is Euclid's lemma. Now assume the statement holds for n and suppose $p \mid a_1 a_2 \dots a_n a_{n+1}$. We can write $a_1 a_2 \dots a_n a_{n+1}$ as $(a_1 a_2 \dots a_n)(a_{n+1})$ and by Euclid's lemma, either $p \mid a_{n+1}$ or $p \mid a_1 a_2 \dots a_n$. In the latter case, by the induction hypothesis, $p \mid a_i$ for some $i = 1, 2, \dots, n$. So overall, $p \mid a_i$ for some $i = 1, 2, \dots, n+1$. Thus the result is true by induction. □

A direct consequence of this is the following:

Corollary 10. *If p is prime and $p \mid q_1 q_2 \dots q_n$, where $q_1, q_2, \dots, q_n$ are all prime, then $p = q_i$ for some $i = 1, 2, \dots, n$.*

Euclid's lemma is one of the most important tools in elementary number theory and we will see it appear again and again.

2.1 The fundamental theorem of arithmetic

In math there are a number of "fundamental theorems." There is the fundamental theorem of algebra which states that every nonconstant polynomial has a root, the fundamental theorem of calculus that relates integration to differentiation, and in number theory, there is the *fundamental theorem of arithmetic*, which states that every integer greater than 1 can be factored uniquely into primes. For instance, we can factor 60 into 2 × 2 × 3 × 5 and there is no other product of primes equal to 60, other than changing the order of 2 × 2 × 3 × 5. Here is the formal statement of the theorem:

Theorem 11. *Every integer $n > 1$ can be written uniquely as a product of primes.*

Here is intuitively why we can write n as a product of primes: Either n itself is prime (in which case we are done) or else n can be factored into a product ab. These integers are either prime or they themselves can be factored. These new factors are in turn either prime or they can be factored. We can continue this process, but eventually it must stop since the factors of a number are smaller than the number itself, and things can't keep getting smaller forever. This can be made formal using induction.

Proof. We use strong induction to show that each number can be written as a product of primes. The base case $n = 2$ is clear. Now assume that each integer $1, 2, \dots, n-1$ can be written as a product of primes. Now either n is prime, in which case n is trivially a product of primes, or else we can write $n = ab$ for some integers a and b in the range from 1 to $n-1$. By the induction hypothesis, a and b can be written as products of primes, so $n = ab$ is a product of primes. Thus the result is true by induction.

To show the representation is unique, suppose $n = p_1 p_2 \dots p_k$ and $n = q_1 q_2 \dots q_m$ are two different representations. From the first representation, we have $p_1 \mid n$, and so $p_1 \mid q_1 q_2 \dots q_m$. By Corollary 10, we must have $p_1 = q_i$ for some i. By rearranging terms, we can assume $i = 1$, so $p_1 = q_1$. By the same argument, we can similarly conclude that $p_2 = q_2$, $p_3 = q_3$, etc. Thus the two representations are the same. □

In the factorization, some of the primes may be the same, like in $720 = 2 \times 2 \times 2 \times 2 \times 3 \times 3 \times 5$. We can gather those factors up and write the factorization as $2^3 \cdot 3^2 \cdot 5$. In general, we can always write an integer $n > 1$ as a unique product of the form $p_1^{e_1} p_2^{e_2} \dots p_k^{e_k}$.

Applications of the fundamental theorem

1. The gcd and lcm can be computed easily from the prime factorization. For example, suppose we have $a = 168$ and $b = 180$. We have $168 = 2^3 \cdot 3 \cdot 7$ and $180 = 2^2 \cdot 3^2 \cdot 5$.

   To get the gcd, we go prime-by-prime through the two representations, always taking the lesser of the two amounts. For instance, both 168 and 180 have a factor of 2: 168 has 2^3 and 180 has 2^2, and so we use the lesser factor, 2^2, in the gcd. Moving on to the factor 3, 168 has 3^1 and 180 has 3^2, so we take 3^1. The next factors are 5 and 7, but they are not common to both 168 and 180, so we ignore them. We end up with $\gcd(168, 180) = 2^2 \cdot 3 = 12$.

   The lcm is done similarly, except that we always take the larger amount. We get $\operatorname{lcm}(168, 180) = 2^3 \cdot 3^2 \cdot 5 \cdot 7 = 2520$.

   Using the prime factorization to find the gcd and lcm is fast if we have the factorization is available. However, finding the prime factorization is a slow process for large numbers. The Euclidean algorithm is orders of magnitude faster.[1]

2. The fundamental theorem is a useful tool in proofs. For instance, let's prove that if n is a perfect square with $n = ab$ and $\gcd(a, b) = 1$, then a and b are perfect squares.

   We can write $n = m^2$ for some m and assume m has the prime factorization $m = p_1^{e_1} p_2^{e_2} \dots p_k^{e_k}$. Then

   $$n = m^2 = p_1^{2e_1} p_2^{2e_2} \dots p_k^{2e_k}.$$

   Since $n = ab$, the prime factorization of a includes some of these primes and the prime factorization of b includes the rest of them (or possibly a or b equals 1, in which case the result is trivial). But if the prime factorization of a includes a prime p_i, then the factorization of b cannot include p_i since $\gcd(a, b) = 1$. Thus, after possibly reordering, there exists some integer j such that we can break the prime factorization of n up into $a = p_1^{2e_1} p_2^{2e_2} \dots p_j^{2e_j}$ and $b = p_{j+1}^{2e_{j+1}} p_{j+2}^{2e_{j+2}} \dots p_k^{2e_k}$. Thus $a = (p_1^{e_1} p_2^{e_2} \dots p_j^{e_j})^2$ and $b = (p_{j+1}^{e_{j+1}} p_{j+2}^{e_{j+2}} \dots p_k^{e_k})^2$ are perfect squares.

3. As another example, let's prove that if $a \mid c$ and $b \mid c$ with $\gcd(a, b) = 1$, then $ab \mid c$.

   Since $a \mid c$, every term, $p_i^{e_i}$, in the factorization of a occurs in the factorization of c. Similarly, since $b \mid c$, every term in the factorization of b occurs in the factorization of c. Since $\gcd(a, b) = 1$, the primes in the factorization of a must be different from the primes in the factorization of b. Thus every term in the factorization of ab occurs in the factorization of c. So $ab \mid c$.

[1] A little more formally, the Euclidean algorithm's running time grows linearly with the number of digits in the number, whereas the running times of the fastest known factoring algorithms grow exponentially with the number of digits.

2.2 There are infinitely many primes

One of the first things we might wonder about prime numbers is how many there are. The ancient Greeks answered this question with a proof somewhat like the one below.

Theorem 12. *There are infinitely many primes.*

Proof. Let $p_1, p_2, \dots p_n$ be primes, and define $P = p_1 p_2 \dots p_n + 1$. By the fundamental theorem, P must be divisible by some prime, and that prime must be different from $p_1, p_2, \dots, p_n$ since for any $i = 1, 2, \dots, n$ the fact that p_i divides P means that p_i cannot divide $P + 1$. Thus, given any list of primes, we can use the list to generate a new prime, meaning the number of primes is infinite. □

An alternate way to do the above proof would be to assume that there were only finitely many primes and to use the process above to derive a contradiction. A quick web search will turn up dozens of other interesting proofs of the infinitude of primes.

It is worth noting that the integer P in the proof above need not be prime itself. It only needs to be divisible by a new prime. Numbers of the form given in the proof above are sometimes called *Euclid numbers* or *primorials*. The first few are

$$
\begin{aligned}
&E_1 = 2 + 1 = 3\\
&E_2 = 2 \cdot 3 + 1 = 7\\
&E_3 = 2 \cdot 3 \cdot 5 + 1 = 31\\
&E_4 = 2 \cdot 3 \cdot 5 \cdot 7 + 1 = 211\\
&E_5 = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 + 1 = 2311\\
&E_6 = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 + 1 = 30031
\end{aligned}
$$

The first five Euclid numbers are prime, but $E_6 = 59 \times 509$ is not. It is not currently known whether infinitely many Euclid numbers are prime. There are not many Euclid numbers that are known to be prime. The next few that are prime are E_7, E_{11}, E_{31}, and E_{379}. Note that E_{379} is already a large number, having about 4300 digits.

2.3 Finding primes

One of the simplest ways to check if a number is prime is *trial division*. Just check to see if it is divisible by any of the integers 2, 3, 4, etc. When considering divisors of n, they come in pairs, one less than or equal to $\sqrt{n}$ and the other greater than or equal to $\sqrt{n}$. For instance, if $n = 30$, we have $\sqrt{30} \approx 5.47$ and we can write 30 as 2×15, 3×10, and 5×6, with 2, 3, and 5 less than $\sqrt{30}$ and 6, 10, 15 greater than $\sqrt{30}$. So, in general, we can stop checking for divisors at $\sqrt{n}$.

This process can be made more efficient by just checking for divisibility by 2 and by odd numbers, or better yet, checking for divisibility only by primes (provided the number is small or we have a list of primes). For example, to check if 617 is prime, we have $\sqrt{617} = 24.84$ and we check to see if it is divisible by 2, 3, 5, 7, 11, 13, 17, 19, and 23. It is not divisible by any of those, so it is prime.

This approach is reasonably fast for small numbers, but for checking the primality of larger numbers, like ones with several hundred digits, there are faster techniques, which we will see later.

To find *all* the primes less than an integer n, we can use a technique called the *sieve of Eratosthenes*. Start by listing the integers from 2 to n and cross out all the multiples of 2 after 2. Then cross out all the multiples of 3 after 3. Then cross out all the multiples of 5 after 5. Note that we don't have to cross out the multiples of 4 since they have all already been crossed out as they are multiples of 2. We keep going, crossing out multiples of 7, 11, 13, etc., until we get to $\sqrt{n}$. At the end, only the primes will be left. Here is what we would get for $n = 100$:

	2	**3**	4	**5**	6	**7**	8	9	10
11	12	**13**	14	15	16	**17**	18	**19**	20
21	22	**23**	24	25	26	27	28	**29**	30
31	32	33	34	35	36	**37**	38	39	40
41	42	**43**	44	45	46	**47**	48	49	50
51	52	**53**	54	55	56	57	58	**59**	60
61	62	63	64	65	66	**67**	68	69	70
71	72	**73**	74	75	76	77	78	**79**	80
81	82	**83**	84	85	86	**87**	88	**89**	90
91	92	93	94	95	96	**97**	98	99	100

Here is how we might code it in Python. We create a list of zeroes and ones, with a zero at index i meaning i is not prime and a one meaning i is prime. The list starts off initially with all ones and we gradually cross off all the composites.

```python
def sieve(n):
    L = [0,0] + [1]*(n-1)
    p = 2
    while p <= n**.5:
        while L[p]==0:
            p = p + 1
        for j in range(2*p,n+1,p):
            L[j] = 0
        p += 1
    return [i for i in range(len(L)) if L[i]==1]
```

The sieve works relatively well for finding small primes. The code above, inefficient though it may be, takes 55 seconds to find all the primes less than 10^8 on my laptop.

2.4 The prime number theorem

A natural question to ask is how common prime numbers are. An answer to this question and some other questions is given by the *prime number theorem*, which states that the number of primes less than n is roughly $n/\ln n$. Formally, we can state it as follows:

Theorem 13. *(Prime number theorem) Let $\pi(n)$ denote the number of primes less than or equal to n. Then*

$$\lim_{n\to\infty} \frac{\pi(n)}{n/\ln n} = 1.$$

For example, for $n =$1,000,000, we have $\pi(n) = 78498$ and $n/\ln n = 72382$. The prime number theorem's estimate is off by about 8% here. A more accurate estimate is $n/(\ln(n) - 1)$, which in this case gives 78030.

The theorem tells us that roughly $100/\ln n$ percent of the numbers less than n are prime. For $n =$1,000,000, the theorem tells us that roughly 7-8% of the numbers less than 1,000,000 are prime, and that around $n =$1,000,000 the average gap between primes is roughly $\ln(1000000) \approx 14$.

The prime number theorem is not easy to prove. The first proofs used sophisticated techniques from complex analysis. There is a proof using only elementary methods, but it is fairly complicated.

A more accurate estimate

The prime number theorem tells us that the probability an integer near x is prime is roughly $1/\ln x$. Summing up all these probabilities for all real x from 2 through n gives us another estimate for the number of primes less than n. Such a sum is a continuous sum (an integral), and we get the following estimate for $\pi(n)$:

$$\pi(n) \approx \int_2^n \frac{1}{\ln x}\,dx.$$

The integral above is called the *logarithmic integral* and is denoted $\mathrm{Li}(n)$. This integral predicts 78,628 primes less than 1,000,000, which is only 130 off from the correct value.

An interesting note about this is that for n up through at least 10^{14} it has been shown that $\mathrm{Li}(n) < \pi(n)$. But it turns out not to be true for all n. In fact, it was proved in 1933 that $\mathrm{Li}(n) - \pi(n)$ changes sign infinitely often. This illustrates an important lesson in number theory: just because something is true for the first trillion (or more) integers, does not mean it is true in general.

The proof is interesting in that it included one of the largest numbers to ever be used in a proof, namely that $\pi(n) < \mathrm{Li}(n)$ for some value less than $e^{e^{e^{79}}}$. More recent research has brought the bound down to e^{728}.

2.5 Twin primes

Twin primes are pairs $(p, p+2)$ with both p and $p+2$ prime. Examples include $(5,7)$, $(11,13)$ and $(41,43)$.

One of the most famous open problems in math is the *twin primes conjecture*, which asks if there are infinitely many twin primes, Most mathematicians think the conjecture is true, though it is considered to be a very difficult problem.

Referring back to the prime number theorem, we know that the probability an integer near x is prime is roughly $1/\ln x$. Assuming independence, the probability that both x and $x+2$ would be prime is then $1/(\ln x)^2$ and summing up these probabilities gives $\int_2^n \frac{1}{(\ln x)^2}\,dx$ as an estimate for the number of twin primes pairs less than n. Independence is not quite a valid assumption here, but it is not too far off. It is currently conjectured that the number of twin prime pairs less than n is approximately $2C_2 \int_2^n \frac{1}{(\ln x)^2}\,dx$, where $C_2 \approx .66016$ is something called the twin prime constant.[1]

So it seems reasonable that there are infinitely many twin primes, but it has turned out to be very difficult to prove. The best result so far is that there are infinitely many pairs $(p, p+2)$ where p is prime and $p+2$ is either prime or the product of two primes, proved by Chen Jingrun in 1973.

There are a number of analogous conjectures. For instance, it is conjectured that there are infinitely many pairs of primes of the form $(p, p+4)$ or infinitely many triples of the form $(p, p+2, p+6)$. Recent work has shown that there are infinitely many primes p such that one of $p+2, p+4, \ldots p+246$ is also prime.

2.6 Prime gaps

It is interesting to look at the gaps between consecutive primes. Here are the gaps between the first 100 primes, from 2 to 541:

1, 2, 2, 4, 2, 4, 2, 4, 6, 2, 6, 4, 2, 4, 6, 6, 2, 6, 4, 2, 6, 4, 6, 8, 4, 2, 4, 2, 4, 14,
4, 6, 2, 10, 2, 6, 6, 4, 6, 6, 2, 10, 2, 4, 2, 12, 12, 4, 2, 4, 6, 2, 10, 6, 6, 6, 2, 6,
4, 2, 10, 14, 4, 2, 4, 14, 6, 10, 2, 4, 6, 8, 6, 6, 4, 6, 8, 4, 8, 10, 2, 10, 2, 6, 4,
6, 8, 4, 2, 4, 12, 8, 4, 8, 4, 6, 12, 2, 18

We see gaps of 2 quite often. These correspond to twin primes. There are a few larger gaps, like a gap of 14 from 113 to 127 and a gap of 18 from 523 to 541. It is not too hard to find gaps that are arbitrarily large. Just find an integer n divisible by $2, 3, 4, 5, \ldots, k$ and then $n+2, n+3, \ldots n+k$ will all be composite.

The prime number theorem tells us that the average gap between a prime p and the next prime is approximately $\ln p$. Thus for p near 1,000,000, we would expect an average gap of about 14, and for $p = 10^{200}$, we would expect an average gap of around 460.

One important result is relating to prime gaps is *Bertrand's postulate*.

Theorem 14. *(Bertrand's postulate) For any integer $n > 1$ there exists a prime p such that $n < p < 2n$.*

For example, for $n = 1000$, we are guaranteed that there is a prime between 1000 and 2000. This is not news, but it is useful in some cases to have an interval on which you are guaranteed to have a prime, even if that interval is rather large.

The proof of Bertrand's postulate is actually not too difficult, but we won't cover it here. There are a number of improvements on Bertrand's postulate. For instance, in 1952 it was proved that for $n \geq 25$ there exists a prime between n and $(1 + \frac{1}{5})n$. The range can be narrowed further for larger n.[2]

A related, but unsolved, question is if there is always a prime between consecutive perfect squares, n^2 and $(n+1)^2$.

2.7 Finding large primes

A popular sport among math enthusiasts is finding large prime numbers. The largest primes known are all *Mersenne primes*, which are primes of the form $2^n - 1$. There is a relatively fast algorithm for checking if a number of the form $2^n - 1$ is prime,

[1] See Section 1.2 of *Prime Numbers: A Computational Perspective*, 2nd edition by Crandall and Pomerance for more on this approach.

[2] In general, it has been proved that for any $\epsilon > 0$ there is an integer N such that for $n \geq N$, there is always a prime between n and $(1+\epsilon)n$. In fact, as $n \to \infty$, the number of primes in that range approaches ∞ as well. In math, there are many results concerning how things behave as n approaches ∞. Results of this sort are called *asymptotic results*.

and that is why people searching for large primes use these numbers. As of early 2016, the largest known prime is $2^{74207281}-1$, a number over 22 million digits long.

Most of the largest primes found recently were found by the Great Internet Mersenne Prime Search (or GIMPS), where volunteers from around the world donate their spare CPU cycles towards checking for primes. Finding large primes usually involves either a combination of sophisticated algorithms and finely-tuned hardware or a distributed computer search like GIMPS.

People also look for large primes of special forms. For instance, as of early 2016, the largest known twin primes are $3756801695685 \cdot 2^{666669} \pm 1$, about 200,000 digits long. See http://primes.utm.edu/largest.html for a nice list of large primes.

2.8 Prime-generating formulas

One of the most remarkable polynomials in all of math is $p(n) = n^2 + n + 41$. It has the property that for each $n = 0, 1, \ldots 39$, $p(n)$ is prime. However, $p(40)$ and $p(41)$ are not prime as $p(40) = 41^2$ and $p(41) = 41^2 + 41 + 41$ is clearly divisible by 41. Still, the polynomial keeps on generating primes at a pretty high rate as $p(n)$ is prime for 34 of the next 39 values of n. In total, 156 of the first 200 values of $p(n)$ are prime, and 581 of the first 1000.

There are a number of other polynomials that are good at generating primes, like $n^2 + n + 11$ and $n^2 + n + 17$, though neither of these is quite as good as $n^2 + n + 41$. The formula $n^2 - 79n + 1601$ generates primes for each integer from $n = 0$ through $n = 79$. It is actually a modification of $n^2 + n + 41$: namely, $n^2 - 79n + 1601 = (n - 40)^2 + (n - 40) + 41$. The 80 primes are the same as the 40 primes from $n^2 + n + 41$, each appearing twice.

The site http://mathworld.wolfram.com/Prime-GeneratingPolynomial.html has a nice list of some other prime-generating polynomials.

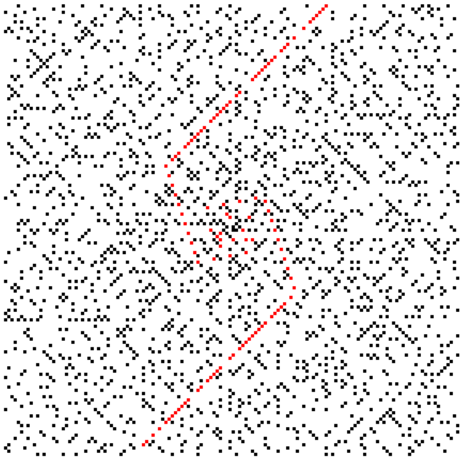
Prime spirals

There is a nice way to visualize prime numbers known as a *prime spiral* or *Ulam spiral*. Start with 1 in the middle, and spiral out from there, like in the figure below on the left. Then highlight the primes, like on the right.

37	36	35	34	33	32	31
38	17	16	15	14	13	30
39	18	5	4	3	12	29
40	19	6	1	2	11	28
41	20	7	8	9	10	27
42	21	22	23	24	25	26
43	44	45	46	47	48	49

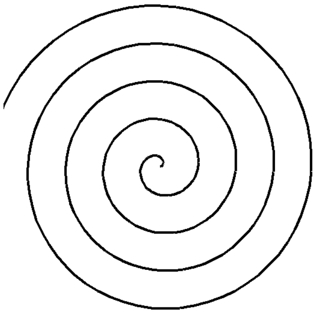
37	36	35	34	33	32	**31**
38	**17**	16	15	14	**13**	30
39	18	**5**	4	**3**	12	**29**
40	**19**	6	1	**2**	**11**	28
41	20	**7**	8	9	10	27
42	21	22	**23**	24	25	26
43	44	45	46	**47**	48	49

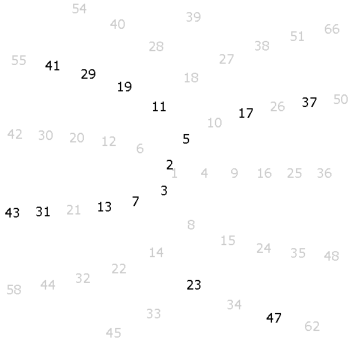
If we expand the view to the first 20,000 primes, we get the figure below.



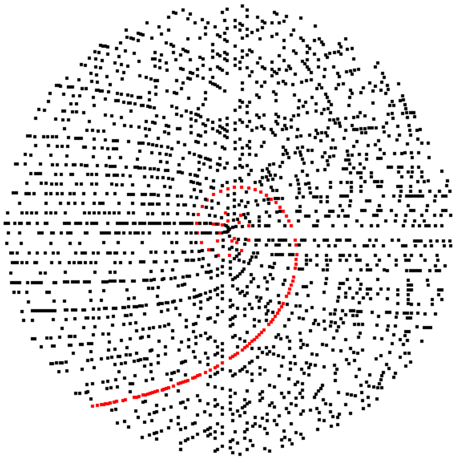
Notice the primes tend to cluster along certain diagonal lines. The dots highlighted in red correspond to Euler's polynomial, n^2+n+41.

An interesting twist on this is something called the *Sacks spiral*. Instead of the rectangular spiral we used above, we instead spiral along an Archimedean spiral, where both the angular and radial velocity are constant, with those constants chosen so that the perfect squares lie along the horizontal axis. Here is a typical Archimedean spiral and the Sacks spiral for the first few primes:





And here it is for a wider range. Euler's polynomial, n^2+n+41, is highlighted in red.



More with primes and polynomials

It is not too difficult to show that there is no nonconstant polynomial that can give only primes. Just plug in the constant term. For instance, if $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, then every term of $p(a_0)$ is divisible by a_0. Thus $p(a_0)$ will be composite, except possibly when $a_0 = 0$ or ± 1. If $a_0 = \pm 1$, then $p(0)$ is not prime, and if $a_0 = 0$, factor out an x and try plugging in the new constant term.

An interesting open problem is whether or not there are infinitely many primes of the form n^2+1.

Other formulas

There are some other interesting formulas for generating primes. For instance, it turns out that there exists a real number r such that $\lfloor r^{3^n} \rfloor$ is prime for any integer n. The exact value of r is unknown, but it is thought to be approximately 1.30637788. Another example is the recurrence $x_n = x_{n-1} + \gcd(n, x_{n-1})$, $x_1 = 7$. The difference between consecutive terms is always either 1 or a prime.

2.9 Primes and arithmetic progressions

One interesting result is Dirichlet's theorem, stated below:

Theorem 15. *(Dirichlet's theorem) If* $\gcd(a,b)=1$*, then there are infinitely many primes of the form* $ak+b$*.*

For example, with $a=4$ and $b=3$, the theorem tells us that there are infinitely many primes of the form $4k+3$. If $a=100$ and $b=1$, the theorem tells us there are infinitely many primes of the form $100k+1$ (i.e., primes that end in 01). The first few are 101, 401, 601, 701, 1201, ... Like the prime number theorem, Dirichlet's theorem is difficult to prove, relying on techniques from analytic number theory.

A related theorem is the Green-Tao theorem, proved in 2004. It concerns arithmetic progressions of primes, sequences of the form p, $p+a$, $p+2a$, $\dots p+(n-1)a$, all of which are prime. In other words, we are looking at sequences of equally spaced primes, like $(3,5,7)$ or $(5,11,17,23,29)$. The Green-Tao theorem states that it is possible to find prime arithmetic progressions of any length. The proof, like many in math, is an existence proof. It shows that these progressions exist but doesn't tell how to find them. According to the Wikipedia page on the Green-Tao theorem, as of 2010 the longest known arithmetic progression was 25 terms long, starting at the integer 43142746595714191.

2.10 Fermat numbers

In the 1600s, Pierre de Fermat studied primes of the form $2^{2^n}+1$. Numbers of this form are now called *Fermat numbers*. The first one is $2^{2^0}+1=3$, and the next few are 5, 17, 257, and 65537. These are all prime, and Fermat conjectured that every Fermat number is prime. However, the next one, $F_5=4294967297$, turns out to have a factor of 641.

The remarkable fact is that there are no other Fermat numbers that are known to be prime. In fact, it is an open question as to whether any other Fermat numbers are prime.

Considerable effort has gone into trying to factor Fermat numbers. This is difficult because of the shear size of the numbers. As of early 2014, F_5 through F_{11} have been completely factored. Partial factorizations have been found for many other Fermat numbers. The smallest Fermat number which is not known to be composite is F_{33}. See www.prothsearch.net/fermat.html for a comprehensive list of results.

2.11 Sophie Germain primes

A Sophie Germain prime is a prime p such that $2p+1$ is also prime. For instance, 11 is a Sophie Germain prime because $2\cdot 11+1=23$ is also prime. The first few Sophie Germain primes are 2, 3, 5, 11, 23, 29, 41, 53, 83, 89. It is not currently known if there are infinitely many, though it is thought that there are.[1]They are named for the 19th century mathematician Sophie Germain, who used them in her work on Fermat's Last Theorem.[2]

If p is a Sophie Germain prime, then $2p+1$ is also prime by definition, and is called a *safe prime*. Safe primes are important in modern cryptography. See Section 4.1.

It is also interesting to create chains where p, $2p+1$, $2(2p+1)+1$, etc. are all prime. Such chains are called *Cunningham chains*. One such chain is 2, 5, 11, 23, 47. It can't be extended any further as $2\cdot 47+1=95$ is not prime. It is thought that there are infinitely many chains of all lengths, but no one knows for sure. According to Wikipedia, the longest chain so far found is 17 numbers long, starting at 2,759,832,934,171,386,593,519.

2.12 Goldbach's conjecture

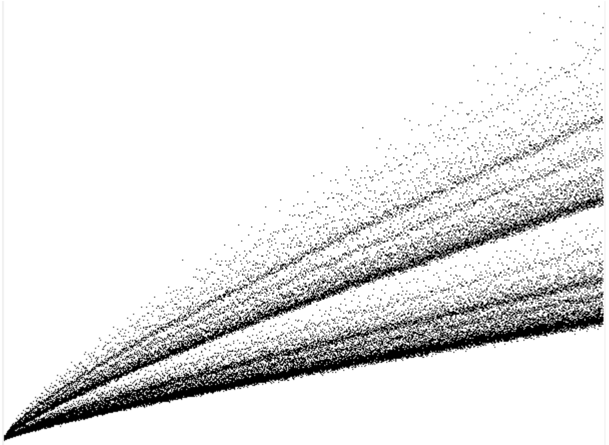
Goldbach's conjecture is one of the most famous open problems in math. It simply states that any even number greater than two is the sum of two primes. For instance, we can write $4=2+2$, $6=3+3$, $8=5+3$, and $10=5+5$ or $7+3$.

[1] In fact, it is suspected that there are about as many Sophie Germain primes as there are twin primes pairs.

[2] Fermat's last theorem states that there are no integer solutions to $x^n+y^n=z^n$ if $n>2$. It was one of the most famous problems in math for a few hundred years.

Goldbach's conjecture has been verified numerically by computer searches up through about 10^{18}. The number of ways to write an even number as a sum of two primes seems to increase quite rapidly. For instance, numbers between 2 and 100 have an average of about four ways to be written as sums of primes. This increases to 18 for numbers between 100 and 1000, 93 for numbers between 1000 and 10,000, and 554 for numbers between 10,000 and 100,000. Moreover, in the range from 10,000 to 100,000 no number can be written in less than 92 ways.

Here is a graph showing how the number of possible ways to write a number as a sum of two primes increases with n. The horizontal axis runs from $n = 4$ to $n =$100,000 and the vertical axis runs to about 2000.



Despite the overwhelming numerical evidence, the Goldbach conjecture is still far from being proved. However, there are a number of partial results. For instance, in the early 1970s Chen Jingrun proved that every sufficiently large even number can be written as a sum $p+q$, where p is prime and q is either prime or a product of two primes.

There is also the *weak Goldbach conjecture* that states that every odd number greater than 7 is the sum of three primes. In the 1930s, I.M. Vinogradov proved that it was true for all sufficiently large integers. It seems that the weak Goldbach conjecture may have been proved in 2013 by Harald Helfgott, though as of this writing, the proof has not been fully checked. Like, Vinogradov's result, Helfgott proved the result true for all sufficiently large integers, but in this case "sufficiently large" was small enough that everything less than it could be checked by computer.

2.13 Some sums

One of the most important functions in analytic number theory is the *zeta function*,

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

For example,

$$\begin{aligned}
\zeta(1) &= \sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots \text{ (diverges)}\\
\zeta(2) &= \sum_{n=1}^{\infty} \frac{1}{n^2} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots = \frac{\pi^2}{6}\\
\zeta(3) &= \sum_{n=1}^{\infty} \frac{1}{n^3} = 1 + \frac{1}{2^3} + \frac{1}{3^3} + \frac{1}{4^3} + \cdots \approx 1.202\\
\zeta(4) &= \sum_{n=1}^{\infty} \frac{1}{n^4} = 1 + \frac{1}{2^4} + \frac{1}{3^4} + \frac{1}{4^4} + \cdots = \frac{\pi^4}{90}.
\end{aligned}$$

Above we have $\zeta(1)$, the well-known harmonic series. We also have $\zeta(2)$ and $\zeta(4)$, whose sums were famously determined by Euler. In general, the even values, $\zeta(2n)$, always sum to some constant times π^{2n}, whereas not too much is known about any of the odd values, except for the harmonic series.

It has been known since at least the 14th century that the harmonic series is divergent. A particularly nice proof of this fact is shown below:

$$\begin{aligned}
&1+\underbrace{\frac{1}{2}+\frac{1}{3}}+\underbrace{\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}}+\underbrace{\frac{1}{9}+\frac{1}{10}+\frac{1}{11}+\frac{1}{12}+\frac{1}{13}+\frac{1}{14}+\frac{1}{15}+\frac{1}{16}}+\dots\\
&\le 1+\left(\frac{1}{2}+\frac{1}{2}\right)+\left(\frac{1}{4}+\frac{1}{4}+\frac{1}{4}+\frac{1}{4}\right)+\left(\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}\right)+\dots\\
&= 1+1+1+1+1+\dots.
\end{aligned}$$

A remarkable fact is that the Nth partial sum of the harmonic series is nearly equal to $\ln(N)$. For instance, we have the following:

N	$\sum_{n=1}^{N}\frac{1}{n}$	$\ln(N)$	Difference
100	5.187378	4.605170	.582207
10,000	9.787606	9.210340	.577266
1,000,000	14.392727	13.815511	.577216
100,000,000	18.997896	18.420681	.577216

In general, we have that $\sum_{n=1}^{N}\frac{1}{n}-\ln(N)$ converges to $\gamma\approx .5772156649$, the *Euler-Mascheroni constant*. This is one of the most famous constants in math, showing up in a number of places in higher mathematics. It is actually not known whether γ is rational or irrational.

Euler also discovered a beautiful relationship between the harmonic series and prime numbers:

$$\sum_{n=1}^{\infty}\frac{1}{n}=\prod_{p\text{ prime}}\frac{1}{1-\frac{1}{p}}.$$

In other words,

$$1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\cdots=\frac{1}{2}\times\frac{2}{3}\times\frac{4}{5}\times\frac{6}{7}\times\frac{10}{11}\times\cdots.$$

Euler then used this fact to show that the following series diverges:

$$\sum_{p\text{ prime}}\frac{1}{p}.$$

This fact gives another proof that there are infinitely many primes, since if there were only finitely many primes, then the series would have to converge. It can in fact be shown that, analogously to the harmonic series, the partial sums of this series satisfy the following:

$$\sum_{p\le N\text{ prime}}\frac{1}{p}-\ln(\ln(N))\to .261497.$$

The constant .261497 is called the Meissel-Mertens constant.

It is interesting to note that a similar sum involving twin primes is actually convergent, namely

$$\sum_{p,\,p+2\text{ prime}}\left(\frac{1}{p}+\frac{1}{p+2}\right)\approx 1.902.$$

The constant 1.902 is called Brun's constant.[1]

2.14 The Riemann hypothesis

Perhaps the most famous unsolved problem in mathematics is the *Riemann hypothesis*, first stated by Bernhard Riemann in 1859. It is a statement involving the zeta function. A process called analytic continuation is used to find a function defined

[1] A 1994 calculation of the constant was responsible for finding a bug in the Pentium processor. See the article *How Number Theory Got the Best of the Pentium Chip* in the January 13, 1995 issue of *Science* magazine.

for most real and complex values that agrees with $\zeta(n)$ wherever $\zeta(n)$ is defined. This new function is called the *Riemann zeta function*.

The Riemann zeta function has zeroes at $-2, -4, -6, \ldots$. These are called its trivial zeros. It has many other complex zeros that have real part 1/2. The line with real part 1/2 is called the critical line. The Riemann hypothesis states that the only nontrivial zeros of the Riemann zeta function lie on the critical line.

It might not be clear at this point what the Riemann hypothesis has to do with primes. Here is the connection, first shown by Euler:

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \frac{1}{1 - \frac{1}{p^s}}.$$

And there are deeper connections. For example, the prime number theorem turns out to be equivalent to the fact that the zeta function has no zeros on the line with real part -1.

If true, the Riemann hypothesis would imply that the prime numbers are distributed fairly regularly, whereas if it were false, then it would mean that prime numbers are distributed considerably more wildly. There are a number of potential theorems in number theory that start "If the Riemann hypothesis is true, then...." So a solution to the Riemann hypothesis would tell us a lot about primes and other things in number theory.

There have been a variety of different approaches to proving the Riemann hypothesis, though none have thus far been successful. Many mathematicians believe it is likely true, but there are some that are not so sure. Numerical computations have shown that the first 10^{13} nontrivial zeroes all lie on the critical line. The Riemann hypothesis is one of the Clay Mathematics Institute's seven Millennium Prize problems, with $1,000,000 offered for its solution.

2.15 Number-theoretic functions

There are a few functions that show up a lot in number theory.

Definition 8. *Let n be positive integer.*

1. *The number of positive divisors of n is denoted by $\tau(n)$.*
2. *The sum of the positive divisors of n is denoted by $\sigma(n)$.*
3. *The number of positive integers less than n relatively prime to n is denoted by $\phi(n)$.*

For example, $n = 12$ has six divisors: 1, 2, 3, 4, 6, and 12. Thus $\tau(12) = 6$ and $\sigma(12) = 1+2+3+4+6+12 = 28$. The only positive integers less than 12 that are relatively prime to 12 are 1, 5, 7, and 11, so $\phi(12) = 4$.

As another example, suppose p is prime. Then the divisors of p are just 1 and p, so $\tau(p) = 2$, $\sigma(p) = p+1$, and every positive integer less than p is relatively prime to it, so $\phi(p) = p - 1$.

The most important of the three functions is $\phi(n)$, called the *totient function* or simply the *Euler phi function*.

Computing $\tau(n)$

Let's start by computing $\tau(1400)$. We have $1400 = 2^3 \cdot 5^2 \cdot 7$. Any divisor will include 0, 1, 2, or 3 twos, 0, 1, or 2 fives, and 0 or 1 seven. So we have 4 choices for the twos, 3 choices for the fives, and 2 choices for the sevens. There are thus $4 \cdot 3 \cdot 2 = 24$ possible divisors in total.

This reasoning works in general. Given the prime factorization $n = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$, we have

$$\tau(n) = (e_1+1)(e_2+1)\cdots(e_k+1) = \prod_{i=1}^{k}(e_i+1).$$

Computing $\sigma(n)$

Let's compute $\sigma(1400)$. Again, we have $1400 = 2^3 \cdot 5^2 \cdot 7$. Consider the following product:

$$(1+2+2^2+2^3)(1+5+5^2)(1+7) = (1)(1)(1)+(1)(1)(7)+(1)(5)(1)+(1)(5)(7)+\cdots+(2^3)(5^2)(7).$$

The right hand side is exactly $\sigma(1400)$; each divisor appears exactly once in the sum. On the other hand, each term on the left side can be rewritten using the geometric series formula. So we end up with

$$\sigma(1400) = \frac{2^4-1}{2-1} \cdot \frac{5^3-1}{5-1} \cdot \frac{7^2-1}{7-1} = 3720.$$

In general, given the prime factorization $n = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$, we have

$$\sigma(n) = \frac{p_1^{e_1+1}-1}{p_1-1} \cdot \frac{p_2^{e_2+1}-1}{p_2-1} \cdot \dots \cdot \frac{p_k^{e_k+1}-1}{p_k-1} = \prod_{i=1}^{k} \frac{p_i^{e_i+1}-1}{p_i-1}.$$

Computing $\phi(n)$

Finding a formula for $\phi(n)$ is considerably more involved. First, if p is prime, then $\phi(p) = p-1$ since every positive integer less than p is relatively prime to p.

Next, consider $\phi(p^i)$, where p is prime. The positive integers less than p^i that are not relatively prime to p^i are precisely the multiples of p. There are p^{i-1} such multiples, namely $1, p, 2p, 3p, \dots, p^{i-1}p$. So $\phi(p^i) = p^i - p^{i-1}$, which we can rewrite as $p^i(1-\frac{1}{p})$.

To extend this to the factorization $n = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$, we will show (in a minute) that $\phi(mn) = \phi(m)\phi(n)$, whenever m and n are relatively prime. Since each of the terms $p_i^{e_i}$ are relatively prime, this will give us

$$\phi(n) = p_1^{e_1}\left(1-\frac{1}{p_1}\right) \cdot p_2^{e_2}\left(1-\frac{1}{p_2}\right) \cdot \dots \cdot p_k^{e_k}\left(1-\frac{1}{p_i}\right) = n\prod_{i=1}^{k} \frac{p_k-1}{p_k}.$$

For example, to compute $\phi(1400)$, we note $1400 = 2^3 \cdot 5^2 \cdot 7$ and compute

$$\phi(1400) = 1400\left(\frac{1}{2}\right)\left(\frac{4}{5}\right)\left(\frac{6}{7}\right) = 480.$$

As another example, to compute $\phi(164934)$, we have $164934 = 2 \cdot 3^2 \cdot 7^2 \cdot 11 \cdot 17$ and so

$$\phi(164394) = 164394\left(\frac{1}{2}\right)\left(\frac{2}{3}\right)\left(\frac{6}{7}\right)\left(\frac{10}{11}\right)\left(\frac{16}{17}\right) = 40320.$$

We now show that $\phi(mn) = \phi(m)\phi(n)$ whenever $\gcd(m,n) = 1$.

First, consider an example: $\phi(30) = 8$. We have $30 = 5 \times 6$, $\phi(5) = 4$, and $\phi(6) = 2$. See the figure below. Notice there are 2 columns with 4 integers relatively prime to 30 in each.

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30

As another example, consider $\phi(33) = 20$. We have $33 = 3 \times 11$, $\phi(3) = 2$, and $\phi(11) = 10$. Notice in the figure below there are 10 columns with 2 integers relatively prime to 33 in each.

1	2	3	4	5	6	7	8	9	10	11
12	13	14	15	16	17	18	19	20	21	22
23	24	25	26	27	28	29	30	31	32	33

This sort of thing happens in general. Suppose we have relatively prime integers m and n, and consider $\phi(nm)$.

$$\begin{array}{cccc}
1 & 2 & \dots & m \\
m+1 & m+2 & \dots & 2m \\
2m+1 & 2m+2 & \dots & 3m \\
\vdots & \vdots & \ddots & \vdots \\
(n-1)m+1 & (n-1)m+2 & \dots & nm
\end{array}$$

We claim that there are $\phi(m)$ columns that each contain $\phi(n)$ integers relatively prime to nm. This tells us that $\phi(nm) = \phi(n)\phi(m)$. The following three steps are enough to prove the claim:

1. If an entry x in row 1 is not relatively prime to m, then none of the entries in the same column as x are relatively prime to m. Thus only $\phi(m)$ columns can contain entries that are relatively prime to mn.

   This is true because all the entries in the column are of the form $x + mk$. Since $\gcd(x, m) \neq 1$ that means some integer d divides both x and m and hence $x + mk$ as well.

2. Each column is a permutation of $0, 1, 2, \ldots, n-1$ modulo n.

   To show this, suppose two entries in the column were congruent modulo n, say $x + mk \equiv x + mj \pmod{n}$ for some i and j. Then $mk \equiv mj \pmod{n}$ and since $\gcd(m, n) = 1$, we can cancel to get $k \equiv j \pmod{n}$, which is to say the entries must have come from the same row. In other words, entries in the column from different rows can't be the same.

3. If we can show that whether an integer is relatively prime to n or not depends only on its congruence class modulo n, then we would be done, since the column is a permutation of $0, 1, 2, \ldots, n-1$ modulo n, and there are $\phi(n)$ integers in that range relatively prime to n.

   To do this, suppose $s \equiv t \pmod{n}$ and s is relatively prime to n. We want to show that t is relatively prime to n. We can write $s - t = nk$ and $sx + ny = 1$ for some integers k, x, and y. Solve the former for s and plug it into the latter to get $(nk + t)x + ny = 1$. Rearrange to get $tx + n(kx + y) = 1$ from which we get that t and n are relatively prime.

Here is an interesting result about the Euler phi function:

Theorem 16. *For any integer $n \geq 1$, $\sum_{d\,|\,n} \phi(d) = n$.*

That is, if we sum $\phi(d)$ over the divisors of n, the result is n. For an idea as to why this is true, take a look at the following example:

d	Integers a with $\gcd(a, 10) = d$	$\phi(10/d)$
1	1, 3, 7, 9	$\phi(10) = 4$
2	2, 4, 6, 8	$\phi(5) = 4$
5	5	$\phi(2) = 2$
10	10	$\phi(1) = 1$

Recall from Section 1.7 that $\gcd(a, n) = d$ if and only if $\gcd(a/d, n/d) = 1$. That is, if we divide a and n by their gcd, then the resulting integers have nothing in common (and the converse holds as well). So for instance, if we want to find all the integers a such that $\gcd(a, 10) = 2$, we find all the integers whose gcd with 10/2 is 1; there are $\phi(5)$ such integers. And this works in both directions. In general then

$$\sum_{d\,|\,n} \phi(d) = \sum_{d\,|\,n} |\{a : \gcd(a, n) = d\}|,$$

and the latter sum must equal n as each integer from 1 through n must fall into exactly one of the sets $\{a : \gcd(a, n) = d\}$.

Multiplicative functions

A little earlier we showed that $\phi(mn) = \phi(m)\phi(n)$ provided $\gcd(m, n) = 1$. This way of breaking up a function is important in number theory. Here is a definition of the concept:

Definition 9. *A function f defined on the positive integers is called* multiplicative *provided $f(mn) = f(m)f(n)$ whenever $\gcd(m, n) = 1$.*

We have the following:

Theorem 17. *The functions τ, σ, and ϕ are multiplicative.*

We already showed ϕ is multiplicative, and it is easy to show τ and σ are multiplicative using the formulas we have for computing them.

The Möbius function

The Möbius function, defined below is important in higher number theory, though we won't use it much in this text.

Definition 10. *The* Möbius function*, denoted $\mu(n)$, is 1 if $n = 1$, $(-1)^k$ if $n = p_1 p_2 \cdots p_k$ is a product of k distinct primes, and is 0 otherwise.*

For instance, $12 = 2 \times 2 \times 3$ is not a product of distinct primes, so $\mu(12) = 0$. On the other hand, $105 = 3 \times 5 \times 7$ is a product of 3 distinct primes, so $\mu(105) = (-1)^3 = -1$.

It easy to show μ is multiplicative from its definition. The following fact, called *Möbius inversion* is an important tool in analytic number theory:

Theorem 18. *(Möbius inversion) If f and g are two number-theoretic functions such that $g(n) = \sum_{d\,|\,n} f(d)$ for every integer $n \geq 1$, then for every integer $n \geq 1$,*

$$f(n) = \sum_{d\,|\,n} \mu(d)\, g\left(\frac{n}{d}\right).$$

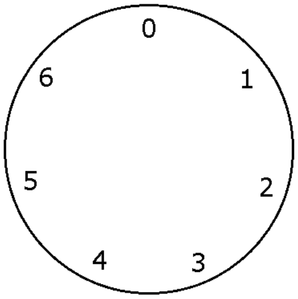
The proof is not difficult and can be found in many textbooks.

Chapter 3

Modular arithmetic

Modular arithmetic is a kind of "wrap around" arithmetic, like arithmetic with time. In 24-hour time, after 23:59, we wrap back around to the start 00:00. After the 7th day of the week (Saturday), we wrap back around to the start (Sunday). After the 365th or 366th day of the year, we wrap back around to the first day of the year. Many things in math and real-life are cyclical and a special kind of math, known as modular arithmetic, is used to model these situations.

Let's look at some examples of arithmetic modulo (mod) 7, where we use the integers 0 through 6. We can think of the integers as arranged on a circle, like below:



We have the following:

1. 7 is the same as 0, 8 is the same as 1, 9 is the same as 2, etc.
2. In general, any multiple of 7 is the same as 0, any number of the form $7k+1$ is the same as 1, any number of the form $7k+2$ is the same as 2, etc.
3. $4+5$ is the same as 2. Adding 5 corresponds to moving around the circle 5 units clockwise.
4. $4-5$ is the same as 6. Subtracting 5 corresponds to moving 5 units counterclockwise.
5. $4+21$ is the same as 4. Adding 21 corresponds to going around the circle 3 times and ending up where you started.

Instead of saying something like "8 is the same as 1," we use the notation $8 \equiv 1 \pmod 7$. This is read as "8 is congruent to 1 mod 7." Such an expression is called a *congruence*. Here is the formal definition:

Definition 11. *Let a, b, and n be integers. We say $a \equiv b \pmod n$ if a and b both leave the same remainders when divided by n. Equivalently, $a \equiv b \pmod n$ provided $n\,|(a-b)$.*[2]

It is not too hard to show that the two definitions are equivalent. We can use whichever one suits us best for a particular situation. The latter definition is useful in that it gives us an equation to work with, namely $a-b=nk$ for some integer k. For example, $29 \equiv 15 \pmod 7$ since both 29 and 15 leave the same remainder when divided by 7. Equivalently, $29 \equiv 15 \pmod 7$ because $29-15$ is divisible by 7.

Modular arithmetic is usually defined by noting that $\equiv$ is an equivalence relation. See Section 3.6 for more on this approach. For now we will just approach things informally.

[2] We can also use $n\,|(b-a)$ in place of $n\,|(a-b)$ in the definition.

A few examples

Here are a few examples to get us some practice with congruences:

1. Find the remainder when $1!+2!+3!+\ldots 100!$ is divided by 12.

   Solution: Notice that $4! = 4\cdot 3\cdot 2\cdot 1$ contains $4\cdot 3$, so it is divisible by 12. Similarly, 5!, 6!, etc. all contain 4×3, so they are all divisible by 12 and hence congruent to 0 modulo 12. Thus $1!+2!+3!+\ldots 100! \equiv 1!+2!+3! \equiv 9 \pmod{12}$

2. A useful fact is that $a \equiv 0 \pmod{n}$ if and only if $n \mid a$. This is useful in computer programs. For instance, to check if an integer a is even in a computer program, we check if `a % 2 == 0`.

3. Mods give an easy way of finding the last digits of a number. The last digit of an integer n is n mod 10. The last two digits are n mod 100, and in general, the last k digits are $n \bmod 10^k$.

   For example, suppose we want the last digit of 2^{1000}. To find it, we compute 2^{1000} modulo 10. Notice that $2^5 \equiv 2 \pmod{10}$. Then $2^{25} = (2^5)^5 \equiv 2^5 \equiv 2 \pmod{10}$. Similarly, $2^{125} \equiv 2 \pmod{10}$, and $2^{1000} = 2^{125\cdot 8} = (2^{125})^8 \equiv 2^8 \pmod{10}$. Since $2^8 = 256 \equiv 6 \pmod{10}$, our answer is 6.

3.1 Working with congruences

Modular arithmetic is like a whole new way of doing math. It's good to have a list of some of the common rules for working with it.

Relationship to the mod operator

Modular arithmetic is related to the mod operation from Section 1.3. For instance, in arithmetic mod 7 we have 1, 8, 15, 22, ... as well as -6, -13, -20, ... all corresponding to the same value. Often, but not always, the most convenient value to use to represent all of them is the smallest positive integer value, in this case 1. To find that value for, we use the mod operation from Section 1.3. For instance, to find the smallest positive integer that 67 is congruent to modulo 3, we can compute 67 mod 3 to get 1.

In general, we have the following:

> Given an integer m, if we want to find the smallest positive integer k such that $m \equiv k \pmod{n}$, we have $k = m \bmod n$.

A similar and useful rule is the following:

> An integer n is of the form $ak+b$ if and only if $n \equiv b \pmod{a}$.

For instance, if a number n is of the form $3k+1$, then $n \equiv 1 \pmod{3}$. And conversely, any number congruent to 1 modulo 3 is of the form $3k+1$.

Algebraic rules

Here are a few rules for working with congruences:

1. The $\equiv$ symbol satisfies three properties that make it an equivalence relation:

   (a) Reflexive property: For all x, $x \equiv x \pmod{n}$.

   (b) Symmetric property: If $x \equiv y \pmod{n}$, then $y \equiv x \pmod{n}$.

   (c) Transitive property: If $x \equiv y \pmod{n}$, and $y \equiv z \pmod{n}$, then $x \equiv z \pmod{n}$.

   These are simple rules that are easy to prove. We will use them without referring to them.

2. $a + cn \equiv a \pmod{n}$.

   In other words, adding a multiple of the modulus n is the same as adding 0. For instance, $2 + 45 \equiv 2 \pmod{9}$ and $17 + 1000 \equiv 17 \pmod{100}$.

3. $n - a \equiv -a \pmod{n}$.

   This rule is a special case of the previous rule and is often useful in computations. For instance, $-1 \equiv 99 \pmod{100}$. So we can replace 99 in computations mod 100 with -1, which is a lot easier to work with. If we needed to compute 99^{50} modulo 100, we could replace 99 with -1 and note that $(-1)^{50}$ is 1, so $99^{50} \equiv 1 \pmod{100}$.

4. We can work with congruences in similar (but not identical) ways to how we work with algebraic equations. For example, we can add or subtract a term on both sides of a congruence, multiply both sides by something, or raise both sides to the same power. That is, if $a \equiv b \pmod{n}$, then for any c

   $$a \pm c \equiv b \pm c \pmod{n}$$
   $$ca \equiv cb \pmod{n}$$
   $$a^c \equiv b^c \pmod{n}.$$

5. We can add two congruences, just like adding two equations. In particular, if $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$, then

   $$a + c \equiv b + d \pmod{n}.$$

Canceling terms

We have to be careful canceling terms. For example, $2 \times 5 \equiv 2 \times 11 \pmod{12}$ but $5 \not\equiv 11 \pmod{12}$. We see that we can't cancel the 2. In general, we have the following theorem:

Theorem 19. *Suppose $ca \equiv cb \pmod{n}$. Then $a \equiv b \pmod{n/\gcd(c,n)}$. In particular, if $\gcd(c,n) = 1$, then we can cancel c to get $a \equiv b \pmod{n}$.*

In the example above, since $\gcd(2,12) \neq 1$, we can't cancel out the 2. However, using the theorem we can say that $7 \equiv 9 \pmod{6}$. On the other hand, given $2 \times 3 \equiv 2 \times 10 \pmod{7}$, since $\gcd(2,7) = 1$, we can cancel out the 2 to get $3 \equiv 10 \pmod{7}$.

Breaking things into cases

Modular arithmetic is useful in that it can break things down into several cases to check. Here are a few examples:

1. Suppose we want to show that n^4 can only end in 0, 1, 5 or 6.

   To solve this, we find the possible values of n^4 modulo 10. There are 10 cases to check, $0^4, 1^4, \ldots, 9^4$, since every integer is congruent to some integer from 0 to 9. We end up with $0^4 \equiv 0 \pmod{10}$, $5^4 \equiv 5 \pmod{10}$, $1^4 \equiv 3^4 \equiv 7^4 \equiv 9^4 \equiv 1 \pmod{10}$, and $2^4 \equiv 4^4 \equiv 6^4 \equiv 8^4 \equiv 1 \pmod{10}$.

2. In Section 1.2 we showed that any perfect square n^2 is of the form $4k$ or $4k + 1$.

   To show this using modular arithmetic, we just consider cases modulo 4. Squaring 0, 1, 2, and 3 modulo 4 gives 0, 1, 0, and 1, so we see that $n^2 \equiv 0 \pmod{4}$ or $n^2 \equiv 1 \pmod{4}$, which is the same as saying that n^2 is of the form $4k$ or $4k + 1$.

3. Show that if $p > 3$ is prime, then $p^2 - 1$ is divisible by 24.

   First, note that if $p > 3$ is prime, then p must be of the form $24k + r$, where $r = 1, 5, 7, 11, 13, 17, 19$, or 23. Any other form is composite (for example $24k + 3$ is divisible by 3 and $24k + 10$ is divisible by 2). Thus $p \equiv r \pmod{24}$ for one of the above values of r, and it is not hard to check that $p^2 - 1 \equiv r^2 - 1 \equiv 0 \pmod{24}$ for each of those values.

In general, if an integer n is of the form $ak + b$, then we can write the congruence $n \equiv b \pmod{a}$. The converse holds as well. For instance, all numbers of the form $5k + 1$ are congruent to 1 modulo 5 and vice-versa.

Breaking up a mod

We have the following useful fact:

Theorem 20. *If $a \equiv b \pmod{m}$ and $a \equiv b \pmod{n}$, with $\gcd(m,n) = 1$, then $a \equiv b \pmod{mn}$.*

This follows from a fact proved in Section 2.1.

We often use this to break up a large mod into smaller mods. For example, suppose we want to show that n^5 and n always end in the same digit. That is, we want to show that $n^5 \equiv n \pmod{10}$. Using the above fact, we can do this by showing $n^5 \equiv n \pmod 2$ and $n^5 \equiv n \pmod 5$. The first congruence is easily seen to be true. For the second, we check the five cases 0^5, 1^5, 2^5, 3^5, and 4^5. A short calculation verifies that the congruence holds for each of them.

The theorem above can be generalized to the following:

Theorem 21. *Consider the system of congruences $a \equiv b \pmod{m_1}$, $a \equiv b \pmod{m_2}$, …, $a \equiv b \pmod{m_k}$. If m_1, m_2, … m_n are all pairwise relatively prime, then $a \equiv b \pmod{m_1 m_2 \dots m_2}$.*

If the moduli are not necessarily pairwise relatively prime, we still have

$$a \equiv b \pmod{\operatorname{lcm}(m_1, m_2, \dots, m_k)}.$$

Working with the definition

One of the most important parts of working with congruences is using the definition. In particular, we have that $x \equiv y \pmod n$ provided $n \,|\, (x-y)$ or equivalently that $x - y = nk$ for some integer k. Here are several examples:

1. Prove that $n^3 - n$ is divisible by 3 for any $n \in \mathbb{Z}$.

   We can turn this into a statement about congruences, namely $n^3 \equiv n \pmod 3$. Modulo 3 there are only three cases to check: $n = 0$, 1, and 2. And we have $0^3 \equiv 0 \pmod 3$, $1^3 \equiv 1 \pmod 3$, and $2^3 \equiv 2 \pmod 3$. Compare this argument to the longer one using the division algorithm in Example 1 of Section 1.2.

2. Prove if $a \equiv b \pmod n$, then $a + c \equiv b + c \pmod n$.

   We can start by writing the first congruence as an equation: $a - b = nk$ for some integer k. Add and subtract c to the left side to get $a - b + (c - c) = nk$. Then rearrange terms to get $(a + c) - (b + c) = nk$. Then rewrite this equation as the congruence $a + c \equiv b + c \pmod n$, which is what we needed to show.

3. Prove if $a \equiv b \pmod n$, then $a^c \equiv b^c \pmod n$.

   This is a little trickier. The definition tells us we need to show $n \,|\, (a^c - b^c)$. The trick is to factor the left side into $(a-b)(a^{c-1} + a^{c-2}b + a^{c-3}b^2 + \dots + ab^{c-2} + b^{c-1})$. Since $a \equiv b \pmod n$, we have $n \,|\, (a - b)$. Since $a - b$ is a factor of $a^c - b^c$, we get $n \,|\, (a^c - b^c)$.

4. Prove if $ca \equiv cb \pmod n$, then $a \equiv b \pmod{n/d}$, where $d = \gcd(c, n)$.

   Start with $nk = ca - cb$ for some integer k. Also, since d is $\gcd(a, b)$, we have $dx = n$ and $dy = c$ for some integers x and y.

   Plugging in, we get $dxk = dya - dyb$. We can cancel d to get $xk = y(a - b)$. We know $\gcd(x, y) = 1$ as otherwise any common factor of x and y could be included in d to get a larger common divisor of n and c. So we can use Euclid's lemma to conclude that $x \,|\, a - b$. Note that $x = n/d$, so we have $a \equiv b \pmod{n/d}$.

3.2 Powers

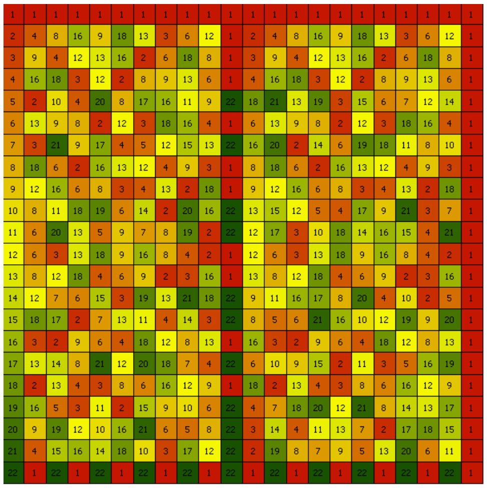
Powers turn out to be interesting in modular arithmetic. For instance, here is a table of powers modulo 7:

	$\mathbf{a^0}$	$\mathbf{a^1}$	$\mathbf{a^2}$	$\mathbf{a^3}$	$\mathbf{a^4}$	$\mathbf{a^5}$	$\mathbf{a^6}$
a = 1	1	1	1	1	1	1	1
a = 2	1	2	4	1	2	4	1
a = 3	1	3	2	6	4	5	1
a = 4	1	4	2	1	4	2	1
a = 5	1	5	4	6	2	3	1
a = 6	1	6	1	6	1	6	1

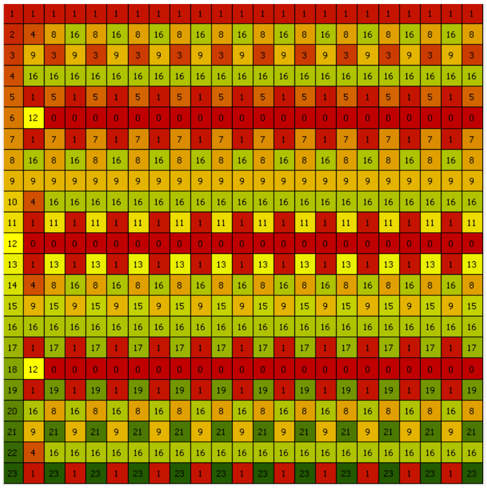
There are a number of interesting things here, some of which we will look at in detail later. We see that $a^6 \equiv 1 \pmod 7$ for all values of a. We can also see that some of the powers run through all the integers 1 through 6, while others don't. In fact, all the powers cycle with period 1, 2, 3, or 6, all of which are divisors of 6.

Note: To create the table, it is not necessary to compute large powers. For instance, instead of computing $5^3 = 125$ and reducing modulo 7, we can instead write $5^3 = 5^2 \cdot 5$. Since 5^2 is 4 modulo 7, we get that $5^3 = 4 \cdot 5$, which is 6 modulo 7.

Below are the tables for modulo 23 and 24. Smaller integers are colored red and larger values are green.[1]



1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	4	8	16	9	18	13	3	6	12	1	2	4	8	16	9	18	13	3	6	12	1
3	9	4	12	13	16	2	6	18	8	1	3	9	4	12	13	16	2	6	18	8	1
4	16	18	3	12	2	8	9	13	6	1	4	16	18	3	12	2	8	9	13	6	1
5	2	10	4	20	8	17	16	11	9	22	18	21	13	19	3	15	6	7	12	14	1
6	13	9	8	2	12	3	18	16	4	1	6	13	9	8	2	12	3	18	16	4	1
7	3	21	9	17	4	5	12	15	13	22	16	20	2	14	6	19	18	11	8	10	1
8	18	6	2	16	13	12	4	9	3	1	8	18	6	2	16	13	12	4	9	3	1
9	12	16	6	8	3	4	13	2	18	1	9	12	16	6	8	3	4	13	2	18	1
10	8	11	18	19	6	14	2	20	16	22	13	15	12	5	4	17	9	21	3	7	1
11	6	20	13	5	9	7	8	19	2	22	12	17	3	10	18	14	16	15	4	21	1
12	6	3	13	18	9	16	8	4	2	1	12	6	3	13	18	9	16	8	4	2	1
13	8	12	18	4	6	9	2	3	16	1	13	8	12	18	4	6	9	2	3	16	1
14	12	7	6	15	3	19	13	21	18	22	9	11	16	17	8	20	4	10	2	5	1
15	18	17	2	7	13	11	4	14	3	22	8	5	6	21	16	10	12	19	9	20	1
16	3	2	9	6	4	18	12	8	13	1	16	3	2	9	6	4	18	12	8	13	1
17	13	14	8	21	12	20	18	7	4	22	6	10	9	15	2	11	3	5	16	19	1
18	2	13	4	3	8	6	16	12	9	1	18	2	13	4	3	8	6	16	12	9	1
19	16	5	3	11	2	15	9	10	6	22	4	7	18	20	12	21	8	14	13	17	1
20	9	19	12	10	16	21	6	5	8	22	3	14	4	11	13	7	2	17	18	15	1
21	4	15	16	14	18	10	3	17	12	22	2	19	8	7	9	5	13	20	6	11	1
22	1	22	1	22	1	22	1	22	1	22	1	22	1	22	1	22	1	22	1	22	1



1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	4	8	16	8	16	8	16	8	16	8	16	8	16	8	16	8	16	8	16	8	16	8
3	9	3	9	3	9	3	9	3	9	3	9	3	9	3	9	3	9	3	9	3	9	3
4	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16
5	1	5	1	5	1	5	1	5	1	5	1	5	1	5	1	5	1	5	1	5	1	5
6	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	1	7	1	7	1	7	1	7	1	7	1	7	1	7	1	7	1	7	1	7	1	7
8	16	8	16	8	16	8	16	8	16	8	16	8	16	8	16	8	16	8	16	8	16	8
9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
10	4	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16
11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	1	13	1	13	1	13	1	13	1	13	1	13	1	13	1	13	1	13	1	13	1	13
14	4	8	16	8	16	8	16	8	16	8	16	8	16	8	16	8	16	8	16	8	16	8
15	9	15	9	15	9	15	9	15	9	15	9	15	9	15	9	15	9	15	9	15	9	15
16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16
17	1	17	1	17	1	17	1	17	1	17	1	17	1	17	1	17	1	17	1	17	1	17
18	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	1	19	1	19	1	19	1	19	1	19	1	19	1	19	1	19	1	19	1	19	1	19
20	16	8	16	8	16	8	16	8	16	8	16	8	16	8	16	8	16	8	16	8	16	8
21	9	21	9	21	9	21	9	21	9	21	9	21	9	21	9	21	9	21	9	21	9	21
22	4	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16
23	1	23	1	23	1	23	1	23	1	23	1	23	1	23	1	23	1	23	1	23	1	23

Computing numbers to rather large powers turns out to be pretty important in number theory and cryptography. Here are a couple of examples:

1. Compute 6^{100} modulo 7.
   We have $6 \equiv -1 \pmod 7$ so $6^{100} \equiv (-1)^{100} \equiv 1 \pmod 7$.
2. Compute 2^{100} modulo 7.
   Note that $2^3 \equiv 1 \pmod 7$. Further, we can write $100 = 3 \times 33 + 1$. Thus we have $2^{100} = 2 \times (2^3)^{33}$. Then
   $$2^{100} \equiv 2 \times (2^3)^{33} \equiv 2 \times 1^{33} \equiv 2 \pmod 7.$$

In general, if we can spot a power that is simple, like that $6^1 \equiv -1 \pmod 7$ or $2^3 \equiv 1 \pmod 7$, then we can leverage that to work out large powers. Otherwise, we can use the technique demonstrated below.

Suppose we want to compute $5^{100} \pmod{11}$. Compute the following:

$$\begin{aligned}
&5^1 \equiv 5 \pmod{11}\\
&5^2 \equiv 3 \pmod{11}\\
&5^4 \equiv (5^2)^2 \equiv 9 \pmod{11}\\
&5^8 \equiv (5^4)^2 \equiv 4 \pmod{11}\\
&5^{16} \equiv (5^8)^2 \equiv 5 \pmod{11}\\
&5^{32} \equiv (5^{16})^2 \equiv 3 \pmod{11}\\
&5^{64} \equiv (5^{32})^2 \equiv 9 \pmod{11}.
\end{aligned}$$

Then we break up 5^{100} as $5^{64+32+4}$, which is $5^{64} \cdot 5^{32} \cdot 5^4$ or $9 \cdot 3 \cdot 9$. This reduces to 1 modulo 11.

This process is called *exponentiation by squaring*. In general, to compute $a^b \pmod n$, we compute a^1, a^2, a^4, a^8, etc., up until the exponent is the largest power of 2 less than b. We then write b in terms of those powers and use the rules of exponents to compute a^b. Writing b in terms of those powers is the same process as converting b to binary. For instance, 100 in binary is 1100100, which corresponds to $64 \cdot 1 + 32 \cdot 1 + 16 \cdot 0 + 8 \cdot 0 + 4 \cdot 1 + 2 \cdot 0 + 1 \cdot 0$, or $64 + 32 + 4$.

Here is how we might code this algorithm in Python:

```python
def mpow(b,e,n):
    prod = 1
    while e > 0:
        if e % 2 == 1:
```

[1] The program that produced these images can be found at http://www.brianheinold.net/mods.html.

```
            prod = (prod * b) % n
        e = e // 2
        b = (b*b) % n
    return prod % n
```

Note, however, that this algorithm is already built into Python with the built-in function `pow`. In particular, `pow(a, n, m)` will compute $a^n \bmod m$. It can handle quite large powers.

3.3 Some further examples of modular arithmetic

1. Unlike with ordinary arithmetic, it is possible for the product of two nonzero integers to be 0 in modular arithmetic. For example, $2 \times 5 \equiv 0 \pmod{10}$.

   Note that this cannot happen if the modulus is prime. This is because if $ab \equiv 0 \pmod{n}$, then we have $p \mid ab$. By Euclid's lemma, since p is prime, either $p \mid a$ or $p \mid b$, which would imply at least one of a and b is congruent to 0 modulo p.[1]

2. An easy way to tell if a number is divisible by 3 is if the sum of its digits is divisible by 3. We can use modular arithmetic to show that this is true. Suppose we have a number n with ones digit d_0, tens digit d_1, etc. We can write that number as

   $$n = 10^k d_k + 10^{k-1} d_{k-1} + 100 d_2 + 10 d_1 + d_0.$$

   The sum of the digits of the number is $S = d_k + d_{k-1} + \cdots + d_1 + d_0$. We can see that $n \equiv S \pmod{3}$ because

   $$n - S \equiv (10^k - 1)d_k + (10^{k-1} - 1)d_{k-1} + \cdots + 999d_3 + 99d_2 + 9d_1 \equiv 0 \pmod{3}.$$

   Since n and S are congruent modulo 3, whenever one is divisible by 3, the other is, too.

   The key here is that when we compute $n - S$, each of the coefficients is a multiple of 3. It is possible to use this idea to develop tests for divisibility by other integers. For instance, for divisibility by 11, we use $S = d_0 - d_1 + d_2 - d_3 + \cdots \pm d_k$, where the last sign is + or -, depending on whether k is even or odd. When we compute $n - S$, the coefficients become 11, 99, 1001, 9999, etc., which are all divisible by 11.

   As another example, suppose we want a test for whether a four-digit number is divisible by 7. The ideas above can be streamlined into the following procedure:

   $$\begin{array}{rrrrr} & 1000 & 100 & 10 & 1 \\ - & 994 & 98 & 7 & 0 \\ \hline & 6 & 2 & 3 & 1 \end{array}$$

   The numbers in the second row are the closest multiples of 7 less than the powers of 10 directly above.

   Our divisibility test for the four-digit number $abcd$ is to check if $6a + 2b + 3c + d$ is divisible by 7. Or, since $6 \equiv -1 \pmod{7}$, we can check if $-a + 2b + 3c + d$ is divisible by 7.

3. Modular arithmetic can be used to find the day of the week of any date. For example, here is how to compute the date of Christmas is any given year Y:

   $$a = \left\lfloor \frac{Y}{100} \right\rfloor \bmod 4 \qquad b = \mathrm{Y} \bmod 100 \qquad c = \left\lfloor \frac{b}{4} \right\rfloor \bmod 7$$

   Reduce $(b + c - 2a + 1)$ modulo 7. A 0 corresponds to Sunday, 1 to Monday, etc.

   For example, if $y = 1977$, then $a = 19 \bmod 4 = 3$, $b = 77$, and $c = \lfloor 77/4 \rfloor \bmod 7 = 5$. Then we get $b + c - 2a + 1 \equiv 0 \pmod{7}$, so Christmas was on a Sunday in 1977.

   A more general process can be used to find the day of the week of any date in history. See *Secrets of Mental Math* by Art Benjamin and Michael Shermer. Modular arithmetic can also be used to determine the date of Easter in a given year.

3.4 Fermat's little theorem

Fermat's Little theorem is a useful rule that is simple to state:

Theorem 22. *(Fermat's little theorem) If p is prime, and p does not divide a, then* $a^{p-1} \equiv 1 \pmod{p}$.

[1] By the more general version of Euclid's lemma (Theorem 7), if n is relatively prime to both a and b, then we can't have $ab \equiv 0 \pmod{n}$.

An equivalent way to state the theorem is: If p is prime, then $a^p \equiv a \pmod{p}$ for any integer a.

To get from the this statement to the original, we can divide both sides through by a, which works as long as $\gcd(p, a) = 1$. To get from the original to this statement, just multiply through by a.

We will now prove Fermat's little theorem. To do so, we will need the following lemma:

Lemma 23. *If p is prime, then $(a+b)^p \equiv a^p + b^p \pmod{p}$.*

Proof. Use the binomial theorem to write

$$(a+b)^p = a^p + \binom{p}{1}a^{p-1}b + \binom{p}{2}a^{p-2}b^2 + \cdots + \binom{p}{p-1}ab^{p-1} + b^p.$$

Each binomial coefficient, $\binom{p}{k}$ with $1 < k < p$, can be written as below:

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

Since p is prime, no term in the denominator will cancel with p, meaning that each binomial coefficient is divisible by p and hence congruent to 0 modulo p. Thus only the a^p and b^p terms survive. □

The lemma above is sometimes called the *freshman's dream* since many freshman calculus and algebra students want to say $(a+b)^2 = a^2 + b^2$, forgetting the $2ab$ term. You can't forget the $2ab$ term in ordinary arithmetic, but you can when working mod 2.

We can now prove the alternate statement of Fermat's little theorem ($a^p \equiv a \pmod{p}$) using the lemma.

Proof. The proof is by induction on a. The base case $a = 1$ is simple. Now assume $a^p \equiv a \pmod{p}$. We need to show $(a+1)^p \equiv a \pmod{p}$. Using the previous lemma $(a+1)^p \equiv a^p + 1 \pmod{p}$ and by the induction hypothesis $a^p \equiv a \pmod{p}$. Thus we have $(a+1)^p \equiv a+1 \pmod{p}$, as desired. □

Here are some examples of Fermat's little theorem in action:

1. If a is not a multiple of 7, then $a^6 \equiv 1 \pmod{7}$. We saw this in the last row of the table of powers we computed in Section 3.2.
2. Find the remainder when 5^{38} is divided by 11.

   By Fermat's little theorem, $5^{10} \equiv 1 \pmod{11}$. Thus $5^{30} \equiv 1 \pmod{11}$ and so

   $$5^{38} \equiv 5^8 \cdot 5^{30} \equiv 5^8 \equiv 5^3 \cdot 5^3 \cdot 5^2 \equiv 4 \times 4 \times 3 \equiv 4 \pmod{11}.$$

3. The *inverse* of an integer a modulo a prime p is an integer a^{-1} such that $aa^{-1} \equiv 1 \pmod{p}$. Show that a^{p-2} is the inverse of a, provided $\gcd(a, p) = 1$.

   By Fermat's little theorem, we have $a \cdot a^{p-2} \equiv a^{p-1} \equiv 1 \pmod{p}$. From this we see that a^{p-2} fits the definition of a^{-1}. As an example, the inverse of 3 modulo 7 is $3^5 \equiv 4 \pmod{7}$.
4. Show that if a not divisible by 7, then $a^3 + 1$ or $a^3 - 1$ is divisible by 7.

   By Fermat's little theorem, $a^6 \equiv 1 \pmod{7}$. From this we get that $7 \,|\, (a^6 - 1)$. We can factor $a^6 - 1$ into $(a^3-1)(a^3+1)$ and use Euclid's lemma to conclude that $7 \,|\, (a^3 - 1)$ or $7 \,|\, (a^3 + 1)$.
5. Show that $p^{q-1} + q^{p-1} \equiv 1 \pmod{pq}$.

   By Fermat's little theorem, we have $p^{q-1} \equiv 1 \pmod{q}$. Also, $q^{p-1} \equiv 0 \pmod{q}$. Adding these gives $p^{q-1} + q^{p-1} \equiv 1 \pmod{q}$. A similar argument shows $p^{q-1} + q^{p-1} \equiv 1 \pmod{p}$. Since the same congruence holds modulo p and modulo q (and $\gcd(p, q) = 1$), it holds modulo pq.
6. Show that any prime other than 2 and 5 divides infinitely many of the numbers 1, 11, 111, 1111, ….

   These numbers are of the form $1 + 10 + 10^2 + 10^3 + \cdots + 10^k$, which can be rewritten as $\frac{10^{k+1}-1}{9}$ using the geometric series formula. By Fermat's little theorem $10^{p-1} \equiv 1 \pmod{p}$ for any prime p such that $\gcd(p, 10) = 1$ (i.e. for any prime besides 2 and 5). Thus also $10^{2(p-1)}$, $10^{3(p-1)}$, etc. are congruent to 1 modulo p. Thus $\frac{10^{k+1}-1}{9}$ will be congruent to 0 modulo p for infinitely many values of k.

   For example, the integers 111111 (6 ones), 111111111111 (12 ones), 111111111111111111 (18 ones) etc. are all divisible by 7 since $10^6 \equiv 1 \pmod{7}$. As another example, numbers that consist of 16, 32, 48, etc. ones are all divisible by 17 since $10^{16} \equiv 1 \pmod{17}$.
7. Fermat's little theorem and a generalization of it called Euler's theorem (see the next section) are a key part of the RSA algorithm, which is of fundamental importance in modern cryptography. See Section 4.2.

3.5 Euler's theorem

Euler's theorem is a generalization of Fermat's little theorem to nonprimes.

Theorem 24. *(Euler's theorem) If* $\gcd(a,n)=1$*, then* $a^{\phi(n)} \equiv 1 \pmod{n}$*.*

Recall that $\phi(n)$ is the Euler phi function from Section 2.15. Since $\phi(p)=p-1$ for any prime p, we see that Euler's theorem reduces to Fermat's little theorem when $n=p$ is prime.

As an example of Euler's theorem, $3^4 \equiv 1 \pmod{10}$ since $\gcd(3,10)=1$ and $\phi(10)=4$.

To understand why this works, recall that 1, 3, 7, and 9 are the $\phi(10)=4$ integers relatively prime to 10. Multiply each of these by $a=3$ to get 3, 9, 21, and 27, which reduce modulo 10 to 3, 9, 1, and 7, respectively. We see these are a rearrangement of the originals. So

$$(3^4)(1\cdot 3\cdot 7\cdot 9) \equiv 21\cdot 3\cdot 27\cdot 9 \equiv (7\cdot 1\cdot 9\cdot 3) \pmod{10}.$$

And we can cancel the common terms to get $3^4 \equiv 1 \pmod{10}$. We can formalize this example into a proof of Euler's theorem.

Proof. Let x_1, x_2, $\dots x_{\phi(n)}$ be the integers from 1 to $n-1$ that are relatively prime to n. Multiply each by a to get ax_1, ax_2, $\dots ax_{\phi(n)}$. We claim this is just a rearrangement of the original values. To show this, we need to show that the ax_i are all distinct and relatively prime to n.

The ax_i are distinct because if $ax_i \equiv ax_j \pmod{n}$, then since $\gcd(a,n)$, we can cancel a to get $x_i \equiv x_j \pmod{n}$.

We have ax_i is relatively prime to n because if some prime p divides ax_i, then by Euclid's lemma, $p \mid a$ or $p \mid x_i$, and as $\gcd(a,n)=\gcd(x_i,n)=1$, p cannot divide n. Thus ax_i and n cannot have any prime factors (and hence any factors besides 1) in common.

Thus we have

$$a^{\phi(n)}(x_1x_2\dots x_{\phi(n)}) \equiv (ax_1)(ax_2)\dots(ax_{\phi(n)}) \equiv x_1x_2\dots x_{\phi(n)} \pmod{n}.$$

Since each x_i is relatively prime to n, we can cancel it from both sides, leaving us with $a^{\phi(n)} \equiv 1 \pmod{n}$. □

3.6 Formal definition and inverses

The notation $\mathbb{Z}_n$ refers to the set $\{0,1,2,\dots,n-1\}$ with all arithmetic done modulo n.[1] More formally, the way $\mathbb{Z}_n$ is defined usually goes as follows:

The relation $\equiv$ is an equivalence relation (it is reflexive, symmetric, and transitive). As such, it partitions $\mathbb{Z}$ into disjoint sets called equivalence classes, where every integer in a given set is congruent to everything else in that set and nothing else.

For instance, here are the sets we get modulo 5:

$$\begin{aligned}
&[0]=\{\dots,-15,-10,-5,0,5,10,15,\dots\}\\
&[1]=\{\dots,-14,-9,-4,1,6,11,16,\dots\}\\
&[2]=\{\dots,-13,-8,-3,2,7,12,17,\dots\}\\
&[3]=\{\dots,-12,-7,-2,3,8,13,18,\dots\}\\
&[4]=\{\dots,-11,-6,-1,4,9,14,19,\dots\}.
\end{aligned}$$

We usually use the smallest nonnegative integer in the set to give set its name. We define $\mathbb{Z}_n$ to be the set $\{[0],[1],\dots,[n-1]\}$, with the addition and multiplication defined by $[a]+[b]=[a+b]$ and $[a]\times[b]=[a\times b]$.

The formal definition is used to make sure everything is on a firm mathematical footing. There is more to show to make sure that everything works out mathematically, but we will skip that here and just think of $\mathbb{Z}_n$ as the set $\{0,1,\dots,n-1\}$ with arithmetic done modulo n.

[1] In many texts the notation $\mathbb{Z}/n\mathbb{Z}$ is used.

Inverses

Some integers have an inverse in $\mathbb{Z}_n$. That is, for some integers a, there exists an integer a^{-1} such that $aa^{-1} \equiv 1 \pmod{n}$. For instance, in $\mathbb{Z}_7$, $3 \cdot 5 \equiv 1 \pmod{7}$, so we can say that the inverse of 3 is 5 (and also that the inverse of 5 is 3). Here is a useful fact about inverses.

Theorem 25. *An integer a has an inverse in $\mathbb{Z}_n$ if and only if* $\gcd(a,n) = 1$. *This inverse is unique.*

Proof. Finding an inverse of a is the same as solving $ax \equiv 1 \pmod{n}$ for x, which is the same as finding x and y such that $ax - ny = 1$. Theorem 6 guarantees that this has a solution if and only if $\gcd(a,n) = 1$.

To see that the inverse is unique, suppose $ax \equiv 1 \pmod{p}$ and $ay \equiv 1 \pmod{p}$. Then $ax \equiv ay \pmod{p}$ and since $\gcd(a,p) = 1$, we can cancel a to get $x \equiv y \pmod{p}$. □

The particular case of interest is $\mathbb{Z}_p$ where p is prime:

Corollary 26. *If p is prime, then each element of $\mathbb{Z}_p$ has a unique inverse.*

For instance, in $\mathbb{Z}_7$, we have $1^{-1} = 1$, $2^{-1} = 4$, $3^{-1} = 5$, $4^{-1} = 2$, $5^{-1} = 3$, and $6^{-1} = 6$. Notice that 1 and 6 (which is -1 mod 7) are their own inverses. We have the following:

Theorem 27. *For any integer n, 1 and -1 are their own inverses in $\mathbb{Z}_n$. If n is prime, then no other integers are their own inverses.*

Proof. Since $1 \cdot 1 \equiv 1 \pmod{n}$, 1 is its own inverse. Similarly, $-1 \cdot -1 \equiv 1 \pmod{n}$, so -1 is its own inverse.

Now assume $n = p$ is prime. If a is its own inverse, then $a \cdot a \equiv 1 \pmod{p}$. From this we get that $p \,|(a^2 - 1)$. We can factor $a^2 - 1$ into $(a-1)(a+1)$. By Euclid's lemma, $p \,|\, a-1$ or $p \,|\, a+1$, which tells us that $a \equiv 1 \pmod{p}$ or $a \equiv -1 \pmod{p}$. □

Modulo a composite, there can be integers besides ±1 that are their own inverses. For instance, $5 \cdot 5 \equiv 1 \pmod{8}$, so 5 is its own inverse mod 8.

3.7 Wilson's theorem

Wilson's theorem is a nice theorem that it gives a simple characterization of prime numbers in terms of modular arithmetic.

Theorem 28. *(Wilson's theorem) An integer p is prime if and only if* $(p-1)! \equiv -1 \pmod{p}$.

This gives us a way to check if a number is prime: just compute $(p-1)!$ modulo p. Its fatal flaw is that $(p-1)!$ is huge and difficult to compute even for relatively small values of p. So Wilson's theorem is not a practical primality test, unless someone were to find an easy way to compute factorials modulo a prime. Still, here is an example with $p = 11$:

$$(11-1)! \equiv (10 \cdot 9)(8 \cdot 7)(6 \cdot 5)(4 \cdot 3 \cdot 2 \cdot 1) \equiv 2 \cdot 1 \cdot 8 \cdot 2 \equiv 32 \equiv -1 \pmod{11}.$$

The proof of Wilson's theorem is interesting. Let's look at some examples to help understand it.

Take $p = 14$, a composite. We have $(14-1)! \equiv 0 \pmod{14}$ since 13! contains $2 \cdot 7 = 14$. This will work in general for a composite number—factor it and find its factors in $(p-1)!$. We have to be a little careful if p is the square of a prime, though. But as long as $p > 4$, we can still find its factors in $(p-1)!$. For instance, if $p = 25$, then $(25-1)! \equiv 0 \pmod{25}$ since 24! is divisible by both 5 and 10 (which contains a factor of 5), so we do get $25 \,|\, 24!$.

Now take $p = 7$, a prime. In $\mathbb{Z}_7$, 2 and 4 are inverses of each other, as are 3 and 5. Then we have

$$(7-1)! \equiv 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 \equiv (6)(5 \cdot 3)(4 \cdot 2)(1) \equiv 6 \cdot 1 \cdot 1 \cdot 1 \equiv -1 \pmod{7}.$$

In general, if p is prime, then in $\mathbb{Z}_p$, we can pair off everything except 1 and $p-1$ into inverse pairs. The numbers in each pair will cancel each other out in $(p-1)!$, leaving just $p-1$.

Here is a formal write-up of the proof:

Proof. It is easy to check that the theorem holds for $p < 5$, so suppose $p \geq 5$.

If p is not prime, then $(p-1)! \equiv 0 \pmod{p}$ since $p \,|(p-1)!$. We have this because we can write $p = ab$ for some positive integers a and b less than p (since p is not prime) and those integers both show up in $(p-1)!$, except possibly if the square of a prime, specifically $p = q^2$. But in that case as long as $p > 4$, we know that both q and $2q$ will show up in $(p-1)!$.

On the other hand, suppose p is prime. By Theorems 25 and 27, if p is prime, then in $\mathbb{Z}_p$ each integer $1, 2, \ldots, p-1$ has a unique inverse, with 1 and $p-1$ being the only integers that are their own inverses. This means that all the other integers come in inverse pairs. Thus, looking at $(p-1)! = (p-1)(p-2)\ldots 3 \cdot 2 \cdot 1$, we know that all the terms in the middle, $p-2$, $p-3, \ldots, 3, 2$ pair off into inverse pairs and cancel each other out, leaving us with $(p-1)! \equiv p-1 \equiv -1 \pmod{p}$. □

Even though Wilson's theorem is not practical for checking primality, it is a useful tool in proofs. There is also an interesting twin-prime analogue of Wilson's Theorem proved by P. Clement in 1949: $(p, p+2)$ are a twin-prime pair if and only if $4(p+1)! \equiv -(p+4) \pmod{p(p+2)}$.

3.8 Solving congruences

In ordinary algebra, solving the linear equation $ax = b$ is very useful and also very easy. Solving the congruence $ax \equiv b \pmod{n}$ is also useful, but a little trickier than solving $ax = b$.

For example, suppose we want to solve $2x \equiv 5 \pmod{11}$. To solve the ordinary equation $2x = 5$, we would divide by 2 to get $x = 2.5$, but in modular arithmetic, we don't quite have division, so we have to find other approaches. A simple approach is trial and error, as there are only 11 values of x to try. The solution ends up being $x = 8$. We will see some faster approaches shortly.

The algebraic equation $ax = b$ always has a single solution (unless $a = 0$), but with $ax \equiv b \pmod{n}$, it often happens that there is no solution or multiple solutions. For example, $2x \equiv 5 \pmod{10}$ has no solution, while $2x \equiv 4 \pmod{10}$ has two solutions modulo 10, namely $x = 2$ and $x = 7$.

Procedure for solving $ax \equiv b \pmod{n}$

Let $d = \gcd(a, n)$.

1. If $d \nmid b$, then there is no solution. Otherwise there are d solutions.
2. If there is a solution, we first find one solution, x_0, and use it to find all the other solutions. To find x_0, a variety of techniques can be used, such as the extended Euclidean algorithm, properties of congruences, and systematic checking of all possibilities (if n is small).
3. Once a solution, x_0, has been found, all solutions are of the form $x_0 + \frac{n}{d}t$ for $t = 0, 1, \ldots d-1$.

The idea for why this works is we can write $ax \equiv b \pmod{n}$ as $ax - nk = b$ for some integer k, Thus, solving the congruence is the same as finding integer solutions to the equation $ax - nk = b$. We know from Theorem 3 that such an equation only has a solution provided $d = \gcd(a, n)$ divides b. Further, the extended Euclidean algorithm can be used to find x and k in the equation.

To see where the formula for all the solutions comes from, note that we can divide $ax \equiv b \pmod{n}$ through by d to get $\frac{a}{d}x \equiv \frac{b}{d} \pmod{\frac{n}{d}}$. By a similar argument to the one used to prove Theorem 25 (about the existence of inverses), this equation must have a unique solution: $x \equiv x_0 \pmod{\frac{n}{d}}$. So x_0 will also be a solution modulo n, as will $x_0 + \frac{n}{d}$, $x_0 + 2\frac{n}{d}$, $x_0 + 3\frac{n}{d}, \ldots$, $x_0 + (d-1)\frac{n}{d}$, which are all less than n and congruent to x_0 modulo $\frac{n}{d}$. This gets us d different solutions. It is not hard to show that these are the only solutions.

A few notes on finding an initial solution

Below are a few different ways to find an initial solution x_0.

1. **Extended Euclidean algorithm** — Probably the most effective way in general to find an initial solution is to use the extended Euclidean algorithm, which was covered in Section 1.6. As an example, suppose we want to solve $8x \equiv 11$

(mod 23). Write it as $8x-23k=11$. Using the Euclidean algorithm on 8 and 23, we get

$$\begin{aligned}23 &= 8\cdot 2+7\\ 8 &= 7\cdot 1+\boxed{1}\end{aligned}$$

We can stop the Euclidean algorithm here as we see that the gcd will be 1. We then write

$$\begin{aligned}1 &= 8-7\\ &= 8-(23-8\cdot 2)\\ &= 8(3)-23(1)\end{aligned}$$

So we have $8(3)-23(1)=1$ and multiplying through by 11 gives $8(33)+23(-11)=11$. So $x_0=33$ is a solution, which we can reduce mod 23 to get $x_0=10$.

2. **Trial and error** — If the modulus is small enough, we can just try things systematically until something works. For example, to solve $5x\equiv 7 \pmod{11}$, we can list all the multiples of 5 until we get to one that is 7 more than a multiple of 11. Doing this, we get 5, 10, 15, 20, 25, 30, 35, 40. We stop here as 40 is a 7 away from 33, which is a multiple of 11. Thus $x_0=8$ is a solution. Another way of doing this is to list all the integers of the form $11k+7$ (18, 29, …) until we get a multiple of 5.

3. **Congruence rules** — We can use rules for working with congruences to manipulate the congruence into giving us an initial solution. For example, to find a solution to $11x\equiv 31 \pmod{45}$, we can add 90 (which is congruent to 0 mod 45) to both sides to get $11x\equiv 121 \pmod{45}$. Then, since $\gcd(11,45)=1$, we can divide both sides by 11 to get $x\equiv 11 \pmod{45}$.

4. **"Division"** — One final method is to use "division." To solve the ordinary linear equation $ax=b$, we would divide by a to get $x=b/a$. We can't divide by a in modular arithmetic, but we can do its equivalent, which is to multiply by a^{-1} (provided that a^{-1} exists). A solution to $ax\equiv n \pmod{b}$ is $x\equiv a^{-1}n \pmod{b}$. For instance, to solve $3x\equiv 4 \pmod{7}$, we note that $3^{-1}=5$ and $5\cdot 4\equiv 6 \pmod{7}$, so that $x_0=6$ is a solution.

   One problem with this is that finding a^{-1} itself takes some work. However, if you have a lot of equations of the form $ax\equiv n \pmod{b}$, where a is fixed, but n varies, then it makes sense to use this method, since all the work goes into finding a^{-1} and once we have it, it is short work to solve all those equations.

   One way to find a^{-1} is to note that if p is prime and $\gcd(a,p)=1$, then by Fermat's little theorem, $a^{-1}\equiv a^{p-2} \pmod{p}$. In general, by Euler's theorem, $a^{-1}\equiv a^{\phi(n)-1} \pmod{n}$ as long as $\gcd(a,n)=1$.

Examples

1. Solve $3x\equiv 11 \pmod{36}$.

   Solution: Since $\gcd(3,36)=3$ and $3\nmid 11$, there is no solution.

2. Solve $14x\equiv 4 \pmod{37}$.

   Solution: Since $\gcd(14,37)=1$, there will be exactly 1 solution.

   Using the Euclidean algorithm on 14 and 37, we get

$$\begin{aligned}37 &= 14\cdot 2+9\\ 14 &= 9\cdot 1+5\\ 9 &= 5\cdot 1+4\\ 5 &= 4\cdot 1+\boxed{1}.\end{aligned}$$

   We can stop the Euclidean algorithm here as we see that the gcd will be 1. Then use the extended Euclidean algorithm to get

$$\begin{aligned}1 &= 5-4\\ &= 5-(9-1\cdot 5)\\ &= 5(2)-9(1)\\ &= (14-1\cdot 9)(2)-9(1)\\ &= 14(2)-9(3)\\ &= 14(2)-(37-2\cdot 14)(3)\\ &= 14(8)-37(3).\end{aligned}$$

   So we have $14(8)-37(3)=1$ and multiplying through by 4 gives $14(32)+37(-12)=4$. From this, we get $x_0=32$ is the solution.

3. Solve $12x \equiv 18 \pmod{30}$.

   Solution: Since $\gcd(12,30) = 6$ and $6 \,|\, 18$, there 6 different solutions mod 30. We can divide the whole equation through by $\gcd(12,30) = 6$ to get $2x \equiv 3 \pmod 5$. By trial and error, we get $x_0 = 4$ is a solution. Then the solutions of the original are

   $$x = 4 + \frac{30}{6}t \qquad \text{for } t = 0,1,2,3,4,5.$$

   In other words, they are 4, 9, 14, 19, 24, and 29.

3.9 Solving linear Diophantine equations

Diophantine equations are algebraic problems where we are looking for integer solutions. They are among some of the trickiest problems in mathematics. For instance, Fermat's Last Theorem, which took 400 years and some seriously high-powered math to solve, is about showing that $x^n + y^n = z^n$ has no integer solutions if $n > 2$. However, linear Diophantine equations of the form $ax + by = c$ can be easily solved since they closely are related to the congruence $ax \equiv c \pmod b$.

Using the formula for the solutions to that congruence gives us the following formula for solutions to $ax + by = c$:

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

for any $t \in \mathbb{Z}$. Here we need some solution (x_0, y_0) to $ax + by = c$ to get us started. Such a solution can be found using the extended Euclidean algorithm, trial and error, or by working with congruences. Note that there is no solution if c is not divisible by $\gcd(a,b)$.

Here are a few example problems:

1. Find all the integer solutions to $37x + 14y = 11$.

   Solution: Notice that this is equivalent to the congruence $37x \equiv 11 \pmod{14}$, which we did earlier. In that example, we found $14(32) + 37(-12) = 4$. From this, we get $x_0 = 32$ and $y_0 = -12$. From here, all the solutions are of the form

   $$x = 32 + 37t$$
   $$y = -12 - 14t$$

   for any $t \in \mathbb{Z}$. Each value of t gives a different solution. For instance $t = 1$ gives $(69, -26)$ and $t = -1$ gives $(-5, 2)$. And we can check our work: $(-5, 2)$ is a solution because $14(-5) + 37(2) = 4$.

2. Find all integer solutions of $12x + 30y = 18$.

   Solution: This equation is also equivalent to a congruence we solved earlier. In that example, we got $x_0 = 4$ and from $12x_0 + 30y_0 = 18$, we get $y_0 = -1$. Then all solutions are of the form

   $$x = 4 + \frac{30}{6}t = 4 + 5t$$
   $$y = -1 - \frac{12}{6}t = -1 - 2t$$

   for any $t \in \mathbb{Z}$. For example, $t = -2$, -1, 0, 1, and 2 give $(-6,3)$, $(-1,1)$, $(4,-1)$, $(9,-3)$, and $(14,-5)$.

3. Linear Diophantine equations are the key to many famous old word problems. As a simple example, suppose apples are 69 cents and oranges are 75 cents. We spend a total of $12.09. How many of each did we buy?

   This reduces to the equation $69x + 75y = 1209$. The extended Euclidean algorithm gives $69(12) + 75(-11) = 3$ (we'll skip the details here; a quick way to get this would be to use the program of Section 1.6). Multiply through by 403 to get $69(4836) + 75(-4433) = 1209$. All the solutions are of the form

   $$x = x_0 + \frac{75}{3}t = 4836 + 25t$$
   $$y = y_0 - \frac{69}{3}t = -4433 - 23t$$

   for any $t \in \mathbb{Z}$. However, not all solutions will make sense since neither x nor y can be negative as they represent the numbers of apples and oranges bought. So we must have $4836 + 25t \geq 0$ and $-4433 - 23t \geq 0$. The first inequality can be solved to give us $t \geq -193.44$. The second gives us $t \leq -192.74$. So only $t = -193$ will give us positive values for x and y, which turn out to be $x = 11$ and $y = 6$.

4. Oystein Ore's *Number Theory and Its History* has a number of interesting old problems from various cultures. One such problem comes from a 12th century Hindu manuscript. Here is a modification of it:

   > A person has 5 rubies, 8 sapphires, 7 pearls, and 92 coins, while the other has 19 rubies, 14 sapphires, 2 pearls, and 4 coins. The combined worth is the same for both people. How many coins must rubies, sapphires, and pearls each be worth?

   Letting r, s, and p denote the values of rubies, sapphires, and pearls, we have $5r + 8s + 7p + 92 = 19r + 14s + 2p + 4$, which becomes $14r + 6s - 5p = 88$. This has three variables, as opposed to our previous examples which all have two. To handle this we start with the first two terms, $14r + 6s$. We have $\gcd(14, 6) = 2$ and we can write $14(1) + 6(-2) = 2$. Thus, the number of rubies and sapphires is always a multiple of 2, say $2n$ for some integer n. Then consider $2n - 5p = 88$. We have $\gcd(2, 5) = 1$ and we can write $2(3) - 5(1) = 1$. Multiply through by 88 to get $2(264) - 5(88) = 88$. Thus all solutions of $2n - 5p = 88$ can be written in the form

   $$n = 264 - 5t$$
   $$p = 88 - 2t$$

   for any $t \in \mathbb{Z}$. Going back to the equation $14(1) + 6(-2) = 2$ and multiplying through by $n = 264 - 5t$ gives $14(264 - 5t) + 6(-528 + 10t) = 2(264 - 5t)$. Thus all solutions of $14r + 6s - 5p = 88$ are of the form

   $$r = 264 - 5t + 3u$$
   $$s = -528 + 10t - 7u$$
   $$p = 88 - 2t$$

   for any $t, u \in \mathbb{Z}$. We are looking for positive integers, so we know that $88 - 2t > 0$ and hence $t < 44$. From this and the equation for s, we get that $u < -14$. Note that we can make things look a little nicer by setting $t = 43 - x$ and $u = -14 - y$. Then we have

   $$r = 4 + 5x - 3y$$
   $$s = 7 - 10x + 7y$$
   $$p = 2 + 2x$$

   for $x, y \geq 0$. For instance, $x = y = 0$ produces $r = 4$, $s = 7$, and $p = 2$, while $x = y = 1$ produces $r = 6$, $s = 4$, $p = 4$. Not all values of x and y will produce positive solutions, but there are infinitely many that do.

Equations with 3 or more unknowns

The example above is an equation of the form $ax + by + cz = d$. The procedure used in that example can be streamlined and used in general:

1. Let $e = \gcd(a, b)$ and $f = \gcd(e, c)$.
2. Find a solution (x_0, y_0) to $ax + by = e$.
3. Find a solution (w_0, z_0) to $ew + cz = d$.
4. Then all solutions are given by

   $$x = x_0\left(w_0 + \frac{c}{f}t\right) + \frac{b}{f}s$$
   $$y = y_0\left(w_0 + \frac{c}{f}t\right) - \frac{a}{f}s$$
   $$z = z_0 - \frac{e}{f}t,$$

   for any $s, t \in \mathbb{Z}$. This works provided $\gcd(a, b, c) \mid d$.

In general $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$ has a solution provided $\gcd(a_1, a_2, \ldots, a_n) \mid b$. The equation can be solved by an iterative process like above.

3.10 The Chinese remainder theorem

Just like we can solve systems of algebraic equations, we can solve systems of congruences. The technique used is called the *Chinese remainder theorem*. The name comes from its appearance in a third century Chinese manuscript.

Theorem 29. *(Chinese remainder theorem) Suppose we have a system like the one below:*

$$\begin{aligned} x &\equiv a_1 \pmod{n_1} \\ x &\equiv a_2 \pmod{n_2} \\ &\vdots \\ x &\equiv a_k \pmod{n_k}, \end{aligned}$$

where the n_i are pairwise relatively prime. Then we can find a solution that is unique modulo the product $n_1 n_2 \cdots n_k$.

To solve such a system, let $N = n_1 n_2 \cdots n_k$. Then for each $i = 1, 2, \ldots, k$, let $N_i = N/n_i$ (the product of all the moduli except n_i), and solve the congruence $N_i x_i \equiv 1 \pmod{n_i}$. The solution to the system is given by $a_1 N_1 x_1 + a_2 N_2 x_2 + \cdots + a_k N_k x_k$, which is unique modulo N.

1. Suppose we have the system

   $$x \equiv 3 \pmod 5 \qquad x \equiv 2 \pmod 7 \qquad x \equiv 4 \pmod 8.$$

   Note that $\gcd(5,7) = \gcd(5,8) = \gcd(7,8) = 1$, so we can use the Chinese remainder theorem. We have $N = 5 \cdot 7 \cdot 8 = 280$ along with $N_1 = 7 \cdot 8 = 56$, $N_2 = 5 \cdot 8 = 40$, and $N_3 = 5 \cdot 7 = 35$. We then solve the following three congruences:

   $$56x_1 \equiv 1 \pmod 5 \qquad 40x_2 \equiv 1 \pmod 7 \qquad 35x_3 \equiv 1 \pmod 8.$$

   The first equation reduces to $x_1 \equiv 1 \pmod 5$, so $x_1 = 1$. The second reduces to $5x_2 \equiv 1 \pmod 7$, from which we get $x_2 = 3$. The last reduces to $3x_3 \equiv 1 \pmod 8$, from which we get $x_3 = 3$. Then the solution is

   $$a_1 N_1 x_1 + a_2 N_2 x_2 + a_3 N_3 x_3 \equiv (3)(56)(1) + (2)(40)(3) + (4)(35)(3) \equiv 828 \equiv 268 \pmod{280}.$$

   We can check that this works: 268 leaves a remainder of 3 when divided by 5, a remainder of 2 when divided by 7, and a remainder of 4 when divided by 8.

2. The Chinese remainder theorem is useful for finding when several cyclical events will all line up. For instance, suppose there are three salesmen that visit a town on different cycles. Salesman A visits every 10 days, B visits every 7 days, and C visits every 3 days. Suppose A was last there 8 days ago, B was last there yesterday, and C is there today. When will all three be in town on the same day?

   We can describe this problem with a system of three congruences:

   $$x \equiv -8 \pmod{10} \qquad x \equiv -1 \pmod 7 \qquad x \equiv 0 \pmod 3.$$

   We have $N = 10 \cdot 7 \cdot 3 = 210$ and $N_1 = 7 \cdot 3 = 21$ and $N_2 = 10 \cdot 3 = 30$. Because of the 0 in the last congruence, there is no need to worry about N_3 and its associated congruence. We then solve the following two congruences:

   $$21x_1 \equiv 1 \pmod{10} \qquad 30x_2 \equiv 1 \pmod 7.$$

   We get $x_1 = 1$ and $x_2 = 4$. The solution to the problem is then

   $$a_1 N_1 x_1 + a_2 N_2 x_2 + a_3 N_3 x_3 \equiv (-8)(21)(1) + (-1)(30)(4) + 0 \equiv -48 \equiv 162 \pmod{210}.$$

   Thus the salesmen were last here together 48 days ago and will all be back 162 days from now. They are all in town together every 210 days thereafter.

Moduli that aren't relatively prime

The Chinese remainder theorem can't be used if the moduli are not relatively prime, but there are things that can be done:

1. A useful trick is to replace the moduli m_i with new moduli c_i. The rules for the new moduli are that c_i must be a divisor of m_i for each i, and the lcm Of the new moduli must be the same as the lcm of the originals. For example, suppose we have the system

   $$x \equiv 2 \pmod 4 \qquad x \equiv 2 \pmod 5 \qquad x \equiv 4 \pmod 6.$$

   The problem is that the moduli 4 and 6 share a factor of 2. We can set $c_1 = m_1$, $c_2 = m_2$, and remove a factor of 2 from $m_3 = 6$ to get $c_3 = 3$. This doesn't change the lcm, as $\operatorname{lcm}(4,5,6) = 60$ and $\operatorname{lcm}(4,5,3) = 60$. We then solve the reduced system

   $$x \equiv 2 \pmod 4 \qquad x \equiv 2 \pmod 5 \qquad x \equiv 4 \pmod 3.$$

   This turns out to have a solution of 58, which is unique modulo 60.
2. Sometimes the above trick is not enough. In that case, it might be possible to combine or eliminate some congruences. For instance, if $x \equiv 1 \pmod 2$ and $x \equiv 3 \pmod 4$, the first congruence is redundant. It tells us that x must be odd, but the second congruence also implies that.
3. Another way of approaching these problems is shown in the following example. Suppose we have $x \equiv 1 \pmod 4$ and $x \equiv 3 \pmod 6$. From the first congruence, we have $x - 1 = 4k$ for some integer k. Plug this into the second to get $4k \equiv 2 \pmod 6$. We can divide through by 2, though this changes the modulus to 3. So we get $2k \equiv 1 \pmod 3$, which we can solve to get $k \equiv 2 \pmod 3$. We can write this as $k = 3j+2$ for some integer j. Thus we have $x = 4k+1 = 12j+9$. In other words, $x \equiv 9 \pmod{12}$ solves the two congruences.

   One way to think about this is if a number is of the form $4k+1$ and $6k+3$, then it is of the form $12k+9$.

   It is not too hard to generalize the procedure above to solve $x \equiv a_1 \pmod{n_1}$ and $x \equiv a_2 \pmod{n_2}$. What we do is set $d = \gcd(n_1, n_2)$, solve $\frac{m_1}{d}k \equiv \frac{a_2 - a_1}{d} \pmod{\frac{m_2}{d}}$ for k, and then the solution to both congruences is $x \equiv a_1 + m_1 k \pmod{\operatorname{lcm}(a_1, a_2)}$. Note that this works provide $d \,|(a_2 - a_1)$.

In general, some combination of these techniques can be used for tricky problems. In fact, we have the following theorem:

Theorem 30. *Suppose we have a system like the one below:*

$$\begin{aligned} x &\equiv a_1 \pmod{n_1} \\ x &\equiv a_2 \pmod{n_2} \\ &\vdots \\ x &\equiv a_k \pmod{n_k}, \end{aligned}$$

The system has a unique solution modulo $\operatorname{lcm}(n_1, n_2, \cdots, n_k)$ *if and only if* $a_i \equiv a_j \pmod{\gcd(n_i, n_j)}$ *for all i and j.*

A classic Chinese remainder theorem problem is the following: There are some eggs in a basket. When they are removed in pairs, there is one left over. When three at a time are removed, there are two left over. When four, five, six, or seven at a time are removed, there remain three, four, five, or zero respectively. How many eggs are in the basket?

This corresponds to the following system:

$$\begin{aligned} x &\equiv 1 \pmod 2 \\ x &\equiv 2 \pmod 3 \\ x &\equiv 3 \pmod 4 \\ x &\equiv 4 \pmod 5 \\ x &\equiv 5 \pmod 6 \\ x &\equiv 0 \pmod 7. \end{aligned}$$

We can't use the Chinese remainder theorem yet as the moduli are not pairwise relatively prime. But we can eliminate some of them. For instance, the first congruence says that the number of eggs is odd. But the third congruence tells us the number of eggs is of the form $4k+3$, which is odd. So we can drop the first congruence.

Then the remaining moduli 3, 4, 5, 6, 7 can be reduced to 3, 4, 5, 6, 7 using the first trick given above. This turns $x \equiv 5 \pmod 6$ into $x \equiv 5 \pmod 3$, which is the same as $x \equiv 2 \pmod 3$, which we already have, so we can drop it. We are thus left with the following:

$$x \equiv 2 \pmod 3 \qquad x \equiv 3 \pmod 4 \qquad x \equiv 4 \pmod 5 \qquad x \equiv 0 \pmod 7.$$

We then compute $N = 3 \cdot 4 \cdot 5 \cdot 7 = 420$, $N_1 = 4 \cdot 5 \cdot 7 = 140$, $N_2 = 3 \cdot 5 \cdot 7 = 105$, $N_3 = 3 \cdot 4 \cdot 7 = 84$. We don't need N_4 because of the 0 in the last congruence. We then solve

$$140x_1 \equiv 1 \pmod 3 \qquad 105x_2 \equiv 1 \pmod 4 \qquad 84x_3 \equiv 1 \pmod 5.$$

To get $x_1 = 2$, $x_2 = 1$, and $x_3 = 4$. The solution is then

$$a_1N_1x_1 + a_2N_2x_2 + a_3N_3x_3 + a_4N_4x_4 \equiv (2)(140)(2) + (3)(105)(1) + (4)(84)(4) + 0 \equiv 2219 \equiv 119 \pmod{420}.$$

A few notes

All of the congruences we considered above are of the form $x \equiv a \pmod m$. It is possible to have more general cases, where we have $cx \equiv a \pmod m$. To handle these, we first have to solve them for x.

The Chinese remainder theorem has a number of applications. As seen above, it is useful any time we need to know when several cyclical events will line up. The Chinese remainder theorem is also an important part of modern cryptography, and it shows up here and there in higher math.

One important use of the Chinese remainder theorem is breaking up composite moduli into smaller pieces that are easier to work with. For instance, if we need to solve a congruence $f(x) \equiv a \pmod{mn}$ with $\gcd(m,n) = 1$, we can solve the congruences $f(x) \equiv a \pmod m$ and $f(x) \equiv a \pmod n$ and combine them by the Chinese remainder theorem to get a solution modulo mn. Note the similarity between this and the fact mentioned on in Section 3.1.

3.11 Order

It is interesting to look at powers modulo an integer. For example, if we look at the powers of 2 modulo 9, we get the repeating sequence 2, 4, 8, 7, 5, 1, 2, 4, 8, 7, 5, 1, ... If we look at the powers of 7 modulo 9, we get the repeating sequence 7, 4, 1, 7, 4, 1, ... The powers of 8 give the repeating sequence 8, 1, 8, 1, ... The goal of this section is to understand a little about these repeating sequences.

In particular, we are interested in the first power that turns out to be 1. This power is called the *order*. We have the following definition:

Definition 12. *Let a and n be relatively prime positive integers. The* order *of a modulo n is the least positive integer k such that $a^k \equiv 1 \pmod n$.*

For instance, the order of 7 modulo 9 is 3, since $7^3 \equiv 1 \pmod 9$ and no smaller positive power of 7 (7^1 or 7^2) is congruent to 1.

If a is not relatively prime to n, the order is undefined, as no power of a beside a^0 can ever be 1. So we will only concern ourselves with values of a that are relatively prime to n.

Euler's theorem tells us that $a^{\phi(n)} \equiv 1 \pmod n$, so the order must always be no greater than $\phi(n)$. But there is an even closer connection between the order and $\phi(n)$, namely that the order must always be a divisor of $\phi(n)$.

For example, take $n = 13$. We have $\phi(13) = 12$. Suppose some integer a had an order that is not a divisor of 12, say order 7. Then we would have $a^7 \equiv 1 \pmod{13}$ and $a^{14} \equiv 1 \pmod{13}$. But since $a^{12} \equiv 1 \pmod{13}$ (by Euler's theorem), we would have

$$1 \equiv a^{14} \equiv a^{12}a^2 \equiv a^2 \pmod{13}.$$

But since the order of a was assumed to be 7, it is not possible to have a power less than the seventh power be congruent to 1. Here is a formal statement and proof of the theorem.

Theorem 31. *The order of an integer a modulo n is a divisor of $\phi(n)$.*

Proof. Let k be the order of a. By the division algorithm, we can write $\phi(n) = kq + r$ for some integers q and r with $0 \le r < k$. Our goal is to show that k is a divisor of $\phi(n)$, which means we need to show that $r = 0$. Euler's theorem tells us $a^{\phi(n)} \equiv 1 \pmod n$, so we have

$$1 \equiv a^{\phi(n)} \equiv a^{kq+r} \equiv a^{kq}a^r \equiv 1 \cdot a^r \pmod n.$$

So we have $a^r \equiv 1 \pmod n$, But since $0 \le r < k$ and the order of a is k, this is only possible if $r = 0$. □

This theorem is a special case of Lagrange's Theorem, an important result in group theory.

3.12 Primitive roots

Below is a table of orders of integers modulo 13. We have $\phi(13) = 12$ and the possible orders are the divisors of 12, namely 1, 2, 3, 4, 6, and 12.

Order	Integers with that order modulo 13
1	1
2	12
3	3, 9
4	5, 8
6	4, 10
12	2, 6, 7, 11

Of particular interest are 2, 6, 7, and 11, which have order 12, the highest possible order. These values are called *primitive roots*. In general, we have the following:

Definition 13. *Let n be a positive integer. An integer a is said to be a* primitive root *of n if the order of a modulo n is $\phi(n)$.*

Not every integer has primitive roots. For instance, 8 doesn't have any. We have $\phi(8) = 4$, but 1, 3, 5, and 7 all have order 1 or 2. We have the following theorem:

Theorem 32. *If $n = 2$, 4, p^k ,or $2p^k$ for some odd prime p and positive integer k, then n has a primitive root. Otherwise it doesn't.*

We won't prove this theorem here as it is a bit involved. However, you can find a complete development of the theorem in many number theory texts. The integers between 2 and 50 that have primitive roots are shown below:

$$2, 3, 4, 5, 6, 7, 9, 10, 11, 13, 14, 17, 18, 19, 22, 23, 25, 26, 27, 29, 31, 34, 37, 38, 41, 43, 46, 47, 49, 50$$

The powers of a primitive root a of n are all unique and run through all of the integers relatively prime to n. For instance, consider $a = 2$ and $n = 13$. The powers of 2 are 2, 4, 8, 3, 6, 12, 11, 9, 5, 10, 7, 1. We see that they run through all the integers relatively prime to 13.[1]

Theorem 33. *If a is a primitive root of n, the integers $a, a^2, \dots, a^{\phi(n)}$ are distinct and hence run through all the integers relatively prime to $\phi(n)$.*

Proof. Suppose i and j are exponents in the range from 1 to $\phi(n)$ with $i \leq j$ and $a^i \equiv a^j \pmod{n}$. Since $\gcd(a, n) = 1$, we must also have $\gcd(a^j, n) = 1$ and we can divide through by a^i to get $a^{j-i} \equiv 1 \pmod{n}$. But $j - i < \phi(n)$ and the order of a is $\phi(n)$, so we must have $j - i = 0$, meaning $i = j$. □

Phrased another way, if a is a primitive root of n, then every integer relatively prime to n is of the form a^k for some k. This gives us a way to determine the orders of other elements modulo n.

As an example, consider an integer n with $\phi(n) = 30$ and suppose we want the order of some integer b that turns out to equal a^8. Since a is a primitive root, its order must be 30. That is $a^{30} \equiv 1 \pmod{n}$, and for that matter a^{60}, a^{90}, a^{120}, etc. are all congruent to 1 modulo n as well. We are looking for the order of b, so we want to find a power of a^8 that is congruent to 1. So we go through a^{16}, a^{24}, a^{32}, etc. until we get to one that matches up with one of the powers 30, 60, 90, etc.

In other words, we are looking for when multiples of 8 match up with multiples of 30. Thus we just need to find $\operatorname{lcm}(8, 30)$, which is 120. Then

$$1 \equiv a^{120} \equiv \left(a^8\right)^{15} \pmod{n}.$$

Then order of b is thus 15, which is $\operatorname{lcm}(8, 30)/8$. In general, we have the following:

[1] Using the terminology of abstract algebra, we can say a is a generator of the multiplicative group of integers relatively prime to n since every integer is a power of a. The group is thus cyclic provided n has a primitive root.

Theorem 34. *Let n be a positive integer with a primitive root a. If $b = a^k$, then the order of b is $\operatorname{lcm}(k, \phi(n))/k$; equivalently, the order is $\phi(n)/\gcd(k, \phi(n))$.*

Proof. By Euler's theorem, the only powers of a that are congruent to 1 are multiples of $\phi(n)$. Thus for a^k to be congruent to 1, we must have k be a multiple of $\phi(n)$. The smallest such multiple is $m = \operatorname{lcm}(k, \phi(n))$. Thus we have $1 \equiv a^m \equiv (b^k)^{m/k}$. So the order of b is m/k. By Theorem 5, we can also write the order of b is $\phi(n)/\gcd(k, \phi(n))$. □

The theorem above actually allows us to determine how many primitive roots there are for a given integer n:

Theorem 35. *If an integer n has a primitive root, then it has $\phi(\phi(n))$ of them.*

Proof. Let a be a primitive root of n, and consider a^k for some integer k. If k is relatively prime to $\phi(n)$, then the order of a_k is $\phi(n)$ by the previous theorem, making a^k a primitive root. There are $\phi(\phi(n))$ such integers k relatively prime to $\phi(n)$. □

We can actually say something a little more general:

Theorem 36. *If n has a primitive root, then there are $\phi(m)$ integers with order m in $\mathbb{Z}_n$.*

Proof. Let a be a primitive root of n. We know the order of a^k is $\phi(n)/\gcd(k, \phi(n))$. Setting this equal to m and rewriting gives $\gcd(k, \phi(n)) = \phi(n)/m$. From the proof of Theorem 16, any integer k whose gcd with $\phi(n)$ is $\phi(n)/m$ must satisfy $\gcd(k, \phi(n)/(\phi(n)/m)) = 1$. In other words, k must be relatively prime to m. There are $\phi(m)$ such integers. □

Finding primitive roots

So we know how many primitive roots any integer must have, but what are they? That is a much harder question to answer. Even simple questions such as which integers have 2 as a primitive root or finding the smallest primitive root for a given integer don't have easy answers. It is suspected that every positive integer that is not a perfect square is a primitive root of infinitely many primes. This has not been proved. It is known as *Artin's conjecture*. It is further suspected that the number of primes less than x for which a non-perfect square a is a primitive root is approximately some constant times $x/\ln(x)$. In terms of partial results, it has been shown that Artin's conjecture is true for almost all primes, specifically that there at most two primes for which Artin's conjecture fails.

For relatively small integers, it is not too difficult to write a program to find the primitive roots. Here is some Python code to do that:

```
from fractions import gcd

def phi(n):
    return len([x for x in range(1,n) if gcd(x,n)==1])

def order(a, n):
    if gcd(a,n) != 1:
        return -1
    c = a % n
    p = 1
    while c != 1:
        c = c*a % n
        p += 1
    return p

def prim_roots(n):
    return [a for a in range(1,n) if order(a, n)==phi(n)]
```

The code above mostly brute-forces things. There are more efficient approaches, like using the formula for calculation $\phi(n)$ from its prime factorization. And for finding the primitive roots, we know that possible orders are divisors of $\phi(n)$, so to check if a is a primitive root, we could just compute a^d for each divisor d of $\phi(n)$ less than $\phi(n)$. If none of those come out to 1, then we know a is a primitive root.

3.13 Decimal expansions

Decimal expansions have interesting connections to modular arithmetic. Here are a few decimal expansions:

1/3	$.\overline{3}$
1/5	$.2$
1/6	$.1\overline{6}$
1/7	$.\overline{142856}$
1/11	$.\overline{09}$
1/13	$.\overline{076923}$
1/17	$.\overline{0588235294117647}$
1/19	$.\overline{052631578947368421}$
1/37	$.\overline{027}$

The overbar indicates a repeating decimal. For instance, $1/11 = .\overline{09}$ is shorthand for the decimal expansion $.09090909\ldots$. To find the decimal expansion of a/b, repeatedly apply the division algorithm as in the example below that computes the decimal expansion of $1/7$:

$$10 = 1\cdot 7+3$$
$$30 = 4\cdot 7+2$$
$$20 = 2\cdot 7+6$$
$$60 = 8\cdot 7+4$$
$$40 = 5\cdot 7+5$$
$$50 = 7\cdot 7+1$$
$$10 = 1\cdot 7+3$$

etc.

The quotients 1, 4, 2, 8, 5, 7, 1, ... are the digits of the decimal expansion. At each stage the remainder term from the previous step is multiplied by 10 and becomes the new dividend. Since the remainder is guaranteed to be between 0 and $b-1$, eventually a remainder will be repeated, causing the decimal expansion to repeat. It is interesting to note that changing the 10 to another integer d will give the base-d decimal expansion of a/b.

Below is a short Python implementation of the algorithm above. It will print the first n digits of the decimal expansion of a/b. With a little work it could be modified to find how long it takes before the expansion repeats itself.

```
for i in range(n):
        print(10 * a // b)
        a = 10 * a % b
```

Every fraction has a decimal expansion that will repeat or terminate. A terminating expansion is a special case of a repeating expansion, with a 0 repeating, like in $1/4 = .250000\cdots$. The expansion of a/b is terminating if and only if b is of form 2^j5^k for some integers j and k. This is not too hard to show using the division algorithm and noting that 2 and 5 are the prime divisors of 10, the base of the decimal system.

Further, only rational numbers have repeating or terminating expansions. The decimal expansion of irrational numbers cannot have an endlessly repeating pattern of this sort. To see why, consider the process below that will find the fraction corresponding to the repeating decimal $x = .345634563456\ldots$.

Multiply both sides by 10000 to get $10000x = 3456.34563456\ldots$. We can rewrite this as $10000x = 3456+x$, which we can solve to get 3456/9999. We can reduce this in lowest terms to 384/1111. A similar process works in general. For example, $.\overline{123}$ would be 123/999, and $.\overline{235711}$ would be 235711/999999. A variation of the process can be used for other numbers where the pattern does not start right away, like $.23\overline{45}$ or $.1\overline{8}$.

This technique can help us find the length of the repeating pattern in the expansion of $1/n$. Suppose we have $1/n = .\overline{d_1d_2\ldots d_k}$. Write $10^k/n = d_1d_2\ldots d_k + 1/n$. From here we get $10^k-1 = (d_1d_2\ldots d_k)n$, which we can write as $10^k \equiv 1 \pmod{n}$. Thus k, the length of the cycle, is given by the order of 10 modulo n.

For example, 10 has order 1 modulo 3, so 1/3 has a decimal expansion with a repeating cycle of length 1. Also, 10 has order 6 modulo 7, so 1/7 has a decimal expansion with a repeating cycle of length 6. In general, since the maximum order of 10 modulo n is $\phi(n)$, that is the maximum length of a repeating cycle.

An interesting note is that the factors of $10^k - 1$ tell us for what primes p that $1/p$ might have a repeating decimal expansion of length k. For instance, $10^3 - 1 = 999$, which factors into $3^3 \times 37$, so only 1/3 and 1/37 can have length 3, and 1/3 has length 1, so only 1/37 has a length of 3.

3.14 Quadratic Reciprocity

This section is about what positive integers are perfect squares in $\mathbb{Z}_n$. For example, in $\mathbb{Z}_7$, squaring 1, 2, 3, 4, 5, and 6 gives 1, 4, 2, 2, 4, and 1. So the only perfect squares are 1, 4, and 9.

As another example, in $\mathbb{Z}_{11}$, the squares of the integers 1 through 10 are 1, 4, 9, 5, 3, 3, 5, 9, 4, and 1, in that order. So the perfect squares are 1, 3, 4, 5, and 9.

Perfect squares modulo n have long been studied and are usually referred to as *quadratic residues*. Here is the formal definition:

Definition 14. *Let a and n be relatively prime integers. Then a is called a* quadratic residue *of n if there exists a b such that $b^2 \equiv a \pmod{n}$. Otherwise, a is called a* quadratic nonresidue *of n.*

To denote whether an integer is a quadratic residue or not, a special notation called the *Legendre symbol* is used. Here is the formal definition:

Definition 15. *Let p be an odd prime and let $a \in \mathbb{Z}$. The* Legendre symbol, $(\frac{a}{p})$, *is defined to be 1 if a is a quadratic residue of p, -1 if a is a quadratic non-residue of p, and 0 if a is 0.*

Basic properties

Here are the squares of the nonzero elements of $\mathbb{Z}_{13}$ in order from 1^2 to 12^2: 1, 4, 9, 3, 12, 10, 10, 12, 3, 9, 4, 1. Notice the symmetry about the middle. This always happens. Notice also that each square appears exactly twice in the list above and that exactly half of the integers from 1 through 12 are squares. This always happens modulo a prime.

Theorem 37. *The following hold in $\mathbb{Z}_p$, where p is prime.*

1. *Every quadratic residue is the square of exactly two elements of $\mathbb{Z}_p$, one of which is less than $p/2$ and the other of which is greater than $p/2$.*
2. *Exactly half of the elements of $\mathbb{Z}_p$ are quadratic residues.*

Proof. If $b^2 \equiv a \pmod{p}$, then $(-b)^2 \equiv a \pmod{p}$ as well. So every quadratic residue is the square of at least two things. Note that if $b < p/2$, then $p - b$, which is congruent to $-b$, is greater than $p/2$. And if $b > p/2$, then $n - b < p/2$.

Now suppose $b^2 \equiv c^2 \pmod{p}$. Then $p \,|(b^2 - c^2) = (b - c)(b + c)$. By Euclid's lemma, $p \,|(b - c)$ or $p \,|(b + c)$. Hence $b \equiv \pm c \pmod{p}$. So each quadratic residue is the square of exactly two integers modulo p. From this, we also get that half of the integers modulo a prime are quadratic residues and the other half are not. □

This doesn't necessarily work for composite moduli. For instance, in $\mathbb{Z}_8$, the integer 1 has four square roots, namely 1, 3, 5, and 7. So in what follows, we will usually be working modulo a prime.

Euler's identity

In $\mathbb{Z}_7$, if we raise the integers from 1 to 6 to the 3rd power, we get 1, 1, -1, 1, -1, 1. In $\mathbb{Z}_{11}$, if we raise the integers from 1 to 10 to the 5th power, we get 1, -1, 1, 1, 1, -1, -1, -1, 1, -1. Do we always get ±1 when raising an integer to the $(p-1)/2$ power modulo a prime p? The answer is yes. Fermat's little theorem tells us that $a^{p-1} \equiv 1 \pmod{p}$, and the two square roots of 1 are ± 1, so $a^{(p-1)/2} \equiv \pm 1 \pmod{p}$.

The interesting fact, known as *Euler's identity*, is that whether $a^{(p-1)/2}$ is 1 or -1 tells us if a is a quadratic residue or not.

Theorem 38. *(Euler's identity) Let p be an odd prime and let $a \in \mathbb{Z}$. Then $(\frac{a}{p}) \equiv a^{(p-1)/2} \pmod{p}$.*

As an example, to tell if 2 is a quadratic residue of 17, we just compute 2^8 modulo 17. Doing so, we get 1, so 2 is a quadratic residue of 17.

To understand how the proof works, consider $p = 11$ and the quadratic residue $a = 5$. We have $5 \equiv 4^2 \pmod{11}$ so $5^{(10-1)/2} \equiv 4^{10-1} \equiv 1 \pmod{11}$ by Fermat's little theorem.

On the other hand, take $a = 2$, which is not a quadratic residue of 11. The integers 1 through 10 pair up into products that equal 3, namely $2 = 1 \cdot 2$, $3 \cdot 8$, $4 \cdot 6$, $5 \cdot 7$, and $9 \cdot 10$. Since a is not a quadratic residue, none of those pairs could have a repeat (like $3 \cdot 3$). Thus we have $10! \equiv 2^5 \pmod{11}$. Notice that 5 is $(p-1)/2$ here and by Wilson's theorem, 10! (which is $(p-1)!$) will be -1.

Here is a formal proof:

Proof. Suppose a is a quadratic residue. Then $a \equiv b^2 \pmod{p}$ for some integer b and we have by Fermat's little theorem that

$$a^{(p-1)/2} \equiv \left(b^{(p-1)/2}\right)^2 \equiv b^{p-1} \equiv 1 \pmod{p}.$$

Now suppose a is a quadratic nonresidue. For any $b = 1, 2, \ldots, p-1$, the congruence $bx \equiv a \pmod{p}$ has a unique solution since p is prime, and since a is a quadratic nonresidue, we can't have $x = b$. As a consequence, the integers 1, 2, …, $p-1$ can be broken up into $(p-1)/2$ pairs whose products all equal a. Therefore, we have $a^{(p-1)/2} \equiv (p-1)! \pmod{p}$ and this is congruent to -1 by Wilson's theorem. □

One nice consequence of Euler's identity is the following:

Theorem 39. *If p is prime, then -1 is a quadratic residue of p if and only if p is of the form* $4k+1$.

Proof. If p is of the form $4k+1$, then using Euler's identity, we get

$$(-1)^{(4k+1-1)/2} \equiv (-1)^{2k} \equiv 1 \pmod{p},$$

so -1 is a quadratic residue of any $4k+1$ prime. On the other hand, for a $4k+3$ prime, a similar computation results in -1, showing -1 is a quadratic nonresidue for $4k+3$ primes. □

One can use this fact to show that there are infinitely many primes of the form $4k+1$. The proof is reminiscent of Euclid's proof that there are infinitely many primes, but with a twist. Suppose $p_1, p_2, \ldots, p_n$ are all $4k+1$ primes. Consider $(2p_1p_2\cdots p_n)^2+1$. It is odd, so it is divisible by some odd prime p, and that prime cannot be any of the p_i. We can then write $(2p_1p_2\cdots p_n)^2 \equiv -1 \pmod{p}$. We have written -1 as the square of an element of $\mathbb{Z}_p$, so -1 is a quadratic residue of p, which means p is a $4k+1$ prime by the theorem above. Thus, given any list of $4k+1$ primes, we can generate another one, giving us infinitely many.

While we're at it, we can also show that there are infinitely many $4k+3$ primes. This proof is also reminiscent of Euclid's proof, but it doesn't require anything about quadratic residues. First note that the product of integers of the form $4k+1$ is also of the form $4k+1$, so any number of the form $4k+3$ can't consist of only $4k+1$ primes; it must be divisible by some prime of the form $4k+3$. Let $p_1, p_2, \ldots, p_n$ be $4k+3$ primes, none of which equal 3. Consider $P = 3p_1p_2\cdots p_n + 2$ and $P' = 3p_1p_2\cdots p_n + 4$. When multiplying numbers of the form $4k+3$ together, the product will either be of the form $4k+1$ or $4k+3$, depending on whether there are an odd or even amount of numbers in the product. So one of P and P' must be of the form $4k+3$. None of the p_i can divide P or P' because they are all divisors of $3p_1p_2\cdots p_n$ and are greater than 4. But, as mentioned, one of P and P' must be divisible some $4k+3$ prime, and prime cannot be one of the p_i. Thus, given any list of $4k+3$ primes, we can generate another one, giving us infinitely many.

Another easy consequence of Euler's identity is that the Legendre symbol is multiplicative.

Theorem 40. *The Legendre symbol is multiplicative. That is,* $(\frac{ab}{p}) = (\frac{a}{p})(\frac{b}{p})$ *whenever* $\gcd(a,b) = 1$.

In other words, if a and b are both quadratic residues or both quadratic nonresidues of p, then their product is a quadratic residue of p. Otherwise their product is a quadratic nonresidue.

Proof. By Euler's formula, we have

$$\left(\frac{ab}{p}\right) \equiv (ab)^{(p-1)/2} \equiv a^{(p-1)/2}b^{(p-1)/2} \equiv \left(\frac{a}{p}\right)\left(\frac{b}{p}\right).$$

□

Gauss's lemma

Another way to tell if something is a quadratic residue is *Gauss's lemma*:

Theorem 41. *(Gauss's lemma) Let p be an odd prime with a relatively prime to p. Consider the following multiples of a: a, $2a$, $\ldots (p-1)/2 \cdot a$. Let n be the number of them that are greater than $p/2$. Then $(\frac{a}{p}) = (-1)^n$.*

To help us understand why this is true, look at the multiples of $a = 2$ in $\mathbb{Z}_{11}$: 2, 4, 6, 8, 10, 1, 3, 5, 7, 9. Replacing the elements greater than 5 with their equivalent negative forms, we get 2, 4, -5, -3, -1, 1, 3, 5, -4, -2. Notice that 1, 2, 3, 4, and 5 or their negatives appear exactly once in the first five multiples and once in the last five. This always happens. If we multiply the first five multiples together, one the one hand we get $2^5 \cdot 5!$, and on the other hand we get $(-1)^3 \cdot 5!$. This gives $2^5 \equiv (-1)^3 \pmod{11}$. By Euler's identity, 2^5 tells us whether 2 is a quadratic residue or not, so here we have another way of computing $(\frac{2}{5})$, by looking at how many negatives (numbers greater than 11/2) appear in the first 5 multiples of 2. Here is a formal proof:

Proof. If two multiples ja and ka are congruent mod p, then we have $ja \equiv ka \pmod{p}$, and we can cancel a to conclude $j \equiv k \pmod{p}$. If we have $ja \equiv -ka \pmod{p}$, then we have $j \equiv -k \pmod{p}$. Thus the absolute values of the multiples $|a|$, $|2a|$, $\ldots$, $|(p-1)/2 \cdot a|$ are distinct. Therefore, multiplying all of the multiples together gives us $(-1)^n \cdot ((p-1)/2)!$. On the other hand, it gives us $a^{(p-1)/2}((p-1)/2)!$. Equating these and canceling out the common terms gives us $a^{(p-1)/2} \equiv (-1)^n \pmod{p}$. By Euler's identity, $(\frac{a}{p}) = a^{(p-1)/2}$, so the result follows. □

A straightforward calculation using Gauss's lemma shows the following:

Theorem 42. *Let p be prime, then 2 is a quadratic residue of p if and only if p is of the form $8k \pm 1$.*

Proof. Looking at the multiples $2, 4, \ldots, 2(p-1)/2$, there are $(p-1)/2$ in total, and $\lfloor (p-1)/4 \rfloor$ that are less than $p/2$. So there are $(p-1)/2 - \lfloor (p-1)/4 \rfloor$ that are greater than $p/2$. If p is of the form $8k+1$, that expression simplifies to $2p$ and since $(-1)^{2p} = 1$, we get that 2 is a quadratic residue of p by Gauss's lemma. The other cases $8k+3$, $8k-3$, and $8k-1$, can be similarly verified. □

Eisenstein's lemma

Eisenstein's lemma builds on Gauss's lemma to give us another way to compute $(\frac{a}{p})$.

Theorem 43. *(Eisenstein's lemma) If p is prime and a is odd, then*

$$\left(\frac{a}{p}\right) = \sum_{k=1}^{(p-1)/2} \left\lfloor \frac{ka}{p} \right\rfloor.$$

To better understand the statement and its proof, consider an example with $a = 5$ and $p = 13$. The terms $\lfloor \frac{ka}{p} \rfloor$ are the quotients in the division algorithm when ka is divided by p. For instance, below is the division algorithm written out for all the multiples, a, $2a$, $\ldots 6a$.

$$\begin{aligned}
&1 \cdot 5 = 0 \cdot 13 + 5 \\
&2 \cdot 5 = 0 \cdot 13 + (13-3) \\
&3 \cdot 5 = 1 \cdot 13 + 2 \\
&4 \cdot 5 = 1 \cdot 13 + (13-6) \\
&5 \cdot 5 = 1 \cdot 13 + (13-1) \\
&6 \cdot 5 = 2 \cdot 13 + 4
\end{aligned}$$

We have chosen to write all of the remainders that are larger than $p/2$ in a particular way. If we add up all these equations, we get

$$5 \cdot (1+2+3+4+5+6) = 13 \cdot (0+0+1+1+1+2) + (5-3+2-6-1+4) + 3 \cdot 13.$$

If we then write this equation as a congruence modulo 2, we get the following

$$(0+0+1+1+1+2) = 3 \pmod{2}.$$

This is because 5 and 13 are both odd an hence congruent to 1 modulo 2, and since any integer is congruent to its negative mod 2, we get that $(1+2+3+4+5+6)$ and $5-3+2-6-1+4$ are congruent. The right side of the congruence, 3, is the number of multiples greater than $p/2$, which we know from Gauss's lemma is equal to $(\frac{a}{p})$.

Here is the formal proof:

Proof. For each $k = 1,2,\dots,(p-1)/2$, we use the division algorithm to write $ka = q_k p + r_k$, where the quotient q_k equals $\lfloor \frac{ka}{p} \rfloor$ and the remainder r_k satisfies $0 \le r_k < p$. For each k, define the modified remainder s_k to be r_k if $r_k < p/2$ and $r_k - p$ if $r_k > p/2$. By Gauss's lemma there are $(\frac{a}{p})$ remainders greater than $p/2$. So we can rewrite $r_1 + r_2 + \dots r_k$ as $s_1 + s_2 + \cdots + s_k + p(\frac{a}{p})$.

Add up all of the equations from the division algorithm to get

$$a(1+2+\cdots+(p-1)/2) = p\left(\left\lfloor \frac{a}{p} \right\rfloor + \left\lfloor \frac{2a}{p} \right\rfloor + \left\lfloor \frac{((p-1)/2)a}{p} \right\rfloor\right) + (s_1+s_2+\cdots+s_k) + p\left(\frac{a}{p}\right).$$

We will now write this equation as a congruence mod 2. Note that a and p are both congruent to 1 mod 2. By the argument used in the proof of Gauss's lemma, $\{s_1, s_2, \dots, s_k\} = \{1,2,\dots,(p-1)/2\}$. Since $x \equiv -x \pmod 2$ for any integer x, we must then have $s_1 + s_2 + \cdots + s_k \equiv 1 + 2 + \cdots + (p-1)/2 \pmod 2$. Thus the equation above, when written as a congruence mod 2, reduce to

$$\left(\frac{a}{p}\right) = \sum_{k=1}^{(p-1)/2} \left\lfloor \frac{ka}{p} \right\rfloor.$$

□

The law of quadratic reciprocity

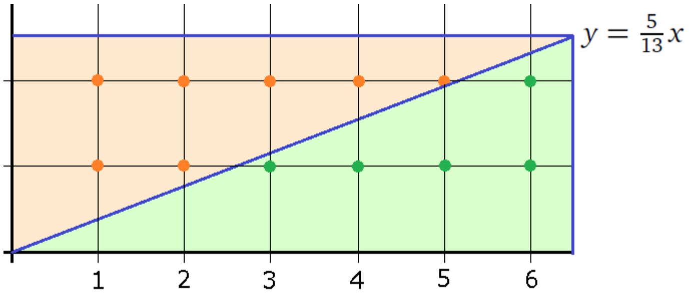
We have been building up to one of the most famous results in number theory, the *law of quadratic reciprocity*:

Theorem 44. *(Law of quadratic reciprocity) Let p and q be odd primes. Then*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}.$$

In short, if either p or q is a $4k+1$ prime, then $(\frac{p}{q}) = (\frac{q}{p})$, and if both p and q are $4k+3$ primes, then $(\frac{p}{q}) = -(\frac{q}{p})$. This is a bit of a surprising connection between prime numbers.

We can use Eisenstein's lemma to prove it. The sum from Eisenstein's lemma, $\sum_{k=1}^{(p-1)/2} \left\lfloor \frac{ka}{p} \right\rfloor$ has a nice geometric interpretation. It is the number of lattice points (points with integer coordinates) under the line $y = \frac{a}{p}x$ and above the x-axis, between $x = 0$ and $x = p/2$. See the figure below for an example with $a = 5$ and $p = 13$.



For example, the height of the line $y = \frac{5}{13}x$ at $x = 6$ is about 2.3 and so there are 2 (that is, $\lfloor \frac{6\times 5}{13} \rfloor$) lattice points below that point and above the axis. Eisenstein's lemma for this example can be rephrased to say that $(\frac{5}{13}) = (-1)^n$, where n is the number of lattice points under $y = \frac{5}{13}x$ and above the axis between $x = 0$ and $x = 13/2$. By symmetry, $(\frac{13}{5})$ is $(-1)^m$, where m is the number of lattice points to the left of $x = \frac{13}{5}y$ and right of the y-axis between $y = 0$ and $y = 5/2$.

We can think of $(\frac{5}{13})$ as coming from a count of the lattice points in the interior of the bottom green triangle shaded in the figure above and $(\frac{13}{5})$ as coming from a count of the lattice points in the top pink triangle. Those two triangles fit together to give a rectangle. Inside that rectangle there are exactly $((5-1)/2)\cdot((13-1)/2) = 12$ lattice points. Therefore, we have

$$\left(\frac{5}{13}\right)\left(\frac{13}{5}\right) = (-1)^m(-1)^n = (-1)^{m+n} = (-1)^{\frac{5-1}{2}\cdot\frac{13-1}{2}}.$$

Here is a proof of the law of quadratic reciprocity:

Proof. Assume p and q are prime. Then $m = \sum_{k=1}^{(p-1)/2} \left\lfloor \frac{kq}{p} \right\rfloor$ is the number of lattice points below the line $y = \frac{q}{p}x$ and above the x-axis, $x = 0$ and $x = p/2$. Similarly, $n = \sum_{k=1}^{(q-1)/2} \left\lfloor \frac{kp}{q} \right\rfloor$ is the number of lattice points to left of the line $x = \frac{q}{p}y$ and right of the y-axis, between $y = 0$ and $y = q/2$. None of the lattice points are the same since none of them can lie on the line $y = \frac{q}{p}x$ (which is the same as $x = \frac{p}{q}y$), since if some lattice point (x_0, y_0) lies on that line, then we would have $py_0 = qx_0$, which can only happen for y_0 a multiple of q and x_0 a multiple of p since p and q are prime.

The lattice points are all the lattice points of the interior of the rectangle running from $x = 0$ to $p/2$ and $y = 0$ to $q/2$. There are $\frac{p-1}{2}\frac{q-1}{2}$ of them, so $m + n = \frac{p-1}{2}\frac{q-1}{2}$. By Eisenstein's lemma, we have $(\frac{p}{q}) = (-1)^m$ and $(\frac{q}{p}) = (-1)^n$. Thus we have

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^m(-1)^n = (-1)^{m+n}(-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}.$$

□

The law of quadratic reciprocity can be rephrased as follows:

> If either p or q is of the form $4k + 1$, then $(\frac{p}{q}) = (\frac{q}{p})$. Otherwise, $(\frac{p}{q}) = -(\frac{q}{p})$.

It's quite surprising in that there is no *a priori* reason why the congruence $x^2 \equiv p \pmod{q}$ should have anything at all to do with the congruence $x^2 \equiv q \pmod{p}$. It is what mathematicians call a *deep theorem*, in that it is not at all obvious. The proof we gave requires Fermat's little theorem, Euler's identity, Gauss's lemma, and some other little parts. Gauss, who came up with the first proof, said he struggled for a year trying to figure out a proof. Euler, who conjectured the result, was unable to prove it. There are well over 100 proofs known for quadratic reciprocity, many of them quite different from the others.

Quadratic reciprocity is responsible for much of modern number theory, as generalizations of it have occupied a lot of number theorists' time. One such generalization was a key part of the proof of Fermat's last theorem.

Here are a couple of applications of quadratic reciprocity.

1. Suppose we want to compute $(\frac{7}{47})$. Both of these are $4k + 3$ primes, so by quadratic reciprocity, we have $(\frac{7}{47}) = -(\frac{47}{7})$. We are trying to see if 47 is a quadratic residue of 7, so we can reduce 47 mod 7 to get 5. So we can compute $(\frac{5}{7})$, which by quadratic reciprocity is the same as $(\frac{7}{5})$. We can reduce this to $(\frac{2}{5})$ and our rule from earlier tells us that 2 is not a quadratic residue of 5 since 5 is not of the form $8k \pm 1$. In short we have

$$\left(\frac{7}{47}\right) = -\left(\frac{47}{7}\right) = -\left(\frac{5}{7}\right) = -\left(\frac{7}{5}\right) = -\left(\frac{2}{5}\right) = -(-1) = 1.$$

   So 7 is a quadratic residue of 47.

2. As another example, we have $(\frac{13}{43}) = -(\frac{43}{13}) = -(\frac{4}{13})$. We can't use quadratic reciprocity to compute $(\frac{4}{13})$ since 4 is not prime, but we can use the fact that the Legendre symbol is multiplicative to write $(\frac{4}{13}) = (\frac{2}{13})(\frac{2}{13})$. Then, since 13 is not of the form $8k \pm 1$, 2 is not a quadratic residue of 13, and so overall we get $(\frac{13}{43}) = -1$. So 13 is not a quadratic residue of 43.

3. In general, we have $(\frac{3}{p}) = 1$ if and only if p is of the form $12k \pm 1$. We can write $(\frac{3}{p}) = (\frac{p}{3})$ if p is of the form $4k + 1$ and $(\frac{3}{p}) = -(\frac{p}{3})$ if p is of the form $4k + 3$. The only quadratic residue of 3 is 1, so $(\frac{p}{3}) = 1$ if and only if p is of the form $3k + 1$. So one way for $(\frac{3}{p})$ to equal 1 is if $p \equiv 1 \pmod{4}$ and $p \equiv 1 \pmod{3}$. Combining these congruences, we get that this will happen when $p \equiv 1 \pmod{12}$. The other way for $(\frac{3}{p})$ to equal 1 is if $p \equiv 1 \pmod{4}$ and $p \equiv 2 \pmod{3}$, which happens when $p \equiv -1 \pmod{12}$.

4. A nice application of quadratic reciprocity is Pepin's test, which is used to tell if a Fermat number, $F_n = 2^{2^n} + 1$ is prime. Fermat numbers are covered in Section 2.10. Pepin's test is as follows:

   > Let $n > 0$. Then F_n is prime if and only if $3^{(F_n-1)/2} \equiv -1 \pmod{F_n}$.

   Here is how we can prove this fact. First suppose F_n is prime. By Euler's identity, $3^{(F_n-1)/2}$ tells us whether 3 is a quadratic residue of F_n. We know that 3 is a quadratic residue of a prime p if and only if p is of the form $12k \pm 1$. However, all Fermat numbers besides F_0 are of the form $12k + 5$. This is because $2^{2^1} \equiv 4 \pmod{12}$ and $2^{2^{n+1}} \equiv \left(2^{2^n}\right)^2 \equiv 4^2 \equiv 4 \pmod{12}$ So we must have $3^{(F_n-1)/2} \equiv -1 \pmod{F_n}$.
   On the other hand, suppose $3^{(F_n-1)/2} \equiv -1 \pmod{F_n}$. Squaring this tells us that $3^{F_n-1} \equiv 1 \pmod{F_n}$, so the order of 3 modulo F_n must be a divisor of $F_n - 1 = 2^{2^n}$, a power of 2. Its proper divisors are 2, 4, 8, 16, …, $(F_n - 1)/2$. But since

$3^{(F_n-1)/2} \equiv -1 \pmod{F_n}$, $3^d \not\equiv 1 \pmod{F_n}$ for any proper divisor d of $F_n - 1$. So the order of 3 is $F_n - 1$. Since the order is a divisor of $\phi(F_n)$, we have $\phi(F_n) = F_n - 1$, which means that F_n is prime.

We can implement Pepin's test in Python like below:

```python
def pepin(n):
        f = 2**(2**n)+1
        return pow(3,(f-1)//2,f) == f-1
```

My laptop was able to verify that F_{14} was not prime in about 10 seconds. It took 80 seconds to verify F_{15} is not prime and about 10 minutes to show F_{16} is not prime. The largest one ever done, according to Wikipedia, was F_{24} in 1999. Note that this requires raising 3 to a humongous power, as F_{24} is over 5 million digits long.

The Jacobi symbol

Just a quick note: There is an important generalization of the Legendre symbol to all positive integers, called the *Jacobi symbol*. If the prime factorization of n is $p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$, then the Jacobi symbol is defined by

$$\left(\frac{a}{n}\right) = \left(\frac{a}{p_1}\right)^{e_1} \left(\frac{a}{p_2}\right)^{e_2} \cdots \left(\frac{a}{p_k}\right)^{e_k},$$

where the symbols $(\frac{a}{p_i})$ are Legendre symbols. The Jacobi symbol has many of the same properties as the Legendre symbol. The one notable exception is that it doesn't tell us wither a is a quadratic residue or not. However, it turns out to have a number of important applications in math and cryptography.

Chapter 4

Cryptography

Since the 1970s number theory has been a critical part of cryptography. The following sections detail two important modern cryptographic algorithms: Diffie-Hellman and RSA.

One thing that should be noted before proceeding is that it is usually recommended that you not try to write your own cryptographic routines in any important program as there are many details to get right (including things that you would probably never think of). Any lapse can make it easy for attackers to break your system.

4.1 Diffie-Hellman key exchange

Most forms of cryptography require a key. For example, one of the simplest methods is the substitution cipher. Under the substitution cipher, each letter of the alphabet is replaced with another letter of the alphabet. For instance, maybe A is replaced by Q, B is replaced by W, and C is replaced by B, like in the figure below:

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
Q	W	B	R	S	E	T	P	U	M	F	X	H	V	Z	O	K	A	G	Y	C	N	D	J	I	L

The message *SECRET* is encoded as *GSBASY*. The substitution cipher's key is the replacement alphabet, QWBRSETPUMFXHVZOKAGYCNDJIL. A major obstacle to the substitution cipher and many more sophisticated methods is that both the person sending the message and the person decrypting the message need to have a copy of the key. Safely transferring that key can be a serious challenge since you would not want your key to be intercepted.

The Diffie-Hellman key exchange algorithm is a way for two people, usually called Alice and Bob, to publicly create a shared secret number (a key) known only to them. Imagine they are shouting some numbers back and forth to each other across a crowded room. Anyone can hear the numbers that Alice and Bob are yelling to each other, but at the end of the process Alice and Bob will have a shared secret number that only they will know.

It starts with Alice and Bob picking a large prime p and a generator g. The generator g is usually a primitive root modulo p, or at least an element of high order. These values are not secret.

Once p and g have been chosen, Alice picks a secret random number a and Bob picks a secret random number b. Alice sends $g^a \bmod p$ to Bob and Bob sends $g^b \bmod p$ to Alice. Alice computes $(g^b)^a = g^{ab} \bmod p$ and Bob computes $(g^a)^b = g^{ab} \bmod p$. Alice and Bob now have the value g^{ab} in common and this is the secret key. Even though g^a and g^b were sent publicly, only someone who knows a or b can easily compute g^{ab}.

We say *easily compute* g^{ab} because, in theory, someone could compute g^{ab} just knowing p, g, g^a, and g^b (which were all sent in public), but if p is sufficiently large, there is no known way to do this efficiently.

Here is an example with $p = 23$. It turns out that $g = 5$ is a primitive root. Alice and Bob agree on these values and don't need to keep them secret. Then Alice and Bob pick their secret numbers, say Alice picks $a = 6$ and Bob picks $b = 7$. Alice then sends Bob $g^a \bmod p$, which is $5^6 \bmod 23$, or 8. Bob sends Alice $g^b \bmod p$, which is $5^7 \bmod 23$, or 17. These values, 8 and 17, are sent publicly. Alice then computes $(g^b)^a \bmod p$, which is $17^6 \bmod 23$, and Bob computes $(g^a)^b \bmod p$, which is $8^7 \bmod 23$. Both of these values work out to $g^{ab} \bmod p$, which is 12. This is the shared key.

Someone monitoring Alice and Bob's exchanges would see the values 23, 5, 8, and 17. For this small example, they could solve $5^a \equiv 8 \pmod{23}$ or $5^b \equiv 17 \pmod{23}$ by brute-force to get a and b, as there are only 22 possible values for each of a and b. Once they have a and b, they can easily compute the shared key, g^{ab}. However, if p is very large, then this brute-force search would be infeasible. There are other techniques that are better than a brute-force search, but if p is large enough, these techniques are still computationally infeasible. The problem of determining g^{ab} from the public information is known as the *Diffie-Hellman problem*. More generally, given g, p, and a value c, solving $g^x \equiv c \pmod{p}$ is called the *discrete logarithm problem*. Both are currently considered intractable for large p.

The shared secret number created by this algorithm can be used as a key for some other cryptographic algorithm, like DES or a more modern variant of it.

Possible problems with Diffie-Hellman

One major problem is that this method is subject to a so-called *man-in-the-middle attack*. This is where a third party, usually called Eve (for eavesdropper) sits in the middle, pretending to be Bob to Alice and pretending to be Alice to Bob. Alice and Bob think they are communicating with each other, but they are really communicating with Eve. Eve ends up with a key in common with Alice and a key in common with Bob. Alice and Bob will only be able to discover that it was Eve they were communicating with once Eve leaves and they find that they can't communicate with each other (because they have different keys). The solution to this problem is to add some form of authentication so that Alice and Bob can be sure that they are communicating with each other.

Another problem is a poor choice of p and g. For example, suppose $p = 61$ and $g = 13$. The problem is that $g^3 = 1$, so that the powers of g get caught in the repeating cycle 1, 13, 47, 1, 13, 47, ... With only three values to choose from, it will be pretty easy to guess what g^{ab} is. This is the reason that g is supposed to be a primitive root; we want the powers of g to take on a wide range of values to make brute-forcing infeasible. Actually, because of this, it is not necessary that g be a primitive root as long as the *order* of g is relatively large.

In practice, p is chosen to be what's called a *safe prime*, a prime of the form $2q+1$, where q is also prime.[1] Since $\phi(p) = p-1 = 2q$ and q is prime, the only possible orders mod p are 1, 2, q, and $2q$ (recall that the order is always a divisor of $\phi(n)$). Only 1 has order 1 and only $p-1$ has order 2 (see Theorem 27). The other elements have order q or $2q$. So any value except 1 and $p-1$ would be fine to use for g.

One further problem is not choosing a large enough value of p. Big primes are usually measured in bits. A 256-bit prime is a prime on the order of 2^{256}. This corresponds to $\log_{10}(2^{256}) \approx 77$ digits. Despite being a very large number, the discrete logarithm problem with primes of this size can be solved using a technique called a number field sieve. 1024-bit primes are a bit safer, but not recommended. Primes on the order of 2048 bits or larger are recommended. Of course, the question arises, why not use a truly huge prime far beyond what technology could ever break? The reason is using such huge numbers slows down the Diffie-Hellman algorithm too much and may require more computational power than a device (like a phone) can deliver. There is a tradeoff between speed and security.

Finally, it is worth noting that Diffie-Hellman can be used with other algebraic structures. The positive integers modulo p are a type of algebraic structure known as a *group*. There are many other types of groups whose elements are numbers or other types of objects, called groups, for which a kind of arithmetic works. Diffie-Hellman can be extended to these more general groups. In particular, *elliptic curve cryptography* involves using groups whose elements are rational points on elliptic curves. See Section 4.3.

4.2 RSA cryptography

RSA is similar to Diffie-Hellman in that both methods rely on the intractability of a number-theoretic problem. Diffie-Hellman relies on the difficulty of the discrete logarithm problem, while RSA relies on the fact that it is difficult to factor very large numbers.

The scenario is this: Alice wants other people to be able to send her secret messages. She posts a public key that anyone can use to encrypt messages. She maintains a private key (related to, but different from, the public key) that only she can use to decrypt the messages.

Alice creates the keys as follows: She picks two large prime numbers, p and q, and computes $n = pq$. The primes p and q must be kept secret, but n is part of the public key. Then Alice picks an integer e between 1 and $(p-1)(q-1)$ that shares no divisors

[1] Note that q is a Sophie Germain prime.

with $(p-1)(q-1)$. That integer is also part of the public key. She then finds a value d such that $de \equiv 1 \pmod{(p-1)(q-1)}$. In other words, d is the inverse of e modulo $(p-1)(q-1)$. This value is found with the extended Euclidean algorithm. And it is kept secret. In summary, n and e are public, while p, q, and d are kept private.

Here is how Bob can encrypt a message using Alice's public key: We'll assume the message is an integer a (text can be encoded as integers in a variety of different ways). Bob computes $a^e \bmod n$ and sends it to Alice. Alice can then decrypt the message using d. In particular, Alice computes $(a^e)^d \equiv a^{ed} \pmod{n}$.

We know that $ed \equiv 1 \pmod{(p-1)(q-1)}$ but not necessarily that $ed \equiv 1 \pmod{n}$. However, we have $\phi(n) = (p-1)(q-1)$, and since $ed \equiv 1 \pmod{\phi(n)}$, we can write $ed = 1 + k\phi(n)$. By Euler's theorem $a^{\phi(n)} \equiv 1 \pmod{n}$, so

$$(a^e)^d \equiv a^{ed} \equiv a^{1+\phi(n)k} = a \cdot (a^{\phi(n)})^k \equiv a \pmod{n}.$$

Here is an example: Suppose Alice chooses $p = 13$ and $q = 17$. Then $n = pq = 221$. Note also that $(p-1)(q-1) = 192$. Alice then chooses an e with no factors in common with 192, say $e = 11$. She then computes d such that $de \equiv 1 \pmod{(p-1)(q-1)}$, which in our case becomes $11d \equiv 1 \pmod{192}$. We get it by using the extended Euclidean algorithm, as follows (starting with the Euclidean algorithm):

$$192 = 11 \cdot 17 + 5$$
$$11 = 5 \cdot 2 + \boxed{1}$$

We can stop the Euclidean algorithm here as we see that the gcd will be 1. We then write

$$\begin{aligned} 1 &= -2 \cdot 5 + 1 \cdot 11 \\ &= -2 \cdot (192 - 11 \cdot 17) + 1 \cdot 11 \\ &= -2 \cdot 192 + 35 \cdot 11 \end{aligned}$$

Thus we have $11 \cdot 35 - 192 \cdot 2 = 1$, so $d = 35$.

Now suppose Bob encrypts a message $a = 65$. He computes $a^e \bmod n$, or $65^{11} \bmod 221$ to get 78. He sends this to Alice. Alice can decrypt it by computing $78^d \bmod n$, which is $78^{35} \bmod 221$ or 65.

Someone observing this communication would see $n = 221$, $e = 11$, as well as $a^e = 78$ pass by. In order to decrypt the message, they would need to solve $a^{11} \equiv 78 \pmod{221}$, factor n, or find the decryption exponent d. These are easy tasks for $n = 221$, but for large values of n there are no known efficient ways to do them.

Possible problems with RSA

There are a number of attacks possible on RSA, some of which are quite devious, so that anyone implementing RSA needs to be careful. Here is a short, incomplete list:

1. The first thing is that p and q need to be large primes in order to make it computationally infeasible to factor $n = pq$.
2. We also need to make sure p and q should not themselves be predictable. If the random number generator used to generate p and q is not completely random, then that leaves an opening for an attacker (and such openings have been exploited in the past).
3. Because of the way some factoring algorithms work, not just any large primes p and q will work. If $p-1$ or $q-1$ have a lot of small prime factors, those algorithms have an easier time factoring n. If $p-1$ and $q-1$ have small prime factors it also makes it more likely that e has a small order (since the order of e divides $\phi(n) = (p-1)(q-1)$).
4. It can be shown that if an attacker is able to figure out just 1/4 of the bits of n (specifically the 1/4 least significant or the 1/4 most significant bits), it is possible to use those bits to efficiently find d.
5. The value of n should not be reused among different people. Suppose Alice has n, d, and e and Bob also uses n but with a different d and e. Bob can use his d to factor n. Then if Bob intercepts a message encrypted with Alice's public key, he can easily decrypt it.
6. If the value of d is too small, there is an efficient way for an attacker to figure out d. The problem with using a large value of d is it might make the RSA algorithm too computationally intense for some devices. One way around this is the Chinese remainder theorem. To compute a^d modulo $n = pq$, compute a^d modulo p and q separately and then use the Chinese remainder theorem to combine the two parts. This delivers a nice speedup and is used in practice.
7. If the message a^e to be sent is less than n, then solving $a^e \equiv b \pmod{n}$ reduces to just finding the ordinary eth root of b, which can be done very easily.

8. RSA must be implemented with a padding scheme, where letters of the message are permuted and random stuff is added to the message (this process can be undone by the receiver). If a padding scheme is not used, then RSA is susceptible to the following attacks.

   (a) There can be a problem if e is too small. Suppose Alice sends the same message a to three different people using $e = 3$ and three values of n, say n_1, n_2, and n_3. We have three congruences: $x \equiv a^3 \pmod{n_1}$, $x \equiv a^3 \pmod{n_2}$, and $x \equiv a^3 \pmod{n_3}$. We can combine those into one congruence: $x \equiv a^3 \pmod{n_1 n_2 n_3}$. The problem with this is since $a < n_1$, $a < n_2$, and $a < n_3$, we have $a < n_1 n_2 n_3$, and so solving $x \equiv a^3 \pmod{n_1 n_2 n_3}$ reduces to just computing an ordinary (not a modular) cube root.

   In general, if we send e or more messages using the same e, then this attack can be applied, unless a padding scheme is used. This is important because small values of e are often used in practice to speed up the RSA algorithm. This is important for devices without a lot of power that can't don't handle serious computations well.[1]

   (b) Another attack, called a chosen plaintext attack, involves intercepting some encrypted text and then trying to encrypt likely messages using the public key until you get something that matches the text that you intercepted. Many messages start with something predictable, like "Dear so and so" or maybe some information about the sender or something like that. Once an attacker knows a few different inputs and their corresponding outputs, they will have an easier time breaking the encryption.

   (c) There is a related attack, called a chosen ciphertext attack. One way this attack works is if an attacker has access to the decryption algorithm, though not the actual values of d, p, or q. The attacker decrypts a bunch of messages in an attempt to learn something about those values. Another way this can work is as follows: Suppose Eve intercepts a message a^e that Bob is sending to Alice. Eve wants to know what a is. She picks some integer b, and sends $(ab)^e$ to Alice. Alice decrypts the message, which should come out as garbage because of the extra factor that Eve added. If Eve can somehow convince Alice to send her the decrypted message (possibly by pretending to be Bob), then Eve can figure out Bob's original message. This is because Alice decrypts $(ab)^e$ into ab and if she sends that to Eve, then Eve can just divide by b to find a.

9. One particularly interesting class of attacks on RSA is what are known as side-channel attacks. In these, an attacker observes the state of a computer's CPU while it is encrypting and decrypting, in particular, when it is raising numbers to powers. The CPU works harder at some points of the process than others, depending on the bit pattern of the key. An attacker can look at where the CPU is working harder and where it isn't in order to determine the bit pattern of the key (and hence the key itself). They can do this, for instance, by placing a cell phone nearby the computer. The CPU makes a high frequency noise that varies based on how hard it is working, and there are programs that can pick up and decode the changes in the sound to figure out the key. Another approach involves putting one hand on the computer and holding a voltmeter in the other to detect small changes in power output of the CPU.[2] Still another approach simply relies on timing how long it takes the CPU to perform various steps of the encryption process. In order to stop this, if you implement RSA, you would need to disguise things so that the CPU is working at the same rate at all times.

10. It is also theoretically possible to factor $n = pq$, even if p and q are large, if you have a quantum computer. A regular computer is based on bits that have one of two states: on or off (0 or 1). A quantum computer, using the properties of quantum physics, has qubits instead of bits, which can exist in a variety of states from 0 to 1. It has been proved that a quantum computer could quickly factor pq even for very large values of p and q. However, the largest quantum computers that have been built consist of only a few qubits and haven't been able to factor numbers larger than 100.

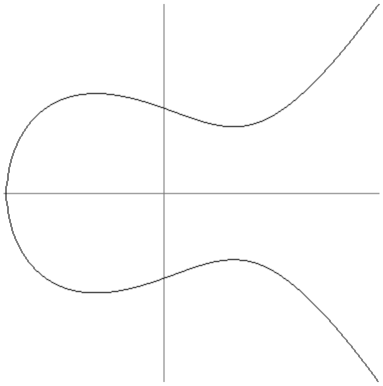
4.3 Elliptic curve cryptography

This section provides a brief introduction to elliptic curve cryptography. We will leave many mathematical details out and oversimplify some things.

Elliptic curves are curves with equations of the form $y^2 = x^3 + ax + b$. A typical elliptic curve looks a lot like the curve shown below:

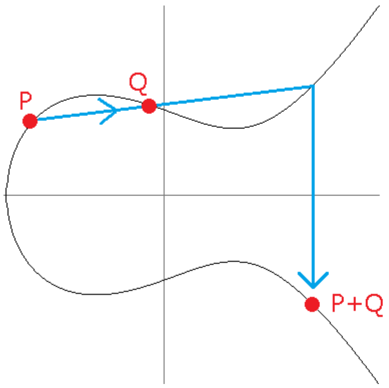
[1] Actually, e is often chosen to be a Fermat prime (a prime of the form $2^{2^n} + 1$) since exponentiation by repeated squaring on Fermat primes is much faster than with other exponents.

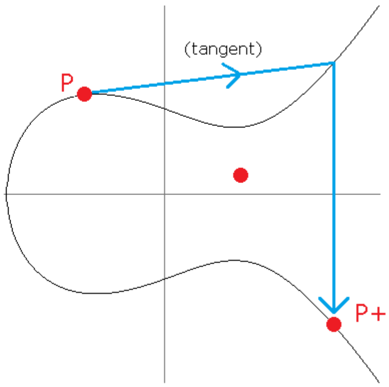
[2] See http://www.cs.tau.ac.il/~tromer/acoustic/.



Mathematicians have been interested for a while in rational points on such curves; namely, points (x, y) whose coordinates are both rational numbers. For instance, the curve $y^2 = x^3 - x + 1$ contains the points $(1/4, 7/8)$ and $(3, 5)$, both of which have rational coordinates. See Section 5.2 for a nice application of using rational points on the unit circle to find Pythagorean triples.

We can define a way to combine rational points on the curve to get new points. Many lines will intersect an elliptic curve at three points. If we connect two rational points on an elliptic curve by a line, that line will meet the curve in one other point, and it's not too hard to show that point will also have rational coordinates. We then reflect that point across the x-axis, using the fact that elliptic curves are symmetric about the x-axis, to get a new point. So given two different points, P and Q, on the curve, $P + Q$ is defined to be this new point. See the figure below on the left.





A line that is tangent to the curve may only intersect the curve in two points. We can use this to define a rule for $P + P$. Namely, we follow the tangent line from P until it hits the curve and then reflect across the x-axis, like in the figure above on the right.

The only other possibilities for lines intersecting the curve are vertical lines, which can meet the curve in one or two points. The key to understanding them is there one other point, called *the point at infinity*, that we need to add to our curve. We can sort of think of it sort of sitting out at infinity. We use the symbol 0 for it. It acts as the additive identity. The vertical lines are important for showing that $P + 0 = P$ and $P - P = 0$, where $-P$ is the point obtained by reflecting P across the x-axis.

We are omitting a lot of technical details here. But the important point is that this addition operation makes the rational points into a mathematical object called an abelian group. Basically, this means that the addition operation behaves nicely, obeying many of the rules that ordinary addition on integers satisfies.[1]

It is possible to work out formulas for $P + Q$ and $P + P$. If P has coordinates (x, y), and Q has coordinates (x', y'), then $P + Q$ has coordinates $(\lambda^2 - x - x', \lambda(x - x') - y)$, where $\lambda = \frac{y'-y}{x'-x}$. And $P + P$ has coordinates $(\mu^2 - 2x, \mu(x - (\mu^2 - 2x)) - y)$, where $\mu = \frac{3x^2+a}{2y}$.

For cryptography, we use modular arithmetic with elliptic curves. For instance, say we use arithmetic in $\mathbb{Z}_7$ on $y^2 = x^3 - x + 1$. In that case, the point $(5, 2)$ is on the curve since $2^2 \equiv 5^3 - 5 + 1 \pmod{7}$. We define addition of points on the curve using the formulas given above, but in place of division we use the modular inverse. For instance, instead of doing $4/3$ in $\mathbb{Z}_7$, we would do $4 \cdot 3^{-1}$, which is $4 \cdot 5 \equiv 6 \pmod{7}$.

[1] An abelian group, roughly speaking, is a set along with an operation that is commutative, associative, has an additive identity akin to the number 0, and every element of the set has an additive inverse.

Finally, to actually do cryptography, we pick a curve (that is, we pick values of a and b in $y^2 = x^3 + ax + b$). There are certain values that people suggest to use. Then we pick a large prime p so that all the arithmetic will be done modulo p. Then we pick a random point G on the curve and a random integer a. We then compute aG, which denotes G added to itself a times. The public key is aG and the private key is a. Note the similarity with Diffie-Hellman, where we have a generator g and a random integer a and g^a is sent publicly. If the group of points on the curve is large, it is thought to be very difficult for someone to recover a from aG (which will appear as just a random point on the curve).

Once we have a way of generating a public and private key like this, we can do all sorts of cryptographic things, including analogs of Diffie-Hellman key exchange, secure communication of messages, and more. The benefits of elliptic curve cryptography over cryptography with ordinary modular arithmetic are that arithmetic on elliptic curves is less computationally intensive than raising numbers to large powers and the best known algorithms for brute-force breaking the private key in elliptic curve cryptography are not as good as the best-known algorithms for brute-force breaking the private key in ordinary modular arithmetic. In short, you can get more security for less computational effort.

For more information, do a web search for *A Tutorial on Elliptic Curve Cryptography* by Fuwen Liu or *A (Relatively Easy To Understand) Primer on Elliptic Curve Cryptography* by Nick Sullivan.

Chapter 5

Special numbers

5.1 Perfect numbers and Mersenne primes

A number is called *perfect* if it equals the sum of its proper divisors. For example, 6 is perfect because 6 is the sum of its proper divisors, 1, 2, and 3. Recall that $\sigma(n)$ denotes the sum of the divisors of n. So we can define perfect numbers as below:

Definition 16. *A positive integer n is called* perfect *if $\sigma(n) = 2n$.*

Perfect numbers have been of interest to mathematicians since at least the time of the ancient Greeks. They knew of four perfect numbers: 6, 28, 496, and 8128. The next perfect number, 33,550,336, was not found until possibly as late as the 15th century. The next one after that, 8,589,869,056 was discovered in the late 16th century.

Perfect numbers that are even are associated with a particular type of number called a *Mersenne number*, named for the 17th century monk and mathematician, Marin Mersenne.

Definition 17. *A positive integer of the form $2^n - 1$ is called a* Mersenne number. *If it is prime, it is called a* Mersenne prime.

The relationship between perfect numbers and Mersenne primes is that whenever $2^n - 1$ is a Mersenne prime, then $2^{n-1}(2^n - 1)$ is perfect. For instance, the first Mersenne prime is $2^2 - 1 = 3$ and $2^1(2^2 - 1) = 6$ is perfect. The next Mersenne prime is $2^3 - 1 = 7$ and $2^2(2^3 - 1) = 28$ is perfect. The next two Mersenne primes are $2^5 - 1$ and $2^7 - 1$, which correspond to the perfect numbers 496 and 8128.

The next several exponents of Mersenne primes are 13, 17, 19, 31, 61, 89, 107, 127. The next one, 521, was found by a computer search in 1952. As of 2019, there were only 51 known Mersenne primes, the largest of which has exponent 82,589,933. It corresponds to a number with close to 25 million digits. It is also the largest prime number that had been found. It is suspected that there are infinitely many Mersenne primes (and hence perfect numbers), but no one has been able to prove it.

One thing to note is that if $2^n - 1$ is prime, then n itself must be prime. This is because if $n = ab$ is not prime, then $2^{ab} - 1$ can be factored into $(2^a - 1)(1 + 2^a + 2^{2a} + 2^{3a} + \cdots + 2^{(b-1)a})$. A similar factorization shows that there are no primes of the form $3^n - 1$, $4^n - 1$, etc. as $m^n - 1$ is divisible by $m - 1$.

We record the relationship between Mersenne primes and perfect numbers in the following theorem. A proof of it appears in Euclid.

Theorem 45. *If $2^n - 1$ is prime, then $2^{n-1}(2^n - 1)$ is perfect.*

Proof. Let $k = 2^{n-1}(2^n - 1)$. We have to show that $\sigma(k) = 2k$. Since σ is multiplicative and $\gcd(2^{n-1}, 2^n - 1) = 1$, we have $\sigma(k) = \sigma(2^{n-1})\sigma(2^n - 1)$. The formula for computing σ tells us that $\sigma(2^{n-1}) = 2^n - 1$, and we have $\sigma(2^n - 1) = 2^n$ since $2^n - 1$ is prime. Putting this together gives us $\sigma(k) = 2k$. □

We can take things one step further.

Theorem 46. *Every even perfect number is of the form $2^{n-1}(2^n - 1)$, where $2^n - 1$ is a Mersenne prime.*

Proof. Let k be an even perfect number. Factor as many twos as possible out to write $k = 2^{n-1}m$, where m is odd and $n \geq 2$. We have

$$\sigma(k) = \sigma(2^{n-1})\sigma(m) = (2^n - 1)\sigma(m).$$

Since k is perfect, $\sigma(k) = 2k = 2^n m$, and so we have

$$2^n m = (2^n - 1)\sigma(m).$$

Thus $2^n - 1$ is a divisor of $2^n m$. By Euclid's lemma, since $\gcd(2^n - 1, 2^n) = 1$, we must have $2^n - 1 \,|\, m$. So we can write $(2^n - 1)j = m$ for some integer j. Plugging this in to the equation above and simplifying gives $2^n j = \sigma(m)$.

We know that both j and m are divisors of m, so $\sigma(m) \geq j + m = j + (2^n - 1)j = 2^n j = \sigma(m)$. But 1 is also a divisor of m and wasn't included in that sum, so we have a contradiction unless $j = 1$. In that case, the only divisors of m are 1 and itself, meaning m is prime and further that $m = 2^n - 1$ and $k = 2^{n-1}(2^n - 1)$. □

The above theorems tell us about even perfect numbers but not about odd perfect numbers. In fact, no odd perfect numbers have ever been found. Mathematicians have not been able to prove that they don't exist, but if they do, it would come as a surprise. There are a number of things that have been proved must be true of an odd perfect number, should one exist:

- It must be larger than 10^{1500}.
- It must be congruent to 1 mod 12, 117 mod 468, or 81 mod 324.
- It must have at least 9 distinct prime factors, the largest of which is greater than 10^8

There are quite a few others. See the Wikipedia page on perfect numbers for more.

An interesting fact about the even perfect numbers we've seen (6, 28, 496, 8128, 33,550,336, and 8,589,869,056) is that they all end in 6 or 8. This is in fact true for all even perfect numbers. Every even perfect number is of the form $2^{n-1}(2^n - 1)$, where n is prime. Since n is prime, either $n = 2$ or n is of the form $4k \pm 1$. If $n = 2$, we get the perfect number 6. If n is of the form $4k + 1$, then a simple calculation allows us to reduce $2^{n-1}(2^n - 1)$ to 6 mod 10. A similar calculation for $n = 4k - 1$ reduces $2^{n-1}(2^n - 1)$ to 8 mod 10.

Finding Mersenne primes

There is a reason why the largest known primes are all Mersenne primes: there is an easy test to tell if a Mersenne number is prime:

Theorem 47. *(Lucas-Lehmer test for Mersenne primes) Let $S_1 = 4$ and $S_{k+1} = S_k^2 - 2$ for $k \geq 1$. Then $2^n - 1$ is prime if and only if $2^n - 1 \,|\, S_{n-1}$.*

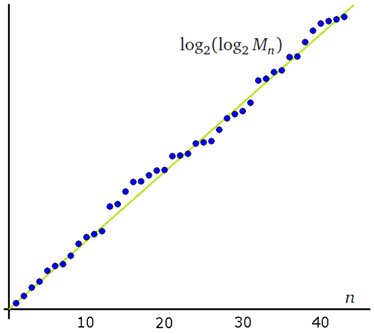
For example, let's use it to show that $2^7 - 1 = 127$ is prime. We have to show that $127 | S_6$, or equivalently that $S_6 \equiv 0 \pmod{127}$. Working mod 127, we start with $S_1 = 4$. We then get the following:

$$\begin{aligned}
&S_2 = S_1^2 - 2 = 14\\
&S_3 = S_2^2 - 2 = 194 \equiv 67 \pmod{127}\\
&S_4 = S_3^2 - 2 = 4487 \equiv 42 \pmod{127}\\
&S_5 = S_4^2 - 2 = 1762 \equiv -16 \pmod{127}\\
&S_6 = S_5^2 - 2 = 254 \equiv 0 \pmod{127}
\end{aligned}$$

As S_6 is congruent to 0 mod 127, we conclude that $2^7 - 1$ is a Mersenne prime.

There are various optimizations that can be made to make this process more efficient. The Lucas-Lehmer test is used by GIMPS, the Great Internet Mersenne Prime Search, which uses the idle time of computers around the world to search for new Mersenne primes. GIMPS has found most of the recent record prime numbers.

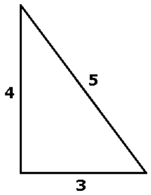
As mentioned earlier, it has not been proved that there are infinitely many Mersenne primes, but it is suspected that there are. A conjecture of Lenstra, Pomerance, and Wagstaff is that there are roughly $e^\gamma \log_2(\log_2 x)$ Mersenne prime numbers less than x, where γ is the Euler-Mascheroni constant. Evidence for this conjecture is shown in the graph below, which shows $\log_2(\log_2 M_n)$, where the M_n run through the first 43 Mersenne primes. Note the very nearly linear relationship.



See http://primes.utm.edu/mersenne/heuristic.html for a simple heuristic argument in favor of the conjecture.

5.2 Pythagorean triples

We've all seen a 3-4-5 triangle, like the one below, where all the sides are integers.



A 3-4-5 triangle is not the only one with integer sides. There are many others, like 6-8-10 and 5-12-13. Each of these triangles has integer sides (a,b,c) that satisfy the Pythagorean theorem $a^2+b^2=c^2$. We make the following definition:

Definition 18. *A triple of positive integers, (a,b,c), is called a* Pythagorean triple *if $a^2+b^2=c^2$.*

Given any triple (a,b,c), we can generate infinitely many other triples by multiplying through by a constant. That is, for any integer $k\in\mathbb{Z}$, (ka,kb,kc) is also a Pythagorean triple. This is because $(ka)^2+(kb)^2=k^2(a^2+b^2)=k^2c^2$. For instance, $(3,4,5)$ leads to $(6,8,10)$, $(9,12,15)$, $(12,16,20)$, etc.

We are not too interested in Pythagorean triples that are multiples of other ones, so we make the following definition.

Definition 19. *A Pythagorean triple (x,y,z) is called* primitive *if* $\gcd(x,y,z)=1$.

So $(3,4,5)$ and $(5,12,13)$ are primitive. Can we find a primitive Pythagorean triple that includes the integer 7? The answer is yes. We want to find b and c such that $7^2+b^2=c^2$. We can rewrite this as $c^2-b^2=7^2$ or $(c-b)(c+b)=49$. Since c and b are integers, $c-b$ will be a divisor of 49 and $c+b$ will be its complement. We have 49 equal to 1×49 or 7×7. The latter doesn't give a solution, but the former does. We have $c-b=1$ and $c+b=49$. Solving this system gives $b=24$ and $c=25$, so $(7,24,25)$ is a new primitive Pythagorean triple.

In general, one way to find Pythagorean triples involving the integer a is to write $a^2=(c-b)(c+b)$ and assign factors of a^2 to $c-b$ and $c+b$. For instance, with $a=15$, one way to factor $a^2=225$ is as 9×25. Setting $c-b=9$ and $c+b=25$ gives $c=17$ and $b=8$. So we get the triple $(8,15,17)$.

There is a nice formula describing all primitive Pythagorean triples.

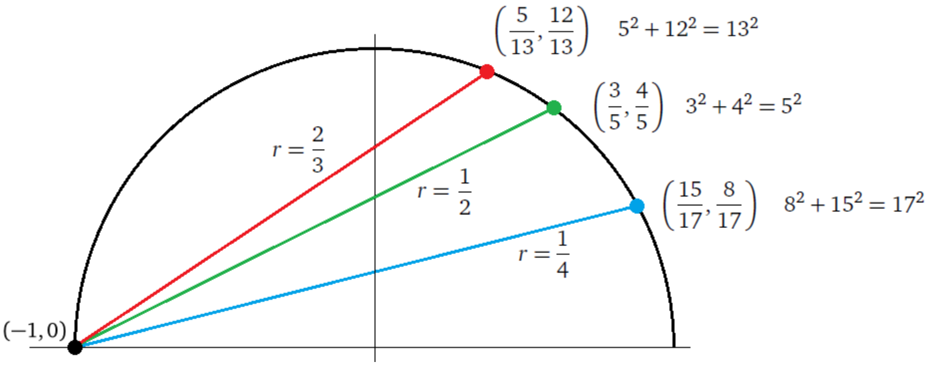
Theorem 48. *(Euclid's formula) Every primitive Pythagorean triple is of the form $(2mn, m^2-n^2, m^2+n^2)$ for some positive integers m and n, where $\gcd(m,n)=1$ and m and n are not both odd.*

For example, with $m=5$ and $n=2$, we get the triple $(20,21,29)$. If we remove the conditions on m and n, we still get Pythagorean triples, just not primitive ones. This formula was first proved by Euclid.

One way to prove it is to use the technique we used in the examples preceding the theorem. Here is a different argument that has connections to higher mathematics. Suppose we have $a^2+b^2=c^2$. Dividing through by c^2 gives $\left(\frac{a}{c}\right)^2+\left(\frac{b}{c}\right)^2=1$.

This process can be reversed, so we see that each Pythagorean triple corresponds to a point on the unit circle with rational coordinates and vice-versa.

It is not too hard to show with a little algebra that if a line with rational slope intersects the unit circle at a rational point, then the other point of intersection must also be rational. Conversely, if we draw a line with rational slope from a rational point on the unit circle, then the other point of intersection with the circle will also be rational. Therefore, if we pick a convenient rational point on the unit circle (like $(-1,0)$) and draw lines with rational slope from that point, we will hit all of the other rational points on the unit circle (and hence find all the Pythagorean triples). See the figure below for a few example slopes and the rational points (and Pythagorean triples) they generate.


$\left(\frac{5}{13}, \frac{12}{13}\right)$
$5^2+12^2=13^2$
$\left(\frac{3}{5}, \frac{4}{5}\right)$
$3^2+4^2=5^2$
$\left(\frac{15}{17}, \frac{8}{17}\right)$
$8^2+15^2=17^2$
$r=\frac{2}{3}$
$r=\frac{1}{2}$
$r=\frac{1}{4}$
$(-1,0)$

A line through $(-1,0)$ with rational slope r has equation $y=r(x+1)$. Plugging this into the unit circle equation $x^2+y^2=1$ gives $x^2+(r(x+1))^2=1$. After a little algebra, we can write this as

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

From this, we get $x=\frac{1-r^2}{1+r^2}$ and plugging back into the line equation gives $y=\frac{2r}{1+r^2}$. Each value of r gives a different rational point (x,y) on the curve. If we write $r=m/n$ for some integers m and n, and convert the rational point into a triple, we get the desired formulas $a=n^2-m^2$, $b=2mn$, $c=n^2+m^2$.

Studying rational points on other curves, especially elliptic curves like $y^2=x^3+ax+b$, is a major focus of modern mathematics.

There are a number of interesting properties that a Pythagorean triple (a,b,c) must satisfy. Here are a few of them:

- One of a and b is even and the other is odd.
- Exactly one of a and b is divisible by 3.
- Exactly one of a and b is divisible by 4.
- Exactly one of a, b, and c is divisible by 5.

See the Wikipedia page on Pythagorean triples for more properties.

Fermat's last theorem

So we have seen that there are infinitely many integer solutions to $x^2+y^2=z^2$. What about $x^3+y^3=z^3$ or $x^4+y^4=z^4$? One of the most famous stories from math concerns these equations. Fermat wrote in the margin of a copy of Diophantus' *Arithmetica* that he had a proof that $x^n+y^n=z^n$ has no integer solutions and said he couldn't include it because the book's margin was too small to hold the proof. People tried for the next 350 years to prove it before it was finally resolved by Andrew Wiles in the 1990s.

Chapter 6

Primality testing and factoring

We will look at two important problems in number theory: determining if a number is prime and factoring a number. The former can be done relatively quickly, even for pretty large numbers, while there is no known fast way to do the latter.

6.1 Primality testing

As mentioned in Section 2.3, one way to tell if a number n is prime be to test if it is divisible by 2, 3, 4, …, $\sqrt{n}$. One improvement is to just check for divisibility by primes, but we might not know all the primes from 2 to $\sqrt{n}$. One thing we can do is to check if n is divisible by 2 or 3 and then check divisibility by all the integers of the form $6k \pm 1$ up through $\sqrt{n}$. These are the integers 5, 7, 11, 13, 17, 19, 23, 25, 29, 31, …. There is no need to check divisibility by integers of the form $6k$, $6k+2$, $6k+3$, or $6k+4$, since if a number is divisible by one of those numbers then it must already be divisible by 2 or 3.

Here is some Python code implementing this method.

```python
def is_prime(n):
    if n in [2,3,5,7,11,13,17,19,23]:
        return True
    if n==0 or n==1 or n%2==0 or n%3==0:
        return False
    d = 5
    step = 2
    stop = n**.5
    while d <= stop:
        if n % d == 0:
            return False
        d += step
        step = 4 if step == 2 else 2
    return True
```

On my laptop here is some data for how long it took this program to verify that various integers are prime:

prime	running time in seconds
100000000003	0.08
1000000000039	0.22
10000000000037	0.91
100000000000031	2.47
1000000000000037	8.14
10000000000000061	24.54

We see that adding a digit multiplies the running time roughly by 3. Modern cryptography needs primes that are hundreds of digits long. This algorithm would take around 10^{40} seconds to verify that a 100-digit number is prime, which is simply unreasonable. The running time here is exponential in the number of digits.

Probabilistic primality tests

Recall Fermat's little theorem, which states that if p is prime, then $a^{p-1} \equiv 1 \pmod{p}$ for any a relatively prime to p. The contrapositive of this statement gives us a way to show a number n is not prime: just find an a such that $a^{n-1} \not\equiv 1 \pmod{n}$. For instance, 10 is not prime because $2^9 \equiv 2 \pmod{10}$. Similarly, $2^{11} \equiv 8 \pmod{12}$, so 12 is not prime. Can we always use 2^{n-1} to show that n is not prime? No, though it usually does work. The smallest value for which this fails is $n = 341$. We have $2^{340} \equiv 1 \pmod{341}$ but 341 is not prime. Because of this, 341 is called a *pseudoprime* to base 2.

Note that even though we couldn't use $a = 2$ to show that 341 is not prime, we can use $a = 3$ since $3^{340} \equiv 56 \pmod{341}$. The question then becomes: if $a^n \equiv a \pmod{n}$ for *every* a relatively prime to n, is n prime? The answer, perhaps surprisingly, is still no. There are numbers, called *Carmichael numbers*, that are not prime and yet $a^n \equiv a \pmod{n}$ for every a relatively prime to n. The first several Carmichael numbers are 561, 1105, 1729, 2465, 2821, 6601, and 8911. Carmichael numbers are considerably more rare than primes, there being only 43 less than 1,000,000 and 105,122 less than 10^{15} (versus about 29 trillion primes less than 10^{15}). Despite their relative rarity, it was proved in 1994 that there are infinitely many of them.

Because they are so rare, one way to test if a number n is prime, is for several values of a to test if $a^n \equiv a \pmod{n}$. If it fails any test, then the number is composite. If it passes every test, then it might not be prime, but there is a good chance that it is.

This is an example of a *probabilistic primality test*. There is a small, but not negligible, chance of it being wrong. We can modify this test, however, to create a test whose probability of being wrong can be made vanishingly small. This new test is based on the following fact:

Theorem 49. *Let n be an odd prime, with $n-1$ factored into $2^s t$ for some integer s and odd integer t. Let a be relatively prime to n. Then one of the following congruences is true: $a^t \equiv 1 \pmod{n}$, $a^t \equiv -1 \pmod{n}$, $a^{2t} \equiv -1 \pmod{n}$, $a^{4t} \equiv -1 \pmod{n}$, …, $a^{2^{s-1}t} \equiv -1 \pmod{n}$.*

Proof. Since $n-1 = 2^s t$, we can factor $a^{n-1}-1$ into $(a^{2^{s-1}t}+1)(a^{2^{s-1}t}-1)$. We can then factor $a^{2^{s-1}t}-1$ into $(a^{2^{s-2}t}+1)(a^{2^{s-2}t}-1)$. Continuing this way, we get

$$a^{n-1}-1 = (a^{2^{s-1}t}+1)(a^{2^{s-2}t}+1)\cdots(a^{4t}+1)(a^{2t}+1)(a^t+1)(a^t-1).$$

By Fermat's little theorem, $a^{n-1}-1 = 0$, so one of the factors of the right side must be 0. Hence once of those congruences must hold. □

We can use this theorem as probabilistic primality test in a similar way that we use Fermat's little theorem. Here is the algorithm to test if n is prime. It is called the *Miller-Rabin probabilistic primality test*.

1. Factor as many twos as possible out of $n-1$ to write it as $n-1 = 2^s t$ with t odd.
2. Repeat the following steps several times:
   (a) Choose a random integer a in the range from 2 through $n-2$.
   (b) Consider the following congruences: $a^t \equiv 1 \pmod{n}$, $a^t \equiv -1 \pmod{n}$, $a^{2t} \equiv -1 \pmod{n}$, $a^{4t} \equiv -1 \pmod{n}$, …, $a^{2^{s-1}t} \equiv -1 \pmod{n}$.
   (c) If none of those congruences are true, then n is composite and we stop the algorithm. If at least one of those congruences is true, then go back to step (a).

Here is a Python implementation of this algorithm:

```python
def miller_rabin(n, tries=10):
    s = 0
    t = n-1
    while t%2 == 0:
        t //= 2
        s += 1

    for i in range(tries):
        a = random.randint(2,n-2)
        b = pow(a,t,n)
        if b!=1 and b!=n-1:
            for j in range(s):
                b = pow(b,2,n)
                if b == n-1:
                    break
            else:
                return False
    return True
```

The basic idea of the algorithm is that we apply the theorem to several random values of a. It can be shown that the probability that at least one of the congruences in the theorem is true and yet the number is still composite is at most 1/4 (and actually often quite a bit less). Assuming independence, if we repeat the process with k values of a, and some of the congruences are true each time, the probability that a composite will pass through undetected will be less than $(1/4)^k$.[1]

The upshot of all of this is that we just perform Step 2 of the algorithm several times (for large enough values even once or twice is enough), and if the algorithm doesn't tell us the number is composite, then we can be *nearly* certain that the number is prime.

It took my laptop a little over 7 seconds with one step of the test to verify (with high probability) that the 4000-digit number $1477! + 1$ is prime. It took about 16 minutes to show that $6380! + 1$ (a 21000-digit number) is prime.

There are efficient primality tests that are not probabilistic, but they are also not as fast as the Miller-Rabin test. If you can live with a very small amount of uncertainty, the Miller-Rabin test is a good way to test primality.[2]

6.2 Factoring

The simple way to factor a number is to check the possible divisors one-by-one. Just like with primality testing, there are more efficient ways to do things than the simple approach.

Fermat's method

Suppose we want to factor 9919. We might notice that it is $10000-81$, which is $100^2 - 9^2$ or $(100-9)(100+9)$. Thus we have written 9919 as 91×109. We could further factor this into $7 \times 13 \times 109$ if we like.

This leads to an approach known as *Fermat's method*. We systematically try to write our integer n as a difference of two squares, which we can then easily factor. This process will always work. Given $n = ab$, we can write $n = \left(\frac{a+b}{2}\right)^2 - \left(\frac{a-b}{2}\right)^2$.

Here is the general process: We want to write $n = x^2 - y^2$. We can rewrite this as $x^2 - n = y^2$. We start with x equal to the smallest perfect square greater than n and continually increment x by 1 unit until $x^2 - n$ is a perfect square. Here is a step-by-step description:

1. Let $x = \lceil \sqrt{n} \rceil$.
2. If $x^2 - n$ is a perfect square, then we can factor n into $(x-y)(x+y)$.
3. Otherwise, increase x by 1 and go to step 2.

For example, let $n = 119{,}143$. We start with $x = \lceil \sqrt{119143} \rceil = 346$. We then compute

$$\begin{aligned}
&346^2 - 119143 = 573 \\
&347^2 - 119143 = 1266 \\
&348^2 - 119143 = 1961 \\
&349^2 - 119143 = 2658 \\
&350^2 - 119143 = 3357 \\
&351^2 - 119143 = 4058 \\
&352^2 - 119143 = 4761 = 69^2.
\end{aligned}$$

[1] Note that since composites are so much more common than primes, the probability that the number is prime is not quite $(1/4)^k$. But actually the 1/4 probability is an extremely conservative estimate. For large numbers, the probability is actually much lower, so in fact the probability will always turn out to be less than $(1/4)^k$. For instance, Pomerance and Crandall in *Number Theory: A Computational Perspective*, 2nd edition, report that for a 150-digit prime, the probability is actually less than $1/4^{28}$, not 1/4.

[2] Note that if an important conjecture known as the extended Riemann hypothesis is true, then the Miller-Rabin test could be turned into a true primality test by running the test for all a less than $2(\ln n)^2$.

We stop at $x = 352$, since we get a perfect square at that step. We then factor n into $352^2 - 69^2$, which is $(352-69)(352+69)$ or 283 × 421, both of which are prime. Notice that this is considerably less steps than trial division. Fermat's method is good for finding factors close to $\sqrt{n}$. It is bad at finding small factors. Thus it is a good complement for trial division. Trial division can be used to find small factors and Fermat's method can be used to find the others.

Note of course that Fermat's method just finds one factor, d. To find more factors, we can run the algorithm or another on n/d. Here is a Python implementation of Fermat's method:

```
from math import floor, ceil

def is_perfect_square(n):
    return abs(n**.5 - floor(n**.5)) < 1e-14

def fermat_factor(n):
    x = ceil(n**.5)
    y = x*x - n
    while not is_perfect_square(y):
        x += 1
        y = x*x - n
    return (x-floor(y**.5), x+floor(y**.5))
```

Note the way we check if an integer n is a perfect square is if $|\sqrt{n} - \lfloor \sqrt{n} \rfloor|$ is less than some (small) tolerance. One way to speed this up a bit would be to save the value of $\sqrt{y}$ instead of computing it three separate times.

The Pollard rho method

We will consider one more factoring technique to give a sense for what factoring techniques are out there. This method is called the *Pollard rho* method.

Say we need to factor 221, which is 13 × 17. Consider iterating the function $f(x) = x^2 + 1$ starting with $x_0 = 1$. We get the following sequence of iterates:

$$1, 2, 5, 26, 14, 197, 135, 104, 209, 145, 31, 78, 118, 2, 5, 26, 14, \ldots$$

Notice that $x_4 - x_0 = 26 - 1$ is divisible by 13. Also, $x_5 - x_1 = 195$ and $x_6 - x_2 = 130$ are divisible by 13, and in general, $x_{m+4} - x_m$ is divisible by 13 for any $m \geq 4$. Further, notice that $x_7 - x_1 = 102$, $x_8 - x_2 = 204$, $x_9 - x_1 = 119$, etc. are all divisible by 17. This sort of thing will always happen for the divisors of an integer. This suggests a way to find factors of an integer n: Look at the sequence of iterates of $x^2 + 1$ and look at $\gcd(n, x_k - x_j)$. Eventually this should (but not always) lead to a factor of n.

To see why this works, consider iterating $f(x) = (x^2 + 1) \bmod 13$ starting with $x = 1$. We get the repeating sequence 1, 2, 5, 0, 1, 2, 5, 0, ... Notice the period of the repeat is 4, which corresponds to differences of the form $x_{m+4} - x_m$ being divisible by 13 in the iteration mod 221. Similarly, iterating $f(x) = (x^2 + 1) \bmod 17$ starting with $x = 1$, gives the sequence, 1, 2, 5, 9, 14, 10, 16, 2, 5, 9, which has a repeating cycle of length 6 starting at the second element. This corresponds to differences of the form $x_{m+6} - x_m$ being divisible by 17 in the iteration mod 221. Note that $x_6 - x_0$ above is not divisible by 17 as the repeating pattern mod 17 doesn't start until the second term of the sequence.

In general, if we iterate $f(x) = (x^2 + c) \bmod n$ for any integers c and n, and any starting value, we will eventually[1] end up in a repeating cycle. This is because the sequence can only take on a finite number of values, so eventually we must get a repeat, and since each term in the sequence is completely determined by the term before it, that means the next term and all subsequent terms must fall into that cycle.

How many iterations do we have to do before we see get a repeated value? If we think of the iteration as generating random numbers between 0 and $m - 1$, then we are looking at how many numbers in that range we can randomly generate before running into a repeat. This is just like the birthday problem from probability, where we want to know how many people we have to have in a room before there is a 50/50 chance that some two people in the room share a birthday. It turns out that we need just 23 people. We can think of the birthday problem as just like our problem with $m = 365$. An analysis of the birthday problem, which we omit here, shows that after roughly $\sqrt{n}$ terms, we should have a good chance of seeing a repeat.

So to find a factor of n, we can iterate $x^2 + 1$ and look at the gcd of various differences, $x_j - x_k$. If one of those is not relatively prime to n, then we have found a divisor of n. The problem is we don't know the length of the cycle we are trying to find

[1] The sequence might not right away start repeating. For instance, for $f(x) = (x^2 + 1) \bmod 11$ starting at 1, we get the sequence 1, 2, 5, 4, 6, 4, 6, 4, 6, ... Notice that the sequence ends up in a repeating cycle. This is where the ρ in the name comes from—the starting values 1, 2, 5 are the tail of the ρ and then the values end up in a cycle, which is the circular part of the ρ.

or where it starts. The Pollard rho method uses a technique called Floyd's cycle-finding algorithm to find a cycle. It searches for cycles by reading through the sequence at two different rates, one unit at a time and two units at a time. So we will be looking at the differences $x_1 - x_2$, $x_2 - x_4$, $x_3 - x_6$, $x_4 - x_8$, etc. This will eventually find a cycle, even if it isn't the shortest possible cycle.

Note It might happen that we have $n = ab$ and the cycle length for $f(x) = (x^2 + 1) \pmod{a}$ is a divisor of the cycle length for $f(x) = (x^2 + 1) \pmod{b}$. This would mean that the gcd we calculate would come out to n and we wouldn't find a nontrivial factor. In this case, we can switch to another function, $f(x) = x^2 + c$, for some other value of c, and try again. Also note that we don't have to start the iteration at $x_0 = 1$. It might be better to start with a random value.

If $n = ab$, with $a < b$, the time it takes to find a factor a should be on the order of $\sqrt{a}$. Here is the Pollard rho algorithm to find a nontrivial factor of n:

1. Pick a random c in the range from 1 to $n - 3$ and a random s in the range from 0 to $n - 1$.
2. Set u and v to s and define a function $f(x) = (x^2 + c) \bmod n$.
3. Set $g = 1$ and compute $u = f(u)$, $v = f(f(v))$, and $g = \gcd(u - v, n)$. Keep repeating the computation until $g \neq 1$. This value of g will be a factor of n. However, if $g = n$, then go back to step 1, as we want a nontrivial factor.

Here is some Python code implementing this algorithm:

```python
from random import randint
from fractions import gcd

def pollard_rho(n):
    g = n
    while g == n:
        c = randint(1,n-3)
        s = randint(0,n-1)
        u = v = s
        f = lambda x:(x*x+c) % n

        g = 1
        while g == 1:
            u = f(u)
            v = f(f(v))
            g = gcd(u-v,n)
    return g
```

On my laptop, this program was able to factor a 20-digit number into two 10-digit primes in a few seconds. It took about six minutes to factor a 30-digit number into two 15-digit primes. Compare this to the Miller-Rabin probabilistic primality test, where my laptop was able to determine (with high probability) that a 21,000-digit number was prime in about 16 minutes. In short, factoring seems to be a lot harder than primality testing.

Note that because of the random choices in the algorithm, if we run the algorithm on the same number multiple times, we might find different factors. Also, just Fermat's method, this method will just return a single factor, a of n. We can repeat the algorithm on n/a to find more factors. Don't try to run this algorithm on a prime, however, as it will end up in an infinite loop.

Appendix of useful algebra

Here are a few tricks that are occasionally useful in number theory:

1. The geometric series $1 + a + a^2 + \cdots + a^n$ can be rewritten as $\dfrac{a^{n+1} - 1}{a - 1}$.
2. We can factor $x^2 - y^2$ into $(x - y)(x + y)$ and more generally we can factor $x^n - y^n$ as

   $$x^n - y^n = (x - y)(x^{n-1} + x^{n-2}y + x^{n-3}y^2 + \cdots + xy^{n-2} + y^{n-1}).$$

   A particularly useful special case of this is $y = 1$.
3. The binomial theorem states that

   $$(x + y)^n = x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y^2 + \cdots + \binom{n}{n-1}xy^{n-1} + y^n.$$

   Recall that $\binom{n}{k}$ is called a *binomial coefficient* and is often read as "n choose k". We have

   $$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdot \cdots \cdot (n-k+1)}{k \cdot (k-1) \cdot \cdots \cdot 1}.$$

   For example,

   $$\binom{7}{3} = \frac{7!}{3!4!} = \frac{7 \cdot 6 \cdot 5}{3 \cdot 2 \cdot 1}.$$

   Note that the number of terms in the numerator and denominator of the reduced form is always the same. Binomial coefficients can also be read off of Pascal's triangle, where $\binom{n}{k}$ is the entry in row n, column k of the triangle (where we start counting rows and columns at index 0 instead of 1).

Bibliography

There are a number of books I used in preparing these notes. Here they are listed, roughly in order of how much I used them.

1. Burton. *Elementary Number Theory*, 5th edition. McGraw-Hill, 2002.
2. Pommersheim, Marks, and Flapan. *Number Theory: A Lively Introduction with Proofs, Applications, and Stories*. Wiley, 2010.
3. Crandall and Pomerance. *Prime Numbers: A Computational Perspective*, 2nd edition. Springer, 2010.
4. Tattersall. *Elementary Number Theory in Nine Chapters*, 2nd edition. Cambridge, 2005.
5. De Koninck and Mercier. *10001 Problems in Classical Number Theory*. American Mathematical Society, 2007.
6. Ore. *Number Theory and Its History.* McGraw-Hill, 1948.
7. Bressoud and Wagon. *A Course in Computational Number Theory*. Key College Publishing, 2000.
8. Niven and Zuckerman. *An Introduction to The Theory of Numbers*. Wiley, 1972.

In addition, I used Wikipedia quite a bit, as well as `http://primes.utm.edu`.