INTRODUCTION TO *GAUSS'S* NUMBER THEORY

Andrew Granville

We present a modern introduction to number theory. There are many introductory number theory books available, mostly developed more-or-less directly from Gauss's book *Disquisitiones Arithmeticae*. The core of our book is no different, indeed we have gone back to *Disquisitiones* for inspiration, but we do not hesitate to bring forward more modern ideas also.

Today's introductory number theory course occupies an anachronistic place in the typical curriculum. Although much of modern mathematics germinated from number theoretic seed, the curriculum places it at the end of an undergraduate program, to give examples of what students have learnt in other pure mathematics courses. In my experience students often fail to appreciate the connection between this and those more abstract courses — it would seem more natural to place number theory as a first year course to inspire the mathematics that is to come. In this book we highlight the connections between introductory number theory and other areas, but written without the assumption of that knowledge, so this book can be used as either a last year or first year text.

There does seem to be a more-or-less standard course, those things a student must know to have a basic grounding in number theory. We present this course as a dozen chapter series at the start of the book. There are fifty additional chapters in the second half of the book. Some of these are meant to further highlight the material in the first 12 chapters, and we will indicate there where they might be included. Others of these chapters could be an additional lecture, or might be used as a student for an independent reading project. Most of these can be read independently of the other chapters; quite a few require the reader to figure out things for themselves, and point the reader to further, deeper references.

We have chosen to give several proofs of various key results, not to confuse the reader but to highlight how well the subject hangs together.

The most unconventional choice in our "basic course" is to give Gauss's original proof of the law of quadratic reciprocity. Almost all textbooks give Eisenstein's proof based on a surprising lattice point counting argument (which we give in section C8); while this is elegant and highlights how different areas of mathematics support one another, it is largely unmotivated and too complicated for a student to fully grasp in an introductory course. Gauss's original proof is much more motivated by the introductory material, and has been reworked so as to be only slightly more complicated than Eisenstein's proof.

Typeset by $\mathcal{AMS}$-TEX

Table of Contents

Additional Sections

Section B. Basics

Section C. Algebra

Section D. Algebra and calculation

Section E. The distribution of Primes

Section F. Analytic number theory

Notation

$\mathbb{N}$ – The *natural numbers*, $1, 2, 3, \dots$
$\mathbb{Z}$ – The *integers*, $\dots, -3, -2, -1, 01, 2, 3, \dots$
Throughout all variables are taken to be integers, unless otherwise specified.
$\mathbb{Q}$ – The *rational numbers*, that is the fractions a/b with $a \in \mathbb{Z}$ and $b \in \mathbb{N}$.
$\mathbb{R}$ – The *real numbers*
$\mathbb{C}$ – The *complex numbers*
$A[x]$ — The set of polynomials with coefficients from the set A, that is $f(x) = \sum_{i=0}^{d} f_i x^i$ where each $f_i \in A$. Mostly we work with $A = \mathbb{Z}$.

$$\sum_{\substack{\text{Some variables:}\\ \text{Certain conditions hold}}} \text{summand} \quad \text{and} \quad \prod_{\substack{\text{Some variables:}\\ \text{Certain conditions hold}}} \text{summand},$$

mean that we sum, or product, the summand over the integer values of some variable, satisfying certain conditions.
$[t]$ — The *integer part of* t. That is, the largest integer $\leq t$.
$\{t\}$ — The *fractional part of* (real number) t. That is $\{t\} = t - [t]$. Notice that $0 \leq \{t\} < 1$.
(a, b) — The greatest common divisor of a and b.
$[a, b]$ — The least common multiple of a and b.
$b|a$ — means b divides a
$p^k \| a$ — means p^k divides a, but not p^{k+1}
$I(a, b)$ — The ideal $\{am + bn : \ m, n \in \mathbb{Z}\}$

1. The Euclidean Algorithm

1.1. Finding the gcd. You probably know the Euclidean algorithm, used to find the greatest common divisor of two given integers. For example, to determine the greatest common divisor of 85 and 48, we begin by subtracting the smaller from the larger, 48 from 85, to obtain $85-48=37$. Now $\gcd(85,48)=\gcd(48,37)$ and we apply the algorithm again to the pair 48 and 37. So we subtract the smaller from the larger to obtain $48-37=11$, so that $\gcd(48,37)=\gcd(37,11)$. Next we should subtract 11 from 37, but then we would only do so again, and a third time, so let's do all that in one go and take $37-3\times 11=4$, to obtain $\gcd(37,11)=\gcd(11,4)$. Similarly we take $11-2\times 4=3$, and then $4-3=1$, so that the gcd of 85 and 48 is 1. This is the Euclidean algorithm that you learnt in school, but did you ever prove that it really works?

To do so, we must first carefully define what we have implicitly used in the above paragraph:

We say that a *is divisible by* b (or a *is a multiple of* b), or b *is a divisor of* a (or b *is a factor of* a), if there exists an integer q such that $a=qb$. For convenience we write "$b|a$".

`Exercise 1.1.1a.` Prove that if b divides a then either $a=0$ or $|a|\geq|b|$.

`Exercise 1.1.1b.` Deduce that if $a|b$ and $b|a$ then $b=\pm a$.

`Exercise 1.1.1c.` Prove that if a divides b and c then a divides $bx+cy$ for all integers x,y.

`Exercise 1.1.1d.` Prove that if a divides b, and b divides c, then a divides c.

In general we have

Lemma 1.1. *If a and $b>0$ are integers then there exist integers q and r, with $0\leq r\leq b-1$, such that $a=qb+r$. We call q the "*quotient*", and r the "*remainder*".*

Proof. Let r be the smallest element of the set $S:=\{a+nb\geq 0 : n\in\mathbb{Z}\}$. Evidently the set is non-empty (as may be seen by selecting n sufficiently large) so that r exists. Now $r\geq 0$ by definition, and if $r=a-qb$ then we have $r<b$ else $a-bq\geq b$ so that $r-b=a-(q+1)b\in S$, contradicting the minimality of r.

`Exercise 1.1.2.` (i) Let $[t]$ be the *integer part* of t, that is the largest integer $\leq t$. Prove that $q=[a/b]$.
(ii) Let $\{t\}$ to be the *fractional part* of t, that is $\{t\}=t-[t]$. Prove that $r=b\{r/b\}=b\{a/b\}$.

We say that d is a *common divisor* of a and b if d divides both a and b. We are interested here in the *greatest common divisor* of a and b, which is often written $\gcd(a,b)$ or simply (a,b).[1]

`Exercise 1.1.3.` Show that if a and b are not both 0, then $\gcd(a,b)$ is a positive integer.

We say that a is *coprime* with b, or a and b are *coprime integers* or *relatively prime* if $(a,b)=1$.

---

[1] In the UK this is known as the *highest common factor* of a and b, and written $\mathrm{hcf}(a,b)$.

Corollary 1.2. *If $a = qb + r$ as in Lemma 1.1, then* $\gcd(a, b) = \gcd(b, r)$.

Proof. Let $g = \gcd(a, b)$ and $h = \gcd(r, b)$. Now g divides a and b, so g divides $a - qb = r$. Therefore g is a common divisor of both r and b, and therefore $g \leq h$. Similarly h divides b and r, so h divides $qb + r = a$ and hence h is a common divisor of both a and b, and therefore $h \leq g$. We have shown that $g \leq h$ and $h \leq g$, which together imply that $g = h$.

`Exercise 1.1.4.` Use Corollary 1.2 to deduce that the Euclidean algorithm indeed yields the greatest common divisor of the two given integers.

1.2. Linear combinations. Another aspect of the Euclidean algorithm is that one can find a linear combination of a and b, over the integers, which equals $\gcd(a, b)$; that is, one can find integers u and v such that

$$au + bv = \gcd(a, b).$$

We proceed as follows in our example above: $1 = 1 \cdot 4 - 1 \cdot 3$, and so we have

$$1 = 1 \cdot 4 - 1 \cdot 3 = 1 \cdot 4 - 1 \cdot (11 - 2 \cdot 4) = 3 \cdot 4 - 1 \cdot 11,$$

as we had $3 = 11 - 2 \cdot 4$. This then implies, since we had $4 = 37 - 3 \cdot 11$, that

$$1 = 3 \cdot (37 - 3 \cdot 11) - 1 \cdot 11 = 3 \cdot 37 - 10 \cdot 11.$$

Continuing in this way, we deduce:

$$1 = 3 \cdot 37 - 10 \cdot (48 - 37) = 13 \cdot 37 - 10 \cdot 48 = 13 \cdot (85 - 48) - 10 \cdot 48 = 13 \cdot 85 - 23 \cdot 48,$$

that is, we have the desired linear combination of 85 and 48.

To prove that this method always works, we use Lemma 1.1 again: Suppose that $a = qb + r$ so that $\gcd(a, b) = \gcd(b, r)$ by Corollary 1.2, and we have $bu - rv = 1$ for some integers u and v. Then

$$\gcd(a, b) = \gcd(b, r) = bu - rv = bu - (a - qb)v = b(u + qv) - av,$$

the desired linear combination of a and b. This yields a proof of the following:

Theorem 1.3. *If a and b are given integers then there exist integers u and v such that*

$$au + bv = \gcd(a, b).$$

`Exercise 1.2.1.` Prove that if there exist integers u and v such that $au + bv = 1$ then $\gcd(a, b) = 1$.

`Exercise 1.2.2.` Prove that if d divides both a and b then d divides $\gcd(a, b)$.

`Exercise 1.2.3.` Prove that if a divides m, and b divides n then $\gcd(a, b)$ divides $\gcd(m, n)$. In particular show that if a divides m, and b divides n where $\gcd(m, n) = 1$ then $\gcd(a, b) = 1$.

Corollary 1.4. *If* $\gcd(a, m) = \gcd(b, m) = 1$ *then* $\gcd(ab, m) = 1$

Proof. By Theorem 1.3 there exist integers r, s, u, v such that $au + mv = br + ms = 1$. Therefore $ab(ur) + m(bvr + aus + msv) = (au + mv)(br + ms) = 1$. Hence $\gcd(ab, m)$ divides 1 by exercise 1.2.2, and the result follows from exercise 1.1.3.

Corollary 1.5. *We have* $\gcd(ma, mb) = m \cdot \gcd(a, b)$ *for all integers* $m \geq 1$*.*

Proof. By Theorem 1.3 there exist integers r, s, u, v such that $au + bv = \gcd(a, b)$ and $(ma)r + (mb)s = \gcd(ma, mb)$. Now $\gcd(ma, mb)$ divides ma and mb so it divides $mau + mbv = m \cdot \gcd(a, b)$. Similarly $\gcd(a, b)$ divides a and b, so that $m \cdot \gcd(a, b)$ divides ma and mb, and therefore $\gcd(ma, mb)$ by exercise 1.2.2. The result follows for exercise 1.1.1b.

`Exercise 1.2.4.` Deduce that if A and B are given integers with $g = \gcd(A, B)$ then $\gcd(A/g, B/g) = 1$. (Hint: Try $m = g,\ A = ma,\ B = mb$ in Corollary 1.4.)

`Exercise 1.2.5.` Show that any rational number u/v where $u, v \in \mathbb{Z}$ with $v \neq 0$, may be written as r/s where r and s are coprime integers with $s > 0$.

We define the set of linear combinations of two integers as follows:

$$I(a, b) := \{am + bn : \ m, n \in \mathbb{Z}\}.$$

This definition can be extended to an arbitrary set of integers in place of $\{a, b\}$; that is

$$I(a_1, \ldots a_k) := \{a_1 m_1 + a_2 m_2 + \ldots + a_k m_k : \ m_1, m_2, \ldots, m_k \in \mathbb{Z}\}.$$

Corollary 1.6. *If a and b are given non-zero integers then we have $I(a, b) = I(g)$ where $g :=\gcd(a, b)$; that is*

$$\{am + bn : \ m, n \in \mathbb{Z}\} = \{gk : \ k \in \mathbb{Z}\}.$$

Proof. By Theorem 1.3 we know that there exist $u, v \in \mathbb{Z}$ such that $au + bv = g$. Therefore $a(uk) + b(vk) = gk$ so that $gk \in I(a, b)$ for all $k \in \mathbb{Z}$; that is $I(g) \subset I(a, b)$. On the other hand, as g divides both a and b, there exist integers A, B such that $a = gA, b = gB$, and so any $am + bn = g(Am + Bn) \in I(g)$. That is $I(a, b) \subset I(g)$. The result now follows from the two inclusions.

`Exercise 1.2.6.` Show that $I(a_1, \ldots a_k) = I(g)$ for any non-zero integers $a_1, \ldots a_k$, where $g = \gcd(a_1, \ldots a_k)$.

`Exercise 1.2.7.` Deduce that if we are given integers $a_1, a_2, \ldots, a_k$, not all zero, then there exist integers $m_1, m_2, \ldots, m_k$ such that

$$m_1 a_1 + m_2 a_2 + \ldots + m_k a_k = \gcd(a_1, a_2, \ldots, a_k).$$

We say that the integers $a_1, a_2, \ldots, a_k$ are *relatively prime* if $\gcd(a_1, a_2, \ldots, a_k) = 1$. We say that they are *pairwise coprime* if $\gcd(a_i, a_j) = 1$ whenever $i \neq j$.

We have seen how the Euclidean algorithm can be used to find the gcd of two given integers a and b, and to find integers u, v such that $au + bv = \gcd(a, b)$. This is more than the mere existence of u and v, which is all that was claimed in Theorem 1.3; the price for obtaining the values of u and v is a somewhat complicated analysis of the Euclidean algorithm. However if we only want to prove that such u and v exist, we can give an easier proof:

Non-constructive proof of Theorem 1.3. Let h be the smallest positive integer that belongs to $I(a,b)$, say $h = au + bv$. Then $g := \gcd(a,b)$ divides h, as g divides both a and b.

Lemma 1.1 implies that there exist integers q and r, with $0 \le r \le h-1$ such that $a = qh + r$. Therefore

$$r = a - qh = a - q(au + bv) = a(1 - qu) + b(-qv) \in I(a,b),$$

which contradicts the minimality of h, unless $r = 0$; that is h divides a. An analogous argument reveals that h divides b, and so h divides g by exercise 1.2.2.

Hence g divides h, and h divides g, so that $g = h$ as desired.

In section C1 we discuss how the sets $I(a,b)$ generalize to other number domains, and discuss some of the basic theory attached to that. This is recommended to be inserted here particularly for classes in which many of the students have had a course in algebra.

Up until now we have considered the Euclidean algorithm, one step at a time. It is convenient to give appropriate notation for the steps of the Euclidean algorithm, so that we can consider all the steps together:

1.3. Continued Fractions. If $a > b > 1$ with $(a,b) = 1$ then Lemmas 1.1 and 1.2 yield that there exists integers q and r, with $b > r \ge 1$ such that

$$\frac{a}{b} = q + \frac{r}{b} = q + \frac{1}{\frac{b}{r}} .$$

And then we can repeat this with the pair of integers b and r. Thus going back to our original example, where we were finding the gcd of 85 and 48, we begin by noting that

$$\frac{85}{48} = 1 + \frac{37}{48},$$

and then

$$\frac{48}{37} = 1 + \frac{11}{37}, \text{ so that } \frac{85}{48} = 1 + \frac{1}{\frac{48}{37}} = 1 + \frac{1}{1 + \frac{11}{37}} .$$

We continue like this:

$$\frac{37}{11} = 3 + \frac{4}{11}, \ \frac{11}{4} = 2 + \frac{3}{4}, \text{ and } \frac{4}{3} = 1 + \frac{1}{3},$$

so that

$$\frac{85}{48} = 1 + \frac{1}{1 + \frac{11}{37}} = 1 + \frac{1}{1 + \frac{1}{3 + \frac{4}{11}}} = 1 + \frac{1}{1 + \frac{1}{3 + \frac{1}{2 + \frac{3}{4}}}} = 1 + \frac{1}{1 + \frac{1}{3 + \frac{1}{2 + \frac{1}{1 + \frac{1}{3}}}}} .$$

This is the *continued fraction* for $\frac{85}{48}$ and is more conveniently written as $[1, 1, 3, 2, 1, 3]$. Notice that this is the sequence of quotients a_i from the various divisions, that is

$$\frac{a}{b} = [a_0,\ a_1,\ a_2,\ \ldots,\ a_k] := a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \cdots + \frac{1}{a_k}}}} .$$

`Exercise 1.4.1.` Show that if $a_k > 1$ then $[a_0,\ a_1, \ldots,\ a_k] = [a_0,\ a_1, \ldots,\ a_k - 1, 1]$. Prove that the set of positive rational numbers are in $1-1$ correspondence with the finite length continued fractions that do not end in 1.

Taking the rationals corresponding to the first part of the continued fraction, namely $[1] = 1, [1, 1] = 2, [1, 1, 3], \ldots$ gives

$$1 + \cfrac{1}{1 + \frac{1}{3}} = \frac{7}{4}, \qquad 1 + \cfrac{1}{1 + \frac{1}{3 + \frac{1}{2}}} = \frac{16}{9}, \qquad 1 + \cfrac{1}{1 + \frac{1}{3 + \frac{1}{2 + \frac{1}{1}}}} = \frac{23}{13},$$

which are increasingly good approximations ($1.75,\ 1.777\ldots,\ 1.7692\ldots$) to $85/48 = 1.770833\ldots$. We call these the *convergents* $r_j/s_j,\ j \geq 1$ for a continued fraction, defined by $\frac{r_j}{s_j} = [a_0,\ a_1,\ a_2,\ \ldots,\ a_j]$, so that $a/b = r_k/s_k$. We will show in section C2 that $r_j s_{j-1} - r_{j-1} s_j = (-1)^{j-1}$, so if $u = (-1)^{k-1} s_{k-1}$ and $v = (-1)^k r_{k-1}$ then

$$au + bv = 1.$$

This is really just a convenient reworking of the Euclidean algorithm, as we explained it above, for finding u and v. Bachet de Meziriac, the celebrated editor and commentator of Diophantus, introduced this method to Renaissance mathematicians in the second edition of his brilliantly named book *Pleasant and delectable problems which are made from numbers* (1624). Such methods had been known from ancient times, certainly to 8th century Indian scholars, probably to Archimedes, and possibly to the Babylonians.

2. Congruences.

2.1. Basic Congruences. If a divides $b-c$ then we write $b \equiv c \pmod a$, and say that b and c are *congruent modulo* a, where a is the *modulus*. The numbers involved should be integers, not fractions, and the modulus can be taken in absolute value; that is $b \equiv c \pmod a$ if and only if $b \equiv c \pmod{|a|}$, by definition.

For example, $-10 \equiv 15 \pmod 5$, and $-7 \equiv 15 \pmod{11}$, but $-7 \not\equiv 15 \pmod 3$. Note that $b \equiv b \pmod a$ for all integers a and b.[2]

The integers $\equiv a \pmod m$ are precisely those of the form $a+km$ where k is an integer. We call this set of integers a *congruence class* or *residue class* mod m, and any particular element of the congruence class is a *residue*.

Theorem 2.1. *Suppose that m is a positive integer. Exactly one of any m consecutive integers is $\equiv a \pmod m$.*

Proof. Suppose that we are given the m consecutive integers $x, x+1, \dots, x+m-1$. One of these is of the form $a+km$, where k is an integer, if and only if there exists an integer k for which

$$x \leq a+km < x+m.$$

Subtracting a from each term here and dividing through by m, we find that this holds if and only if

$$\frac{x-a}{m} \leq k < \frac{x-a}{m} + 1.$$

Hence k must be an integer from an interval of length one with just one endpoint included the interval. One easily sees that such an integer k is unique, indeed it is the smallest integer that is $\geq \frac{x-a}{m}$.

A number theoretic proof that there is at most one such integer goes as follows: If $x+i \equiv a \pmod m$ and $x+j \equiv a \pmod m$, where $0 \leq i < j \leq m-1$ then $i \equiv a-x \equiv j \pmod m$, so that m divides $j-i$ which is impossible as $1 \leq j-i \leq m-1$.

Theorem 2.1 implies that any m consecutive integers yields a *complete set of residues* $\pmod m$; that is every congruence class $\pmod m$ is represented by one element of the given set of m integers. For example, every integer has a unique residue amongst the *least non-negative residues* $\pmod m$,

$$0,\ 1,\ 2,\ \dots (m-1)$$

(which is also a direct consequence of Theorem 1.1), amongst the *least positive residues* $\pmod m$, $1, 2, \dots, m$, and also amongst $-(m-1),\ -(m-2),\ \dots,\ -2,\ -1,\ 0$. If the residue is not 0 then these residues occur in pairs, one positive the other negative, and one of each pair is $\leq m/2$ in absolute value, which we call the *absolutely least* residue $\pmod m$ (and when m is even we select $m/2$ rather than $-m/2$). For example 2 is the absolutely least residue of $-13 \pmod 5$, whereas -3 is the least negative residue. 5 is

[2]We adopt the symbol $\equiv$ because of the analogies between equality and congruence; to avoid ambiguity we have made a minor distinction between the two notations, by adding the extra bar.

its own least positive residue mod 7, and -2 is the least negative residue as well as the absolutely least.

We defined a *complete set of residues* to be any set of representatives for the residue classes mod m, one for each residue class. A *reduced set of residues* has representatives only for the residue classes that are coprime with m. For example $\{0, 1, 2, 3, 4, 5\}$ is a complete set of residues $\pmod 6$, whereas $\{1, 5\}$ is a reduced set of residues.

`Exercise 2.1.1.` Prove that the set of integers in the congruence class $a \pmod d$ can be partitioned into the set of integers in the congruence classes $a \pmod{kd}$, $a + d \pmod{kd}, \dots, a + (k-2)d \pmod{kd}$ and $a + (k-1)d \pmod{kd}$.

`Exercise 2.1.2.` Show that if $a \equiv b \pmod m$ then $(a, m) = (b, m)$.

`Exercise 2.1.3.` Prove that the property of congruence modulo m is an *equivalence relation* on the integers. To prove this one must establish (i) $a \equiv a \pmod m$; (ii) $a \equiv b \pmod m$ implies $b \equiv a \pmod m$; and (iii) $a \equiv b \pmod m$ and $b \equiv c \pmod m$ imply $a \equiv c \pmod m$.

One consequence of this is that congruent numbers have the same least residues, whereas non-congruent numbers have different least residues.

The main use of congruences is that it simplifies arithmetic when we are looking into questions about remainders. This is because the usual rules for addition, subtraction and multiplication work for congruences; division is a little more complicated, as we shall see.

Lemma 2.2. *If* $a \equiv b \pmod m$ *and* $c \equiv d \pmod m$ *then*

$$\begin{aligned} a + c &\equiv b + d \pmod m \\ a - c &\equiv b - d \pmod m \\ \text{and} \quad ac &\equiv bd \pmod m. \end{aligned}$$

Proof. By hypothesis there exist integers u and v such that $a - b = um$ and $c - d = vm$. Therefore

$$(a + c) - (b + d) = (a - b) + (c - d) = um + vm = (u + v)m$$

so that $a + c \equiv b + d \pmod m$,

$$(a - c) - (b - d) = (a - b) - (c - d) = um - vm = (u - v)m$$

so that $a - c \equiv b - d \pmod m$, and

$$ac - bd = a(c - d) + d(a - b) = a \cdot vm + b \cdot um = (av + bu)m$$

so that $ac \equiv bd \pmod m$.

`Exercise 2.1.4.` Show that for any integers k and l we have $ka + lc \equiv kb + ld \pmod m$.

To see that division does not work so easily, we try to divide each side of $8 \equiv 2 \pmod 6$ by 2, which yields the incorrect "$4 \equiv 1 \pmod 6$". To make this correct we need to divide the modulus through by 2 also so as to obtain $4 \equiv 1 \pmod 3$. However even this is not the whole story, for if we wish to divide both sides of $21 \equiv 6 \pmod 5$ through by 3, we cannot also divide the modulus, since 3 does not divide 5. However, in this case one does not need to divide the modulus through by 3, indeed $7 \equiv 2 \pmod 5$. So what is the general rule? We shall return to this question in Lemma 3.5. For now the only observation we make is the following easy exercise:

`Exercise 2.1.5.` Prove that if $a \equiv b \pmod m$ then $a \equiv b \pmod d$ for any divisor d of m.

Let $\mathbb{Z}[x]$ denote the set of polynomial with integer coefficients.

Corollary 2.3. *If $f(x) \in \mathbb{Z}[x]$ and $a \equiv b \pmod m$ then $f(a) \equiv f(b) \pmod m$.*

Proof. Since $a \equiv b \pmod m$ we have $a^2 \equiv b^2 \pmod m$ by Lemma 2.2, and then $a^k \equiv b^k \pmod m$ for all integers k, by induction. Now, writing $f(x) = \sum_{i=0}^{d} f_i x^i$ where each f_i is an integer, we have

$$f(a) = \sum_{i=0}^{d} f_i a^i \equiv \sum_{i=0}^{d} f_i b^i = f(b) \pmod m,$$

by exercise 2.1.4.

This result can be extended to polynomials in many variables.

`Exercise 2.1.6.` Prove that if $f(t) \in \mathbb{Z}[t]$ and $r, s \in \mathbb{Z}$ then $r - s$ divides $f(r) - f(s)$.

Therefore, for any given polynomial $f(x) \in \mathbb{Z}[x]$, the sequence $f(0), f(1), f(2), f(3), \dots$ modulo m is *periodic* of period m, that is the values repeat every mth time, repeated indefinitely. More precisely $f(n + m) \equiv f(n) \pmod m$ for all integers n.

Example: If $f(x) = x^3 - 8x + 6$ and $m = 5$ then we get the sequence

$$f(0), f(1), \dots = 1, 4, 3, 4, 3, 1, 4, 3, 4, 3, 1 \dots$$

and the first five terms $1, 4, 3, 4, 3$ repeat infinitely often. Moreover we get the same pattern if we run though the consecutive negative integer values for x.

Note that in this example $f(x)$ is never 0 or 2 (mod 5). Thus neither of the two equations

$$x^3 - 8x + 6 = 0 \qquad \text{and} \qquad x^3 - 8x + 4 = 0$$

can have solutions in integers.

`Exercise 2.1.7.` Let $f(x) \in \mathbb{Z}[x]$. Suppose that $f(r) \not\equiv 0 \pmod m$ for all integers r in the range $0 \le r \le m - 1$. Deduce that there does not exist an integer n for which $f(n) = 0$.

2.2. Tests for divisibility. There are easy tests for divisibility based on ideas from this section. For instance since

$$a + 10b + 100c + \dots \equiv a + b + c + \dots \pmod 9$$

we can test the first number for divisibility by 9, by testing the latter. Similarly we can use this same test for divisibility by 3, since 3 divides 9. For example, is 7361842509 divisible by 9? This holds if and only if $7+3+6+1+8+4+2+5+0+9=45$ is divisible by 9, which holds if and only if $4+5=9$ is divisible by 9, which it is.

For the modulus 11 we have that $10^2 = 100 \equiv 1 \pmod{11}$ and, in general, that

$$10^{2k} = (10^2)^k \equiv 1^k \equiv 1 \pmod{11} \quad \text{and} \quad 10^{2k+1} = 10^{2k} \cdot 10 \equiv 1 \cdot (-1) \equiv -1 \pmod{11}.$$

Therefore

$$a + 10b + 100c + \ldots \equiv a - b + c \ldots \pmod{11}.$$

Therefore 7361842509 is divisible by 11 if and only if $7-3+6-1+8-4+2-5+0-9=1$ divisible by 11, which it is not.

One may deduce similar rules to test for divisibility by *any* integer, though we will need to develop our theory of congruences. We return to this theme in section 7.7.

`Exercise 2.2.1.` Invent tests for divisibility by 2 and 5 (easy), and also by 7 and 13 (similar to the above). Try and make one test that tests for divisibility by 7, 11 and 13 simultaneously.

3. The basic algebra of number theory

A *prime number* is an integer $n > 1$ whose only positive divisors are 1 and n. Hence $2, 3, 5, 7, 11, \dots$ are primes. Integer $n > 1$ is *composite* if it is not prime.

3.1. The Fundamental Theorem of Arithmetic. All the way back to ancient Greek times, mathematicians recognized that abstract lemmas allowed them to *prove* sophisticated theorems. The archetypal result is "Euclid's Lemma", an important result that first appeared in Euclid's "*Elements*" (Book VII, No. 32).

Euclid's Lemma. *If c divides ab and $\gcd(c, a) = 1$ then c divides b.*

This has the following important consequence, taking $c = p$ prime:

Theorem 3.1. *If prime p divides ab then p must divide at least one of a and b.*

The hypothesis in Theorem 3.1 that p is prime, and the hypothesis in Euclid's Lemma are certainly necessary, as may be understood from the example where 4 divides $2 \cdot 6$, but 4 does not divide either 2 or 6.

We begin by giving Gauss's proof of Theorem 3.1, which is (arguably) more intuitive than the usual proof of Euclid's lemma:

Gauss's proof of Theorem 3.1. Suppose that this is false so there exist positive integers a and b that are not divisible by p, and yet ab is divisible by p (if a or b is negative, replace them by $-a$ or $-b$, respectively). Pick the counterexample with b as small as possible, and note that $0 < b < p$ else if n is the least residue of b mod p, then $n \equiv b \not\equiv 0 \pmod p$ and $an \equiv ab \equiv 0 \pmod p$, contradicting the minimality of b.

We also have $b > 1$ else p divides $a \cdot 1 = a$.

Let B be the least positive residue of $p \pmod b$, so that $1 \leq B < b < p$, and therefore $p \nmid B$. Writing $B = p - kb$ for some integer k we have

$$aB = a(p - kb) = pa - (ab)k \equiv 0 \pmod p,$$

since ab is divisible by p. However p does not divide either a or b, and so this contradicts the minimality of b.

The slick, but unintuitive proof of Euclid's lemma. Since $\gcd(c, a) = 1$ there exist integers m and n such that $cm + an = 1$ by Theorem 1.3. Hence c divides

$$c \cdot bm + ab \cdot n = b(cm + an) = b.$$

Corollary 3.2. *If $am = bn$ then $a/\gcd(a, b)$ divides n.*

Proof. Let $a/\gcd(a, b) = A$ and $b/\gcd(a, b) = B$ so that $(A, B) = 1$ by exercise 1.2.4, and $Am = Bn$. Therefore $A|Bn$ with $(A, B) = 1$ so that $A|n$ by Euclid's Lemma.

`Exercise 3.1.1.` Prove that if prime p divides $a_1 a_2 \dots a_k$ then p divides a_j for some j, $1 \leq j \leq k$.

With this preparation we are ready to prove the first great theorem of number theory, which appears in Euclid's "*Elements*":

The Fundamental Theorem of Arithmetic. *Every integer $n > 1$ can be written as a product of primes in a unique way (up to re-ordering).*

By "re-ordering" we mean that although one can write 12 as $2 \times 2 \times 3$, or $2 \times 3 \times 2$, or $3 \times 2 \times 2$, we count all of these as the same product, since they involve the same primes, each the same number of times.

Proof. We first show that there is a factorization of n into primes. We prove this by induction on n: If n is prime then we are done; since $n = 2, 3$ are prime this also starts our induction hypothesis. If n is composite then it must have a divisor a for which $1 < a < n$, and so $b = n/a$ is an integer for which $1 < b < n$. Then, by the induction hypothesis, both a and b can be factored into primes, and so $n = ab$ equals the product of these two factorizations.

Now we prove that there is just one factorization for each $n \geq 2$. If this is not true then let n be the smallest integer ≥ 2 that has two distinct factorizations,

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

where the p_i and q_j are primes. Now prime p_r divides $q_1 q_2 \cdots q_s$, and so p_r divides q_j for some j, by exercise 3.1.1. Re-ordering the q_j if necessary we may assume that $j = s$, and we have that $p_r = q_s$ since q_s is a prime and hence its only prime divisor is itself. If we divide through both factorizations by $p_r = q_s$ we have two distinct factorizations of

$$n/p_r \ = \ p_1 p_2 \cdots p_{r-1} \ = \ q_1 q_2 \cdots q_{s-1},$$

which contradicts the minimality of n unless $n/p_r = 1$. But then $n = p_r$ is prime, and by the definition (of primes) it can have no other factor.

It is useful to write the factorizations of natural numbers in a standard form, like

$$n = 2^{n_2} 3^{n_3} 5^{n_5} 7^{n_7} \dots ,$$

with each $n_i \geq 0$, and where only finitely many of the n_i are non-zero. Usually we only write down those prime powers where $n_i \geq 1$, for example $12 = 2^2 \cdot 3$ and $50 = 2 \cdot 5^2$.

`Exercise 3.1.2.` Prove that every natural number has a unique representation as $2^k m$ with $k \geq 0$ and m an odd natural number.

`Exercise 3.1.3.` Show that if all of the prime factors of an integer n are $\equiv 1 \pmod m$ then $n \equiv 1 \pmod m$. Deduce that if $n \not\equiv 1 \pmod m$ then n has a prime factor that is $\not\equiv 1 \pmod m$.

`Exercise 3.1.4.` Show that if all of the prime factors of an integer n are $\equiv 1$ or $3 \pmod 8$ then $n \equiv 1$ or $3 \pmod 8$. Prove this with 3 replaced by 7. Generalize this as much as you can.

We write $p^e \| n$ if p^e is the highest power of p that divides n; thus $3^2 \| 18$ and $11^1 \| 1001$. Suppose that $n = \prod_{p \text{ prime}} p^{n_p}, \ a = \prod_p p^{a_p}, \ b = \prod_p p^{b_p}$.[3] If $n = ab$ then

$$2^{n_2} 3^{n_3} 5^{n_5} \cdots \ = \ 2^{a_2} 3^{a_3} 5^{a_5} \cdots 2^{b_2} 3^{b_3} 5^{b_5} \cdots \ = \ 2^{a_2+b_2} 3^{a_3+b_3} 5^{a_5+b_5} \cdots ,$$

[3]Here, and often hereafter, we suppress writing "prime" in the subscript of $\prod$, for convenience.

so by the fundamental theorem of arithmetic we have $n_p = a_p + b_p$ for each prime p. As $a_p, b_p \geq 0$ for each prime p, we can deduce that $0 \leq a_p \leq n_p$. In the other direction if $a = 2^{a_2}3^{a_3}5^{a_5}\ldots$ with each $0 \leq a_p \leq n_p$ then a divides n since we can write $n = ab$ where $b = 2^{n_2-a_2}3^{n_3-a_3}5^{n_5-a_5}\ldots$

Therefore the number of divisors a of n is equal to the number of possibilities for the exponents a_p, that is each a_p is an integer in the range $0 \leq a_p \leq n_p$. There are therefore $(n_p + 1)$ possibilities for the exponent for each prime p making

$$(n_2 + 1)(n_3 + 1)(n_5 + 1)\ldots$$

in total. Hence if we write $\tau(n)$ for the number of divisors of n, then

$$\tau(n) = \prod_{\substack{p \text{ prime} \\ p^{n_p} \| n}} \tau(p^{n_p}).$$

Functions like this, in which we can break up the value of the function at n, via the factorization of n, into the value of the function at the maximum prime powers that divide n, are called *multiplicative functions*.

`Exercise 3.1.5.` Reprove Corollary 1.4 using the Fundamental Theorem of Arithmetic.

`Exercise 3.1.6.` Prove that if $(a, b) = 1$ then $(ab, m) = (a, m)(b, m)$.

`Exercise 3.1.7.` Use the description of the divisors of a given integer to prove the following: Suppose that we are given positive integers $m = \prod_p p^{m_p}$ and $n = \prod_p p^{n_p}$. Then

(i) $\gcd(m, n) = \prod_p p^{\min\{m_p,n_p\}}$ and (ii) $\text{lcm}[m, n] = \prod_p p^{\max\{m_p,n_p\}}$.

Here $\text{lcm}[m, n]$ denotes the *least common multiple* of m and n, that is the smallest positive integer which is divisible by both m and n.

`Exercise 3.1.8.` Prove that d divides $\gcd(a, b)$ if and only if d divides both a and b.
Prove that $\text{lcm}(a, b)$ divides m if and only if a and b both divide m.

`Exercise 3.1.9.` Deduce that $mn = \gcd(m, n) \cdot \text{lcm}(m, n)$ for all pairs of natural numbers m and n.

`Exercise 3.1.10.` Prove that $\text{lcm}[ma, mb] = m \cdot \text{lcm}[a, b]$ for any positive integer m.

`Exercise 3.1.11.` Prove that $\gcd(a, b, c) \cdot \text{lcm}(a, b, c) = abc$ if and only if a, b and c are pairwise coprime.

`Exercise 3.1.12.` Prove that for any integers a, b, m, n there exists an integer c such that $\frac{a}{m} + \frac{b}{n} = \frac{c}{L}$ where $L = \text{lcm}[m, n]$. Show that $\text{lcm}[m, n]$ is the smallest positive integer with this property. For this reason we often call $\text{lcm}[m, n]$ the *lowest common denominator* of the fractions $1/m$ and $1/n$.

One can obtain the gcd and lcm for any number of integers by similar means:

Example: If $A = 504 = 2^3 \cdot 3^2 \cdot 7$, $B = 2880 = 2^6 \cdot 3^2 \cdot 5$ and $C = 864 = 2^5 \cdot 3^3$, then the greatest common divisor is $2^3 \cdot 3^2 = 72$ and the least common multiple is $2^6 \cdot 3^3 \cdot 5 \cdot 7 = 60480$.

This method for finding the gcd of two integers appears to be much simpler than that discussed previously, using the Euclidean algorithm. However in order to make this method work one needs to be able to factor the integers involved, which puts severe limitations on the size of numbers for which the method will easily work. On the other hand, the Euclidean algorithm is very efficient for finding the gcd of two given integers without needing to know anything else about those numbers.

`Exercise 3.1.13.` Reprove exercise 1.2.4, that if $(a, b) = g$ then $(a/g, b/g) = 1$, using the factorizations of a, b and g.

`Exercise 3.1.14.` Prove that if each of $a, b, c, \dots$ is coprime with m then so is $abc \dots$

`Exercise 3.1.15.` Prove that if $a, b, c, \dots$ are pairwise coprime and they each divide m, then $abc \dots$ divides m.

`Exercise 3.1.16.` Deduce that if $m \equiv n \pmod a$ and $m \equiv n \pmod b$ and $m \equiv n \pmod c$, $\dots$, where $a, b, c, \dots$ are coprime with one another, then $m \equiv n \pmod{abc} \dots$

`Exercise 3.1.17.` Prove that each of $a, b, c, \dots$ divides m if and only if $\text{lcm}[a, b, c, \dots]$ divides m. What is the analogous strengthening of the result in exercise 3.1.16?

Gauss's proof of Euclid's Lemma. Since ab is divisible by both a and c, and since $(a, c) = 1$, therefore ab is divisible by ac by exercise 3.1.15. Therefore $ab/ac = b/c$ is an integer, and so c divides b.

Using the representation of an integer in terms of its prime power factors can be useful when considering powers:

`Exercise 3.1.18.` Prove that n divides the exponent of all of the prime power factors of A, if and only if A is the nth power of an integer.

`Exercise 3.1.19.` Prove that if $a, b, c, \dots$ are pairwise coprime, positive integers and their product is an nth power then they are each an nth power.

`Exercise 3.1.20.` Prove that if ab is a square then $a = \pm gA^2$ and $b = \pm gB^2$ where $g = \gcd(a, b)$. (Hint: Use exercise 3.1.13.)

`Exercise 3.1.21.` Let p be an odd prime. Suppose that x, y and z are integers for which $x^p + y^p = z^p$. Show that there exist an integer r such that $z - y = r^p$, pr^p or $p^{p-1}r^p$. (Hint: Factor $z^p - y^p = (z - y)(z^{p-1} + z^{p-2}y + \dots + zy^{p-2} + y^{p-1})$ and find the possible gcds of the two factors.) Rule out the possibility that $z - y = pr^p$. (This last part is not easy – you may wish to use Lemma 7.12.)

3.2. Irrationality. Are there irrational numbers? How about $\sqrt{2}$?

Proposition 3.3. *There does not exist a rational number a/b for which $\sqrt{2} = a/b$. That is, $\sqrt{2}$ is irrational.*

Proof. We may assume, as in any fraction, that $(a, b) = 1$ so that a and b are minimal, and that $b \geq 1$ (and so $a \geq 1$). Now if $\sqrt{2} = a/b$ then $a = b\sqrt{2}$ and so $a^2 = 2b^2$.

Write the factorizations

$$a = \prod_p p^{a_p},\ b = \prod_p p^{b_p} \text{ so that } \prod_p p^{2a_p} = 2\prod_p p^{2b_p},$$

and therefore $2a_2 = 1 + 2b_2$ which is impossible mod 2.

More generally we have

Proposition 3.4. *If d is an integer for which $\sqrt{d}$ is rational, then $\sqrt{d}$ is an integer. Therefore if integer d is not the square of an integer than $\sqrt{d}$ is irrational.*

Proof. We may write $\sqrt{d} = a/b$ where a and b are coprime positive integers, and $a^2 = db^2$. Write $a = \prod_p p^{a_p},\ b = \prod_p p^{b_p},\ d = \prod_p p^{d_p}$ where each $a_p, b_p, d_p \geq 0$, so that $2a_p = 2b_p + d_p$ for each prime p, as $a^2 = db^2$. Therefore if $b_p > 0$ or $d_p > 0$ then $a_p = b_p + d_p/2 > 0$, and so $b_p = 0$ as $(a, b) = 1$; but then $d_p = 2a_p$. Therefore $b = 1$ and $d = a^2$.

3.3. Dividing in congruences. We are now ready to return to the topic of dividing both sides of a congruence through by a given divisor.

Lemma 3.5. *If d divides both a and b and $a \equiv b \pmod m$ where $g = \gcd(d, m)$ then*

$$a/d \equiv b/d \pmod{m/g}.$$

Proof. We may write $a = dA$ and $b = dB$ for some integers A and B, so that $dA \equiv dB \pmod m$. Hence m divides $d(A-B)$ and therefore $\frac{m}{g}$ divides $\frac{d}{g}\ (A-B)$. Now $\gcd(\frac{m}{g}, \frac{d}{g}) = 1$ by Corollary 1.4, and so $\frac{m}{g}$ divides $A - B$ by Euclid's Lemma.

Corollary 3.6. *If $(a, m) = 1$ then $u \equiv v \pmod m$ if and only if $au \equiv av \pmod m$.*

Proof. First use the third part Lemma 2.2 to verify that if $u \equiv v \pmod m$ then $au \equiv av \pmod m$. Then let $\{a, b\} = \{au, av\}$ and $d = a$ in Lemma 3.5, so that $g = (a, m) = 1$, to verify that if $au \equiv av \pmod m$ then $u \equiv v \pmod m$.

Corollary 3.6 implies that if $(a, m) = 1$ then

$$a.0,\ a.1,\ \ldots,\ a.(m-1)$$

is a complete set of residues $\pmod m$, since there are m of them and no two of them are congruent. In particular one of these is congruent to $1 \pmod m$; and so we deduce:

Corollary 3.7. *If $(a, m) = 1$ then there exists an integer r such that $ar \equiv 1 \pmod m$. We call r the* inverse *of $a \pmod m$. We often denote this by $1/a \pmod m$.*

Third Proof of Theorem 1.3. For any given integers A, M let $A = ag, M = mg$ where $g = \gcd(A, M)$ so that $(a, m) = 1$. Then, by Corollary 3.7, there exists an integer r such that $ar \equiv 1 \pmod m$, and so there exists an integer s such that $ar - 1 = ms$; that is $ar - ms = 1$. Hence $Ar - Ms = g(ar - ms) = g = \gcd(A, M)$.

This also goes in the other direction:

Second proof of Corollary 3.7. By Theorem 1.3 there exist integers u and v such that $au + mv = 1$ and so

$$au \equiv au + mv = 1 \pmod{m}.$$

`Exercise 3.3.1.` Prove that if $(a, m) = 1$ and b is an integer then

$$a.0 + b,\ a.1 + b,\ \ldots,\ a(m-1) + b$$

is a complete set of residues $\pmod{m}$.

`Exercise 3.3.2.` Deduce that, whenever $(a, m) = 1$, for all given integers b and c, there is a unique value of $x \pmod{m}$ for which $ax + b \equiv c \pmod{m}$.

`Exercise 3.3.3.` Prove that if $\{r_1, \ldots r_k\}$ is a reduced set of residues mod m, and $(a, m) = 1$ then $\{ar_1, \ldots ar_k\}$ is also a reduced set of residues mod m

3.4. Linear equations in two unknowns. Given integers a, b, c can we find all solutions in integers m, n to

$$am + bn = c\ ?$$

Theorem 3.8. *Let a, b, c be given integers. There are solutions in integers m, n to $am + bn = c$ if and only if* $\gcd(a, b)$ *divides c. If there are solutions then one solution, call it r, s, can be found using the Euclidean algorithm. All other integer solutions are given by*

$$m = r + \ell \frac{b}{(a,b)}, \quad n = s - \ell \frac{a}{(a,b)} \qquad \textit{where } \ell \textit{ is an integer.}$$

Proof 1. If there are solutions m, n then $\gcd(a, b)$ divides $am + bn = c$ by exercise 1.1.1c. Hence there are no solutions when $\gcd(a, b)$ does not divide c. On the other hand, we have seen that there exists integers u, v such that $au + bv = (a, b)$ and so if $c = k(a, b)$ then $a(ku) + b(kv) = c$.

Given one solution r, s to $ar + bs = c$ we can find all other solutions by noting that if $am + bn = c = ar + bs$ then

$$a(m - r) = b(s - n).$$

Hence $b/(a, b)$ divides $m - r$ by Corollary 3.2, so we can write $m = r + \ell b/(a, b)$ for some integer ℓ, and then $n = s - \ell a/(a, b)$.

Note that the real solutions to $ax + by = c$ are given by $x = r + kb,\ y = s - ka,\ k \in \mathbb{R}$. The integer solutions come when $k = \ell/(a, b)$ where $\ell \in \mathbb{Z}$.

An equation involving a congruence is said to be *solved* when integer values can be found for the variables so that the congruence is satisfied. For example $6x + 5 \equiv 13 \pmod{11}$ has the unique solution $x \equiv 5 \pmod{11}$, that is all integers of the form $11k + 5$.

Proof 2. For a given integer m there exists an integer n such that $am + bn = c$ if and only if $am \equiv c \pmod{b}$. In that case $c \equiv am \equiv 0 \pmod{(a, b)}$ as $(a, b)|b$. If so write

$a = (a, b)A$, $b = (a, b)B$, $c = (a, b)C$ and then we are looking for solutions to $Am \equiv C \pmod B$ where $(A, B) = 1$. If $q \equiv 1/A \pmod B$ then this is equivalent to

$$m \equiv qAm \equiv qC \pmod B.$$

That is the set of solutions m is a residue class mod $B = b/(a, b)$ and the result follows.

There is another way to interpret our Theorem:

The Local-Global Principal for Linear Equations. *Let a, b, c be given integers. There are solutions in integers m, n to $am + bn = c$ if and only if for all positive integers r there exist residue classes $u, v \pmod r$ such that $au + bv \equiv c \pmod r$.*

Proof. If $am + bn = c$ then $am + bn \equiv c \pmod r$ for all $r \geq 1$. On the other hand if $au + bv \equiv c \pmod b$ and m is any integer $\equiv u \pmod{b/(a, b)}$ then $am \equiv au + bv \equiv c \pmod b$, as $a \cdot b/(a, b) = b \cdot a/(a, b) \equiv 0 \pmod b$, and so there exists an integer n such that $am + bn = c$.

Remark. Note that it suffices to take only the modulus $r = b$ in this result.

The Frobenius postage stamp problem: If we only have postage stamps worth a cents and b cents where $(a, b) = 1$, what amounts can we make up? That is, what is the set

$$\mathcal{P}(a, b) := \{am + bn : \ m, n \in \mathbb{Z}, \ m, n \geq 0\} \ \ ?$$

(Note that in $\mathcal{P}(a, b)$ we only allow non-negative coefficients for a and b in our linear combinations, whereas in $I(a, b)$ there is no such restriction.) Suppose that r is an integer with $0 \leq r \leq b - 1$. If $N = am + bn \in \mathcal{P}(a, b)$ with $N \equiv ar \pmod{ab}$ then $am \equiv N \equiv ar \pmod n$ so that $m \equiv r \pmod b$ and hence $m = r + bk$ for some integer $k \geq 0$. Therefore $N = am + bn = ar + b(n + ak)$, and so the elements of $\mathcal{P}(a, b)$ in the arithmetic progression $ar \pmod b$ are all those elements of the arithmetic progression that are $\geq ar$. Hence $a(b-1) - b = ab - a - b$ is the largest integer that is not in $\mathcal{P}(a, b)$.

`Exercise 3.4.1.` Show that if $1 \leq M, N \leq ab$ with $(M, ab) = 1$ and $M + N = ab$ then exactly one of M and N is in $\mathcal{P}(a, b)$. (Hint: Given a representation of M, find one of N.)

Determining, in general, the largest integer that does not belong $\mathcal{P}(a, b, c)$, is an open problem.

3.5. Congruences to several moduli. What are the integers that satisfy given congruences to two different moduli?

Lemma 3.9. *Suppose that a, A, b, B are integers. There exists an integer x satisfying $x \equiv a \pmod A$ and $x \equiv b \pmod B$ if and only if $b \equiv a \pmod{\gcd(A, B)}$. If so, this holds for all those x belonging to a unique residue class $\pmod{\operatorname{lcm}[A, B]}$.*

Proof. Integers x satisfying $x \equiv a \pmod A$ can be written as $x = Ay + a$ for an arbitrary integer y, and then $Ay + a = x \equiv b \pmod B$ has solutions if and only $\gcd(A, B)$ divides $b - a$ by Theorem 3.8 (as in the second proof). Moreover Theorem 3.8 gives us that y is any element of a particular residue class mod $B/(A, B)$, and so $x = Ay + a$ is any element of a particular residue class modulo $AB/(A, B) = [A, B]$.

The generalization of this last result is most elegant when we restrict to moduli that are pairwise coprime.

The Chinese Remainder Theorem. *Suppose that $m_1, m_2, \dots, m_k$ are a set of pairwise coprime positive integers. For any set of residue classes $a_1 \pmod{m_1}$, $a_2 \pmod{m_2}, \dots, a_k \pmod{m_k}$, there exists a unique residue class $x \pmod{m}$, where $m = m_1 m_2 \dots m_k$, such that $x \equiv a_j \pmod{m_j}$ for each j.*

Proof. We can map $x \pmod{m}$ to the vector $(x \pmod{m_1},\ x \pmod{m_2}, \dots, x \pmod{m_k})$. There are $m_1 m_2 \dots m_k$ different such vectors and each different x mod m maps to a different one, for if $x \equiv y \pmod{m_j}$ for each j then $x \equiv y \pmod{m}$ by exercise 3.1.16. Hence there is a suitable 1-to-1 correspondence between residue classes mod m and vectors, which implies the result.

This is known as the *Chinese Remainder Theorem* because of the ancient Chinese practice (as discussed in Sun Tzu's 4th century *Classic Calculations*) of counting the number of soldiers in a platoon by having them line up in three columns and seeing how many are left over, then in five columns and seeing how many are left over, and finally in seven columns and seeing how many are left over, etc. For instance, if there are a hundred soldiers then one has 1,0 and 2 soldiers left over respectively; and the next smallest number of soldiers you would have to have for this to be true is 205 ... Presumably an experienced commander can eyeball the difference between 100 soldiers and 205! Primary school children in China learn a song that celebrates this contribution.

In order to make the Chinese Remainder Theorem practical we need an algorithm for determining x, given $a_1, a_2, \dots, a_k$. This can be done be constructing a formula for x: Since $(m/m_j, m_j) = 1$ there exists an integer b_j such that $b_j \cdot \frac{m}{m_j} \equiv 1 \pmod{m_j}$ for each j, by Corollary 3.7. Then

$$x \equiv a_1 b_1 \cdot \frac{m}{m_1} + a_2 b_2 \cdot \frac{m}{m_2} + \dots + a_k b_k \cdot \frac{m}{m_k} \pmod{m}. \tag{3.5}$$

We can verify that this works, since m_j divides m/m_i for each $i \neq j$, and therefore $x \equiv a_j \cdot b_j \frac{m}{m_j} \equiv a_j \cdot 1 \equiv a_j \pmod{m_j}$ for each j. Note that the b_j can all be determined using the Euclidean algorithm, so x can be determined rapidly in practice.

In Gauss's 1801 book he gives an example involving what was then a practical question, but one that is long forgotten today. Before pocket watches and cheap printing, people were perhaps more aware of solar cycles and the moon's phases then what year it actually was. Moreover from Roman times to Gauss's childhood, taxes were difficult to collect since travel was difficult and expensive, and so were not paid annually but rather on a multiyear cycle. Gauss explained how to use the Chinese Remainder Theorem to deduce the year in the Julian calendar from this information: The three pieces of information given were

- The *indiction*, which is $\equiv$ year $+ 3 \pmod{15}$, was used from 312 to 1806 to specify the position of the year in a 15 year taxation cycle.
- The *golden number*, which is $\equiv$ year $+ 1 \pmod{19}$, since the moon's phases and the days of the year repeat themselves every 19 years.[4]
- The *solar cycle*, which is $\equiv$ year $+ 9 \pmod{28}$, since the days of the week and the dates of the year repeat in cycles of 28 years in the Julian calender.[5]

[4] Meton of Athens (5th century BC) observed that 19 (solar) years is less than two hours out from being a whole number of lunar months.

[5] Since there are seven days in a week, and leap years occur every four years.

Taking $m_1 = 15,\ m_2 = 19,\ m_3 = 28$, we observe that

$$\begin{aligned}
b_1 &\equiv \frac{1}{19\cdot 28} \equiv \frac{1}{4\cdot(-2)} \equiv -2 \pmod{15} \quad \text{and} \quad b_1 \cdot \frac{m}{m_1} = -2\cdot 19\cdot 28 = -1064;\\
b_2 &\equiv \frac{1}{15\cdot 28} \equiv \frac{1}{(-4)\cdot 9} \equiv \frac{1}{2} \equiv 10 \pmod{19} \quad \text{and} \quad b_2 \cdot \frac{m}{m_2} = 10\cdot 15\cdot 28 = 4200;\\
b_3 &\equiv \frac{1}{15\cdot 19} = \frac{1}{(14+1)\cdot 19} \equiv \frac{1}{14+19} \equiv \frac{1}{5} \equiv -11 \pmod{28} \quad \text{and} \quad b_3 \cdot \frac{m}{m_3} = -3135.
\end{aligned}$$

Therefore if the indiction is a, the golden number is b, and the solar cycle is c then the year is

$$\equiv -1064a + 4200b - 3135c \pmod{7980}.$$

`Exercise 3.5.1.` Use this method to give a general formula for $x \pmod{1001}$ if $x \equiv a \pmod 7$, $x \equiv b \pmod{11}$ and $x \equiv c \pmod{13}$.

`Exercise 3.5.2.` Suppose that $p_1 < p_2 < \ldots < p_k$ are primes, and that $f(x) \in \mathbb{Z}[x]$. Prove that there exist integers $a_1, \ldots a_k$ such that $f(a_i) \equiv 0 \pmod{p_i}$ for $1 \leq i \leq k$, if and only if there exists an integer a such that $f(a) \equiv 0 \pmod{p_1 p_2 \ldots p_k}$.

There is more discussion of the Chinese Remainder Theorem in section B2.

4. Multiplicative functions

A function f is *multiplicative* if $f(mn) = f(m)f(n)$ for all pairwise coprime positive integers m, n; and *totally multiplicative* if $f(mn) = f(m)f(n)$ for all $m, n \geq 1$. We already saw the example $\tau(n)$, which counts the number of divisors of n, which is multiplicative but not totally multiplicative, since $\tau(p^a) = a + 1$. One key observation is that being multiplicative allows one to evaluate $f(n)$ in terms of the $f(p^e)$ for the prime powers p^e dividing n.

`Exercise 4.1.1.` Show that if f is multiplicative, and $n = \prod_{p \text{ prime}} p^{n_p}$ then

$$f(n) = \prod_{p \text{ prime}} f(p^{n_p}).$$

Deduce that if f is totally multiplicative then $f(n) = \prod_p f(p)^{n_p}$.

We will focus in this section on two further examples of multiplicative functions of great interest.

4.1. Euler's ϕ-function. It is natural that we wish to know the value of

$$\phi(n) := \#\{m : \ 1 \leq m \leq n \text{ and } (m, n) = 1\}$$

for any $n \geq 1$. We have already seen that there are always $\phi(m)$ elements in a reduced system of residues. Evidently $\phi(1) = 1$.

Lemma 4.1. *$\phi(n)$ is a multiplicative function.*

Proof. Suppose that $n = mr$ where $(m, r) = 1$. By the Chinese Remainder Theorem there is natural bijection between the integers $a \pmod n$ with $(a, n) = 1$, and the pairs of integers $(b \pmod m, c \pmod r)$ with $(b, m) = (c, r) = 1$. Therefore $\phi(n) = \phi(m)\phi(r)$.

Hence to evaluate $\phi(n)$ for all n we simply need to evaluate it on the prime powers, which is straightforward: If $n = p$ is prime then $\phi(p)$ is simply the number of integers $1, 2, \ldots, p-1$; that is $\phi(p) = p-1$. If $n = p^a$ is a prime power then we count every integer $1 \leq m \leq p^a$ except those that are a multiple of p, that is except for $p, 2p, 3p, \ldots, (p^{a-1})p$. Therefore

$$\phi(p^a) = p^a - p^{a-1} = p^{a-1}(p-1) = p^a \left(1 - \frac{1}{p}\right).$$

Hence we deduce

Theorem 4.2. *If $n = \prod_{p \text{ prime}} p^{n_p}$ then*

$$\phi(n) = \prod_{\substack{p \text{ prime} \\ p|n}} (p^{n_p} - p^{n_p - 1}) = \prod_{\substack{p \text{ prime} \\ p|n}} p^{n_p} \left(1 - \frac{1}{p}\right) = n \prod_{\substack{p \text{ prime} \\ p|n}} \left(1 - \frac{1}{p}\right).$$

Example: $\phi(60) = 60 \cdot \left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{3}\right)\left(1 - \frac{1}{5}\right) = 16$, the numbers being
$1, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 49, 53$ and 59.

`Exercise 4.1.2.` Show that there are $[x/d]$ natural numbers $\leq x$ that are divisible by d.

We give an alternative proof of Theorem 4.2, based on the inclusion-exclusion principle, in section F1.

If one looks at values of $\phi(n)$ one makes a surprising observation:

Proposition 4.3. *We have* $\sum_{d|n} \phi(d) = n$.

Example: For $n = 30$, we have $\phi(1)+\phi(2)+\phi(3)+\phi(5)+\phi(6)+\phi(10)+\phi(15)+\phi(30) = 1+1+2+4+2+4+8+8 = 30$

Proof. For any integer m one knows that (m,n) is a divisor n, and so $d = n/(m,n)$ is an integer for which $(m,n) = n/d$. The n integers m in the range $1 \leq m \leq n$ can be counted by summing up the number of integers $1 \leq m \leq n$ for which $(m,n) = n/d$, for each divisor d of n. Writing $m = an/d$, this equals the number of integers $1 \leq a \leq d$ for which $(a,d) = 1$, which is $\phi(d)$ by definition. The result follows.

`Exercise 4.1.3.` Prove that $\prod_{d|n} d = n^{\tau(n)/2}$.

`Exercise 4.1.4.` The function $\phi(m)$ is fundamental in number theory. Looking at its values, Carmichael came up with the conjecture that for all integers m there exists an integer $n \neq m$ such that $\phi(n) = \phi(m)$. By considering $n = 2m$ and $n = 3m$ show that Carmichael's conjecture is true if m is odd or if m is not divisible by 3. Can you find other classes of m for which it is true? Carmichael's conjecture is still an open problem but it is known that if it is false then the smallest counterexample is $> 10^{10^{10}}$?

4.2. Perfect numbers. 6 is a *perfect number*, the sum of its smaller divisors, since

$$6 = 1 + 2 + 3.$$

> "*Six is a number perfect in itself, and not because God created all things in six days; rather, the converse is true. God created all things in six days because the number is perfect...*" — from *The City of God* by Saint Augustine (354-430).

The next perfect number is $28 = 1+2+4+7+14$ which is the number of days in a lunar month. However the next, $496 = 1 + 2 + 4 + 8 + 16 + 31 + 62 + 124 + 248$, appears to have little cosmic relevance, though we will be interested in trying to classify them all. To create an equation we will add the number to both sides to obtain that n is perfect if and only if

$$2n = \sigma(n), \text{ where } \sigma(n) := \sum_{d|n} d.$$

`Exercise 4.2.1.` Show that n is perfect if and only if $\sum_{d|n} \frac{1}{d} = 2$.

`Exercise 4.2.2.` Prove that if $(a,b) = 1$ then each divisor of ab can be written as ℓm where $\ell|a$ and $m|b$.

By this last exercise we see that if $(a,b) = 1$ then

$$\sigma(ab) = \sum_{d|ab} d = \sum_{\ell|a,\ m|b} \ell m = \sum_{\ell|a} \ell \cdot \sum_{m|b} m = \sigma(a)\sigma(b),$$

proving that σ is a multiplicative function. Now

$$\sigma(p^k) = 1 + p + p^2 + \ldots + p^k = \frac{p^{k+1}-1}{p-1}$$

by definition, and so if $n = \prod_i p_i^{k_i}$ then

$$\sigma(n) = \prod_i \frac{p_i^{k_i+1} - 1}{p_i - 1} .$$

Proposition 4.4. (Euclid) *If $2^p - 1$ is a prime number then $2^{p-1}(2^p - 1)$ is a perfect number.*

Note that the cases $p = 2,\ 3,\ 5$ give the three examples above (and then 7 and 13).

Proof. Since σ is multiplicative we have, for $n = 2^{p-1}(2^p - 1)$,

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

After extensive searching one finds that these appear to be the only perfect numbers. Euler succeeded in proving that these are the only *even* perfect numbers, and we believe that there are no *odd* perfect numbers.

Theorem 4.5. (Euler) *If n is an even perfect number then there exists a prime number of the form $2^p - 1$ such that $n = 2^{p-1}(2^p - 1)$.*

Proof. Write $n = 2^{k-1}m$ where m is odd and $k \geq 2$, so that

$$2^k m = \sigma(n) = \sigma(2^{k-1})\sigma(m) = (2^k - 1)\sigma(m).$$

Now $(2^k - 1, 2^k) = 1$ and so $2^k - 1$ divides m. Writing $m = (2^k - 1)M$ we find that $\sigma(m) = 2^k M = m + M$. That is $\sigma(m)$, which is the sum of all of the divisors of m, equals the sum of just two of its divisors, implying that those are the only two divisors of m. But the only integers with just two divisors are the primes, so that m is a prime and $M = 1$ and the result follows.

We will discuss numbers of the form $2^m - 1$ in more detail in the next chapter. In exercise 5.1.4 we will show that if $2^p - 1$ is prime then p must itself be prime.

It is widely believed that the only perfect numbers were those identified by Euclid; that is that there are no odd perfect numbers. It has been proved that if there is an odd perfect number then it is $> 10^{300}$.

`Exercise 4.2.3.` Prove that if p is odd and k is odd then $\sigma(p^k)$ is even. Deduce that if n is an odd perfect then $n = pm^2$ where p is a prime $\equiv 1 \pmod 4$.

5. The Distribution of Prime Numbers

Once one begins to determine which integers are primes, one quickly finds that there are many of them, though as we go further and further, they seem to be a smaller and smaller proportion of the positive integers. It is also tempting to look for patterns amongst the primes: Can we find a formula that describes all of the primes? Or at least some of them? Are there actually infinitely many? And, if so, can we quickly determine how many there are up to a given point? Or at least give a good estimate? Once one has spent long enough determining primes, one cannot help but ask whether it is possible to recognize prime numbers quickly and easily? These questions motivate different parts of this section and of section 10.

5.1. Proofs that there are infinitely many primes. The first known proof appears in Euclid's *Elements*, Book 9 Proposition 20:

Theorem 5.1. *There are infinitely many primes.*

Proof 1. Suppose that there are only finitely many primes, which we will denote by $2 = p_1 < p_2 = 3 < \ldots < p_k$. What are the prime factors of $p_1p_2\ldots p_k+1$? Since this number is > 1 it must have a prime factor by the Fundamental Theorem of Arithmetic, and this must be p_j for some $j,\ 1 \leq j \leq k$, since *all* primes are contained amongst $p_1, p_2, \ldots, p_k$. But then p_j divides both $p_1p_2\ldots p_k$ and $p_1p_2\ldots p_k + 1$, and hence p_j divides their difference, 1, by exercise 1.1.1c, which is impossible.

`Exercise 5.1.1.` (Proof # 2) Suppose that there are only finitely many primes, the largest of which is n. Show that this is impossible by considering the prime factors of $n! - 1$.

`Exercise 5.1.2.` Prove that there are infinitely many composite numbers.

Our first proof that there are infinitely many primes proceeds by showing that it is impossible that there are finitely many. It is mildly disturbing that we do not show how to find primes in proving that that there are infinitely many primes, nor give an algorithm for how to find infinitely many. We correct this deficiency by defining the sequence

$$a_1 = 2, a_2 = 3 \text{ and then } \ a_n = a_1a_2\ldots a_{n-1} + 1 \text{ for each } n \geq 2.$$

For each $n \geq 1$ let p_n be some prime divisor of a_n. We claim that the p_n are all distinct so we have an infinite sequence of distinct primes. We know that these primes are distinct else if $p_m = p_n$ with $m < n$ then p_m divides $(a_m, a_n) = (a_m, 1) = 1$ by exercise 2.1.2, since $a_n \equiv 1 \pmod{a_m}$ by our construction, which is impossible.

Fermat conjectured that the integers $F_n = 2^{2^n} + 1$ are primes for all $n \geq 0$. His claim starts off correct: $3, 5, 17, 257, 65537$ are all prime, but is false for $F_5 = 641 \times 6700417$, as Euler famously noted. It is an open question as to whether there are infinitely many primes of the form F_n.[6] Nonetheless we can prove that if p_n is some prime divisor of F_n for

[6] There are no Fermat primes, $2^{2^n} + 1$, known other than for $n \leq 4$, and we know that the Fermat numbers, $2^{2^n} + 1$, are composite for $5 \leq n \leq 30$ and many other n besides. It is always a significant moment when a Fermat number is factored for the first time. It could be that all $F_n, n > 4$ are composite, or they might all be prime for some sufficiently large n. Currently, we have no way of knowing what is the truth.

each $n \geq 0$ then $p_0, p_1, \dots$ is an infinite sequence of distinct primes, in this case because $F_n = F_1 F_2 \dots F_{n-1} + 2$ for each $n \geq 1$, and so $(F_m, F_n) = (F_m, 2) = 1$ for all $m < n$, since $F_n \equiv 2 \pmod{F_m}$. (This proof appeared in a letter from Goldbach to Euler in July 1730.)

`Exercise 5.1.3.` Suppose that $p_1 = 2 < p_2 = 3 < \dots$ is the sequence of prime numbers. Use the fact that every Fermat number has a distinct prime divisor to prove that $p_n \leq 2^{2^n} + 1$. What can one deduce about the number of primes up to x?

The *Mersenne numbers* take the form $M_n = 2^n - 1$.

`Exercise 5.1.4.` Prove that if n is composite then M_n is composite, by showing that M_a divides M_{ab}. Show that if m is not a power of 2 then $2^m + 1$ is composite by showing that if $2^a + 1$ divides $2^{ab} + 1$ whenever b is odd.

Hence if M_n is prime then $n = p$ is prime. It is conjectured that there are infinitely many Mersenne primes $M_p = 2^p - 1$.[7] We saw in section 4.2 that the Mersenne primes are in 1-to-1 correspondance with the even perfect numbers.

Also if $2^m + 1$ is prime then m equals 0 or a power of 2, say 2^n, and therefore we have a Fermat number $F_n = 2^{2^n} + 1$.

Furstenberg's extraordinary proof that there are infinitely many primes, using point set topology. Define a topology on the set of integers $\mathbb{Z}$ in which a set S is open if it is empty or if for every $a \in S$ there is an arithmetic progression $\mathbb{Z}_{a,q} := \{a + nq : n \in \mathbb{Z}\}$ which is a subset of S. Evidently each $\mathbb{Z}_{a,q}$ is open, and it is also closed since

$$\mathbb{Z}_{a,q} = \mathbb{Z} \setminus \bigcup_{b:\ 0 \leq b \leq q-1,\ b \neq a} \mathbb{Z}_{b,q}.$$

If there are only finitely many primes p then $A = \cup_p \mathbb{Z}_{0,p}$ is also closed, and so $\mathbb{Z} \setminus A = \{-1, 1\}$ is open, but this is obviously false since A does not contain any arithmetic progression $\mathbb{Z}_{1,q}$. Hence there are infinitely many primes.

Remark: This is Euclid's proof in heavy disguise: In effect Furstenberg's proof states that the integer $1 + \ell p_1 p_2 \dots p_k$ is evidently not in any of the arithmetic progressions $\mathbb{Z}_{0,p_i}$ so cannot be divisible by any prime, contradiction. More simply, a proof that lies somewhere between those of Euclid and Furstenberg:

Proof # 6. Suppose that there are only finitely many primes, namely $p_1, p_2, \dots, p_k$. Let $m = p_1 p_2 \cdots p_k$ be their product. If r is an integer with $(r, m) = 1$ then r cannot divisible by any primes (since they all divide m), and so must equal -1 or 1. Therefore $\phi(m) = \#\{1 \leq r \leq m : \ (r, m) = 1\} = 1$, but this is easily seen to contradict our formula in Theorem 4.2

5.2. The Sieve of Eratosthenes. The sieve of Eratosthenes yields an efficient method to find all of the primes up to x. We begin by writing down every integer up to x and then deleting every composite even number, that is one deletes every second integer up to

[7] It is known that $2^p - 1$ is prime for $p = 2, 3, 5, 7, 13, 17, 19, \dots, 43112609$, a total of 46 values as of late 2008. There is a long history of the search for Mersenne primes, from the first serious computers to the first great distributed computing project, GIMPS (The Great Internet Mersenne Prime Search).

x after 2. The first undeleted integer > 2, is 3; one then deletes every composite integer divisible by 3, that is every third integer up to x after 3. The next undeleted integer is 5 and one deletes every fifth integer subsequently. One keeps on going like this, finding the next undeleted integer, call it p, which must be prime, and then delete every pth integer beyond p and up to x. We stop once $p > \sqrt{x}$ and then the undeleted integers are the primes $\leq x$. There are about $x \log\log x$ steps in this algorithm, so it is remarkably efficient.

`Exercise 5.2.1.` Use this method to find all of the primes up to 100

The number of integers left after one removes the multiples of 2 is roughly $\frac{1}{2} \cdot x$. After one removes the multiples of 3, one expects that there are about $\frac{2}{3} \cdot \frac{1}{2} \cdot x$ integers left. In general removing the multiples of p removes, we expect, about $1/p$ of the integers, and so leaves a proportion $1 - \frac{1}{p}$. Therefore we expect that the number of primes up to x, which equals the number of integers left, up to x, by the sieve of Eratosthenes, is about

$$x \prod_{\substack{p \leq \sqrt{x} \\ p \text{ prime}}} \left(1 - \frac{1}{p}\right).$$

The product $\prod_{p \leq y}(1 - \frac{1}{p})$ is well approximated by $c/\log y$, where $c \approx 0.5614594836$,[8] so one might guess from these sieve methods that the number of primes up to x is approximately

$$2c\ \frac{x}{\log x}\ . \tag{5.2}$$

5.3. Primes in certain arithmetic progressions. How are the primes split between the arithmetic progressions modulo 3? Or 4? Or modulo any given integer m? Evidently every integer in the arithmetic progression $0 \pmod 3$ (that is integers of the form $3k$) is divisible by 3, so the only prime in that arithmetic progression is 3 itself. There are no such divisibility restrictions for the arithmetic progressions $1 \pmod 3$ and $2 \pmod 3$ and if we calculate the primes up to 100 we find

Primes $\equiv 1 \pmod 3$: $7, 13, 19, 31, 37, 43, 61, 67, 73, 79, 97, \dots$
Primes $\equiv 2 \pmod 3$: $5, 11, 17, 23, 29, 41, 47, 53, 59, 71, 83, 89, \dots$

There seem to be lots of primes in either arithmetic progression, and they seem to be roughly equally split between the two. Let's see what we can prove. First let's deal, in general, with the analogy to the case $0 \pmod 3$. This includes not only $0 \pmod m$ but also cases like $2 \pmod 4$:

`Exercise 5.3.1.` Prove that any integer $\equiv a \pmod m$ is divisible by (a, m). Deduce that if $(a, m) > 1$ then there cannot be more than one prime $\equiv a \pmod m$. Give examples of arithmetic progressions which contain exactly one prime, and examples which contain none.

Thus all but finitely many primes are distributed among the $\phi(m)$ arithmetic progressions $a \pmod m$ with $(a, m) = 1$. How are they distributed? If the $m = 3$ case is anything to go by it appears that there are infinitely many in each such arithmetic progression and maybe even roughly equal numbers.

[8] This is a fact that is beyond of this book.

We will prove that there are infinitely many primes in each of the two feasible residue classes mod 3 (see Theorems 5.2 and 7.17). Proving that there are roughly equally many in each progression is rather more difficult.

`Exercise 5.3.2.` Use exercise 3.1.3 to show that if $n \equiv -1 \pmod 3$ then there exists a prime factor p of n which is $\equiv -1 \pmod 3$.

Theorem 5.2. *There are infinitely many primes* $\equiv -1 \pmod 3$.

Proof. If there are only finitely many, say $p_1, p_2, \dots, p_k$, then $N = 3p_1p_2 \dots p_k - 1$ must have a prime factor $q \equiv -1 \pmod 3$, by exercise 5.3.2. However q divides both N and $N+1$ (since it must be one of the primes p_i), and hence q divides their difference 1, which is impossible.

`Exercise 5.3.3.` Prove that there are infinitely many primes $\equiv -1 \pmod 4$.

`Exercise 5.3.4.` Prove that there are infinitely many primes $\equiv 5 \pmod 6$. (Hint: Consider splitting arithmetic progressions mod 3 into several arithmetic progressions mod 6.)

`Exercise 5.3.5.` Prove that at least two of the arithmetic progressions mod 8 contain infinitely many primes (one might use exercise 3.1.4 in this proof).

In an exercise in section B4 one can generalize this considerably, using basically the same ideas.

The 1837 Dirichlet showed that whenever $(a, q) = 1$ there are infinitely many primes $\equiv a \pmod q$. We discuss this deep result in section E4. In fact we know that the primes are roughly equally distributed amongst these arithmetic progressions. In other words, half the primes are $\equiv 1 \pmod 3$ and half are $\equiv -1 \pmod 3$. Roughly 1% of the primes are $\equiv 69 \pmod{101}$ and indeed in each arithmetic progression a mod 101 with $1 \le a \le 100$.

5.4. How many primes are there up to x? When people started to develop large tables of primes, perhaps looking for a pattern, they discovered no patterns, but did find that the proportion of integers that are prime is gradually diminishing. In 1808 Legendre suggested that there are roughly $\frac{x}{\log x}$ primes up to x, and even the more precise assertion that there exists a constant B such that $\pi(x)$, the number of primes up to x, is well approximated by $x/(\log x - B)$ for large enough x. A few years earlier, aged 15 or 16, Gauss had already made a much better guess, based on studying tables of primes:

> "*In 1792 or 1793 ... I turned my attention to the decreasing frequency of primes ... counting the primes in intervals of length* 1000. *I soon recognized that behind all of the fluctuations, this frequency is on average inversely proportional to the logarithm...*" — from a letter to Encke by K.F. Gauss (Christmas Eve 1849).

His observation may be best phrased as

About 1 in $\log x$ *of the integers near* x *are prime.*

This suggests that a good approximation to the number of primes up to x is $\sum_{n=2}^{x} \frac{1}{\log n}$. Now $\frac{1}{\log t}$ is does not vary much for t between n and $n+1$, and so Gauss deduced that

$\pi(x)$ should be well approximated by

$$\int_2^x \frac{dt}{\log t}\ . \tag{5.4.1}$$

We denote this quantity by $\mathrm{Li}(x)$ and call it *the logarithmic integral*. Here is a comparison of Gauss's prediction with the actual count of primes up to various values of x:

x	$\pi(x) = \#\{\text{primes} \le x\}$	Overcount: $\mathrm{Li}(x) - \pi(x)$
10^3	168	10
10^4	1229	17
10^5	9592	38
10^6	78498	130
10^7	664579	339
10^8	5761455	754
10^9	50847534	1701
10^{10}	455052511	3104
10^{11}	4118054813	11588
10^{12}	37607912018	38263
10^{13}	346065536839	108971
10^{14}	3204941750802	314890
10^{15}	29844570422669	1052619
10^{16}	279238341033925	3214632
10^{17}	2623557157654233	7956589
10^{18}	24739954287740860	21949555
10^{19}	234057667276344607	99877775
10^{20}	2220819602560918840	222744644
10^{21}	21127269486018731928	597394254
10^{22}	201467286689315906290	1932355208
10^{23}	1925320391606818006727	7236148412

TABLE 1. Primes up to various x, and the overcount in Gauss's prediction.

We see that Gauss's prediction is amazingly accurate. It does seem to always be an overcount, and since the width of the last column is about half that of the central one it appears that the difference is no bigger than $\sqrt{x}$, perhaps multiplied by a constant. The data certainly suggests that $\pi(x)/\mathrm{Li}(x) \to 1$ as $x \to \infty$.

Exercise 5.4.1. Integrate (5.4.1) by parts to prove that $\mathrm{Li}(x) = \frac{x}{\log x} - \frac{2}{\log 2} + \int_2^x \frac{dt}{(\log t)^2}$. By bounding $1/\log t$ by a constant in the range $2 \le t \le \sqrt{x}$, and by $1/\log\sqrt{x}$ for larger t, show that there exists a constant κ_N such that $\int_2^x \frac{dt}{(\log t)^2} < \kappa_2 \frac{x}{(\log x)^2}$ for all $x \ge 2$. Deduce that

$$\mathrm{Li}(x)\ \Big/\ \frac{x}{\log x} \to 1 \quad \text{as} \quad x \to \infty.$$

Exercise 5.4.2. What is the best choice of B in Legendre's assertion stated above.

Combining the result of exercise 5.4.1 with Gauss's prediction (5.4.1) gives that $\pi(x)/(\frac{x}{\log x}) \to 1$ as $x \to \infty$. The notation of limits is rather cumbersome notation – it is easier to write

$$\pi(x) \sim \frac{x}{\log x} \tag{5.4.2}$$

as $x \to \infty$, "$\pi(x)$ *is asymptotic to* $x/\log x$". (In general, $A(x) \sim B(x)$ is equivalent to $\lim_{x\to\infty} A(x)/B(x) = 1$.) This is different by a multiplicative constant from (5.2), our guesstimate based on the sieve of Eratosthenes. The data here makes it clear that the constant 1 given here, rather than $2c$, is correct.

The asymptotic (5.4.2) is called *The Prime Number Theorem* and its proof had to wait until the end of the nineteenth century, requiring various remarkable developments. The proof was a high point of nineteenth century mathematics and there is still no straightforward proof. There are reasons for this: Surprisingly the prime number theorem is equivalent to a statement about zeros of an analytic continuation, and although a proof can be given that hides this fact, it is still lurking somewhere just beneath the surface, perhaps inevitably so. A proof of the prime number theorem is beyond the scope of this book.

In section E1 we will prove Chebyshev's 1850 result that there exist constant $c_2 > c_1 > 0$ such that

$$c_1 \frac{x}{\log x} \leq \pi(x) \leq c_2 \frac{x}{\log x}$$

for all $x \geq 100$.

`Exercise 5.4.3.` Let $p_1 = 2 < p_2 = 3 < \dots$ be the sequence of primes. Prove that the prime number theorem is equivalent to the asymptotic

$$p_n \sim n \log n \quad \text{as } n \to \infty.$$

`Exercise 5.4.4.` Assuming the prime number theorem, show that for all $\epsilon > 0$ there are primes between x and $x + \epsilon x$ is x is sufficiently large. Deduce that $\mathbb{R}_{\geq 0}$ is the set of limit points of the set $\{p/q : \ p, q \text{ primes}\}$.

The average gap between consecutive primes up to x is

$$\frac{1}{N-1} \sum_{n=1}^{N-1} (p_{n+1} - p_n) = \frac{p_N - p_2}{N-1} \sim \frac{x}{x/\log x} = \log x$$

where p_N is the largest prime $\leq x$, as the sum is telescoping and $N = \pi(x) \sim \frac{x}{\log x}$ by the prime number theorem, and $p_N \sim x$ by exercise 5.4.3. One might ask whether there are gaps that are much smaller and whether there are gaps that are much larger?

One can easily prove that there are arbitrarily long gaps between consecutive primes, since if $2 \leq j \leq m$ then j divides $m! + j$, and so

$$m! + 2,\ m! + 3, \dots, m! + m$$

are all composite. Hence if p is the largest prime $\leq m! + 1$ and if q is the next largest prime so that $q \geq m! + m + 1$, then $q - p \geq m$. One can extend this argument to prove that

$$\limsup_{n\to\infty} \frac{p_{n+1} - p_n}{\log p_n} = \infty.$$

What about small gaps between primes?

`Exercise 5.4.5.` Prove that 2 and 3 are the only two primes that differ by 1.

There are many pairs of primes that differ by two, namely 3 and 5, 5 and 7, 11 and 13, 17 and 19, etc., seemingly infinitely many, and this *twin prime conjecture* remains an open problem. It is also open as to whether there are infinitely many pairs of primes that differ by no more than 100, and until recently, that differ by no more than 1/4 of the average. However in 2009, Goldston, Pintz and Yildirim showed that

$$\liminf_{n\to\infty} \frac{p_{n+1}-p_n}{\log p_n} = 0.$$

Other famous open problems include:

– There are infinitely many pairs of primes $p, 2p+1$ (*Sophie Germain primes*),
– There are infinitely many primes of the form n^2+1,
– There are infinitely many primes of the form 2^p-1.
– *Goldbach's conjecture* that every even number ≥ 4 is the sum of two primes. This has been verified for all even numbers $\leq 10^{18}$.

5.5. Formulas for primes. Are there formulas that only yield prime values? For example can we give a polynomial $f(x)$ of degree ≥ 1 such that $f(n)$ is prime for every integer n? The example $6n+5$ has values $5, 11, 17, 23, 29$ which are all prime, before getting to $35 = 5 \times 7$. Continuing on gives primes $41, 47, 53, 59$ till we hit $65 = 5 \times 13$, another multiple of 5. One observes that every fifth term of the arithmetic progression is divisible by 5, since $6(5k)+5 = 5(6k+1)$. More generally $qn+a$ is a multiple of a whenever n is a multiple of a, since $q(ak)+a = a(qk+1)$, and so is composite whenever $|a| > 1$ and $k \geq 1$. When $a = 0$ we see that qn is composite for all $n > 1$ provided $q > 1$. When $a = \pm 1$ we need to proceed a little differently. For example the progression $6m-1$ is the same, we observe, as $6n+5$ (by taking $m = n+1$), and we have already dealt with this example. In general, when $n = (q-1)k+1$ we have $qn-1 = (q-1)(qk+1)$, and when $n = (q+1)k+1$ we have $qn+1 = (q+1)(qk+1)$. We develop this argument to work for all polynomials, but will need the following result (which is proved in section A3):

The Fundamental Theorem of Algebra. *If $f(x) \in \mathbb{C}[x]$ has degree $d \geq 1$ then $f(x)$ has no more than d distinct roots in $\mathbb{C}$.*

Proposition 5.3. *If $f(x) \in \mathbb{Z}[x]$ has degree $d \geq 1$ then there are infinitely many integers n such that $|f(n)|$ is composite.*

Proof. There are at most d roots of each of the polynomials $f(x),\ f(x)-1,\ f(x)+1$ by the Fundamental Theorem of Algebra. Select an integer m which is not the root of any of these polynomials so that $|f(m)| > 1$. Now $kf(m)+m \equiv m \pmod{|f(m)|}$ and so, by Corollary 2.3, we have

$$f(kf(m)+m) \equiv f(m) \equiv 0 \pmod{|f(m)|}.$$

There are a most $3d$ values of k for which $kf(m)+m$ is a root of one of $f(x)-|f(m)|,\ f(x)$ or $f(x)+|f(m)|$, by the Fundamental Theorem of Algebra. For any other k we have that

$|f(kf(m)+m)| > |f(m)|$, and so $|f(kf(m)+m)|$ factors into $|f(m)|$ and a complementary factor, and is therefore composite.

So we see that there is no polynomial that only takes primes values.

`Exercise 5.5.1.` Show that if $f(x,y) \in \mathbb{Z}[x,y]$ has degree $d \geq 1$ then there are infinitely many pairs of integers m, n such that $|f(m,n)|$ is composite.

We saw that nine of the first ten values of the polynomial $6n+5$ are primes. Even better is the polynomial n^2+n+41, discovered by Euler in 1772, which is prime for $n = 0, 1, 2, \dots, 39$, and the square of a prime for $n = 40$. However, as in the proof of Proposition 5.3, we know that it is composite whenever n is a positive multiple of 41. See section 12.3 for more on such prime rich polynomials.

Earlier we discussed Fermat numbers $2^{2^n}+1$ which Fermat had believed to all be prime. He was wrong, but perhaps there are other formulas, more exotic than mere polynomials, which yield only primes? One intriguing possibility stems from the fact that

$$2^2-1,\ 2^{2^2-1}-1,\ 2^{2^{2^2-1}-1}-1 \text{ and } 2^{2^{2^{2^2-1}-1}-1}-1$$

are all prime. Could every term in this sequence be prime? No one knows and the next example is so large that one will not be able to determine whether or not it is prime in the foreseeable future. (Draw lessons on the power of computation from this example!)

Actually with a little imagination it is not so difficult to develop formulae that easily yield all of the primes. For example if $p_1 = 2 < p_2 = 3 < \dots$ is the sequence of primes then define

$$\alpha = \sum_{m \geq 1} \frac{p_m}{10^{m^2}} = .2003000050000007000000011\dots$$

One can read off the primes from the decimal expansion of α, the mth prime coming from the few digits to the right of m^2th digit; or, more formally,

$$p_m = [10^{m^2}\alpha] - 10^{2m-1}[10^{(m-1)^2}\alpha].$$

Is α truly interesting? If one could easily describe α (other than by the definition that we gave) then it might provide an easy way to determine the primes. But with its artificial definition it does not seem like it can be used in any practical way. There are other such constructions (see, e.g., exercise 7.3.3).

In a rather different vein, Matijasevič, while working on Hilbert's tenth problem, discovered that there exist polynomials f in many variables, such that the set of positive values taken by f when each variable is set to be an integer, is precisely the set of primes.[9] One can find many different polynomials for the primes; we will give one with 26 variables of degree 21. (One can cut the degree to as low as 5 at the cost of having an enormous

[9]One can also construct such polynomials so as to yield the set of Fibonacci numbers, or the set of Fermat primes, or the set of Mersenne primes, or the set of even perfect numbers, and indeed any *Diophantine* set (and see section 6 for more on "Diophantine").

number of variables. No one knows the minimum possible degree, nor the minimum possible number of variables): Our polynomial is $k+2$ times

$$
\begin{aligned}
&1-(n+l+v-y)^2-(2n+p+q+z-e)^2-(wz+h+j-q)^2-(ai+k+1-l-i)^2\\
&-((gk+2g+k+1)(h+j)+h-z)^2-(z+pl(a-p)+t(2ap-p^2-1)-pm)^2\\
&-(p+l(a-n-1)+b(2an+2a-n^2-2n-2)-m)^2-((a^2-1)l^2+1-m^2)^2\\
&-(q+y(a-p-1)+s(2ap+2a-p^2-2p-2)-x)^2-((a^2-1)y^2+1-x^2)^2\\
&-(16(k+1)^3(k+2)(n+1)^2+1-f^2)^2-(e^3(e+2)(a+1)^2+1-o^2)^2\\
&-(16r^2y^4(a^2-1)+1-u^2)^2-(((a+u^2(u^2-a))^2-1)(n+4dy)^2+1-(x+cu)^2)^2.
\end{aligned}
$$

Stare at this for a while and try to figure out how it works: The key is to determine when the displayed polynomial takes positive values; note that it is equal to 1 minus a sum of squares. Understanding much beyond this seems difficult, and it seems that the only way to appreciate this polynomial is to understand its derivation – see [JSW]. In the current state of knowledge it seems that this absolutely extraordinary and beautiful polynomial is entirely useless in helping us better understand the distribution of primes!

6. Diophantine problems

Diophantus lived in Alexandria in the third century A.D. His thirteen volume *Arithmetica* dealt with solutions to equations in integers and rationals (though only parts of six of the volumes have survived). Diophantus's work was largely forgotten in Western Europe during the Dark Ages, as ancient Greek became much less studied; but interest in *Arithmetica* was revived by Bachet's 1621 translation into Latin.[10] In his honour, a *Diophantine equation* is a polynomial equation for which we are searching for integer or rational solutions.

6.1. The Pythagorean equation. We wish to find all solutions in integers x, y, z to

$$x^2 + y^2 = z^2.$$

We may assume that x, y, z are all positive and so $z > x, y$. Given any solution we may divide through by any common factor of x, y and z to obtain a solution where $(x, y, z) = 1$.

`Exercise 6.1.1.` Prove that if $(x, y, z) = 1$ and $x^2 + y^2 = z^2$ then x, y and z are pairwise coprime.

Now x and y cannot both be odd else $z^2 = x^2 + y^2 \equiv 1 + 1 \equiv 2 \pmod 8$, which is impossible. Interchanging x and y if necessary we may assume that x is even and y and z are odd. Now

$$(z - y)(z + y) = z^2 - y^2 = x^2.$$

We prove that $(z - y, z + y) = 2$: Since y and z are both odd, we know that 2 divides $(z-y, z+y)$. Moreover $(z-y, z+y)$ divides $(z+y)-(z-y) = 2y$ and $(z+y)+(z-y) = 2z$, and hence $(2y, 2z) = 2(y, z) = 2$. Therefore, by exercise 3.1.20, there exist integers r, s such that

$$z - y = 2s^2 \text{ and } z + y = 2r^2,$$

so that

$$x = 2rs,\ y = r^2 - s^2, \text{ and } z = r^2 + s^2.$$

To ensure these are pairwise coprime we need $(r, s) = 1$ and $r + s$ odd. If we now add back in any common factors we get the general solution

$$\boxed{x = 2grs,\ y = g(r^2 - s^2), \text{ and } z = g(r^2 + s^2).} \tag{6.1}$$

There is also a nice geometric proof of this parametrization:

`Exercise 6.1.2.` Prove that the integer solutions to $x^2 + y^2 = z^2$ with $z \neq 0$ and $(x, y, z) = 1$ are in 1-to-1 correspondence with the rational solutions u, v to $u^2 + v^2 = 1$.

Where else does a line going though $(1, 0)$ intersect the circle $x^2 + y^2 = 1$? Unless the line is vertical it will hit the unit circle in exactly one other point, which we will denote (u, v). Note that $u < 1$. If the line has slope t then $t = v/(u - 1)$ is rational if u and v are.

[10] Translations of ancient Greek texts into Latin helped inspire the Renaissance.

In the other direction, the line through $(1,0)$ of slope t is $y = t(x-1)$ which intersects $x^2+y^2=1$ where $1-x^2 = y^2 = t^2(x-1)^2$ so that either $x=1$ or $1+x = t^2(1-x)$. Hence

$$u = \frac{t^2-1}{t^2+1} \text{ and } v = \frac{-2t}{t^2+1}$$

are both rational if t is. We have therefore proved that $u, v \in \mathbb{Q}$ if and only if $t \in \mathbb{Q}$. Writing $t = -r/s$ where $(r,s)=1$ we have $u = \frac{r^2-s^2}{r^2+s^2}$ and $v = \frac{2rs}{r^2+s^2}$, the same parametrization to the Pythagorean equation as in (6.1) when we clear out denominators.

In around 1637, Pierre de Fermat was studying the proof of (6.1) in his copy of Bachet's translation of Diophantus's *Arithmetica*. In the margin he wrote:

> "*I have discovered a truly marvelous proof that it is impossible to separate a cube into two cubes, or a fourth power into two fourth powers, or in general, any power higher than the second into two like powers. This margin is too narrow to contain it.*" — by P. de Fermat (1637), in his copy of *Arithmetica*.

In other words, Fermat claimed that for every integer $n \geq 3$ there do not exist positive integers x, y, z for which

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

Fermat did not subsequently mention this problem or his truly marvelous proof elsewhere, and the proof has not, to date, been re-discovered, despite many efforts.

6.2. No solutions to a Diophantine equation through prime divisibility. One can sometimes show that a Diophantine equation has no non-trivial solutions by considering the divisibility of the variables by various primes. For example we will give such a proof that $\sqrt{2}$ is irrational.

Proof of Proposition 3.3 by 2*-divisibility*: Let us recall that if $\sqrt{2}$ is rational then we can write it as a/b so that $a^2 = 2b^2$. Let us suppose that (b,a) give the smallest non-zero solutions to $y^2 = 2x^2$ in non-zero integers. Now 2 divides $2b^2 = a^2$ so that $2|a$. Writing $a = 2A$, thus $b^2 = 2A^2$, and so $2|b$. Writing $b = 2B$ we obtain a solution $A^2 = 2B^2$ where A and B are half the size of a and b, contradicting minimality.

`Exercise 6.2.1.` Show that there are no non-zero integer solutions to $x^3+3y^3+9z^3=0$.

6.3. No solutions through geometric descent. We will give yet another proof of both Propositions 3.3 and 3.4 on irrationality, this time using geometric descent.

Proof of Proposition 3.3 by geometric descent: Again we may assume that $\sqrt{2} = a/b$ with a and b positive integers, where a is minimal. Hence $a^2 = 2b^2$ which gives rise to the smallest right-angle, isosceles triangle, OPQ with integer side lengths $\overline{OP} = \overline{OQ} = b$, $\overline{PQ} = a$ and angles $P\hat{O}Q = 90^o$, $P\hat{Q}O = Q\hat{P}O = 45^o$. Now mark a point R which is b units along PQ from Q and then drop a perpendicular to meet OP at the point S. Now $R\hat{P}S = Q\hat{P}O = 45^o$, and so $R\hat{S}P = 180^o - 90^o - 45^o = 45^o$ by considering the angles in the triangle RSP, and therefore this is a smaller isosceles, right-angled triangle. This implies that $\overline{RS} = \overline{PR} = a - b$. Now two sides and an angle are the same in OQS and RQS so these triangles are congruent; in particular $\overline{OS} = \overline{SR} = a - b$ and therefore $\overline{PS} = \overline{OP} - \overline{OS} = b - (a - b) = 2b - a$. Hence RSP is a smaller isosceles, right-angled triangle than OPQ with integer side lengths, giving a contradiction.

This same proof can be written: As $a^2 = 2b^2$, so $a > b > a/2$. Now

$$(2b - a)^2 = a^2 - 4ab + 2b^2 + 2b^2 = a^2 - 4ab + 2b^2 + a^2 = 2(a - b)^2.$$

However $0 < 2b - a < a$ contradicting the minimality of a.

Proof of Proposition 3.4 by geometric descent: Suppose that a is the smallest integer for which $\sqrt{d} = a/b$ with a and b positive integers. Let r be the smallest integer $\geq db/a$, so that $\frac{db}{a} + 1 > r \geq \frac{db}{a}$, and therefore $a > ra - db \geq 0$. Then

$$\begin{aligned}(ra - db)^2 &= da^2 - 2rdab + d^2b^2 + (r^2 - d)a^2 \\ &= da^2 - 2rdab + d^2b^2 + (r^2 - d)db^2 = d(rb - a)^2\end{aligned}$$

However $0 \leq ra - db < a$ contradicting the minimality of a, unless $ra - db = 0$. In this case $r^2 = d \cdot db^2/a^2 = d$.

6.4. Fermat's "infinite descent".

Theorem 6.1. *There are no solutions in non-zero integers x, y, z to*

$$x^4 + y^4 = z^2.$$

Proof. Let x, y, z give the solution in positive integers with z minimal.. We may assume that $\gcd(x, y) = 1$ else we can divide out the common factor. Here we have

$$(x^2)^2 + (y^2)^2 = z^2 \quad \text{with} \quad \gcd(x^2, y^2) = 1,$$

and so, by (6.1), there exist integers r, s with $(r, s) = 1$ and $r + s$ odd such that

$$x^2 = 2rs, \ y^2 = r^2 - s^2, \ \text{and } z = r^2 + s^2.$$

Now $s^2 + y^2 = r^2$ with y odd and $(r, s) = 1$ and so, by (6.1), there exist integers a, b with $(a, b) = 1$ and $a + b$ odd such that

$$s = 2ab, \ y = a^2 - b^2, \ \text{and } r = a^2 + b^2,$$

and so

$$x^2 = 2rs = 4ab(a^2 + b^2).$$

Now a, b and $a^2 + b^2$ are pairwise coprime integers whose product is a square so they must each be squares by exercise 3.1.19, say $a = u^2, \ b = v^2$ and $a^2 + b^2 = w^2$ for some positive integers u, v, w. Therefore

$$u^4 + v^4 = a^2 + b^2 = w^2$$

yields another solution to the original equation with

$$w \leq w^2 = a^2 + b^2 = r < r^2 + s^2 = z,$$

contradicting the minimality of z.

6.5. Fermat's Last Theorem.

Corollary 6.2. *There are no solutions in non-zero integers x, y, z to*

$$x^4 + y^4 = z^4.$$

`Exercise 6.5.1.` Prove this! (Hint: Use Theorem 6.1.)

Proposition 6.3. *If Fermat's Last Theorem is false then there exists an odd prime p and pairwise coprime non-zero integers x, y, z such that*

$$x^p + y^p + z^p = 0.$$

Hence, to prove Fermat's Last Theorem, one can restrict attention to odd prime exponents.

Proof. Suppose that $x^n + y^n = z^n$ with $x, y, z > 0$ and $n \geq 3$. If two of x, y have a common factor then it must divide the third and so we can divide out the common factor. Hence we may assume that x, y, z are pairwise coprime positive integers. Now any integer $n \geq 3$ has a factor m which is either $= 4$ or is an odd prime. Hence, if $n = dm$ then $(x^d)^m + (y^d)^m = (z^d)^m$, so we get a solution to Fermat's Last Theorem with exponent m. We can rule out $m = 4$ by Corollary 6.2. If $m = p$ is prime and we are given a solution to $a^p + b^p = c^p$ then $a^p + b^p + (-c)^p = 0$ as desired.

There is a great history of Fermat's Last Theorem, some of which we will discuss in the additional sections. For a very long time Fermat's Last Theorem was the best known and most sought after open question in number theory. It inspired the development of much great mathematics, in many different directions. For example ideal theory, as we will see in section C1.

In 1994 Andrew Wiles announced that he had finally proved Fermat's Last Theorem, from an idea of Frey and Serre involving modular forms, a subject far removed from the original. The proof is extraordinarily deep, involving some of the most profound themes in arithmetic geometry. If the whole proof were written in the leisurely style of, say, this book, it would probably take a couple of thousand pages. This could not be the proof that Fermat believed that he had – could Fermat have been correct? Could there be a short, elementary, marvelous proof still waiting to be found? Such a proof came to Lisbeth Salander in *The girl who played with fire* just as she went into the final tense moments of that novel — can truth follow fiction, as it so often does, or will this always remain a mystery?

7. Power Residues

7.1. Generating the multiplicative group of residues.

Lemma 7.1. *For any integer a, with $(a, m) = 1$, there exists an integer k, $1 \le k \le \phi(m)$ for which $a^k \equiv 1 \pmod m$.*

Proof. Each term of the sequence $1, a, a^2, a^3, \ldots$ is coprime with m by exercise 3.1.14. But then each is congruent to some element from any given reduced set of residues mod m (which has size $\phi(m)$). Therefore, by the pigeonhole principle, there exist i and j with $0 \le i < j \le \phi(m)$ for which $a^i \equiv a^j \pmod m$. Finish off the proof:

`Exercise 7.1.1.` Deduce that $a^k \equiv 1 \pmod m$ where $1 \le k = j - i \le \phi(m)$. (Hint: Let b be the inverse of $a \pmod m$ so that $b^i a^i \equiv 1 \pmod m$.)

Another proof of Corollary 3.7. If $r = a^{k-1}$ then $ar = a^k \equiv 1 \pmod m$.

Examples. Consider the geometric progression $2, 4, 8, \ldots$. The first term $\equiv 1 \pmod{13}$ is $2^{12} = 4096$. The first term $\equiv 1 \pmod{23}$ is $2^{11} = 2048$. Similarly $5^6 = 15625 \equiv 1 \pmod 7$ but $5^5 \equiv 1 \pmod{11}$. Hence we see that in some cases the power needed is as big as $\phi(p) = p - 1$, but not always.

If $a^k \equiv 1 \pmod m$, then $a^{k+j} \equiv a^j \pmod m$ for all $j \ge 0$, and so the geometric progression, modulo m, has period k. Thus if $u \equiv v \pmod k$ then $a^u \equiv a^v \pmod m$. Therefore one can easily determine the residues of powers $\pmod m$. For example, to compute $3^{1000} \pmod{13}$, first note that $3^3 \equiv 1 \pmod{13}$. Now $1000 \equiv 1 \pmod 3$, and so $3^{1000} \equiv 3^1 = 3 \pmod{13}$.

If $(a, m) = 1$ then let $\text{ord}_m(a)$, the *order* of $a \pmod m$, denote the smallest positive integer k for which $a^k \equiv 1 \pmod m$.

`Exercise 7.1.2.` Prove that $a^j \equiv a^i \pmod m$ if and only if $j \equiv i \pmod{\text{ord}_m(a)}$.

`Exercise 7.1.3.` Deduce that $1, a, a^2, \ldots, a^{\text{ord}_m(a)-1}$ are distinct $\pmod m$; and, for any j, that $a^j \equiv a^i \pmod m$, where i is the least non-negative residue of $j \pmod{\text{ord}_m(a)}$.

Lemma 7.2. *n is an integer for which $a^n \equiv 1 \pmod m$ if and only if $\text{ord}_m(a)$ divides n.*

Proof # 1. This follows immediately from exercise 7.1.2.

Proof # 2. There exist integers q and r such that $n = q \cdot \text{ord}_m(a) + r$ where $0 \le r \le \text{ord}_m(a) - 1$. Hence $a^r = a^n/(a^{\text{ord}_m(a)})^q \equiv 1/1^q \equiv 1 \pmod m$. Therefore $r = 0$ by the minimality of $\text{ord}_m(a)$, and so $\text{ord}_m(a)$ divides n as claimed.

In the other direction we have $a^n = (a^{\text{ord}_m(a)})^{n/\text{ord}_m(a)} \equiv 1 \pmod m$.

Theorem 7.3. *If p is a prime and p does not divide a then $\text{ord}_p(a)$ divides $p - 1$.*

Proof. Let $A = \{1, a, a^2, \ldots, a^{\text{ord}_p(a)-1} \pmod p\}$. For any non-zero $b \pmod p$ define the set $bA = \{ba \pmod p : a \in A\}$. In the next paragraph we will prove that for any two non-zero elements $b, b' \pmod p$, either $bA = b'A$ or $bA \cap b'A = \emptyset$, so that the bA partition the non-zero elements mod p. In other words, the residues $1, \ldots, p - 1 \pmod p$ may be partitioned into disjoint *cosets* bA, of A, each of which has size $|A|$; and therefore $|A| = \text{ord}_p(a)$ divides $p-1$.

Now if $bA \cap b'A \neq \emptyset$ then there exist $0 \leq i, j \leq \text{ord}_p(a) - 1$ such that $ba^i \equiv b'a^j \pmod p$. Therefore $b' \equiv ba^k \pmod p$ where k is the least non-negative residue of $i - j \pmod{\text{ord}_p(a)}$. Hence

$$b'a^\ell \equiv \begin{cases} ba^{k+\ell} \pmod p & \text{if } 0 \leq \ell \leq \text{ord}_p(a) - 1 - k \\ ba^{k+\ell-\text{ord}_p(a)} \pmod p & \text{if } \text{ord}_p(a) - k \leq \ell \leq \text{ord}_p(a) - 1 \end{cases}$$

We deduce that $bA = b'A$,

The beauty of this proof, from Gauss's *Disquisitiones Arithmeticae*, is that it works in any finite group, as we will see in section B4.

Fermat's "Little" Theorem. *If p is a prime and a is an integer that is not divisible by p then*

$$\boxed{p \text{ divides } a^{p-1} - 1.}$$

Proof. If $d = \text{ord}_p(a)$ then d divides $p-1$ by Theorem 7.3. Therefore

$$a^{p-1} = (a^d)^{\frac{p-1}{d}} \equiv 1^{\frac{p-1}{d}} = 1 \pmod p.$$

Exercise 7.1.4. Show that $p | a^{p-1} - 1$ whenever $p \nmid a$ if and only if p divides $a^p - a$ for every integer a.

Euler's 1741 Proof: We shall show that $a^p - a$ is divisible by p for every integer $a \geq 0$ that is not divisible by p, then we divide through by a to deduce the above result. We proceed by induction on a: For $a = 1$ we have $1^{p-1} - 1 = 0$, and so the result is trivial. Otherwise

$$(a+1)^p - a^p - 1 = \sum_{i=1}^{p-1} \binom{p}{i} a^i \equiv 0 \pmod p,$$

as p divides the numerator but not the denominator of $\binom{p}{i}$ for each $i, 1 \leq i \leq p-1$, so that

$$(a+1)^p - (a+1) \equiv (a^p + 1) - (a+1) \equiv a^p - a \equiv 0 \pmod p$$

by the induction hypothesis.

Combinatorial proof. The numerator of the multinomial coefficient $\binom{p}{a,b,c,\dots}$ is divisible by p, by not the denominator, unless all but one of $a, b, c, \dots$ equals 0 and the other p, in which case the multinomial coefficient equals 1. Therefore

$$(a + b + c + \dots)^p \equiv a^p + b^p + c^p + \dots \pmod p.$$

Taking $a = b = c = \dots = 1$ gives $k^p \equiv k \pmod p$ for all k.

Another proof of Theorem 7.3. The last two proofs of Fermat's Little Theorem do not use Theorem 7.3, so we have proved that $a^{p-1} \equiv 1 \pmod p$ independent of Theorem 7.3. But then Theorem 7.3 follows from Fermat's Little Theorem and Lemma 7.2.

"Sets of reduced residues" proof. In exercise 3.3.3 we saw that $\{a \cdot 1, a \cdot 2, \dots, a \cdot (p-1)\}$ form a reduced set of residues. The residues of these integers mod p, are therefore the same as the residues of $\{1, 2, \dots, p-1\}$ although in a different order. However since the two sets are the same mod p, the product of the elements of each set are equal mod p, and so

$$(a \cdot 1)(a \cdot 2) \dots (a \cdot (p-1)) \equiv 1 \cdot 2 \cdots (p-1) \pmod p.$$

Therefore

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

and, as $(p, (p-1)!) = 1$, we can divide the $(p-1)!$ out from both sides to obtain the desired

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

`Exercise 7.1.5.` The argument in this last proof works for any symmetric function of the elements of two given sets of reduced residues. Use this to show that for any integer $k \geq 1$ and any a which is not divisible by p we have

$$\text{Either } a^k \equiv 1 \pmod p, \text{ or } 1^k + 2^k + \dots + (p-1)^k \equiv 0 \pmod p.$$

Deduce that $1^k + 2^k + \dots + (p-1)^k \equiv 0$ or $p-1 \pmod p$.

Let us return to the problem of determining large powers in modular arithmetic, for example $2^{1000001} \pmod{31}$. Now $2^{30} \equiv 1 \pmod{31}$ by Fermat's Little Theorem, and so, as $1000001 \equiv 11 \pmod{30}$, we obtain $2^{1000001} \equiv 2^{11} \pmod{31}$ and it remains to do the final calculation. On the other hand, it is not hard to show that $\text{ord}_{31}(2) = 5$, so that $2^5 \equiv 1 \pmod{31}$ and, as $1000001 \equiv 1 \pmod 5$ we obtain $2^{1000001} \equiv 2^1 \equiv 2 \pmod{31}$. We see that using the order makes this calculation significantly easier.

It is worth stating the converse to Fermat's Little Theorem:

Corollary 7.4. *If $(a, n) = 1$ and $a^{n-1} \not\equiv 1 \pmod n$ then n is composite.*

For example $(2, 15) = 1$ and $2^4 = 16 \equiv 1 \pmod{15}$ so that $2^{14} \equiv 2^2 \equiv 4 \pmod{15}$. Hence 15 is a composite number. The surprise here is that we have proved that 15 is composite without having to factor 15. Indeed whenever the Corollary is applicable we will not have to factor n to show that it is composite. This is important because we do not know a fast way to factor an arbitrary integer n, but one can compute rapidly with this Corollary. We will discuss such compositeness tests in section 7.6.

Theorem 7.3 generalizes easily to: *For any $m > 1$ if $(a, m) = 1$ then* $\text{ord}_m(a)$ *divides* $\phi(m)$ by the analogous proof; and hence we can deduce, in the same manner as the first proof above:

Euler's Theorem. *For any $m > 1$ if $(a, m) = 1$ then $a^{\phi(m)} \equiv 1 \pmod m$.*

And this generalizes even further, to any finite group, as we will discuss section B4.

`Exercise 7.1.6.` Determine the last two decimal digits of 3^{8643}.

7.2. Special primes and orders. We now look at prime divisors of the Mersenne and Fermat numbers using our results on orders.

`Exercise 7.2.1.` Show that if p is prime, and q is a prime dividing $2^p - 1$, then $\mathrm{ord}_q(2) = p$.

Hence if q divides $2^p - 1$ then p divides $q - 1$ by Theorem 7.3.

Proof # 7 that there are infinitely many primes. If p is the largest prime, and q is a prime factor of $2^p - 1$, then we have just seen that p divides $q - 1$ so that $p \leq q - 1 < q$ which contradicts the hypothesis that p is the largest prime.

`Exercise 7.2.2.` Show that if prime p divides $F_n = 2^{2^n} + 1$ then $\mathrm{ord}_p(2) = 2^{n+1}$. Deduce that $p \equiv 1 \pmod{2^{n+1}}$.

Theorem 7.5. *Fix $k \geq 2$. There are infinitely many primes $\equiv 1 \pmod{2^k}$.*

Proof. Let p_n be a prime factor of $F_n = 2^{2^n} + 1$. We saw that these are all distinct in section 5.1. By exercise 7.2.2 we see that $p_n \equiv 1 \pmod{2^k}$ for all $n \geq k - 1$.

7.3. Further observations. We begin with another Theorem due to Lagrange which is proved in section A3.

Lagrange's Theorem. *Let $f(x)$ be a polynomial mod p of degree $d \geq 1$ (that is, p does not divide the coefficient of x^d in f). There are no more than d distinct roots $m \pmod{p}$ of $f(m) \equiv 0 \pmod{p}$.*

Corollary 7.6. *If p is an odd prime then there are exactly two square roots of $1 \pmod{p}$, namely 1 and -1.*

Proof # 1. As $1^2 = (-1)^2 = 1$, both 1 and -1 are roots of $x^2 - 1$, not only over $\mathbb{C}$, but also mod m for any m. Now 1 and -1 are distinct mod m if $m > 2$. Moreover there are no more than two roots of $x^2 - 1 \equiv 0 \pmod{p}$ when $m = p$ is prime, by Lagrange's Theorem. Combining these two facts gives the result.

There can be more than two square roots of 1 if the modulus is composite. For example $1, 3, 5$ and 7 are all roots of $x^2 \equiv 1 \pmod 8$; $1, 4, -4$ and -1 are all roots of $x^2 \equiv 1 \pmod{15}$; and $\pm 1, \pm 29, \pm 34, \pm 41$ are all square roots of $1 \pmod{105}$.

Proof # 2. If $x^2 \equiv 1 \pmod{p}$ then $p | (x^2 - 1) = (x-1)(x+1)$ and so p divides either $x - 1$ or $x + 1$ by Theorem 3.1. Hence $x \equiv 1$ or $-1 \pmod{p}$.

Fermat's Little Theorem tells us that $1, 2, 3, \ldots, p-1$ are $p - 1$ distinct roots of $x^{p-1} - 1 \pmod{p}$, and are therefore all the roots, by Lagrange's Theorem. Therefore the polynomials $x^{p-1} - 1$ and $(x-1)(x-2)\ldots(x-(p-1))$ mod p are the same up to a multiplicative constant. But since they are both *monic*,[11] they must be identical; that is

$$x^{p-1} - 1 \equiv (x-1)(x-2)\ldots(x-(p-1)) \pmod{p}, \tag{7.1}$$

or

$$x^p - x \equiv x(x-1)(x-2)\ldots(x-(p-1)) \pmod{p}.$$

[11] A polynomial $\sum_{i=0}^{d} c_i x^i$ with leading coefficient $c_d \neq 0$ is *monic* if $c_d = 1$.

Wilson's Theorem. *For any prime p we have $(p-1)! \equiv -1 \pmod p$.*

Proof. Take $x = 0$ in (7.1), and note that $(-1)^{p-1} \equiv 1 \pmod p$, even for $p = 2$.

Gauss's proof of Wilson's theorem. Let S be the set of pairs (a, b) for which $1 \le a < b < p$ and $ab \equiv 1 \pmod p$; that is, every residue is paired up with its inverse unless it equals its inverse. Now if $a \equiv a^{-1} \pmod p$ then $a^2 \equiv 1 \pmod p$, in which case $a \equiv 1$ or $p-1 \pmod p$ by Corollary 7.6. Therefore

$$1 \cdot 2 \cdots (p-1) = 1 \cdot (p-1) \cdot \prod_{(a,b)\in S} ab \equiv 1 \cdot (-1) \cdot \prod_{(a,b)\in S} 1 \equiv -1 \pmod p.$$

Example: For $p = 13$ we have

$$12! = 12(2 \times 7)(3 \times 9)(4 \times 10)(5 \times 8)(6 \times 11) \equiv -1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \equiv -1 \pmod{13}$$

`Exercise 7.3.1.` Show that $n \ge 2$ is prime if and only if n divides $(n-1)! + 1$. (Hint: Show that if $a|n$ then $(n-1)! + 1 \equiv 1 \pmod a$, and so deduce the result for composite n.)

`Exercise 7.3.2.` Show that if $n > 2$ is composite then n divides $(n-1)!$.

Combining Wilson's Theorem with the last exercise we have an indirect primality test for integers $n > 2$: Compute $(n-1)! \pmod n$. If it is $\equiv -1 \pmod n$ then n is prime; if it is $\equiv 0 \pmod n$ then n is composite. Note however that in determining $(n-1)!$ we need to do $n-2$ multiplications, so that this primality test takes far more steps than trial division!

`Exercise 7.3.3.` Show that the number of primes up to N equals, exactly,

$$\sum_{2\le n\le N} \frac{n}{n-1} \cdot \left\{ \frac{(n-1)!}{n} \right\} - \frac{2}{3}.$$

Compare this with the formulae at the end of section 5.

`Exercise 7.3.4.` Use the idea in the proof of Wilson's Theorem to show that

$$\prod_{\substack{1\le a\le n\\ (a,n)=1}} a \equiv \prod_{\substack{1\le b\le n\\ b^2\equiv 1 \pmod n}} b \pmod n.$$

`Exercise 7.3.5.` Determine the product of the square roots of $1 \pmod n$. One idea to begin is to multiply the square root b with another square root $n-b$. One can also try to pair the square roots in some other way.

7.4. The number of elements of a given order. In Theorem 7.3 we saw that the order modulo p of any integer a (which is coprime to p) divides $p-1$.

Example: For $p = 19$ we have

Order (mod 19)	a (mod 19)
1	1
2	18
3	7, 11
6	8, 12
9	4, 5, 6, 9, 16, 17
18	2, 3, 10, 13, 14, 15

Theorem 7.7. *If m divides $p-1$ then there are exactly $\phi(m)$ elements $a \pmod p$ of order m. If m does not divide $p-1$ then there are no elements $\pmod p$ of order m.*

A *primitive root* a mod p is an element of order $p-1$, so that $\{1, a, a^2, \dots, a^{p-2}\}$ is a reduced set of residues mod p. For example, $2, 3, 10, 13, 14, 15$ are the primitive roots mod 19. We can verify that the powers of 3 mod 19 are the reduced set of residues:

$$1, 3, 9, 27 \equiv 8, 5, -4, 7, 2, 6, -1, -3, -9, -8, -5, 4, -7, -2, -6, 1, \dots \pmod{19}$$

Taking $d = p-1$ in Theorem 7.7 we obtain.

Corollary 7.8. *For every prime p there exists a primitive root mod p. In fact there are $\phi(p-1)$ distinct primitive roots mod p.*

Proof of Theorem 7.7. By induction on m dividing $p-1$. The only element of order 1 is 1 $\pmod p$. Therefore we assume that $m > 1$ and $\psi(d) := \#\{1 \le a \le p-1 : \ \mathrm{ord}_p(a) = d\}$ equals $\phi(d)$ for all $d < m$ that divide $p-1$.

We saw in (7.1) that

$$x^{p-1} - 1 = (x^m - 1)(x^{p-1-m} + x^{p-1-2m} + \ldots + x^{2m} + x^m + 1)$$

factors into distinct linear factors mod p, and so $x^m - 1$ does also. By Lemma 7.2 we know that the set of roots of $x^m - 1 \pmod p$ is precisely the union of the sets of elements of order d, over each d dividing m. Therefore the number of roots of $x^m - 1 \pmod p$ is

$$m = \sum_{d|m} \psi(d) = \psi(m) + \sum_{\substack{d|m \\ d<m}} \psi(d) = \psi(m) + \sum_{\substack{d|m \\ d<m}} \phi(d) = \psi(m) + m - \phi(m)$$

by the induction hypothesis and Proposition 4.3. The result follows.

Although there are many primitive roots mod p it is not obvious how to always find one rapidly. However in special cases this is not difficult:

`Exercise 7.4.1.` Show that if $p = 2q + 1$ where p and q are primes with $p \equiv 3 \pmod 8$ then 2 is a primitive root mod p.

It is believed that 2 is a primitive root mod p for infinitely many primes p though this remains an open question. In fact it is conjectured that every prime q is a primitive root mod p for infinitely many primes p, and it is known that this is true for all, but at most two, primes.

Corollary 7.9. *For every prime p we have*

$$1^k + 2^k + \ldots + (p-1)^k \equiv \begin{cases} 0 & \text{if } p-1 \nmid k \\ -1 & \text{if } p-1 \mid k \end{cases} \pmod p.$$

Proof. Let a be a primitive root in exercise 7.1.5 so that $a^k \not\equiv 1 \pmod p$ when $p-1 \nmid k$. If $p-1$ divides k then each $j^k \equiv 1 \pmod p$ and the result follows.

`Exercise 7.4.2.` Write each reduced residue mod p as a power of the primitive root a, and use this to evaluate $1^k + 2^k + \ldots + (p-1)^k \pmod p$ directly, so as to give another proof of Corollary 7.9.

7.5. Primitive roots, indices and orders. If a is a primitive root $\pmod p$ then the least residues of the powers $1, a, a^2, a^3, \ldots, a^{p-2} \pmod p$ are distinct and so must equal $1, 2, \ldots, p-1$. Thus any number, not divisible by p, is congruent to some power of a. This property is extremely useful for it allows us to treat multiplication as addition of exponents in the same way that the introduction of logarithms simplifies usual multiplication.

So if $b \equiv a^e \pmod p$ where a is a primitive root mod p, then e is the *index* or *discrete logarithm* of b in base a, denoted $\mathrm{ind}_p(b)$. Note that its value is only determined mod $p-1$. It is a challenging open problem to determine the discrete logarithm of a given residue in a given base in a short amount of time.

`Exercise 7.5.1.` Show that $\mathrm{ind}_p(a \pm b) \equiv \mathrm{ind}_p(a) \pm \mathrm{ind}_p(b) \pmod{p-1}$. Deduce that $\mathrm{ind}_p(a^n) \equiv n\, \mathrm{ind}_p(a) \pmod{p-1}$.

`Exercise 7.5.2.` Show that $\mathrm{ind}_p(1) = 0$ and $\mathrm{ind}_p(-1) = (p-1)/2$, irrespective of the base used.

`Exercise 7.5.3.` Show that if m divides $p-1$ then a is mth power mod p if and only if m divides $\mathrm{ind}_p(a)$.

We also give an important practical way to recognize primitive roots mod p:

Corollary 7.10. *Suppose that p is a prime that does not divide integer a. Then a is not a primitive root* $\pmod p$ *if and only if there exists a prime q dividing $p-1$, such that*

$$a^{(p-1)/q} \equiv 1 \pmod p.$$

Proof. By definition a is not a primitive root $\pmod p$ if and only if $m := \mathrm{ord}_p(a) < p-1$. If so then let q be a prime factor of $(p-1)/m$, so that m divides $(p-1)/q$, and therefore $a^{(p-1)/q} \equiv 1 \pmod p$ by Lemma 7.2. On the other hand if $a^{(p-1)/q} \equiv 1 \pmod p$ then m divides $(p-1)/q$ by Lemma 7.2; in particular, $m \le (p-1)/q < p-1$.

Define *Carmichael's λ-function* $\lambda(m)$ to be the maximal order of an element a mod m for which $(a, m) = 1$. In fact $\lambda(p^e) = \phi(p^e)$ for all odd prime powers p^e as well as for $p^e = 2$ or 4, and $\lambda(2^e) = 2^{e-2}$ for all $e \ge 3$.

A *primitive root* mod m, is a residue $g \pmod m$ whose powers generate all of the $\phi(m)$ reduced residues mod m.

`Exercise 7.5.4.` Use Euler's Theorem and Lemma 7.2 to prove that $\lambda(m)$ divides $\phi(m)$. Prove also that there is a primitive root mod m if and only if $\lambda(m) = \phi(m)$.

Proposition 7.11. $\lambda(m) = \text{lcm}[\lambda(p^e):\ p^e\|m]$.

Proof. Let r and s be coprime integers. Suppose that a has order $\lambda(r)$ mod r, and b has order $\lambda(s)$ mod s. Select $n \equiv a \pmod r$ and $\equiv b \pmod s$ by the Chinese Remainder Theorem. If k is the order of $n \pmod{rs}$ then $a^k \equiv n^k \equiv 1 \pmod r$ so that $\lambda(r)|k$, and $b^k \equiv n^k \equiv 1 \pmod s$ so that $\lambda(s)|k$, and therefore $L|k$ where $L := \text{lcm}[\lambda(r), \lambda(s)]$. On the other hand for any m with $(m, rs) = 1$ we have $m^L = (m^{\lambda(r)})^{L/\lambda(r)} \equiv 1 \pmod r$ and similarly $m^L = (m^{\lambda(s)})^{L/\lambda(s)} \equiv 1 \pmod s$, so that $m^L \equiv 1 \pmod{rs}$. Hence we have proved that if $(r, s) = 1$ then $\lambda(rs) = L = \text{lcm}[\lambda(r), \lambda(s)]$. The result follows by induction on the number of distinct prime factors of m.

`Exercise 7.5.5.` Prove that if $pq|m$, where $p < q$ are odd primes, then $\lambda(m) < \phi(m)$. (Hint: Consider the power of 2 dividing $\lambda(m)$.)

`Exercise 7.5.6.` Prove that if $4p|m$, where p is an odd prime, then $\lambda(m) < \phi(m)$.

`Exercise 7.5.7.` Deduce that if m has a primitive root then m is a prime power or twice a prime power.

Lemma 7.12. *Let p be an odd prime. If $a = b + p^k m$ where $k \geq 1$ and $p \nmid bm$ then $a^p = b^p + p^{k+1}M$ for some integer M that is not divisible by p*

Proof. Using the binomial theorem we have

$$a^p \ = \ (b+p^k m)^p \ = \ b^p + pb^{p-1}p^k m + p\ \frac{p-1}{2}\ (p^k m)^2 + \ldots \equiv b^p + p^{k+1}b^{p-1}m \pmod{p^{k+2}},$$

and the result follows.

Theorem 7.13. *There is a primitive root mod m if and only if $m = 2$, or 4, or p^k or $2p^k$ where p is an odd prime.*

Proof. Let g be a primitive root mod p so that $\text{ord}_p(g) = p-1$. Hence $g^{p-1} = 1 + p\ell$ for some integer ℓ. If $p \nmid \ell$ then let $a = g$, else let $a = g + p$ and then

$$a^{p-1} \ = \ (g+p)^{p-1} \ = \ g^{p-1} + (p-1)g^{p-2}p + \binom{p-1}{2} g^{p-3}p^2 + \ldots \equiv 1 - pg^{p-2} \pmod{p^2}.$$

Either way $a^{p-1} = 1 + pm_0$ where $p \nmid m_0$. We now apply Lemma 7.12 and, by induction on $k \geq 0$, deduce that

$$a^{p^{k-1}(p-1)} \ = \ 1 + p^k m_k \ \text{ where } \ p \nmid m_k.$$

Now let m be the order of a mod p^k. Since $a^m \equiv 1 \pmod{p^k}$ therefore $g^m \equiv a^m \equiv 1 \pmod p$ and so $p-1$ divides m by Lemma 7.2. We also have that m divides $p^{k-1}(p-1)$ by Lemma 7.2 and Euler's Theorem, so that $m = p^j(p-1)$, for some j, $0 \leq j \leq k-1$. However $a^{p^{k-2}(p-1)} \ = \ 1 + p^{k-1}m_{k-1} \not\equiv 1 \pmod{p^k}$ and so $j = k-1$; that is a has order $p^{k-1}(p-1)$ mod p^k and so is a primitive root mod p^k.

`Exercise 7.5.7.` Complete the proof of the Theorem.

7.6. Testing for composites, pseudoprimes and Carmichael numbers. In the converse to Fermat's Little Theorem, Corollary 7.4, we saw that if integer n *does not divide* $a^{n-1}-1$ for some integer a coprime to n, then n is composite. For example, taking $a=2$ we calculate that

$$2^{1000} \equiv 562 \pmod{1001},$$

so we know that 1001 is composite. We might ask whether this always works. In other words,

Is it true that *if* n *is composite then* n *does not divide* 2^n-2?

For, if so, we have a very nice way to distinguish primes from composites. Unfortunately the answer is "no" since, for example,

$$2^{340} \equiv 1 \pmod{341},$$

but $341 = 11 \times 31$. We call 341 a *base-2 pseudoprime*. Note though that

$$3^{340} \equiv 56 \pmod{341},$$

and so the converse to Fermat's Little Theorem, with $a=3$, implies that 341 is composite.

So then we might ask whether there is always *some* value of a that helps us prove that a given composite n is indeed composite, via the converse to Fermat's Little Theorem. In other words, we are asking whether or not there are any *Carmichael numbers*, composite numbers n for which $a^{n-1} \equiv 1 \pmod{n}$ for all integers a coprime to n; one can think of these as composite numbers that "masquerade" as primes.

`Exercise 7.6.1.` Show that if $\lambda(n)$ divides $n-1$ then n is a Carmichael number.

There are indeed Carmichael numbers, the smallest of which is $561 = 3 \cdot 11 \cdot 17$, and this can be proved to be a Carmichael number since $\lambda(561) = [2, 10, 16] = 80$ which divides 560. The next few Carmichael numbers are $1105 = 5 \cdot 13 \cdot 17$, then $1729 = 7 \cdot 13 \cdot 19$, etc. Carmichael numbers are a nuisance, masquerading as primes like this, though computationally they only appear rarely. Unfortunately it was recently proved that there are infinitely many of them, and that when we go out far enough they are not so rare as it first appears. Here is an elegant way to recognize Carmichael numbers:

Lemma 7.14. *n is a Carmichael number if and only if n is squarefree, and $p-1$ divides $n-1$ for every prime p dividing n.*

Proof. Suppose that n is a Carmichael number. If prime p divides n then $a^{n-1} \equiv 1 \pmod{p}$ for all integers a coprime to n. If a is a primitive root mod p then $p-1 = \mathrm{ord}_p(a)$ divides $n-1$ by Lemma 7.2. If $p^2|n$ then let a be a primitive root mod p^2, so that $p(p-1) = \mathrm{ord}_{p^2}(a)$ divides $n-1$. However this implies that p divides $n-1$, as well as n, and hence their difference, 1, which is impossible. Therefore n must be squarefree.

In the other direction if $(a,n)=1$ and prime p divides n, then $\mathrm{ord}_p(a)|p-1$ by Theorem 7.3 which divides $n-1$, and so $a^{n-1} \equiv 1 \pmod{p}$ by Lemma 7.2. Therefore $a^{n-1} \equiv 1 \pmod{n}$ by the Chinese Remainder Theorem.

`Exercise 7.6.2.` Show that composite n is a Carmichael number if and only if n divides $a^n - a$ for all integers a.

`Exercise 7.6.3.` Show that if p is prime then the Mersenne number $2^p - 1$ is either a prime or a base-2 pseudoprime.

7.7. Divisibility tests, again. In section 2.2 we found simple tests for the divisibility of integers by $7, 9, 11$ and 13, promising to return to this theme later. The key to these earlier tests was that $10 \equiv 1 \pmod 9$ and $10^3 \equiv -1 \pmod{7 \cdot 11 \cdot 13}$; that is $\mathrm{ord}_9(10) = 1$ and $\mathrm{ord}_7(10) = \mathrm{ord}_{11}(10) = \mathrm{ord}_{13}(10) = 6$. For all primes $p \neq 2$ or 5 we know that $k := \mathrm{ord}_p(10)$ is an integer dividing $p - 1$. Hence

$$n = \sum_{j=0}^{d} n_j 10^j \equiv \sum_{m \geq 0} \sum_{i=0}^{k-1} n_{km+i} 10^i \pmod p,$$

since if $j = km + i$ then $10^j \equiv 10^i \pmod p$. In the displayed equation we have cut up integer n, written in decimal, into blocks of digits of length k and add these blocks together, which is clearly an efficient way to test for divisibility. The length of these blocks, k, is always $\leq p - 1$ no matter what the size of n.

If $k = 2\ell$ is even we can do a little better (as we did with $p = 7, 11$ and 13), namely that

$$n = \sum_{j=0}^{d} n_j 10^j \equiv \sum_{m \geq 0} \left(\sum_{i=0}^{\ell-1} n_{km+i} 10^i - \sum_{i=0}^{\ell-1} n_{km+\ell+i} 10^i \right) \pmod p,$$

thus breaking n up into blocks of length $\ell = k/2$.

7.8. The decimal expansion of fractions. The fraction $\frac{1}{3} = .3333\ldots$ is given by a recurring digit 3, so we write it as $.\overline{3}$. More interesting to us are the set of fractions

$$\frac{1}{7} = .\overline{142857}, \quad \frac{2}{7} = .\overline{285714}, \quad \frac{3}{7} = .\overline{428571}, \quad \frac{4}{7} = .\overline{571428}, \quad \frac{5}{7} = .\overline{714285}, \quad \frac{6}{7} = .\overline{857142}.$$

Notice that the decimal expansions of the six fractions $\frac{a}{7}$, $1 \leq a \leq 6$, are each periodic of period length 6, and each contain the same six digits in the same order but starting at a different place. Starting with the period for $1/7$ we find that we go through the fractions $a/7$ with $a = 1, 3, 2, 6, 4, 5$ when we rotate the period one step at a time. Do you recognize this sequence of numbers? These are the least positive residues of $10^0, 10^1, 10^2, 10^3, 10^4, 10^5 \pmod 7$. To prove this, note that

$$\frac{10^6}{7} = 142857.\overline{142857}, \text{ so that } \frac{10^6 - 1}{7} = \frac{10^6}{7} - \frac{1}{7} = 142857.$$

That is $10^6 \equiv 1 \pmod 7$ and the period 142857 is the quotient. What happens when we multiply $1/7$ through by 10^k? For example, if $k = 4$ then

$$\frac{10^4}{7} = 1428.\overline{571428} = 1428 + \frac{4}{7};$$

The part after the decimal point is always $\{\frac{10^k}{7}\}$ which equals $\frac{\ell}{7}$ where ℓ is the least positive residue of $10^k \pmod 7$. We can now prove two results.

Proposition 7.15. *Suppose that p is an odd prime, $p \neq 5$. If $1 \leq a \leq p-1$ then the decimal expansion of the period for a/p is periodic with period of length* $\text{ord}_p(10)$.

Proof. If $a/p = .\overline{m}$ where m has length n, then $10^n a/p = m.\overline{m}$, so that $(10^n - 1)a/p = m$. That is $p|a(10^n - 1)$ and so $p|(10^n - 1)$ which implies that $\text{ord}_p(10)|n$. On the other hand if $10^n \equiv 1 \pmod p$ then $(10^n - 1)a/p = m$ for some integer m. Dividing through by 10^n, then 10^{2n}, then 10^{3n}, etc. and adding, we obtain that $a/p = .\overline{m}$.

Theorem 7.16. *Suppose that p is an odd prime for which 10 is a primitive root. If m is the periodic part of $1/p$, and if a is the least residue of $10^k \pmod p$, then a/p has periodic part m_k, which is given by taking m, removing the leading k digits and concatenating them on to the end.*

`Exercise 7.8.1.` Prove this!

7.9. Primes in arithmetic progressions, revisited. We can use the ideas in this section to prove that there are infinitely many primes in certain arithmetic progressions 1 (mod m).

Theorem 7.17. *There are infinitely many primes* $\equiv 1 \pmod 3$.

Proof. Suppose that there are finitely many primes $\equiv 1 \pmod 3$, say $p_1, p_2, \dots, p_k$. Let $a = 3p_1 p_2 \cdots p_k$, and q be a prime dividing a^2+a+1. Now $q \neq 3$ as $a^2+a+1 \equiv 1 \pmod 3$. Moreover q divides $a^3-1 = (a-1)(a^2+a+1)$, but not $a-1$ (else $0 \equiv a^2+a+1 \equiv 1+1+1 \equiv 3 \pmod q$ but $q \neq 3$). Therefore $\text{ord}_q(a) = 3$ and so $q \equiv 1 \pmod 3$ by Theorem 7.3. Hence $q = p_j$ for some j, so that q divides a and thus $(a^2 + a + 1) - a(a + 1) = 1$, which is impossible.

`Exercise 7.9.1.` Generalize this argument to primes that are 1 (mod 4), 1 (mod 5), 1 (mod 6), etc. Can you prove that there are infinitely many primes $\equiv 1 \pmod m$ for arbitrary m?

In order to generalize this argument to primes $\equiv 1 \pmod m$, we need to replace the polynomial $a^2 + a + 1$ by one that recognizes when a has order m. Evidently this must be a divisor of the polynomial $a^m - 1$, indeed $a^m - 1$ divided through by all of the factors corresponding to orders which are proper divisors of m. So define the *cyclotomic polynomials* $\phi_n(t) \in \mathbb{Z}[t]$, inductively, by the requirement

$$t^m - 1 = \prod_{d|m} \phi_d(t) \qquad \text{for all } m \geq 1,$$

with each $\phi_d(t)$ monic. Therefore

$$\phi_1(t) = t-1,\ \phi_2(t) = t+1,\ \phi_3(t) = t^2+t+1,\ \phi_4(t) = t^2+1,\ \phi_5(t) = t^4+t^3+t^2+t+1, \dots$$

`Exercise 7.9.2.` Prove that $\phi_m(t)$ has degree $\phi(m)$. (Hint: Use the definition together with Proposition 4.3.)

We will discuss cyclotomic polynomials in detail at the end of section A3.

8. Quadratic residues

We are interested in understanding the squares mod m; that is the residues $a \pmod m$ for which there exists $b \pmod m$ with $b^2 \equiv a \pmod m$. By the Chinese Remainder Theorem we know that a is a square mod m if and only if a is a square modulo every prime power factor of m, so it suffices to study only the case where m is a prime power. We begin by considering only $m = p$ an odd prime.

8.1. Squares mod p. We define those non-zero residues $a \pmod p$ congruent to a square modulo p to be "*quadratic residues* $\pmod p$". All other numbers are "*quadratic non-residues*". If there is no ambiguity we simply say "residues" and "non-residues". Note that 0 is always a square mod p (as $0^2 \equiv 0 \pmod p$). Examples:

Modulus	Quadratic residues
5	1, 4
7	1, 2, 4
11	1, 3, 4, 5, 9
13	1, 3, 4, 9, 10, 12
17	1, 2, 4, 8, 9, 13, 15, 16

In each case we see that there are $\frac{p-1}{2}$ quadratic residues mod p. One sees immediately that $(p-b)^2 \equiv b^2 \pmod p$ so the distinct quadratic non-residues are $1^2, 2^2, \dots, \left(\frac{p-1}{2}\right)^2 \pmod p$.

Lemma 8.1. *The distinct quadratic residues mod p are given by* $1^2, 2^2, \dots, \left(\frac{p-1}{2}\right)^2 \pmod p$.

Proof. If $r^2 \equiv s^2 \pmod p$ where $1 \leq s < r \leq p-1$ then $p \mid r^2 - s^2 = (r-s)(r+s)$ and so $p \mid r-s$ or $p \mid r+s$. Now $-p < r-s < p$ and so if $p|r-s$ then $r = s$. Moreover $0 < r+s < 2p$ and so if $p|r+s$ then $r+s = p$. Hence the residues of $1^2, 2^2, \dots, \left(\frac{p-1}{2}\right)^2 \pmod p$ are distinct, and if $s = p-r$ then $s^2 \equiv (-r)^2 \equiv r^2 \pmod p$.

`Exercise 8.1.1.` One can write each non-zero residue mod p as a power of a primitive root. Prove that the quadratic residues are precisely those residues that have even index, and the quadratic non-residues are those that have odd index.

`Exercise 8.1.2.` Are primitive roots ever quadratic residues?

`Exercise 8.1.3.` Prove that for every $m \pmod p$ there exist a and b mod p such that $a^2+b^2 \equiv m \pmod p$. (Hint: Consider the sizes of the set of residues $a^2 \pmod p$ and the set of residues $m - b^2 \pmod p$, as a and b vary.) Deduce that there are three squares, not all divisible by p, whose sum is divisible by p.

Define the *Legendre symbol* as follows:

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{if } a \equiv 0 \pmod p \\ 1 & \text{if } a \text{ is a quadratic residue } \pmod p, \\ -1 & \text{if } a \text{ is a quadratic non-residue } \pmod p. \end{cases}$$

`Exercise 8.1.4.` Prove that if $a \equiv b \pmod p$ then $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$.

`Exercise 8.1.5.` Prove that $\sum_{a=0}^{p-1} \left(\frac{a}{p}\right) = 0$.

Corollary 8.2. *There are exactly* $1 + \left(\frac{a}{p}\right)$ *residues classes* $b \pmod p$ *for which* $b^2 \equiv a \pmod p$.

Proof. This is immediate if a is a quadratic non-residue. For $a = 0$ if $b^2 \equiv 0 \pmod p$ then $b \equiv 0 \pmod p$ so there is just one solution. If a is a quadratic residue then, by definition, there exists b such that $b^2 \equiv a \pmod p$, and then there are the two solutions $(p-b)^2 \equiv b^2 \equiv a \pmod p$ and no others, by the proof in Lemma 8.1 (or by Lagrange's Theorem).

Theorem 8.3. $\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$ *for any integers* a, b. *That is:*
i) *The product of two quadratic residues* $\pmod p$ *is a quadratic residue;*
ii) *The product of a quadratic residue and a non-residue, is itself a non-residue.*
iiI) *The product of two quadratic non-residues* $\pmod p$ *is a quadratic residue;*

Proof. (i) If $a \equiv A^2$ and $b \equiv B^2$ then $ab \equiv (AB)^2 \pmod p$.

Therefore if $(a/p) = 1$ then $\{ab : (b/p) = 1\}$ is a set of $\frac{p-1}{2}$ distinct quadratic residues mod p, and therefore all of the quadratic residues by lemma 8.1.
(ii) But then none of the elements of the set $\{ac : (c/p) = -1\}$ can be quadratic residues, and hence they are all quadratic non-residues.
(iii) In (ii) we saw that if $(c/p) = -1$ then the elements of the set $\{ca : (a/p) = 1\}$ are all quadratic non-residues, and hence give all $\frac{p-1}{2}$ distinct quadratic non-residues mod p. Therefore all of the elements cd with $(d/p) = -1$ must be quadratic residues mod p.

`Exercise 8.1.6.` Prove Theorem 8.3 by considering the parity of $\mathrm{ind}_p(a)$ and $\mathrm{ind}_p(b)$.

`Exercise 8.1.7.` What is the value of $\left(\frac{a/b}{p}\right)$? (Hint: Compare this to $\left(\frac{ab}{p}\right)$).

We deduce from the theorem that $\left(\frac{\cdot}{p}\right)$ is a multiplicative function. Therefore if we have a factorization of a into prime factors as $a = q_1^{e_1} q_2^{e_2} \dots q_k^{e_k}$ then

$$\left(\frac{a}{p}\right) = \prod_{\substack{i=1 \\ e_i \text{ odd}}}^{k} \left(\frac{q_i}{p}\right).$$

This implies that, in order to determine $\left(\frac{a}{p}\right)$ for all integers a, it is only really necessary to know the values of $\left(\frac{-1}{p}\right)$, and $\left(\frac{q}{p}\right)$ for all primes q.

8.2. Squares mod m. We show how to recognize squares modulo prime powers, in terms of the squares mod p:

Proposition 8.4. *Suppose that r is not divisible by prime p. If r is a square mod p^k then r is a square mod p^{k+1} whenever $k \geq 1$, except perhaps in the cases $p^k = 2$ or 4.*

Proof. Let x be an integer, coprime with p, such that $x^2 \equiv r \pmod{p^k}$, so that there exists an integer n for which $x^2 = r + np^k$. Therefore

$$(x - jp^k)^2 \;=\; x^2 - 2jxp^k + x^2p^{2k} \;\equiv r + (n - 2jx)p^k \pmod{p^{k+1}};$$

and this is $\equiv r \pmod{p^{k+1}}$ for $j \equiv n/2x \pmod p$ when p is odd. If $p = 2$ then

$$(x - n2^{k-1})^2 \;=\; x^2 - nx2^k + x^2 2^{2k-2} \;\equiv r \pmod{2^{k+1}},$$

provided $k \geq 3$.

`Exercise 8.2.1.` Deduce that integer r is a quadratic residue mod p^k if and only if r is a quadratic residue mod p, when p is odd, and if and only if $r \equiv 1 \pmod{\gcd(2^k, 8)}$ where $p = 2$.

Notice that this implies that exactly half of the reduced residue classes mod p^k are quadratic residues, when p is odd, and exactly one quarter when $p = 2$ and $k \geq 3$.

Using the Chinese Remainder Theorem we deduce from exercise 8.2.1 that if $(a, m) = 1$ then a is a square mod m if and only if $\left(\frac{a}{p}\right) = 1$ for every odd prime p dividing m, and $a \equiv 1 \pmod{\gcd(m, 8)}$.

8.3. The Jacobi symbol. It is useful to extend the definition of the Legendre symbol as follows: If m is odd, with $m = \prod_p p^{e_p}$ then

$$\left(\frac{a}{m}\right) = \prod_p \left(\frac{a}{p}\right)^{e_p}.$$

Observe that if a is a square mod m then $(a/p) = 1$ for all $p|m$ and so $(a/m) = 1$. However the converse is not always true: The squares mod 15 that are prime to 15 are $(\pm 1)^2 \equiv (\pm 4)^2 \equiv 1 \pmod{15}$ and $(\pm 2)^2 \equiv (\pm 7)^2 \equiv 4 \pmod{15}$. Therefore 2 is not a square mod 15 but

$$\left(\frac{2}{3}\right) = \left(\frac{2}{5}\right) = -1 \text{ so that } \left(\frac{2}{15}\right) = \left(\frac{2}{3}\right)\left(\frac{2}{5}\right) = 1.$$

`Exercise 8.3.1.` Prove that $\left(\frac{ab}{m}\right) = \left(\frac{a}{m}\right)\left(\frac{b}{m}\right)$.

`Exercise 8.3.2.` Prove that $\left(\frac{a}{m}\right) = \left(\frac{b}{m}\right)$ whenever $a \equiv b \pmod m$.

`Exercise 8.3.3.` For how many residues a mod m do we have $(a/m) = 1$?

8.4. The quadratic character of a residue. We have seen that $p-1$th power of any reduced residue mod p is congruent to 1 (mod p) but are there perhaps other patterns amongst the lower powers?

a	a^2	a^3	a^4
1	1	1	1
2	-1	-2	1
-2	-1	2	1
-1	1	-1	1

The powers of a mod 5.

a	a^2	a^3	a^4	a^5	a^6
1	1	1	1	1	1
2	-3	1	2	-3	1
3	2	-1	-3	-2	1
-3	2	1	-3	2	1
-2	-3	-1	2	3	1
-1	1	-1	1	-1	1

The powers of a mod 7.

As expected the $p-1$st column is all 1s, but one also observes that the middle column, $a^2 \pmod 5$ and $a^3 \pmod 7$, is all -1s and 1s. This column represents the least residues of numbers of the form $a^{\frac{p-1}{2}} \pmod p$. Euler showed that, not only is this always -1 or 1, but that it determines the value of the Legendre symbol:

Euler's criterion. $\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod p$, *for all primes p and integers a.*

Proof 1. If $\left(\frac{a}{p}\right) = 1$ then there exists b such that $b^2 \equiv a \pmod p$ so that $a^{\frac{p-1}{2}} \equiv b^{p-1} \equiv 1 \pmod p$, by Fermat's Little Theorem.

If $\left(\frac{a}{p}\right) = -1$ then we proceed as in the proof of Wilson's Theorem by defining

$$S = \{(r,s) : \ 1 \le r < s \le p-1, \ rs \equiv 1 \pmod p\}.$$

Each integer m, $1 \le m \le p-1$, appears in exactly one such pair, for it is paired with the least positive residue of $a/m \pmod p$, and no residue is paired with itself else $m^2 \equiv a \pmod p$ which is impossible as a is a quadratic non-residue mod p. Hence

$$(p-1)! = \prod_{(r,s)\in S} rs \equiv a^{|S|} = a^{\frac{p-1}{2}} \pmod p,$$

and the result follows from Wilson's Theorem.

`Exercise 8.4.1.` Evaluate $(p-1)! \pmod p$ as in the second part of this proof, when $(a/p) = 1$?

Proof 2: $\left(\frac{a}{p}\right) = 1$ if and only if a is a quadratic residue. We know that there are $\frac{p-1}{2}$ quadratic residues $\pmod p$ and that these are all roots of $x^{\frac{p-1}{2}} - 1 \pmod p$ (since if $a \equiv b^2 \pmod p$ then $a^{\frac{p-1}{2}} \equiv b^{p-1} \equiv 1 \pmod p$). However all of the reduced residues are roots of $x^{p-1} - 1 = (x^{\frac{p-1}{2}} - 1)(x^{\frac{p-1}{2}} + 1)$ by Fermat's Theorem. Thus the quadratic non-residues are the roots of $x^{\frac{p-1}{2}} + 1 \pmod p$, and the result follows.

Example: $\left(\frac{3}{13}\right) = 1$ since $3^6 = 27^2 \equiv 1^2 \equiv 1 \pmod{13}$, but $\left(\frac{2}{13}\right) = -1$ since $2^6 = 64 = -1 \pmod{13}$.

`Exercise 8.4.2.` Explain how one can determine $\left(\frac{a}{p}\right)$ by knowing the least residue of $a^{\frac{p-1}{2}} \pmod p$.

`Exercise 8.4.3.` Prove Euler's criterion by considering the parity of $\mathrm{ind}_p(a)$.

One of the beautiful consequences of Euler's criterion is that one can test whether a is a square mod p without determining the square root of $a \pmod p$ (which may be difficult). Taking a high power of a mod p is not difficult using the method of section A5. However when $p \equiv 3 \pmod 4$ it is easy to find the square root of $a \pmod p$:

`Exercise 8.4.4.` Let p be a prime $\equiv 3 \pmod 4$. Show that if $\left(\frac{a}{p}\right) = 1$ and $x \equiv a^{\frac{p+1}{4}} \pmod p$ then $x^2 \equiv a \pmod p$. Can one adapt this method when $p \equiv 1 \pmod 4$?

Although half of the residues mod p are quadratic non-residues we do not know how to find one quickly (and thus we do not know how to find primitive roots quickly either)

`Exercise 8.4.5.` If a is coprime to p, consider the permutation σ of the reduced residues given by the map $n \to an \pmod p$. Show that $\left(\frac{a}{p}\right) = (-1)^{T(\sigma)}$, where $T(\sigma)$ is the number of transpositions in σ. (Hint: You might use the proof of Theorem 7.3.)

Given an integer m it is easy to determine all of the quadratic residues $\pmod m$, by simply computing $a^2 \pmod m$ for each $(a, m) = 1$. However finding all primes p for which m is a quadratic residue $\pmod p$ is considerably more difficult. We start examining this question now.

8.5. The residue -1.

Theorem 8.5. *-1 is a quadratic residue $\pmod p$ if and only if $p = 2$ or $p \equiv 1 \pmod 4$.*

Proof. By Euler's criterion, and exercise 8.4.2.

Proof 2. In exercise 7.5.2 we saw that $-1 \equiv g^{(p-1)/2} \pmod p$ for any primitive root g mod p. Now if $-1 \equiv (g^k)^2 \pmod p$ for some integer k then $\frac{p-1}{2} \equiv 2k \pmod{p-1}$, and there exists such an integer k if and only if $\frac{p-1}{2}$ is even.

Proof 3. (Euler) If a is a quadratic residue then so is $1/a \pmod p$. Thus we may "pair up" the quadratic residues $\pmod p$, except those for which $a \equiv 1/a \pmod p$. The only solutions of $a \equiv 1/a \pmod p$ (that is $a^2 \equiv 1 \pmod p$) are $a \equiv 1$ and $-1 \pmod p$. Therefore

$$
\begin{aligned}
\frac{p-1}{2} &= \#\{a \pmod p : a \text{ is a quadratic residue } \pmod p\} \\
&\equiv \#\{a \in \{1,-1\} : a \text{ is a quadratic residue } \pmod p\} \pmod 2 \\
&= \begin{cases} 2 & \text{if } \left(\frac{-1}{p}\right) = 1; \\ 1 & \text{if } \left(\frac{-1}{p}\right) = -1 \end{cases} \pmod 2
\end{aligned}
$$

and the result follows.

Proof 4. The first part of the last proof also show us that the product of the quadratic residues mod p is congruent to $-(-1/p)$. On the other hand the roots of $x^{\frac{p-1}{2}} - 1 \pmod p$

are precisely the quadratic residues mod p, and so, taking $x = 0$, we see that $(-1)^{\frac{p-1}{2}}$ times the product of the quadratic residues mod p is congruent to -1. Hence $(-1/p) \equiv (-1)^{\frac{p-1}{2}} \pmod p$.

Proof 5. The number of quadratic non-residues $\pmod p$ is $\frac{p-1}{2}$, and so, by Wilson's Theorem, we have

$$\left(\frac{-1}{p}\right) = \left(\frac{(p-1)!}{p}\right) = \prod_{a \pmod p} \left(\frac{a}{p}\right) \equiv (-1)^{\frac{p-1}{2}}.$$

Theorem 8.5 implies that if $p \equiv 1 \pmod 4$ then $\left(\frac{-r}{p}\right) = \left(\frac{r}{p}\right)$; and if $p \equiv -1 \pmod 4$ then $\left(\frac{-r}{p}\right) = -\left(\frac{r}{p}\right)$.

Corollary 8.6. *If n is odd then*

$$\left(\frac{-1}{n}\right) = \begin{cases} 1 & \text{if } n \equiv 1 \pmod 4; \\ -1 & \text{if } n \equiv -1 \pmod 4. \end{cases}$$

`Exercise 8.5.1.` Prove this.

8.6. The law of quadratic reciprocity. We have already seen that if p is an odd prime then

$$\left(\frac{-1}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1 \pmod 4; \\ -1 & \text{if } p \equiv -1 \pmod 4. \end{cases}$$

In the next section we will show that

$$\left(\frac{2}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1 \text{ or } -1 \pmod 8; \\ -1 & \text{if } p \equiv 3 \text{ or } -3 \pmod 8. \end{cases}$$

To be able to evaluate Legendre symbols we will also need the *law of quadratic reciprocity*. This states that if p and q are distinct odd primes then

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = \begin{cases} -1 & \text{if } p \equiv q \equiv -1 \pmod 4 \\ 1 & \text{otherwise.} \end{cases}$$

These rules, taken together, allow us to rapidly evaluate any Legendre symbol, as follows: Suppose that we wish to evaluate (m/p). First we reduce m mod p, so that $(m/p) = (n/p)$ where $n \equiv m \pmod p$ and $|n| < p$. Next we factor n and, by the multiplicativity of the Legendre symbol, as discussed at the end of section 8.1, we can evaluate (n/p) in terms of $(-1/p), (2/p)$ and the (q/p) for those primes q dividing n. We can easily determine the values of $(-1/p)$ and $(2/p)$ from determining $p \pmod 8$, and then we need to evaluate each (q/p) where $q \le |n| < p$. We do this by the law of quadratic reciprocity so that $(q/p) = \pm(p/q)$ depending only on the values of p and q mod 4. We repeat the procedure on each (p/q). Clearly this process will quickly finish as the numbers involved are always getting smaller.

This is an efficient procedure provided that one is capable of factoring the numbers n that arise. Although this may be the case for small examples, it is not practical for large examples. We can by-pass this difficulty by using the Jacobi symbol. The three rules above hold just as well provided p and q are any two odd coprime integers. Hence to evaluate (m/p) we find $n \equiv m \pmod p$ with $|n| < p$ as above, and then write $n = qN$, where $q = \pm$ a power of 2, and N is an odd positive integer, so that $N \leq |n| < p$. Therefore $(m/p) = (n/p)$ which may be evaluated in terms of $(-1/p), (2/p)$ and the (N/p). This last equals $\pm(p/N)$ depending only on p and N mod 4, and then we repeat the procedure with (p/N). This process only involves dividing out powers of 2 and a possible minus sign, so goes fast and avoids serious factoring; in fact it is guaranteed to go at least as fast as the Euclidean algorithm since it involves very similar steps. Let us work through some examples.

$$\begin{aligned}
\left(\frac{111}{71}\right) &= \left(\frac{-1}{71}\right)\left(\frac{31}{71}\right) && \text{as } 111 \equiv -31 \pmod{71} \\
&= (-1)\cdot(-1)\cdot\left(\frac{71}{31}\right) && \text{as } 71 \equiv 31 \equiv -1 \pmod 4 \\
&= \left(\frac{9}{31}\right) \ = \ 1 && \text{as } 71 \equiv 9 \pmod{31}.
\end{aligned}$$

Next we give an alternate evaluation, without explaining each step:

$$\left(\frac{111}{71}\right) = \left(\frac{40}{71}\right) = \left(\frac{2}{71}\right)^3\left(\frac{5}{71}\right) = 1^3\cdot 1\cdot\left(\frac{71}{5}\right) = \left(\frac{1}{5}\right) = 1.$$

These examples all work with only the Legendre symbol; but here is one where we use the Jacobi symbol:

$$\left(\frac{106}{71}\right) = \left(\frac{35}{71}\right) = -\left(\frac{71}{35}\right) = -\left(\frac{1}{35}\right) = -1.$$

In the next few subsections we will prove the results used above. Our approach will not be the one mostly seen in textbooks today, but rather (a version of) the original proof of Gauss.

8.7. The residues $+2$ and -2. By computing, one finds that the odd primes $p < 100$ for which $\left(\frac{2}{p}\right) = 1$ are $p = 7, 17, 23, 31, 41, 47, 71, 73, 79, 89, 97$. These are exactly the primes < 100 that are $\equiv \pm 1 \pmod 8$. The values of $p < 100$ for which $\left(\frac{-2}{p}\right) = 1$ are $p = 3, 11, 17, 19, 41, 43, 59,\ 67, 73, 83, 89, 97$. These are exactly the primes < 100 that are $\equiv 1$ or $3 \pmod 8$. These observations are established as facts in the following result.

Theorem 8.7. *If n is odd then*

$$\left(\frac{2}{n}\right) = \begin{cases} 1 & \text{if } n \equiv 1 \text{ or } -1 \pmod 8; \\ -1 & \text{if } n \equiv 3 \text{ or } -3 \pmod 8. \end{cases}$$

Proof. By induction on n. If n is composite with $n \equiv \pm 1 \pmod 8$ then write $n = ab$ with $1 < a, b < n$. Then $b \equiv a^2 b = an \equiv \pm a \pmod 8$, as $a^2 \equiv 1 \pmod 8$, and so $\left(\frac{2}{a}\right) = \left(\frac{2}{b}\right)$ by the induction hypothesis. Hence $\left(\frac{2}{n}\right) = \left(\frac{2}{a}\right)\left(\frac{2}{b}\right) = 1$.

Proceeding similarly when $n \equiv \pm 3 \pmod 8$ we find that $\left(\frac{2}{a}\right) = -\left(\frac{2}{b}\right)$ by the induction hypothesis, and so $\left(\frac{2}{n}\right) = \left(\frac{2}{a}\right)\left(\frac{2}{b}\right) = -1$.

We may therefore assume that $n = p$ is prime. For $p \equiv 1 \pmod 4$ we have an element r such that $r^2 \equiv -1 \pmod p$ by Theorem 8.5. We now show that 2 is a square mod p if and only if r is a square mod p For if $a^2 \equiv 2 \pmod p$ then

$$((1+r)/a)^2 = (1 + r^2 + 2r)/a^2 \equiv 2r/2 \equiv r \pmod p;$$

and if $b^2 \equiv r$ then

$$((1+r)/b)^2 = (1 + r^2 + 2r)/b^2 \equiv 2r/r \equiv 2 \pmod p.$$

Now r is a square mod p if and only if there is an element of order 8 mod p, and this holds if and only $8|p-1$ by Theorem 7.7. The result thus follows when $p \equiv 1$ or $5 \pmod 8$.

We now exhibit an argument that we will later use to prove the full law of quadratic reciprocity. If the result is false for $p \equiv \pm 3 \pmod 8$ then $\left(\frac{2}{p}\right) = 1$. Select $a^2 \equiv 2 \pmod p$ with a odd and minimal, so that $1 \leq a \leq p - 1$.[12] Write $a^2 - 2 = pr$. Evidently $pr \equiv a^2 - 2 \equiv -1 \pmod 8$ and so $r \equiv \pm 3 \pmod 8$. But then $a^2 \equiv 2 \pmod r$ and so $\left(\frac{2}{r}\right) = 1$ with $r = \frac{a^2-2}{p} < p$, which contradicts the induction hypothesis.

If the result is false for $p \equiv 5$ or $7 \pmod 8$ then $\left(\frac{-2}{p}\right) = 1$. Select $a^2 \equiv -2 \pmod p$ with a minimal and odd, so that $1 \leq a \leq p-1$. Write $a^2+2 = pr$. Evidently $pr \equiv a^2+2 \equiv 3 \pmod 8$ and so $r \equiv 5$ or $7 \pmod 8$. But then $a^2 \equiv 2 \pmod r$ and so $\left(\frac{2}{r}\right) = 1$ with $r = \frac{a^2+2}{p} < p$, which contradicts the induction hypothesis.

Combining these cases gives the result for all odd primes p.

There is a well-known proof of Theorem 8.7 when $n = p$ that can be deduced from Euler's criterion. We discuss this in section C8.

Corollary 8.8. *If n is odd then*

$$\left(\frac{-2}{n}\right) = \begin{cases} 1 & \text{if } n \equiv 1 \text{ or } 3 \pmod 8; \\ -1 & \text{if } n \equiv 5 \text{ or } 7 \pmod 8. \end{cases}$$

`Exercise 8.7.1.` Deduce this from Corollary 8.6 and Theorem 8.7.

[12] If b is the smallest positive integer for which $b^2 \equiv 2 \pmod p$, so that $1 \leq b \leq p-1$, then let $a = b$ if b is odd, and $a = p - b$ if b is even.

8.8. Small residues and non-residues. 1 is always a quadratic residue mod p, as are $4, 9, 16, \ldots$ If 2 and 3 are quadratic non-residues then $2 \cdot 3 = 6$ is a quadratic residue, by Theorem 8.3(iii). Hence one is always guaranteed lots of small quadratic residues. How about quadratic non-residues? Since half the residues are quadratic non-residues one might expect to find lots of them, but a priori one is only guaranteed to find one $\leq \frac{p+1}{2}$. Can one do better? This is an important question in number theory, and one where the best results known are surprisingly weak (see section F3 for more discussion).

One amusing problem is to find strings of consecutive quadratic residues. Developing the discussion in the last paragraph prove the following:

`Exercise 8.8.1.` Prove that for every prime $p \geq 7$ there exists an integer $n = n_p \leq 9$ for which one has $\left(\frac{n}{p}\right) = \left(\frac{n+1}{p}\right) = 1$. Can you extend this result to three consecutive quadratic residues?

More complicated is to ask whether, for a given integer m, one can find small primes p and q for which

$$\left(\frac{m}{p}\right) = 1 \quad \text{and} \quad \left(\frac{m}{q}\right) = -1.$$

We shall study here just the second of the two problems (more on the first in section F3):

Theorem 8.9. *If p is a prime $\equiv 1 \pmod 4$ there exists a prime $q < p$ such that $\left(\frac{p}{q}\right) = -1$.*

Actually we will get much better bounds on q than this.

Part I. *If $p \equiv 5 \pmod 8$ then there exists a prime $q < 2(\sqrt{2p} - 1)$ with $\left(\frac{p}{q}\right) = -1$.*

Proof. Choose integer a as large as possible so that $2a^2 < p$; in particular we may choose $a > (p/2)^{1/2} - 1$. Now $p - 2a^2 \equiv 3$ or $5 \pmod 8$ and so has a prime divisor $q \equiv 3$ or $5 \pmod 8$. But then, by Theorem 8.7, we have $\left(\frac{2}{q}\right) = -1$ and so $\left(\frac{p}{q}\right) = \left(\frac{2a^2}{q}\right) = -1$. Finally

$$q \leq p - 2a^2 < 2(\sqrt{2p} - 1).$$

The next case involves a remarkable proof given by Gauss:

Part II. *If $p \equiv 1 \pmod 8$ then there exists an odd prime $q < 2\sqrt{p} + 1$ with $\left(\frac{p}{q}\right) = -1$.*

Proof. Let $m = [\sqrt{p}]$ and consider the product $(p - 1^2)(p - 2^2)\ldots(p - m^2)$, under the assumption that $\left(\frac{p}{q}\right) = 1$ for all $q \leq 2m + 1$. Now since $\left(\frac{p}{q}\right) = 1$ there exists a such that $p \equiv a^2 \pmod q$; in fact there exists a_q such that $p \equiv a_q^2 \pmod{q^n}$ for any given integer $n \geq 1$ (by the discussion in section 8.2). Since this is true for each $q \leq 2m + 1$, and since $(2m + 1)!$ is divisible only by powers of primes $q \leq 2m + 1$, we use the Chinese Remainder Theorem to construct an integer A for which $p \equiv A^2 \pmod{(2m + 1)!}$. Thus

$$\begin{aligned}(p - 1^2)(p - 2^2)\ldots(p - m^2) &\equiv (A^2 - 1^2)(A^2 - 2^2)\ldots(A^2 - m^2) \\ &\equiv \frac{(A + m)!}{(A - m - 1)!} \cdot \frac{1}{A} \qquad \pmod{(2m + 1)!}.\end{aligned}$$

Now $(p,(2m+1)!) = 1$ and so $(A,(2m+1)!) = 1$; moreover $\binom{A+m}{2m+1}$ is an integer, and so

$$\frac{(A+m)!}{(A-m-1)!}\cdot\frac{1}{A} = \frac{1}{A}\cdot(2m+1)!\binom{A+m}{2m+1} \equiv 0 \pmod{(2m+1)!}.$$

Therefore $(2m+1)!$ divides $(p-1^2)(p-2^2)\dots(p-m^2)$. However $p<(m+1)^2$ and so

$$\begin{aligned}(2m+1)! &\le (p-1^2)(p-2^2)\dots(p-m^2)\\ &< ((m+1)^2-1^2)((m+1)^2-2^2)\dots((m+1)^2-m^2) = \frac{(2m+1)!}{m+1}\end{aligned}$$

giving a contradiction.

8.9. Proof of the law of quadratic reciprocity. Gauss gave four proofs of the law of quadratic reciprocity, and there are now literally hundreds of proofs. None of the proofs are easy. For an elementary textbook like this one wishes to avoid any deeper ideas, which considerably cuts down the number of choices. The one that has been long preferred stems from an idea of Eisenstein and is discussed in section C8. It ends up with an elegant lattice point counting argument though the intermediate steps are difficult to follow and motivate. Gauss's very first proof was long and complicated yet elementary and the motivation is quite clear. Subsequent authors [Savitt] have shortened Gauss's proof and we present a version of that proof here. We will prove that for any odd integers m and n with $(m,n)=1$ we have

$$\left(\frac{m}{n}\right)\left(\frac{n}{m}\right) = \begin{cases} -1 & \text{if } m\equiv n\equiv -1 \pmod 4\\ 1 & \text{otherwise}\end{cases}$$

where we define $\left(\frac{m}{-n}\right) = \left(\frac{m}{n}\right)$. Note that we can write the right side as $(-1)^{\frac{m-1}{2}\cdot\frac{n-1}{2}}$.

We prove this by induction on $\max\{|m|,|n|\}$. It is already proved if one of m and n equals 1 or -1. If $m=ab$ is composite with $1<a,b<m$ then

$$\left(\frac{m}{n}\right)\left(\frac{n}{m}\right) = \left(\frac{a}{n}\right)\left(\frac{n}{a}\right)\cdot\left(\frac{b}{n}\right)\left(\frac{n}{b}\right) = (-1)^{\frac{a-1}{2}\cdot\frac{n-1}{2}}\cdot(-1)^{\frac{b-1}{2}\cdot\frac{n-1}{2}},$$

and the result follows since:

Exercise 8.9.1. Prove that $\frac{a-1}{2}+\frac{b-1}{2}\equiv\frac{ab-1}{2} \pmod 2$ for any odd integers a,b.

A similar proof works if n is composite. We can assume that m and n are positive for if $m<0$ then we can write $m=-b$ with $b>0$ and follow the above argument through with $a=-1$. Therefore we are left with the case that $m=p<n=q$ are primes, that is we wish to prove that

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = \begin{cases} -1 & \text{if } p\equiv q\equiv -1 \pmod 4\\ 1 & \text{otherwise.}\end{cases}$$

The proof is modeled on that of the last two cases in the proof of Theorem 8.7. There are two cases here:

• When $\left(\frac{p}{q}\right)=1$ or $\left(\frac{-p}{q}\right)=1$, let $\ell=p$ or $-p$, respectively, so that $\left(\frac{\ell}{q}\right)=1$. Then there exists an even integer $e,\ 1\leq e\leq q-1$ such that $e^2\equiv \ell \pmod q$, and therefore there exists an integer s with

$$e^2=\ell+qs.$$

Now $|s|=\left|\frac{e^2-\ell}{q}\right|<\frac{(q-1)^2+q}{q}<q$, so the reciprocity law works for the pair ℓ,s by the induction hypothesis. Observing that $e^2\equiv \ell \pmod s$ and $e^2\equiv qs \pmod \ell$ we deduce that $\left(\frac{\ell}{s}\right)=\left(\frac{qs}{\ell}\right)=1$ assuming $p=|\ell|$ does not divide s. We therefore deduce:

$$\left(\frac{\ell}{q}\right)\left(\frac{q}{\ell}\right)=1\cdot\left(\frac{q}{\ell}\right)\cdot\left(\frac{qs}{\ell}\right)\cdot\left(\frac{\ell}{s}\right)=\left(\frac{s}{\ell}\right)\cdot\left(\frac{\ell}{s}\right)=(-1)^{\frac{\ell-1}{2}\cdot\frac{s-1}{2}}$$

Now $\ell+qs=e^2\equiv 0 \pmod 4$, and the result follows as $q\equiv s \pmod 4$ if $\ell\equiv -1 \pmod 4$.

If $p|s$ we write $s=\ell S,\ e=\ell E$ to obtain $\ell E^2=1+qS$, and so $\left(\frac{\ell}{S}\right)=\left(\frac{-qS}{\ell}\right)=1$. Therefore

$$\left(\frac{\ell}{q}\right)\left(\frac{q}{\ell}\right)=\left(\frac{q}{\ell}\right)\cdot\left(\frac{-qS}{\ell}\right)\cdot\left(\frac{\ell}{S}\right)=\left(\frac{-S}{\ell}\right)\cdot\left(\frac{\ell}{-S}\right)=(-1)^{\frac{\ell-1}{2}\cdot\frac{S+1}{2}}$$

and the result follows since $S\equiv -q \pmod 4$.

• When $\left(\frac{p}{q}\right)=\left(\frac{-p}{q}\right)=-1$, we have $\left(\frac{-1}{q}\right)=-1$ so that $q\equiv 1 \pmod 4$. Therefore there exists a prime $\ell<q$ such that $\left(\frac{q}{\ell}\right)=-1$ by Theorem 8.9. Moreover $\left(\frac{\ell}{q}\right)=-1$ else, since we have already proved the reciprocity law when $\left(\frac{\ell}{q}\right)=1$, this would imply that $\left(\frac{q}{\ell}\right)=1$ as $q\equiv 1 \pmod 4$.

Therefore $\left(\frac{p\ell}{q}\right)=1$ and so there exists an even integer $e,\ 1\leq e\leq q-1$ such that $e^2\equiv p\ell \pmod q$, and therefore there exists an integer s with

$$e^2=p\ell+qs.$$

Note that $|s|=\left|\frac{e^2-p\ell}{q}\right|\leq\left|\frac{\max\{(q-1)^2,p\ell\}}{q}\right|<q$, so the reciprocity law works for any two of ℓ,p,s by the induction hypothesis.

We proceed much as above but now there are four possibilities for $d=(p\ell,qs)=(p\ell,s)$, which we handle all at once: Since d is squarefree and $d|p\ell+qs=e^2$, hence $d|e$. We write $e=dE,\ p\ell=dL$ and $s=dS$ so that $dE^2=L+qS$. But then

$$\left(\frac{-LqS}{d}\right)=\left(\frac{dqS}{L}\right)=\left(\frac{dL}{q}\right)=\left(\frac{dL}{S}\right)=1.$$

Multiplying these all together and re-organizing, and using that $p\ell=dL$, we obtain

$$\left(\frac{-L}{d}\right)\left(\frac{d}{-L}\right)\cdot\left(\frac{q}{p\ell}\right)\left(\frac{p\ell}{q}\right)\cdot\left(\frac{S}{p\ell}\right)\left(\frac{p\ell}{S}\right)=1.$$

Now $\left(\frac{q}{\ell}\right)\left(\frac{\ell}{q}\right) = 1$ by the choice of ℓ, and using the induction hypothesis for the pairs $(-L, d), (p, S), (\ell, S)$ we obtain

$$\begin{aligned}\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) &= \left(\frac{-L}{d}\right)\left(\frac{d}{-L}\right)\cdot\left(\frac{S}{p}\right)\left(\frac{p}{S}\right)\cdot\left(\frac{S}{\ell}\right)\left(\frac{\ell}{S}\right)\\ &= (-1)^{\frac{L+1}{2}\cdot\frac{d-1}{2}+\frac{p-1}{2}\cdot\frac{S-1}{2}+\frac{\ell-1}{2}\cdot\frac{S-1}{2}}.\end{aligned}$$

Now $S \equiv -qL \equiv -L \pmod 4$ and $dp\ell = d^2L \equiv L \pmod 4$, so the above exponent is $\equiv \frac{L+1}{2}\cdot\frac{dp\ell-1}{2} \equiv \frac{L+1}{2}\cdot\frac{L-1}{2} \equiv 0 \pmod 2$, and the result follows.

Another proof for $(2/n)$. By induction on n. The result is easily proved for $n = 1$. For odd $n > 1$ we have, using the law of quadratic reciprocity,

$$\left(\frac{2}{n}\right) = \left(\frac{-1}{n}\right)\left(\frac{n-2}{n}\right) = \left(\frac{-1}{n}\right)\left(\frac{n}{n-2}\right) = \left(\frac{-1}{n}\right)\left(\frac{2}{n-2}\right),$$

as one of n and $n-2$ is $\equiv 1 \pmod 4$.

`Exercise 8.9.2.` Complete the proof, which proceeds via an analysis of the four cases mod 8.

9. Quadratic equations

9.1. Sums of two squares. What primes are the sum of two squares? If we start computing we find that

$$2 = 1^2+1^2,\ 5 = 1^2+2^2,\ 13 = 2^2+3^2,\ 17 = 1^2+4^2,\ 29 = 5^2+2^2,\ 37 = 1^2+6^2,\ 41 = 5^2+4^2, \dots$$

so we might guess that the answer is 2 and any prime $\equiv 1 \pmod 4$.

Proposition 9.1. *If p is an odd prime that is the sum of two squares then $p \equiv 1 \pmod 4$.*

Proof. If $p = a^2 + b^2$ then $p \nmid a$, else $p|b$ and so $p^2|a^2 + b^2 = p$ which is impossible, and similarly $p \nmid b$. Now $a^2 \equiv -b^2 \pmod p$ so that

$$1 = \left(\frac{a}{p}\right)^2 = \left(\frac{-1}{p}\right)\left(\frac{b}{p}\right)^2 = \left(\frac{-1}{p}\right),$$

and therefore $p \equiv 1 \pmod 4$.

The other direction is more complicated

Theorem 9.2. *Any prime $p \equiv 1 \pmod 4$ can be written as the sum of two squares.*

Proof. Since $p \equiv 1 \pmod 4$ we know that there exists an integer b such that $b^2 \equiv -1 \pmod p$. Consider now the set of integers

$$\{i + jb :\ 0 \le i, j \le [\sqrt{p}]\}$$

The number of pairs i, j used in the construction of this set is $([\sqrt{p}] + 1)^2 > p$, and so by the pigeonhole principle, two must be congruent mod p; say that

$$i + jb \equiv I + Jb \pmod p$$

where $0 \le i, j, I, J \le [\sqrt{p}]$ and $\{i, j\} \ne \{I, J\}$. Let $r = i - I$ and $s = J - j$ so that

$$r \equiv bs \pmod p$$

where $|r|, |s| \le [\sqrt{p}] < \sqrt{p}$, and r and s are not both 0. Now

$$r^2 + s^2 \equiv (bs)^2 + s^2 = s^2(b^2 + 1) \equiv 0 \pmod p$$

and $0 < r^2 + s^2 < \sqrt{p}^2 + \sqrt{p}^2 = 2p$. The only multiple of p between 0 and $2p$ is p, and therefore $r^2 + s^2 = p$.

Exercise 9.1.1. Suppose that $b \pmod p$ is given, and that $R \ge 1$ and S are positive numbers such that $RS = p$. Prove that there exist integers r, s with $|r| \le R, 0 < s \le S$ such that $b \equiv r/s \pmod p$.

What integers can be written as the sum of two squares? Note the identity

$$(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2. \tag{9.1}$$

`Exercise 9.1.2.` Use this to show that the product of two or more integers that are the sum of two squares is itself the sum of two squares.

We see that (9.1) is a useful identity, yet we simply gave it without indicating how one might find such an identity. Let i be a complex number for which $i^2 = -1$. Then we have $x^2 + y^2 = (x + iy)(x - iy)$, a factorization in the set $\{a + bi : \ a, b \in \mathbb{Z}\}$. Therefore

$$\begin{aligned}(a^2 + b^2)(c^2 + d^2) &= (a + bi)(a - bi)(c + di)(c - di) = (a + bi)(c + di)(a - bi)(c - di) \\ &= ((ac - bd) + (ad + bc)i)((ac - bd) - (ad + bc)i) \\ &= (ac - bd)^2 + (ad + bc)^2,\end{aligned}$$

and so we get (9.1). A different re-arrangement leads to a different identity:

$$(a^2 + b^2)(c^2 + d^2) = (a + bi)(c - di)(a - bi)(c + di) = (ac + bd)^2 + (ad - bc)^2. \tag{9.2}$$

`Exercise 9.1.3.` Prove that if prime $p = a^2 + b^2$ is coprime with $c^2 + d^2$ then $\frac{ac-bd}{ad+bc} \equiv \frac{a}{b} \pmod p$ in (9.1); and $\frac{ac+bd}{ad-bc} \equiv -\frac{a}{b} \equiv \frac{b}{a} \pmod p$ in (9.2).

In Theorem 9.2 we saw that every prime $p \equiv 1 \pmod 4$ can be written as the sum of two squares. A few examples indicate that perhaps there is a unique such representation, up to signs and changing the order of the squares. This will now be proved:

`Exercise 9.1.4.` Suppose that p is a prime $\equiv 1 \pmod 4$ with $p = a^2 + b^2 = c^2 + d^2$ where $a, b, c, d > 0$.

(i) Prove that $(a, b) = (c, d) = 1$.
(ii) Prove that $a/b \equiv c/d$ or $-c/d \pmod p$.
(iii) Assuming that $a/b \equiv c/d \pmod p$ in (ii), use (9.2) to deduce that $p|(ac + bd)$.
(iv) Use (iii) and (9.2) to deduce that $ad = bc$, and then (i) to deduce that $a = c$ and $b = d$.
(v) Work through the case where $a/b \equiv -c/d \pmod p$ using (9.1).

Exercise 9.1.4 tells us that any prime $p \equiv 1 \pmod 4$ can be written as the sum of two squares in a unique way, thus $5 = 1^2 + 2^2$, $13 = 2^2 + 3^2$, $17 = 1^2 + 4^2$ and there are no other representations. For a composite number like 65 we can use the formulae (9.1) and (9.2) to obtain that $65 = 1^2 + 8^2 = 7^2 + 4^2$, and indeed any composite that is the product of two distinct primes $\equiv 1 \pmod 4$ can be written as the sum of two squares in exactly two ways, for examples $85 = 7^2 + 6^2 = 9^2 + 2^2$ and $221 = 13 \cdot 17 = 14^2 + 5^2 = 11^2 + 10^2$. We will discuss the number of representations further in section F5.

Theorem 9.3. *Positive integer n can be written as the sum of two squares of integers if and only if for every prime $p \equiv 3 \pmod 4$ which divides n, the exact power of p dividing n is even.*

Proof. Suppose that $n = a^2 + b^2$ where $(a, b) = 1$. This implies that $(b, n) = 1$ else if prime $q|(b, n)$ then $q|(n - b^2) = a^2$ and so $q|a$ implying that $q|(a, b)$. Therefore if odd prime p divides n then let c be the inverse of $b \pmod n$ so that $(ca)^2 = c^2(n - b^2) \equiv -(bc)^2 \equiv -1 \pmod p$. Hence $(-1/p) = 1$ and so $p \equiv 1 \pmod 4$.

Now suppose that $N = A^2 + B^2$ where $g = (A, B)$, and suppose that p is a prime $\equiv 3 \pmod 4$ which divides N. Writing $A = ga, B = gb$ and $n = N/g^2$, we have $n = a^2 + b^2$

with $(a, b) = 1$ so that $p \nmid n$ by the previous paragraph. Hence $p|g$, and so the power of p dividing N is even, as claimed.

In the other direction, write $n = mg^2$ where m is squarefree. By hypothesis m has no prime factors $\equiv 3 \pmod 4$. Now by Theorem 9.2 we know that every prime factor of m can be written as the sum of two squares. Hence m can be written as the sum of two squares by exercise 9.1.2, and so n can be, multiplying each square through by g^2.

Exercise 9.1.5. Deduce that positive integer n can be written as the sum of two squares of *rationals* if and only if n can be written as the sum of two squares of integers.

In section 6.1 we saw how to find all solutions to $x^2 + y^2 = 1$ in rationals x, y. How about all rational solutions to $x^2 + y^2 = n$? It is not difficult to do this in the case that $n = p$ prime, and this argument can be generalized to arbitrary n:

Proposition 9.4. *Suppose that prime p can be written as $a^2 + b^2$. Then all solutions in rationals x, y to $x^2 + y^2 = p$ are given by the parametrization:*

$$x = \frac{2ars + b(s^2 - r^2)}{r^2 + s^2}, \quad y = \frac{2brs + a(r^2 - s^2)}{r^2 + s^2}, \tag{9.3}$$

or the same with b replaced by $-b$.

Proof sketch. Let x, y be rationals for which $x^2 + y^2 = p$. Let z be the smallest integer such that $X = xz, Y = yz$ are both integers, so that $X^2 + Y^2 = pz^2$. Now $(X, Y)^2 | X^2 + Y^2 = pz^2$ so that $(X, Y)|z$. Therefore $Z = z/(X, Y)$ is an integer with $X/(X, Y) = xZ$, $Y/(X, Y) = yZ$ both integers implying, by the minimality of z that $Z = z$ and so $(X, Y) = 1$.

Now $X^2 + Y^2 \equiv 0 \pmod p$, and so $(X/Y)^2 \equiv -1 \pmod p$ as $(X, Y) = 1$. But then $X/Y \equiv \pm a/b \pmod p$, say '+', so that $p|(bX - aY)$. Now

$$p^2 z^2 = (a^2 + b^2)(X^2 + Y^2) = (aX + bY)^2 + (aY - bX)^2$$

and so $p|(aX + bY)$. Hence $z^2 = ((aX + bY)/p)^2 + ((aY - bX)/p)^2$, and so by (6.1) there exist integers g, r, s such that

$$aX + bY = 2pgrs, \; aY - bX = pg(r^2 - s^2), \; z = g(r^2 + s^2).$$

The result follows.

9.2. The values of $x^2 + dy^2$. How about $x^2 + 2y^2$? We have the identity

$$(a^2 + 2b^2)(c^2 + 2d^2) = (ac + 2bd)^2 + 2(ad - bc)^2,$$

analogous to (9.1), so can focus on what primes are represented. Now if odd prime $p = x^2 + 2y^2$ then $(-2/p) = 1$. On the other hand if $(-2/p) = 1$ then select $b \pmod p$ such that $b^2 \equiv -2 \pmod p$. We take $R = 2^{1/4}\sqrt{p}$, $S = 2^{-1/4}\sqrt{p}$ in exercise 9.1.1, so that p divides $r^2 + 2s^2$, which is $\leq 2^{3/2}p < 3p$. Hence $r^2 + 2s^2 = p$ or $2p$. In the latter case $2|2p - 2s^2 = r^2$ so that $2|r$. Writing $r = 2R$ we have $s^2 + 2R^2 = p$. Hence we have proved that p can be written as $m^2 + 2n^2$ if and only if $p = 2$ or $p \equiv 1$ or $3 \pmod 8$.

`Exercise 9.2.1.` What integers can be written as $x^2 + 2y^2$?

`Exercise 9.2.2.` Fix integer $d \geq 1$. Give an identity showing that the product of two integers of the form $a^2 + db^2$ is also of this form.

`Exercise 9.2.3.` Try to determine what primes are of the form $a^2 + 3b^2$, and $a^2 + 5b^2$, $a^2 + 6b^2$, etc.

9.3. Solutions to quadratic equations. It is easy to see that there do not exist non-zero integers a, b, c such that $a^2 + 5b^2 = 3c^2$. For if we take the smallest non-zero solution then we have

$$a^2 \equiv 3c^2 \pmod 5$$

and since $(3/5) = -1$ this implies that $a \equiv c \equiv 0 \pmod 5$ and so $b \equiv 0 \pmod 5$. Therefore $a/5, b/5, c/5$ gives a smaller solution to $x^2 + 5y^2 = 3z^2$, contradicting minimality.

Another proof stems from looking at the equation mod 4 since then $a^2 + b^2 + c^2 \equiv 0 \pmod 4$, and thus $2|a, b, c$ as 0 and 1 are the only squares mod 4, and so $a/2, b/2, c/2$ gives a smaller solution, contradicting minimality.

In general there are an even number of proofs modulo powers of different primes that a given quadratic equation has no solutions if there are none. These are not difficult to identify. On the other hand, what is remarkable, is that if there are no such "mod p^k obstructions", then there are non-zero integer solutions:

The Local-Global Principal for Quadratic Equations. *Let a, b, c be given integers. There are solutions in integers ℓ, m, n to $a\ell^2 + bm^2 + cn^2 = 0$ if and only if there are real numbers λ, μ, ν for which $a\lambda^2 + b\mu^2 + c\nu^2 = 0$, and for all positive integers r there exist residue classes $u, v, w \pmod r$, not all $\equiv 0 \pmod r$, such that $au^2 + bv^2 + cw^2 \equiv 0 \pmod r$.*

Notice the similarity with the Local-Global Principal for Linear Equations given in section 3.4. Just as there, we can restrict our attention to just one modulus r. We may also restrict the set of a, b, c without loss of generality:

`Exercise 9.3.1.` Show that we may assume a, b, c are squarefree, without loss of generality. (Hint: Suppose that $a = Ap^2$ for some prime p and establish a 1-to-1 correspondence with the solutions for A, b, c.)

`Exercise 9.3.2.` Show that we may also assume that a, b, c are pairwise coprime.

The Local-Global Principal for Quadratic Equations. (Legendre, 1785) *Suppose that squarefree non-zero integers a, b, c are pairwise coprime. Then the equation*

$$a\ell^2 + bm^2 + cn^2 = 0$$

has solutions in integers, other than $\ell = m = n = 0$ if and only if $-bc$ is a square mod a, $-ac$ is a square mod b, and $-ab$ is a square mod c, and a, b and c do not all have the same sign.

We can restate again the criterion asking only for solutions to $a\ell^2 + bm^2 + cn^2 \equiv 0 \pmod{abc}$ with $(\ell mn, abc) = 1$.

Proof $\Longrightarrow$. If a, b, c all have the same sign, then so do $a\ell^2, bm^2, cn^2$ and then the only solution is $\ell = m = n = 0$. Otherwise, suppose that we have the minimal non-zero solution.

We show that $(m, a) = 1$: If not $p|(m, a)|a\ell^2 + bm^2 = -cn^2$ and so $p|n$ as $(a, c) = 1$. Moreover $p^2|bm^2 + cn^2 = -a\ell^2$ and so $p|\ell$ as a is squarefree. But then $\ell/p, m/p, n/p$ yields a smaller solution, contradicting minimality.

Now $bm^2 \equiv -cn^2 \pmod a$ and as $(m, a) = 1$ there exists r such that $rm \equiv 1 \pmod a$. Therefore $-bc \equiv -bc(rm)^2 = cr^2 \cdot (-bm^2) \equiv cr^2 \cdot cn^2 = (crn)^2 \pmod a$.

An analogous argument works mod b and mod c.

Proof $\Longleftarrow$. Interchanging a, b, c, and multiplying through by -1, as necessary, we can assume that $a, b > 0 > c$.

Suppose that α, β, γ are integers such that

$$\alpha^2 \equiv -bc \pmod a, \quad \beta^2 \equiv -ac \pmod b, \quad \gamma^2 \equiv -ab \pmod c.$$

Construct, using the Chinese Remainder Theorem integers u, v, w for which

$$u \equiv \begin{cases} \gamma \pmod c \\ c \pmod b \end{cases}, \quad v \equiv \begin{cases} \alpha \pmod a \\ a \pmod c \end{cases}, \quad w \equiv \begin{cases} \beta \pmod b \\ b \pmod a \end{cases}.$$

`Exercise 9.3.3.` Working mod a, b, c separately and then using the Chinese Remainder Theorem, verify that

$$au^2 + bv^2 + cw^2 \equiv 0 \pmod{abc};$$

And show that if x, y, z are integers for which $aux + bvy + cwz \equiv 0 \pmod{abc}$ then

$$ax^2 + by^2 + cz^2 \equiv 0 \pmod{abc}.$$

Now consider the set of integers

$$\{aui + bvj + cwk : \ 0 \le i \le \sqrt{|bc|}, \ 0 \le j \le \sqrt{|ac|}, \ 0 \le k \le \sqrt{|ab|}\}.$$

The number of i values is $1 + [\sqrt{|bc|}] > \sqrt{|bc|}$; and similarly the number of j and k values, so that the number of elements of the set is $> \sqrt{|bc|} \cdot \sqrt{|ac|} \cdot \sqrt{|ab|} = |abc|$. Hence two different elements of the set are congruent mod abc, say $aui + bvj + cwk \equiv auI + bvJ + cwK \pmod{abc}$. Then $x = i - I$, $y = j - J$, $z = k - K$ are not all zero, and $aux + bvy + cwz \equiv 0 \pmod{abc}$. By the previous exercise we deduce that $ax^2 + by^2 + cz^2 \equiv 0 \pmod{abc}$. Now $|x| \le \sqrt{|bc|}$, $|y| \le \sqrt{|ac|}$, $|z| \le \sqrt{|ab|}$ and so

$$-abc = 0 + 0 - abc \le ax^2 + by^2 + cz^2 \le abc + abc + 0 = 2abc.$$

Since $|bc|, |ac|, |ab|$ are squarefree integers by hypothesis, if we get equality in either inequality here then $a = b = 1$, but this case is settled by Theorem 9.3. Hence we may assume that

$$ax^2 + by^2 + cz^2 \equiv 0 \pmod{abc}, \quad \text{and} \quad -abc < ax^2 + by^2 + cz^2 < 2abc,$$

so that $ax^2 + by^2 + cz^2 = 0$ as desired or $ax^2 + by^2 + cz^2 = abc$. The first case gives us the theorem with excellent bounds on the solutions. In the second we make an unintuitive transformation to note that

$$a(xz + by)^2 + b(yz - ax)^2 + c(z^2 - ab)^2 = (z^2 - ab)(ax^2 + by^2 + cz^2 - abc) = 0$$

In 1950, Holzer showed that if there are solutions then the smallest non-zero solution satisfies

$$|a\ell^2|, \; |bm^2|, \; |cn^2| \leq |abc|.$$

In 1957, Selmer showed that the Local-Global Principal does not necessarily hold for cubic equations since $3x^3 + 4y^3 + 5z^3 = 0$ has solutions in the reals, and mod r for all $r \geq 1$, yet has no integer solutions.

10. Square Roots and Factoring

10.1. Square roots mod p. How difficult is it to find square roots mod n? The first question to ask is how many square roots does a square have mod n?

Lemma 10.1. *If n is a squarefree odd integer with k prime factors, and A is a square mod n with $(A, n) = 1$, then there are exactly 2^k residues mod n whose square is $\equiv A \pmod n$.*

Proof. Suppose that $b^2 \equiv A \pmod n$ where $n = p_1 p_2 \dots p_k$, and each p_i is odd and distinct. If $x^2 \equiv A \pmod n$ then $n|(x^2 - b^2) = (x-b)(x+b)$ so that p divides $x - b$ or $x + b$ for each $p|n$. Now p cannot divide both else p divides $(x + b) - (x + b) = 2b$ and so $4A \equiv (2b)^2 \equiv 0 \pmod p$, which contradicts that fact that $(p, 2A)|(n, 2A) = 1$. So let

$$d = (n, x - b), \quad \text{and therefore} \quad n/d = (n, x + b).$$

Then $x \equiv b_d \pmod n$ where b_d is that unique residue class mod n for which

$$b_d \equiv \begin{cases} b \pmod d \\ -b \pmod{n/d}. \end{cases}$$

Note that the b_d are well-defined by the Chinese Remainder Theorem, are distinct, and that $x^2 \equiv b_d^2 \equiv b^2 \equiv A \pmod n$ for each d.

Now suppose that one has a fast algorithm for finding square roots mod n; that is, given a square A mod n the algorithm finds a square root, say $b \pmod n$. We claim that one can then rapidly find a non-trivial factor of n: Take a random number $x \pmod n$ and let $A \equiv x^2 \pmod n$. Apply the algorithm to obtain $b \pmod n$ such that $b^2 \equiv A \pmod n$. By the proof of the Lemma we know that $x \equiv b_d \pmod n$ for some $d|n$; and since x was chosen at random, each d is possible with probability $1/2^k$. Note that $d = (n, x - b)$ and $n/d = (n, x+b)$ so we have a non-trivial factorization of n provided $d \neq 1, n$. This happens with probability $1 - 2/2^k \geq 1/2$ for n composite. If one is unlucky, that is, if $d = 1$ or n. then we repeat the process. choosing our new value of x independently of the first round.

On the other hand if we can find a non-trivial factor d of n and we already have a square root b of A, then it is easy to find another square-root b_d, and this is $\not\equiv \pm b \pmod n$.

Hence we have shown that finding square roots mod n, and factoring n are, more-or-less, equally difficult.

10.2. Cryptosystems. Cryptography has been around for as long as the need to communicate secrets at a distance. Julius Caesar, on campaign, communicated military messages by creating *cyphertext* by replacing each letter with that letter which is three further on in the alphabet. Thus A becomes D, B becomes E, etc. For example,

$$THISISVERYINTERESTING \text{ becomes } WKLVLVYHUBLQWHUHVWLQJ.$$

(Notice that Y became B, since we wrap around to the beginning of the alphabet. It is essentially the map $x \to x + 3 \pmod{26}$.) At first sight an enemy might regard $WKLV \dots WLQJ$ as gibberish even if the message was intercepted. It is easy enough to decrypt the cyphertext, simply by going back three places in the alphabet for each letter, to reconstruct the original message. The enemy could easily do this if (s)he guessed that the *key* is to rotate the letters by three places in the alphabet, or even if they guessed that one rotates letters, since there would only be 26 possibilities to try. So in classical cryptography it is essential to keep the key secret and probably even the general technique by which the key was created.[13]

One can generalize to arbitrary *substitution cyphers* where one replaces the alphabet by some permutation of the alphabet. There are 26! permutations of our alphabet, which is around 4×10^{26} possibilities, enough one might think to be safe. And it would be if the enemy went through each possibility, one at a time. However the clever *cryptographer* will look for patterns in the cyphertext. In the above short message we see that L appears four times amongst the 21 letters, and H, V, W three times each, so it is likely that these letters each represent one of A, E, I, S, T. By looking for multiword combinations (like the cyphertext for THE) one can quickly break any cyphertext of around one hundred letters.

To combat this, armies in the First World War used longer cryptographic keys, rather than of length 1. That is they would take a word like $ABILITY$ and since A is letter 1 in the alphabet, B is letter 2, and $ILITY$ are letters 9,12,9,20,25, respectively, they would rotate on through the alphabet by $1, 2, 9, 12, 9, -6, -1$ letters to *encrypt* the first seven letters, and then repeat this process on the next seven. This can again be "broken" by statistical analysis, though the longer the key length the harder. Of course using a long key on a battlefield would be difficult, so one needed to compromise between security and practicality. A *one-time pad*, where one uses such a long key that one never repeats a pattern, is unbreakable by statistical analysis. This might have been used by spies during the cold war, and was perhaps based on the letters in an easily obtained book, so that the spy would not have to possess any obviously incriminating evidence.

During the Second World War the Germans came up with an extraordinary substitution cypher that involved changing several settings on a specially built typewriter (an *Enigma machine*). The number of possibilities were so large that the Germans remained confident that it could not be broken, and even changed the settings every day so as to ensure that it would be extremely difficult. The Poles managed to obtain an early Enigma machine and send it to London during their short part in the war and so, since they had a good idea how these machines worked, a great amount of effort was put in by the UK and US to be able to break German codes quickly enough to be useful. Early successes led to the Germans becoming more cautious, and led to horrific decisions having to be made by the Allied leaders to safeguard this most precious secret.[14]

The Allied cryptographers would cut down the number of possibilities (for the settings

---

[13] *Steganography*, hiding secrets in plain view, is another method for communicating secrets at a distance. In 499 BC, Histiaeus shaved the head of his most trusted slave, tattooed a message on his bald head, and then sent the slave to Aristagoras, once the slave's hair had grown back. Aristagoras then shaved the slave's head again to recover the secret message telling him to revolt against the Persians. In more recent times, cold war spies reportedly used "microdots" to transmit information, and Al-Qaeda supposedly notifies its terrorist cells via messages hidden in images on certain webpages.

[14] The ability to crack the Enigma code might have allowed leaders to save lives, but had they done so too often, making it obvious that they had broken the code, then the Germans were liable to have moved on to a different cryptographic method, which the Allied codebreakers might have been unable to crack. Hence the leadership was forced to use its knowledge sparingly so that it would be available in the militarily most advantageous situations.

on the Enigma machine) to a few million, and then their challenge became to build a machine to try out many possibilities very rapidly. Up until then one would have to change, by hand, external settings on the machine to try each possibility; it became a goal to create a machine where one could change what it was doing *internally*, by what became known as a *program*, and this stimulated in part the creation of the first modern computers.

10.3. RSA. In the theory of cryptography we always have two people, Alice and Bob, attempting to share a secret over an open communication channel, and the evil Oscar listening in, attempting to figure out what the message says. We will begin by describing a *private key* scheme for exchanging secrets based on the ideas in our number theory course:

Suppose that prime p is given and integers d and e such that $de \equiv 1 \pmod{p-1}$. Alice knows p and e but not d, while Bob knows p and d but not e. The numbers d and e are kept secret by whoever knows them. Thus if Alice's secret message is M, she *encrypts* M by computing $x \equiv M^e \pmod{p}$. She sends the *cyphertext* x over the open channel. Then Bob *decrypts* by raising x to the dth power mod p, since

$$x^d \equiv (M^e)^d \equiv M^{de} \equiv M \pmod{p}$$

as $de \equiv 1 \pmod{p-1}$. Now if Oscar steals the values of p and e from Alice, he will be able to determine d, since d is the inverse of e mod $p-1$, and this can be determined by the method of section 1.1.

This is the problem with most classical cryptosystems; once one knows the encryption method it is not difficult to determine the decoding method. In 1975 Diffie and Hellman proposed a sensational idea: Can one give a cryptographic scheme in which the encryption method gives no help in determining a decryption method? If one could, one would then have a *public key* cryptographic scheme. What they realized is that this would be exactly what is needed in our age of electronic information, in particular allowing people to use passwords in public places (for instance when using an ATM) without fear that any lurking Oscar will be able to impersonate them.

In 1977 Rivest, Shamir and Adleman realized this ambition, via a minor variation of the above private key cryptosystem: Now let $p \neq q$ be two large primes and $n = pq$. Select integers d and e such that $de \equiv 1 \pmod{\phi(pq)}$. Alice knows pq and e but not d, while Bob knows pq and d. Thus if Alice's secret message is M, the cyphertext is $x \equiv M^e \pmod{pq}$, and Bob decrypts this by taking $x^d \equiv (M^e)^d \equiv M^{de} \equiv M \pmod{pq}$ as $de \equiv 1 \pmod{\phi(pq)}$ using Euler's Theorem.

Now, if Oscar steals the values of pq and e from Alice, will he be able to determine d, the inverse of e mod $(p-1)(q-1)$? When the modulus was the prime p, Oscar had no difficulty in determining $p-1$. Now that the modulus is pq, can Oscar easily determine $(p-1)(q-1)$? If so, then since he already knows pq, he would be able to determine $pq+1-\phi(pq) = p+q$ and hence p and q, since they are the roots of $x^2-(p+q)x+pq = 0$. In other words, if Oscar could "break" the RSA algorithm, then he could factor pq, and vice-versa.[15]

[15] This is somewhat misleading. We have not proved that the only way to determine d is via knowing the value of $(p-1)(q-1)$; however I cannot think of another way.

If breaking RSA is as difficult as factoring, then we believe that RSA is secure only if we believe that it is difficult to factor... Is it? No one knows. Certainly we do not know any very efficient ways to factor large numbers, but that does not necessarily mean that there is no quick way to do so. So why do we put our faith (and secrets and fortunes) in the difficulty of factoring? The reason is that many of the greatest minds in history, from Gauss onwards, have looked for an efficient factoring algorithm and failed. Is this a good basis to have faith in RSA? Probably not, but we have no better.[16] More on this at the end of section A5.

10.4. Certificates and the complexity classes `P` and `NP`. Algorithms are typically designed to work on any of an arbitrarily large class of examples, and one wishes them to work as fast as possible. If the example is input in ℓ characters, and the function calculated is genuinely a function of all the characters of the input, then one cannot hope to compute the answer any quicker than the length, ℓ, of the input. A *polynomial time algorithm* is one in which the answer is computed in no more than $c\ell^A$ steps, for some fixed $c, A > 0$, no matter what the input. These are considered to be quick algorithms. There are many simple problems that can be answered in polynomial time (the set of such problems is denoted by `P`); see section A5 for more details. In modern number theory, because of the intrinsic interest as well as because of the applications to cryptography, we are particularly interested in the running times of factoring and primality testing algorithms.

At the 1903 meeting of the American Mathematical Society, F.N. Cole came to the blackboard and, without saying a word, wrote down

$$2^{67} - 1 = 147573952589676412927 = 193707721 \times 761838257287,$$

long-multiplying the numbers out on the right side of the equation to prove that he was indeed correct. Afterwards he said that figuring this out had taken him "three years of Sundays". The moral of this tale is that although it took Cole a great deal of work and perseverance to find these factors, it did not take him long to justify his result to a room full of mathematicians (and, indeed, to give a proof that he was correct). Thus we see that one can provide a short proof, even if finding that proof takes a long time.

In general one can exhibit factors of a given integer n to give a short proof that n is composite. Such proofs, that can be checked in polynomial time, are called *certificates* (The set of problems which can be checked in polynomial time is denoted by `NP`.) Note that it is not necessary to exhibit factors to give a short proof that a number is composite. Indeed, we already saw in the converse to Fermat's Little Theorem, Corollary 7.4, that one can exhibit an integer a coprime to n for which n does not divide $a^{n-1} - 1$ to provide a certificate that n is composite.

What about primality testing? If someone gives you an integer and asserts that it is prime, can you quickly check that this is so? Can they give you better evidence than their say-so that it is a prime number? Can they provide some sort of certificate that gives you all the information you need to quickly verify that the number is indeed a prime? We had hoped (see section 7.6) that we could use the converse of Fermat's Little Theorem to

[16] It is a notoriously difficult open problem to find mathematical problems for which one can prove that there is no efficient algorithm. No one has yet succeeded.

establish a quick primality test, but we saw that Carmichael numbers seem to stop that idea from reaching fruition. Here we are asking for less, for a short certificate for a proof of primality. It is not obvious how to construct such a certificate; certainly not so obvious as with the factoring problem. It turns out that some old remarks of Lucas from the 1870's can be modified for this purpose:

First note that n is prime if and only if there are precisely $n-1$ integers a in the range $1 \le a \le n-1$ which are coprime to n. Therefore if we can show the existence of $n-1$ distinct values mod n which are coprime to n, then we have a proof that n is prime. So to prove that n is prime we could exhibit a primitive root g, along with a proof that it is indeed a primitive root. Corollary 7.10 shows that g is not a primitive root mod n if and only if $g^{(n-1)/q} \equiv 1 \pmod n$ for some prime q dividing $n-1$. Thus a "certificate" to show that n is prime would consist of g and $\{q \text{ prime} : q \text{ divides } n-1\}$, and the checker would need to verify that $g^{n-1} \equiv 1 \pmod n$ whereas $g^{(n-1)/q} \not\equiv 1 \pmod n$ for all primes q dividing $n-1$, something that can be quickly accomplished using fast exponentiation (as explained in section A5).

There is a problem though: One needs certification that each such q is prime. The solution is to iterate the above algorithm; and one can show that no more than $\log n$ odd primes need to be certified prime in the process of proving that n is prime. Thus we have a short certificate that n is prime.

At first one might hope that this also provides a quick way to test whether a given integer n is prime. However there are several obstacles. The most important is that we need to factor $n-1$. When one is handed the certificate $n-1$ is already factored, so that is not an obstacle to the use of the certificate; however it is a fundamental impediment to the rapid creation of the certificate.

10.5. Polynomial time Primality testing. Although the converse to Fermat's Little Theorem does not provide a polynomial time primality test, one can further develop this idea. For example, we know that $a^{\frac{p-1}{2}} \equiv -1 \text{ or } 1 \pmod p$ by Euler's criterion, and hence if $a^{\frac{n-1}{2}} \not\equiv \pm 1 \pmod n$ then n is composite. This identifies even more composite n then Corollary 7.4 alone, but not necessarily all n. We develop this idea further in section D3 to find a criterion of this type that is satisfied by all primes but not by any composites. However we are unable to prove that this is indeed a polynomial time primality test without making certain assumptions that are, as yet, unproved.

There have indeed been many ideas for establishing a primality test, which is provably polynomial time, but this was not achieved until 2002. This was of particular interest since the proof was given by a professor, Manindra Agrawal, and two undergraduate students, Kayal and Saxena, working together on a summer research project. Their algorithm is based on the following elegant characterization of prime numbers.

Agrawal, Kayal and Saxena. *For given integer $n \ge 2$, let r be a positive integer $< n$, for which n has order $> (\log n)^2$ modulo r. Then n is prime if and only if*

- *n is not a perfect power,*
- *n does not have any prime factor $\le r$,*
- $(x+a)^n \equiv x^n + a \mod (n, x^r - 1)$ *for each integer* $a, 1 \le a \le \sqrt{r} \log n$.

At first sight this might seem to be a rather complicated characterization of the prime numbers. However this fits naturally into the historical progression of ideas in this subject, is not so complicated (compared to some other ideas in use), and has the great advantage that it is straightforward to develop into a fast algorithm for proving the primality of large primes.

10.6. Factoring methods.

"*The problem of distinguishing prime numbers from composite numbers, and of resolving the latter into their prime factors is known to be one of the most important and useful in arithmetic. It has engaged the industry and wisdom of ancient and modern geometers to such an extent that it would be superfluous to discuss the problem at length. Nevertheless we must confess that all methods that have been proposed thus far are either restricted to very special cases or are so laborious and difficult that even for numbers that do not exceed the limits of tables constructed by estimable men, they try the patience of even the practiced calculator. And these methods do not apply at all to larger numbers ... It frequently happens that the trained calculator will be sufficiently rewarded by reducing large numbers to their factors so that it will compensate for the time spent. Further, the dignity of the science itself seems to require that every possible means be explored for the solution of a problem so elegant and so celebrated ... It is in the nature of the problem that any method will become more complicated as the numbers get larger. Nevertheless, in the following methods the difficulties increase rather slowly ... The techniques that were previously known would require intolerable labor even for the most indefatigable calculator.*" — from article 329 of *Disquisitiones Arithmeticae* (1801) by C. F. Gauss.

After trial division, which looks for small factors first, perhaps the first factoring technique was given by Fermat: His goal was to write n as $x^2 - y^2$. He started with m, the smallest integer $\geq \sqrt{n}$, and then looked to see if $m^2 - n$ is a square. Fermat simplified this, by testing whether $m^2 - n$ is a square modulo various small primes. If $m^2 - n$ is not a square then he tested whether $(m+1)^2 - n$ is a square; if that failed, whether $(m+2)^2 - n$ is a square, or $(m+3)^2 - n, \dots$, etc. Since Fermat computed by hand he also noted the trick that $(m+1)^2 - n = m^2 - n + (2m+1)$, $(m+2)^2 - n = (m+1)^2 - n + (2m+3)$, etc. For example, Fermat factored $n = 2027651281$ so that $m = 45030$. Then

$$
\begin{aligned}
45030^2 - n &= 49619 \text{ which is not a square mod } 100;\\
45031^2 - n &= 49619 + 90061 = 139680 \text{ which is divisible by } 2^5, \text{ not } 2^6;\\
45032^2 - n &= 139680 + 90063 = 229743 \text{ which is divisible by } 3^3, \text{ not } 3^4;\\
45033^2 - n &= 229743 + 90065 = 319808 \text{ which is not a square mod } 3; \text{ etc}\\
&\vdots
\end{aligned}
$$

up until $45041^2 - n = 1020^2$, so that $n = 2027651281 = 45041^2 - 1020^2 = 44021 \times 46061$.

Gauss and other authors further developed Fermat's ideas, most importantly realizing that if $x^2 \equiv y^2 \pmod n$ with $x \not\equiv \pm y \pmod n$ and $(x, n) = 1$, then

$$\gcd(n, x - y) \cdot \gcd(n, x + y)$$

gives a non-trivial factorization of n.

Several factoring algorithms work by generating a pseudo-random sequence of integers $a_1, a_2, ...,$ with each

$$a_i \equiv b_i^2 \pmod n,$$

for some known integer b_i, until some subsequence of the a_i's has product equal to a square, say

$$y^2 = a_{i_1} \cdots a_{i_r},$$

and set $x^2 = (b_{i_1} \cdots b_{i_r})^2$. Then $x^2 \equiv y^2 \pmod n$ and there is a good chance that $\gcd(n, x-y)$ is a non-trivial factor of n.

We want to generate the a_is so that it is not so difficult to find a subsequence whose product is a square; to do so, we need to be able to factor the a_i. This is most easily done by only keeping those a_i that have all of their prime factors $\leq y$. Suppose that the primes up to y are $p_1, p_2, \ldots, p_k$. If $a_i = p_1^{a_{i,1}} p_2^{a_{i,2}} \cdots p_k^{a_{i,k}}$ then let $v_i = (a_{i,1}, a_{i,2}, \ldots, a_{i,k})$, which is a vector with entries in $\mathbb{Z}$.

`Exercise 10.6.1.` Show that $\prod_{i \in I} a_i$ is a square if and only if $\sum_{i \in I} v_i \equiv (0, 0, \ldots, 0) \pmod 2$.

Hence to find a non-trivial subset of the a_i whose product is a square, we simply need to find a non-trivial linear dependency mod 2 amongst the vectors v_i. This is easily achieved through the methods of linear algebra, and guaranteed to exist once $r > k$.

The quadratic sieve factoring algorithm selects the b_i so that it is easy to find the small prime factors of the a_i, using Corollary 2.3. There are other algorithms that attempt to select the b_i so that the a_i are small and therefore more likely to have small prime factors. The best algorithm, the number field sieve, is an analogy to the quadratic sieve algorithm, over number fields.

10.7. Cryptosystems based on discrete logarithms. Another elementary number theory problem that appears to be difficult is the "discrete log problem", so that it used as the basis for various cryptographic protocols. The issue is that given a primitive root g mod p and an integer k (preferably coprime with $p-1$) one can easily determine $a :\equiv g^k \pmod p$, whereas it is not obvious, given a primitive root g mod p and a residue a, how to find $k := \mathrm{ind}_p(a)$, the discrete log of a mod p in base g.

The Diffie-Hellman key exchange. Alice and Bob wish to create a secret number that they both know, without meeting, so to do so, they must share information across on open channel with Oscar listening in:

1. They agree upon a large prime p and primitive root g.
2. Alice picks a secret exponent a, and Bob picks a secret exponent b.
3. Alice transmits the least positive residue of $g^a \pmod p$ to Bob, and Bob transmits the least positive residue of $g^b \pmod p$ to Alice.
4. The secret key is the least residue of $g^{ab} \pmod p$. Alice computes it as $g^{ab} \equiv (g^b)^a \pmod p$, and Bob computes it as $g^{ab} \equiv (g^a)^b \pmod p$.

Oscar has access to p as well as to g, g^a and $g^b \pmod p$, and he wishes to determine $g^{ab} \pmod p$. The only obvious way to proceed, with the information that he has, is to compute the discrete logarithms of g^a or g^b in base g, to recover a or b and hence determine $(g^b)^a$

or $(g^a)^b \pmod p$. Notice that, in this exchange, Bob never knows Alice's secret exponent a, and Alice never knows Bob's secret exponent b.

There is a lot more that can be said, for example how to stop a "man-in-the-middle" attack. That is Oscar can get in-between Alice and Bob on their communication channel, and hence pretend to be Bob when dealing with Alice and send her his own g^b; and similarly pretend to be Alice when dealing with Bob and send him his own g^a. It is difficult to stop, or even to recognize, such fraudulent behaviour, but there are well-established protocols for dealing with this, and other, difficult situations.

The El Gamal cryptosystem. This is another public key cryptosystem (like RSA), in which one uses the secret key, $g^{ab} \pmod p$, above:

(1) Alice wishes to transmit a message M to Bob. She creates the cyphertext $x \equiv M/g^{ab} \pmod p$, and transmits that.
(2) Bob determines the original message M by computing $xg^{ab} \pmod p$.

As before Oscar has access to p as well as to g, g^a and $g^b \pmod p$, and the cyphertext x. Determining $M \equiv xg^{ab} \pmod p$ is therefore equivalent to determining $g^{ab} \pmod p$. Oscar finds himself back with the same mathematical problem as in the Diffie-Hellman key exchange.

Why does one choose one cryptosystem over another? This is an important practical question, especially given that we are unable to prove that any particular cryptosystem is truly secure from an intelligent attack (that is, there may be polynomial time algorithms for factoring or for solving the discrete log problem). Most people who are not directly involved in selling a particular product, would guess that RSA is the safest, since factoring is a much better explored problem than discrete logs. However RSA has a distinct disadvantage as compared with the El Gamal system, which is the quantity and difficulty of the calculation involved in implementing the algorithms: Let us compare the cryptosystems in the situation that Alice is regularly communicating with Bob. In RSA she must raise M to the power e each time she transmits a message, which requires around $\log e$ multiplications mod p. Typically we choose e to be large, that is of length comparable to the length of p, otherwise RSA will not be very secure. On the other hand, in the El Gamal cryptosystem, Alice can compute A and B^{-a} mod p, once and for all, so that when we she transmits her cyphertext she simply multiplies M by $B^{-a} \pmod p$, one multiplication. This difference is not important if Alice works with a large computer, but many applications today use hand-held devices, like a cellphone or a smartcard, which have limited computing capacity, so this time difference is sigificant.

11. The pigeonhole principle

11.1. Rational approximations to real numbers. We are interested in how close the integer multiples of a given real number α can get to an integer; that is, are there integers m, n such that $n\alpha - m$ is small? It is obvious that if $\alpha = p/q$ is rational then $n\alpha = m$ whenever $n = kq$ for some integer k, so that $m = kp$. How about irrational α?

Dirichlet's Theorem. *Suppose that α is a given real number. For every integer $N \geq 1$ there exists a positive integer $n \leq N$ such that*

$$|n\alpha - m| < \frac{1}{N},$$

for some integer m.

Proof. The $N+1$ numbers $\{0 \cdot \alpha\}$, $\{1 \cdot \alpha\}$, $\{2 \cdot \alpha\}, \dots$, $\{N \cdot \alpha\}$ all lie in the interval $[0, 1)$. The intervals

$$\left[0, \frac{1}{N}\right), \left[\frac{1}{N}, \frac{2}{N}\right), \dots, \left[\frac{N-1}{N}, 1\right)$$

partition $[0, 1)$,[17] and so each of our $N + 1$ numbers lies in exactly one of the N intervals. Hence some interval contains at least two of our numbers, say $\{i\alpha\}$ and $\{j\alpha\}$ with $0 \leq i < j \leq N$, so that $|\{i\alpha\} - \{j\alpha\}| < \frac{1}{N}$. Therefore, if $n = j - i$ then $1 \leq n \leq N$, and if $m := [j\alpha] - [i\alpha] \in \mathbb{Z}$ then

$$n\alpha - m = (j\alpha - i\alpha) - ([j\alpha] - [i\alpha]) = \{j\alpha\} - \{i\alpha\},$$

and the result follows.

Corollary 11.1. *If α is a real irrational number then there are infinitely many pairs m, n of coprime positive integers for which*

$$\left|\alpha - \frac{m}{n}\right| < \frac{1}{n^2} .$$

Proof. By Dirichlet's Theorem we have that for any $N > 1$, there exists $n \leq N$ such that

$$\left|\alpha - \frac{m}{n}\right| < \frac{1}{nN} \leq \frac{1}{n^2} .$$

Suppose that we already have the pairs (m_j, n_j), $1 \leq j \leq k$ and then let N be the smallest integer $\geq 1/\min_{1 \leq j \leq k}\{|n_j\alpha - m_j|\}$. Applying Dirichlet's Theorem we find that there exists a pair of integers m, n for which

$$|n\alpha - m| < \frac{1}{N} \leq |n_j\alpha - m_j| \text{ for all } j,$$

[17] That is each point of $[0, 1)$ lies in exactly one of these intervals, and the union of these intervals exactly equals $[0, 1)$.

and so this gives a new such pair.

`Exercise 11.1.1.` How can we guarantee that $\min_{1\leq j\leq k}\{|n_j\alpha - m_j|\} \neq 0$ so that N is well-defined?

Another Proof of Corollary 3.7. Let $\alpha = \frac{a}{m}$ and $N = m-1$ in Liouville's Theorem so that there exists integers $r \leq m-1$ and s such that $|ra/m - s| < 1/(m-1)$; that is $|ra - sm| < m/(m-1) \leq 2$. Hence $ra - sm = -1, 0$ or 1. It cannot equal 0 else $m|sm = ar$ and $(m,a) = 1$ so that $m|r$ which is impossible as $r < m$. Hence $ra \equiv \pm 1 \pmod m$ and so $\pm r$ is the inverse of $a \pmod m$.

For irrational α one might ask how the numbers $\{\alpha\}$, $\{2\alpha\}, \dots$, $\{N\alpha\}$ are distributed in $[0,1)$ as $N \to \infty$, for α irrational. In an section G3 we will show that the values are dense and see how this ties in with the geometry of the torus, and exponential sum theory.

We saw an important use of the pigeonhole principle in number theory in the proof of Theorem 9.2, and this idea was generalized significantly by Minkowski and others.

11.2. Pell's equation. Perhaps the most researched equation in the early history of number theory is the so-called Pell's equation: Are there integer solutions x, y to

$$x^2 - dy^2 = 1?$$

We will show in Theorem 11.2 that the answer is "yes" for any non-square positive integer d. In section C4 we will see that solutions can always be found using the continued fraction for $\sqrt{d}$. This was evidently known to Brahmagupta in India in 628 A.D., and one can guess that it was well understood by Archimedes, judging by his "Cattle Problem":

The Sun god's cattle, friend, apply thy care
to count their number, hast thou wisdom's share.
They grazed of old on the Thrinacian floor
of Sic'ly's island, herded into four,
colour by colour: one herd white as cream,
the next in coats glowing with ebon gleam,
brown-skinned the third, and stained with spots the last.
Each herd saw bulls in power unsurpassed,
in ratios these: count half the ebon-hued,
add one third more, then all the brown include;
thus, friend, canst thou the white bulls' number tell.
The ebon did the brown exceed as well,
now by a fourth and fifth part of the stained.
To know the spottedall bulls that remained
reckon again the brown bulls, and unite
these with a sixth and seventh of the white.
Among the cows, the tale of silver-haired
was, when with bulls and cows of black compared,
exactly one in three plus one in four.
The black cows counted one in four once more,
plus now a fifth, of the bespeckled breed
when, bulls withal, they wandered out to feed.

The speckled cows tallied a fifth and sixth
of all the brown-haired, males and females mixed.
Lastly, the brown cows numbered half a third
and one in seven of the silver herd.
Tell'st thou unfailingly how many head
the Sun possessed, o friend, both bulls well-fed
and cows of ev'ry colourno-one will
deny that thou hast numbers' art and skill,
though not yet dost thou rank among the wise.
But come! also the foll'wing recognise.

Whene'er the Sun god's white bulls joined the black,
their multitude would gather in a pack
of equal length and breadth, and squarely throng
Thrinacia's territory broad and long.
But when the brown bulls mingled with the flecked,
in rows growing from one would they collect,
forming a perfect triangle, with ne'er
a diff'rent-coloured bull, and none to spare.
Friend, canst thou analyse this in thy mind,
and of these masses all the measures find,
go forth in glory! be assured all deem
thy wisdom in this discipline supreme!

— from an epigram written to Eratosthenes (of Cyrene)
by Archimedes (of Alexandria), 250 B.C.

The first paragraph involves only linear equations. To resolve the second, one needs to find a non-trivial solution in integers u, v to

$$u^2 - 609 \cdot 7766v^2 = 1.$$

The first solution is enormous, the smallest herd having about 7.76×10^{206544} cattle: It wasn't until 1965 that anyone was able to write down all 206545 decimal digits! How did Archimedes know that the solution would be ridiculously large? We don't know, though presumably he did not ask this question by chance.

Theorem 11.2. *Let $d \geq 2$ be a given non-square integer. There exist integers x, y for which*

$$x^2 - dy^2 = 1,$$

with $y \neq 0$. If x_1, y_1 are the smallest solutions in positive integers, then all other solutions are given by the recursion $x_{n+1} = x_1x_n + dy_1y_n$ and $y_{n+1} = x_1y_n + y_1x_n$ for $n \geq 1$.

Proof. We begin by showing that there exists a solution with $y \neq 0$. By Corollary 11.1, there exists infinitely many pairs of integers $(m_j, n_j),\ j = 1, 2, \dots$ such that $|\sqrt{d} - \frac{m}{n}| \leq \frac{1}{n^2}$. Therefore

$$|m^2 - dn^2| = n^2 \left|\sqrt{d} - \frac{m}{n}\right| \cdot \left|\sqrt{d} + \frac{m}{n}\right| \leq \left|\sqrt{d} + \frac{m}{n}\right| \leq 2\sqrt{d} + \left|\sqrt{d} - \frac{m}{n}\right| \leq 2\sqrt{d} + 1.$$

Since there are only finitely many possibilities for $m^2 - dn^2$ there must be some integer r, with $|r| \leq 2\sqrt{d} + 1$ such that there are infinitely many pairs of positive integers m, n for which $m^2 - dn^2 = r$.

Since there are only r^2 pairs of residue classes ($m \bmod r,\ n \bmod r$) there must be some pair of residue classes a, b such that there are infinitely many pairs of integers m, n for which $m^2 - dn^2 = r$ with $m \equiv a \pmod r$ and $n \equiv b \pmod r$. Let m_1, n_1 be the smallest such pair, and m, n any other such pair, so that $m_1^2 - dn_1^2 = m^2 - dn^2 = r$ with $m_1 \equiv m \pmod r$ and $n_1 \equiv n \pmod r$. This implies that $r|(m_1n - n_1m)$ and

$$(m_1m - dn_1n)^2 - d(m_1n - n_1m)^2 = (m_1^2 - dn_1^2)(m^2 - dn^2) = r^2,$$

so that $r^2|(r^2 + d(m_1n - n_1m)^2) = (m_1m - dn_1n)^2$ and thus $r|(m_1m - dn_1n)$. Therefore $x = |m_1m - dn_1n|/r$ and $y = |m_1n - n_1m|/r$ are integers for which $x^2 - dy^2 = 1$.

`Exercise 11.2.1.` Show that $y \neq 0$ using the fact that $(m, n) = 1$ for each such pair m, n.

Let x_1, y_1 be the solution in positive integers with $x_1 + \sqrt{d}y_1$ minimal. Note that this is $\geq 1 + \sqrt{d} > 1$. We claim that all other such solutions take the form $(x_1 + \sqrt{d}y_1)^n$. If not let x, y be the counterexample with $x + \sqrt{d}y$ smallest, and $X = x_1x - dy_1y$ and $Y = x_1y - y_1x$. Then $X^2 - dY^2 = (x_1^2 - dy_1^2)(x^2 - dy^2) = 1$, and

$$X + \sqrt{d}Y = (x_1 - \sqrt{d}y_1)(x + \sqrt{d}y) = \frac{x + \sqrt{d}y}{x_1 + \sqrt{d}y_1} < x + \sqrt{d}y.$$

Since x, y was the smallest counterexample, hence $X + \sqrt{d}Y = (x_1 + \sqrt{d}y_1)^n$ for some integer $n \geq 1$, and therefore $x + \sqrt{d}y = (x_1 + \sqrt{d}y_1)(X + \sqrt{d}Y) = (x_1 + \sqrt{d}y_1)^{n+1}$

`Exercise 11.2.2.` This proof is not quite complete since we have not yet shown that X and Y are both positive. Remedy this problem. (Proving that $X > 0$ is not difficult, from the fact that $x^2 = dy^2 + 1 > dy^2$. One might prove that $Y > 0$ by establishing that $x_1/y_1 - \sqrt{d} > x/y - \sqrt{d}$.)

One of the fascinating things about Pell's equation is the size of the smallest solution, as we saw in the example given by Archimedes. We will indicate in section E4, that the smallest solution is $\leq e^{c\sqrt{d}}$ for some constant $c > 0$. However what is surprising is that usually the smallest solution is really this large. This is not something that has been proved; indeed understanding the distribution of sizes of the smallest solutions to Pell's equation is an outstanding open question in number theory.

Another issue is whether there is a solution to $u^2 - dv^2 = -1$. Notice, for example, that $2^2 - 5 \cdot 1^2 = -1$. Evidently if there is a solution then -1 is a square mod d, so that d has no prime factors $\equiv -1 \pmod 4$. Moreover d is not divisible by 4 else $u^2 \equiv -1 \pmod 4$ which is impossible. We saw that $x^2 - dy^2 = 1$ has solutions for every non-square $d > 1$, and one might have guessed that there would be some simple criteria to decide whether there are solutions to $u^2 - dv^2 = -1$, but there does not appear to be. Even the question of whether there are solutions for "many" d has only recently been resolved by Fouvry and Kluners.

11.3. Transcendental numbers. In section 3.2 we showed that $\sqrt{d}$ is irrational if d is an integer that is not the square of an integer. We can also show that there exist irrational numbers simply by how well they can be approximated by rationals:

Proposition 11.3. *Suppose that α is a given real number. If for every integer $q \geq 1$ there exist integers m, n such that*

$$0 < |n\alpha - m| < \frac{1}{q},$$

then α is irrational.

Proof. If α is rational then $\alpha = p/q$ for some coprime integers p, q with $q \geq 1$. For any integers m, n we then have $n\alpha - m = (np - mq)/q$. Now, the value of $np - mq$ is an integer $\equiv np \pmod q$. Hence $|np - mq| = 0$ or is an integer ≥ 1, and therefore $|n\alpha - m| = 0$ or is $\geq 1/q$.

There are several other methods to prove that numbers are irrational, but more challenging is to prove that a number is *transcendental*; that is, that it is not the root of a polynomial with integer coefficients (such a root is called *an algebraic number*).

Liouville's Theorem. *Suppose that α is the root of an irreducible polynomial $f(x) \in \mathbb{Z}[x]$ of degree ≥ 2. There exists a constant $c_\alpha > 0$ such that for any rational p/q with $(p, q) = 1$ and $q \geq 1$ we have*

$$\left|\alpha - \frac{p}{q}\right| \geq \frac{c_\alpha}{q^d}\ .$$

Proof. Since $I := [\alpha - 1, \alpha + 1]$ is a closed interval, there exists a bound $B \geq 1$ such that $|f'(t)| \leq B$ for all $t \in I$. Let $c_\alpha = 1/B$. If $p/q \notin I$ then $|\alpha - p/q| \geq 1 \geq c_\alpha \geq c_\alpha/q^d$ as desired. Henceforth we may assume that $p/q \in I$.

If $f(x) = \sum_{i=0}^d f_i x^i$ then $q^d f(p/q) = \sum_{i=0}^d f_i p^i q^{d-i} \in \mathbb{Z}$. Now $f(p/q) \neq 0$ since f is irreducible of degree ≥ 2 and so $|q^d f(p/q)| \geq 1$.

The mean value theorem tells us that there exists t lying between α and p/q, and hence in I, such that

$$f'(t) = \frac{f(\alpha) - f(p/q)}{\alpha - p/q}\ .$$

Therefore

$$\left|\alpha - \frac{p}{q}\right| = \frac{|q^d f(p/q)|}{q^d |f'(t)|} \geq \frac{1}{Bq^d} = \frac{c_\alpha}{q^d}\ .$$

One usually first proves that there exist transcendental numbers by simply showing that the set of real numbers is uncountable, and the set of algebraic numbers is countable, so that the vast majority of real numbers are transcendental. However it is unsatisfying that this method yields that most real numbers are transcendental, without actually constructing any! As a consequence of Liouville's Theorem it is not difficult to construct transcendental numbers, for example

$$\alpha = \frac{1}{10} + \frac{1}{10^{2!}} + \frac{1}{10^{3!}} + \dots$$

since if p/q with $q = 10^{(n-1)!}$ is the sum of the first $n - 1$ terms then $0 < \alpha - p/q < 2/q^n$, and α cannot be an algebraic number by Liouville's Theorem.

Liouville's Theorem has been improved to its, more-or-less, final form:

Roth's Theorem. (1955) *Suppose that α is a real algebraic number. For any fixed $\epsilon > 0$ there exists a constant $c_{\alpha,\epsilon} > 0$ such that for any rational p/q with $(p, q) = 1$ and $q \geq 1$ we have*

$$\left|\alpha - \frac{p}{q}\right| \geq \frac{c_{\alpha,\epsilon}}{q^{2+\epsilon}} .$$

Evidently this cannot be improved much since, by Corollary 11.1, we know that if α is real, irrational then there are infinitely many p, q with $\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{q^2}$. In section C2 we will show that all p/q for which $\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{2q^2}$ can be easily identified from the continued fraction of α. Moreover we will see that if α is a quadratic, real irrational then there exists a constant $c_\alpha > 0$ such that $\left|\alpha - \frac{p}{q}\right| \geq \frac{c_\alpha}{q^2}$ for all p/q. The most amusing example is where $\alpha = \frac{1+\sqrt{5}}{2}$ and the best approximations are given by F_{n+1}/F_n where F_n is the nth Fibonacci numbers (see section A1 for details). One can show that

$$\left|\frac{1+\sqrt{5}}{2} - \frac{F_{n+1}}{F_n}\right| = \frac{1}{5F_n^2} + \frac{c_n}{F_n^4},$$

where $|c_n| < 1$.

12. Binary quadratic forms

12.1. Representation of integers by binary quadratic forms. We have already seen that the integers that can be represented by the binary linear form $ax+by$ are those integers divisible by $\gcd(a,b)$.

Exercise 12.1.1. Show that if N can be represented by $ax+by$ then there exist coprime integers m and n such that $am+bn=N$.

Now we let a,b,c be given integers, and ask what integers can be represented by the *binary quadratic form* $ax^2+bxy+cy^2$? That is, for what integers N do there exist coprime integers m,n such that

$$N = am^2+bmn+cn^2 \quad ? \tag{12.1}$$

We may reduce to the case that $\gcd(a,b,c)=1$ by dividing though by $\gcd(a,b,c)$. One idea is to complete the square to obtain

$$4aN = (2am+bn)^2 - dn^2$$

where the *discriminant* $d := b^2-4ac$. Hence $d \equiv 0$ or $1 \pmod 4$. When $d<0$ the right side of the last displayed equation can only take positive values, which makes our discussion easier than when $d>0$. For this reason we will restrict ourselves to the case $d<0$ here, and revisit the case $d>0$ in section C4. In section 9 we already worked with a few basic examples, and we will now see how this theory develops.

Exercise 12.1.2. Show that if $d<0$ then $am^2+bmn+cn^2$ has the same sign as a, no matter what the choices of integers m and n.

We replace a,b,c by $-a,-b,-c$ if necessary, to ensure that the value of $am^2+bmn+cn^2$ is always ≥ 0, and so we call this a *positive definite* binary quadratic form.

Exercise 12.1.3. Show that if $ax^2+bxy+cy^2$ is positive definite then $a,c>0$.

The key idea stems from the observation that x^2+y^2 represents the same integers as $X^2+2XY+2Y^2$. This is easy to see for if $N=m^2+n^2$ then $N=(m-n)^2+2(m-n)n+2n^2$, and similarly if $N=u^2+2uv+2v^2$ then $N=(u+v)^2+v^2$. The reason is that the substitution

$$\begin{pmatrix} x \\ y \end{pmatrix} = M \begin{pmatrix} X \\ Y \end{pmatrix} \qquad \text{where } M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

transforms x^2+y^2 into $X^2+2XY+2Y^2$, and the transformation is invertible, since $\det M = 1$. Much more generally define

$$\mathrm{SL}(2,\mathbb{Z}) = \left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} : \ \alpha,\beta,\gamma,\delta \in \mathbb{Z} \text{ and } \alpha\delta-\beta\gamma=1 \right\}.$$

Exercise 12.1.4. Prove that the binary quadratic form $ax^2+bxy+cy^2$ represents the same integers as the binary quadratic form $AX^2+BXY+CY^2$ whenever $\begin{pmatrix} x \\ y \end{pmatrix} = M \begin{pmatrix} X \\ Y \end{pmatrix}$ with $M \in \mathrm{SL}(2,\mathbb{Z})$. We say that these two quadratic forms are *equivalent*. This yields an equivalence relation and splits the binary quadratic forms into equivalence classes.

`Exercise 12.1.5.` Show that two equivalent binary quadratic forms represent each integer in the same number of different ways.

We can write $ax^2+bxy+cy^2 = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$ and note that the discriminant is -4 times the determinant of $\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}$. We deduce that

$$AX^2 + BXY + CY^2 = \begin{pmatrix} X & Y \end{pmatrix} M^{\mathrm{T}} \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} M \begin{pmatrix} X \\ Y \end{pmatrix},$$

and so $A = a\alpha^2 + b\alpha\gamma + c\gamma^2$ and $C = a\beta^2 + b\beta\delta + c\delta^2$ as

$$\begin{pmatrix} A & B/2 \\ B/2 & C \end{pmatrix} = M^{\mathrm{T}} \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} M. \tag{12.2}$$

`Exercise 12.1.6.` Use (12.2) to show that two equivalent binary quadratic forms have the same discriminant.

12.2. Equivalence classes of binary quadratic forms. Now $29X^2+82XY+58Y^2$ is equivalent to x^2+y^2 so when we are considering representations, it is surely easier to work with the latter form rather than the former. Gauss observed that every equivalence class of binary quadratic forms (with $d < 0$) contains a unique reduced representative, where the quadratic form $ax^2 + bxy + cy^2$ with discriminant $d < 0$ is *reduced* if

$$-a < b \leq a \leq c, \text{ and } b \geq 0 \text{ whenever } a = c.$$

For a reduced binary quadratic form, $|d| = 4ac - (|b|)^2 \geq 4a \cdot a - a^2 = 3a^2$ and hence

$$a \leq \sqrt{|d|/3}.$$

Therefore for a given $d < 0$ there are only finitely many a, and so b (as $|b| \leq a$), but then $c = (b^2-d)/4a$ is determined, and so there are only finitely many reduced binary quadratic forms of discriminant d. Hence $h(d)$, the *class number*, which is the number of equivalence classes of binary quadratic forms of discriminant d, is finite when $d < 0$. In fact $h(d) \geq 1$ since we always have the *principal* form (for both positive and negative discriminants),

$$\begin{cases} x^2 - (d/4)y^2 & \text{when } d \equiv 0 \pmod 4, \\ x^2 + xy + \frac{(1-d)}{4}\, y^2 & \text{when } d \equiv 1 \pmod 4. \end{cases}$$

`Exercise 12.2.1.` Show that there are no other binary quadratic forms $x^2+bxy+cy^2$, up to equivalence.

Theorem 12.1. *Every positive definite binary quadratic form is properly equivalent to a reduced form.*

Proof. We will define a sequence of properly equivalent forms; the algorithm terminates when we reach one that is reduced. Given a form (a, b, c):[18]

i) If $c < a$ the transformation $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix}$, yields the form $(c, -b, a)$ which is properly equivalent to (a, b, c).

ii) If $b > a$ or $b \leq -a$ then select b' to be the least residue, in absolute value, of $b \pmod{2a}$, so that $-a < b' \leq a$, say $b' = b - 2ka$. Hence the transformation matrix will be $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & -k \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix}$. The resulting form (a, b', c') is properly equivalent to (a, b, c).

iii) If $c = a$ and $-a < b < 0$ then we use the transformation $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix}$ yielding the form $(a, -b, a)$.

If the resulting form is not reduced then repeat the algorithm. If none of these hypotheses holds then one can easily verify that the form is reduced. To prove that the algorithm terminates in finitely many steps we follow the leading coefficient a: a starts as a positive integer. Each transformation of type (i) reduces the size of a. It stays the same after transformations of type (ii) or (iii), but after a type (iii) transformation the algorithm terminates, and after a type (ii) transformation we either have another type (i) transformation, or else the algorithm stops after at most one more transformation. Hence the algorithm finishes in no more than $2a + 1$ steps.

Example: Applying the reduction algorithm to the form $(76, 217, 155)$ of discriminant -31, one finds the sequence of forms $(76, 65, 14), (14, -65, 76), (14, -9, 2), (2, 9, 14), (2, 1, 4)$, the sought after reduced form. Similarly the form $(11, 49, 55)$ of discriminant -19, gives the sequence of forms $(11, 5, 1), (1, -5, 11), (1, 1, 5)$.

The very precise condition in the definition of "reduced" were so chosen because every positive definite binary quadratic form is properly equivalent to a *unique* reduced form, which the enthusiastic reader will now prove:

`Exercise 12.2.2.` (i) Show that the least values taken by the reduced form $am^2 + bmn + cn^2$ with $(m, n) = 1$, are $a \leq c \leq a - |b| + c$, each represented twice (the last four times if $b = 0$). (Hint: One might use the inequality $am^2 + bmn + cn^2 \geq am^2 - |b| \max\{m, n\}^2 + cn^2$, to show that if the value is $am^2 + bmn + cn^2 \leq a - |b| + c$ then $|m|, |n| \leq 1$.)

- (ii) Use this, and exercise 12.1.5, to show that if two different reduced forms are equivalent then they must be $ax^2 + bxy + cy^2$ and $ax^2 - bxy + cy^2$, and thus $a < c$ since these are both reduced.
- (iii) Suppose that $M \in \mathrm{SL}(2, \mathbb{Z})$ transforms one into the other. Given that we know all the representations of a and c by $ax^2 + bxy + cy^2$, use (12.2) to deduce that $M = \pm I$.
- (iv) Deduce that $b = -b$ so that $b = 0$. Therefore no two reduced forms can be equivalent.

Together with Theorem 12.1 this implies that every positive definite binary quadratic form is properly equivalent to a unique reduced form.

---

[18] Which we write for convenience in place of $ax^2 + bxy + cy^2$.

What restrictions are there on the values that can be taken by a binary quadratic form?

Proposition 12.2. *Suppose $d = b^2 - 4ac$ with $(a, b, c) = 1$, and p is a prime.* (i) *If $p = am^2 + bmn + cn^2$ for some integers m, n then d is a square mod $4p$.* (ii) *If d is a square mod $4p$ then there exists a binary quadratic form of discriminant d that represents p.*

Proof. (i) Note that $(m, n)^2 | am^2 + bmn + cn^2 = p$ so that $(m, n) = 1$.

Now $d = b^2 - 4ac \equiv b^2 \pmod 4$, and even mod $4p$ if $p|ac$. If $p|d$ then d is a square mod p and the result then follows unless $p = 2$. But if $2|d = b^2 - 4ac$ then b is even; therefore $d = b^2 - 4ac \equiv 0$ or $4 \pmod 8$ and hence is a square mod 8.

If $p = 2 \nmid acd$ then b is odd, and so $am^2 + bmn + cn^2 \equiv m^2 + mn + n^2 \not\equiv 0 \pmod 2$ as $(m, n) = 1$.

So suppose that $p \nmid 2ad$ and $p = am^2 + bmn + cn^2$. Therefore $4ap = (2am + bn)^2 - dn^2$ and so dn^2 is a square mod $4p$. Now $p \nmid n$ else $p | 4ap + dn^2 = (2am + bn)^2$ so that $p|2am$ which is impossible as $p \nmid 2a$ and $(m, n) = 1$. We deduce that d is a square mod p.

(ii) If $d \equiv b^2 \pmod{4p}$ then $d = b^2 - 4pc$ for some integer c, and so $px^2 + bxy + cy^2$ is a quadratic form of discriminant d which represents $p = p \cdot 1^2 + b \cdot 1 \cdot 0 + c \cdot 0^2$.

12.3. Class number one.

Corollary 12.3. *Suppose that $h(d) = 1$. Then p is represented by the form of discriminant d if and only if d is a square mod $4p$.*

Proof. This follows immediately from Proposition 12.2, since there is just one equivalence class of quadratic forms of discriminant d, and forms in the same equivalence class represent the same integers by exercise 12.1.4.

The proof that the number of reduced forms is finite can also by turned into an algorithm to find all the reduced binary quadratic forms of a given negative discriminant.

Example: If $d = -163$ then $|b| \le a \le \sqrt{163/3} < 8$. But b is odd so $|b| = 1, 3, 5$ or 7. Therefore $ac = (b^2 + 163)/4 = 41, 43, 47$ or 53, a prime, with $a < c$ and hence $a = 1$. However all such forms are equivalent to the principal form, by exercise 12.2.1, and therefore $h(-163) = 1$. This implies, by Corollary 12.3, that if $(-163/p) = 1$ then p can be represented by the binary quadratic form $x^2 + xy + 41y^2$.

`Exercise 12.3.1.` Determine $h(d)$ for $-20 \le d \le -1$ as well as for $d = -43$ and -67.

Typically one restricts attention to *fundamental discriminants*, which means that if $q^2|d$ then $q = 2$ and $d \equiv 8$ or $12 \pmod{16}$. It turns out that the only fundamental $d < 0$ with $h(d) = 1$ are $d = -3, -4, -7, -8, -11, -19, -43, -67, -163$. Therefore, as in the example above, if $p \nmid d$ then

p is represented by $x^2 + y^2$ if and only if $(-1/p) = 1$,
p is represented by $x^2 + 2y^2$ if and only if $(-2/p) = 1$,
p is represented by $x^2 + xy + y^2$ if and only if $(-3/p) = 1$,
p is represented by $x^2 + xy + 2y^2$ if and only if $(-7/p) = 1$,
p is represented by $x^2 + xy + 3y^2$ if and only if $(-11/p) = 1$,
p is represented by $x^2 + xy + 5y^2$ if and only if $(-19/p) = 1$,

p is represented by $x^2 + xy + 11y^2$ if and only if $(-43/p) = 1$,
p is represented by $x^2 + xy + 17y^2$ if and only if $(-67/p) = 1$,
p is represented by $x^2 + xy + 41y^2$ if and only if $(-163/p) = 1$.

Euler noticed that the polynomial x^2+x+41 is prime for $x = 0, 1, 2, \dots, 39$, and similarly the other polynomials above. Rabinowiscz proved that this is an "if and only if" condition; that is

Rabinowiscz's criterion. *We have $h(1-4A) = 1$ for $A \geq 2$ if and only if x^2+x+A is prime for $x = 0, 1, 2, \dots, A-2$.*

Note that $(A-1)^2 + (A-1) + A = A^2$. We will prove Rabinowiscz's criterion below

The proof that the above list gives all of the $d < 0$ with $h(d) = 1$ has an interesting history. By 1934 it was known that there is no more than one further such d, but that putative d could not be ruled out by the method. In 1952, Kurt Heegner, a German school teacher proposed an extraordinary proof that there are no further d; at the time his paper was ignored since it was based on a result from an old book (of Weber) whose proof was known to be incomplete. In 1966 Alan Baker gave a very different proof that was acknowledged to be correct. However, soon afterwards Stark realized that the proofs in Weber are easily corrected, so that Heegner's work had been fundamentally correct. Heegner was subsequently given credit for solving this famous problem, but sadly only after he had died. Heegner's paper contains a most extraordinary construction, widely regarded to be one of the most creative and influential in the history of number theory, that we will discuss again in section H2 on elliptic curves.

What about when the class number is not one? In the first example, $h(-20) = 2$, the two reduced forms are x^2+5y^2 and $2x^2+2xy+3y^2$. By Proposition 12.2(i), p is represented by at least one of these two forms if and only if $(-5/p) = 0$ or 1, that is, if $p \equiv 1, 3, 7$ or $9 \pmod{20}$ or $p = 2$ or 5. So can we tell which of these primes are represented by which of the two forms? Note that if $p = x^2 + 5y^2$ then $(p/5) = 0$ or 1 and so $p = 5$ or $p \equiv \pm 1 \pmod 5$, and thus $p \equiv 1$ or $9 \pmod{20}$. If $p = 2x^2 + 2xy + 3y^2$ then $2p = (2x+y)^2 + 5y^2$ and so $p = 2$ or $(2p/5) = 1$, that is $(p/5) = -1$, and hence $p \equiv 3$ or $7 \pmod{20}$. Hence we have proved

p is represented by $x^2 + 5y^2$ if and only if $p = 5$, or $p \equiv 1$ or $9 \pmod{20}$;
p is represented by $2x^2 + 2xy + 3y^2$ if and only if $p = 2$, or $p \equiv 3$ or $7 \pmod{20}$.

That is, we can distinguish which primes can be represented by which binary quadratic form of discriminant -20 through congruence conditions, despite the fact that the class number is not one. However we cannot always distinguish which primes are represented by which binary quadratic form of discriminant d. It is understood how to recognize those discriminants for which this is the case, indeed these *idoneal numbers* were recognized by Euler. He found 65 of them, and no more are known – it is an open conjecture as to whether Euler's list is complete. It is known that there can be at most one further idoneal number.

Proof of Rabinowiscz's criterion. We begin by showing that $f(n) := n^2 + n + A$ is prime for $n = 0, 1, 2, \dots, A-2$, if and only if $d = 1 - 4A$ is not a square mod $4p$ for all primes $p < A$. For if $n^2 + n + A$ is composite, let p be its smallest prime factor so that $p \leq f(n)^{1/2} < f(A-1)^{1/2} = A$. Then $(2n+1)^2 - d = 4(n^2+n+A) \equiv 0 \pmod{4p}$ so that d is a square mod $4p$. On the other hand if d is a square mod $4p$ where p is a prime $\leq A-1$, select n to be the smallest integer ≥ 0 such that $d \equiv (2n+1)^2 \bmod 4p$. Then

$0 \le n \le p-1 \le A-2$, and p divides n^2+n+A with $p < A = f(0) < f(n)$ so that n^2+n+A is composite.

Now we show that $h(d) = 1$ if and only if $d = 1-4A$ is not a square mod $4p$ for all primes $p < A$. If $h(d) > 1$ then there exists a reduced binary quadratic $ax^2+bxy+cy^2$ of discriminant d with $1 < a \le \sqrt{|d|/3}$. If p is a prime factor of a then $p \le a < A$ and $d = b^2-4ac$ is a square mod $4p$. On the other hand if d is a square mod $4p$, and $h(d) = 1$ then p is represented by $x^2+xy+Ay^2$ by Proposition 12.2(ii). However the smallest values represented by this form are 1 and A, by exercise 12.2.2(i), and this gives a contradiction since $1 < p < A$. Hence $h(d) > 1$.