INTRODUCTION TO *GAUSS'S* NUMBER THEORY

The optional sections

Abstract. These are the first drafts of the many optional sections. Since we are close to the middle of the semester it seemed to better to provide these now rather than wait any longer. However they are not complete, and not polished. Nonetheless they will hopefully be useful.

A. Elementary

A1. Fibonacci numbers and linear recurrence sequences. The *Fibonacci numbers*, $F_0 = 0, F_1 = 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, \dots$ appear in many places in mathematics and its applications, especially in sequences whose evolution depends on their past. This is because the rule that the terms in the sequence depend on the sequence's own recent history:

$$F_n = F_{n-1} + F_{n-2} \quad \text{for all } n \geq 2.$$

It is not difficult to find a formula for F_n:

$$F_n = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^n - \left(\frac{1-\sqrt{5}}{2} \right)^n \right) \quad \text{for all } n \geq 0. \tag{A1.1}$$

`Exercise A1.1.` Prove this is correct by verifying that it holds for $n = 0, 1$ and then by induction.

In more generality if a, b, x_0, x_1 are given and

$$x_n = ax_{n-1} + bx_{n-2} \quad \text{for all } n \geq 2,$$

then there exist coefficients c_α, c_β such that

$$x_n = c_\alpha \alpha^n + c_\beta \beta^n \quad \text{for all } n \geq 0, \tag{A1.2}$$

where $x^2 - ax - b = (x - \alpha)(x - \beta)$ assuming that $\alpha \neq \beta$. We determine c_α and c_β so that

$$c_\alpha + c_\beta = x_0 \quad \text{and} \quad c_\alpha \alpha + c_\beta \beta = x_1.$$

The result follows by induction on n: It is evidently true for $n = 0$ and 1 by the definitions of c_α and c_β. For given $n \geq 2$, by the induction hypothesis

$$\begin{aligned} ax_{n-1} + bx_{n-2} &= a(c_\alpha \alpha^{n-1} + c_\beta \beta^{n-1}) + b(c_\alpha \alpha^{n-2} + c_\beta \beta^{n-2}) \\ &= c_\alpha \alpha^{n-2}(a\alpha + b) + c_\beta \beta^{n-2}(a\beta + b) \\ &= c_\alpha \alpha^{n-2} \cdot \alpha^2 + c_\beta \beta^{n-2} \cdot \beta^2 = c_\alpha \alpha^n + c_\beta \beta^n, \end{aligned}$$

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

as $a\alpha + b = \alpha^2$ and $a\beta + b = \beta^2$.

`Exercise A1.2.` Show that if $\alpha \neq \beta$ with $x_0 = 0, x_1 = 1$ then $x_n = \frac{\alpha^n - \beta^n}{\alpha - \beta}$ for all $n \geq 0$.

`Exercise A1.3.` Prove that $\alpha = \beta$ if and only if $a^2 + 4b = 0$, and then $\alpha = a/2$ and $x_n = (cn + x_0)\alpha^n$ where $c = x_1/\alpha - x_0$. Deduce that if $\alpha = \beta$ with $x_0 = 0, x_1 = 1$ then $x_n = n\alpha^{n-1}$ for all $n \geq 0$.

An alternate view on these recurrences is via generating functions:

$$\begin{aligned}(1 - x - x^2)\sum_{n\geq 0} F_n x^n &= F_0 + (F_1 - F_0)x + (F_2 - F_1 - F_0)x^2 + \dots \\ &\quad + (F_n - F_{n-1} - F_{n-2})x^n + \dots \\ &= 0 + (1 - 0)x + (1 - 1 - 0)x^2 + \dots + 0 \cdot x^n + \dots \;=\; x.\end{aligned}$$

Hence if $\alpha = \frac{1+\sqrt{5}}{2}$ and $\beta = \frac{1-\sqrt{5}}{2}$ then

$$\begin{aligned}\sum_{n\geq 0} F_n x^n \;&= \frac{x}{1 - x - x^2} = \frac{1}{\alpha - \beta}\left(\frac{\alpha x}{1 - \alpha x} - \frac{\beta x}{1 - \beta x}\right) \\ &= \frac{1}{\alpha - \beta}\left(\sum_{m\geq 1} \alpha^m x^m - \sum_{m\geq 1} \beta^m x^m\right) = \sum_{m\geq 1} \frac{\alpha^m - \beta^m}{\alpha - \beta}\, x^m,\end{aligned}$$

and the result follows, again.

Both of these methods generalize to arbitrary linear recurrences of degree n.

Theorem A1.1. *Suppose that $a_1, a_2, \dots, a_d$ and $x_0, x_1, \dots, x_{d-1}$ are given, and that*

$$x_n = a_1 x_{n-1} + a_2 x_{n-2} + \dots + a_d x_{n-d} \quad \textit{for all } n \geq d.$$

Suppose that $X^d - a_1 X^{d-1} - a_2 X^{d-2} + \dots - a_{d-1}X - a_d = \prod_{j=1}^{k} (X - \alpha_j)^{e_j}$. Then there exist polynomials $P_1, \dots, P_k$ where P_j has degree $e_j - 1$ such that

$$x_n \;=\; \sum_{j=1}^{k} P_j(n)\alpha_j^n \quad \textit{for all } n \geq 0. \tag{A1.3}$$

Moreover the coefficients of the P_j (and hence the polynomials p_j themselves) can all be determined by solving the linear equations obtained by taking this for $n = 0, 1, 2, \dots, d-1$.

`Exercise A1.4.` Prove this one way or another.

Other recurrences. It is easily seen that the recurrence $x_{n+1} = 2x_n + 1$ with $x_0 = 0$ is satisfied by $x_n = 2^n - 1$.

`Exercise A1.5.` Find a formula for x_n if x_0 is given and $x_{n+1} = ax_n + b$. (Hint: If $a \neq 1$ write $b = (a-1)c$ and add c to both sides. Treat the case $a = 1$ separately.)

Division sequences. Recall that we saw that the Mersenne number M_r divides M_{kr} for all positive integers k and r, so that M_{kr} is composite.

`Exercise A1.6.` Show that if $M_0 = 0, M_1 = 1$ and $M_n = 3M_{n-1} - 2M_{n-2}$ then M_n is the nth Mersenne number.

Any sequence of integers $x_0, x_1, \dots$ with the property that x_m divides x_{km} for all positive integers k and m is called a *division sequence*. We have seen that the Mersenne numbers from a division sequence; this is true for all second order linear recurrence sequences that begin $0, 1$:

Proposition A1.2. *Suppose that a, b are given integers with $x_0 = 0$ and $x_1 = 1$, and $x_n = ax_{n-1} + bx_{n-2}$ for all $n \geq 2$. Then x_m divides x_{km} for all positive integers k and m.*

`Exercise A1.7.` Proof. First prove, by induction, that $x_{m+\ell} \equiv x_{m+1}x_\ell \pmod{x_m}$ for all $\ell \geq 0$. Then prove the assertion by induction on $k \geq 1$.

Corollary A1.3. *Suppose that $x_0, x_1, \dots$ is a division sequence, and $|x_n|$ is increasing as $n \to \infty$. If $|x_n|$ is prime then n is prime.*

Thus, for example, if $2^n - 1$ or F_n is prime then n is prime.

Proof. Now $|x_1| > |x_0| \geq 0$ and so $|x_1| \geq 1$. Similarly $|x_2| \geq |x_1| + 1 \geq 2$, and $|x_m| \geq m$ for all $m \geq 0$ by induction. Now if $n = mk$ is composite then $|x_m|$ divides $|x_n|$, and so $|x_n|/|x_m|$ is an integer, which is > 1 as the sequence increases, and hence $|x_n| = |x_m| \cdot |x_n|/|x_m|$ is composite.

It is conjectured that there are infinitely many Mersenne primes $2^p - 1$ as well as Fibonacci primes F_p. There are 33 known Fibonacci primes. The first few are $F_3 = 2,\ F_4 = 3,\ F_5 = 5,\ F_7 = 13,\ F_{11} = 89,\ F_{13} = 233,\ F_{17} = 1597,\ F_{23} = 28657, \dots$ Notice that $F_{19} = 4181 = 37 \times 113$ is composite.

We will see later that solutions to Pell's equation, and the co-ordinates of points on elliptic curves yield division sequences.

The Fibonacci numbers mod p. If $(5/p) = 1$ then there exists $b \bmod p$ such that $b^2 \equiv 5 \pmod p$. By induction one easily proves that

$$F_n \equiv \frac{1}{b}\left(\left(\frac{1+b}{2}\right)^n - \left(\frac{1-b}{2}\right)^n\right) \pmod p \text{ for all } n \geq 0.$$

In particular note that

$$F_{p-1} \equiv \frac{1}{b}(1-1) \equiv 0 \equiv F_0 \pmod p \text{ and } F_p \equiv \frac{1}{b}\left(\frac{1+b}{2} - \frac{1-b}{2}\right) \equiv 1 \equiv F_1 \pmod p,$$

by Fermat's Little Theorem.

`Exercise A1.8.` Deduce that if $(5/p) = 1$ then $F_n \pmod p$ is periodic of period dividing $p - 1$.

We cannot proceed this way when $(5/p) = -1$.

Exercise A1.9. Prove that it is impossible for $F_n \equiv F_{n+1} \equiv 0 \pmod{p}$, and so can exclude this case.

There are only $p^2 - 1$ non-zero pairs $(F_n, F_{n+1}) \pmod{p}$ and so at least two of the pairs for $n = 0, 1, \ldots, p^2 - 1$ must be the same.

Exercise A1.10. Deduce that if $(5/p) = -1$ then $F_n \pmod{p}$ is periodic of period $\leq p^2 - 1$.

Since each binomial coefficient $\binom{p}{j}$ for $1 \leq j \leq p-1$ is divisible by p, hence

$$(x+y)^p \;=\; \sum_{j=0}^{p} \binom{p}{j} x^j y^{p-j} \;\equiv\; x^p + y^p \pmod{p}.$$

One can apply this, using Euler's criterion, to note that

$$(1+\sqrt{5})^p \;\equiv\; 1^p + \sqrt{5}^p \;=\; 1 + 5^{\frac{p-1}{2}}\sqrt{5} \;\equiv\; 1 + \left(\frac{5}{p}\right)\sqrt{5} \;\equiv\; 1 - \sqrt{5} \pmod{p}.$$

Exercise A1.11. Deduce that if $(5/p) = -1$ $F_p \equiv -1$, $F_{p+1} \equiv 0 \pmod{p}$, $F_{p+2} \equiv -1 \pmod{p}$ and then $F_{2p+2} \equiv 0 \pmod{p}$, $F_{2p+3} \equiv 1 \pmod{p}$ using exercise A1.7. Deduce further that the period of $F_n \pmod{p}$ divides $2p + 2$.

A2. Formulae for sums of powers of integers. As a five year old, Gauss quickly added up the numbers from 1 to 100, by noting that $1+100 = 2+99 = 3+98 = \ldots = 99+2 = 100+1 = 101$, so that $1+2+\ldots+100$ equals $\frac{1}{2}$ of 100 times 101. In general one has the formula

$$\sum_{n=0}^{N-1} n = \frac{(N-1)N}{2}\ .$$

Similarly one has

$$\sum_{n=0}^{N-1} n^2 = \frac{(N-1)N(2N-1)}{6} \quad \text{and} \quad \sum_{n=0}^{N-1} n^3 = \left(\frac{(N-1)N}{2}\right)^2.$$

`Exercise A2.1.` Prove the last formula by induction. Then prove it by replacing n by $N-n$ and using the previous two identities.

Are there such formulas for the sums of the kth powers of the integers, for every $k \geq 1$? And, if so, can we easily find the formula? Our first hint for such a formula come from noting that since t^k is an increasing function of t, hence $\int_{n-1}^{n} t^k dt < n^k < \int_n^{n+1} t^k dt$, so that

$$\frac{N^{k+1}}{k+1} = \int_0^N t^k dt \ < \ \sum_{n=1}^{N} n^k \ < \ \int_1^{N+1} t^k dt < \frac{(N+1)^{k+1}}{k+1},$$

so we can be sure that the leading term of the polynomial, if it exists, is $N^{k+1}/(k+1)$.

In fact such a polynomial does exist, but proving that it does is not so straightforward. We will begin with an easier proof that such polynomials exist, but in which it is not easy to identify the polynomials, and then a less intuitive proof that gives the polynomials explicitly.

We begin with some linear algebra. The polynomials of degree d in $\mathbb{R}[x]$ can be viewed as a vector space over $\mathbb{R}$ with basis $\{1, x, x^2, \ldots, x^d\}$. There are many other possible bases for a given vector space; in this case any sequence of polynomials, one of each degree. We choose the binomial coefficients $\{\binom{x}{0}, \binom{x}{1}, \binom{x}{2}, \ldots, \binom{x}{d}\}$ (so that $\binom{x}{k} := \frac{(x(x-1)\ldots(x-k+1)}{k!}$). Therefore there exist rational numbers $a_0, \ldots, a_d$ such that $x^d = \sum_{j=0}^{d} a_j \binom{x}{j}$.

`Exercise A2.2.` Prove that $\sum_{n=0}^{N-1} \binom{n}{k} = \binom{N}{k+1}$. One idea is to do so by induction. Another is to study the coefficients of the identity $(1-t)^{-k-1} \cdot (1-t)^{-1} = (1-t)^{-k-2}$.

We deduce that

$$\sum_{n=0}^{N-1} n^d = \sum_{j=0}^{d} a_j \sum_{n=0}^{N-1} \binom{n}{j} = \sum_{j=0}^{d} a_j \binom{N}{j+1},$$

which equals a polynomial in N of degree $d+1$, once we have written out each binomial coefficient as a polynomial. To compute our summatory polynomial we therefore need to understand these coefficients (the *Stirling numbers of the first kind*), as well as the a_j (the *Stirling numbers of the second kind*), and how to combine them. We shall not do this as there is an easier route to the eventual solution.

To understand the coefficients, we begin by defining the *Bernoulli numbers*, B_n, as the coefficients in the power series:

$$\frac{X}{e^X-1}=\sum_{n\ge 0} B_n\ \frac{X^n}{n!}\ .$$

The first few Bernoulli numbers are $B_0=1, B_1=-\frac{1}{2}, B_2=\frac{1}{6}, B_3=0, B_4=-\frac{1}{30}$, $B_5=0,\ B_6=\frac{1}{42},\ B_7=0,\ B_8=-\frac{1}{30},\ B_9=0,\ B_{10}=\frac{5}{66},\dots$ From this data one can make a few guesses as to what they look like.

`Exercise A2.3.` Prove that the Bernoulli numbers are all rational numbers.

One guesses, from the data that $B_n=0$ if n is odd and >1, and this is easily proved since

$$\sum_{\substack{n\ge 0\\ n\ \text{odd}}} 2B_n\ \frac{X^n}{n!}=\frac{X}{e^X-1}-\frac{(-X)}{e^{-X}-1}=\frac{X}{e^X-1}-\frac{(Xe^X)}{e^X-1}=-X.$$

Other facts include that $(-1)^n B_{2n}<0$ for all $n\ge 1$ and, the more subtle Von Staudt-Clausen Theorem that the set of primes dividing the denominator of B_{2n} is precisely the set of primes p for which $p-1$ divides $2n$, and that the denominator is always squarefree. In fact

$$pB_{2n}\ +\sum_{\substack{p\ \text{prime}\\ p-1|2n}}\frac{1}{p}\quad \text{is an integer for all } n\ge 1. \tag{A2.1}$$

We will see later that the Bernoulli numbers occur in various different areas of number theory.

Next we define the *Bernoulli polynomials*, $B_n(t)$, as the coefficients in the power series:

$$\frac{Xe^{tX}}{e^X-1}=\sum_{n\ge 0} B_n(t)\ \frac{X^n}{n!},$$

and therefore $B_n(0)=B_n$. To verify that these are really polynomials, note that

$$\sum_{k\ge 0} B_k(t)\ \frac{X^k}{k!}=\ e^{tX}\cdot\ \frac{X}{e^X-1}=\ \sum_{m\ge 0}\frac{(tX)^m}{m!}\cdot\sum_{n\ge 0} B_n\ \frac{X^n}{n!}=\ \sum_{m\ge 0}\sum_{n\ge 0} B_n t^m\ \ \frac{X^{m+n}}{m!n!}\ .$$

Here we change variable, writing $k=m+n$, and then the coefficient of X^k, times $k!$, is

$$B_k(t)\ =\ \sum_{\substack{m,n\ge 0\\ m+n=k}}\frac{k!}{m!n!}\ B_n t^m\ =\ \sum_{n=0}^{k}\binom{k}{n}\ B_n t^{k-n}.$$

Theorem A2.1. *For any integers $k \ge 1$ and $N \ge 1$ we have*

$$\sum_{n=0}^{N-1} n^{k-1} = \frac{1}{k}\,(B_k(N) - B_k)$$

Proof. If N is an integer ≥ 1 then

$$\begin{aligned}\sum_{k\ge 0}(B_k(N) - B_k)\,\frac{X^k}{k!} &= \frac{X(e^{NX}-1)}{e^X - 1} = X\sum_{n=0}^{N-1} e^{nX}\\ &= X\sum_{n=0}^{N-1}\sum_{j\ge 0}\frac{(nX)^j}{j!} = \sum_{j\ge 0}\left(\sum_{n=0}^{N-1} n^j\right)\frac{X^{j+1}}{j!}\\ &= \sum_{k\ge 1}\left(k\sum_{n=0}^{N-1} n^{k-1}\right)\frac{X^k}{k!}\end{aligned}$$

taking $k = j+1$. The result follows by comparing the coefficients on both sides.

`Exercise A2.4.` We shall use Theorem A2.1 to partly prove (A2.1), the von-Staudt Clausen Theorem.

(i) Use Corollary 7.9 to show that $\frac{1}{k}\,(B_k(p) - B_k) \equiv 0 \pmod p$ for each $k \ge 1$, unless $k > 1$ and $p-1$ divides $k-1$, in which case it is $\equiv p-1 \pmod p$.
Let us try to prove the result by induction on $k \ge 1$. Suppose that the result is proved for all B_n with $n \le k-1$ and we now try to prove the result for B_k.

(ii) If p does not divide k then deduce that $\frac{1}{k}\,(B_k(p) - B_k) \equiv B_{k-1}p \pmod p$, and thence the von Staudt-Clausen Theorem.

(iii) Explain what remains to be proved.

A3. The number of distinct roots of polynomials. If $f(x) = \sum_{j=0}^{d} f_j x^j$ where $f_d \neq 0$ then $f(x)$ has *degree* d, and *leading coefficient* f_d. We say that $f(x)$ is *monic* if $f_d = 1$.

`Exercise A3.1.` (i) Show that if f is monic and has a rational root then that root must be an integer.
(ii) Show that if f has an integer root n then $f(n) \equiv 0 \pmod m$ for any integer m.
(iii) Show that if f has a rational root r/d, where r and d are coprime integers, and $(d, m) = 1$, then there exists an integer n such that $f(n) \equiv 0 \pmod m$.
(iv) For each integer m give an example of a polynomial f which has a rational root r/m with $(r, m) = 1$, but for which there does not exist an integer n such that $f(n) \equiv 0 \pmod m$. (Hint: $f(x) = 3x + 1$ has the rational root $-\frac{1}{3}$ yet $f(n) \equiv 1 \pmod 3$, for all integers n.)

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 $\mathbb{C}$.*

Proof. By induction. For $d = 1$ we note that $ax + b$ has the unique root $-b/a$. For higher degree, if f has root α then subtract multiples of $x - \alpha$ from $f(x)$ to find polynomials $q(x), r(x)$ such that $f(x) = (x - \alpha)q(x) + r(x)$; here we can assume that $r(x)$ has degree < 1, and so is a constant, which we denote by r. But then $r = r(x) = f(\alpha) = 0$. Now $q(x)$ has degree $d - 1$ so, by induction, has $\leq d - 1$ distinct roots, which implies that $f(x) = (x - \alpha)q(x)$ has $\leq 1 + (d - 1)$ distinct roots.

Proposition A3.1. *Let $\alpha \in \mathbb{C}$ and $f(x) \in \mathbb{Z}[x]$ be the polynomial of minimal degree for which $f(\alpha) = 0$.*[1] *If $g(x) \in \mathbb{Z}[x]$ with $g(\alpha) = 0$ then $f(x)$ divides $g(x)$.*

Proof. There exist $q(x), r(x) \in \mathbb{Z}[x]$ and $k \in \mathbb{Z}$ with $0 \leq \deg r \leq \deg f - 1$, such that $kg(x) = q(x)f(x) + r(x)$. Hence $r(\alpha) = kg(\alpha) - q(\alpha)f(\alpha) = 0$ and r has smaller degree than f, so the only possibility is that $r(x) = 0$. Hence $kg(x) = q(x)f(x)$ and the result follows.

Another result that will be useful is:

Lemma A3.2. *If $f(x), g(x) \in \mathbb{Q}[x]$ are monic and $f(x)g(x) \in \mathbb{Z}[x]$ then $f(x)$ and $g(x) \in \mathbb{Z}[x]$.*

Proof. Suppose that the conclusion is false so that some coefficient of $f(x)$ is not an integer. Let p be a prime dividing the denominator of a coefficient of $f(x)$. Let p^a and p^b be the highest powers of p dividing the denominator of any coefficient of $f(x)$ and $g(x)$, respectively, so that $a \geq 1$ and $b \geq 0$. Therefore we may write $p^a f(x) \equiv f_d x^d + \dots \pmod p$ where $f_d \not\equiv 0 \pmod p$, and similarly $p^b g(x) \equiv g_k x^k + \dots \pmod p$ where $g_k \not\equiv 0 \pmod p$. Now $a + b \geq 1$ and $f(x)g(x) \in \mathbb{Z}[x]$ so that

$$0 \equiv p^{a+b} f(x)g(x) \equiv (f_d x^d + \dots)(g_k x^k + \dots) \equiv f_d g_k x^{d+k} + \dots \pmod p,$$

which implies that p divides $f_d g_k$, a contradiction.

[1] We call f the *minimum polynomial* for α.

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 $f(m) \equiv 0 \pmod p$.*

Proof. If $f(x)$ has degree 1 then we have seen that the congruence has exactly one root $\pmod p$ just after (2.1). We proceed by induction on degree d: Suppose now that $f(x)$ has degree $d \geq 2$ and root $a \pmod p$. Let

$$g(x) = f(x+a) \equiv \sum_{i=0}^{d} g_i x^i \pmod p.$$

Then $g_0 = g(0) = f(a) \equiv 0 \pmod p$ so we may write $g(x) \equiv xh(x) \pmod p$ where $h(x)$ has degree $d-1$. Now suppose $f(y) \equiv 0 \pmod p$. Then $(y-a)h(y-a) \equiv g(y-a) = f(y) \equiv 0 \pmod p$, and so either $y - a \equiv 0 \pmod p$ or $h(y-a) \equiv 0 \pmod p$ by Theorem 3.1. But the number of possible y values $\pmod p$ is then $\leq 1 + (d-1) = d$, by the induction hypothesis.

Notice how close this proof is to the proof of the Fundamental Theorem of Algebra.

`Exercise A3.2.` Suppose that $f(x) \in \mathbb{Z}[x]$ has degree d. Show that $f(x)$ is irreducible if and only if $x^d f(1/x)$ is irreducible. Moreover if $\alpha, \beta, \gamma, \delta \in \mathbb{Z}$ with $\alpha\delta - \beta\gamma = 1$ show that $f(x)$ is irreducible if and only if $(\gamma x + \delta)^d f(\frac{\alpha x+\beta}{\gamma x+\delta})$ is irreducible. (Remark: The easy way to prove this uses the generators of $\mathrm{SL}(2,\mathbb{Z})$ — see section C5)

Cyclotomic polynomials. Let $\zeta_m = e^{2i\pi/m}$. Then $\zeta^m = e^{2i\pi} = 1$, and so $(\zeta^j)^m = (\zeta^m)^j = 1$ for all integers j. Hence the ζ^j are all mth roots of unity.

`Exercise A3.3.` Show that $\zeta^i = \zeta^j$ if and only if $i \equiv j \pmod m$.

Therefore $1, \zeta, \zeta^2, \ldots, \zeta^{m-1}$ are distinct and so denote all of the m roots of $x^m - 1$, by the Fundamental Theorem of Algebra. We call them the mth roots of unity. If α is an mth root of unity, but not an rth root of unity for any r, $1 \leq r \leq m-1$ then α is a *primitive* mth root of unity.

Now suppose that α is an mth root of unity and let r be the minimal integer ≥ 1 for which $\alpha^r = 1$. Selecting integers u, v such that $um + vr = \gcd(r,m)$ we have $\alpha^{\gcd(r,m)} = (\alpha^m)^u(\alpha^r)^v = 1$; and so $\gcd(r,m) = r$ by the minimality of r, that is r divides m. We define here (differently from section 7.9) the *cyclotomic polynomials*

$$\phi_m(x) := \prod_{\substack{\alpha \text{ a primitive} \\ m\text{th root of unity}}} (x - \alpha).$$

Every root of $\phi_m(x)$ is a root of $x^m - 1$, and so if d divides m then $\phi_d(x)$ divides $x^d - 1$, which divides $x^m - 1$. The polynomials $\phi_d(x)$ all have distinct roots and so $\prod_{d|m} \phi_d(x)$ divides $x^m - 1$. On the other hand the roots of $x^m - 1$ are all mth roots of unity, and hence are each a primitive dth root of unity for some d dividing m. Since $x^m - 1$ has no

repeated roots (else the root would also be a root of its derivative, mx^{m-1}), we deduce that

$$x^m - 1 = \prod_{d|m} \phi_d(x).$$

In section 7.9 we saw that $\phi_m(x)$ has degree $\phi(m)$. In fact the roots can be written more explicitly as $\{\zeta^j : 1 \le j \le m \text{ and } (j,m) = 1\}$.

We call p a *primitive prime factor* of $a^m - 1$ if p divides $a^m - 1$ but does not divide $a^r - 1$ for any $1 \le r \le m-1$. In other words a has order m mod p and so $p \equiv 1 \pmod m$. Note also that p divides $\phi_m(a)$ but not $\phi_r(a)$ for any $1 \le r \le m-1$.

Proposition A3.3. *If prime p divides $\phi_m(a)$ then either $p \equiv 1 \pmod m$ or p divides m.*

Proof. Suppose that prime p divides $\phi_m(a)$. Then p divides $a^m - 1$, and hence $\mathrm{ord}_p(a)$ divides m by Lemma 7.2. If $\mathrm{ord}_p(a) = m$ then $m = \mathrm{ord}_p(a)$ divides $p-1$, that is $p \equiv 1 \pmod m$. If, on the other hand, $\mathrm{ord}_p(a) = d < m$ then p divides $a^d - 1$, and p divides $\phi_m(a)$ which divides $\frac{a^m-1}{a^d-1}$. Therefore

$$\begin{aligned} 0 \equiv \frac{a^m-1}{a^d-1} &= 1 + a^d + a^{2d} + a^{3d} + \ldots + a^{m-d} \\ &\equiv 1 + 1 + 1 + \ldots + 1 = \frac{m}{d} \qquad (\mathrm{mod}\ a^d - 1), \end{aligned}$$

and hence mod p, as p divides $a^d - 1$. Hence p divides m/d which divides m.

`Exercise A3.4.` (i) Prove that $\phi_m(0) = -1$ or 1.
(ii) Show that if $p|a$ then $\phi_m(a) \equiv \pm 1 \pmod p$, and so $p \not| \phi_m(a)$.
(iii) Deduce that if $m|a$ and prime $p|\phi_m(a)$ then $p \equiv 1 \pmod m$.

Primitive prime factors of linear recurrence sequences. Suppose that a, b are coprime integers with $x_0 = 0$ and $x_1 = 1$, and $x_n = ax_{n-1} + bx_{n-2}$ for all $n \ge 2$. Then

$$(x_n, b) = (ax_{n-1}, b) = (x_{n-1}, b) = \ldots = (x_1, b) = 1,$$

and so

$$(x_n, x_{n-1}) = (bx_{n-2}, x_{n-1}) = (x_{n-1}, x_{n-2}) = \ldots = (x_1, x_0) = 1.$$

Now

$$x_m = x_{m+1}x_0 + x_m x_1 \text{ and } x_{m+1} = x_{m+1}x_1 + x_m x_0,$$

so that $x_{m+2} = x_{m+1}x_2 + x_m x_1$ and hence, by induction,

$$x_{m+\ell} = x_{m+1}x_\ell + x_m x_{\ell-1} \text{ for all } \ell \ge 1. \tag{A3.1}$$

`Exercise A3.5.` Use this to reprove exercise A1.7.

We know that $x_m | x_{km}$ for all k. In the next exercise we show how to understand p-divisibility of second order linear recurrences that do not begin with $0, 1$.

Exercise A3.6. Suppose $(by_0, y_1) = 1$, and $y_n = ay_{n-1} + by_{n-2}$ for all $n \geq 2$.

(1) Show that $y_{r+\ell} = y_{r+1}x_\ell + y_r x_{\ell-1}$ for all $r \geq 0,\ \ell \geq 1$.
Suppose that q is an integer which divides $x_m,\ y_r$ with $m \geq 1,\ r \geq 0$ minimal.
(2) Prove that $(q, y_{r+1}) = (q, x_{m-1}) = 1$.
(3) Deduce that $q|y_{r+n}$ if and only if $q|x_n$, and that $q|y_{n+m}$ if and only if $q|y_n$.
(4) Finally deduce that $q|y_n$ if and only if $n \equiv r \pmod{m}$.

A4. Binomial coefficients, Lucas's Theorem etc, self-similarity. One of the first objects of number theory interest that one encounters are the binomial coefficients

$$\binom{n}{m} = \frac{n!}{m!(n-m)!} = \frac{n(n-1)\cdots(n-m+1)}{m(m-1)\cdots 2\cdot 1}$$

which is the number of different ways of choosing m objects from n. It is surprising that these counts should give numbers that factor in this way, and that they are the coefficients in the binomial theorem,

$$(x+y)^n = \sum_{m=0}^{n} \binom{n}{m} x^{n-m} y^m.$$

This formula implies that each binomial coefficient is a non-negative integer, which is not obvious when one encounters the definition, above, of one big product divided by another.

Upper bounds. Taking $x = y = 1$ in the binomial theorem one deduces easily that

$$\binom{n}{m} \le \sum_{j=0}^{n} \binom{n}{j} = 2^n.$$

`Exercise A4.1.` Use the fact that $\binom{n}{m} = \binom{n}{n-m}$ to prove that if $n \neq 2m$ then $\binom{n}{m} \le 2^{n-1}$. If $n = 2m$ then prove this inequality by comparing $\binom{n}{m}$ with $\binom{n}{m-1} + \binom{n}{m+1}$.

Lower bounds. Through somewhat more involved arguments we now give a lower bound on the same product.

`Exercise A4.2.` Prove that $\binom{n}{m+1} \ge \binom{n}{m}$ if and only if $n - m \ge m + 1$. Deduce that the maximum of $\binom{n}{m}$ as m varies is attained when m is the closest integer to $n/2$ (that is $m = [n/2]$ and $[(n+1)/2]$).

Now $\binom{n}{0} + \binom{n}{n} = 1 + 1 = 2 \le \binom{n}{1} \le \binom{n}{[n/2]}$ if $n \ge 2$. Hence

$$2^n = \sum_{m=0}^{n} \binom{n}{m} \le 2 + \sum_{m=1}^{n-1} \binom{n}{m} \le n\binom{n}{[n/2]},$$

by the last exercise.

The prime powers dividing a given binomial coefficient.

`Exercise A4.3.` Show that there are exactly $[n/q]$ integers $m,\ 1 \le m \le n$ that are divisible by q, for any integers $q, n \ge 1$.

Lemma A4.1. *The largest power of prime p that divides $n!$ is $\sum_{k\ge 1}[n/p^k]$. In other words*

$$n! = \prod_{p \text{ prime}} p^{\left[\frac{n}{p}\right]+\left[\frac{n}{p^2}\right]+\left[\frac{n}{p^3}\right]+\cdots}$$

Proof. We wish to determine the power of p dividing $n! = 1\cdot 2\cdot 3\cdots(n-1)\cdot n$. If p^k is the power of p dividing m then we will count 1 for p dividing m, then 1 for p^2 dividing m, ... , and finally 1 for p^k dividing m. Therefore the power of p dividing $n!$ equals the number of integers $m,\ 1 \le m \le n$ that are divisible by p, plus the number of integers $m,\ 1 \le m \le n$ that are divisible by p^2, plus etc. The result then follows from the previous exercise.

Kummer's Theorem. *The largest power of prime p that divides the binomial coefficient $\binom{a+b}{a}$ is given by the number of carries when adding a and b in base p.*

Example: To recover the factorization of $\binom{14}{6}$ we add 6 and 8 in each prime base ≤ 14:

$$\begin{array}{cccccc} 0101 & 020 & 11 & 06 & 06 & 06 \\ 1000_2 & 022_3 & 13_5 & 11_7 & 08_{11} & 08_{13} \\ \hline 1101 & 112 & 24 & 20 & 13 & 11 \end{array}$$

We see that there are no carries in base 2, 1 carry in base 3, no carries in base 5, 1 carry in base 7, 1 carry in base 11, and 1 carry in base 13, so we deduce that $\binom{14}{6} = 3^1 \cdot 7^1 \cdot 11^1 \cdot 13^1$.

Proof. For given integer $k \ge 1$, let $q = p^k$. Then let A and B be the least non-negative residue of a and $b \pmod q$, respectively, so that $0 \le A, B \le q-1$. Note that A and B give the first k digits (from the right) of a and b in base p. If C is the first k digits of $a+b$ in base p then C is the least non-negative residue of $a+b \pmod q$, that is of $A+B \pmod q$. Now $0 \le A+B < 2q$:

• If $A+B < q$ then $C = A+B$ and there is no carry in the kth digit when we add a and b in base p.

• If $A+B \ge q$ then $C = A+B-q$ and so there is a carry of 1 in the kth digit when we add a and b in base p.

We need to relate these observations to the formula in the lemma. The trick comes in noticing that $A = a - p^k \left[\frac{a}{p^k}\right]$, and similarly $B = b - p^k \left[\frac{b}{p^k}\right]$ and $C = a+b-p^k \left[\frac{a+b}{p^k}\right]$. Therefore

$$\left[\frac{a+b}{p^k}\right] - \left[\frac{a}{p^k}\right] - \left[\frac{b}{p^k}\right] = \frac{A+B-C}{p^k} = \begin{cases} 1 & \text{if there is a carry in the kth digit;} \\ 0 & \text{if not;} \end{cases}$$

and so

$$\sum_{k \ge 1} \left(\left[\frac{a+b}{p^k}\right] - \left[\frac{a}{p^k}\right] - \left[\frac{b}{p^k}\right] \right)$$

equals the number of carries when adding a and b in base p. However this equals the exact power of p dividing $\frac{(a+b)!}{a!b!} = \binom{a+b}{a}$ by lemma A4.1, and the result follows.

Corollary A4.2. *If p^e divides the binomial coefficient $\binom{n}{m}$ then $p^e \le n$.*

Proof. There are $k+1$ digits in the base p expansion of n when $p^k \le n < p^{k+1}$. When adding m and $n-m$ there can be carries in every digit except the $(k+1)$st (which corresponds to the number of multiples of p^k). Therefore that are no more than k carries when adding m to $n-m$ in base p, and so the result follows from Kummer's Theorem.

Kummer's Theorem shows us how to determine what power of prime p divides a given binomial coefficient, and the next result shows us how to find the value mod p when the binomial coefficient is not divisible by p. It is convenient to define $\binom{n}{m} = 0$ when $m > n$.

Lucas' Theorem. *Write $n = n_0 + n_1 p + \ldots + n_d p^d$ in base p (so that $0 \le n_j \le p-1$ for each j) and similarly $m = m_0 + m_1 p + \ldots + m_d p^d$. Then*

$$\binom{n}{m} \equiv \binom{n_0}{m_0}\binom{n_1}{m_1}\cdots\binom{n_d}{m_d} \pmod{p}.$$

Proof. Using Fermat's Little Theorem $(1+x)^{p^j} \equiv 1 + x^{p^j} \pmod{p}$, and so

$$\begin{aligned}\sum_{m=0}^{n}\binom{n}{m}x^m = (1+x)^n &= \prod_{j=0}^{d}(1+x)^{n_j p^j} \\ &\equiv \prod_{j=0}^{d}(1+x^{p^j})^{n_j} \pmod{p} \equiv \prod_{j=0}^{d}\sum_{m_j=0}^{n_j}\binom{n_j}{m_j}x^{m_j p^j} \\ &= \sum_{\substack{0\le m_j\le n_j \\ \text{for } j=0,1,2,\ldots,d}} \binom{n_0}{m_0}\binom{n_1}{m_1}\cdots\binom{n_d}{m_d}\, x^{m_0+m_1 p+\ldots+m_d p^d} \pmod{p}.\end{aligned}$$

The result follows by comparing the coefficient of x^m on either side.

We have seen that $(x+1)^p \equiv x^p + 1 \pmod{p}$. Hence

$$\sum_{j=0}^{p-1}\binom{p-1}{j}x^j \;=\; (x+1)^{p-1} \;=\; \frac{(x+1)^p}{x+1} \equiv \frac{x^p+1}{x+1} \;=\; \sum_{j=0}^{p-1}(-x)^j \pmod{p},$$

so that

$$\binom{p-1}{j} \equiv (-1)^j \pmod{p}.$$

Another proof simply comes from expanding $\binom{p-1}{j} = \frac{p-1}{1}\frac{p-2}{2}\cdots\frac{p-j}{j} \equiv (-1)^j \pmod{p}$, since each $\frac{p-i}{i} \equiv -1 \pmod{p}$. One nice consequence is that

$$\left(\frac{p-1}{2}\right)!^2 \;=\; (p-1)!\Big/\binom{p-1}{\frac{p-1}{2}} \equiv (-1)^{\frac{p+1}{2}} \pmod{p}$$

using Wilson's Theorem. Hence if $p \equiv 1 \pmod 4$ then $\left(\frac{p-1}{2}\right)!$ is a square root of $-1 \pmod{p}$.

The binomial coefficients $\binom{p}{m}, 1 \le m \le p-1$ are all divisible by p. If we divide through by p we obtain

$$\frac{1}{p}\binom{p}{m} = \frac{(p-1)(p-2)\ldots(p-(m-1))}{m!} \equiv \frac{(-1)^{m-1}}{m} \pmod{p}.$$

Therefore

$$\mathcal{L}_p(1-x) := \frac{(1-x)^p - 1 + x^p}{p} = \frac{1}{p}\sum_{m=1}^{p-1}\binom{p}{m}(-x)^m \;\equiv\; -\sum_{m=1}^{p-1}\frac{x^m}{m} \pmod{p}.$$

For example if $x = -1$ then $\frac{2^p-2}{p} \equiv -\sum_{m=1}^{p-1} \frac{(-1)^m}{m} \pmod p$, and also if $x = 2$ then $\frac{2^p-2}{p} \equiv -\sum_{m=1}^{p-1} \frac{2^m}{m} \pmod p$. Note that $\mathcal{L}_p(1-x)$ is a truncation of the expansion for the logarithm function, $\log(1-x) = -\sum_{m\ge 1} \frac{x^m}{m}$. There is a further connection: If $\ell_p(x) := \frac{x^{p-1}-1}{p}$ so that $x^{p-1} = 1 + p\ell_p(x)$ then

$$1 + p\ell_p(ab) = (ab)^{p-1} = (1 + p\ell_p(a))(1 + p\ell_p(b)) \equiv 1 + p(\ell_p(a) + \ell_p(b)) \pmod{p^2},$$

and so $\ell_p(ab) \equiv \ell_p(a) + \ell_p(b) \pmod p$, much like the logarithm function. It is not obvious what the connection is between these two appearances of the logarithm function!

A5. Taking powers efficiently, P and NP. How can we raise an integer to the nth power "quickly", when n is very large? In 1785 Legendre computed high powers mod p by a method that we now call *fast exponentiation*: Suppose we have to determine $5^{65} \pmod{161}$. We begin by writing 65 in base 2, that is $65 = 2^6 + 2^1$. Let $f_0 = 5$, $f_1 \equiv f_0^2 \equiv 5^2 \equiv 25 \pmod{161}$, $f_2 \equiv f_1^2 \equiv 25^2 \equiv 142 \pmod{161}$, $f_3 \equiv 142^2 \equiv 39 \pmod{161}$, $f_4 \equiv 72$, $f_5 \equiv 32$, $f_6 \equiv 58 \pmod{161}$ and so $5^{65} = 5^{64+1} \equiv f_6 \cdot f_0 \equiv 58 \cdot 5 \equiv 129 \pmod{161}$. We have determined the value of $5^{65} \pmod{161}$ in seven multiplications mod 161, as opposed to 65 multiplications by the more obvious algorithm.

In general to compute $a^n \pmod m$ quickly: Define $f_0 = a$ and then $f_j \equiv f_{j-1}^2 \pmod m$ for each $j = 1, 2, \dots, j_1$, where j_1 is the largest integer for which $2^{j_1} \le n$.

Writing n in binary, say as $n = 2^{j_1} + 2^{j_2} + \cdots + 2^{j_\ell}$ with $j_1 > j_2 > \cdots > j_\ell \ge 0$, let $g_1 = f_{j_1}$ and then $g_i \equiv g_{i-1} f_{j_i} \bmod m$ for $j = 2, 3, \dots, \ell$. Therefore

$$g_\ell \equiv f_{j_1} \cdot f_{j_2} \cdots f_{j_\ell} \equiv a^{2^{j_1}+2^{j_2}+\cdots+2^{j_\ell}} = a^n \pmod m.$$

This involves $j_\ell + \ell - 1 \le 2j_\ell \le \frac{2 \log n}{\log 2}$ multiplications mod m as opposed to n by the more obvious algorithm.

Running time: The inputs into this algorithm are the integers a and m and the exponent n. We may assume that $1 \le a \le m$, and that all of the residues f_j and g_i can be taken to lie in $[1, m]$. If m has d digits when written down (so that d is proportional to $\log m$) then the usual multiplication algorithm, multiplying digit by digit, requires roughly d^2 steps, and reducing an integer with no more than $2d + 1$ digits, mod m, requires about $2d$ steps. Therefore the total number of steps in the above algorithm is proportional to

$$(\log m)^2 \log n.$$

The length D of the inputs is the number of digits when we write them down (in base 2 or 10), and so is proportional to $\log (mn)$. Hence the running time of the algorithm is no more than some constant times D^3, that is a polynomial in D. We therefore call this a *polynomial time algorithm*. We cannot hope for an algorithm to take less time than it takes to read the inputs, so any polynomial time algorithm is considered to be pretty fast.

`Exercise A5.1.` Show that there are polynomial time algorithms for both addition and multiplication.

`Exercise A5.2.` Prove that the Euclidean algorithm works in polynomial time.

One should distinguish between the mathematical problem and the algorithm for resolving the problem. There may be many choices of algorithm and one wishes, of course, to find a fast one. We denote by `P` the class of problems that can be resolved by an algorithm that runs in polynomial time. There are very few mathematical problems which belong to `P`.

In section 10.5 we discussed problems that have been resolved and for which the answer can be quickly checked. For example one can exhibit factors of a given integer n to give a short proof that n is composite. We also saw Lucas' short proof that a number is prime based on the fact that only prime numbers n have primitive roots generating $n - 1$

elements. By "short" we mean that the proof can be verified in polynomial time, and we say that such problems are in the class NP ("*non-deterministic polynomial time*"[2]). In these cases we are not suggesting that the proof can be found in polynomial time, only that the proof can be checked in polynomial time; indeed we have no idea whether it is possible to factor numbers in polynomial time, and this is now the outstanding number theory problem of this area.

By definition P⊆NP; and of course we believe that there are problems, for example the factoring problem, which are in NP, but not in P; however *this has not been proved*, and it is now perhaps the outstanding unresolved question of theoretical computer science. This is another of the Clay Mathematics Institute's million dollar problems, and perhaps the most likely to be resolved by someone with less formal training, since the experts seem to have few plausible ideas for attacking this question.

It had better be the case that P≠NP, else there is little chance that one can have safe public key cryptography (see, e.g., section 10.3) or that one could build a highly unpredictable (pseudo-)random number generator[3], or that we could have any one of several other necessary software tools for computers. Notice that one implication of the "P≠NP" question remaining unresolved is that no fast public key cryptographic protocol is, as yet, provably safe!

Difficult problems: There are only a finite number of possible commands for each line of a computer program, which therefore induce a finite number of possible states for the number and values of the variables.[4] It is therefore easy to show that most problems need exponential length programs to be solved:

We consider the set of problems where we input N bits and output one bit, that is functions

$$f:\ \{0,1\}^N \to \{0,1\}.$$

Since there are 2^N possible inputs, and for each the function can have two possible outputs, hence the number of such functions is 2^{2^N}.

If a computer language allows M different possible statements, then the number of programs containing k lines is M^k, and this is therefore a bound on the number of functions that can be calculated by a computer program that is k lines long. Therefore if $k < c_M \cdot 2^N$, where $c_M := \frac{\log 2}{2\log M}$ then this accounts for $\le 2^{2^{N-1}}$ functions; hence the vast majority of such problems require a program of length at least $c_M \cdot 2^N$. (Notice here that M, and thus C_M, is fixed by the computer, and N is varying).

Since almost all problems require such long programs, exponential in the length of the input, one would think that it would be easy to specify problems that needed longish

[2]Note that NP is **not** "non-polynomial time", a common source of confusion. In fact it is "non-deterministic" because the method for discovering the proof is not necessarily determined.

[3]So-called "random number generators" written in computer software are not random since they need to work on a computer where everything is designed to be determined! Thus what are called "random numbers" are typically a sequence of numbers, determined in a totally predictable manner, but which appear to be random when subjected to "randomness tests" in which the tester does not know how the sequence was generated.

[4]Here we are talking about a classical computer. As yet impractical *quantum computers* face less restrictions and thus, perhaps, will allow more things to be computed rapidly.

programs. However this is a wide open problem. Indeed even finding specific problems that cannot be resolved in polynomial time is open, or even problems that really require more than linear time! This is the pathetic state of our knowledge on lower bounds for running times, in practice. So if you ever hear claims that some secret code is provably difficult to break, that your secrets are perfectly safe, then either there has been a major scientific breakthrough, or you are hearing salesmanship, not mathematical proof.

A6. Solving the cubic. The roots of a quadratic polynomial $ax^2 + bx + c = 0$ are

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

The easy way to prove this is to put the equation into a form that is easy to solve. One begins by dividing through by a, to get $x^2+(b/a)x+c/a = 0$, so that the leading coefficient is 1. Next we make a change variable, letting $y = x + b/2a$ to obtain

$$y^2 - (b^2 - 4ac)/4a^2 = 0.$$

Having removed the y^1 term, we can simply take square-roots to obtain the possibilities for y, and hence the possible values for x.

We call $\Delta := b^2 - 4ac$ the *discriminant* of our polynomial. Note that if $f(x) = ax^2 + bx + c$ then $f'(x) = 2ax + b$. We apply the Euclidean algorithm on these two polynomials: $2f(x) - xf'(x) = bx + 2c$ and so $2a(bx + 2c) - b(2ax + b) = -\Delta$, which yields

$$\Delta = -4a(ax^2 + bx + c) + (2ax + b)^2.$$

Thus Δ is the smallest positive integer in the ideal generated by f and f' over $\mathbb{Z}[a, b, x]$.

Can one similarly find the roots of a cubic? We can certainly begin the same way.

`Exercise A6.1.` Show that we can easily deduce the roots of any given cubic polynomial, from the roots of some cubic polynomial of the form $x^3 + ax + b$.

We wish to find the roots of $x^3 + ax + b = 0$. This does not look so easy since we cannot simply take cube roots unless $a = 0$. Cardano's trick (1545) is a little surprising: Write $x = u + v$ so that

$$x^3 + ax + b = (u + v)^3 + a(u + v) + b = (u^3 + v^3 + b) + (u + v)(3uv + a).$$

This equals 0 when $u^3 + v^3 = -b$ and $3uv = -a$; in other words

$$u^3 + v^3 = -b \text{ and } u^3v^3 = -a^3/27. \tag{1}$$

Hence $(X - u^3)(X - v^3) = X^2 + bX - a^3/27$ and so, using the formula for the roots of a quadratic polynomial, yields

$$u^3, v^3 = \frac{-b \pm \sqrt{b^2 + 4a^3/27}}{2}$$

Taking cube roots yield values for u and v for which (1) holds but it is not clear that $3uv = -a$. Indeed what we do have is that if $\alpha = -3uv/a$ then $\alpha^3 = -27u^3v^3/a^3 = 1$ by (1), and so α is one of the three cube roots of unity, and not necessarily 1. To rectify this we replace v by $\alpha^2 v$. Hence the roots of $x^3 + ax + b$ are given by

$$u + v,\ \omega u + \omega^2 v,\ \omega^2 u + \omega v,$$

where ω is a primitive cube root of unity. Now the discriminant is

$$\Delta := 4a^3 + 27b^2 = (6ax^2 - 9bx + 4a^2)(3x^2 + a) - 9(2ax - 3b)f(x),$$

where $f(x) = x^3 + ax + b$, the smallest positive integer in the ideal generated by f and f' over $\mathbb{Z}[a, b, x]$, and so $u^3, v^3 = \frac{-b \pm \sqrt{\Delta/27}}{2}$.

The important thing to notice here is that the solution to a cubic is given in terms of both cube roots and square roots, not just cube roots.

How about the roots of a quartic polynomial? Can these be found in terms of fourth roots, cube roots and square roots? And similarly roots of quintics and higher degree polynomials?

The general quartic: We begin, as above, by rewriting the equation in the form $x^4 + ax^2 + bx + c = 0$. Following Ferrari (1550s) we add an extra variable y to obtain the equation

$$(x^2 + a + y)^2 = (a + 2y)x^2 - bx + ((a + y)^2 - c),$$

and then select y to make the right side the square of a linear polynomial in x (and so we would have $(x^2 + a + y)^2 = (rx + s)^2$ and hence x can be deduced as a root of one of the quadratic polynomials $(x^2 + a + y) \pm (rx + s)$). The right side is a square of a linear polynomial if and only if its discriminant is 0, that is $b^2 - 4(a+2y)((a+y)^2 - c) = \Delta = 0$. But this is a cubic in y, and we have just seen how to find the roots of a cubic polynomial.

Example: We want the roots of $X^4 + 4X^3 - 37X^2 - 100X + 300$. Letting $x = X + 1$ yields $x^4 - 43x^2 - 18x + 360$. Proceeding as above leads to the cubic equation $2y^3 - 215y^2 + 6676y - 64108 = 0$. Dividing through by 2 and then changing variable $y = t + 215/6$ gives the cubic $t^3 - (6169/12)t - (482147/108) = 0$. This has discriminant $-4(6169/12)^3 + 27(482147/108)^2 = -(2310)^2$. Hence $u^3, v^3 = (482147 \pm 27720\sqrt{-3})/216$. Unusually this has an exact cube root in terms of $\sqrt{-3}$; that is $u, v = \omega^*(-37 \pm 40\sqrt{-3})/6$. Now $-3(-37 + 40\sqrt{-3})/6 \cdot (-37 - 40\sqrt{-3})/6 = -6169/12 = a$. Therefore we can take $u, v = (-37 \pm 40\sqrt{-3})/6$, and the roots of our cubic are $t = u + v = -37/3$, $\omega u + \omega^2 v = 157/6$, $\omega^2 u + \omega v = -83/6$ so that $y = 47/2, 62, 22$. From these Ferrari's equation becomes $(x^2 - 39/2) = \pm(2x + 9/2)$ for $y = 47/2$ and so the possible roots $-5, 3; -4, 6$; or $(x^2 + 19) = \pm(9x + 1)$ for $y = 62$ and so the possible roots $-5, -4; 3, 6$; or $(x^2 - 21) = \pm(x + 9)$ for $y = 22$ and so the possible roots $-5, 6; 3, -4$. For each such y we get the same roots $x = 3, -4, -5, 6$, yielding the roots $X = 2, -5, -6, 5$ of the original quartic.

Example: Another fun example is to find the fifth roots of unity. That is those x satisfying $\phi_5(x) = \frac{x^5 - 1}{x - 1} = x^4 + x^3 + x^2 + x + 1 = 0$. Proceeding as above we find the roots

$$\frac{\sqrt{5} - 1 \pm \sqrt{-2\sqrt{5} - 10}}{4}, \quad \frac{-\sqrt{5} - 1 \pm \sqrt{2\sqrt{5} - 10}}{4} \quad .$$

Example: Gauss's favourite example was the expression in surds was for $\cos \frac{2\pi}{2^k}$, which we will denote by $c(k)$. A double angle formula states that $\cos 2\theta = 2\cos^2 \theta - 1$, and so taking

$\theta = 2\pi/2^k$ we have $c(k-1) = 2c(k)^2 - 1$, or $c(k) = \frac12\sqrt{2+2c(k-1)}$. Note that $c(k) \ge 0$ for $k \ge 2$ and $c(2) = 0$. Hence $c(3) = \frac12\sqrt{2}$, $c(4) = \frac12\sqrt{2+\sqrt{2}}$, $c(5) = \frac12\sqrt{2+\sqrt{2+\sqrt{2}}}$ and, in general,

$$\cos\left(\frac{2\pi}{2^k}\right) = \frac12 \underbrace{\sqrt{2+\sqrt{2+\sqrt{2+\sqrt{2+\dots\sqrt{2}}}}}}_{k-2 \text{ times}} \qquad \text{for each } \ k \ge 3.$$

Why surds? (A *surd* is a number of the form $n^{1/k}$ where n and k are positive integers.) Why do we wish to express roots of all polynomials in terms of square roots, cube roots, etc.? That is, surds. After all, is a solution in terms of $\sqrt{7}$ any more enlightening then in terms of the second-largest root of $x^5 - 3x^2 + 2x - 11$? By this I mean we have an expression that gives each of these numbers "exactly", though that expression is not something that is a solid integer, just something that one can approximate speedily and accurately. But there are methods to quickly approximate the roots of any given polynomial to any desired level of accuracy, so why the obsession with surds? The answer is more aesthetic than anything else – we have a comfort level with surds that we do not have with more complicated expressions. One can rephrase the question: *Can we describe the roots of any given polynomial, $x^d + a_1x^{d-1} + \dots + a_d$ as a polynomial, with rational coefficients, in roots of polynomials of the form $x^k - n$?* One can see that this is probably wishful thinking since we wish to express a root that is given in terms of d coefficients, in terms of something much simpler, and it is perhaps miraculous that we have succeeded with all polynomials where $d \le 4$. After this discussion it may not come as such a surprise that there are degree five polynomials whose roots cannot be expressed as a rational expression in surds. This was understood by Gauss in 1804, but waited for a magnificent proof by Galois in 1829 at the age of 18. More on that in a moment.

The theory of symmetric polynomials. It is difficult to work with algebraic numbers since one cannot necessarily evaluate them precisely. However for many of the reasons we use them we do not need to actually work with complex numbers, but rather we work with the set of roots of a polynomial. It was Sir Isaac Newton who recognized the following result. We say that $P(x_1, x_2, \dots, x_n)$ is a *symmetric polynomial* if $P(x_k, x_2, \dots, x_{k-1}, x_1, x_{k+1}, \dots, x_n) = P(x_1, x_2, \dots, x_n)$ for each k.

`Exercise A6.2.` Show that for any permutation σ if $1, 2, \dots, n$ and any symmetric polynomial P we have $P(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}) = P(x_1, x_2, \dots, x_n)$.

The fundamental theorem of symmetric polynomials. *Let $f(x) = \prod_{i=1}^d (x - \alpha_i) = \sum_{i=0}^d a_i x^i$ be a monic polynomial with integer coefficients. Then any symmetric polynomial in the roots of f can be expressed as a polynomial in the a_i.*

Let us see the idea. First, we know by multiplying out $\prod_{i=1}^d (x - \alpha_i)$ that

$$\sum_i \alpha_i = -a_1, \ \sum_{i<j} \alpha_i\alpha_j = a_2, \ \sum_{i<j<k} \alpha_i\alpha_j\alpha_k = -a_3, \ \dots, \ \alpha_1\alpha_2\dots\alpha_n = \pm a_n.$$

Now define $s_k := \sum_{i=1}^d \alpha_i^k$. Since $\frac{f'(x)}{f(x)} = \sum_{i=1}^d \frac{1}{x-\alpha_i}$ have

$$\frac{\sum_{j=0}^d j a_j x^{d-j}}{\sum_{i=0}^d a_i x^{d-i}} = \frac{x^{d-1}}{x^{d-1}} \cdot \frac{f'(1/x)}{x f(1/x)} = \sum_{i=1}^d \frac{1}{1-\alpha_i x} = \sum_{i=1}^d \sum_{k\geq 0} (\alpha_i x)^k = \sum_{k\geq 0} s_k x^k.$$

This implies that $\sum_{j=0}^d (d-j) a_{d-j} x^j = \sum_{i=0}^d a_{d-i} x^i \cdot \sum_{k\geq 0} s_k x^k$, so that, comparing the coefficients of x^k, we obtain (as $a_d = 1$)

$$s_k = -\sum_{i=1}^{\min\{d,k\}} a_{d-i} s_{k-i} + \begin{cases} (d-k)a_{d-k} & \text{if } k < d; \\ 0 & \text{if } k \geq d. \end{cases}$$

Hence, by induction on k, we see that the s_k are polynomials in the a_j.

`Exercise A6.3.` If f is not monic, develop analogous results by working with $g(x)$ defined by $g(a_d x) = a_d^{d-1} f(x)$.

Now that we have obtained the s_k, we can prove Newton's result for arbitrary symmetric polynomials, by showing that every symmetric polynomial is a polynomial in the s_k, which implies the theorem. We proceed by induction on the number of variables in the monomials of the symmetric polynomial. The result for the s_k is precisely the case where each monomial has one variable. Now, for the proof by induction, suppose that the symmetric polynomial under question has monomial $\alpha_{i_1}^{k_1} \alpha_{i_2}^{k_2} \dots \alpha_{i_r}^{k_r}$ with each $k_i \geq 1$ and summed over all possibilities of $i_1, i_2, \dots, i_r$ being distinct elements of $1, 2, \dots, n$. We subtract $s_{k_1} s_{k_2} \dots s_{k_r}$,[5] and we are left with various cross terms, in which two or more of the variables α_j are equal. Hence in the remaining expression each monomial contains fewer variables and the result follows by induction.

Example: Look at $\sum_{i,j,k} \alpha_i \alpha_j^2 \alpha_k^3$. Subtract $s_1 s_2 s_3$ and we have to account the cases where $i = j$ or $i = k$ or $j = k$. Hence what remains is $-\sum_{i,k} \alpha_i^3 \alpha_k^3 - \sum_{i,j} \alpha_i^4 \alpha_j^2 - \sum_{i,j} \alpha_i \alpha_j^5 + 2s_6$ where in the first sum we have $i = j$, in the second $i = k$, then $j = k$ and $i = j = k$. Proceeding the same way again we have $\sum_{i,j} \alpha_i^4 \alpha_j^2 = s_4 s_2 - s_6$, $\sum_{i,j} \alpha_i \alpha_j^5 = s_1 s_5 - s_6$ and $\sum_{i,k} \alpha_i^3 \alpha_k^3 = (s_3^2 - s_6)/2$, the last since in s_3^2 the cross term $\alpha_i^3 \alpha_k^3$ appears also as $\alpha_k^3 \alpha_i^3$. Collecting this all together yields $\sum_{i,j,k} \alpha_i \alpha_j^2 \alpha_k^3 = s_1 s_2 s_3 - s_1 s_5 - s_2 s_4 - s_3^2/2 + 9s_6/2$. Notice that in each term here the sum of the indices is 6, the degree of the original polynomial.

Some special cases: If α is a root of an irreducible polynomial $f(x) = a \prod_{i=1}^d (x - \alpha_i)$ then there are two particular symmetric polynomials of the roots of special interest:

The *trace* of α is $\alpha_1 + \alpha_2 + \ldots + \alpha_d$, the sum of the roots of f.

The *norm* of α is $\alpha_1 \alpha_2 \dots \alpha_d$, the product of the roots of f.

[5] Actually this is only really correct if the k_j are distinct. To correct we divide through by $\prod_i m_i!$ where m_i is the number of k_j that equal i.

A7. Constructibility. The ancient Greeks were interested in what could be constructed using only an unmarked ruler (i.e. a straight edge) and a compass. Three questions stumped them:

(1) *Quadrature of the circle*: Draw a square that has area equal to that of a given circle.
(2) *Duplication of the cube*: Construct a cube that has twice the volume of a given cube.
(3) *Trisection of the angle*: Construct an angle which is one third the size of a given angle.

Let us formulate these problems algebraically.

(1) The area of the square is π, so we need to be able to find a root of $x^2 - \pi$.

(2) Starting with a cube of side length 1, the new cube would have side length $2^{1/3}$; in other words we need to find the real root of $x^3 - 2$.

(3) Constructing an angle θ as is difficult as constructing a right angled triangle containing that angle, and so the triangle with side lengths $\sin\theta$, $\cos\theta$, 1. Hence if we start with angle 3θ and wish to determine θ, we will need to be able to determine $\cos\theta$ from $\cos 3\theta$ and $\sin 3\theta$. But these are linked by the formula $\cos 3\theta = 4\cos^3\theta - 3\cos\theta$, that is we need to find the root $x = 2\cos\theta$ of $x^3 - 3x - A$ where $A = 2\cos 3\theta$. So if $\theta = \pi/9$ this yields the equation $x^3 - 3x - 1$.

The first is impossible because π is transcendental (something we may prove later). The next two, whether we can construct the roots of the polynomials $x^3 - 2$ and $x^3 - 3x - 1$, we shall discuss now:

Our first goal is to understand the algebra of a new point constructed from given points and lengths.

Proposition A7.1. *Given a set of known points on known lines, and a set of lengths, any new points that can be constructed using ruler and compass will have coordinates that can be determined as roots of degree one or two polynomials whose coefficients are rational functions of the already known coordinates.*

Proof. Lines are defined by pairs of points: Given the points $A = (a_1, a_2)$ and $B = (b_1, b_2)$ the line between them is $(b_1 - a_1)(y - a_2) = (b_2 - a_2)(x - a_1)$.

`Exercise A7.1.` Show that the coefficients of the equation of this line can be determined by a degree one equation in already known coordinates.

`Exercise A7.2.` Prove that any two (non-parallel) lines intersect in a point that can be determined by a degree one equation in the coefficients of the equations of the lines.

Given a point $C = (c_1, c_2)$ and a radius r, we can draw a circle $(x-c_1)^2+(y-c_2)^2 = r^2$.

`Exercise A7.3.` Prove that the points of intersection of this circle with a given line can be given by a degree two equation in already known coordinates. (Hint: Substitute the value of y given by the line, into the equation of the circle.)

`Exercise A7.4.` Prove that the points of intersection of two circles can be given by a degree two equation in already known coordinates. (Hint: Subtract the equations for the two circles.)

So to show that one cannot duplicate the cube, or trisect an angle, we need to have a theory that shows that the roots of irreducible polynomials of degree three cannot be determined in terms of a (finite) succession of roots of linear or quadratic polynomials whose coefficients are already constructed. This is the beginning of Galois theory.

Fields. A field is a set of objects amongst which we can apply the usual operations of arithmetic (i.e. addition, subtraction, multiplication and division).[6] In number theory, the basic field is the set of rationals, $\mathbb{Q}$. We can adjoin an irrational to $\mathbb{Q}$, like $\sqrt{2}$, to obtain $\mathbb{Q}(\sqrt{2})$, the set of all arithmetic expressions in $\sqrt{2}$ with rational coefficients.

`Exercise A7.5.` Show that $\mathbb{Q}(\sqrt{2}) = \{r + s\sqrt{2} : \ r, s \in \mathbb{Q}\}$.

One can even adjoin several irrationals to $\mathbb{Q}$, for example to obtain $\mathbb{Q}(2^{1/2}, 3^{1/3}, 5^{1/5}, 7^{1/7})$. One might ask whether there exists α such that this field can be written as $\mathbb{Q}(\alpha)$ (so that every element of the field can be given as a polynomial in α with rational coefficients)? If so then the field extension is *simple*, but this is not always the case. In our cases of interest, like in the example of the fifth roots of unity, we start with $K = \mathbb{Q}(\sqrt{5})$, and then the fifth roots of unity all live in $L = K(\sqrt{-2\sqrt{5} - 10}, \sqrt{2\sqrt{5} - 10})$. In fact if we call these two elements α, β then $\alpha\beta = 4\sqrt{5} \in K$, and so $L = K(\alpha)$ so L/K is a simple extension.

`Exercise A7.6.` Verify that $\sqrt{5} = -\alpha^2/2 - 5$ and show how to find β as a function of α with rational coefficients. Deduce that $L/\mathbb{Q}$ is a simple extension.

The degree of the *field extension* $\mathbb{Q}(\alpha)$ of $\mathbb{Q}$ is the degree of the minimal polynomial for α over $\mathbb{Q}$. Since any field extension K of $\mathbb{Q}$ can be obtained as $K = \mathbb{Q}(\alpha_1, \alpha_2, \dots, \alpha_n)$, let $K_0 = \mathbb{Q}$, $K_1 = \mathbb{Q}(\alpha_1)$, $K_2 = K_1(\alpha_2), \dots, K_n = K$, we obtain the degree of $K/\mathbb{Q}$ as the product of the degrees of K_{j+1}/K_j. If L is a subfield of K then we can find numbers $\beta_1, \dots, \beta_m$ such that $K = L(\beta_1, \dots, \beta_m)$.

With this we can give a more precise description of constructibility: A number γ is *constructible* (by ruler and compass) if there exists a field K such that $\gamma \in K$ and K can be written as $K = \mathbb{Q}(\alpha_1, \alpha_2, \dots, \alpha_n)$ where each K_{j+1}/K_j has degree 2. Hence the degree of K of $\mathbb{Q}$ is a power of 2. Let $L = \mathbb{Q}(\gamma)$. We see from the above that the degree of $L/\mathbb{Q}$, times the degree of K/L equals the degree of $K/\mathbb{Q}$, which is a power of 2. Hence the degree of $L/\mathbb{Q}$ must itself be a power of 2; that is the degree of the minimum polynomial of γ must be a power of 2. We deduce that if the minimum polynomial of γ has degree 3 then γ is not constructible.

`Exercise A7.7.` Deduce that one cannot duplicate the cube nor trisect the angle $\pi/3$. (You will need to show that the relevant cubic polynomials are irreducible over $\mathbb{Z}$. To do this you might use Lemma A3.2.)

As we have discussed in section 11.3, an *algebraic number* is a number $\alpha \in \mathbb{C}$ which satisfies a polynomial with integer coefficients. An *algebraic integer* is a number $\alpha \in \mathbb{C}$ which satisfies a *monic* polynomial with integer coefficients.

`Exercise A7.8.` Let $f(x)$ be a polynomial in $\mathbb{Z}[x]$ of minimal degree for which $f(\alpha) = 0$, where the gcd of the coefficients of f is 1.

(1) Show that if $g(x) \in \mathbb{Z}[x]$ with $g(\alpha) = 0$ then $f(x)|g(x)$.

[6] Technically, the objects are organized into both additive and multiplicative groups — see section B4 for more details.

(2) Deduce that $f(x)$ is well-defined and unique, and so can be called the *minimum polynomial* of α.
(3) Show that if f has leading coefficient a then $a\alpha$ is an algebraic integer.
(4) Show that if $g(x) \in \mathbb{Z}[x]$ with $g(\alpha) = 0$ is monic then α is an algebraic integer.

If α is an algebraic integer then so is $m\alpha + n$ for any integers m, n; for if $f(x)$ is the minimal polynomial of α and has degree d then $F(x) := m^d f(\frac{x-n}{m})$ is a monic polynomial in $\mathbb{Z}[x]$ with root $m\alpha + n$.

Suppose that α and β are algebraic integers with minimal polynomials f and g. Then

$$\prod_{\substack{u:\ f(u)=0\\ v:\ g(v)=0}} (x - (u+v)) = \prod_{u:\ f(u)=0} g(x-u).$$

By the fundamental theorem of symmetric polynomials this has rational coefficients, and so $\alpha + \beta$ is an algebraic number.

`Exercise A7.9.` Prove that $\alpha\beta$ is an algebraic number.

In the fundamental theorem of arithmetic we ignored negative integers. If we seek to generalize the fundamental theorem then we cannot do this. The right way to think about this is that every non-zero integer is of the form un where n is a positive integer and $u = -1$ or 1. These two values for u are the only integers that divide 1, and it is for this reason they are a bit exceptional. In general we define a *unit* to be an algebraic integer that divides 1, that is an algebraic integer u is a unit if and only if there exists an algebraic integer v such that $uv = 1$.

`Exercise A7.10.` Show that if $f(x)$, the minimum polynomial for u, has degree d, then $x^d f(1/x)$ is the minimum polynomial for $1/u$. Deduce that u is a unit if and only if $f(0)$ equals 1 or -1.

A8. Resultants and Discriminants. (should be in section B) In Theorem 3.8 we showed that all solutions m, n to $am + bn = c$ are given by

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

given some initial solution r, s. Are there solutions with $(m, n) = 1$? The first thing to note is that (m, n) divides $am + bn = c$, so $(m, n) = 1$ if and only if for each prime factor p of c we have that $p \not| m$ or $p \not| n$. Now $(\frac{a}{(a,b)}, \frac{b}{(a,b)}) = 1$ so p does not divide at least one of them, say $p \not| \frac{a}{(a,b)}$. Then, by the remarks are Corollary 3.6, there exists a residue class $\ell_p \pmod p$ such that $s - \ell_p \frac{a}{(a,b)} \equiv 1 \pmod p$. (And there is an analogous construction when $p \not| \frac{b}{(a,b)}$.) Now taking $\ell \equiv \ell_p \pmod p$ for each prime p dividing c, we will obtain pairwise coprime integers m, n for which $am + bn = c$. Or, given one solution m, n, we can find infinitely many solutions to $aM + bN = c$ with $(M, N) = 1$, by taking $M = m + bck,\ N = n - ack$ for any integer k, since $(M, N) = (M, N, c) = (m, n, c) = 1$.

Suppose that we have two polynomials $f(x) = f_0 x^D + \dots$ and $g(x) = g_0 x^d + \dots \in \mathbb{Z}[x]$ where $D \geq d$ and f_0, g_0 are non-zero. We can apply the Euclidean algorithm even in $\mathbb{Z}[x]$, subtracting an appropriate polynomial multiple of the polynomial of smaller degree, from a constant multiple of the polynomial of larger degree, to reduce the degree of the polynomial of larger degree; that is take $h(x) = g_0 f(x) - f_0 x^{D-d} g(x)$ to get a new polynomial in $\mathbb{Z}[x]$ of degree $< D$. Moreover if we define $\gcd(f(x), g(x))$ to be the polynomial in $\mathbb{Z}[x]$ of largest degree that divides both $f(x)$ and $g(x)$, then the same proof as in the integers yields that $\gcd(f(x), g(x)) = \gcd(g(x), h(x))$, so we can iterate our procedure until one of the two entries is 0. Evidently $\gcd(f(x), 0) = f(x)$. Hence this implies (as in the integers) that we have polynomials $a(x), b(x) \in \mathbb{Z}[x]$ such that

$$a(x)f(x) + b(x)g(x) = R \gcd_{\mathbb{Z}[x]}(f(x), g(x))$$

for some constant R. One can show that $\deg a < \deg g$ and $\deg b < \deg f$

The most interesting case for us is when $\gcd(f(x), g(x)) = 1$, that is when f and g have no common root, and we divide any common integer factors out from the three terms, to obtain

$$a(x)f(x) + b(x)g(x) = R,$$

where R is the *resultant* of a and b. Now, let us suppose that there exists an integer m such that $f(m) \equiv g(m) \equiv 0 \pmod p$. Substituting in $x = m$ we see that p divides R. This argument can be generalized, using some algebraic number theory, to show that if f and g have any common factor mod p (not just a linear polynomial) than p divides R.

Now suppose that prime p divides R so that $a(x)f(x) \equiv -b(x)g(x) \pmod p$. Hence $f(x)$ divides $b(x)g(x) \pmod p$, but f has higher degree than b and so it must have some factor in common with $g(x) \pmod p$. Thus we have an "if and only if" criterion:

Proposition A8.1. *Suppose that $f(x), g(x) \in \mathbb{Z}[x]$ have no common roots. Then prime p divides the resultant of f and g if and only if f and g have a common polynomial factor mod p.*

A particularly interesting special case of Proposition A8.1 is where we take $g(x) = f'(x)$. The resultant of f and f' is the discriminant of f. Let us check this: If $f(x) = ax^2 + bx + c$ then $f'(x) = 2ax + b$ and so

$$(2ax + b)(2ax + b) - 4a(ax^2 + bx + c) = b^2 - 4ac.$$

If $f(x) = x^3 + ax + b$ then $f'(x) = 3x^2 + a$ and so

$$9(3b - 2ax)(x^3 + ax + b) + (6ax^2 - 9bx + 4a^2)(3x^2 + a) = 4a^3 + 27b^2.$$

Corollary A8.2. *Suppose that $f(x) \in \mathbb{Z}[x]$ has no repeated roots. Then prime p divides Δ, the discriminant of f if and only if f has a repeated polynomial factor mod p.*

We should also note that the polynomial common factor of highest degree of f and f' can be obtained by using the Euclidean algorithm but can also be described as

$$\gcd_{\mathbb{Z}[t]}(f(t), f'(t)) = c' \prod_{i=1}^{k} (t - \alpha_i)^{e_i - 1}, \text{where } f(t) = c \prod_{i=1}^{k} (t - \alpha_i)^{e_i}$$

and c' divides c. In the case that $f(x)$ has no repeated roots, so that $\gcd_{\mathbb{Z}[t]}(f(t), f'(t)) \in \mathbb{Z}$, let us write

$$a(x)f(x) + b(x)f'(x) = \Delta$$

so that $f'(x) = c \sum_{j=1}^{d} \prod_{1 \le i \le d,\ i \ne j} (x - \alpha_i)$. Hence $f'(\alpha_j) = c \prod_{i:\ i \ne j} (\alpha_j - \alpha_i)$ and note that $\Delta = b(\alpha_j) f'(\alpha_j)$ so that $f'(\alpha_j)$ divides Δ. In fact one can determine the discriminant of f as

$$\pm c^{2d-2} \prod_{1 \le i < j \le d} (\alpha_i - \alpha_j)^2 \ = \ \pm c^{d-2} \prod_{j=1}^{d} f'(\alpha_j).$$

`Exercise A8.1.` By multiplying $f(x)$ through by a constant, establish that if such a formula is true then one must have an initial terms of a^{2d-2}.

`Exercise A8.2.` Show that if $f(t) = \prod_{i=1}^{k} (t - \alpha_i)^{e_i}$ then $\prod_{j=1}^{d} f'(\alpha_j)$ is an integer, by using the theory of symmetric polynomials.

A9. Möbius transformations: Lines and circles go to lines and circles. In both the Euclidean algorithm and in working with binary quadratic forms we have seen maps $(x, y) \to (\alpha x + \beta y, \gamma x + \delta y)$. These *linear transformations* have various nice properties, one of which is that a line is mapped to a line under such transformations.

A *Möbius transformation* acts on the complex plane (plus the "point" ∞). It is a map of the form

$$z \to \frac{\alpha z + \beta}{\gamma z + \delta} \quad \text{where} \quad \alpha\delta - \beta\gamma \neq 0.$$

Hence we see that $\infty \to \alpha/\gamma$ (and $\infty \to \infty$ if $\gamma = 0$), and $-\delta/\gamma \to \infty$.

`Exercise A9.1` Show that if one composes two Möbius transformations one gets another one.

`Exercise A9.2` Show that the Möbius transformations $z \to z + 1$, $z \to -1/z$ and $z \to \lambda z$ compose to give all Möbius transformations.

Let's study the two basic shapes, a line and a circle, and how they map under Möbius transformations. Certainly under translations $z \to z + k$, and dilations $z \to \lambda z$, it is clear geometrically that lines map to lines and circles map to circles.

We now focus on the map $z \to -1/z$. Notice that if we apply the map twice then we get back to the original point: A circle centered at the origin of radius r has equation $|z| = r$, and is mapped to the circle, $|z| = 1/r$, centered at the origin of radius $1/r$.

`Exercise A9.3` Show that A line through the origin has equation $\overline{z} = \alpha z$ where $|\alpha| = 1$. Hence this gets mapped to the line $\overline{z} = (1/\alpha)z$.

`Exercise A9.4` Show that any line in the complex plane that does not go through the origin can be viewed as the set of points equi-distant from 0 and some other point $\alpha \neq 0$.

This last exercise implies that any line that does not go through the origin may be written as $|z| = |\alpha - z|$ for some $\alpha \neq 0$. Under the map $z \to -1/z$ we get $|z - \beta| = |\beta|$ where $\beta = -1/\alpha$, the circle centered at $-1/\alpha$ that goes through the origin. Applying the map again we find that any circle that goes through the origin gets mapped back to a line that does not pass through the origin.

Finally we must deal with circles that do not pass through the origin nor have their centers at the origin; that is $|z - \alpha| = r$, where $|\alpha| \neq 0, r$. Under the map $z \to -1/z$ this goes to $|z - \beta| = t|z|$ where $\beta = -1/\alpha$ and $t = r/|\alpha| \neq 1$.

`Exercise A9.5` Show that if $\beta = -(t^2 - 1)\gamma$ with $t \neq 1$ then $|z - \beta| = t|z|$ is the same as $|z - \gamma| = t|\gamma|$, and is therefore a circle.

`Exercise A9.6` Prove that to determine a Möbius transformation one need only know the pre-images of $0, 1$ and ∞.

A10. Egyptian fractions. The ancient Egyptians represented all fractions as a sum of distinct fractions of the form $1/n$. It is amusing to determine how difficult it is to represent fractions a/b with $(a,b)=1$, as a sum of *Egyptian fractions*. For example if n is odd and $n+1=2m$ then

$$\frac{2}{n}=\frac{1}{m}+\frac{1}{mn}\,.$$

`Exercise A10.1.` Show that a/b may be written as $\sum_{i=1}^{k}1/n_i$ with the n_i distinct. (Hint: One method is to proceed by induction on a. If the result is true for $a-1$ then write a/b as $1/b$ plus $(a-1)/b$, and proceed from there.)

Our goal is to find the shortest such representation.

Note that if $a/b=\sum_{i=1}^{k}1/n_i$ then $a/\ell b=\sum_{i=1}^{k}1/\ell n_i$ so we can focus on prime denominators.

There can be many ways to write a fraction as a sum of Egyptian fractions, for since $1=\frac12+\frac13+\frac16$, we can replace $1/n$ by $1/2n+1/3n+1/6n$.

For two term representations we have denominators gr and gs, say, with $(r,s)=1$. We let $k=(r+s,g)$ so that $r+s=ka$ and $g=kb$ where $(a,b)=1$ and therefore $\frac{1}{gr}+\frac{1}{gs}=\frac{a}{brs}$ where $(a,brs)=1$. Hence if n is coprime to a then a/n can be written as the sum of two Egyptian fractions if and only if n has coprime divisors r,s such that $a|r+s$. If $a=3$ and n is a prime $\equiv 1 \pmod 3$ then no such r,s exist and so we see that there are fractions $3/n$ that cannot be written as the sum of two Egyptian fractions.

Writing $3/n$, with $n>2$, as the sum of three Egyptian fractions, or less, is easy:

– If $3|n$ then we have $\frac{3}{n}=\frac{1}{m}$ where $n=3m$.

– If n has an odd prime divisor $2m-1$ with $m>1$ then, writing $n=r(2m-1)$, we have $\frac{3}{n}=\frac{1}{rm}+\frac{1}{n}+\frac{1}{nm}$, which are distinct as $rm<n<nm$.

– If $n=2^{k+2}$ with $k\ge 0$ then we divide $\frac{3}{4}=\frac12+\frac16+\frac{1}{12}$ through by 2^k.

The Erdős-Strauss conjecture states that $4/n$ can always be written as the sum of three Egyptian fractions or less. This remains open, though it is know to be true for all $n<10^{14}$. We may restrict our attention to when $n=p$ is prime since then any $4/mp$ can be represented as $1/m$ times the representation of $4/p$. We can get representations in several cases:

– If $p=4m-1$ we have $\frac{4}{p}=\frac{1}{m}+\frac{1}{4m^2}+\frac{1}{4m^2p}$.

– If $p=3m-1$ then $\frac{4}{p}=\frac{1}{p}+\frac{1}{m}+\frac{1}{mp}$.

We are left with only with the primes $p\equiv 1 \pmod{12}$.

B. Basics

B1. Linear congruences (material from Gauss).

Composite moduli. If the modulus m is composite then we can solve any linear congruence question, "one prime at a time", as in the following example: To solve

$$19x \equiv 1 \pmod{140}$$

we first do so $\pmod 2$ to get $x \equiv 1 \pmod 2$. Substituting $x = 1 + 2y$ into the original equation we get

$$38y \equiv -18 \pmod{140} \text{ or, equivalently, } 19y \equiv -9 \pmod{70}.$$

Since 2 divides 70 we again view this $\pmod 2$ to get $y \equiv 1 \pmod 2$. Substituting $y = 1+2z$ into this equation we get

$$38z \equiv -28 \pmod{70} \text{ and thus } 19z \equiv -14 \pmod{35}.$$

Viewing this $\pmod 5$ gives $-z \equiv 1 \pmod 5$, and so substitute $z = -1 + 5w$ to get

$$95w \equiv 5 \pmod{35} \text{ so that } 5w \equiv 19w \equiv 1 \pmod 7.$$

Therefore we get $w \equiv 3 \pmod 7$ which implies, successively that

$$z \equiv -1 + 5 \cdot 3 \equiv 14 \pmod{35}, \qquad y \equiv 1 + 2 \cdot 14 \equiv 29 \pmod{70}$$
$$\text{and} \quad x \equiv 1 + 2 \cdot 29 \equiv 59 \pmod{140}.$$

If $(a, m) = 1$ then we can (unambiguously) express the root of $ax \equiv b \pmod m$ as $b/a \pmod m$; we take this to mean any integer $\equiv b/a \pmod m$. For example $19/17 \equiv 11 \pmod{12}$. Such quotients share all the properties described in Lemma 2.2.

Linear congruences with several unknowns. We will restrict our attention to the case that there are as many congruences as there are unknowns. That is we wish to find all integer (vector) solutions $x \pmod m$ to $Ax \equiv b \pmod m$, where A is a given n-by-n matrix of integers, and b is a given vector of n integers.

Let a_i be the ith column vector of A. Let $V_j = \{v \in \mathbb{R}^n : \ v \cdot a_i = 0 \text{ for all } i \neq j\}$. Basic linear algebra gives us that V_j is itself a vector space of dimension $\geq n-(n-1) = 1$, and has a basis made up of vectors with only integer entries. Hence we may take a non-zero vector in V_j with integer entries, and divide through by the gcd of those entries to obtain a vector c_j whose entries are coprime. Therefore $c_j \cdot a_i = 0$ for all $i \neq j$. Let $d_j = c_j \cdot a_j \in \mathbb{Z}$. Let C be the matrix with ith row vector c_i, and D be the diagonal matrix with (j, j)th entry d_j. Then

$$Dx = (CA)x = C(Ax) \equiv Cb = y \pmod m, \text{ say.}$$

This has solutions if and only if there exists a solution x_j to $d_j x_j \equiv y_j \pmod m$ for each j. As we saw earlier there are solutions if and only if (d_j, m) divides (y_j, m) for each j, and we have also seen how to find all solutions.

Example: Given $\begin{pmatrix} 1 & 3 & 1 \\ 4 & 1 & 5 \\ 2 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \equiv \begin{pmatrix} 1 \\ 7 \\ 3 \end{pmatrix} \pmod 8$. Therefore we have

$$\begin{pmatrix} -15 & 0 & 0 \\ 0 & 15 & 0 \\ 0 & 0 & 15 \end{pmatrix} x = \begin{pmatrix} 9 & 1 & -14 \\ 6 & -1 & -1 \\ 6 & 4 & -11 \end{pmatrix} \begin{pmatrix} 1 & 3 & 1 \\ 4 & 1 & 5 \\ 2 & 2 & 1 \end{pmatrix} x$$
$$\equiv \begin{pmatrix} 9 & 1 & -14 \\ 6 & -1 & -1 \\ 6 & 4 & -11 \end{pmatrix} \begin{pmatrix} 1 \\ 7 \\ 3 \end{pmatrix} = \begin{pmatrix} -26 \\ -4 \\ 1 \end{pmatrix} \pmod 8,$$

so that $x \equiv \begin{pmatrix} -2 \\ 4 \\ -1 \end{pmatrix} \pmod 8$. This gives all solutions mod 8.

Example: Given $\begin{pmatrix} 3 & 5 & 1 \\ 2 & 3 & 2 \\ 5 & 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \equiv \begin{pmatrix} 4 \\ 7 \\ 6 \end{pmatrix} \pmod{12}$ we have

$$\begin{pmatrix} 4 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 28 \end{pmatrix} x = \begin{pmatrix} 1 & -2 & 1 \\ 1 & 1 & -1 \\ -13 & 22 & -1 \end{pmatrix} \begin{pmatrix} 3 & 5 & 1 \\ 2 & 3 & 2 \\ 5 & 1 & 3 \end{pmatrix} x$$
$$\equiv \begin{pmatrix} 1 & -2 & 1 \\ 1 & 1 & -1 \\ -13 & 22 & -1 \end{pmatrix} \begin{pmatrix} 4 \\ 7 \\ 6 \end{pmatrix} = \begin{pmatrix} -4 \\ 5 \\ 96 \end{pmatrix} \pmod{12},$$

and so $x_1 \equiv -1 \pmod 3$, $x_2 \equiv -1 \pmod{12}$, $x_3 \equiv 0 \pmod 3$. To obtain all solutions mod 12 we substitute $x_1 = 2+3t$, $x_2 = -1$, $x_3 = 3u$ into the original equations, we obtain $\begin{pmatrix} 3 & 1 \\ 2 & 2 \\ 5 & 3 \end{pmatrix} \begin{pmatrix} t \\ u \end{pmatrix} \equiv \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \pmod 4$ which is equivalent to $t \equiv u - 1 \pmod 4$. So we end up with $x_3 \equiv 0 \pmod 3$ with $x_1 \equiv x_3 - 1$, $x_2 \equiv -1 \pmod{12}$.

B2. The Chinese Remainder Theorem in general.

When the moduli are not coprime. We began section 3.5 by considering two moduli that are not necessarily coprime, but then proved the Chinese Remainder Theorem for pairwise coprime moduli. If we drop the assumption that the moduli are pairwise coprime then the statement of the Theorem becomes more complicated (though one can see that it is a direct generalization of Lemma 3.9):

The Chinese Remainder Theorem revisited. *Suppose that $m_1, m_2, \dots, m_k$ are a set of 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 exist integers x such that $x \equiv a_j \pmod{m_j}$ for each j if and only if $a_i \equiv a_j \pmod{(m_i, m_j)}$ for all $i \neq j$. In this case the integers are those that belong to a unique residue class mod $m = \operatorname{lcm}[m_1, m_2, \dots, m_k]$.*

Proof. By induction on $k \geq 2$. It is proved for $k = 2$ in Lemma 3.9. If there is a solution then $a_i \equiv x \equiv a_j \pmod{(m_i, m_j)}$ for all $i \neq j$. If so then by the induction hypothesis there exists a unique residue class $a_0 \bmod m_0 := [m_2, \dots, m_k]$, for which $a_0 \equiv a_j \pmod{m_j}$ for each $j \geq 2$. Now $a_1 \equiv a_j \equiv a_0 \pmod{(m_1, m_j)}$ for all $j \geq 2$, and so, by exercise 3.1.10 the lcm of the $(m_1, m_j) : \ j \geq 2$, which equals (m_1, m_0), divides $a_1 - a_0$. Hence $a_0 \equiv a_1 \pmod{(m_0, m_1)}$ and so, by Lemma 3.9, there exists a unique residue class $x \bmod \operatorname{lcm}[m_0, m_1] = m$, such that $x \equiv a_1 \pmod{m_1}$ and $x \equiv a_0 \pmod{m_1}$ which is $\equiv a_j \pmod{m_j}$ for each $j \geq 2$.

Example: Can one find integers z for which $z \equiv 17 \pmod{504}$, $z \equiv -4 \pmod{35}$, $z \equiv 1 \pmod{16}$? The first two congruences combine to give $z \equiv 521 \pmod{2520}$. Combining this with the congruence $z \equiv 1 \pmod{16}$, we get $z \equiv 3041 \pmod{5040}$.

Given $a_1 \pmod{m_1}$, $a_2 \pmod{m_2}, \dots, a_k \pmod{m_k}$, how do we find that $x \bmod m = [m_1, m_2, \dots, m_k]$ such that $x \equiv a_j \pmod{m_j}$ for each j ? One idea is to re-write each congruence $x \equiv a_j \pmod{m_j}$ as the set of congruences

$$x \equiv a_j \pmod{p_1^{e_{j,1}}},\ x \equiv a_j \pmod{p_2^{e_{j,2}}}, \dots,\ x \equiv a_j \pmod{p_r^{e_{j,r}}},$$

where $p_1, \dots, p_r$ are the distinct primes dividing m, and $m_j = \prod_{i=1}^{r} p_i^{e_{j,i}}$. Now for each i, we have the set of congruences

$$x \equiv a_1 \pmod{p_i^{e_{1,i}}},\ x \equiv a_2 \pmod{p_i^{e_{2,i}}}, \dots,\ x \equiv a_k \pmod{p_i^{e_{k,i}}},$$

and evidently, if $e_i := \max\{e_{j,i} : \ 1 \leq j \leq k\} = e_{j(i),i}$, then there exists such an x if and only if $a_j \equiv a_{j(i)} \pmod{p_i^{e_{j,i}}}$ for each j. If so then this holds exactly for those $x \equiv a_{j(i)} \pmod{p_i^{e_i}}$, and we now have coprime moduli (the $p_i^{e_i}$) so can use the algorithm described in section 3.5 to construct the congruence class for $x \pmod{m}$.

Using this technique on our earlier example, the congruences may be re-written as

$$\begin{aligned} z \equiv 17 \pmod{504} &\Leftrightarrow z \equiv 17 \pmod{8},\ z \equiv 17 \pmod{9} \ \&\ z \equiv 17 \pmod{7}; \\ z \equiv -4 \pmod{35} &\Leftrightarrow z \equiv -4 \pmod{5} \ \&\ z \equiv -4 \pmod{7}; \\ z \equiv 1 \pmod{16} &\Leftrightarrow z \equiv 1 \pmod{16}. \end{aligned}$$

For each prime we need to verify that the congruences to that prime's powers can all be satisfied simultaneously. We thus get

$$z \equiv 1 \pmod{16},\ z \equiv -1 \pmod{9},\ z \equiv 1 \pmod{5},\ z \equiv 3 \pmod{7}$$

which are consistent with all six congruences above and, with these, we now obtain

$$z \equiv 3041 \pmod{5040} \qquad \text{where} \qquad 5040 = 16 \cdot 9 \cdot 5 \cdot 7.$$

The big problem with the method just described is that we need to factor the moduli m_j to construct our congruence class mod m. It is possible to proceed without factoring, which is preferable since factoring large numbers can often be difficult (as discussed in section 10.3). We show how to do this with two moduli; more moduli can be added by iterating this algorithm:

Suppose that we wish to combine the pair of congruences $x \equiv a \pmod{A}$ and $x \equiv b \pmod{B}$. We have seen that there is a solution if and only if $a \equiv b \pmod{g}$ where $g = (A, B)$. If so then we determine, using the Euclidean algorithm, integers r and s for which $Ar + Bs = g$. Now let $c = a + \frac{(b-a)}{g} \cdot Ar$ and $C = [A, B]$. Evidently $c \equiv a \pmod{A}$, and also $c = b + \frac{(a-b)}{g} \cdot Bs \equiv b \pmod{B}$. Hence $x \equiv a \pmod{A}$ and $x \equiv b \pmod{B}$ if and only if $x \equiv c \pmod{C}$.

`Exercise B2.1.` Prove that $a \equiv b \pmod{m}$ if and only if $\frac{a}{m} - \frac{b}{m}$ is an integer. With an abuse of our notation we can write this as $\frac{a}{m} \equiv \frac{b}{m} \pmod{1}$; or even that $\frac{a}{m} = \frac{b}{m}$ in $\mathbb{R}/\mathbb{Z}$ (that is, $\mathbb{R} \pmod{\mathbb{Z}}$)

Going back to the Chinese Remainder Theorem, we see that (3.5) is equivalent to the equation

$$\frac{x}{m} \equiv \frac{a_1 b_1}{m_1} + \frac{a_2 b_2}{m_2} + \ldots + \frac{a_k b_k}{m_k} \qquad \text{in } \mathbb{R}/\mathbb{Z}.$$

If the difference between the two sides is k (which must be an integer) then we can replace $a_1 b_1$ by $a_1 b_1 + km$ in the first fraction on the right side so that the two sides are equal numbers in $\mathbb{R}$. This shows us how we can always decompose a fraction with a composite denominator $\prod_p p^{e_p}$ into a sum of fractions whose denominators are the prime powers p^{e_p}, and whose numerators are fixed mod p^{e_p}.

B3. Combinatorics and the multiplicative group mod m.

Card Shuffling. The cards in a 52 card deck can be arranged in $52! \approx 8 \times 10^{67}$ different orders. Between card games we shuffle the cards, in order to make the order of the cards unpredictable. But what if someone can shuffle perfectly? How unpredictable will the order of the cards become? Let's analyze this, by carefully figuring out what happens in a "perfect shuffle": In a riffle shuffle one splits the deck in two, places the two halves in either hand and then drops the cards, using one's thumbs, in order to more-or-less interlace the cards from the two decks.

If that is all done perfectly, one cuts the cards into two 26 card halves, one half with the cards that were in positions 1 through 26, the other half with the cards that were in positions 27 through 52; one then interlaces the two halves so that the new order of the cards becomes (from the top) those that were in positions $1, 27, 2, 28, 3, 29, 4, 30, \ldots$ Viewed the other way around, the cards that were in positions $1, 2, 3, \ldots 26$ go to positions $1, 3, 5, \ldots 51$, that is $k \to 2k-1$ for $1 \le k \le 26$; and the cards that were in positions $27, 28, \ldots 52$ go to positions $2, 4, \ldots 52$, that is $k \to 2k-52$ for $27 \le k \le 52$. Note that the top and bottom cards do not move, that is $1 \to 1$ and $52 \to 52$, so we focus on understanding the permutation of the other fifty cards:

Let us define σ so that $1+m \to \sigma(1+m)$ for $1 \le m \le 50$. Whether m is even or odd, we find that $\sigma(1+m) \equiv 1+2m \pmod{51}$ in either case. We can change this "$\equiv$" to "$=$" if we take $2m \pmod{51}$ to be the least positive residue of $2m$ mod 51. So what happens after two or more shuffles? Card 1 remains at the top of deck, card 52 remains at the bottom. For the others, note that $\sigma^2(1+m) = \sigma(\sigma(1+m)) \equiv \sigma(1+2m) \equiv 1+4m \pmod{51}$; $\sigma^3(1+m) \equiv \sigma(1+4m) \equiv 1+8m \pmod{51}$; and in general $\sigma^r(1+m) \equiv 1+2^r m \pmod{51}$ for all $r \ge 1$. Now $2^8 \equiv 1 \pmod{51}$, and so $\sigma^8(1+m) \equiv 1+m \pmod{51}$. In other words eight perfect riffle shuffles returns the deck to its original state – so much for the 52! possible orderings!

One should note that 8 more perfect riffle shuffles will also return the deck to its original state, a total of 16 perfect riffle shuffles, and also 24, or 32, or 40, etc. Indeed any multiple of 8. So we see that the order of 2 (mod 51) is 8, and that $2^r \equiv 1 \pmod{51}$ if and only if r is divisible by 8. This shows, we hope, why the notion of order is interesting and exhibits one of the key results about orders.

The "necklace proof" of Fermat's Little Theorem. In a bead shop there are beads of a different colours. You wish to make a necklace with p beads, using as many beads of each colour as you like. How many *different* necklaces can be made? To understand what is meant by "different" we note that the necklace "Red-Blue-Green" is the same as the necklace "Green-Red-Blue", since these are the same three colours in the same order when written around a circle. So if we start with all of the a^p ordered sequences of p beads, we need to determine which sequences yield the same necklace. Omitting the a necklaces where the a beads are all the same colour, we claim that for all of the remaining sequences there are exactly p sequences which yield the same necklace. Let c be such a sequence, with elements $c(1), \ldots, c(p)$, where $c(j)$, represented by a number between 1 and a, corresponds to the colour of the jth bead in the sequence. We define c_i to be the sequence of coloured beads in which the jth bead has the same colour as the bead in the ℓth place of c, where ℓ is the least positive residue of $i+j \pmod{p}$. Hence c_i yields the

same necklace as c with the beads rotated i places; we write $c_i(j) = c(j+i \pmod p)$. We claim the $c_i, 0 \le i \le p-1$ are all distinct for if $c_i = c_k$ then

$$c(j+i \pmod p) = c(j+k \pmod p) \text{ for all } j;$$

taking $d = k - i$ and $j = nd - i$ we have

$$c((n+1)d \pmod p) = c(nd \pmod p) \text{ for all } n,$$

and so $c(nd \pmod p) = c(0)$ by an induction argument. Therefore as $d \not\equiv 0 \pmod p$ we have $c(m) = c(0)$ for all m (by the remarks after Corollary 3.6) and hence the beads on the necklace c all have the same colour, which is false.

We deduce that the total number of different necklaces in which the beads do not all have the same colour, is the total number, $a^p - a$, of such sequences, divided by the number that yield the same necklace, that is

$$\frac{a^p - a}{p},$$

and therefore this must be an integer. That is p divides $a^p - a$ for all a, as desired.

`Exercise B3.1` By noticing that if we reverse the order of the beads we also get the same necklace prove that $2p$ divides $a^p - a$ if $p \ge 3$.

B4. Groups. We discuss the abstract notion of a *group* because it is a structure that occurs often in number theory (and throughout mathematics). We can prove results for groups in general, and then these results apply for all examples of groups that arise (one can waste a lot of energy giving the same proof, with minor variations, in each case that a group arises). Many of the main theorems about groups were first proved in a number theory context and then found to apply elsewhere. The main examples of groups are additive groups such as the integers, the rationals, the complex numbers, the integers mod p, the polynomials of given degree, matrices of given size, etc, and multiplicative groups such as the rationals, the complex numbers, the integers mod p, invertible matrices of given dimensions, but not the integers or polynomials.

A *group* is defined to be a set of objects G, and an operation, call it $*$, such that:

(i) If $a, b \in G$, then $a * b \in G$. We say that G is *closed* under $*$.

(ii) If $a, b, c \in G$, then $(a * b) * c = a * (b * c)$; that is, when we multiply three elements of G together it does not matter which pair we multiply first. We say that G is *associative*.

(iii) There exists an element $0 \in G$ such that for every $a \in G$ we have $a * 0 = a$. We call 0 the *identity element* of G for $*$.

(iv) For every $a \in G$ there exists $b \in G$ such that $a * b = b * a = 0$. We say that b is the *inverse* of a. We sometimes write $-a$ or a^{-1}.

One can check that the examples of groups given above satisfy these criteria. We see that there are both finite and infinite groups. However, there is one property that one is used to with numbers and polynomials that is not used in the definition of the a group, and that is that $a * b = b * a$, that a and b *commute*. Although this often holds, there are some simple counterexamples, for instance 2-by-2 matrices:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 2 \\ -1 & 2 \end{pmatrix} \quad \text{whereas} \quad \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

We develop the full theory for 2-by-2 matrices at the end of this subsection. If all pairs of elements of a group commute then we call the group *commutative* or *abelian*.

A given group G can contain other, usually smaller, groups H, which are called *subgroups*. Every group G contains the subgroup given by the identity element, $\{0\}$, and also the subgroup G. It can also contain others. For example the additive group of integers mod 6 with elements $\{0, 1, 2, 3, 4, 5\}$ contains the four subgroups $\{0\}$, $\{0, 3\}$, $\{0, 2, 4\}$, $\{0, 1, 2, 3, 4, 5\}$. Note that every group, and so subgroup, contains the identity element. Infinite groups can also contain subgroups, indeed

$$\mathbb{C} \supset \mathbb{R} \supset \mathbb{Q} \supset \mathbb{Z}.$$

If H is a subgroup of G then we define a *left coset* to be the set $a * H = \{a * h : h \in H\}$ for any $a \in G$. (Right cosets are analogously defined, and the two types are indistinguishable if G is a commutative group). In Theorem 7.3 we saw the prototype of the following result:

Proposition B4.1. *Let H be a subgroup of G. The left cosets of H in G are disjoint. Moreover if G is finite then they partition G, and hence the size of H, $|H|$, divides $|G|$.*

Proof. Suppose that $a*H$ and $b*H$ have a common element c. Then there exists $h_1, h_2 \in H$ such that $a*h_1 = c = b*h_2$. Therefore $b = a*h_1*(h_2)^{-1}$ so that $b \in a*H$ as $h_1*(h_2)^{-1} \in H$

since H is closed. Writing $b = a * k,\ k \in H$, suppose that $g \in b * H$ so that $g = b * h$ for some $h \in H$. Then $g = (a * k) * h = a * (k * h) \in a * H$ by associativity and closure of H. Hence $b * H \subset a * H$. By an analogous proof we have $a * H \subset b * H$, and hence $a * H = b * H$. Therefore any two left cosets of H in G are either disjoint or identical.

Suppose that G is finite, and let $a_1 * H, a_2 * H, \dots, a_k * H$ be a maximal set of disjoint cosets of H inside G. If their union does not equal G then there exists $a \in G$ which is in none of these cosets. But then the coset $a * H$ is disjoint from these cosets (by the first part), and this contradicts maximality.

We have encountered the cosets of the subgroup $\mathbb{Z}$ of the additive group $\mathbb{R}$. Since the cosets look like $a + \mathbb{Z}$, they are all represented by a number in $[0, 1)$, that is by $\{a\}$, the fractional part of a. We write $\mathbb{R}/\mathbb{Z}$ which is also an additive group. This can be represented by wrapping the real numbers around the unit circle; the line segment from 0 to 1 representing one complete revolution. Hence to find the coset representation of a given real number t we simply go round the circle this many times. We are familiar with this when working with the exponential function, since $e^{2i\pi t} = e^{2i\pi\{t\}}$ as $e^{2i\pi} = 1$. (For convenience we will often write $e(t)$ in place of $e^{2i\pi t}$.)

In exercise B2.1 we saw that $a \equiv b \pmod m$ if and only if $\frac{a}{m} = \frac{b}{m}$ in $\mathbb{R}/\mathbb{Z}$; that is a/m and b/m belong to the same coset of $\mathbb{R}/\mathbb{Z}$.

`Exercise B4.1.` Prove that if H is a subgroup of a finite abelian group G then the cosets $a * H$ themselves form a group. We call this the *quotient group* G/H. (We just encountered the example $\mathbb{R}/\mathbb{Z}$.) Show that every element G can be written in a unique way as $a * h$ where $h \in H$ and $a \in G/H$, which we write as $G \cong H \oplus G/H$ (we say that G is the *direct sum* of H and G/H). If G is finite show that $|G/H| = |G|/|H|$.

The most common type of group encountered in number theory is the additive group of integers mod m. One way to view this is as a map from the integers onto the residue classes mod m, from integer a to its congruence class mod m. We write this $\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$. Here $m\mathbb{Z}$ denotes "m times the integers", that is the integers divisible by m, each of which map to 0 (mod m). The Chinese Remainder Theorem states that there is a 1-to-1 correspondence between the residue classes $a \pmod m$, and the "vector" of residue classes $(a_1 \pmod{m_1},\ a_2 \pmod{m_2}, \dots, a_k \pmod{m_k})$, when the m_is are pairwise coprime and their product equals m. This is usually written

$$\begin{aligned} \mathbb{Z}/m\mathbb{Z} \ &\cong \quad \mathbb{Z}/m_1\mathbb{Z} \quad \oplus \quad \mathbb{Z}/m_2\mathbb{Z} \quad \oplus \dots \oplus \quad \mathbb{Z}/m_k\mathbb{Z} \\ a \pmod m &\leftrightarrow (a_1 \pmod{m_1},\ a_2 \pmod{m_2}, \dots, a_k \pmod{m_k}). \end{aligned}$$

The beauty of this is that most arithmetic operations mod m can be "broken down" into the same arithmetic operations modulo each m_i performed componentwise. This is particularly useful when $m = \prod_p p^{e_p}$ and then the m_i are the individual p^{e_p}, since some arithmetic operations are much easier to do modulo prime powers than modulo composites. Besides addition the most important of these operations is multiplication. Thus the above correspondence gives a 1-to-1 correspondence between the reduced residue classes mod m, and the reduced residue classes mod the m_i; we write this as

$$(\mathbb{Z}/m\mathbb{Z})^* \ \cong \ (\mathbb{Z}/m_1\mathbb{Z})^* \ \oplus \ (\mathbb{Z}/m_2\mathbb{Z})^* \ \oplus \ \dots \ \oplus \ (\mathbb{Z}/m_k\mathbb{Z})^*,$$

considering these now as groups under multiplication; and again the operation (of multiplication) can be understood componentwise. Typically we write 0 for the identity of an additive group, and 1 for the identity of a multiplicative group.

We say that two groups G and H are *isomorphic*, and write $G \cong H$ if there is a 1-to-1 correspondence $\phi : G \to H$ such that $\phi(a *_G b) = \phi(a) *_H \phi(b)$ for every $a, b \in G$, where $*_G$ is the group operation in G, and $*_H$ is the group operation in H.

`Exercise B4.2.` Let H be a subgroup of $(\mathbb{Z}/m\mathbb{Z})^*$.

(1) Prove that if n is an integer coprime to m but which is not in a residue class of H, then n has a prime factor which is not in a residue class of H.
(2) Show that if integers $q = p_1 \cdots p_k$ and a are coprime to m then there are infinitely many integers $n \equiv a \pmod m$ such that $(n, q) = 1$.
(3) Prove that if H is not all of $(\mathbb{Z}/m\mathbb{Z})^*$ then there are infinitely many primes which do not belong to any of the residue classes of H. (Hint: Modify the proof(s) of exercises 5.3.3,4,5.)

Proposition B4.1 implies the following generalization of Fermat's Little Theorem:

Corollary B4.2. (Lagrange's Theorem) *For any element a of any finite multiplicative group G we have $a^{|G|} = 1$.*

Proof. Let m be the order of a in G; that is, the least positive integer for which $a^m = 1$.

`Exercise B4.3.` Prove that $H := \{1, a, a^2, \dots, a^{m-1}\}$ is a subgroup of G.

By Proposition B4.1 we know that $m = |H|$ divides $|G|$, and so

$$a^{|G|} = (a^m)^{|G|/m} = 1^{|G|/m} = 1.$$

To deduce Euler's Theorem let $G = (\mathbb{Z}/m\mathbb{Z})^*$ so that $|G| = \phi(m)$.

`Exercise B4.4.` Deduce that if $|G|$ is a prime then G is cyclic.

`Exercise B4.5.` Show that the product of the elements in the cyclic group H in exercise B4.3 is a if m is even, and 1 if a is odd.

Wilson's Theorem for finite abelian groups. *The product of the elements of any given finite abelian group equals 1 unless the group contains exactly one element, ℓ, of order two, in which case the product equals ℓ.*

Proof. As in the proof of Wilson's Theorem we partition the elements, each element with its inverse, providing that they are distinct, since these multiply together to give 1, and hence the product of all of them gives 1. This leaves the product of the elements which are their own inverses; that is the roots of $x^2 = 1$ in the group. Now if $\ell \neq 1$ and $\ell^2 = 1$ then we partition these elements into pairs, $x, \ell x$. The product of each such pair equals ℓ, and therefore the product of all the $2N$ roots of $x^2 = 1$ equals ℓ^N. Now if N is even this equals 1, as $\ell^2 = 1$, and if N is odd then this equals ℓ. In this case the only roots of $x^2 = 1$ are 1 and ℓ, for if $m^2 = 1,\ m \neq 1, \ell$, then the product would also equal m and hence $m = \ell$, a contradiction.

The group $\mathbb{Z}/m\mathbb{Z}$, sometimes written C_m, is called the *cyclic* group of order m, which means that the elements of the group are precisely $\{0 \cdot a, 1 \cdot a, 2 \cdot a, \dots, (m-1) \cdot a\}$, the multiples of the *generator* a (in this case we can take $a = 1$). We now find the structure of all finite abelian groups:

Fundamental Theorem of Abelian Groups. *Any finite abelian group G may be written as*

$$\mathbb{Z}/m_1\mathbb{Z} \oplus \mathbb{Z}/m_2\mathbb{Z} \oplus \dots \mathbb{Z}/m_k\mathbb{Z}.$$

In other words every element of G may be written in the form $g_1^{e_1} g_2^{e_2} \dots g_k^{e_k}$ where g_j has order m_j. We write $G = \langle g_1, g_2, \dots, g_k \rangle$.

Proof. By induction on the size of G. Let a be the element of highest order in G, say of order m. By exercise B4.3 we know that $H := \{1, a, a^2, \dots, a^{m-1}\}$ is a subgroup of G. If $1 < m < |G|$ then, by induction, both H and G/H can be written as the direct sum of cyclic groups, and therefore G since $G \cong H \oplus G/H$ by exercise B4.1.

We saw above that we can write each of the $\mathbb{Z}/m_1\mathbb{Z}$ as a direct sum of cyclic groups of prime power order. Hence, by the Fundamental Theorem of Abelian Groups, we can write any finite abelian group as a direct sum of cyclic groups of prime power order. For each given prime we can put the powers of that prime in descending order, that is the p-part of the group is $\mathbb{Z}/p^{e_1}\mathbb{Z} \oplus \mathbb{Z}/p^{e_2}\mathbb{Z} \oplus \dots \mathbb{Z}/p^{e_\ell}\mathbb{Z}$ where $e_1 \ge e_2 \ge \dots$ Now if we take the components of largest prime power orders we can recombine these, and then those of second highest order, etc, that is we can write

$$\oplus_p \ \mathbb{Z}/p^{e_r}\mathbb{Z} \cong \mathbb{Z}/n_r\mathbb{Z}$$

for $r = 1, 2, 3, \dots$ so that

$$G \cong \mathbb{Z}/n_1\mathbb{Z} \oplus \mathbb{Z}/n_2\mathbb{Z} \oplus \dots \mathbb{Z}/n_\ell\mathbb{Z} \ \text{ where } \quad n_\ell | n_{\ell-1} | \dots | n_2 | n_1.$$

Explicit decomposition of $(\mathbb{Z}/m\mathbb{Z})^*$ as a direct sum of cyclic groups We saw above that, via the Chinese Remainder Theorem, this is the direct sum of groups of the form $(\mathbb{Z}/p^r\mathbb{Z})^*$ for each prime power $p^r \| m$.

In Theorem 7.13 we saw that if p is an odd prime then there exists a primitive root g mod p^r. That is $(\mathbb{Z}/p^r\mathbb{Z})^* = \{g^k : 1 \le k \le \phi(p^r)\}$, and when we multiply two reduced residues together we have $g^a \cdot g^b \equiv g^k \pmod{p^r}$ where $k \equiv a + b \pmod{\phi(p^r)}$. Hence to understand the group action in $(\mathbb{Z}/p^r\mathbb{Z})^*$ we can simply focus on the indices, and then we can work entirely in the additive group mod $\phi(p^r)$. This proves that

$$(\mathbb{Z}/p^r\mathbb{Z})^* \text{ as a multiplicative group } \ \cong \ \mathbb{Z}/\phi(p^r)\mathbb{Z} \text{ as an additive group.}$$

Hence any $(\mathbb{Z}/p^r\mathbb{Z})^*$, where p is an odd prime, is a cyclic group, and its generators are the primitive roots.

In section D2 we will see that if $r \ge 3$ then the elements of $(\mathbb{Z}/2^r\mathbb{Z})^*$ can all be written in the form $\pm g^k \pmod{2^r}$ for some integer k, $0 \le k \le 2^{r-2} - 1$, for some integer $g \equiv \pm 3 \pmod 8$ which has order 2^{r-2} mod 2^r. This implies that

$$(\mathbb{Z}/2^r\mathbb{Z})^* \text{ as a multiplicative group } \ \cong \ \mathbb{Z}/2^{r-2}\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \ \text{ as an additive group.}$$

When we study $(\mathbb{Z}/m\mathbb{Z})^*$ through the Fundamental Theorem of abelian groups, we see that $\lambda(m)$, the largest order of an element mod m, equals m_1 (where $m_k|\dots|m_2|m_1$). Hence $(\mathbb{Z}/m\mathbb{Z})^*$ has a primitive root, that is it is cyclic, if and only if $k = 1$. From our construction above this can happen only if no prime appears twice when we decompose $(\mathbb{Z}/m\mathbb{Z})^*$ into prime power cyclic groups. Just considering the 2-power cyclic subgroups, we easily prove, as in section 7.5, that m must equal $2, 4$ or p^r or $2p^r$. In those cases we see above that m is indeed a cyclic group.

`Exercise B4.6.` Show that the product of the reduced residues mod m is $-1 \pmod m$ when there is a primitive root mod m, and $1 \pmod m$ otherwise.

$$H := G^2 = \{g^2 : g \in G\} \ = \ \left\{a \pmod p : \left(\frac{a}{p}\right) = 1\right\},$$

is a subgroup of $G = (\mathbb{Z}/p\mathbb{Z})^*$ of size $(p-1)/2$, partitioning G into two cosets H and nH, where $(n/p) = -1$. If $a \in H$ then $(a/p) = 1$, and if $a \in nH$ then $(a/p) = -1$. Hence the Legendre symbol distinguishes between the two equivalence classes in G/H which is isomorphic to $\mathbb{Z}/2\mathbb{Z}$ written in the multiplicative form with representatives -1 and 1. We will develop these ideas in the next subsection, which is on Dirichlet characters.

As promised we finish this section by determining *What commutes with a given 2-by-2 matrix?* We will now explore which 2-by-2 matrices commute with a given 2-by-2 matrix, M.

`Exercise B4.7.` Prove that if A and B commute with M then so does $rA + sB$ for any real numbers r and s.

It is evident that I and M commute with M, and hence any linear combinations of I and M. We will show that this is all, unless M is a multiple of the identity. Let $\mathcal{M}_2$ be the set of 2-by-2 matrices with entries in $\mathbb{C}$.

Proposition B4.3. *Given $M \in \mathcal{M}_2$, let $C(M) := \{A \in \mathcal{M}_2 : \ AM = MA\}$. If $M = aI$ for some constant a then $C(M) = \mathcal{M}_2$. Otherwise $C(M) = \{rI + sM : \ r, s \in \mathbb{C}\}$.*

Proof. Let $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. If $A \in C(M)$ then so is $B = A - rI - sM$ for any r and $s \in \mathbb{C}$.

`Exercise B4.8.` Prove that if $a \neq d$ then we can select r and s so that the diagonal of B is all 0s.

If $a \neq d$ then write $B = \begin{pmatrix} 0 & x \\ y & 0 \end{pmatrix}$, so that

$$\begin{pmatrix} cx & dx \\ ay & by \end{pmatrix} = \begin{pmatrix} 0 & x \\ y & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = BM = MB = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 0 & x \\ y & 0 \end{pmatrix} = \begin{pmatrix} by & ax \\ dy & cx \end{pmatrix}.$$

The off diagonal terms yield that $x = y = 0$ and so $A = rI + sM$.

Now suppose that $a = d$ and $M \neq aI$. If $b \neq 0$ then we can select r and s so that the top row of B is all 0s; that is $B = \begin{pmatrix} 0 & 0 \\ x & y \end{pmatrix}$; then the top row of $MB = BM$ yields that $x = y = 0$ and so $A = rI + sM$. Otherwise $c \neq 0$ and an analogous argument works.

`Exercise B4.9.` Prove that the units form a multiplicative group.

B5. Dirichlet characters. We saw in chapter 8 how useful the Legendre and Jacobi symbols are. The key is that they are multiplicative functions defined on the integers modulo some integer m. By this we mean that $\chi(rs) = \chi(r)\chi(s)$ for all integers r and s, and $\chi(a+m) = \chi(a)$ for all integers a. We now wish to find all such non-zero functions, which we call *Dirichlet characters*, and we begin by reducing to the case of prime powers using the Chinese Remainder Theorem:

Proposition B5.1. *If $(r,s) = 1$ then the characters mod rs are in 1-to-1 correspondence with the pairs of character mod r and mod s.*

Proof. Given $\chi \pmod m$, let ρ be the function mod r where we define $\rho(a) := \chi(A)$ where we choose $A \equiv a \pmod r$ and $\equiv 1 \pmod s$; and similarly let σ be the function mod s where we define $\sigma(b) := \chi(B)$ where we choose $B \equiv 1 \pmod r$ and $\equiv b \pmod s$.

`Exercise B5.1.` Verify that ρ and σ are characters.

Given $n \pmod m$, if $n \equiv A \pmod r$ and $\equiv B \pmod s$ then $n \equiv AB \pmod m$ and so $\chi(n) = \chi(A)\chi(B) = \rho(n)\sigma(n)$. Hence every character mod m can be written as the product of a character mod r and a character mod s.

`Exercise B5.2.` Verify that ρ and σ are the unique such characters.

On the other hand given characters $\rho \pmod r$ and $\sigma \pmod s$, we can always construct the character $\chi \pmod m$, given by $\chi = \rho\sigma$.

We already saw this idea that one can multiply together characters of different moduli when we created the Jacobi symbol.

Following this Proposition we can focus on characters mod p^r for some prime p.

`Exercise B5.3.` Prove that $\chi(0) = 0$, $\chi(p) = 0$ and $\chi(1) = 1$. (Hint: Use that $a \cdot 1 = a$ for some a with $\chi(a) \neq 0$.) Prove that $\chi(-1) = \pm 1$.

`Exercise B5.4.` Deduce that for any Dirichlet character mod m we have $\chi(b) = 0$ if $(b,m) > 1$.

If p is an odd prime then we know that there exists a primitive root g mod p^r by Theorem 7.13. Now every reduced residue $a \pmod{p^r}$ can be written as $a \equiv g^k \pmod{p^r}$ for some integer k, and therefore $\chi(a) = \chi(g^k) = \chi(g)^k$, that is χ is completely determined by the value of $\chi(g)$. What restrictions are there on the value of $\chi(g)$? The key one is that $\chi(g)^{\phi(p^r)} = \chi(g^{\phi(p^r)}) = \chi(1) = 1$, so that $\chi(g)$ is a $\phi(p^r)$th root of unity. Let $\gamma = e\left(\frac{1}{\phi(p^r)}\right)$ be a primitive $\phi(p^r)$th root of unity.

Lemma B5.2. *If p is an odd prime and g is a primitive root mod p^r then the set of Dirichlet characters $\chi_j \pmod{p^r}$ are given by $\chi_j(g) = \gamma^j$ for $j = 0, 1, 2, \dots, \phi(p^r) - 1$.*

Note that if $a \equiv g^k \pmod{p^r}$ then $\chi_j(a) = \chi_j(g^k) = \chi_1(g)^{jk} = \chi_1(g^k)^j = \chi_1(a)^j$; that is one can write $\chi_j = \chi_1^j$. But then the characters $\pmod{p^r}$ form a cyclic group of order $\phi(p^r)$ with generator χ_1; that is the *character group* is isomorphic to the group of reduced residues mod p^r. By the construction in the Proposition B5.1, this is also true for the characters mod m, whenever m is odd. In the character group, the element of order 1

is the *principal character* denoted χ_0; in fact

$$\chi_0(a) = \begin{cases} 1 & \text{if } (a,m)=1, \\ 0 & \text{if } (a,m)>1. \end{cases}$$

These cyclic groups of characters $\pmod{p^r}$ are isomorphic to the group of reduced residues, and so have a unique element of order 2 by Theorem 7.7. We have already encountered the character $\pmod p$ of order two; the Legendre symbol. In fact Proposition 8.4 tells us that this is also how to recognize squares mod p^r; in other words, the Legendre symbol also is a character mod p^r.

The Jacobi symbol arises as a character in two ways. Firstly $\left(\frac{\cdot}{m}\right)$ is a character mod m. Secondly if m is odd then $\left(\frac{m}{\cdot}\right)$ is a character mod m or mod $4m$, as $m \equiv 1$ or $3 \pmod 4$, by the law of quadratic reciprocity.

The above arguments work mod 2 and mod 4, but there is no primitive root mod 2^r with $r \geq 3$. In that case we saw in the previous subsection that all reduced residues take the form $\pm g^k \pmod{2^r}$ for some residue g of order 2^{r-2}.

`Exercise B5.5.` Prove that the set of characters χ mod 2^r can be given as $\chi_{i,j}$ with $i = 0$ or 1, and $0 \leq j \leq 2^{r-2}-1$, defined by $\chi_{i,j}(-1) = (-1)^i$ and $\chi_{i,j}(g) = \gamma^j$ where $\gamma = e\left(\frac{1}{\phi(2^{r-2})}\right)$. Deduce that the character group mod 2^r is isomorphic to the group of reduced residues mod 2^r. Finally deduce that for every positive integer m, the character group mod m is isomorphic to the group of reduced residues mod m.

`Exercise B5.6.` (i) Prove that if $\chi \neq \chi_0$ then there exists $a \pmod m$ for which $\chi(a) \neq 0$ or 1. (ii) Prove that if $(a,m)=1$ and $a \not\equiv 1 \pmod m$ then there exists a character $\chi \pmod m$ for which $\chi(a) \neq 0$ or 1.

The same ideas work for any finite abelian group G with the same proof that the character group of G is isomorphic to G.

Suppose that q divides m, that ψ a character $\pmod q$ and χ_0 the principal character $\pmod m$. The character $\chi := \psi\chi_0$ is a character $\pmod{\text{lcm}[q,m]}$. In fact $\chi(a) = \psi(a)$ if $(a,m)=1$ and $\chi(a)=0$ if $(a,m)>1$, so we say that χ is *induced* by ψ. Note that there are always $\phi(q)$ characters mod m that are induced by characters mod q. Any character that is not induced by a character with a smaller modulus is called *primitive*. The modulus for a character is sometimes called the *conductor* of the character.

`Exercise B5.7.` A *real* character is a character that only takes on real values. Prove that χ is a real character if and only if it has order one or two. Prove that if the conductor m is odd then the real characters are the principal character, and the characters induced by Jacobi symbols modulo the divisors of m.

A *complex* character is a character that is not a real character. The *conjugate* $\overline{\chi}$ of χ is that character for which $\overline{\chi}(n) = \overline{\chi(n)}$ for all integers n. Notice that $\overline{\chi} = \chi$ if and only if χ has order 1 or 2; i.e. χ is a real character.

Now suppose that $\chi \pmod q$ is a character of order m (which must divide $\lambda(q)$). Then for each reduced residue $a \pmod q$ we have $\chi(a)^m = (\chi^m)(a) = \chi_0(a) = 1$, and so $\chi(a)$ is an mth root of unity. Then

$$H := \{a \pmod q : \ \chi(a) = 1\}$$

is a subgroup of $G := (\mathbb{Z}/q\mathbb{Z})^*$. For $0 \le j \le m-1$ select some $b_j \pmod q$ such that $\chi(b_j) = e(\frac{j}{m})$. Then $b_0H = H, b_1H, b_2H, \dots, b_{m-1}H$ partition the reduced residues mod q, and $b_jH = \{a \pmod q : \ \chi(a) = e(\frac{j}{m})\}$. This all in direct analogy with the comments about the Legendre symbol near the end of the last subsection.

The main reason to develop the theory of Dirichlet characters is to identify the elements of an arithmetic progression $a \pmod q$, when $(a,q) = 1$, using the following identity:

$$\frac{1}{\phi(q)} \sum_{\chi \pmod q} \chi(n) = \begin{cases} 1 & \text{if } n \equiv 1 \pmod q \\ 0 & \text{otherwise.} \end{cases} \tag{B5.1}$$

To prove this note first that if $n \equiv 1 \pmod q$ then $\chi(n) = 1$ and the result follows. Otherwise select a character $\psi \pmod q$ such that $\psi(n) \neq 1$ (as in exercise B5.6). As the characters form a group, we have

$$\{\psi\chi : \ \chi \pmod q\} = \{\chi : \ \chi \pmod q\}.$$

(This is analogous to the 'set of reduced residues" proof of Fermat's Little Theorem in section 7.1.) Hence

$$\psi(n) \sum_{\chi \pmod q} \chi(n) = \sum_{\chi \pmod q} (\psi\chi)(n) = \sum_{\chi \pmod q} \chi(n),$$

and the result follows.

We now use (B5.1) to identify integers n that are $\equiv a \pmod q$ when $(a,q) = 1$. The idea is that $m \equiv n/a \pmod q$ then $m \equiv 1 \pmod q$ if and only if $n \equiv a \pmod q$, and moreover $\chi(m) = \overline{\chi}(a)\chi(n)$. Hence

$$\frac{1}{\phi(q)} \sum_{\chi \pmod q} \overline{\chi}(a)\chi(n) = \begin{cases} 1 & \text{if } n \equiv a \pmod q \\ 0 & \text{otherwise.} \end{cases} \tag{B5.2}$$

Therefore if P is some set of integers n with associated weights $w(n)$ then

$$\begin{aligned} \sum_{\substack{n \in P \\ n \equiv a \pmod q}} w(n) &= \sum_{n \in P} w(n) \frac{1}{\phi(q)} \sum_{\chi \pmod q} \overline{\chi}(a)\chi(n) \\ &= \frac{1}{\phi(q)} \sum_{\chi \pmod q} \overline{\chi}(a) \left(\sum_{n \in P} \chi(n)w(n) \right), \end{aligned} \tag{B5.3}$$

so long as all the sums converge absolutely. We have thus changed our problem to several new weighted sums over elements of P, but now we no longer have to concern ourselves with the relatively difficult restriction to an arithmetic progression.

`Exercise B5.8.` Prove the following "dual" identity to (B5.1):

$$\frac{1}{\phi(q)} \sum_{a \pmod q} \chi(a) = \begin{cases} 1 & \text{if } \chi = \chi_0 \\ 0 & \text{otherwise.} \end{cases} \tag{B5.4}$$

One final observation. Corollary 8.2 generalizes rather beautifully to: If $m|p-1$ then

$$\#\{b \pmod p:\ b^m \equiv a \pmod p\} = 1 + \sum_{\substack{\chi \pmod p \\ \chi^m=\chi_0,\ \chi\neq\chi_0}} \chi(a). \tag{B5.5}$$

`Exercise B5.9.` First establish this for $a = 0$. Now let g be a primitive root $\pmod p$ and $a \equiv g^k \pmod p$, and note that $1 = \chi_0(a)$.

(1) Show that there exists a character ψ of order m for which $\psi(g) = e(\frac{1}{m})$.
(2) Show that $\{\chi:\ \chi^m = \chi_0\} = \{\psi^j:\ 0 \le j \le m-1\}$.
(3) Show that the right side of (B5.5) is m if $m|k$, otherwise it equals 0.
(4) Show that the left side of (B5.5) is m if $m|k$, otherwise it equals 0.

Additive characters. Dirichlet characters mod q respect the multiplicative group mod q (that is, they are an *homomorphism* from $(\mathbb{Z}/q\mathbb{Z})^*$ to $\mathbb{C}$). One might ask whether there are characters that respect the additive group mod q, and whether they end up being as useful. So to formulate our problem we want a function with the property that $f(a) = f(b)$ if $a \equiv b \pmod q$, and such that $f(a+b) = f(a)f(b)$. We immediately deduce that $f(0) = 1$, and since the groups $\mathbb{Z}/q\mathbb{Z}$ are cyclic, generated by 1, so $f(a) = f(1)^a$ for all a, so the function depends entirely on the value of $f(1)$. Now $f(1)^q = f(q) = 1$ and so $f(1)$ is a qth root of unity; it turns out that any qth root of unity will do. Let $\psi(1) = e(1/q)$, so that $\psi(n) = e(n/q)$. The set of possible *additive characters* are

$$\psi_a(n) := e\left(\frac{an}{q}\right) \text{ for } 0 \le n \le q-1,$$

defined for $0 \le n \le q-1$. Note that $\psi_a = \psi^a$

In section B5 we saw how to pick out terms of the arithmetic progression $a \pmod q$, when $(a,q) = 1$ using Dirichlet characters. These characters are a homomorphism on the multiplicative group mod q. There is another way to pick out terms of the arithmetic progression $a \pmod q$, whether or not $(a,q) = 1$, using additive characters, that is homomorphisms on the additive group mod q. These are simply the functions

$$E_a(n) := e\left(\frac{an}{q}\right) \text{ for } 0 \le a \le q-1.$$

If we define $e_p = E_1$ then $E_a = e_p^a$ for each a.

The additive characters can also be used to pick out arithmetic progressions, since the sum of the distinct qth roots of unity equals 1.

`Exercise B5.10.` Prove that for any a we have

$$\frac{1}{q}\sum_{m=0}^{q-1} e\left(\frac{-ma}{q}\right) e\left(\frac{mn}{q}\right) = \begin{cases} 1 & \text{if } n \equiv a \pmod q \\ 0 & \text{otherwise.} \end{cases}$$

Deduce that

$$\sum_{\substack{n\in P \\ n\equiv a \pmod q}} w(n) = \frac{1}{q}\sum_{m=0}^{q-1} e\left(\frac{-ma}{q}\right)\left(\sum_{n\in P} w(n) e\left(\frac{mn}{q}\right)\right).$$

B6. Insolvability of the quintic. Suppose that α is the root of the irreducible polynomial $f(x)$ which has integer coefficients. Most equations involving α can be written in the form $G(\alpha) = 0$, and so $f(x)$ divides $G(x)$ by Proposition A3.1. Now, if β is any other root of f then $G(\beta) = 0$ also, since $f(x)$ divides $G(x)$. We call β a *conjugate* of α. Since the actual root of f that we are using is irrelevant, we might as well be working in $\mathbb{Z}[x]$ mod $f(x)$, which is often written as $\mathbb{Z}[x]/(f(x))$.

One might ask if one can extend this. For example if $H(x, y) \in Z[x, y]$ and $H(\alpha, \beta) = 0$ for two given roots α, β of f, is it true that $H(\alpha', \beta') = 0$ for any other two given roots α', β' of f? The answer in general is no. For example the roots of $x^4 + 1$ can be written as $\alpha, \alpha^3, \alpha^5, \alpha^7$ or $\frac{\pm 1 \pm i}{\sqrt{2}}$, and if $H(x, y) = xy - 1$ then $H(\alpha, \alpha^7) = H(\alpha^3, \alpha^5) = 0$, but $H(\alpha, \alpha^3) = -2$ and $H(\alpha, \alpha^5) \neq 0$. However it is evidently interesting to discover which roots can by replaced by which other roots to keep satisfying an equation. For example given $H(x_1, x_2, x_3, x_4)$ with $H(\alpha, \alpha^3, \alpha^5, \alpha^7) = 0$ we can simply let $g(t) = H(t, t^3, t^5, t^7)$ so that $g(\alpha) = 0$. Therefore $g(\alpha^3) = g(\alpha^5) = g(\alpha^7) = 0$ by the remarks of the previous paragraph, and so $H(\alpha^3, \alpha, \alpha^7, \alpha^5) = g(\alpha^3) = 0$ and similarly $H(\alpha^5, \alpha^7, \alpha, \alpha^3) = H(\alpha^7, \alpha^5, \alpha^3, \alpha) = 0$. The key to better understanding solutions to equations, is to understand how the solutions to such polynomials can be mapped. More precisely, if $f(x)$ is irreducible of degree d, with roots $\alpha_1, \dots, \alpha_d$ then let G be the set of permutations σ of $1, 2, \dots, d$ for which $H(\alpha_{\sigma(1)}, \dots, \alpha_{\sigma(d)}) = 0$ for every $H(x_1, \dots, x_d) \in Z[x_1, \dots, x_d]$ for which $H(\alpha_1, \dots, \alpha_d) = 0$.

`Exercise B6.1.` Prove that G is a group (which is called the *Galois group* associated to f).

For "most" polynomials f, the associated Galois group is the set of all permutations of the roots.

Solvability in terms of surds. The ideas used to determine constructibility can also be developed to try to understand when the root of a polynomial can be determined in terms of surds. The set of mth roots of n are given by $n^{1/m}$ times each of the mth roots of 1. If α is a number that can be expressed in terms of surds then it must belong to some field created out of surds. The details of how one can find α for which this is impossible are discussed in any good book on Galois theory. Here we will just sketch the main ideas.

The key trick that Galois came up with was to study the Galois group as above. Let α be a primitive mth root of unity. The roots of $\phi_m(x)$ are α^k, $1 \leq k \leq m$, $(k, m) = 1$ (as we saw at the end of section A3). Now if $G(x, y) = xy - 1$ then $G(\alpha, \alpha^{m-1}) = 0$ but $G(\alpha, \alpha^k) \neq 0$ for all other such k. Hence we see that our group, for surds, is very limited. Without getting into the (complicated) details of the definition, the group associated to extensions created by surds is always *solvable*. Moreover all subfields of extensions created by surds also have groups that are solvable. The easiest group that is not solvable is the set of all permutations on five elements, and then one shows that there are irreducible polynomials of degree five such that the field created by adjoining the roots of this polynomial to $\mathbb{Q}$, has this group – for example $x^5 - 6x + 3$.

C. Algebra

C1. Ideals. Let R be a set of numbers that is closed under addition and subtraction; for example $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ or $\mathbb{C}$, but not $\mathbb{N}$. We define the *ideal* generated by $a_1, \dots, a_k$ over R, to be the set of linear combinations of $a_1, \dots, a_k$ with coefficients in R; that is

$$I_R(a_1, \dots, a_k) = \{r_1a_1 + r_2a_2 + \dots + r_ka_k : \ r_1, \dots, r_k \in R\}.$$

(Note $a_1, \dots, a_k$ are not necessarily in R.) In Corollary 1.6 and the exercise 1.2.6 we saw that any ideal over $\mathbb{Z}$ can be generated by just one element. The reason we take such interest in this definition is that this is not necessarily true when the a_i are taken from other domains. For example if $R = \mathbb{Z}[\sqrt{-5}]$, that is the numbers of the form $u + v\sqrt{-5}$ where u and v are integers, then the ideal $I_R(2, 1 + \sqrt{-5})$ cannot be generated by just one element, as we will see below. A *principal ideal* is an ideal that can be generated by just one element.

Ideals in quadratic fields. We saw that any ideal in $\mathbb{Z}$ may be generated by just one element. We will now prove that any ideal in a quadratic ring of integers:

$$R := \{a + b\sqrt{d} : \ a, b \in \mathbb{Z}\}$$

can be generated by at most two integers. Suppose an ideal $I \subset R$ is given. Either $I \subset \mathbb{Z}$ in which case it is a principal ideal, or there exists some element $u + v\sqrt{d} \in I$ with $v \neq 0$. We may assume that $v > 0$ by replacing $u + v\sqrt{d}$ with $-(u + v\sqrt{d})$ if v was negative.

Now select that $r + s\sqrt{d} \in I$ with $s > 0$ minimal. Such an s exists since it must be a positive integer in $\{1, 2, \dots |v|\}$. Note that if $u + v\sqrt{d} \in I$ then s divides v, for if not select $k, \ell \in \mathbb{Z}$ for which $ks + \ell v = g := \gcd(s, v)$ and then

$$(kr + \ell u) + g\sqrt{d} = k(r + s\sqrt{d}) + \ell(u + v\sqrt{d}) \in I$$

contradicting the minimality of s.

If $u + v\sqrt{d} \in I$ then let $m = v/s$, so that $(u + v\sqrt{d}) - m(r + s\sqrt{d}) = u - mr$. Therefore every element of the ideal I may be written as $m(r + s\sqrt{d}) + n$ where $n \in I \cap \mathbb{Z}$, and m is an arbitrary integer. Now $I \cap \mathbb{Z}$ is an ideal in $\mathbb{Z}$ so must be principal, generated by some integers $g \geq 1$. Therefore

$$I = \{m(r + s\sqrt{d}) + ng : \ m, n \in \mathbb{Z}\} = I_{\mathbb{Z}}(r + s\sqrt{d}, g).$$

So we have achieved our goal, I has been shown to be generated by just two elements; and better yet we have proved that we only need to take linear combinations of those two elements with coefficients in $\mathbb{Z}$ to obtain the whole of I. However, we can simplify even more:

Since $\sqrt{d} \in R$, hence $g\sqrt{d} \in I$ and $sd + r\sqrt{d} \in I$, and so s divides both g and r. Therefore $r = sb$ and $g = sa$ for integers a and b. Finally $s(b^2 - d) = (r + s\sqrt{d})(b - \sqrt{d}) \in I \cap \mathbb{Z}$ and so $s(b^2 - d)$ is a multiple of $g = sa$; hence a divides $b^2 - d$. Therefore

$$I = I_{\mathbb{Z}}(s(b + \sqrt{d}), sa) \ \text{ which we write as } \ s \cdot I_{\mathbb{Z}}(b + \sqrt{d}, a),$$

for some integers s, a, b where a divides $b^2 - d$.

Non-principal ideals. Let $R = \mathbb{Z}[\sqrt{-d}]$ with $d \geq 2$. Which ideals $I := I_R(p, r + s\sqrt{-d})$ are principal, where p is a prime in $\mathbb{Z}$ that divides $r^2 + ds^2$, but does not divide s? (This includes the example $I_R(2, 1 + \sqrt{-5})$).

Theorem C1.1. *Let $R = \mathbb{Z}[\sqrt{-d}]$ with $d \geq 2$. Suppose that p is a prime in $\mathbb{Z}$ which divides $r^2 + ds^2$ but not s. Then the ideal $I := I_R(p, r + s\sqrt{-d})$ is principal if and only if $p = I_R(a + b\sqrt{-d})$ where $p = a^2 + db^2$ with $a, b \in \mathbb{Z}$ (and $a/b \equiv r/s \pmod p$).*

We will use the following result:

Lemma C1.2. *If integer prime p equals the product of two elements of $\mathbb{Z}[\sqrt{-d}]$ then it is either as $(\pm 1) \cdot (\pm p)$, or as $p = (a + b\sqrt{-d})(a - b\sqrt{-d})$ where $p = a^2 + db^2$.*

Proof of Lemma C1.2. Suppose that $p = (a + b\sqrt{-d})(u + v\sqrt{-d})$ where $a, b, u, v \in \mathbb{Z}$. Now $\gcd(a, b) \cdot \gcd(u, v)$ divides p, then at least one of these gcds equals 1, say $(a, b) = 1$. Now $p = (au - dbv) + (av + bu)\sqrt{-d}$, so the coefficient of the imaginary part is $av + bu = 0$. Therefore $a|bu$ and therefore $a|u$ as $(a, b) = 1$. Writing $u = ak$ we have $v = -bk$ and therefore $p = k(a + b\sqrt{-d})(a - b\sqrt{-d}) = k(a^2 + db^2)$.

Now $a^2 + db^2$ is a positive divisor of p so must equal either 1 or p. If $b \neq 0$ then $a^2 + db^2 \geq d > 1$, and so if $a^2 + db^2 = 1$ then $a = \pm 1,\ b = 0$. Otherwise $a^2 + db^2 = p$.

Proof of Theorem C1.1. Suppose that $I := I_R(p, r + s\sqrt{-d})$ is principal, say, $I = I_R(g)$ where $g \in R$. Then we can write p as the product of two elements of $\mathbb{Z}[\sqrt{-d}]$, including g, and so by the lemma, $g = \pm 1$ or $\pm p$ or $a \pm b\sqrt{-d}$ where $p = a^2 + db^2$. We cannot have $g = \pm p$ for if $r + s\sqrt{-d} = \pm p(u + v\sqrt{-d})$ then we would have that p divides r and s contrary to the hypothesis.

Now let $t \equiv 1/s \pmod p$ and $m \equiv rt \pmod p$ so that $m + \sqrt{-d} \equiv t(r + s\sqrt{-d}) \pmod p$, and $r + s\sqrt{-d} \equiv s(m + \sqrt{-d}) \pmod p$ as $sm \equiv rst \equiv r \pmod p$. Moreover $m^2 + d \equiv t^2(r^2 + ds^2) \equiv 0 \pmod p$, so that $I = I_R(p, m + \sqrt{-d})$. This is presented in the form at the end of the last subsection and so $I = I_{\mathbb{Z}}(p, m + \sqrt{-d})$. Notice that ± 1 is not in this ideal (since it is not divisible by p).

We are left with the only possibility that $g = a \pm b\sqrt{-d}$. In this case $(mb)^2 - a^2 = b^2(m^2 + d) - (a^2 + db^2) \equiv 0 \pmod p$ so that p divides $(mb - a)(mb + a)$. Hence either $m \equiv a/b$ or $-a/b \pmod p$, and we choose the sign of b so that $a \equiv bm \pmod p$. Therefore $a + b\sqrt{-d} \equiv b(m + \sqrt{-d}) \pmod p$ and so $a + b\sqrt{-d} \in I$.

Example: Both $I_R(2, 1 + \sqrt{-5})$ and $I_R(3, 1 + \sqrt{-5})$ are non-principal since there do not exist integers a, b for which $a^2 + 5b^2 = 2$ or 3.

C2. Continued Fractions.

C2.1. The Euclidean Algorithm: The Matrix version. We return to the continued fraction description given in section 1.3. The equation

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

can be re-written as

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} q & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} b \\ r \end{pmatrix},$$

especially if we adopt the convention that the matrix $\begin{pmatrix} a \\ b \end{pmatrix}$ represented $\frac{a}{b}$. But with this representation, it is easy to see what happens when we iterate the algorithm: If we write $a_0 = q = [a/b]$, and then $a_1 = [b/r]$ so that $b = a_1 r + s$, and define $a_2, a_3, \dots$ iteratively (as in section 1.3), then

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} b \\ r \end{pmatrix} = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} r \\ s \end{pmatrix} = \dots =$$
$$= \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_k & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

since the Euclidean algorithm ends with the pair $(\gcd(a, b), 0) = (1, 0)$. Noting that $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is the first column of I, we can define

$$\begin{pmatrix} p_j & p_{j-1} \\ q_j & q_{j-1} \end{pmatrix} = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_j & 1 \\ 1 & 0 \end{pmatrix}$$

so that $a = p_k$ and $b = q_k$. Taking determinants we have

$$p_j q_{j-1} - p_{j-1} q_j = (-1)^{j+1};$$

in particular $au + bv = 1$ where $u = (-1)^{k-1} q_{k-1}$ and $v = (-1)^k p_{k-1}$. This gives a compact, straightforward way to formulate the Euclidean algorithm.

C2.2. Tiling a rectangle with squares. Given, say, a 9-by-7 rectangle we will tile it, greedily, with squares. The largest square that we can place inside a 9-by-7 rectangle is the 7-by-7 square. If we place a 7-by-7 square at one end, then it goes across the breadth of the rectangle and most of the length, leaving 2 units at the end. That is, we have yet to tile the remaining 7-by-2 rectangle. The largest square that can be placed inside this rectangle is a 2-by-2 square, in fact we have room for three of them, leaving us with a 1-by-2 rectangle which we can cover with two 1-by-1 squares. Hence we have tiled the 9-by-7 rectangle by one 7-by-7, three 2-by-2, and two 1-by-1 squares. The area of the rectangle can be computed in term of the areas of each of the squares, that is

$$9 \cdot 7 = 1 \cdot 7^2 + 3 \cdot 2^2 + 2 \cdot 1^2.$$

What has this to do with the Euclidean algorithm? At a given step we have an a-by-b rectangle, with $a > b \geq 1$, and then we can remove q b-by-b squares, where $a = qb + r$ with $0 \leq r \leq a - 1$ leaving an r-by-a rectangle, and so proceed by induction. Writing $9 = 1 \cdot 7 + 2,\ 7 = 3 \cdot 2 + 1$ and $2 = 2 \cdot 1 + 0$ yields the example above.

`Exercise C2.2.1.` Given an a-by-b rectangle show how to write $a \cdot b$ as a sum of squares, as above, in terms of the quotients and partial convergents of the continued fraction for a/b.[7]

C2.3. Continued fractions for real numbers. One can define the continued fraction for any real number $\alpha = \alpha_0$: Let $a_0 := [\alpha_0]$. If $\alpha_0 - a_0 = 0$, that is α_0 is an integer we stop; otherwise we repeat the process with $\alpha_1 = 1/(\alpha_0 - a_0)$. This yields a unique continued fraction for each real number α. In fact $\alpha_j - a_j = \alpha_j - [\alpha_j] = \{\alpha_j\} \in [0, 1)$, so that each $\alpha_{j+1} \geq 1$ for all $j \geq 0$. Hence a_j is a positive integer for each $j \geq 1$.

`Exercise C2.3.1.` Prove that if α has a finite length continued fraction then the last term is an integer ≥ 2.

To determine the value of $[a_0,\ a_1,\ a_2,\ \ldots,]$, where the integer $a_0 \geq 0$ and each other integer $a_i \geq 1$, we define the *convergents* $p_n/q_n := [a_0,\ a_1,\ \ldots,\ a_n]$ for each $n \geq 0$ as above by

$$\begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix}.$$

Note that $p_n = a_n p_{n-1} + p_{n-2}$ and $q_n = a_n q_{n-1} + q_{n-2}$ for all $n \geq 2$, so that the sequences $p_1, p_2, \ldots$ and $q_1, q_2, \ldots$ are increasing. Taking determinants yields that

$$\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} = \frac{(-1)^{n+1}}{q_{n-1} q_n} \tag{C2.1}$$

for each $n \geq 1$.

`Exercise C2.3.2.` Deduce that

$$\frac{p_0}{q_0} < \frac{p_2}{q_2} < \cdots < \frac{p_{2j}}{q_{2j}} < \cdots\cdots < \frac{p_{2j+1}}{q_{2j+1}} < \cdots < \frac{p_3}{q_3} < \frac{p_1}{q_1};$$

and that p_n/q_n tends to a limit as $n \to \infty$.

We have proved that if n is finite then the value given by the continued fraction is indeed α, but this is not so obvious if n is infinite (i.e. α is irrational). We now prove this, and as a consequence one can deduce that the positive real numbers are in 1-to-1 correspondence with the continued fractions.

`Exercise C2.3.3.` Show that if a, b, A, B, u, v are positive reals then $\frac{au+Av}{bu+Bv}$ lies between $\frac{a}{b}$ and $\frac{A}{B}$.

Now $\alpha = [a_0,\ a_1,\ a_2,\ \ldots,\ a_n,\ \alpha_{n+1}]$, so that $\alpha = R/S$ where

$$\begin{pmatrix} R & p_n \\ S & q_n \end{pmatrix} = \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} \begin{pmatrix} \alpha_{n+1} & 1 \\ 1 & 0 \end{pmatrix},$$

[7] I would like to thank Dusa MacDuff and Dylan Thurston for bringing my attention to this beautiful application.

and hence

$$\alpha = \frac{R}{S} = \frac{\alpha_{n+1}p_n + p_{n-1}}{\alpha_{n+1}q_n + q_{n-1}} \tag{C2.2}$$

lies between $\frac{p_{n-1}}{q_{n-1}}$ and $\frac{p_n}{q_n}$ for each $n \geq 1$, by the previous exercise.

`Exercise C2.3.4.` Deduce that $\alpha = \lim_{n\to\infty} p_n/q_n$.

`Exercise C2.3.5.` Also deduce that $\left|\alpha - \frac{p_n}{q_n}\right| \leq \frac{1}{q_n q_{n+1}}$ for all $n \geq 0$.

Now $\pi = [3, 7, 15, 1, 292, 1, \dots]$ which leads to the convergents

$$3 < \frac{333}{106} < \dots < \pi < \dots < \frac{355}{113} < \frac{22}{7} .$$

Archimedes knew that $|\pi - \frac{355}{113}| < 3 \cdot 10^{-7}$.[8] The continued fraction for e displays an interesting pattern: $e = [2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, \dots]$. One can generalize the notion of continued fractions to obtain

$$\frac{\pi}{4} = \cfrac{1}{1+\cfrac{1^2}{2+\cfrac{3^2}{2+\cfrac{5^2}{2+\dots}}}} \qquad \text{or} \qquad \pi = \cfrac{4}{1+\cfrac{1^2}{3+\cfrac{2^2}{5+\cfrac{3^2}{7+\dots}}}} .$$

By (C2.2) and then (C2.1) we have

$$\alpha - \frac{p_n}{q_n} = \frac{\alpha_{n+1}p_n + p_{n-1}}{\alpha_{n+1}q_n + q_{n-1}} - \frac{p_n}{q_n} = \frac{(-1)^n}{q_n(\alpha_{n+1}q_n + q_{n-1})} .$$

Now $a_{n+1} \leq \alpha_{n+1} < a_{n+1} + 1$ and so $q_{n+1} \leq \alpha_{n+1}q_n + q_{n-1} < q_{n+1} + q_n < 2q_{n+1}$, yielding

$$\frac{1}{2q_nq_{n+1}} < \left|\alpha - \frac{p_n}{q_n}\right| \leq \frac{1}{q_nq_{n+1}} . \tag{C2.3}$$

This is a good approximation, but are there better? Lagrange showed that there are not:

Theorem C2.1. *If $1 \leq q < q_{n+1}$ then $|q_n\alpha - p_n| \leq |q\alpha - p|$.*

Proof. Let $x = (-1)^n(pq_{n+1} - qp_{n+1})$ and $y = (-1)^n(pq_n - qp_n)$, so that $p_nx - p_{n+1}y = p$ and $q_nx - q_{n+1}y = q$ as $p_nq_{n+1} - q_np_{n+1} = (-1)^n$. We observe that $x \neq 0$ else q_{n+1} divides $q_{n+1}y = -q$ so that $q_{n+1} \leq q$ contradicting the hypothesis.

[8] Around 1650 B.C., ancient Egyptians approximated π by regular octagons obtaining 256/81, a method developed further by Archimedes in the third century B.C. and Liu Hui in China in the third century A.D. In 1168 B.C. the Talmudic scholar Maimonides asserted that π can only be known approximately, perhaps a claim that it is irrational. In the ninth century B.C. the Indian astronomer Yajnavalkya arguably gave the approximation 333/106 in *Shatapatha Brahmana*; in the 14th century A.D., Madhava of the Kerala school in India indicated how to get arbitrarily good approximations to π.

Now $q_nx = q_{n+1}y + q$ where $q < q_{n+1} \le |q_{n+1}y|$ if $y \neq 0$, and so q_nx and $q_{n+1}y$ have the same sign, and therefore x and y have the same sign. We saw earlier that $q_n\alpha - p_n$ and $q_{n+1}\alpha - p_{n+1}$ have opposite signs, and so $x(q_n\alpha - p_n)$ and $y(q_{n+1}\alpha - p_{n+1})$ have opposite signs. Now $q\alpha - p = x(q_n\alpha - p_n) - y(q_{n+1}\alpha - p_{n+1})$ and so

$$|q\alpha - a| = |x(q_n\alpha - p_n)| + |y(q_{n+1}\alpha - p_{n+1})| \ge |q_n\alpha - p_n|,$$

with equality implying that $|x| = 1$ and $y = 0$, so that $\{p, q\} = \{p_n, q_n\}$.

`Exercise C2.3.6.` Deduce that if $1 \le q < q_n$ then $\left|\alpha - \frac{p_n}{q_n}\right| < \left|\alpha - \frac{p}{q}\right|$.

Corollary C2.2. *If* $\left|\alpha - \frac{p}{q}\right| < \frac{1}{2q^2}$ *then* $\frac{p}{q}$ *is a convergent for* α.

Proof. If $q_n \le q < q_{n+1}$ then $|q_n\alpha - p_n| < 1/2q$ by Theorem C2.1. Hence $p/q = p_n/q_n$ else

$$\frac{1}{qq_n} \le \left|\frac{p}{q} - \frac{p_n}{q_n}\right| \le \left|\alpha - \frac{p}{q}\right| + \left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{2q^2} + \frac{1}{2qq_n},$$

which is impossible.

Suppose that we have a solution to Pell's equation, that is $p^2 - dq^2 = \pm 4$ with $p, q > 0$. Therefore $|\sqrt{d} + p/q| > \sqrt{d}$ so that

$$\left|\sqrt{d} - \frac{p}{q}\right| = \frac{|p^2 - dq^2|}{q^2(\sqrt{d} + p/q)} < \frac{4}{\sqrt{d}q^2}\ .$$

If $d \ge 64$ then this $< 1/2q^2$ and so p/q is a convergent for $\sqrt{d}$.

`Exercise C2.3.7.` Show that if $0 < p^2 - dq^2 \le \sqrt{d}$ with $p, q \ge 1$ then p/q is a convergent for $\sqrt{d}$.

`Exercise C2.3.8.` Suppose that $p, q \ge 1$ and $-\sqrt{d} \le p^2 - dq^2 < 0$. Show that $-1 < p/q - \sqrt{d} < 0$. Deduce that if $0 < -(p^2 - dq^2) \le \sqrt{d} - \frac{1}{2}$ then p/q is a convergent for $\sqrt{d}$.

Lemma C2.3. *The inequality* $\left|\alpha - \frac{p}{q}\right| \le \frac{1}{2q^2}$ *is satisfied for at least one of* $p/q = p_n/q_n$ *or* p_{n+1}/q_{n+1} *for each* $n \ge 0$.

Proof. If not then, since $\alpha - \frac{p_n}{q_n}$ and $\alpha - \frac{p_{n+1}}{q_{n+1}}$ have opposite signs, hence

$$\frac{1}{q_nq_{n+1}} = \left|\frac{p_n}{q_n} - \frac{p_{n+1}}{q_{n+1}}\right| = \left|\alpha - \frac{p_n}{q_n}\right| + \left|\alpha - \frac{p_{n+1}}{q_{n+1}}\right| > \frac{1}{2q_n^2} + \frac{1}{2q_{n+1}^2},$$

which is false for any positive reals q_n, q_{n+1}.

Hurwitz showed that for at least one of every three convergents one can improve this to $\le 1/(\sqrt{5}q^2)$ and that this is best possible, since this is the constant for the golden ratio $\frac{1+\sqrt{5}}{2}$:

`Exercise C2.3.9.` Show that $\frac{1+\sqrt{5}}{2} = [1, 1, 1, 1, \dots]$ and so the convergents are F_{n+1}/F_n where F_n is the nth Fibonacci numbers. By using the general formula for Fibonacci numbers, determine how good these approximations are; i.e. prove a strong version of the formula at the end of chapter 11:

$$\left|\frac{1+\sqrt{5}}{2} - \frac{F_{n+1}}{F_n} + \frac{(-1)^n}{\sqrt{5}F_n^2}\right| \le \frac{1}{2F_n^4}\ .$$

C2.4. Quadratic irrationals and periodic continued fractions. We just saw that the continued fraction for $\frac{1+\sqrt{5}}{2}$ is just 1 repeated infinitely often. What are the values of continued fractions in which the entries are periodic? We use the notation $\alpha = [\overline{a_0,\ a_1,\ \dots,\ a_n}]$ to mean $\alpha = [a_0,\ a_1, \dots,\ a_n,\ a_0,\ a_1, \dots,\ a_n,\ a_0,\ a_1, \dots,\ a_n, \dots]$ is periodic with period $a_0,\ a_1,\ \dots,\ a_n$. This means that $\alpha = [a_0,\ a_1,\ a_2,\ \dots,\ a_n,\ \alpha]$; that is $\alpha_{n+1} = \alpha$ and so, as above,

$$\alpha \ = \ \frac{\alpha p_n + p_{n-1}}{\alpha q_n + q_{n-1}} \ . \tag{C2.4}$$

This implies that $q_n\alpha^2 + (q_{n-1} - p_n)\alpha - p_{n-1} = 0$, that is α satisfies a quadratic equation. This equation must be irreducible, that is α is irrational, else the continued fraction would be of finite length (as we saw in section 1.3).

If $\gamma = [\overline{a_n,\ a_{n-1},\ \dots,\ a_0}]$ then, since

$$\begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a_{n-1} & 1 \\ 1 & 0 \end{pmatrix}\cdots\begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} = \left(\begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix}\cdots\begin{pmatrix} a_k & 1 \\ 1 & 0 \end{pmatrix}\right)^T = \begin{pmatrix} p_k & q_k \\ p_{k-1} & q_{k-1} \end{pmatrix},$$

hence $\gamma \ = \ \frac{\gamma p_n + q_n}{\gamma p_{n-1} + q_{n-1}}$ and so $p_{n-1}\gamma^2 + (q_{n-1} - p_n)\gamma - q_n \ = \ 0$, which implies that $q_n(-1/\gamma)^2 + (q_{n-1} - p_n)(-1/\gamma) - p_{n-1} = 0$; that is $-1/\gamma$ satisfies the same quadratic equation as α. However these are two distinct roots since both $\alpha > 0 > -1/\gamma$. We call $-1/\gamma$ the *conjugate* of α.

It may be that $\alpha = [a_0,\ a_1,\ a_2,\ \dots,\ a_m, \overline{b_0,\ b_1,\ \dots,\ b_n}]$ is eventually periodic. In that case $\beta := [\overline{b_0,\ b_1,\ \dots,\ b_n}]$ is quadratic irrational, and hence so is $\alpha = \frac{\beta p_m + p_{m-1}}{\beta q_m + q_{m-1}}$.

Let us suppose that $\alpha = u + v\sqrt{d}$, with d squarefree, has a periodic continued fraction of period m. Then (C2.4) is satisfied whenever n is a multiple of m. Hence

$$u + v\sqrt{d} \ = \ \alpha \ = \ \frac{p_n - q_{n-1} + \sqrt{(q_{n-1} - p_n)^2 + 4p_{n-1}q_n}}{2q_n}, \tag{C2.5}$$

so that $(q_{n-1} + p_n)^2 + 4(-1)^n = (q_{n-1} - p_n)^2 + 4p_{n-1}q_n = d(2q_nv)^2$. Since the left side is an integer, so is the right side, and so we have infinitely many solutions to *Pell's equation*

$$x^2 - dy^2 = \pm 4.$$

A continued fraction $\beta = [b_0,\ \dots,\ b_m, \overline{a_0,\ a_1,\ \dots,\ a_n}]$, for any $m \geq 0$ is called *eventually periodic*. Note that if $\alpha = [\overline{a_0,\ a_1,\ \dots,\ a_n}]$ then

$$\beta \ = \ \frac{\alpha p_m + p_{m-1}}{\alpha q_m + q_{m-1}} \ .$$

Theorem C2.4. *Any quadratic irrational real number has a continued fraction that is eventually periodic.*

Proof. Suppose that α has minimal polynomial $ax^2+bx+c = a(x-\alpha)(x-\beta)$, and define

$$f(x,y) := ax^2+bxy+cy^2 = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

By (C2.4), $\begin{pmatrix} \alpha \\ 1 \end{pmatrix} = \kappa \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} \begin{pmatrix} \alpha_{n+1} \\ 1 \end{pmatrix}$ for some $\kappa \neq 0$, and so if we define

$$\begin{pmatrix} A_n & B_n/2 \\ B_n/2 & C_n \end{pmatrix} := \begin{pmatrix} p_n & q_n \\ p_{n-1} & q_{n-1} \end{pmatrix} \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix}$$

(so that $b^2-4ac = B_n^2 - 4A_nC_n$ by taking determinants of both sides) then

$$\begin{aligned} A_n\alpha_{n+1}^2 + B_n\alpha_{n+1} + C_n &= \begin{pmatrix} \alpha_{n+1} & 1 \end{pmatrix} \begin{pmatrix} A_n & B_n/2 \\ B_n/2 & C_n \end{pmatrix} \begin{pmatrix} \alpha_{n+1} \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} \alpha_{n+1} & 1 \end{pmatrix} \begin{pmatrix} p_n & q_n \\ p_{n-1} & q_{n-1} \end{pmatrix} \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} \begin{pmatrix} \alpha_{n+1} \\ 1 \end{pmatrix} \\ &= \kappa^2 \begin{pmatrix} \alpha & 1 \end{pmatrix} \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \begin{pmatrix} \alpha \\ 1 \end{pmatrix} = \kappa^2 f(\alpha,1) = 0. \end{aligned}$$

Therefore $f_n(x) := A_nx^2+B_nx+C_n$ has root α_{n+1}. Now $A_n = f(p_n,q_n)$ and $C_n = A_{n-1}$. By (C2.3), $\left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{q_n^2} \leq 1$, and $\left|\beta - \frac{p_n}{q_n}\right| \leq |\beta-\alpha| + \left|\alpha - \frac{p_n}{q_n}\right| \leq |\beta-\alpha|+1$, so that

$$|A_n| = |f(p_n,q_n)| = aq_n^2 \left|\alpha - \frac{p_n}{q_n}\right| \left|\beta - \frac{p_n}{q_n}\right| \leq a(|\beta-\alpha|+1),$$

Since the A_n are all integers, there are only finitely many possibilities for the values of A_n and C_n. Moreover, given these values there are only two possibilities for B_n, as $B_n^2 = b^2-4ac+4A_nC_n$. Hence there are only finitely many possible triples $f_n(x)$ and each corresponds to at most two roots, so one such root must repeat infinitely often. That is, there exists $m<n$ such that $\alpha_m = \alpha_n$.

`Exercise C2.4.1` Deduce that the continued fraction for α is eventually periodic.

Proposition C2.5. *Suppose that α is a real quadratic irrational number with conjugate β. Then α has a periodic continued fraction if and only if $\alpha > 1$ and $0 > \beta > -1$.*

Proof. By Theorem C2.4 the continued fraction of α is eventually periodic. This implies that each $\alpha_n > a_n \geq 1$ for all $n \geq 1$ and we now show that if $0 > \beta > -1$ then $0 > \beta_n > -1$ for all $n \geq 1$ by induction. Since $\alpha_{n-1} = a_{n-1} + 1/\alpha_n$ by definition, we have $\beta_{n-1} = a_{n-1} + 1/\beta_n$ by taking conjugates. This means that $a_{n-1} = -1/\beta_n + \beta_{n-1}$ is an integer in $(-1/\beta_n - 1, -1/\beta_n)$ and so $a_{n-1} = [-1/\beta_n]$ and hence $-1/\beta_n > 1$ implying that $0 > \beta_n > -1$. Since the continued fraction for α is periodic, there exists $0 \leq m < n$ with

$\alpha_m = \alpha_n$; select m to be the minimal integer ≥ 0 for which such an n exists. Then $m = 0$ else taking conjugates gives $\beta_m = \beta_n$, so that $a_{m-1} = [-1/\beta_m] = [-1/\beta_n] = a_{n-1}$ and hence $\alpha_{m-1} = a_{m-1} + 1/\alpha_m = a_{n-1} + 1/\alpha_n = \alpha_{n-1}$, contradicting the minimality of m.

On the other hand if the continued fraction is purely periodic of period n then, as above $f(x) := q_n x^2 + (q_{n-1} - p_n)x - p_{n-1} = 0$ for $x = \alpha$ and β. Now $f(0) = -p_{n-1} < 0$ and $f(-1) = (q_n - q_{n-1}) + (p_n - p_{n-1}) > 0$, and so f has a root in $(-1, 0)$. This root cannot be α which is $\geq a_0 = a_n \geq 1$ so must be β.

C2.5. Pell's equation. Here are some examples of the continued fraction for $\sqrt{d}$:

$\sqrt{2} = [1, \overline{2}]$, $\sqrt{3} = [1, \overline{1, 2}]$, $\sqrt{5} = [2, \overline{4}]$, $\sqrt{6} = [2, \overline{2, 4}]$, $\sqrt{7} = [2, \overline{1, 1, 1, 4}]$, $\sqrt{8} = [2, \overline{1, 4}]$, $\sqrt{10} = [3, \overline{6}]$, $\sqrt{11} = [3, \overline{3, 6}]$, $\sqrt{12} = [3, \overline{2, 6}]$, $\sqrt{13} = [3, \overline{1, 1, 1, 1, 6}], \dots$

These examples seem to suggest that $\sqrt{d} = [a_0, \overline{a_1, \dots, a_n}]$ where $a_n = 2a_0 = 2[\sqrt{d}]$. Let us suppose, for now, that this is true, so that $\sqrt{d} + [\sqrt{d}]$ and $1/(\sqrt{d} - [\sqrt{d}])$ are (purely) periodic.

`Exercise C2.5.1.` Show that $\sqrt{d} + [\sqrt{d}]$ is indeed periodic.

If $\sqrt{d} = [a_0, a_1, \dots]$ then $\sqrt{d} + [\sqrt{d}] = [\overline{2a_0, a_1, \dots, a_{n-1}}]$ for some n, by exercise C2.5.1, so that $\sqrt{d} = [a_0, \overline{a_1, \dots, a_n}]$ where $a_n = 2a_0$ (as suggested by the examples). Now if P_k/Q_k are the convergents for $\sqrt{d} + [\sqrt{d}]$ then we deduce from the coefficients in (C2.5) that

$$(Q_{n-2} + P_{n-1})^2 - d(2Q_{n-1})^2 = 4(-1)^n \quad \text{and} \quad P_{n-1} - Q_{n-2} = 2a_0 Q_{n-1}.$$

`Exercise C2.5.2.` Show that $P_k/Q_k = p_k/q_k + a_0$ for all k (Hint: Use matrices to evaluate the P_k, Q_k, p_k, q_k.); that is $Q_k = q_k$ and $P_k = p_k + a_0 q_k$. Deduce from this and the last displayed equation that $Q_{n-2} + P_{n-1} = 2p_{n-1}$ and so

$$p_{n-1}^2 - dq_{n-1}^2 = (-1)^n.$$

Hence we have seen, in exercise C2.3.7, that if $d \geq 64$ and $p^2 - dq^2 = \pm 4$ with $p, q \geq 1$ then p/q is a convergent to $\sqrt{d}$. Now we see that each period of the continued fraction of $\sqrt{d}$ gives rise to another solution of the Pell equation.

If one takes a slightly larger example like $\sqrt{43} = [6, \overline{1, 1, 3, 1, 5, 1, 3, 1, 1, 12}]$ one cannot help but notice that the period is symmetric, that is $a_j = a_{n-j}$ for $j = 1, 2, \dots, n-1$. To prove this is straightforward: At the beginning of section C2.4 we saw that if we have $\gamma = [\overline{a_{n-1}, a_{n-2}, \dots, a_1, 2a_0}]$, then $-1/\gamma$ is the conjugate of $\sqrt{d} + [\sqrt{d}]$, that is $1/\gamma = \sqrt{d} - [\sqrt{d}]$ and therefore

$$\begin{aligned}[\overline{2a_0, a_1, \dots, a_{n-1}}] &= \sqrt{d} + [\sqrt{d}] = 2a_0 + 1/\gamma \\ &= [2a_0, \overline{a_{n-1}, \dots, a_1, 2a_0}] = [\overline{2a_0, a_{n-1}, \dots, a_1}].\end{aligned}$$

Remark: We have yet to show that the solutions to Pell's equation are precisely those that come from the period.

C2.6. The size of solutions to Pell's equation. As in the proof of Theorem C2.4 but now noting that $\left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{q_n q_{n+1}} < \frac{1}{a_n q_n^2}$ since $q_{n+1} = a_n q_n + q_{n-1} > a_n q_n$, we obtain that

$$1 \le |A_n| = aq_n^2 \left|\alpha - \frac{p_n}{q_n}\right| \left|\beta - \frac{p_n}{q_n}\right| \le aq_n^2 \cdot \frac{1}{q_n q_{n+1}} \left(|\beta - \alpha| + 1\right) < \frac{(a + \sqrt{d})}{a_n};$$

as $(a(\beta - \alpha))^2 = b^2 - 4ac = d$. For $\alpha = \sqrt{d} + [\sqrt{d}]$ we have $a = 1$, and so $1 \le a_n \le 2\sqrt{d} + 1$, for all $n \ge 1$. This allows us to deduce upper and lower bounds on p_n and q_n:

`Exercise C2.6.1.` Suppose that $x_{n+1} = a_n x_n + x_{n-1}$ for all $n \ge 1$, with x_0, x_1 positive integers, not both 0.

(i) Use that each $a_n \ge 1$ to deduce that $x_n \ge F_n$ for all $n \ge 0$.

(ii) Use that each $a_n \le B$ $(= \sqrt{d} + 1)$ to deduce that $x_n \le (B+1)^{n-1}(x_1 + x_0)$ for all $n \ge 1$.

Hence if the continued fraction for $\sqrt{d}$ has period ℓ and this gives rise to a solution x, y to Pell's equation, then $\phi^\ell \ll \epsilon_d := x + y\sqrt{d} \ll (\sqrt{d} + 1)^\ell$ where $\phi = \frac{1+\sqrt{5}}{2}$. Hence there is a direct link between the size of the smallest solution to Pell's equations and the length of the continued fraction.

C2.7. More examples of Pell's equation. Here we give only the longest continued fractions and the largest fundamental solutions.

$$\begin{aligned}
\sqrt{2} &= [1, \overline{2}], & 1^2 - 2 \cdot 1^2 &= -1 \\
\sqrt{3} &= [1, \overline{1, 2}], & 2^2 - 3 \cdot 1^2 &= 1 \\
\sqrt{6} &= [2, \overline{2, 4}], & 5^2 - 6 \cdot 2^2 &= 1 \\
\sqrt{7} &= [2, \overline{1, 1, 1, 4}], & 8^2 - 7 \cdot 3^2 &= 1 \\
\sqrt{13} &= [3, \overline{1, 1, 1, 1, 6}], & 18^2 - 13 \cdot 5^2 &= -1 \\
\sqrt{19} &= [4, \overline{2, 1, 3, 1, 2, 8}], & 170^2 - 19 \cdot 39^2 &= 1 \\
\sqrt{22} &= [4, \overline{1, 2, 4, 2, 1, 8}], & 197^2 - 22 \cdot 42^2 &= 1 \\
\sqrt{31} &= [5, \overline{1, 1, 3, 5, 3, 1, 1, 10}], & 1520^2 - 31 \cdot 273^2 &= 1 \\
\sqrt{43} &= [6, \overline{1, 1, 3, 1, 5, 1, 3, 1, 1, 12}], & 3482^2 - 43 \cdot 531^2 &= 1 \\
\sqrt{46} &= [6, \overline{1, 3, 1, 1, 2, 6, 2, 1, 1, 3, 1, 12}], & 24335^2 - 46 \cdot 3588^2 &= 1 \\
\sqrt{76} &= [8, \overline{1, 2, 1, 1, 5, 4, 5, 1, 1, 2, 1, 16}], & 57799^2 - 76 \cdot 6630^2 &= 1 \\
\sqrt{94} &= [9, \overline{1, 2, 3, 1, 1, 5, 1, 8, 1, 5, 1, 1, 3, 2, 1, 18}], & 2143295^2 - 94 \cdot 221064^2 &= 1 \\
\sqrt{124} &= [11, \overline{7, 2, 1, 1, 1, 3, 1, 4, 1, 3, 1, 1, 1, 2, 7, 22}], & 4620799^2 - 124 \cdot 414960^2 &= 1 \\
\sqrt{133} &= [11, \overline{1, 1, 7, 5, 1, 1, 1, 2, 1, 1, 1, 5, 7, 1, 1, 22}], & 2588599^2 - 133 \cdot 224460^2 &= 1 \\
\sqrt{139} &= [11, \overline{1, 3, 1, 3, 7, 1, 1, 2, 11, 2, 1, 1, 7, 3, 1, 3, 1, 22}], & 77563250^2 - 139 \cdot 6578829^2 &= 1.
\end{aligned}$$

These are the champions up to 150. After that we list the continued fraction lengths and the fundamental solutions for the champions up to 1000:

$$\text{Length} = 20: \quad 1728148040^2 - 151 \cdot 140634693^2 = 1$$
$$\text{Length} = 22: \quad 1700902565^2 - 166 \cdot 132015642^2 = 1$$
$$\text{Length} = 26: \quad 278354373650^2 - 211 \cdot 19162705353^2 = 1$$
$$\text{Length} = 26: \quad 695359189925^2 - 214 \cdot 47533775646^2 = 1$$
$$\text{Length} = 26: \quad 5883392537695^2 - 301 \cdot 339113108232^2 = 1$$
$$\text{Length} = 34: \quad 2785589801443970^2 - 331 \cdot 153109862634573^2 = 1$$
$$\text{Length} = 37: \quad 440424456968214 18^2 - 421 \cdot 2146497463530785^2 = -1$$
$$\text{Length} = 40: \quad 840560915469529337 75^2 - 526 \cdot 36650197573242955 32^2 = 1$$
$$\text{Length} = 42: \quad 181124355061630786130^2 - 571 \cdot 75798183506289825 87^2 = 1$$
$$\text{Length} = 44: \quad 59729912963116831 99^2 - 604 \cdot 243037569063951720^2 = 1$$
$$\text{Length} = 48: \quad 489615753129986500355 60^2 - 631 \cdot 194912953757515103 6427^2 = 1$$
$$\text{Length} = 52: \quad 729331846679488242441896 0^2 - 751 \cdot 26613697067720602445 6793^2 = 1$$
$$\text{Length} = 54: \quad 774352459305765585163776 5^2 - 886 \cdot 26014879646402419485 0378^2 = 1$$
$$\text{Length} = 60: \quad 448160301093711945155126372 0^2 - 919 \cdot 14783444239653675978149958 9^2 = 1$$
$$\text{Length} = 60: \quad 379516400906811930638014896080^2 - 991 \cdot 120557357903313594474425387 67^2 = 1$$

Notice that the length of the continued fractions here are around $2\sqrt{d}$, and the size of the fundamental solutions $10^{\sqrt{d}}$.

C3. Unique Factorization. The proof of the Fundamental Theorem of Arithmetic appears to use very few ideas, and so one might expect that it generalizes into all sorts of other domains. For example, do polynomials factor in a unique way into irreducibles? Or numbers of the form $\{a + b\sqrt{d} : \ a, b \in \mathbb{Z}\}$? Or other simple arithmetic sets?

One good example is the set of positive integers, $\mathcal{F}$, which are $\equiv 1 \pmod 4$. We note that $\mathcal{F}$ is closed under multiplication by exercise 3.1.2 and contains 1 (just like the positive integers). We know that this is an artificial set, in the sense that 21 factors over the integers into $3 \cdot 7$, but not in $\mathcal{F}$ since neither 3 nor 7 belongs to $\mathcal{F}$.[9] Indeed 21 is not the product of two smaller elements of $\mathcal{F}$ and so we call it *irreducible* in $\mathcal{F}$.[10] We ask whether factorization into irreducibles is unique *in* $\mathcal{F}$? A few calculations and we find that the answer is "no", since

$$441 = 9 \cdot 49 = 21 \cdot 21 \quad \text{or} \quad 693 = 9 \cdot 77 = 21 \cdot 33,$$

and each of $9, 21, 33, 49$ and 77 are *irreducible* in $\mathcal{F}$. What has gone wrong? What is the structural difference between $\mathcal{F}$ and $\mathbb{Z}$? One key difference is that $\mathbb{Z}$ has an additive structure (that is, any two elements of $\mathbb{Z}$ add to another), whereas $\mathcal{F}$ does not. So even though factorization is a multiplicative property, it somehow needs additive structure to be unique.

Before embarking on the next example we need to exclude examples like $3 = (-1)\cdot(-3)$ and $-3 = (-1)\cdot 3$, which make it appear that every integer can be factored into at least two others. The issue here is that one of the factors, -1, divides 1 in $\mathbb{Z}$, that is $(-1)\cdot(-1) = 1$, so division by such a number does not really reduce the size of numbers that we are working with. Elements of a ring that divide 1 are called *units* and will be excluded from the notion of factorization.

How about rings like $\{a + b\sqrt{d} : \ a, b \in \mathbb{Z}\}$ which have are closed under addition as well as multiplication? In $R := \{a + b\sqrt{-5} : \ a, b \in \mathbb{Z}\}$ we have the example

$$6 = 2 \cdot 3 = (1 + \sqrt{-5}) \cdot (1 - \sqrt{-5}).$$

Now suppose that prime $p = (a+b\sqrt{-5})(c+d\sqrt{-5})$ for $p = 2$ or 3. Then $(a, b)(c, d)$ divides p, so at least one of these gcds equals 1, say $(a, b) = 1$. The coefficient of the imaginary part is $ad + bc = 0$, and so $a|bc$ and therefore $a|c$. Writing $c = ak$ we have $d = -bk$ and therefore $p = k(a+b\sqrt{-5})(a-b\sqrt{-5}) = k(a^2+5b^2)$. Therefore $a^2+5b^2 = 1, 2$ or 3, so that $b = 0,\ a = \pm 1$, yielding the uninteresting factorization $(\pm 1)(\pm p)$, and hence neither 2 nor 3 can be factored in R. Therefore 2 and 3 are irreducible in R. Also, by taking complex conjugates $1+\sqrt{-5} = (a+b\sqrt{-5})(c+d\sqrt{-5})$ if and only if $1-\sqrt{-5} = (a-b\sqrt{-5})(c-d\sqrt{-5})$, and hence $6 = (1 + \sqrt{-5})(1 - \sqrt{-5}) = (a + b\sqrt{-5})(c + d\sqrt{-5})(a - b\sqrt{-5})(c - d\sqrt{-5}) = (a^2 + 5b^2)(c^2 + 5d^2)$. Since $a^2 + 5b^2$ and $c^2 + 5d^2$ are both positive and $\neq 2$ or 3, with product 6, one must equal 1, the other 6, say $|a| = |b| = |c| = 1,\ d = 0$, which shows that $1 + \sqrt{-5}$ and $1 - \sqrt{-5}$ are both irreducible in R. So we *do not have* unique factorization of elements of R.

---

[9] However, one might similarly argue that just as $\mathcal{F} \subset \mathbb{Z}$ so $\mathbb{Z} \subset \mathbb{Q}$, and thus we could argue 3 reduces further as $2 \cdot (3/2)$.

[10] Since we hesitate to use the word "prime" in this context.

Ideals, again. Let us suppose that in the ring $R = \mathbb{Z}[\sqrt{-d}] := \{a + b\sqrt{-d} : \ a, b \in \mathbb{Z}\}$, where $d > 1$, there are two ways of factoring an integer of R into irreducibles, say

$$p \cdot q = (a + b\sqrt{-d})(a - b\sqrt{-d}),$$

where p and q are distinct primes of $\mathbb{Z}$. If we had such an equation over the integers, say $pq = rs$ we might proceed by first computing $g = \gcd(p, r)$ (that is the (unique) generator of $I_{\mathbb{Z}}(p, r)$), and then writing $p = gP$, $r = gR$ so that $Pq = Rs$ where $(P, R) = 1$. Hence $P|Rs$ and $(P, R) = 1$ so that $P|s$. Writing $s = PS$ we obtain $q = RS$. Hence $pq = rs$ further factors as $gPRS = (gP)(RS) = (gR)(PS)$.

If we proceed like this in R then we need that the ideal generated by p and $a+b\sqrt{-d}$, namely $I_R(p, a+b\sqrt{-d})$, is principal (so we can divide through by the generator). We saw in the Theorem above, that either p divides a and b, whence $p^2|a^2 + db^2 = pq$ which is impossible, or p can be written as u^2+dv^2 where $u/v \equiv a/b \pmod p$ and $I_R(p, a+b\sqrt{-d}) = I_R(u + v\sqrt{-d})$.

Let us write $a+b\sqrt{-d} = (u+v\sqrt{-d})(\ell+m\sqrt{-d})$, so that $a-b\sqrt{-d} = (u-v\sqrt{-d})(\ell-m\sqrt{-d})$ by taking complex conjugates. Substituting in above we have

$$q = \frac{pq}{u^2 + dv^2} = \frac{a + b\sqrt{-d}}{u + v\sqrt{-d}} \cdot \frac{a - b\sqrt{-d}}{u - v\sqrt{-d}} = (\ell + m\sqrt{-d})(\ell - m\sqrt{-d}) = \ell^2 + dm^2.$$

Hence $pq = (a + b\sqrt{-d})(a - b\sqrt{-d})$ factors further as

$$\big((u + v\sqrt{-d})(u - v\sqrt{-d})\big) \cdot \big((\ell + m\sqrt{-d})(\ell - m\sqrt{-d})\big) = \big((u + v\sqrt{-d})(\ell + m\sqrt{-d})\big) \cdot \big((u - v\sqrt{-d})(\ell - m\sqrt{-d})\big).$$

`Exercise C3.1.` Show that the only units of R are 1 and -1. What if we allow $d = 1$?

Factoring into ideals. We need arithmetic to work in $R = \mathbb{Z}[\sqrt{-d}]$ even though, as we have seen we cannot always factor uniquely into irreducibles. It turns out that the way to proceed is to replace all the numbers in the ring by the ideals that they generate. To do so we need to be able to multiply ideals, and it is easy to show from their definition that this works out by multiplying generators: For any $\alpha, \beta, \gamma, \delta \in R$ we have

$$I_R(\alpha, \beta) \cdot I_R(\gamma, \delta) = I_R(\alpha\gamma, \alpha\delta, \beta\gamma, \beta\delta).$$

Therefore if $n = ab$ in R then $I_R(n) = I_R(a)I_R(b)$. There are several desirable properties of ideals: All issues with units disappear for if I is an ideal and u a unit then $I = uI$. Ideals can be factored into prime ideals in a unique way; in all "number rings" R we get unique factorization. Note though that primes are no longer elements of the ring, or even necessarily principal ideals of the ring.

In our example $6 = 2 \cdot 3 = (1 + \sqrt{-5}) \cdot (1 - \sqrt{-5})$ above, all of $2, 3, 1 + \sqrt{-5}, 1 - \sqrt{-5}$ are irreducibles of $\mathbb{Z}[\sqrt{-5}]$ but none generate prime ideals. In fact we can factor the ideals

they generate into prime ideals as

$$\begin{aligned} I_R(2) &= I_R(2, 1+\sqrt{-5}) \cdot I_R(2, 1-\sqrt{-5}) \\ I_R(3) &= I_R(3, 1+\sqrt{-5}) \cdot I_R(3, 1-\sqrt{-5}) \\ I_R(1+\sqrt{-5}) &= I_R(2, 1+\sqrt{-5}) \cdot I_R(3, 1+\sqrt{-5}) \\ I_R(1-\sqrt{-5}) &= I_R(2, 1-\sqrt{-5}) \cdot I_R(3, 1-\sqrt{-5}). \end{aligned}$$

None of these prime ideals are principal by Theorem *.*. We do call an element of R prime if it generates a prime ideal.

Here we see that the notion of "irreducible" and "prime" are not in general the same. In fact any prime of R is irreducible, but not vice-versa. One fun question of Davenport is to determine the most prime ideal factors an irreducible integer can have in R.

Maybe we should have Lenstra's brilliant example for $x^2 + 19 = y^3$.

C4. Binary quadratic forms with positive discriminant, and continued fractions. When $d > 0$, Gauss defined $ax^2 + bxy + cy^2$ to be *reduced* when

$$0 < \sqrt{d} - b < 2|a| < \sqrt{d} + b. \tag{C4.1}$$

This implies that $0 < b < \sqrt{d}$ so that $|a| < \sqrt{d}$ and therefore there are only finitely many reduced forms of positive discriminant d. Note that $ax^2 + bxy + cy^2$ is reduced if and only if $cx^2 + bxy + ay^2$ is. The first inequality implies that $ac = (b^2 - d)/4 < 0$.

Let $\rho_1 := \frac{-b+\sqrt{d}}{2a}$ and $\rho_2 := \frac{-b-\sqrt{d}}{2a}$ be the two roots of $at^2 + bt + c = 0$. Then (C4.1) holds if and only if $|\rho_1| < 1 < |\rho_2|$ and $\rho_1\rho_2 < 0$.

Forms $ax^2 + bxy + cy^2$ and $cx^2 + b'xy + c'y^2$ are *neighbours* (and equivalent) if they have the same discriminant and $b + b' \equiv 0 \pmod{2c}$, since they are equivalent under the transformation $\begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} 0 & -1 \\ 1 & k \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$ where $b + b' = 2ck$.

The reduction algorithm proceeds as follows: Given $ax^2 + bxy + cy^2$ we select a neighbour as follows: Let b_0' be the least residue in absolute value of $-b \pmod{2c}$ so that $|b_0'| \le c$.

- If $|b_0'| > \sqrt{d}$ then let $b' = b_0'$. Note that $0 < (b')^2 - d \le c^2 - d$ so that $|c'| = ((b')^2 - d)/4|c| < |c|/4$.
- If $|b_0'| < \sqrt{d}$ then select $b' \equiv -b \pmod{2c}$ with b' as large as possible so that $|b'| < \sqrt{d}$. Note that $-d \le (b')^2 - d = 4cc' < 0$. If $2|c| > \sqrt{d}$ then $|c'| \le |d/4c| < |c|$.

Otherwise $\sqrt{d} \ge 2|c|$ and $\sqrt{d} - 2|c| < |b'| < \sqrt{d}$, and therefore the neighbour is reduced. Thus we see that the absolute values of the coefficients a and c of the binary quadratic form are reduced at each step of the algorithm until we obtain a reduced form.

The major difference between this, the $d > 0$ case, and the $d < 0$ case is that there is not necessarily a unique reduced form in a given class of binary quadratic forms of positive discriminant. Rather, when we run Gauss's algorithm we eventually obtain a *cycle* of reduced forms, which must happen since every reduced form has a unique right and a unique left reduced neighbouring form, and there are only finitely many reduced forms. Given a quadratic form $a_0x^2 + b_0xy + a_1y^2$ we define a sequence of forms, in the following notation:

$$a_0 \ {}^{b_0} \ a_1 \ {}^{b_1} \ a_2 \ {}^{b_2} \ a_3 \ \dots \ .$$

This represents, successively, the forms $a_0x^2 + b_0xy + a_1y^2$, $a_1x^2 + b_1xy + a_2y^2$, $a_2x^2 + b_2xy + a_3y^2$, $\dots$, of equal discriminant, where a form is the unique reduced right neighbour of its predecessor, and then $a_{i+1} = (b_i^2 - d)/4a_i$. For example, when $d = 816$,

$$5 \ {}^{26} \ -7 \ {}^{16} \ 20 \ {}^{24} \ -3 \ {}^{24} \ 20 \ {}^{16} \ -7 \ {}^{26} \ 5 \ {}^{24} \ -12 \ {}^{24} \ 5 \ {}^{26} \ -7 \ \dots$$

which is a cycle of period 8.

A solution to Pell's Equation, $v^2 - dw^2 = \pm 4$ yields a map $\begin{pmatrix} X \\ Y \end{pmatrix} \to \begin{pmatrix} \frac{v-bw}{2} & -cw \\ aw & \frac{v+bw}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$, for which $aX^2 + bXY + cY^2 = \pm(ax^2 + bxy + cy^2)$, which is an automorphism only when

$v^2 - dw^2 = 4$. Any solution to Pell's Equation yields a good rational approximation $\frac{v}{w}$ to $\sqrt{d}$, in fact with $|\frac{v}{w} - \sqrt{d}| < \frac{1}{2w^2}$ if $d \geq 19$. This implies that $\frac{v}{w}$ is a convergent for the continued fraction of $\sqrt{d}$. For $\alpha = \sqrt{d}$ let $c_n := p_n^2 - dq_n^2$, so that $c_n c_{n+1} < 0$ and $|c_n| < 2\sqrt{d}+1$, and that there is a cycle of reduced forms $c_0\ ^{b_0}\ c_1\ ^{b_1}\ c_2\ ^{b_2}\ c_3\ \ldots$ of discriminant d. For example $\sqrt{60} = [7, \overline{1, 2, 1, 14}]$ gives rise to the cycle $-11\ ^4\ 4\ ^4\ -11\ ^7\ 1\ ^7\ -11\ ^4\ 4$, and the first 4 corresponds to the unit $\frac{8+\sqrt{60}}{2} = 4 + \sqrt{15}$. In general if $\frac{p_n}{q_n}$ is the nth convergent to $\frac{\sqrt{d}-b}{2|a|}$ then define $c_n = ap_n^2 \pm bp_nq_n + cq_n^2$ where $\pm$ represents the sign of a, and we have such a cycle. For example $\frac{\sqrt{97}-9}{8} = [0, \overline{9, 2, 2, 1, 4, 4, 1, 2, 2}]$, which gives the cycle $-1\ ^9\ 4\ ^7\ -3\ ^5\ 6\ ^7\ -2\ ^9\ 2\ ^7\ -6\ ^5\ 3\ ^7\ -4\ ^9\ 1\ ^9\ -4\ ^7\ 3\ ^5\ -6\ ^7\ 2\ ^9\ -2\ ^7\ 6\ ^5\ -3\ ^7\ 4\ ^9\ -1\ ^9\ 4 \ldots$.

The *fundamental unit* is that solution $\epsilon_d := \frac{v_0+\sqrt{d}w_0}{2}$ which is minimal and > 1. We call $\frac{v^2-dw^2}{4}$ the *norm* of ϵ_d. All other solutions of (4.6.2) take the form

$$\frac{v + \sqrt{d}w}{2} = \pm\epsilon_d^k, \tag{C4.1}$$

for some $k \in \mathbb{Z}$ (for a proof see exercise 4.6c). We let ϵ_d^+ be the smallest unit > 1 with norm 1. One can deduce from (C4.1) that $\epsilon_d^+ = \epsilon_d$ or ϵ_d^2, depending on whether the norm of ϵ_d is 1 or -1.

`Exercise C4.2.` Prove that every reduced form of positive discriminant has a unique right and a unique left reduced neighbouring form.

The class number one problem in real quadratic fields. Although $h(-d)$ gets large, roughly of size $\sqrt{d}$ as d gets larger, surprisingly $h(d)$ seems to mostly remains quite small. What we do know is that $h(d)\log \epsilon_d$ is roughly of size $\sqrt{d}$ as d gets larger, so that computational data suggests that ϵ_d is often around $e^{\sqrt{d}}$ whereas $h(d)$ stays small. There are exceptions; for example if $d = m^2 + 1$ then $\epsilon_d = m + \sqrt{d}$ and so we can prove, for such d, that $h(d)$ gets large (like $\sqrt{d}$).

Hooley, and Cohen and Lenstra, made some attempts to guess at how often $h(d)$ is small. One can show that there are distinct binary quadratic forms for each odd squarefree divisor of d and so $h(d) \geq 2^{\nu(d)}$, where $\nu(d)$ is the number of odd prime factors of d. Therefore the smallest that $h(d)$ can be is $2^{\nu(d)}$ (we call these the *idoneal numbers*) and therefore if $h(d) = 1$ then d must be prime. Gauss observed that $h(p) = 1$ for what seemed to be a positive proportion of primes p and this is still an open problem today. Even proving that there are infinitely many primes p for which $h(p) = 1$, is open.

Cohen and Lenstra made the following conjectures

$$\text{The proportion of } d \text{ for which } p \text{ divides } h(d) = 1 - \prod_{k\geq 2}\left(1 - \frac{1}{p^k}\right)$$

The proportion of primes p for which $h(p) = 1$ is $\lambda := \prod_{k\geq 2}\left(1 - \frac{1}{2^k}\right)\zeta(k) = .7544581517\ldots$ Then the proportion of primes p for which $h(p) = q$, where q is prime, equals $\lambda/q(q-1)$.

Finally if most of the class numbers are small, but the occasional one is as big as $\sqrt{p}$ then which dominates in the average? The conjecture is that for the primes $p \le x$ with $p \equiv 1 \pmod 4$, we have that $h(p) \sim \frac{1}{8} \log x$ on average (and thus, the big class numbers are very rare).

C5. SL$(2,\mathbb{Z})$-transformations. Forms-Ideals-Transformations.

Generators of SL$(2,\mathbb{Z})$. We will show that $\mathrm{SL}(2,\mathbb{Z})$ is generated by the two elements $\begin{pmatrix}1&1\\0&1\end{pmatrix}, \begin{pmatrix}0&1\\-1&0\end{pmatrix}$. Given $\begin{pmatrix}\alpha&\beta\\\gamma&\delta\end{pmatrix}$ of determinant 1, we shall perform the Euclidean algorithm on α/γ when $\gamma\neq 0$: Select integer a so that $\gamma' := \alpha - a\gamma$ has the same sign as α and $0\le|\gamma'|<\gamma$. If $\alpha' = -\gamma$ then $\begin{pmatrix}0&-1\\1&-a\end{pmatrix}\begin{pmatrix}\alpha&\beta\\\gamma&\delta\end{pmatrix} = \begin{pmatrix}\alpha'&\beta'\\\gamma'&\delta'\end{pmatrix}$. Other than the signs this is the same process as the Euclidean algorithm, and we reduce the size of the pair of numbers in the first column. Moreover the matrix $\begin{pmatrix}0&-1\\1&-a\end{pmatrix}$ has determinant 1, and therefore so does $\begin{pmatrix}\alpha'&\beta'\\\gamma'&\delta'\end{pmatrix}$. We repeat this process as long as we can; evidently this is impossible once $\gamma=0$. In that case α and δ are integers for which $\alpha\delta=1$ and therefore our matrix is $\pm I$. Hence we have that there exists integers $a_1, a_2, \dots, a_k$ such that

$$\begin{pmatrix}0&-1\\1&-a_k\end{pmatrix}\begin{pmatrix}0&-1\\1&-a_{k-1}\end{pmatrix}\cdots\begin{pmatrix}0&-1\\1&-a_1\end{pmatrix}\begin{pmatrix}\alpha&\beta\\\gamma&\delta\end{pmatrix} = \pm\begin{pmatrix}1&0\\0&1\end{pmatrix}.$$

Now $\begin{pmatrix}a&-1\\1&0\end{pmatrix}\begin{pmatrix}0&-1\\1&-a\end{pmatrix} = -I$, and so we deduce that

$$\begin{pmatrix}\alpha&\beta\\\gamma&\delta\end{pmatrix} = \pm\begin{pmatrix}a_1&-1\\1&0\end{pmatrix}\begin{pmatrix}a_2&-1\\1&0\end{pmatrix}\cdots\begin{pmatrix}a_k&-1\\1&0\end{pmatrix}$$

Now $\begin{pmatrix}a&-1\\1&0\end{pmatrix} = -\begin{pmatrix}1&a\\0&1\end{pmatrix}\begin{pmatrix}0&1\\-1&0\end{pmatrix} = -\begin{pmatrix}1&1\\0&1\end{pmatrix}^a\begin{pmatrix}0&1\\-1&0\end{pmatrix}$

`Exercise C5.1.` Complete the proof that $\mathrm{SL}(2,\mathbb{Z})$ is generated by the two elements $\begin{pmatrix}1&1\\0&1\end{pmatrix}, \begin{pmatrix}0&1\\-1&0\end{pmatrix}$.

We consider the binary quadratic form $f(x,y) := ax^2+bxy+cy^2$. We saw that two forms f and g are equivalent, written $f\sim g$, if there exists $\begin{pmatrix}\alpha&\beta\\\gamma&\delta\end{pmatrix}\in\mathrm{SL}(2,\mathbb{Z})$ such that $g(x,y) = f(\alpha x+\beta y, \gamma x+\delta y)$.

The root $z_f := \frac{-b+\sqrt{d}}{2a}$ of f is a point in $\mathbb{C}$, the sign of $\sqrt{d}$ chosen, when $d<0$, to be in the upper half plane. Two points in the complex plane z, z' are said to be equivalent if there exists $\begin{pmatrix}\alpha&\beta\\\gamma&\delta\end{pmatrix}\in\mathrm{SL}(2,\mathbb{Z})$ such that $z' = u/v$ where $\begin{pmatrix}u\\v\end{pmatrix} = \begin{pmatrix}\alpha&\beta\\\gamma&\delta\end{pmatrix}\begin{pmatrix}z\\1\end{pmatrix}$. Hence $z\sim z+1$ and $z\sim -1/z$.

The ideal $I_f := (2a, -b+\sqrt{d})$ corresponds to f. Note that $I_f = 2a(1, z_f)$. Two ideals I, J are said equivalent if there exists $\alpha\in\mathbb{Q}(\sqrt{d})$ such that $J=\alpha I$. Hence $I_f\sim(1,z_f)$.

The generators of $\mathrm{SL}(2,\mathbb{Z})$ correspond to two basic operations in Gauss's reduction algorithm for binary quadratic forms:

The first is $x\to x+y, y\to y$, so that

$$f(x,y)\sim g(x,y) := f(x+y,y) = ax^2+(b+2a)xy+(a+b+c)y^2.$$

Note that $I_g = (2a, -(b+2a)+\sqrt{d}) = I_f$, and $z_g = \frac{-b-2a+\sqrt{d}}{2a} = z_f - 1$.

The second is $x \to y,\ y \to -x$ so that

$$f(x,y) \sim h(x,y) := f(y,-x) = cx^2 - bxy + ay^2.$$

Note that $I_h = (2c, b+\sqrt{d})$, and $z_h = \frac{b+\sqrt{d}}{2c}$. First observe that

$$z_f \cdot z_h = \frac{-b+\sqrt{d}}{2a} \cdot \frac{b+\sqrt{d}}{2c} = \frac{d-b^2}{4ac} = -1$$

that is $z_h = -1/z_f$. Then

$$I_h \sim (1, z_h) = (1, -1/z_f) \sim (1, -z_f) = (1, z_f) \sim I_f.$$

Since any $\mathrm{SL}(2,\mathbb{Z})$-transformation can be constructed out of the basic two transformation we deduce

Theorem C5.1. *$f \sim f'$ if and only if $I_f \sim I_{f'}$ if and only if $z_f \sim z_{f'}$.*

It is amazing that this fundamental, non-trivial, equivalence can be understood in three seemingly very different ways. Which is the best? That is hard to say; each has their uses, but what is good one can translate any question into the setting in which it is most natural. For example the notion of reduced binary quadratic form seems a little unnatural; however in the context of points in the upper half plane it translates to the points

$$z \in \mathbb{C} :\ \mathrm{Im}(z) > 0,\ -\frac{1}{2} \le \mathrm{Re}(z) < \frac{1}{2},\ |z| \ge 1,\ \text{if } |z| = 1 \text{ then } \mathrm{Re}(z) \le 0.$$

Be careful here; we are out by a factor of 2, and we might wish to place z on the right not the left

C6. Minkowski and lattices. A *lattice* Λ in $\mathbb{R}^n$ is the set of points generated by n linearly independent vectors, with basis $x_1, x_2, \dots, x_n$ say. In other words

$$\Lambda := \{a_1x_1 + a_2x_2 + \ldots + a_nx_n : \ a_1, a_2, \dots, a_n \in \mathbb{Z}\}.$$

One can see that Λ is an additive group, but it also has some geometry connected to it. The *fundamental parallellopiped* of Λ with respect to $x_1, x_2, \dots, x_n$ is the set $P = \{a_1x_1 + a_2x_2 + \cdots + a_nx_n : 0 \le a_i < 1\}$. The sets $x + P$, $x \in \Lambda$ are disjoint and their union is $\mathbb{R}^n$. The *determinant* $\det(\Lambda)$ of Λ is the volume of P; in fact $\det(\Lambda) = |\det(A)|$, where A is the matrix with column vectors $x_1, x_2, \dots, x_n$ (written as vectors in $\mathbb{R}^n$). A *convex body* K is a bounded convex open subset of $\mathbb{R}^n$.

We define $A - B$ to be the set of points that can be expressed as $a - b$. A key result is:

Blichfeldt's Lemma. *Let $K \subset \mathbb{R}^n$ be a measurable set, and Λ a lattice such that* $\mathrm{vol}(K) > \det(\Lambda)$. *Then $K - K$ contains a non-zero point of Λ.*

Proof. (By the pigeonhole principle.) Let L be the set of points $\ell \in P$ such that there exists $x \in \Lambda$ for which $\ell + x \in K$. We claim that there are two such x for at least one point in L, else $\mathrm{vol}\ K = \mathrm{vol}\ L \le \mathrm{vol}\ P = \det(\Lambda) < \mathrm{vol}(K)$, by hypothesis, a contradiction. Therefore for $k_x := \ell + x \ne k_y := \ell + y \in K$ with $x, y \in \Lambda$ we have $k_x - k_y = x - y \in \Lambda$ which is the result claimed.

`Exercise C6.1` Show that if $\mathrm{vol}(K) > m\ \det(\Lambda)$. Then $K - K$ contains at least m non-zero points of Λ.

We deduce:

Minkowski's First Theorem. *If K is a centrally symmetric convex body with* $\mathrm{vol}(K) > 2^n\det(\Lambda)$ *then K contains a non-zero point of Λ.*

Proof. As K is convex and centrally symmetric, $K = \frac{1}{2}\ K - \frac{1}{2}\ K$. However, $\mathrm{vol}(\frac{1}{2}\ K) > \det(\Lambda)$, so the result follows by Blichfeldt's Lemma.

Another proof of the sum of two squares theorem. Suppose that p is a prime $\equiv 1 \pmod 4$ so that there exist integers a, b such that $a^2 + b^2 \equiv 0 \pmod p$. Let Λ be the lattice in $\mathbb{Z}^2$ generated by $(a, b), (-b, a)$.

`Exercise C6.2.` Prove that $\det(\Lambda) = p$. Show that if $(u, v) \in \Lambda$ then $u^2 + v^2 \equiv 0 \pmod p$.

Let $K := \{(x, y : \ x^2 + y^2 < 2p\}$ so that $\mathrm{vol}(K) = 2\pi p > 2^2\det(\Lambda)$. Minkowski's First Theorem implies that there exists a non-zero $(u, v) \in K \cap \Lambda$, so that $0 < u^2 + v^2 < 2p$ and $u^2 + v^2 \equiv 0 \pmod p$, which implies that $u^2 + v^2 = p$.

Another proof of the local-glocal principle for diagonal quadratic forms. Let a, b, c be given integers such that abc is coprime and all the residue symbols work out. Let Λ be the lattice in $\mathbb{Z}^3$ generated by solutions x, y, z to $ax^2 + by^2 + cz^2 \equiv 0 \pmod{4abc}$. We will prove that $\det(\Lambda) = 4|abc|$:

The first observation is that if, say, $p|a$ then we know, by hypothesis that there exist u, v with $bu^2 + cv^2 \equiv 0 \pmod p$, etc (To be understood).

Now let $K := \{(x, y, z : \ |a|x^2 + |b|y^2 + |c|z^2 < 4|abc|\}$ so that $\text{vol}(K) = \frac{8\pi}{3} \cdot 4|abc| > 2^3\det(\Lambda)$. Hence Minkowski's First Theorem implies that there exists a non-zero $(u, v, w) \in K \cap \Lambda$, such that $au^2 + bv^2 + cw^2 \equiv 0 \pmod{4abc}$ with $|au^2 + bv^2 + cw^2| \le |a|u^2 + |b|v^2 + |c|u^2 < 4|abc|$.

Hence we have shown that there exists a non-zero integer solution to

$$ax^2 + by^2 + cz^2 = 0, \text{with} \ \ |a|x^2 + |b|y^2 + |c|z^2 < 4|abc|.$$

`Exercise C6.3.` Can you improve the 4 in the last displayed equation?

`Exercise C6.4.` We may assume, wlog, that $a, b, c > 0$ and we are looking for solutions to $ax^2 + by^2 = cz^2$. Now try $\Lambda := \{(x, y, z) : \ ax^2 + by^2 + cz^2 \equiv 0 \pmod{2abc}\}$ with $K := \{(x, y, z : \ ax^2 + by^2, \ cz^2 < 2|abc|\}$. What do you get?

I am not sure whether we need this so we will not delete it for now! For a centrally symmetric convex body K define λ_k to be the infimum of those λ for which λK contains k linearly independent vectors of Λ. We call $\lambda_1, \lambda_2, \dots, \lambda_n$ the *successive minima* of K with respect to Λ. Let $b_1, b_2, \dots, b_n \in \mathbb{R}^n$ be linearly independent vectors with $b_k \in \lambda_k \overline{K} \cap \Lambda$ for each k. The proof of the next result, and much more, can be found in [15].

Minkowski's Second Theorem. *If $0 < \lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$ are the successive minima of convex body K with respect to Λ then $\lambda_1\lambda_2 \ldots \lambda_n \text{vol}(K) \le 2^n \det(\Lambda)$.*

Let $r_1, r_2, \dots, r_k \in \mathbb{Z}/N\mathbb{Z}$ and $\delta > 0$ be given. We define the *Bohr neighbourhood*

$$B(r_1, r_2, \dots, r_k; \delta) := \{s \in \mathbb{Z}/N\mathbb{Z} : \|r_i s/N\| \le \delta \text{ for } i = 1, 2, \dots, k\};$$

that is, the least residue, in absolute value, of each $r_i s \pmod N$ is $< \delta N$ in absolute value.

C7. Connection between sums of 3 squares and $h(d)$. We have seen which integers are representable as the sum of two squares. How about three?

• The only squares mod 4 are 0 and 1. Therefore if n is divisible by 4 and is the sum of three squares then all three squares must be even. Hence if $n = 4m$ then to obtain every representation of n as the sum of three squares, we just take every representation of m as the sum of three squares, and double the number that are being squared.

• The only squares mod 8 are $0, 1$ and 4. Therefore no integer $\equiv 7 \pmod 8$ can be written as the sum of three squares (of integers). By the previous remark no integer of the form $4^k(8m + 7)$ can be written as the sum of three squares.

Legendre's Theorem. (1798) *A positive integer n can be written as the sum of three squares of integers if and only if it is not of the form $4^k(8m + 7)$.*

We will not prove this as all known proofs are too complicated for a first course.

One might ask how many ways are there to write an integer as the sum of three squares? Gauss proved the following remarkable theorem (for which there is still no easy proof): Suppose that n is squarefree.[11] If $n \equiv 3 \pmod 8$ then there are $8h(-4n)$ ways in which n can be written as the sum of three squares; if $n \equiv 1$ or $2 \pmod 4$, $n > 1$ there are $12h(-4n)$ ways.

[11] That is $p^2 \not| n$ for all primes p.

C8. Eisenstein's proof of quadratic reciprocity. There are many proofs of the law of quadratic reciprocity, something like 233 at last count (see the list at `http://www.rzuser.uni-heidelberg.de/~hb3/fchrono.html`). One of the most elegant is due to Eisenstein (1844).

A lemma of Gauss gives a complicated formula to determine (a/p):

Gauss's Lemma. *For $(a,p)=1$, let r_n be the absolute least residue of $an \pmod p$, and $\mathcal{N}$ be the set of integers $1 \le n \le \frac{p-1}{2}$ such that $r_n < 0$. Then $\left(\frac{a}{p}\right) = (-1)^{|\mathcal{N}|}$.*

Proof. For each $m, 1 \le m \le \frac{p-1}{2}$ there is exactly one integer n, $1 \le n \le \frac{p-1}{2}$ such that $r_n = m$ or $-m \pmod p$ (for if $an \equiv \pm an' \pmod p$ then $p|a(n \mp n')$ so $p|n \mp n'$ which is possible in this range only if $n = n'$). Therefore

$$\left(\frac{p-1}{2}\right)! = \prod_{1\le m\le \frac{p-1}{2}} m = \prod_{\substack{1\le n\le \frac{p-1}{2}\\ n\notin\mathcal{N}}} r_n \cdot \prod_{\substack{1\le n\le \frac{p-1}{2}\\ n\notin\mathcal{N}}} (-r_n)$$

$$\equiv \prod_{\substack{1\le n\le \frac{p-1}{2}\\ n\notin\mathcal{N}}} (an) \cdot \prod_{\substack{1\le n\le \frac{p-1}{2}\\ n\notin\mathcal{N}}} (-an) = a^{\frac{p-1}{2}}(-1)^{|\mathcal{N}|} \cdot \left(\frac{p-1}{2}\right)! \pmod p.$$

The result follows from Euler's criterion.

Proof of Theorem 8.7 for primes. If $a = 2$ in Gauss's Lemma, $\mathcal{N} = \{1 \le n \le \frac{p-1}{2} :\ 2n > \frac{p}{2}\}$ so that $|\mathcal{N}| = \frac{p-1}{2} - [\frac{p-1}{4}]$, which equals $\frac{p-1}{4}$ if $p \equiv 1 \pmod 4$, and $\frac{p+1}{4}$ if $p \equiv 3 \pmod 4$. The result follows from Gauss's Lemma.

`Exercise C8.1.` Let r be the absolute least residue of $N \pmod p$. Prove that

$$N - p\left[\frac{N}{p}\right] = \begin{cases} r & \text{if } r \ge 0; \\ p + r & \text{if } r < 0. \end{cases}$$

By the last exercise we have

$$\sum_{n=1}^{\frac{p-1}{2}} \left(an - p\left[\frac{an}{p}\right]\right) = \sum_{n\notin\mathcal{N}} r_n + \sum_{n\in\mathcal{N}} (p + r_n)$$

where the sums are all restricted to n in the range $1 \le n \le \frac{p-1}{2}$. We will take a and p odd and study this equation mod 2. It is convenient to let $T \equiv \sum_{n=1}^{\frac{p-1}{2}} n \pmod 2$. Then the equation becomes

$$T + \sum_{n=1}^{\frac{p-1}{2}} \left[\frac{an}{p}\right] \equiv \sum_n r_n + |\mathcal{N}| \pmod 2.$$

In the proof of Gauss's Lemma we saw that

$$\{r_n :\ n \in \mathcal{N}\} \cup \{-r_n :\ 1 \le n \le \frac{p-1}{2},\ n \notin \mathcal{N}\} = \{m :\ 1 \le m \le \frac{p-1}{2}\},$$

so that

$$\sum_{n=1}^{\frac{p-1}{2}} r_n \equiv \sum_{n\in\mathcal{N}} r_n + \sum_{n\notin\mathcal{N}} (-r_n) = \sum_{m=1}^{\frac{p-1}{2}} m \equiv T \pmod 2.$$

Hence

$$|\mathcal{N}| \equiv \sum_{n=1}^{\frac{p-1}{2}} \left[\frac{an}{p}\right] \pmod 2.$$

We deduce from Gauss's lemma that

$$\left(\frac{a}{p}\right) = (-1)^{\sum_{n=1}^{\frac{p-1}{2}} \left[\frac{an}{p}\right]}. \tag{C8.1}$$

`Exercise C8.2.` Show that the number of lattice points $(x, y) \in \mathbb{Z}^2$ for which $py < qx$ with $0 < x < p/2$ is

$$\sum_{n=1}^{\frac{p-1}{2}} \left[\frac{qn}{p}\right].$$

There are $\frac{p-1}{2} \cdot \frac{q-1}{2}$ lattice points $(x, y) \in \mathbb{Z}^2$ for which $1 \le x \le \frac{p-1}{2}$ and $1 \le y \le \frac{q-1}{2}$. We split this region into the two parts on either side of the line $py = qx$. In the last exercise we saw how many such lattice points satisfy $py < qx$, and the same exercise, with the roles of p and q reversed, gives us the number of lattice points for which $py > qx$. There are none with $py = qx$. Hence we have

$$\sum_{m=1}^{\frac{p-1}{2}} \left[\frac{qm}{p}\right] + \sum_{n=1}^{\frac{q-1}{2}} \left[\frac{pn}{q}\right] = \frac{p-1}{2} \cdot \frac{q-1}{2}$$

Exponentiating this, and applying (C8.1) with $a = q$, and then with the roles of p and q reversed, we obtain the law of quadratic reciprocity:

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

`Exercise C8.3.` Prove that for every prime $p \ge 7$ there exists an integer $n \le 9$ such that n and $n + 1$ are quadratic residues mod p.

`Exercise C8.4.` Can you prove an analogous result for triples of quadratic residues? (Hint: Think mod 3, and try to generalize this)

C9. Higher reciprocity laws. Discuss that there is no simple law for cubic reciprocity. 2 is a fourth power mod p if and only $p = x^2 + 64y^2$

C10. Finite fields.

Congruences in number rings are a little more subtle. For example, what are the set of residue classes $m + ni \pmod 3$ with $m, n \in \mathbb{Z}$? It is evident that any such number is congruent to some $a + bi \pmod 3$ with $a, b \in \{0, 1, 2\}$; are any of these nine residue classes congruent? If two are congruent then take their difference to obtain $u \equiv iv \pmod 3$ for some integers u, v, not both 0, and each ≤ 2 in absolute value. But then $u - iv = 3(r - is)$ for some integers r, s and hence $3|u, v$ and hence $u = v = 0$, a contradiction

`Exercise C10.1.` Generalize this argument to show that there are exactly p^2 distinct residue classes amongst the integers $a + bi \pmod p$ for any prime p.

When we work mod p in $\mathbb{Z}$ we have all of the usual rules of addition, multiplication and even division. If we look back at Lemma 3.5 we only stop working mod p, when we divide through by an integer that is divisible by p, that is by an integer $\equiv 0 \pmod p$; expressed like this it still likes a regular rule of arithmetic, not dividing through by 0. Working mod a composite number $n = ab$ we see that dividing by a means that we stop working mod n, and yet $a \not\equiv 0 \pmod n$. The core issue is that the group $(\mathbb{Z}/n\mathbb{Z})^*$ has zero divisors; that is $ab \equiv 0 \pmod n$ with neither a nor b are $\equiv 0 \pmod n$.

What about mod p when working in $\mathbb{Z}[i]$?[12] Now if $p \equiv 1 \pmod 4$ we can write $p = a^2 + b^2$ and so $(a + bi)(a - bi) = p \equiv 0 \pmod p$ yet neither $a + bi$ nor $a - bi$ are $\equiv 0 \pmod p$. So the primes $\equiv 1 \pmod 4$ are composite number in $\mathbb{Z}[i]$, and in this sense behave that way. On the other hand for the primes $p \equiv 3 \pmod 4$, we have that if $(a+bi)(c+di) \equiv 0 \pmod p$ then p divides $(a+bi)(c+di)(a-bi)(c-di) = (a^2+b^2)(c^2+d^2)$. Hence p divides one of $a^2 + b^2$ and $c^2 + d^2$, say the first, so that p divides both a and b. But then $a + bi \equiv 0 \pmod p$, and so we have proved that there are no zero divisors mod p.

A set of numbers in which all the usual rules of addition, subtraction, multiplication and division hold is called a *field*. Its definition is that it is a set F, where F has an additive group and $F \setminus \{0\}$ has a multiplicative group, both commutative, their identity elements, denoted 0 and 1 respectively, are distinct, and that $a \cdot (b + c) = a \cdot b + a \cdot c$. Since $F \setminus \{0\}$ is a multiplicative group hence F has no zero divisors.

We now suppose that F is finite.

`Exercise C10.2.` By Lagrange's Theorem we know that $|F| \cdot 1 = 0$.

(1) Show that if prime q divides $|F|$ then either $q \cdot 1 = 0$ or $|F|/q \cdot 1 = 0$. Use an induction hypothesis to show that there exists a prime p such that $p \cdot 1 = 0$ in F.
(2) Now show that this prime p is unique. (For example, show that if $p \cdot 1 = q \cdot 1 = 0$ where p and q are distinct primes then deduce that $1 = 0$ (which is impossible).)
(3) Begin with a non-zero element $a_1 \in F$. Let $P = \{1, 2, \dots, p\}$. If $a_2 \notin I_P(a_1)$ then show that $I_P(a_1, a_2)$ has p^2 distinct elements. Hence by induction show that there are p^r distinct elements of F given by $I_P(a_1, a_2, \dots, a_r)$.

[12] This needs explaining earlier. The polynomials in i with integer coefficients; since we can replace i^2 by -1 we see that all elements here can be written as $a + bi$ with $a, b \in \mathbb{Z}$.

Hence we deduce that the only finite fields have $q = p^r$ elements for some prime p and integer $r \geq 1$. It can be shown that, up to isomorphism, there is just one such field for each prime power. We denote this field as $\mathbb{F}_q$.

The easiest way to construct a finite field of p^r elements is to use a root α of a polynomial $f(x)$ of degree r which is irreducible in $\mathbb{F}_p$. (roughly 1 in r polynomials of degree r are irreducible). Then we can represent the elements of the finite field as $a_0 + a_1\alpha + \ldots + a_{r-1}\alpha^{r-1}$ where we take the $a_i \in \mathbb{F}_p$.

`Exercise C10.3.` Verify that this indeed gives the field on p^r elements.

`Exercise C10.4.` Show that the multiplicative group is indeed cyclic (??) We call this a primitive root

The multiplicative group of $\mathbb{F}_{p^r}$ has $p^r - 1$ elements so that $a^{p^r-1} = 1$ for all $a \in F$ by Lagrange's Theorem. Therefore $a^{p^r} = a$. Hence the map $x \to x^p$ partitions the field into orbits of size $\leq r$ of the form $a, a^p, a^{p^2}, a^{p^3}, \ldots, a^{p^{r-1}}$. In particular note that since $p = 0$ we can use the multinomial theorem to note that $f(x^p) = f(x)^p$. Hence if a is a root of polynomial over $\mathbb{F}_p$ of degree r then $a, a^p, a^{p^2}, a^{p^3}, \ldots, a^{p^{r-1}}$ are r distinct roots of the polynomial. This implies, for instance, that if g is a primitive root then g is the root of an irreducible polynomial of degree r over $\mathbb{F}_p$.

The integers mod p are not only a field but isomorphic to $\mathbb{F}_p$.

`Exercise C10.5.` Show that the finite field on p^2 elements is quite different from the integers mod p^2.

C11. Affine vs. Projective. When we discussed the pythagorean equation $x^2+y^2 = z^2$ we saw a correspondence between the integer solutions with $z \neq 0$ and $\gcd(x, y, z) = 1$ and the rational points on $u^2 + v^2 = 1$. Let us look at this a little more closely.

To deal with the uninteresting fact that we get infinitely many solutions by scaling a given solution of $x^2 + y^2 = z^2$ through by a constant, we usually impose a condition like $\gcd(x, y, z) = 1$, and stick with integer solutions or, when $z \neq 0$, divide out by z and get a rational solution. The first is arguably unsatisfactory since we select one of an infinite class of solutions somewhat arbitrarily; moreover we haven't really decided between (x, y, z) and $(-x, -y, -z)$, and when we ask the same question say in $\mathbb{Z}[\sqrt{5}]$ then there will infinitely many such equivalent solutions (i.e. we can multiply through by $(2+\sqrt{5})^k$ for any k). One can overcome these issues by treating solutions as the same if the ratios $x : y : z$ are the same. This equivalence class of solutions is called a *projective* solution to the Diophantine equation. This is only possible if the different monomials in the equation all have the same degree.

This almost the same thing as dividing out by the z-value. This reduces the number of variables in the equation by 1, and the different monomials do not all have the same degree. The solutions here are *affine* solutions. Often it is more convenient to work with rational solutions to an affine equation, but it does have the disadvantage that we have "lost" the solutions where $z = 0$. One way to deal with this is to ask oneself what was so special about z? Why not divide through by y and get a different affine equation, and recover all the solutions except those with $y = 0$? Or do the same with x. It seems like a bit of overkill for what turns out to be just one or two solutions, but this discussion does

make the point that there can be several affine models for a given projective equation. What is typically done is to work with one affine model, say our first, and treat the lost solutions separately, calling them *the points at infinity* as if we divided through by $z = 0$. It is good to keep track of them since then all affine models of the same equation, have the same solutions!

Affine equations in two variables are curves, and so projective equations in three variables are also known as curves.

There are, moreover, other ways to transform equations. We saw in section A that it is much easier to reduce the number of monomials in a problem by suitable linear transformations of the variables. In this case the rational solutions are be mapped 1-to-1, though the integer solutions are not necessarily. One special case is the projective equation $x^2 - dy^2 = z^2$. If we divide through by z we get the Pell equation $u^2 - dv^2 = 1$ and we will see that this has $p - (d/p)$ solutions mod p. When $z = 0$ we are asking for solutions to $x^2 = dy^2$, and this evidently has $1 + (d/p)$ solutions mod p. Hence the total number of solutions, counting those at infinity, is $p + 1$. More on points at infinity later.

The transformations in section A kept the number of variables the same, which is different from the above transformations. For example solutions to $z^2 = x^2 + 2xy + 2y^2$ and in 1-1 correspondence with solutions to $z^2 = v^2 + y^2$ taking $v = x + y$. However other linear transformations can be a little confusing. For example if we are looking at solutions to $y^4 = x^4 + x$ we might take the change of variables $y = v/u$ and $x = 1/u$ to obtain the equation $v^4 = u^3 + 1$. At first sight appears to be of lower degree than the original equation, which implies that to understand rational points we probably need more subtle invariants than degree. The *genus* of the curve handles this for us, though it's definition involves more algebraic geometry than we want to discuss in this book. For now it suffices to note that linear and quadratic equations have genus zero and, if solvable, will have parameterized families of solutions, like the Pythagorean equation. Equations of degree 3 like $y^2 = x^3 + ax + b$ and $x^3 + y^3 = 1$ have genus one, and if solvable non-trivially, typically have infinitely many solutions which can be determined from the first – we will discuss this in some detail in chapter *. Higher degree curves usually have genus > 1 and, as we shall see, typically have only finitely many rational solutions, though this is a very deep result.

C12. Descent and the quadratics.. A famous problem asks to prove that if a and b are positive integers for which $ab+1$ divides a^2+b^2 is an integer then prove that the quotient is a square. One can approach this as follows: Suppose that $a \geq b \geq 1$ and $a^2+b^2 = k(ab+1)$ for some positive integer k. Then a is the root of the quadratic $x^2 - kbx + (b^2 - k)$. If c is the other root then $a + c = kb$ so that c is also an integer for which $b^2 + c^2 = k(bc + 1)$. We shall now prove that this is a "descent": If $c = 0$ then $k = b^2$ and $a = b^3$ and we have a solution. Otherwise $c \geq 1$ else $bc + 1 \leq 0$ and so $b^2 + c^2 \leq 0$ which is impossible. But then $b^2 - k = ac > 0$ and so $c = (b^2 - k)/a < b^2/b = b$. Hence (b, c) gives a smaller pair of solutions than (a, b). We deduce that all solutions can be obtained by iterating the map

$$(b, c) \to (kb - c, b)$$

starting from initial solutions $(d, 0)$ with $k = d^2$.

Other quadratics have a similar property. Perhaps the most famous is the Markov

equation: Find positive integers x, y, z for which

$$x^2 + y^2 + z^2 = 3xyz.$$

One finds many solutions: $(1,1,1), (1,1,2), (1,2,5), (1,5,13), (2,5,29), (1,13,34), (1,34,89), (2,29,169), (5,13,194), (1,89,233), (5,29,433), (89,233,610)$. Given one solution (x,y,z) one has that x is a root of a quadratic, the other root being $3yz - x$, and so we obtain a new solution $(3yz - x, y, z)$. (And one can do the same procedure with y or z. If we fix one co-ordinate we see that if there is one solution there are infinitely many. For example, taking $z = 1$ yields the equation $x^2 + y^2 = 3xy - 1$.

`Exercise C12.1` Determine what solutions are obtained from (1,1,1) by using the maps $(x,y) \to (3y-x, y)$ and $(x,y) \to (x, 3x-y)$.

One open question is to determine all of the integers that appear in a Markov triple. The first few are $1, 2, 5, 13, 29, 34, 89, 169, 194, 233, 433, 610, 985, 1325, \ldots$; it is believed that they are quite sparse.

Arguably the most beautiful such problem is the Apollonian circle packing problem.[13] Take three circles that touch each other (for example, take three coins and push them together). In between the circles one has a crescent type shape (*a hyperbolic triangle*), and one can inscribe a (unique) circle that touches all three of the original circles. What s the relationship between the radius of the new circle and the radii of the original circles? If we define the *curvature* of the circles to be $1/r$ (where r is the radius) then Descartes observed, in 1643, that the four curvatures satisfy the equation

$$2(a^2+b^2+c^2+d^2) = (a+b+c+d)^2, \text{ that is } a^2+b^2+c^2+d^2-2(ab+bc+cd+da+ac+bd) = 0.$$

We see that given b, c, d there are two possibilities for a, since this is a quadratic equation, the other is the circle that contains the three original circles and touches them all. We scale up the first three curvatures so that they are integers (with $\gcd(b,c,d) = 1$). We will focus on the case that a is also an integer, for example if we start with $b = c = 2$ and $d = 3$ we have $a^2 - 14a - 15 = 0$ so that $a = -1$ or $a = 15$. Evidently $a = -1$ corresponds to the outer circle (the negative sign comes from the fact that the circle contains the original circles), and $a = 15$ the inner one. In general if we have a solution (a,b,c,d) then we also have a solution (A,b,c,d) with $A = 2(b+c+d) - a$. Yet again we can iterate this (perhaps using the variables b, c or d) and obtain infinitely many Apollonian circles. But there is another interpretation of this, since each time we have a crescent in-between three existing circles we fill part of it in with a new circle, and we are eventually *tiling* the while of the original circle (see the enclosed pictures). There are many questions that can be asked: What integers appear as curvatures in a given packing? There are some integers that cannot appear because of congruence restrictions. For example if a, b, c, d are all odd, then all integers that arise as curvatures in this packing will be odd. The conjecture is that all sufficiently large integers that satisfy these trivial congruence constraints (mod 24) will appear as curvatures in the given packing. Although this is an open question, we do know

[13] Apollonius lived in Perga, 262-190 BC.

that a positive proportion of integers appear in any such packing, that the total number of circles in packing with curvature $\leq x$ is $\sim cT^\alpha$ where $|alpha = 1.30568\ldots$, and that the Apollonian twin prime conjecture holds: that there are infinitely many pairs of touching circles in the packing whose curvatures are both primes.

This last question is accessible because we see that any given solution (a,b,c,d) is mapped to another solution by any permutation of the four elements, as well as the matrix $\begin{pmatrix} -1 & 2 & 2 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. These (linear) transformations generate a subgroup of $\mathrm{SL}(4,\mathbb{Z})$, and one can proceed by considering orbits under the actions of this subgroup.

D. Algebra and calculation

D1. Primitive roots, indices and orders. Given one element of order m mod p it is not difficult to find them all:

`Exercise D1.1.` Show that if a has order m mod p then $\{a^k \pmod p : 1 \le k \le m,\ (a,m)=1\}$ is the set of residues mod p of order m. Use this to describe the set of primitive roots mod p, given one primitive root.

There are several parameters that go into the definition of index. The one that appears at first sight like it should be of some concern is the choice of primitive root to use as a base. The next result shows that there is little difference between the choice of one basis and another.

`Exercise D1.2.` Suppose that g and h are two primitive roots mod p, where $h \equiv g^\ell \pmod p$. Show that $(\ell, p-1)=1$. Show that the index with respect to g is ℓ times the index with respect to h, mod p.

We have described residues in terms of index and in terms of order. What is the link between the two?

Proposition D1.1. *For any reduced residue a mod p we have*

$$\mathrm{ord}_p(a) \cdot (p-1, \mathrm{ind}_p(a)) = p-1.$$

Proof. Let g be a primitive root with $k = \mathrm{ind}_p(a)$ and let $m = \mathrm{ord}_p(a)$. This means that m is the smallest integer for which $g^{km} \equiv a^m \equiv 1 \pmod p$; that is the smallest integer for which $p-1$ divides km, by Lemma 7.2. The result then follows from Corollary 3.2.

`Exercise D1.3.` Suppose that m divides $p-1$. Show that a is mth power mod p if and only if m divides $\mathrm{ind}_p(a)$.

> *The problem of finding primitive roots is one of the deepest mysteries of numbers*
> — from *Opuscula Analytica* 1, 152 by L. Euler

There are $\phi(p-1)$ primitive roots $\pmod p$, and so the proportion of reduced residues that are primitive roots, $\frac{\phi(p-1)}{p-1}$, is rarely small. Therefore if we select several random elements mod p we should quickly be lucky and find a primitive root. However Gauss described a search method that is more efficient than this, stemming from a different description of the primitive roots.

Proposition D1.3. *Suppose that $p-1 = \prod_q q^b$. The set of primitive roots mod p is precisely the set*

$$\left\{ \prod_{q \mid p-1} A_q : \ A_q \text{ has order } q^b \pmod p \right\}$$

To prove this we need the following:

Lemma D1.4. (Legendre) *If* $\mathrm{ord}_m(a) = k$ *and* $\mathrm{ord}_m(b) = \ell$ *where* $(k,\ell) = 1$ *then* $\mathrm{ord}_m(ab) = k\ell$.

Proof. Since $(ab)^{k\ell} = (a^k)^\ell (b^\ell)^k \equiv 1^\ell 1^k \equiv 1 \pmod m$, we see that $\mathrm{ord}_m(ab) | k\ell$, so we may write $\mathrm{ord}_m(ab) = k_1\ell_1$ where $k_1|k$ and $\ell_1|\ell$ (by exercise 4.2.2). Now

$$a^{k_1\ell} \equiv a^{k_1\ell}(b^\ell)^{k_1} = ((ab)^{k_1\ell_1})^{\ell/\ell_1} \equiv 1 \pmod m,$$

so that $k|k_1\ell$ by Lemma 7.2. As (k,ℓ) we deduce that $k|k_1$ and so $k_1 = k$. Analogously we have $\ell_1 = \ell$ and the result follows.

Proof of Proposition D1.3. We see from Lemma D1.4, and by induction on the number of prime factors of $p-1$, that each $\prod_{q|p-1} A_q$ is a primitive root mod p. These are all distinct for if $\prod_{q|p-1} A_q \equiv \prod_{q|p-1} B_q \pmod p$ then, raising this to the power ℓ where $\ell \equiv 0 \pmod{(p-1)/q^b}$ and $\equiv 1 \pmod{q^b}$, we see that each $A_q \equiv (\prod_{q|p-1} A_q)^\ell \equiv (\prod_{q|p-1} B_q)^\ell \equiv B_q \pmod p$. Finally, by Theorem 7.7 we know that there are $\phi(q^b)$ such A_q, and therefore a total of $\prod_{q^b\|p-1} \phi(q^b) = \phi(p-1)$ such products, that is they give all of the $\phi(p-1)$ primitive roots.

Proposition D1.3 provides a satisfactory way to construct primitive roots provided we can find the A_q of order q^b.

Lemma D1.5. *Suppose that* $a^{(p-1)/q} \not\equiv 1 \pmod p$. *If* q^b *divides* $p-1$ *then* $A_q :\equiv a^{(p-1)/q^b} \pmod p$ *has order* q^b *mod* p.

Proof. Now $A_q^{q^b} \equiv a^{p-1} \equiv 1 \pmod p$ and $A_q^{q^{b-1}} \equiv a^{(p-1)/q} \not\equiv 1 \pmod p$. Therefore $\mathrm{ord}_p(A)$ divides q^b and not q^{b-1}, so the result follows.

Gauss's algorithm to find primitive roots goes as follows: For each prime power $q^b\|p-1$,

(1) Find an integer a_q for which $a_q^{(p-1)/q} \not\equiv 1 \pmod p$.
(2) Let $A_q \equiv a_q^{(p-1)/q^b} \pmod p$.

Then $\prod_{q|p-1} A_q$ is a primitive root $\pmod p$.

How do we find appropriate a_q? Actually the proportion of a that fail to be appropriate is $1/q$; that is most a are appropriate. We can try to find a_q by trying $2, 3, 5, 7, \ldots$ until one finds an appropriate number, but there are no guarantees that this will succeed in a reasonable time period. However if we select values of a at random then the probability that we fail to find an appropriate a_q after k tries is $1/q^k$, which is negligible for $k > 20$.

.

Finding nth roots mod p. In Proposition 8.2 we understood how many square roots a residue has mod p. Now we look at how many nth roots a residue has.

We now show that we can find all solutions to $x^n \equiv a \pmod p$ directly and easily from all solutions to $y^d \equiv a \pmod p$ where $d = (n, p-1)$:

Proposition D1.6. *Suppose that* $(a,p) = 1$ *and let* $d = (n, p-1)$.

(1) *There are solutions* $x \pmod p$ *of* $x^n \equiv a \pmod p$ *if and only if* $a^{(p-1)/d} \equiv 1 \pmod p$.

(2) *Given one solution x_0, the set of all solutions is given by $x_0u \pmod p$ as u runs through the d roots of $u^d \equiv 1 \pmod p$.*
(3) *Given $x \pmod p$ for which $x^n \equiv a \pmod p$ we can find $y \pmod p$ for which $y^d \equiv a \pmod p$; and vice-versa.*

Proof. If $k = \mathrm{ind}_p(a)$ then $x^n \equiv a \pmod p$ with $x = g^t$ if and only if $nt \equiv k \pmod{p-1}$. This has solutions if and only if $d = (n, p-1)|k$.

(1) Now $d|\mathrm{ind}_p(a)$ if and only if $a^{(p-1)/d} \equiv 1 \pmod p$ by Proposition D1.1.

(2) Writing $n = dm$ where $(m, \frac{p-1}{d}) = 1$ we have $mt \equiv k/d \pmod{\frac{p-1}{d}}$, in which case all solutions are given by $t \equiv \ell \cdot k/d \pmod{\frac{p-1}{d}}$, where $\ell m \equiv 1 \pmod{\frac{p-1}{d}}$. Hence the set of solutions takes the form x_0u for $x_0 = g^{\ell k/d}$ and any u for which $\frac{p-1}{d} |\mathrm{ind}_p(u)$, that is whenever u is a dth root of unity mod p (by exercise D1.3).

(3) If $x^n \equiv a \pmod p$ then $y \equiv x^m \pmod p$ satisfies $y^d \equiv a \pmod p$. In the other direction if $y^d \equiv a \pmod p$ and $x \equiv y^\ell \pmod p$ then $x^n \equiv y^{\ell md} \equiv y^d \equiv a \pmod p$.

Notice that the case $n = d = 2$ of Proposition D1.6(i) yields Euler's criterion.

By Proposition D1.6 we can restrict our attention to two problems: If n divides $p-1$ and if a is a nth power mod p then

(1) Find one solution to $x^n \equiv a \pmod p$;
(2) Find all $u \pmod p$ for which $u^n \equiv 1 \pmod p$.

`Exercise D1.4.` Solve the second problem using an efficient variant of Gauss's algorithm for finding primitive roots.

We now re-interpret what was done in exercise 8.4.4 where we saw how to solve the first question in a special case with $n = 2$:

(1) Suppose there exists an integer k such that $2k \equiv 1 \pmod{\frac{p-1}{2}}$ (which is possible if and only if $(\frac{p-1}{2}, 2) = 1$).
(2) Let $x \equiv a^k \pmod p$. We know that $a^{\frac{p-1}{2}} \equiv 1 \pmod p$ so that $x^2 \equiv a^{2k} \equiv a \pmod p$.

Imitating that construction we have:

Proposition D1.7. *Suppose that n divides $p-1$, with $(n, \frac{p-1}{n}) = 1$, and that a is a nth power mod p. If k is an integer for which $nk \equiv 1 \pmod{\frac{p-1}{n}}$ and $x \equiv a^k \pmod p$ then $x^n \equiv a \pmod p$.*

Proof. Since a is a nth power mod p we know that $a^{\frac{p-1}{n}} \equiv 1 \pmod p$, and therefore $x^n \equiv a^{nk} \equiv a \pmod p$.

Unfortunately it is not always true that $(n, \frac{p-1}{n}) = 1$, for example when $p \equiv 1 \pmod 4$ with $n = 2$. In that case we can still often find solutions. For example, if $p \equiv 5 \pmod 8$ then $(2, \frac{p-1}{4}) = 1$, and so if $2k \equiv 1 \pmod{\frac{p-1}{4}}$, that is $2k = 1 + m\,\frac{p-1}{4}$ for some integer m then $(a^k)^2 \equiv ab^m \pmod p$ where $b \equiv a^{\frac{p-1}{4}} \pmod p$. Now $b^2 \equiv a^{\frac{p-1}{2}} \equiv 1 \pmod p$ and so $b \equiv \pm 1 \pmod p$; that is $(a^k)^2 \equiv \pm a \pmod p$. Therefore either a^k or ia^k is a square root of $a \pmod p$, where $i^2 \equiv -1 \pmod p$. Now i is a fourth root of unity, and the fourth roots of unity are not difficult to find, as in exercise D1.4. We can generalize this:

Proposition D1.8. *Suppose that n divides $p-1$, and that a is a nth power mod p. Let N be the smallest positive integer such that $(n, \frac{p-1}{N}) = 1$ and $b \equiv a^{\frac{p-1}{N}} \pmod p$ so that b has order dividing N/n. Let k be an integer for which $nk \equiv 1 \pmod{\frac{p-1}{N}}$, so that one can determine an integer m for which $nk = 1 - m\,\frac{p-1}{N}$. If $r^n \equiv b \pmod p$ and $x \equiv r^m a^k \pmod p$ then $x^n \equiv a \pmod p$.*

`Exercise D1.5.` Show that N is the largest divisor of $p-1$ with exactly the same prime factors as n.

Proof. Since a is a nth power mod p we know that $b^{N/n} \equiv a^{\frac{p-1}{n}} \equiv 1 \pmod p$. Moreover $(r^m a^k)^n \equiv (r^n)^m \cdot a^{kn} \equiv b^m \cdot ab^{-m} \equiv a \pmod p$.

By Proposition D1.8, Gauss showed that in finding solutions to $x^n \equiv a \pmod p$ for arbitrary a, we can restrict our attention to those a of order dividing N/n. In fact if $r^n \equiv b \pmod p$ then $r^N \equiv (r^n)^{N/n} \equiv b^{N/n} \equiv 1 \pmod p$, so the value of r is an Nth root of unity. Therefore if N is not large then finding r is easily done by exercise D1.4.

`Exercise D1.6.` Given a value for x in the hypothesis of Proposition D1.8 given a formula for $r \pmod p$. Hence finding r and x are "equivalent".

Example: We want to solve $x^2 \equiv a \pmod{29}$ where $a^{14} \equiv 1 \pmod{29}$. Then $n = 2$ and $N = 4$, and we take $b \equiv a^7 \pmod{29}$, so that b has order dividing 2. We wish to solve $2k \equiv 1 \bmod 7$ giving $k = 4$, in fact $2 \cdot 4 = 1 + 7$ so that $m = -1$. Now if $r^2 \equiv b \equiv a^7 \pmod{29}$ then $(r^{-1}a^4)^2 \equiv r^{-2}a^8 \equiv a^{-7}a^8 = a \pmod{29}$. In the other direction if $x^2 \equiv a \pmod{29}$ then $(x^{-1}a^4)^2 \equiv x^{-2}a^8 \equiv a^{-1}a^8 = a^7 \equiv b \pmod{29}$.

Now if $a^7 \equiv 1 \pmod{29}$ then $r^2 \equiv 1 \pmod{29}$, that is $r \equiv \pm 1 \pmod{29}$ and so $x \equiv \pm a^4 \pmod{29}$.

If $a^7 \equiv -1 \pmod{29}$ then $r^2 \equiv -1 \pmod{29}$, that is $r \equiv \pm 12 \pmod{29}$ and so $x \equiv \pm 12a^4 \pmod{29}$.

Example: Solve $x^3 \equiv 31 \pmod{37}$. Here $p - 1 = 36,\ n = 3,\ N = 9$. We want $3k \equiv 1 \pmod 4$ giving $k = 3, m = -2$. Now $b \equiv 31^4 \equiv 1 \pmod{37}$, and so if $r^3 \equiv 1 \pmod{37}$ and $x \equiv r^{-2}31^3 \equiv 6r \pmod{37}$ then $x^3 \equiv 31 \pmod{37}$. Hence the three solutions are $6, 6r$ and $6r^2 \pmod{37}$. Since $37|111$ one can take $r = 10$ and hence our solutions are $6, 23$ and $8 \pmod{37}$.

`Exercise D1.7.` Determine the square roots of $3 \pmod{37}$ as above.

D2. Lifting solutions. Gauss discovered that if an equation has solutions mod p then one can often use those solutions to determine solutions to the same equation mod p^k. In Proposition 8.4 we saw how to do this for quadratic equations. We can directly generalize that proof to nth powers:

Proposition D2.1. *Suppose that p does not divide a and that $u^n \equiv a \pmod p$. If p does not divide n then, for each integer $k \geq 2$, there exists a unique congruence class $b \pmod{p^k}$ such that $b^n \equiv a \pmod{p^k}$ and $b \equiv u \pmod p$.*

Proof. We prove this by induction on $k \geq 2$. We may assume that there exists a unique congruence class $b \pmod{p^{k-1}}$ such that $b^n \equiv a \pmod{p^{k-1}}$ and $b \equiv u \pmod p$. Therefore if $B^n \equiv a \pmod{p^k}$ and $B \equiv u \pmod p$ then $B^n \equiv a \pmod{p^{k-1}}$ and so $B \equiv b \pmod{p^{k-1}}$. Writing $B = b + mp^{k-1}$ we have

$$B^n = (b + mp^{k-1})^n \equiv b^n + nmp^{k-1}b^{n-1} \pmod{p^k}$$

which is $\equiv a \pmod{p^k}$ if and only if

$$m \equiv \frac{a - b^n}{np^{k-1}b^{n-1}} \equiv \frac{u}{an} \cdot \frac{a - b^n}{p^{k-1}} \pmod p,$$

as $ub^{n-1} \equiv u^n \equiv a \pmod p$.

`Exercise D2.1.` Show that if prime $p \not| an$ then the number of solutions $x \pmod{p^k}$ to $x^n \equiv a \pmod{p^k}$ does not depend on k.

Starting with the root $b_1 = u \pmod p$ to $x^n \equiv a \pmod p$, Proposition D2.1 gives us a root b_k to $x^n \equiv a \pmod{p^k}$, where $b_i \equiv b_j \pmod{p^k}$ for all $i, j \geq k$. We can define a p-adic norm of $p^k r$ where $p \not| r$ as $|r|_p := p^{-k}$. With this norm we have that $|b_i - b_j|_p \leq p^{-k}$ whenever $i, j \geq k$, so that $\lim_{k\to\infty} b_k$ exists if we complete the space. The completion is called the *p-adic integers* and can be written in the form

$$a_0 + a_1 p + a_2 p^2 + \dots \text{ with each } \ 0 \leq a_i \leq p - 1.$$

Thus Proposition D2.1 implies that the roots of $x^n = a$ in the p-adics are in 1-to-1 correspondence with the solutions to $x^n \equiv a \pmod p$.

`Exercise D2.2.` If prime $p \not| a$, show that the sequence $a_n = a^{p^n}$ converges in the p-adics. Show that $\alpha := \lim_{n\to\infty} a_n$ is a $(p-1)$st root of unity, and that all solutions to $x^{p-1} - 1$ in $\mathbb{Q}_p$ can be obtained in this way. Conclude that $i := \lim_{n\to\infty} 2^{5^n}$ is a square root of -1 in $\mathbb{Q}_5$.

We can find p-adic roots of most equations.

Theorem D2.2. *Suppose that $f(x) \in \mathbb{Z}[x]$ and that p is an odd prime. If a is an integer for which $f(a) \equiv 0 \pmod p$ and $f'(a) \not\equiv 0 \pmod p$ then there is a unique p-adic root α to $f(\alpha) = 0$ with $\alpha \equiv a \pmod p$. On the other hand if α is a p-adic root of $f(\alpha) = 0$ with $|f'(\alpha)|_p = 1$ then $f(a) \equiv 0 \pmod p$ where $a \equiv \alpha \pmod p$.*

This follows immediately from the following result:

Proposition D2.3. *Suppose that $f(x) \in \mathbb{Z}[x]$ and that p is an odd prime. If $f(a) \equiv 0 \pmod p$ and $f'(a) \not\equiv 0 \pmod p$ then for each integer k there exists a unique residue class $a_k \pmod{p^k}$ with $a_k \equiv a \pmod p$ for which $f(a_k) \equiv 0 \pmod{p^k}$.*

Proof. The Taylor expansion of polynomial $f(x)$ at a is simply the expansion of f as a polynomial in $x - a$. In fact

$$f(x) = f(a) + f'(a)(x-a) + f^{(2)}(a)\,\frac{(x-a)^2}{2!} + \ldots + f^{(k)}(a)\,\frac{(x-a)^k}{k!}\,.$$

Now, proceeding by induction on $k \geq 2$ we see that if $f(A) \equiv 0 \pmod{p^{k-1}}$ we can write $A = a_{k-1} + rp^{k-1}$ for some integer r. Using the Taylor expansion we deduce that $0 \equiv f(A) \equiv f(a_{k-1} + rp^{k-1}) \equiv f(a_{k-1}) + f'(a_{k-1})rp^{k-1} \pmod{p^k}$, as p is odd. Hence r is uniquely determined to be $\equiv -f(a_{k-1})/f'(a_{k-1})p^{k-1} \equiv -(f(a_{k-1})/p^{k-1})/f'(a) \pmod p$.

`Exercise D2.3.` Show that if f has no repeated roots then there are only finitely many primes p for which there exists an integer a_p with $f(a_p) \equiv f'(a_p) \equiv 0 \pmod p$.

The reduced residues modulo a power of 2. We are interested in the structure of $(\mathbb{Z}/2^k\mathbb{Z})^*$ for $k \geq 3$. Now, since $x^2 \equiv 1 \pmod 8$ for $x \equiv 1, 3, 5$ or $7 \pmod 8$, we see that the powers of x mod 2^k are all $\equiv 1$ or $x \pmod 8$ and so cannot include half of the possible residue classes. Moreover this implies that the largest possible order of an element mod 2^k is actually $\phi(2^k)/2 = 2^{k-2}$:

Proposition D2.4. *If $a \equiv \pm 3 \pmod 8$ then a has order 2^{k-2} mod 2^k whenever $k \geq 3$. Hence all of the residue classes mod 2^k can be written in the form $\pm 3^j \pmod{2^k}$.*

Proof. We prove by induction on $k \geq 3$ that $a^{2^{k-2}} \equiv 1 + 2^k \pmod{2^{k+1}}$, which implies the result. For $k = 3$ we have $a = 3 + 8b$ so that $a^2 = 9 + 48b + 64b^2 \equiv 1 + 2^3 \pmod{2^4}$; thus a has order 4. Assuming the result for k, and writing $a^{2^{k-2}} \equiv 1 + 2^k + b2^{k+1} \pmod{2^{k+2}}$, we have

$$\begin{aligned} a^{2^{k-1}} &\equiv (a^{2^{k-2}})^2 \equiv (1 + 2^k + b2^{k+1})^2 \equiv 1 + 2(2^k + b2^{k+1}) + 2^{2k}(1+2b)^2 \pmod{2^{k+2}} \\ &\equiv 1 + 2^{k+1} \pmod{2^{k+2}}. \end{aligned}$$

Hence we have proved $(\mathbb{Z}/2^k\mathbb{Z})^* \cong \mathbb{Z}/2^{k-2}\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$, as claimed in section B4. Describing square roots mod 2^k is trickier (Proposition D2.1 does not apply since 2 divides the exponent $n = 2$). Technically the issue becomes that if $a \equiv b \pmod{2^{k-1}}$ then $a^2 \equiv b^2 \pmod{2^k}$, for each $k \geq 2$.

Proposition D2.5. *Suppose that $a \equiv 1 \pmod 8$. For $b = 1$ or 3 and $k \geq 3$ there is a unique residue class $b_k \pmod{2^{k-1}}$ with $b_k \equiv b \pmod 4$ for which $b_k^2 \equiv a \pmod{2^k}$.*

Proof. By induction on k. This is trivially true for $k = 3$. Now suppose it is true for $k - 1$ and that $B^2 \equiv a \pmod{2^k}$ with $B \equiv b \pmod 4$. By the induction hypothesis $B \equiv b_{k-1} \pmod{2^{k-2}}$ so we can write $B = b_{k-1} + r2^{k-2}$ for some integer r. Hence $a \equiv B^2 = (b_{k-1} + r2^{k-2})^2 \equiv b_{k-1}^2 + rb_{k-1}2^{k-1} + r^2 2^{2k-4}) \equiv b_{k-1}^2 + rb_{k-1}2^{k-1} \pmod{2^k}$ (as $k \geq 4$); and hence r is uniquely determined mod 2, that is $r \equiv (a - b_{k-1}^2)/b_{k-1}2^{k-1} \pmod 2$, and therefore b_k is uniquely determined $\pmod{2^{k-1}}$.

D3. Square Roots of 1. Lemma 4.1 implies that there are *at least* four distinct square roots of 1 (mod n), for any odd n which is divisible by two distinct primes. Thus we might try to prove n is composite by finding a square root of 1 (mod n) which is neither 1 nor -1; though the question becomes, how do we efficiently search for a square root of 1?

Our trick is to again use Fermat's Little Theorem, since if p is prime > 2, then $p-1$ is even, and so a^{p-1} is a square. Hence $(a^{\frac{p-1}{2}})^2 = a^{p-1} \equiv 1 \pmod p$, so $a^{\frac{p-1}{2}} \pmod p$ is a square root of 1 (mod p) and must be 1 or -1. Therefore if $a^{\frac{n-1}{2}} \pmod n$ is neither 1 nor -1 then n is composite. Let's try an example: We have $64^{948} \equiv 1 \pmod{949}$, and the square root $64^{474} \equiv 1 \pmod{949}$. Hmmmm, we failed to prove 949 is composite like this but, wait a moment, since 474 is even so we can take the square root again, and a calculation reveals that $64^{237} \equiv 220 \pmod{949}$, so that 949 is composite since $220^2 \equiv 1 \pmod{949}$. More generally, using this trick of repeatedly taking square roots (as often as 2 divides $n-1$), we call integer a a *witness* to n being composite if the finite sequence

$$a^{n-1} \pmod n,\ a^{(n-1)/2} \pmod n, \ldots,\ a^{(n-1)/2^k} \pmod n$$

(where $n-1 = 2^k m$ with m odd) is not equal to either $1, 1, \ldots, 1$ or $1, 1, \ldots, 1, -1, *, \ldots, *$ (which are the only two possibilities were n a prime). One can compute high powers modulo n very rapidly using "fast exponentiation", a technique we discussed in section A5.

It is easy to show that at least one-half of the integers a, $1 \le a \le n$ are witnesses for n, for each odd composite n. So can we find a witness "quickly" if n is composite?

• The most obvious idea is to try $a = 2, 3, 4, \ldots$ consecutively until we find a witness. It is believed that there is a witness $\le 2(\log n)^2$, but we cannot prove this (though we can deduce this from a famous conjecture, the Generalized Riemann Hypothesis[14]).

• Pick integers $a_1, a_2, \ldots, a_\ell, \ldots$ from $\{1, 2, 3, \ldots, n-1\}$ at random until we find a witness. By what we wrote above, if n is composite then the probability that none of $a_1, a_2, \ldots, a_\ell$ are witnesses for n is $\le 1/2^\ell$. Thus with a hundred or so such tests we get a probability that is so small that it is inconceivable that it could occur in practice; so we believe that any integer n for which none of a hundred randomly chosen a's is a witness, is prime. We call such n "*industrial strength primes*".

In practice the witness test allows us to accomplish Gauss's dream of quickly distinguishing between primes and composites, for either we will quickly get a witness to n being composite or, if not, we can be almost certain that our industrial strength prime is indeed prime. Although this solves the problem in practice, we cannot be absolutely certain that we have distinguished correctly when we claim that n is prime since we have no proof, and mathematicians like proof. Indeed if you claim that industrial strength primes are prime, without proof, then a cynic might not believe that your randomly chosen a are so random, or that you are unlucky, or ... No, what we need is a proof that a number is prime when we think that it is.

[14]We discuss the Riemann Hypothesis, and its generalizations, in section E3. Suffice to say that this is one of the most famous and difficult open problems of mathematics, so much so that the Clay Mathematics Insitute has now offered one million dollars for its resolution (see `http://www.claymath.org/millennium/`).

Random polynomial time algorithms: We just saw that if n is composite then there is a probability of at least $1/2$ that a random integer a is a witness for the compositeness of n, and if so then it provides a short certificate verifying that n is composite. Such a test is called a *random polynomial time* test for compositeness (denoted `RP`). As noted if n is composite then the randomized witness test is almost certain to provide a short proof of that fact in 100 runs of the test. If 100 runs of the test do not produce a witness then we can be almost certain that n is prime, but we cannot be *absolutely* certain since no proof is provided.

On the difficulty of finding non-squares (mod p): For a given odd prime p it is easy to find a square mod p: take 1 or 4 or 9, or indeed any $a^2 \pmod p$. Exactly $(p-1)/2$ of the non-zero values mod p are squares mod p, and so exactly $(p-1)/2$ are not squares mod p. One might guess that they would also be easy to find, but we do not know a surefire way to quickly find such a value for each prime p (though we do know a quick way to identify a non-square once we have one).

Much as in the search for witnesses discussed in section 2.4, the most obvious idea is to try $a = 2, 3, 4, \dots$ consecutively until we find a non-square. It is believed that there is a non-square $\le 2(\log p)^2$, but we cannot prove this (though we can also deduce this from the Generalized Riemann Hypothesis).

Another way to proceed is to pick integers $a_1, a_2, \dots, a_k, \dots$ from $\{1, 2, 3, \dots, n-1\}$ at random until we find a non-square. The probability that none of $a_1, a_2, \dots, a_k$ are non-squares mod p is $\le 1/2^k$, so with a hundred or so such choices it is inconceivable that we could fail!

D4. Primality testing and Carmichael numbers. Described enough in sections 7.6, 10.1, 10.4, 10.5.

D5. Quadratic sieve and beyond. In section 10.6 we outlined the key ideas in the quadratic sieve type algorithms. The key question that remains is *how*, explicitly, one selects $b_1, b_2, \dots$ so that if a_i is the least positive residue of $b_i^2 \pmod n$ then there is a good chance that all of the prime factors of a_i are $\le y$ (for a certain pre-chosen value of y). Here are a few methods:

Random squares: Pick the b_i at random in $[1, n]$ so we would guess that the probability that a_i is y-smooth,[15] is roughly the same as for a random integer $\le x$.

The Continued Fractions method: In section C4 we saw that if p/q is a convergent to $\sqrt{n}$ then $|p^2 - nq^2| < 2\sqrt{n} + 1$. Hence above we can take $b_i = p_i$ so that $|a_i| < 2\sqrt{n} + 1$. We discussed earlier that for most n the continued fraction for $\sqrt{n}$ has period length about $\sqrt{n}$, so this algorithm gives us many values of a_i, in fact far more than we will typically need. The sizes of p_i and q_i grow exponentially with i which is not good for computations, but since we only need $p_i \bmod n$ we can work mod n when computing the p_i; that is we simply compute $p_{i+1} \equiv r_{i+1}p_i + p_{i-1} \pmod n$ and, similarly for $q_i \pmod n$, where $\sqrt{d} = [r_0, r_1, \dots]$. We can determine the r_i as in section C4, so that the numbers involved in the calculation are all $\le n$.

[15] That is, all of its prime factors are $\le y$.

Polynomial values: Let $m = [\sqrt{n}]$, and then let $b_i = m + i$ so that $a_i = (m+i)^2 - n$. Now $a_i = i^2 + 2im + (m^2 - n) \leq 2im + i^2 \leq (2i+1)\sqrt{n}$, provided $i < n^{1/4}$ (which it is in practice), so that the a_i are not much bigger than $\sqrt{n}$. The probability of a random number up to $n^{1/2+\epsilon}$ being y-smooth is significantly higher than for a random number up to n. Another issue, that we had not mentioned before, is how to determine whether the a_i are y-smooth. In the random squares method one simply has to test divide to see whether the a_i are y-smooth.[16] Here we have a better idea which a_j are divisible by p:

`Exercise D5.1.` Suppose that i_p is the smallest positive integer i for which $p|a_i$. Prove that $p|a_j$ if and only if $j \equiv i_p$ or $i'_p := 2m - i_p \pmod{p}$.

Hence, for each prime $p \leq y$ we determine the smallest such i_p and then we know precisely all of the j for which $p|a_j$ without test division; simply look at every pth value starting with i_p and i'_p. Carefully storing such data leads to an efficient algorithm to determine which a_j are y-smooth, and this is why the method is known as the *quadratic sieve*.

Large prime variation: By the end of the quadratic sieve process one has divided out the y-smooth part of a_i to be left with an integer r_i. If $r_i = 1$ then a_i is y-smooth. It has proved to be useful to retain r_i if it is itself a prime not too much larger than y, for:

`Exercise D5.2.` Show that if $r_i = r_j$ is prime then $a_i a_j$ is a square times a y-smooth integer.

D6. Discrete Logs. See section 10.7

[16] Or come up with some other method, but one always has the disadvantage that you have no prior knowledge of the prime factors of the a_i.

E. The distribution of Primes

E1. Binomial Coefficients and bounds on the number of primes.
Upper bounds.

We use the first result in A4 to get an upper bound on the number of primes in an interval:

Lemma E1.1. *The product of the primes in* $[n+1, 2n]$ *is* $\le 4^{n-1}$ *for* $n \ge 2$.

Proof. Each prime in $[n+1, 2n]$ appears in the numerator of the binomial coefficient $\binom{2n-1}{n}$ but not the denominator. Hence their product divides $\binom{2n-1}{n}$ and so is $\le \binom{2n-1}{n} \le 2^{2n-2}$.

We can then use this result to get an upper bound on the number of primes up to a given point:

Proposition E1.2. *The product of the primes up to* N *is* $\le 4^N$ *for all* $N \ge 1$.

Proof. By induction on $N \ge 1$. The result is straightforward for $N = 1, 2$ by calculation. If $N = 2n$ or $2n-1$ then the product of the primes up to N is at most the product of the primes up to n, times the product of the primes in $[n+1, 2n]$. The first product is $\le 4^{n-1}$ by the induction hypothesis, and the second $< 2^{2n-2}$ by Lemma E1.1. Combining these gives the bound $\le 4^{N-2}$.

If we take logarithms in the lemma we obtain

$$\sum_{\substack{p \text{ prime} \\ n<p\le 2n}} \log p \le (n-1)\log 4.$$

Each term of the left side is $> \log n$ and therefore

$$\#\{p \text{ prime}: \ n < p \le 2n\} \le \frac{n}{\log n} \cdot \log 4.$$

`Exercise E1.1.` Deduce that $\pi(x) \le (\log 4 + \epsilon)\ \frac{x}{\log x}$. (Hint: Consider only those primes p in $[\epsilon x, x]$ and give a lower bound on $\log p$. Then sum the contributions of such intervals $[\epsilon x, x], [\epsilon^2 x, \epsilon x], [\epsilon^3 x, \epsilon^2 x], \dots$)

Lower bounds. We know that all of the primes in $((n+1)/2, n]$ divide $\binom{n}{[n/2]}$ so if we somehow show that the contribution of the other primes is negligible, then we can obtain a lower bound on the number of primes in this interval. It is evident from the definition of binomial coefficients that only primes $\le n$ divide $\binom{n}{m}$; it was Kummer who showed an easy method to determine to which power, as we saw in section A4.

Combining the Corollary to Kummer's Theorem with the inequality above we deduce that if p^{e_p} is the largest power of p dividing $\binom{n}{[n/2]}$ then

$$\frac{2^n}{n} \le \binom{n}{[n/2]} = \prod_{\substack{p \text{ prime} \\ p\le n}} p^{e_p} \le n^{\#\{p \text{ prime: } p\le n\}},$$

so that

$$\#\{p \text{ prime}: \ p \le n\} \ge (\log 2)\ \frac{n}{\log n} - 1.$$

It is perhaps of more interest to show that there are primes near to a given n. This was conjectured in 1845 by Bertrand on the basis of calculations up to a million, proved in 1850 by Chebyshev; we follow Erdős's 1932 proof from when he was 19 years old:

BERTRAND'S POSTULATE. *For every integer $n \geq 1$, there is a prime number between n and $2n$.*

`Exercise E1.2.` Show that p does not divide $\binom{2n}{n}$, when $2n/3 < p \leq n$. (Either use Kummer's Theorem, or consider directly how often p divides the numerator and denominator of $\binom{2n}{n}$.)

Proof. Let p^{e_p} be the exact power of prime p dividing $\binom{2n}{n}$. We know that

- $e_p = 1$ if $n < p \leq 2n$ by Kummer's Theorem,
- $e_p = 0$ if $2n/3 < p \leq n$ by the last exercise,
- $e_p \leq 1$ if $\sqrt{2n} < p \leq 2n$ by the Corollary,
- $p^{e_p} \leq 2n$ if $p \leq 2n$ by the Corollary.

Combining these gives, and using Lemma *, we obtain

$$\frac{2^{2n}}{2n} \leq \binom{2n}{n} = \prod_{p \leq 2n} p^{e_p} \leq \prod_{n<p\leq 2n} p \prod_{p\leq 2n/3} p \prod_{p\leq\sqrt{2n}} 2n$$
$$\leq \left(\prod_{n<p\leq 2n} p\right) \times 4^{2n/3-1} \times (2n)^{(\sqrt{2n}+1)/2},$$

since the number of primes up to $\sqrt{2n}$ is no more than $(\sqrt{2n}+1)/2$ (as neither 1 nor any even integer > 2 is prime). Taking logarithms we deduce that

$$\sum_{\substack{p \text{ prime} \\ n<p\leq 2n}} \log p > \frac{\log 4}{3}\ n - \frac{\sqrt{2n}+3}{2}\ \log(2n).$$

This implies that

$$\sum_{\substack{p \text{ prime} \\ n<p\leq 2n}} \log p \geq \frac{1}{3}\ n \tag{E1.1}$$

for all $n \geq 2349$, which implies Bertrand's postulate in this range. It is a simple matter to write a computer program to check that (E1.1) holds for all n in the range $1 \leq n \leq 2348$. Therefore (E1.1) holds for all $n \geq 1$ which implies a strong form of Bertrand's postulate.

`Exercise E1.3.` Verify Bertrand's postulate for all n up to 20000 using *only* the primes $2, 3, 5, 7, 13, 23, 43, 83, 163, 317, 631, 1259, 2503, 5003, 9973, 10007$.

`Exercise E1.4.` Prove that there are infinitely many primes p with a 1 as the leftmost digit in their decimal expansion.

`Exercise E1.5.` Use Bertrand's postulate to show, by induction, that every integer $n > 6$ can be written as the sum of distinct primes. (Hint: Use induction to show that, for each $n \geq 6$, every integer in $[7, 2p_n+6]$ is the sum of distinct primes in $\{2, 3 \ldots, p_n\}$, where p_n is the nth smallest prime.)

Further remarks. Can one give an upper bound on gaps between primes? It could be that there exists a constant $c > 0$ such that

$$p_{n+1} - p_n < c\log^2 p_n$$

and, if not, perhaps something only slightly weaker. Cramer conjectured one could take any $c > 1$ for sufficiently large n, whereas recent work suggests that gaps can get a little larger (as big as $1.1\log^2 p_n$). Here are the record breaking gaps:

p_n	$p_{n+1} - p_n$	$(p_{n+1} - p_n)/\log^2 p_n$
113	14	.6264
1327	34	.6576
31397	72	.6715
370261	112	.6812
2010733	148	.7026
20831323	210	.7395
25056082087	456	.7953
2614941710599	652	.7975
19581334192423	766	.8178
218209405436543	906	.8311
1693182318746371	1132	.9206

TABLE 2. (Known) record-breaking gaps between primes.

Evidently the constant is slowly creeping upwards but will it ever reach 1? And will it go beyond? We don't know.

The best upper bound that has been proved, to date, on the gap between consecutive primes, is $p_{n+1} - p_n < p_n^{.535}$. There are no good ideas to improve the exponent to $\frac{1}{2}$ or less. Therefore we have no idea how to prove Lagrange's conjecture that there is always a prime between consecutive squares.

E2. Dynamical systems and primes. The prime divisors of a sequence of integers, all > 1, form an infinite sequence of distinct primes if the integers in the sequence are pairwise coprime. We will generalize the constructions from section 5.1. We begin by simplifying the description of Euler's sequence a_n and the Fermat numbers F_n.

$$a_{n+1} - 1 = a_1 a_2 \dots a_n = (a_1 a_2 \dots a_{n-1})a_n = (a_n - 1)a_n,$$

so that $a_{n+1} = f(a_n)$ where $f(t) := t^2 - t + 1$. We now prove that $a_n \equiv 1 \pmod{a_m}$ for all $n > m$: To start with

$$a_{m+1} = f(a_m) \equiv f(0) = 1 \pmod{a_m},$$

by Corollary 2.3, and then, by induction

$$a_{n+1} = f(a_n) \equiv f(1) = 1 \pmod{a_m}.$$

Hence $(a_n, a_m) = (1, a_m) = 1$. Similarly $F_{n+1} = g(F_n)$ where $g(t) := t^2 - 2t + 2$, and $F_n \equiv g(0)$ or $g(2) \equiv 2 \pmod{F_m}$ whenever $n > m$.

How do we generalize this proof? Let $f(t) \in \mathbb{Z}[t]$. Consider the sequence $a, f(a), f(f(a)), \dots$ (we write $f_1(a) = f(a)$ and then $f_{n+1}(a) = f(f_n(a))$). We call a a *periodic point* if $f_m(a) = a$ for some $m \geq 1$, and we call m the *period of* a if m is the smallest such integer.

`Exercise E2.1.` Show that if $f_m(a) = a$ then $f_{m+n}(a) = f_n(a)$ for all $n \geq 0$.

We call a *pre-periodic* if there exist $n > m \geq 1$ such that $f_m(a) = f_n(a)$, but a is not a *periodic point*. The key technical step in the proofs above was that 0 is a pre-periodic point for both $t^2 - t + 1$ and $t^2 - 2t + 2$:

Proposition E2.1. *Let $f(t) \in \mathbb{Z}[t]$ have degree > 1, positive leading coefficient, and $f(0) \neq 0$. Suppose that 0 is a pre-periodic point for f, and let ℓ be the least common multiple of the integers in the sequence $f_n(0), n \geq 1$. If $a_0 \in \mathbb{Z}$ with $a_{n+1} = f(a_n)$ for all $n \geq 0$, and $(a_n, \ell) = 1$ for all $n \geq 0$, then we obtain an infinite sequence of distinct primes by selecting one prime factor from each a_n.*

Proof. Let $w_0 = 0$ and $w_{n+1} = f(w_n)$ for all $n \geq 0$, so that $a_{m+1} = f(a_m) \equiv f(0) = w_1 \pmod{a_m}$ and, thereafter, $a_{m+j+1} = f_1(a_{m+j}) \equiv f(w_j) = w_{j+1} \pmod{a_m}$ by induction on $j \geq 1$. Therefore if $m < n$ then $(a_m, a_n) = (a_m, w_{n-m})$ which divides (a_m, ℓ), which equals 1 by the hypothesis. The rest of the proof follows as above.

To apply Proposition E2.1 we need to determine when 0 is a pre-periodic point:

Proposition E2.2. *Suppose that 0 is a pre-periodic point for $f(t) \in \mathbb{Z}[t]$. Then there exists $m = 1$ or 2 such that $f_{n+m}(0) = f_n(0)$ for all sufficiently large n.*

Proof. Let m be the smallest integer ≥ 1 such that $f_{n+m}(0) = f_n(0)$ for all sufficiently large n. Let $u_k = f_{n+k}(0)$, so that our period is $u_0, u_1, \dots, u_{m-1}$. Now $x - y$ divides $f(x) - f(y)$ for any integers x, y; in particular $u_{n+1} - u_n$ divides $f(u_{n+1}) - f(u_n) = u_{n+2} - u_{n+1}$. Therefore $u_1 - u_0$ divides $u_2 - u_1$, which divides $u_3 - u_2$, ..., which divides $u_m - u_{m-1}$, which divides $u_{m+1} - u_m = u_1 - u_0$. That is, we have a sequence of integers that all divide

one another and so must all be equal in absolute value. If they are all 0 then $m = 1$. If not then they cannot all be equal, say to $d \neq 0$, else $0 = (u_1 - u_0) + (u_2 - u_1) + (u_3 - u_2) + \cdots + (u_m - u_{mp-1}) = md$. Therefore two consecutive terms must have opposite signs, yet have the same absolute value, so that $u_{n+2} - u_{n+1} = -(u_{n+1} - u_n)$ and thus $u_{n+2} = u_n$. Now, applying f, $p - n$ times to both sides, we deduce that $u_2 = u_0$ and therefore $m = 2$.

This allows us to classify all such polynomials f:

Theorem E2.3. *Suppose that* 0 *is a pre-periodic point for* $f(t) \in \mathbb{Z}[t]$. *The basic possibilities are:*

a) The period has length 1, and either $f(t) = u$ *with* $0 \to u \to u \to \dots$, *or*
$f(t) = (2/u)t^2 - u$ *where* $u = 1$ *or 2, with* $0 \to -u \to u \to u \to \dots$; *or*

b) The period has length 2, and either $f(t) = 1 + ut - t^2$ *with* $0 \to 1 \to u \to 1 \to \dots$, *or*
$f(t) = 1 + t + t^2 - t^3$ *with* $0 \to 1 \to 2 \to -1 \to 2 \to \dots$.

All other examples arise by replacing $f(t)$ *by* $-f(-t)$, *or by adding a polynomial multiple of* $\prod_{i=1}^{k}(t - a_i)$ *where the* a_i *are the distinct integers in the orbit of* 0.

Proof by **Exercises**:

`Exercise E2.2.` Let $f(t) \in \mathbb{Z}[t]$, and assume that f has a period of length 1, say $f(u) = u$. Then
a) f must be of the form $f(t) = u + (t - u)g(t)$ for some $g(t) \in \mathbb{Z}[t]$.

b) If $f(v) = u$ with $v \neq u$ then $f(t) = u + (t - u)(t - v)g(t)$ for some $g(t) \in \mathbb{Z}[t]$.

c) If $f(w) = v$ then $v = f(w) = u + (w - u)(w - v)g(w)$ so that $(v - w)(w - u)$ divides $v - u$. Deduce that $v - w = w - u = \pm 1$ or ± 2, equals δ say and $g(t) = 2/\delta + (t - w)h(t)$ for some $h(t) \in \mathbb{Z}[t]$.

d) If $f(x) = w$ then $(x - u)(x - v)$ divides $(w - u)$, which is impossible.

`Exercise E2.3.` Assume that $f(t) \in \mathbb{Z}[t]$, and f has a period of length 2, say $f(u) = v$ and $f(v) = u$. Then
a) f must be of the form $f(t) = v + u - t + (t - u)(t - v)g(t)$ for some $g(t) \in \mathbb{Z}[t]$.

b) If $f(w) = v$ then $w - v = \pm 1$, so that $g(t) = w - v + (t - w)h(t)$ for some $h(t) \in \mathbb{Z}[t]$.

c) If $f(x) = w$ then $x - u = \pm 1$. If $x - u = w - v = \delta$ then $2 = (x - v)(w - v + (x - w)h(x))$; this implies that $x - v = \delta, 2\delta, -\delta$ or -2δ each of which can be ruled. If $x - u = -(w - v)$ then u, x, w, v are consecutive integers (in this order), and $h(t) = -1 + (t - x)j(t)$ for some $j(t) \in \mathbb{Z}[t]$.

d) Show that if $f(y) = x$ then $y - u$ divides $|x - v| = 2$, and $y - v$ divides $|x - u| = 1$, which is impossible.

Deduce the cases of the theorem by setting $x = 0$, then $w = 0$ and then $v = 0$.

This section was motivated by examples of the first case in (a), that is, $f(u) = u + t(t - u)$. An example in the second case of (a) is given by $f(t) = t^2 - 2$, so that $0 \to -2 \to 2 \to 2 \to \dots$ Let $a_0 = 4$ in Theorem E2.1, and note that 2 divides each x_n, $n \geq 1$ but never 4, so a minor modification of our argument above works to prove that there are infinitely many primes. This sequence also appears in a result of Lucas showing that the Mersenne number $2^n - 1$ is prime if and only if it divides a_{n-2}.

`Exercise E2.4.` Now suppose that u_0 (which is not necessarily an integer) has period p, so that it is a root of the polynomial $f_p(x) - x$. Prove that if f is monic then $\frac{u_j - u_i}{u_1 - u_0}$ is a unit for all $0 \leq i < j \leq p - 1$.

We have considered iterations of the map $n \to f(n)$ where $f(t) \in \mathbb{Z}[t]$. If one allows $f(t) \in \mathbb{Q}[t]$ then it is an open question as to the possible period lengths in the integers.

Even the simplest case, $f(x) = x^2 + c$, with $c \in \mathbb{Q}$, is not only open but leads to the magnificent world of dynamical systems (see []). It would certainly be interesting to know what primes divide the numerators when iterating, starting from a given integer.

E3. Euler's proof of the infinitude of primes and the Riemann zeta-function. In the seventeenth century Euler gave a different proof that there are infinitely many primes, one which would prove highly influential in what was to come later. Suppose again that the list of primes is $p_1 < p_2 < \cdots < p_k$. Euler observed that the fundamental theorem of arithmetic implies that there is a 1-to-1 correspondence between the sets $\{n \geq 1 : n \text{ is a positive integer}\}$ and $\{p_1^{a_1} p_2^{a_2} \dots p_k^{a_k} : \ a_1, a_2, \dots, a_k \geq 0\}$. Thus a sum involving the elements of the first set should equal the analogous sum involving the elements of the second set:

$$\sum_{\substack{n\geq 1\\ n \text{ a positive integer}}} \frac{1}{n^s} = \sum_{a_1,a_2,\dots,a_k\geq 0} \frac{1}{(p_1^{a_1}p_2^{a_2}\dots p_k^{a_k})^s}$$
$$= \left(\sum_{a_1\geq 0} \frac{1}{(p_1^{a_1})^s}\right)\left(\sum_{a_2\geq 0} \frac{1}{(p_2^{a_2})^s}\right)\cdots\left(\sum_{a_k\geq 0} \frac{1}{(p_k^{a_k})^s}\right)$$
$$= \prod_{j=1}^{k}\left(1-\frac{1}{{p_j}^s}\right)^{-1}.$$

The last equality holds because each sum in the second-to-last line is over a geometric progression. Euler then noted that if we take $s = 1$ then the right side equals some rational number (since each $p_j > 1$) whereas the left side equals ∞, a contradiction (and thus there cannot be finitely many primes). We prove that $\sum_{n\geq 1} 1/n$ diverges in exercise * below

What is wonderful about Euler's formula is that something like it holds without assumption, involving the infinity of primes; that is

$$\sum_{\substack{n\geq 1\\ n \text{ a positive integer}}} \frac{1}{n^s} = \prod_{p \text{ prime}} \left(1-\frac{1}{p^s}\right)^{-1}. \tag{E3.1}$$

One does need to be a little careful about convergence issues. It is safe to write down such a formula when both sides are "absolutely convergent", which takes place when $s > 1$; that is the sum of the absolute values of the terms converges. In fact they are absolutely convergent even if s is a complex number so long as $\mathrm{Re}(s) > 1$, for if $s = \sigma + it$ with $\sigma > 1$ then

$$\sum_{n\geq 1}\left|\frac{1}{n^s}\right| = \sum_{n\geq 1}\frac{1}{n^\sigma} \leq 1 + \int_1^\infty \frac{dt}{t^\sigma} = 1 + \frac{1}{\sigma-1} = \frac{\sigma}{\sigma-1}.$$

Here we have used that $1/n^\sigma < \int_{n-1}^{n} dt/t^\sigma$ since $1/t^\sigma$ is a decreasing function in t.

We have just seen that (E3.1) makes sense when s is to the right of the horizontal line in the complex plane going through the point 1. Like Euler, we want to be able to interpret what happens to (E3.1) when $s = 1$. To not fall afoul of convergence issues we need to take the limit of both sides as $s \to 1^+$, since (E3.1) holds for real values of $s > 1$.

But now we can simply note that $1/n^\sigma > \int_n^{n+1} dt/t^\sigma$ for each n and so

$$\zeta(\sigma) = \sum_{n\geq 1} \frac{1}{n^\sigma} \geq \int_1^\infty \frac{dt}{t^\sigma} = \frac{1}{\sigma - 1} \ .$$

This diverges as $\sigma \to 1^+$. We deduce that

$$\prod_{p \text{ prime}} \left(1 - \frac{1}{p}\right) = 0 \tag{E3.2}$$

which, upon taking logarithms, implies that

$$\sum_{p \text{ prime}} \frac{1}{p} = \infty. \tag{E3.3}$$

So how numerous are the primes? One way to get an idea is to determine the behaviour of the sum analogous to (E3.3) for other sequences of integers. For instance $\sum_{n\geq 1} \frac{1}{n^2}$ converges, so the primes are, in this sense, more numerous than the squares. We can do better than this from our observation, just above, that $\sum_{n\geq 1} \frac{1}{n^s} \approx \frac{1}{s-1}$ is convergent for any $s > 1$ (see exercise E3.2 below). In fact, since $\sum_{n\geq 1} \frac{1}{n(\log n)^2}$ converges, we see that the primes are in the same sense more numerous than the numbers $\{n(\log n)^2 : n \geq 1\}$, and hence there are infinitely many integers x for which there are more than $x/(\log x)^2$ primes $\leq x$.

There is another derivation of (E3.1) that is worth seeing. One begins with $\sum_{n\geq 1} \frac{1}{n^s}$, the sum of $1/n^s$ over all integers n. Now suppose that we wish to remove the even integers from this sum. Their contribution to this sum is

$$\sum_{\substack{n\geq 1 \\ n \text{ even}}} \frac{1}{n^s} = \sum_{m\geq 1} \frac{1}{(2m)^s} = \frac{1}{2^s} \sum_{m\geq 1} \frac{1}{m^s}$$

writing even n as $2m$, and hence

$$\sum_{\substack{n\geq 1 \\ (n,2)=1}} \frac{1}{n^s} = \sum_{n\geq 1} \frac{1}{n^s} - \sum_{\substack{n\geq 1 \\ n \text{ even}}} \frac{1}{n^s} = \left(1 - \frac{1}{2^s}\right) \sum_{n\geq 1} \frac{1}{n^s} \ .$$

If we wish to remove the multiples of 3 we can proceed similarly, to obtain

$$\sum_{\substack{n\geq 1 \\ (n,2\cdot 3)=1}} \frac{1}{n^s} = \left(1 - \frac{1}{2^s}\right)\left(1 - \frac{1}{3^s}\right) \sum_{n\geq 1} \frac{1}{n^s};$$

and for arbitrary y, letting $m = \prod_{p\leq y}$,

$$\sum_{\substack{n\geq 1 \\ (n,m)=1}} \frac{1}{n^s} = \prod_{p\leq y} \left(1 - \frac{1}{p^s}\right) \cdot \sum_{n\geq 1} \frac{1}{n^s} \ .$$

As $y \to \infty$, the left side becomes the sum over all integers $n \geq 1$ which do not have any prime factors: the only such integer is $n = 1$ so the left hand side becomes $1/1^s = 1$. Hence

$$\prod_{p \text{ prime}} \left(1 - \frac{1}{p^s}\right) \cdot \sum_{n \geq 1} \frac{1}{n^s} = 1$$

an alternative formulation of (E3.1). The advantage of this proof is that we see what happens when we "sieve" by various primes, that is remove the integers from our set that are divisible by the given prime.

`Exercise E3.1.` Show that if $\text{Re}(s) > 1$ then

$$\left(1 - \frac{1}{2^s}\right) \sum_{n \geq 1} \frac{1}{n^s} = \sum_{\substack{n \geq 1 \\ n \text{ odd}}} \frac{1}{n^s} - \sum_{\substack{n \geq 1 \\ n \text{ even}}} \frac{1}{n^s}.$$

`Exercise E3.2.` The box with corners at $(n, 0), (n+1, 0), (n, 1/n), (n+1, 1/n)$ has area $1/n$ and contains the area under the curve $y = 1/x$ between $x = n$ and $x = n + 1$. Therefore $\sum_{n \leq N} 1/n \geq \int_1^{N+1} \frac{1}{t}\, dt = \log(N+1)$. Deduce that the sum of the reciprocals of the positive integers diverges. Now draw the box of height $1/n$ and width 1 to the left of the line $x = n$, and obtain the upper bound $\sum_{n \leq N} 1/n \leq \log(N) + 1$.

`Exercise E3.3.` Given that $\sum_p 1/p$ diverges, deduce that there are arbitrarily large values of x for which $\#\{p \leq x : \ p \text{ prime}\} \geq \sqrt{x}$. Improve the $\sqrt{x}$ here as much as you can using these methods.

The sieve of Eratosthenes and estimates for the primes up to x. Fix $\epsilon > 0$. By (E3.2) we know that there exists y such that

$$\prod_{p \leq y} \left(1 - \frac{1}{p}\right) < \frac{\epsilon}{3}.$$

Let m be the product of the primes $\leq y$, and select $x > 3y/\epsilon$. If $k = [x/m]$, so that $km \leq x < (k+1)m < 2km$, then the number of primes up to x is no more than the number of primes up to $(k+1)m$, which is no more than the number of primes up to y plus the number of integers up to $(k+1)m$ which have all of their prime factors $> y$. Since there are no more than y primes up to y, and since the set of integers up to $(k+1)m$ is $\{jm + i : \ 1 \leq i \leq m, \ 0 \leq j \leq k\}$, we deduce that the number of primes up to x is

$$\leq y + \sum_{j=0}^{k} \sum_{\substack{1 \leq i \leq m \\ (jm+i, m) = 1}} 1 = y + (k+1)\phi(m)$$

$$< \frac{\epsilon}{3} x + 2km \prod_{p \leq y} \left(1 - \frac{1}{p}\right) < \frac{\epsilon}{3} x + \frac{2\epsilon}{3} x = \epsilon x.$$

In other words

$$\lim_{x \to \infty} \frac{1}{x} \ \#\{p \leq x : \ p \text{ prime}\} \to 0. \tag{E3.4}$$

There is more than one way to estimate $\sum_{n\le N}\log n = \log N!$. Since $\log n$ is an increasing function we have $\int_{n-1}^{n}\log t\ dt < \log n < \int_{n}^{n+1}\log t\ dt$ and so

$$\log 1+\int_1^N \log t\ dt < \sum_{n\le N}\log n < \int_1^N \log t\ dt + \log N$$

and therefore $0 < \sum_{n\le N}\log n - N(\log N-1)+1 < \log N$. We will write

$$\frac{1}{N}\sum_{n\le N}\log n = \log N - 1 + O\left(\frac{\log N}{N}\right),$$

the big Oh meaning that this is a term that is bounded by a constant multiple times $\frac{\log N}{N}$. On the other hand we can write $\log n = \sum_{p^k|n}\log p$ and so

$$\begin{aligned}\sum_{n\le N}\log n &= \sum_{n\le N}\sum_{p^k|n}\log p = \sum_{p^k\le N}\log p\sum_{\substack{n\le N\\ p^k|n}} 1 = \sum_{p^k\le N}\log p\left[\frac{N}{p^k}\right]\\ &= \sum_{p^k\le N}\log p\left(\frac{N}{p^k}+O(1)\right) = N\sum_{p^k\le N}\frac{\log p}{p^k}+O\left(N\right).\end{aligned}$$

Now noting that $\sum_{p^k,\ k\ge 2}\frac{\log p}{p^k} = \sum_p \frac{\log p}{p(p-1)} \le \sum_{n\ge 2}\frac{\log n}{n(n-1)} = O(1)$, we deduce from all of the above that

$$\sum_{p\le N}\frac{\log p}{p} = \log N + O(1). \tag{E3.5}$$

Now if $\pi(x)\sim Lx/\log x$ then

$$\begin{aligned}\sum_{p\le x}\frac{\log p}{p} &= \frac{\log x}{x}\,\pi(x)+\int_1^x \frac{\log t-1}{t^2}\,\pi(t)dt\\ &\sim L + L\int_1^x \frac{\log t-1}{t^2}\frac{t}{\log t}\,dt \sim L\log x.\end{aligned}$$

Comparing this to (E3.5), we deduce that $L=1$.

`Exercise E3.4.` Explain why this does not prove the prime number theorem.

Our next goal is to prove strong versions of (E3.2) and (E3.3).

`Exercise E3.5.` Verify the identity

$$\sum_{p\le x}\frac{1}{p} = \frac{1}{\log x}\sum_{p\le x}\frac{\log p}{p}+\int_2^x\sum_{p\le t}\frac{\log p}{p}\ \frac{dt}{t(\log t)^2}\ .$$

Then substitute in the estimate in (E3.5) to deduce

$$\sum_{p\le x}\frac{1}{p} = \log\log x + O(1). \tag{E3.6}$$

One deduce from (E3.6) that $\prod_{p\le x}(1-\frac{1}{p})$ lies between two constants times $1/\log x$. To be more precise one needs a more accurate estimate in (E3.6). If one does this, one eventually proves *Merten's Theorem*:

$$\prod_{p\le x}(1-\frac{1}{p}) \sim \frac{e^{-\gamma}}{\log x}, \tag{E3.7}$$

where γ is the Euler-Mascheroni constant.[17]

Frequency of p-divisibility of Fermat quotients and class numbers. We have that p divides $2^{p-1}-1$ for every prime p. Does p^2 ever divide $2^{p-1}-1$? The only two known examples are 1093 and 3511, even though searches have gone on as far as $6.7\cdot 10^{15}$. Let $q_p(2) := (2^{p-1}-1)/p$ so that p^2 divides $2^{p-1}-1$ if and only if $q_p(2)\equiv 0 \pmod p$. We do not know much about the value of $q_p(2) \pmod p$; our best guess is that it looks kind of randomly distributed, whatever that means. So if we guess that the "probability" that $q_p(2)\equiv 0 \pmod p$ is roughly $1/p$, then the expected number of primes up to x for which p^2 divides $2^{p-1}-1$ is roughly

$$\sum_{p\le x}\frac{1}{p} = \log\log x + c,$$

by (E3.6). Now $\log\log(6.7\cdot 10^{15}) \approx 3.5$ so is having just two found so far reasonable? To expect a further example we will need to go beyond 10^{43} so it seems unlikely that we will ever compute another example, even if they exist as frequently as expected!

How about p^2 divides $a^{p-1}-1$? For how many a ?

Similar remarks can be made about Bernoulli numbers. For want of better information the "probability" that p divides the numerator of B_{2n} can be taken to be $1/p$. We will see that the case $2n = p-3$ is particularly interesting in section H4.

One can also ask whether p divides the numerator of B_{2n} for any n such that $2\le 2n\le p-3$. If these probabilities are "independent" then the "probability" that p divides none of these denominators is

$$\left(1-\frac{1}{p}\right)^{\frac{p-3}{2}} \approx e^{-1/2} = 0.6065306597\ldots$$

[17]It seems like an improbable co-incidence that this constant appears here, but there does not seem any intuitive reason that it does. One simply obtain γ from a complicated, and unmotivated, calculation.

E4. Primes in arithmetic progressions. The 1837 Dirichlet showed that whenever $(a,q)=1$ there are infinitely many primes $\equiv a \pmod q$. Dirichlet's starting point was the formula (B5.3). From that we obtain, for s such that $\mathrm{Re}(s)>1$,

$$\sum_{\substack{p \text{ prime, } m\ge 1\\ p^m\equiv a \pmod q}} \frac{1}{p^{ms}} = \frac{1}{\phi(q)} \sum_{\chi \pmod q} \overline{\chi}(a) \left(\sum_{\substack{p \text{ prime}\\ m\ge 1}} \frac{\chi(p^m)}{p^{ms}} \right).$$

Now from

$$L(s,\chi) := \sum_{n\ge 1} \frac{\chi(n)}{n^s} = \prod_{p \text{ prime}} \left(1 - \frac{\chi(p)}{p^s}\right)^{-1},$$

we can take logarithms to obtain

$$\log L(s,\chi) = \sum_{\substack{p \text{ prime}\\ m\ge 1}} \frac{\chi(p^m)}{mp^{ms}}.$$

Therefore

$$\sum_{\substack{p \text{ prime, } m\ge 1\\ p^m\equiv a \pmod q}} \frac{1}{p^{ms}} = \frac{1}{\phi(q)} \sum_{\chi \pmod q} \overline{\chi}(a) \log L(s,\chi). \tag{E4.1}$$

For now let us assume that

(H1) if $\chi \ne \chi_0$ then $L(s,\chi) \to L(1,\chi)$, a non-zero real number, as $s \to 1^+$.

On the other hand

$$L(s,\chi_0) = \zeta(s) \prod_{p\mid q} \left(1 - \frac{1}{p^s}\right), \quad \text{which diverges as } s \to 1^+,$$

as we have seen. Entering this information into the equation above implies that $\sum 1/p^{ms}$ diverges as $s \to 1^+$. The contribution of the prime powers is

$$\le \sum_p \sum_{m\ge 2} \frac{1}{p^m} = \sum_p \frac{1}{p(p-1)} \le \sum_{n\ge 2} \frac{1}{n(n-1)} = 1,$$

and so we deduce that

$$\sum_{\substack{p \text{ prime}\\ p\equiv a \pmod q}} \frac{1}{p} = \infty;$$

which implies that there are infinitely many primes $\equiv a \pmod q$.

`Exercise E4.1.` Use (E3.6) to show, by a small modification of the above argument, that whenever $(a,q)=1$,

$$\sum_{\substack{p \text{ prime, } p\le x\\ p\equiv a \pmod q}} \frac{1}{p} = \frac{1}{\phi(q)} \log\log x + O(1).$$

Note that the constant implicit in the $O(1)$ depends on q. This results indicates that perhaps the primes are roughly equally distributed amongst the arithmetic progressions $a \pmod q$ with $(a,q)=1$.

Analytic continuation. To prove (H1) we need to show both that the $L(1,\chi)$ converge and that they take non-zero values; this is challenging. The key to such results is the notion of *analytic continuation*. The Riemann-zeta function, and the Dirichlet L-functions are only well-defined for those complex numbers s for which $\mathrm{Re}(s) > 1$ (in that the series defining them is absolutely convergent). We need to understand their values at $s = 1$, and so up to now we have looked at the limit as we come into $s = 1$ from the right. However we can circumvent that difficulty by coming up with new definitions for $\zeta(s)$ and $L(s,\chi)$ that converge in a much wider range. One way to do this, for the $L(s,\chi)$ with $\chi \neq \chi_0$, is by grouping the terms together: We write

$$L(s,\chi) = \left(\sum_{n=1}^{q} \frac{\chi(n)}{n^s}\right) + \left(\sum_{n=q+1}^{2q} \frac{\chi(n)}{n^s}\right) + \left(\sum_{n=2q+1}^{3q} \frac{\chi(n)}{n^s}\right) + \dots$$

This evidently equals $L(s,\chi)$ when $\mathrm{Re}(s) > 1$; and we will show that if we consider here a term to be each sum in parentheses, then this definition is absolutely convergent in $\mathrm{Re}(s) > 0$. To prove this we need to get a good bound on each sum in parenthesis, and we use the fact that $\sum_{n=kq+1}^{(k+1)q} \chi(n) = \sum_{j=1}^{q} \chi(kq+j) = \sum_{j=1}^{q} \chi(j) = 0$. Therefore

$$\text{(E4.2)} \qquad \left|\sum_{n=kq+1}^{(k+1)q} \frac{\chi(n)}{n^s}\right| = \left|\sum_{n=kq+1}^{(k+1)q} \chi(n)\left(\frac{1}{n^s} - \frac{1}{(kq)^s}\right)\right| \le \sum_{j=1}^{q} \left|\frac{1}{(kq+j)^s} - \frac{1}{(kq)^s}\right|.$$

Taking $s = 1$ we have $\left|\sum_{n=kq+1}^{(k+1)q} \frac{\chi(n)}{n}\right| \le q\left|\frac{1}{kq} - \frac{1}{(k+1)q}\right| = \frac{1}{k(k+1)}$.

`Exercise E4.2.` Deduce that $|L(1,\chi)| \le \log q + 2$.

This argument generalizes: In (E4.2) we pull out a factor $|(kq)^{-s}| = (kq)^{-\sigma}$ where $s = \sigma + it$, and then we need to bound $|(1 + j/kq)^{-s} - 1|$.

`Exercise E4.3.` Show that if $|sz| \le 1$ then $|(1+z)^s - 1| \le c|sz|$ for some constant $c > 0$.

`Exercise E4.4.` Combining our various bounds, show that the quantity in (E4.2) is $\le c|s|q^{1-\sigma}/k^{1+\sigma}$; and deduce that the new definition of $L(s,\chi)$ is absolutely convergent for all s with positive real part. We call this an *analytic continuation* of $L(s,\chi)$ to $\mathrm{Re}(s) > 0$.

`Exercise E4.5.` Use exercise E3.1 to provide an analytic continuation of $(1 - 2^{1-s})\zeta(s)$ to $\mathrm{Re}(s) > 0$.

This way of extending the domain of definition of a function seems to be very ad hoc: Why one method, and not another? And if you have two different methods that allow you to extend the domain of a function, might they not be different outside the region where they are both initially defined? There are several miracles connected to analytic continuation. The first is that any analytic continuations of a function take the same values; that is there is a unique function that extends the definition of the original function. It is for this reason that we write $\zeta(s)$ to mean the function that is the analytic continuation of the function that we defined (by that name) on $\mathrm{Re}(s) > 1$. Hence when we write $\zeta(-1)$ **we do not mean** $1 + 2 + 3 + \dots$ but rather the (convergent) function that is well-defined at that point.

One might reasonably ask for the values of $\zeta(-n)$ for the negative integers $-n$. Euler showed that

$$\zeta(-n) = -\frac{B_{n+1}}{n+1} \text{ for all } n \geq 1,$$

so that $\zeta(-2k) = 0$ for $k \geq 1$. These are the *trivial* zeros of $\zeta(s)$.[18]

The second miracle of analytic continuation is that any such function has a Taylor series at every point. That is, if $f(s)$ is an analytic function on all of $\mathbb{C}$ then, for any $a \in \mathbb{C}$, we can evaluate all of the derivatives of f at a, that is $f'(a), f^{(2)}(a), \dots$, and then

$$f(a+h) = f(a) + hf'(a) + \frac{h^2}{2!} f^{(2)}(a) + \dots \tag{E4.3}$$

Given these properties of analytic functions, you might guess that they are very special functions, and you would be correct! But what we find is that many of the functions that are defined naturally for number theoretic reasons, can be analytically continued to the whole complex plane.

Back to $L(1,\chi)$. We have now seen that each $L(s,\chi)$ is well-defined at $s=1$ as well as $(1-2^{1-s})\zeta(s)$. Notice that $(1-2^{1-s})$ has the Taylor series $(s-1)\log 2 + c_2(s-1)^2 + \dots$ so we see that $(s-1)\zeta(s)$ has a Taylor series. Now above we saw that $1 < (\sigma-1)\zeta(\sigma) < \sigma$ for real $\sigma > 1$, so taking the limit as $\sigma \to 1^+$ we deduce that $(s-1)\zeta(s) = 1 + \kappa_0(s-1) + \dots$. Also if $L(s,\chi)$ has a zero of order ρ_χ at $s=1$, then $L(s,\chi) = c_\chi(s-1)^{\rho_\chi}(1+\kappa_\chi(s-1)+\dots)$ at $s=1$. Multiplying these all together we find that, close to $s=1$,

$$\prod_{\chi \pmod q} L(s,\chi) = c(s-1)^{\sum_\chi \rho_\chi - 1}(1 + \kappa(s-1) + \dots).$$

Now, taking $a=1$ in (E4.1) we have that

$$\prod_{\chi \pmod q} L(s,\chi) = \exp\left(\sum_{\substack{p \text{ prime},\ m \geq 1 \\ p^m \equiv 1 \pmod q}} \frac{1}{p^{ms}} \right) \geq 1$$

for all real $s > 1$ and, in particular, is non-zero. Combining these last two displays, letting $s \to 1^+$, implies that $\sum_\chi \rho_\chi - 1 \leq 0$. Hence at most one of the $L(1,\chi)$'s equals 0.

`Exercise E4.6.` Prove that if $L(1,\chi) = 0$ then $L(1,\overline{\chi}) = 0$.

If $L(1,\chi) = 0$ then we get two zeros, which is impossible, unless $\chi = \overline{\chi}$, that is χ is real-valued, and evidently not the principal character. Hence we are left to prove that $L(1,\chi) \neq 0$ when χ is a real character. Dirichlet eventually gave a proof of this which left the realm of questions about primes and established an unforeseeable link between L-functions and the algebra of quadratic forms.

[18] Notice that $\zeta(-1) = -\frac{1}{12}$, leading some to write $1 + 2 + 3 + \dots = -\frac{1}{12}$.

Dirichlet's class number formula. In 1832 Jacobi conjectured that the class number $h(-p)$, when $p \equiv 3 \pmod 4$, is given by

$$h(-p) \ = \frac{1}{p} \sum_{n=1}^{p-1} \left(\frac{n}{p}\right) n.$$

`Exercise E4.7.` Show that the right side is an integer using Euler's criterion and Corollary 7.9.

`Exercise E4.8.` Let $S := \sum_{n=1}^{(p-1)/2} \left(\frac{n}{p}\right)$ and $T := \sum_{n=1}^{(p-1)/2} \left(\frac{n}{p}\right) n$.

(1) Show that $S = 0$ when $p \equiv 1 \pmod 4$. Henceforth assume that $p \equiv 3 \pmod 4$.
(2) Note that $\left(\frac{p-n}{p}\right)(p-n) = \left(\frac{n}{p}\right)(n-p)$. Use this to evaluate the sum $\sum_{n=1}^{p-1} \left(\frac{n}{p}\right) n$ in terms of S and T by pairing up the nth and $(p-n)$th term, for $n = 1, 2, \dots, \frac{p-1}{2}$.
(3) Do this taking $n = 2m,\ m = 1, 2, \dots, \frac{p-1}{2}$ to deduce that

$$h(-p) \ = \frac{1}{\left(\frac{2}{p}\right) - 2} \sum_{n=1}^{(p-1)/2} \left(\frac{n}{p}\right).$$

In 1838 Dirichlet gave a proof of Jacobi's conjecture an much more. His miraculous class number formula links algebra and analysis in an unforeseen way that was foretaste of many of the most important works in number theory, including Wiles' proof of Fermat's Last Theorem. We will simply state the formulae here: If $d > 0$ then

$$h(d) \log \epsilon_d \ = \ \sqrt{d}\, L\left(1, \left(\frac{d}{\cdot}\right)\right).$$

If $d < -4$ then

$$h(d) \ = \frac{1}{\pi} \ \sqrt{|d|}\, L\left(1, \left(\frac{d}{\cdot}\right)\right).$$

Note that $h(d) \geq 1$ for all d since we always have the principal form. Hence the formulae imply that $L\left(1, \left(\frac{d}{\cdot}\right)\right) > 0$ for all d, as desired in the proof that there are infinitely many primes in arithmetic progressions. In fact these formulae even give lower bounds; for example when $d < -4$ we have $L\left(1, \left(\frac{d}{\cdot}\right)\right) \geq \pi/\sqrt{|d|}$. Getting a significantly better lower bound for all d is a very difficult problem, though Heilbronn showed that there exists a constant $c > 0$ such that $L\left(1, \left(\frac{d}{\cdot}\right)\right) \geq c/\log |d|$ (and hence $h(d) > c'\sqrt{|d|}/\log |d|$) with very few exceptions (in fact no more than one value of d in the range $D < d \leq 2D$ for any D).

The size of a fundamental unit. By exercise E4.2 we have for $d > 0$, since $h(d) \geq 1$,

$$\log \epsilon_d \leq h(d) \log \epsilon_d \ = \ \sqrt{d}\, L\left(1, \left(\frac{d}{\cdot}\right)\right) \leq \sqrt{d}(\log 4d + 2)$$

since $\left(\frac{d}{\cdot}\right)$ is a character mod $4d$, but not necessarily mod d. Hence $\epsilon_d \leq (4e^2 d)^{\sqrt{d}}$. In calculations one finds that ϵ_d is often around $e^{c\sqrt{d}}$. If that is true for d then Dirichlet's class number formula implies that $h(d) < c\log d$ for some constant $c > 0$. Actually $L\left(1, \left(\frac{d}{\cdot}\right)\right)$ is much more usually close to 1; that is, it is between $\frac{1}{10}$ and 10 for more than 99% of the values of d. Hence if ϵ_d is typically around since $e^{c\sqrt{d}}$ then $h(d)$ is typically bounded.

The prime number theorem and the Riemann Hypothesis. In 1859, Riemann wrote a nine page memoir that was to shape the future of number theory. This was his only paper in number theory, but it was to shape the approach to studying the distribution of prime numbers from then on. In effect, Riemann proposed a plan to prove Gauss's guesstimate for the number of primes up to x, discussed in section 5.4. This involved moving the question from number theory to analysis, via the theory of analytic continuation.

The first observation is a simple one. The function $\mathrm{Li}(x)$, defined in (5.4.1), is not an easy one to work with (see Exercise 5.4.1). To improve on this, notice that if, as Gauss asserted, the density of primes at around x is $1/\log x$, then if we sum $\log p$ over the primes up to x, then the expected value of the sum is about x (by Gauss's statement). In fact:

`Exercise E4.9.` Prove that the prime number theorem is equivalent to the estimate $\sum_{p\le x} \log p \sim x$. (Hint: Show that the contribution of the primes $\le x/\log x$ in either sum is small, so discard them and compare the remaining two sums.)

We have seen how Dirichlet studied the distribution of primes using $\log L(s,\chi)$. For various analytic reason is much easier to work with $L'(s,\chi)/L(s,\chi)$.

`Exercise E4.10.` Show that for complex numbers s with $\mathrm{Re}(s) > 1$ we have

$$\frac{\zeta'(s)}{\zeta(s)} = \sum_{\substack{p \text{ prime} \\ m\ge 1}} \frac{\log p}{p^{ms}} .$$

(Hint: Use the Euler product formula for $\zeta(s)$.)

It is too complicated to go into all of the details here,[19] but let us just say that Riemann use this last identity to relate the number of prime powers up to x to the zeros of $\zeta(s)$. This is a surprising thing to do. After all, $\zeta(s)$ has no zeros s with $\mathrm{Re}(s) > 1$, where it is naturally defined; all of its zeros lie in the domain of the analytic continuation of $\zeta(s)$. Riemann's amazing exact formula is:

$$\sum_{\substack{p \text{ prime} \\ m\ge 1 \\ p^m \ge x}} \log p \;=\; x - \sum_{\rho:\ \zeta(\rho)=0} \frac{x^\rho}{\rho} - \frac{\zeta'(0)}{\zeta(0)} .$$

This is a little crazy. We take a nice elementary question, the count of the primes up to x, and relate it to the (infinite set of) zeros of an analytic continuation. To be able to use this formula we will need to know how many zeros ρ there are, and where they are located, both difficult problems. The size of the terms in this formula will also play a role. For example, how big is x^ρ? Does it compare to x?

`Exercise E4.11.` Prove that $|x^\rho| = x^{\mathrm{Re}(\rho)}$.

We remarked that $\zeta(s)$ has no zeros s with $\mathrm{Re}(s) > 1$, and so we know that $\mathrm{Re}(\rho) \le 1$ for each ρ. If $\mathrm{Re}(\rho) = 1$ then the x^ρ/ρ term would have size comparable to x, so we need to show that this is impossible. In fact by the end of the 19th century, researchers

[19] Though see [Da] for a wonderful introduction.

proved that if one could show that $\mathrm{Re}(\rho) < 1$ for all zeros ρ of $\zeta(\rho) = 0$ then the prime number theorem would follow. This was achieved by Hadamard and de la Vallée Poussin, independently, in 1896.

It is not difficult to show that $\zeta(s)$ has zeros at $s = -2, -4, \dots$ and, other than that all of its zeros ρ satisfy $0 \le \mathrm{Re}(\rho) \le 1$, the *critical strip*. Riemann made a few calculations of the zeros of $\zeta(s)$ and all the real parts seemed to be $1/2$. This led to him to:

The Riemann Hypothesis. *If* $\zeta(\rho) = 0$ *with* $0 \le \mathrm{Re}(\rho) \le 1$ *then* $\mathrm{Re}(\rho) = \frac{1}{2}$.

If the Riemann Hypothesis is true that each $|x^\rho| = x^{1/2}$, by the exercise, and in fact one can then deduce that there exists a constant $C > 0$ such that

$$\left| \sum_{p \le x} \log p - x \right| \le C x^{1/2} (\log x)^2.$$

This in turn implies that $|\pi(x) - \mathrm{Li}(x)| \le C' x^{1/2} \log x$. Riemann's formula implies more: These two estimates actually imply the Riemann Hypothesis. That is the Riemann Hypothesis is equivalent to the estimate given by the prime number theorem with a very strong error term.

Riemann's formula shows that speaking about the number of primes up to x, and understanding the zeros of $\zeta(s)$ are more-or-less tautologous. This led leading number theorists in the first half of the twentieth century to believe that it would be impossible to find a proof of the prime number theorem that avoids the zeros of $\zeta(s)$ – and since the zeros belong properly only to the analytic continuation of $\zeta(s)$ that any proof must therefore by non-elementary. It thus came as a great shock when, in 1949, Selberg and Erdős gave an elementary proof.[20] At the heart of both of their proofs is Selberg's extraordinary formula for the number of integers up to x that are the product of two primes, appropriately weighted, which Selberg proved in a very straightforward (though highly ingenious) way:

$$\sum_{\substack{p \le x \\ p \text{ prime}}} (\log p)^2 + \sum_{\substack{pq \le x \\ p,q \text{ primes}}} (\log p)(\log q) = 2x \log x + O(x).$$

Include range issues, eg BV, etc

[20] Elementary, but complicated!

E5. The number of prime factors of an integer. One might count $12 = 2^2 \times 3$ as having two or three primes factors depending on whether one counts the 2^2 as one or two primes. So define

$$\omega(n) = \sum_{\substack{p \text{ prime} \\ p|n}} 1 \quad \text{and} \quad \Omega(n) = \sum_{\substack{p \text{ prime}, a\geq 1 \\ p^a|n}} 1.$$

On average the difference between these two is

$$\begin{aligned}\frac{1}{x}\sum_{n\leq x}\sum_{\substack{p \text{ prime}, a\geq 2 \\ p^a|n}} 1 &= \frac{1}{x}\sum_{\substack{p \text{ prime} \\ a\geq 2}}\sum_{\substack{n\leq x \\ p^a|n}} 1 = \frac{1}{x}\sum_{\substack{p \text{ prime} \\ a\geq 2}}\left[\frac{x}{p^a}\right] \\ &\leq \sum_{p \text{ prime}, a\geq 2}\frac{1}{p^a} = \sum_{p \text{ prime}}\frac{1}{p(p-1)} \leq \sum_{n\geq 2}\frac{1}{n(n-1)} = 1\end{aligned}$$

so we can work with either, as is convenient. Note that here we used the fact that the number of integers divisible by d is $[x/d]$. Now,

$$\sum_{n\leq x}\omega(n) = \sum_{n\leq x}\sum_{\substack{p \text{ prime} \\ p|n}} 1 = \sum_{p \text{ prime}}\sum_{\substack{n\leq x \\ p|n}} 1 = \sum_{p \text{ prime}}\left[\frac{x}{p}\right]$$

Hence the average is approximately

$$\frac{1}{x}\sum_{p \text{ prime}}\frac{x}{p} = \log\log x + o(1),$$

as we saw in (*).[21] The error in this approximation is no more than 1 for each prime p, and so in total $\pi(x)/x = o(1)$.

We are going to go one step further and ask how much $\omega(n)$ varies from its mean, that is we are going to compute the statistical quantity, the *variance*. We begin with a standard identity for the variance:

`Exercise E5.1.` If $a_1, \ldots, a_N$ have average m show that $\frac{1}{N}\sum_{n\leq N}(a_n - m)^2 = \frac{1}{N}\sum_{n\leq N} a_n^2 - m^2$.

This implies that

$$\frac{1}{x}\sum_{n\leq x}\left(\omega(n) - \frac{1}{x}\sum_{m\leq x}\omega(m)\right)^2 = \frac{1}{x}\sum_{n\leq x}\omega(n)^2 - \left(\frac{1}{x}\sum_{m\leq x}\omega(m)\right)^2. \tag{E5.1}$$

[21] The notation "$o(1)$", which we will use repeatedly, stands for a function $A(x)$ for which $A(x) \to 0$ as $x \to \infty$. We use this in the context that we do not much care what the function is, simply that it goes to 0 as x goes to ∞.

Now the first term here is

$$\frac{1}{x}\sum_{n\le x}\omega(n)^2 = \frac{1}{x}\sum_{n\le x}\sum_{\substack{p \text{ prime}\\ p|n}}\sum_{\substack{q \text{ prime}\\ q|n}} 1 = \frac{1}{x}\sum_{p \text{ prime}}\left(\sum_{\substack{q \text{ prime}\\ q=p}}\left[\frac{x}{p}\right] + \sum_{\substack{q \text{ prime}\\ q\neq p}}\left[\frac{x}{pq}\right]\right)$$

$$\le \sum_{\substack{p \text{ prime}\\ p\le x}}\frac{1}{p} + \sum_{p \text{ prime}}\sum_{\substack{q \text{ prime},\ q\neq p\\ pq\le x}}\frac{1}{pq} \le \sum_{\substack{p \text{ prime}\\ p\le x}}\frac{1}{p} + \left(\sum_{\substack{p \text{ prime}\\ p\le x}}\frac{1}{p}\right)^2.$$

Hence the variance is $o((\log\log x)^2)$, and so

$$\frac{1}{x}\sum_{n\le x}(\omega(n)-\log\log x)^2 = o((\log\log x)^2).$$

Exercise E5.2. Show that this implies that, for any fixed $\epsilon > 0$, if x is sufficiently large then there are $< (1-\epsilon)x$ integers $n \le x$ for which $|\omega(n) - \log\log x| \ge \epsilon\log\log x$. In other words there $\sim x$ integers $n \le x$ for which $\omega(n) \sim \log\log x$.

Exercise E5.3. Deduce that there are $\log\log x$ integers $n \le x$ for which $\omega(n), \Omega(n) \sim \log\log n$. Colloquially speaking "Almost all integers n have about $\log\log n$ prime factors". (This is a famous result of Hardy and Ramanujan.)

Exercise E5.4. Apply the above argument carefully to show that the variance is $\le C\log\log x$.

When you were very young you probably had to learn the multiplication table off by heart. Perhaps all the values of $a \times b$ for $1 \le a, b \le 12$. Perhaps you wrote these in a grid, like:

1	2	3	4	5	6	7	8	9	10	11	12
2	4	6	8	10	12	14	16	18	20	22	24
3	6	9	12	15	18	21	24	27	30	33	36
4	8	12	16	20	24	28	32	36	40	44	48
5	10	15	20	25	30	35	40	45	50	55	60
6	12	18	24	30	36	42	48	54	60	66	72
7	14	21	28	35	42	49	56	63	70	77	84
8	16	24	32	40	48	56	64	72	80	88	96
9	18	27	36	45	54	63	72	81	90	99	108
10	20	30	40	50	60	70	80	90	100	110	120
11	22	33	44	55	66	77	88	99	110	121	132
12	24	36	48	60	72	84	96	108	120	132	144

If you were Paul Erdős you might have quickly got bored waiting for the others to learn it, and asked yourself other questions. For example, how many different integers are there in the table? There is the obvious symmetry down the diagonal, meaning that one only need look at the upper triangle for distinct entries. One spots other co-incidences, like $3 \times 4 = 2 \times 6$ and $4 \times 5 = 2 \times 10$, and wonders how many there are. We ask the following precise question: *What percentage of the integers up to N^2 appear in the N-by-N multiplication table, that is equal ab where $1 \le a, b \le N$?* The phrasing of the question presupposes that the percentage exists (but it does). Let us look at some data: For $N = 6$ we have 18 distinct entries, that is $1/2$ of N^2; for $N = 10$ we have 42 distinct entries, that is .42 of N^2; for $N = 12$ we have 59, just below 41%. Let $p(N) = \#\{\text{Distinct entries in } N\text{-by-}N \text{ table}\}/N^2$. Then $p(25) = .36$, $p(50) = .32$, $p(75) \approx .306$, $p(100) \approx .291$, $p(250) \approx .270$, $p(500) \approx .259$, $p(1000) \approx .248$. Can one guess what the limit is? In fact Erdős proved that the limit is 0, and his proof is extraordinary.[22] The idea is simply that almost all integers up to N have about $\log\log N$ prime factors and so the product of two such integers has about $2\log\log N$ prime factors. However almost all integers up to N have about $\log\log N^2 = \log\log N + \log 2$ prime factors, and so are not the product of two typical integers $\le N$.

`Exercise E5.5.` Give a rigorous proof of Erdős's Theorem. To do so you might define $G(x)$ to be the set of integers $n \le x$ for which $|\Omega(n) - \log\log x| \le \epsilon \log\log x$, and go from there.

One might also ask for the number of integers, $N(x, k)$, up to x with exactly k prime factors, including multiplicity. Note that $N(x, 1) = \pi(x) \sim x/\log x$.

Lemma E5.1. *Uniformly, for all x,*

$$\sum_{p \le \sqrt{x}} \frac{\log x}{p \log (x/p)} = \log\log x + O(1).$$

In particular there exists a constant $c_1 > 0$ such that we have that upper bound $\le \log\log x + c_1$ for all x.

Proof. Subtracting $\sum_{p \le \sqrt{x}} 1/p$ we obtain

$$\sum_{p \le \sqrt{x}} \frac{\log p}{p \log (x/p)} \le \frac{2}{\log x} \sum_{p \le \sqrt{x}} \frac{\log p}{p} \ll 1$$

by (E3.5), and the result then follows from (E3.5).

We may write each integer up to x with two prime factors as pq with $p \le q$. Therefore $p \le \sqrt{x}$ and $p \le q \le x/p$. The number of such q is $N(x/p, 1) - N(p, 1)$.

`Exercise E5.6.` Show that $\sum_{p \le x^{1/2}} N(p, 1) \le 2 \frac{x}{\log x}$.

[22] All true mathematicians are motivated by elegant proofs, none more so than the great Paul Erdős. Erdős used to say that "the supreme being" kept a book which contained all of the most beautiful proofs of each theorem and just occasionally we mortals are allowed to glimpse this book, as we discover an extraordinary proof. Erdős's proof of the multiplication table theorem is truly from the book. (See Aigner and Ziegler [] for more examples.)

The main term here is, therefore

$$\sim \sum_{p\le x^{1/2}} \frac{x/p}{\log\,(x/p)} \sim \frac{x}{\log\,x}\,\log\log\,x$$

by the lemma. Hence we have proved that

$$N(x,2)\ \sim \frac{x}{\log\,x}\,\log\log\,x.$$

One can continue like this, by induction on $k\ge 2$, to prove that if k is fixed ≥ 1 then

$$N(x,k)\ \sim \frac{x}{\log\,x}\ \frac{(\log\log\,x)^{k-1}}{(k-1)!}\quad . \tag{E5.2}$$

It is possible to extend this to k that go to infinity with x, though in a limited range, so long as $k/\log\log\,x\to 0$ as $x\to\infty$. It is also possible to prove (E5.2) when k is close to the mean value $\log\log\,x$, in fact whenever $k/\log\log\,x\to 1$ as $x\to\infty$. This lead people (including, it seems, Ramanujan) to believe that (E5.2) holds for all k in a wide range, but in 1950 Sathé showed, to great surprise, that this is false. His proof is extremely complicated, and Selberg gave a beautiful four page proof. Here is the truth: For $1\le k\le (2-\epsilon)\log\log\,x$ we have

$$N(x,k)\ \sim\ F\left(\frac{k-1}{\log\log\,x}\right)\ \frac{x}{\log\,x}\ \frac{(\log\log\,x)^{k-1}}{(k-1)!}\quad . \tag{E5.3}$$

where

$$F(s)=\prod_p \left(1-\frac{s}{p}\right)\Big/\left(1-\frac{1}{p}\right)^s.$$

In particular $F(0)=F(1)=1$ and these are the only two values for which $F(s)=1$.

We see that the end result is rather tricky and unlikely to be easily provable by straightforward methods. However Hardy and Ramanujan had already proved a general upper bound that it is straightforward: There exists constant c_0 such that

$$N(x,k)\ \le\ c_0\ \frac{x}{\log\,x}\ \frac{(\log\log\,x+c_1)^{k-1}}{(k-1)!},$$

with c_1 as in the lemma. We prove this by induction on $k\ge 1$. For $k=1$ we get this from the prime number theorem, or even the Chebyshev bounds on $\pi(x)$. For larger k, suppose that our integer with k prime factors is $n=p_1p_2\dots p_k\le x$ with $p_1\le p_2\le\dots\le p_k$. If $j\le k-1$ then $p_j^2\le p_{k-1}p_k\le x$ and so $p_j\le x$. We will give an upper bound on the number of such n by counting the number of $p_jm_j\le x$ for $1\le j\le k-1$, where $\Omega(m_j)=k-1$. Hence

$$(k-1)N(x,k)\ \le\ \sum_{p\le\sqrt{x}} N(x/p,k-1)\ \le\ \sum_{p\le\sqrt{x}} c_0\ \frac{x/p}{\log\,x/p}\ \frac{(\log\log\,(x/p)+c_1)^{k-2}}{(k-2)!},$$

by the induction hypothesis. Now we use the bound $\log\log\,(x/p)\le\log\log\,x$, and then the result follows by applying the lemma.

`Exercise E5.7.` Given another proof of exercise E5.2 using the Hardy-Ramanujan inequality.

E6. Covering sets of congruences. Are there infinitely many primes of the form $k \cdot 2^n \pm 1$ or of the form $k \pm 2^n$ for given integer k. At first sight this seems like a much more difficult question than asking about primes of the form $2^n \pm 1$, but Erdős showed, ingeniously, how these questions can be resolved for certain integers k:

Let $F_n = 2^{2^n} + 1$ be the Fermat numbers (remember that F_0, F_1, F_2, F_3, F_4 are prime and $F_5 = 641 \times 6700417$), and let k be any positive integer such that $k \equiv 1 \pmod{641F_0F_1F_2F_3F_4}$ and $k \equiv -1 \pmod{6700417}$. Now

- if $n \equiv 1 \pmod 2$ then $k \cdot 2^n + 1 \equiv 1 \cdot 2^1 + 1 = F_0 \equiv 0 \pmod{F_0}$;
- if $n \equiv 2 \pmod 4$ then $k \cdot 2^n + 1 \equiv 1 \cdot 2^2 + 1 = F_1 \equiv 0 \pmod{F_1}$;
- if $n \equiv 4 \pmod 8$ then $k \cdot 2^n + 1 \equiv 1 \cdot 2^{2^2} + 1 = F_2 \equiv 0 \pmod{F_2}$;
- if $n \equiv 8 \pmod{16}$ then $k \cdot 2^n + 1 \equiv 1 \cdot 2^{2^3} + 1 = F_3 \equiv 0 \pmod{F_3}$;
- if $n \equiv 16 \pmod{32}$ then $k \cdot 2^n + 1 \equiv 1 \cdot 2^{2^4} + 1 = F_4 \equiv 0 \pmod{F_4}$;
- if $n \equiv 32 \pmod{64}$ then $k \cdot 2^n + 1 \equiv 1 \cdot 2^{2^5} + 1 = F_5 \equiv 0 \pmod{641}$; and
- if $n \equiv 0 \pmod{64}$ then $k \cdot 2^n + 1 \equiv -1 \cdot 2^0 + 1 = 0 \pmod{6700417}$.

Every integer n belongs to one of these arithmetic progressions (these are called a *covering system* of congruences), and so we have exhibited a prime factor of $k \cdot 2^n + 1$ for every integer n. Therefore we have shown that for a positive proportion of integers k, there is no prime p such that $(p-1)/k$ is a power of 2.

`Exercise E6.1.` Deduce that $k \cdot 2^n + 1$ is composite for every integer $n \geq 0$ (with k as defined above).

`Exercise E6.2.` Prove that $2^n + k$ is composite for every integer $n \geq 0$. (That is, there is no prime p equal to k plus a power of 2.)

`Exercise E6.3.` Let ℓ be any positive integer for which $\ell \equiv -k \pmod{F_6 - 2}$. Prove that $\ell \cdot 2^n - 1$ and $|2^n - \ell|$ are composite for every integer $n \geq 0$. Deduce that a positive proportion of odd integers m cannot be written in the form $p + 2^n$ with p prime.

John Selfridge showed that at least one of the primes $3, 5, 7, 13, 19, 37$, and 73 divides $78557 \cdot 2^n + 1$ for every integer $n \geq 0$. This is the smallest k known for which $k \cdot 2^n + 1$ is always composite. It is an open problem as to whether this the smallest such k.

`Exercise E6.4.` Prove that $F_n - 2 = F_0F_1 \dots F_{n-1}$ cannot be written in the form $p + 2^k + 2^\ell$ where p is prime and $k > \ell \geq 0$. (Hint: Consider divisibility by F_r where 2^r is the highest power of 2 dividing $k - \ell$.)

E7. Prime patterns paper. Are there arbitrarily many consecutive primes in arithmetic progression? That is, can we find nonzero integers a, d such that

$$a, a+d, \dots, a+(k-1)d$$

are all prime? The smallest arithmetic progression of ten primes is given by $199, 409, 619, 829, 1039, 1249, 1459, 1669, 1879, 2089$, which we can write as $199 + 210n$, $0 \le n \le 9$. The smallest examples of k-term arithmetic progression of primes, with k between 3 and 21, are given by:

Length k	Arithmetic Progression ($0 \le n \le k-1$)	Last Term
3	$3 + 2n$	7
4	$5 + 6n$	23
5	$5 + 6n$	29
6	$7 + 30n$	157
7	$7 + 150n$	907
8	$199 + 210n$	1669
9	$199 + 210n$	1879
10	$199 + 210n$	2089
11	$110437 + 13860n$	249037
12	$110437 + 13860n$	262897
13	$4943 + 60060n$	725663
14	$31385539 + 420420n$	36850999
15	$115453391 + 4144140n$	173471351
16	$53297929 + 9699690n$	198793279
17	$3430751869 + 87297210n$	4827507229
18	$4808316343 + 717777060n$	17010526363
19	$8297644387 + 4180566390n$	83547839407
20	$214861583621 + 1884649767 0n$	572945039351
21	$5749146449311 + 26004868890n$	6269243827111

The k-term arithmetic progression of primes with smallest last term.

This famous problem was resolved by Green and Tao in 2008. One might guess that there is a k-term arithmetic progression of primes all $\le k! + 1$, for each $k \ge 3$. Green and Tao gave the bound

$$2^{2^{2^{2^{2^{2^{2^{100k}}}}}}},$$

a spectacular achievement. We will now find some surprising consequences of Green and Tao's Theorem:

There are squares filled with primes, which are in arithmetic progression, when one looks along any row, or any column, like:

5	17	29
47	59	71
89	101	113

29	41	53
59	71	83
89	101	113

503	1721	2939	4157
863	2081	3299	4517
1223	2441	3659	4877
1583	2801	4019	5237

and we want to know whether there are such squares of arbitrary size, and even such cubes of high dimension? We can answer such questions all in one go by looking for primes

$$a + n_1 b_1 + n_2 b_2 + \cdots + n_d b_d,$$
$$\text{for every } \ 0 \le n_1 \le N-1, \ 0 \le n_2 \le N-1, \ldots, 0 \le n_d \le N-1,$$

a d-dimensional cube, where this is the $(n_1, n_2, \ldots, n_d)$ entry.

To do this let $k = N^d$; by the Green-Tao theorem there exists a k-term arithmetic progression of primes, $a + jq$, $0 \le j \le k-1$. Let $b_i = N^{i-1}q$ for each i. Therefore if $j = n_1 + n_2 N + n_3 N^2 + \cdots + n_d N^{d-1}$ in base N then

$$a + n_1 b_1 + n_2 b_2 + \cdots + n_d b_d = a + jq$$

is prime for each entry of our d-dimensional cube.

Arithmetic progressions $a + nd$, $n = 1, 2, \ldots$ can be viewed as the values of a degree one polynomial. Hence the Green-Tao Theorem can be rephrased as stating that for any k there are infinitely many different degree one polynomials such that their first k values are prime. How about degree two polynomials? A famous example is the infamous quadratic polynomial $X^2 + X + 41$, which is prime for $X = 0, 1, \ldots 39$. We saw in our discussion of Rabinowiscz's criterion, in section 12.3, that 41 is the largest integer m for which $X^2 + X + m$ is prime for $X = 0, 1, \ldots m-2$. However, for each k, do there exist m such that $X^2 + X + m$ is prime for $X = 0, 1, \ldots k$? Or perhaps an arbitrary degree d polynomial whose first k values are prime? To show this we know, by the Green-Tao theorem that there exists a k^d-term arithmetic progression of primes, $a + jb$, $0 \le j \le k^d - 1$. Then $a + bi^d$ is prime for $0 \le i \le k-1$, that is the first k values of the polynomial $bx^d + a$ are all prime. Note that this technique does not yield prime values of monic polynomials; we will discuss this further in a moment.

A *magic square* is an n-by-n array of distinct integers such that the sum of the numbers in any row, or in any column, or in either diagonal, equal the same constant. These have been very popular in the recreational mathematics literature. Here are two small examples.

17	89	71
113	59	5
47	29	101

41	89	83
113	71	29
59	53	101

Examples of 3*-by-*3 *magic squares of primes.*

Do you recognize the primes involved? Do you notice any similarities with the examples of 3-by-3 squares of primes above? The reason is that every 3-by-3 square of integers in arithmetic progressions along each row and column, can be rearranged to form a 3-by-3 magic square and vice-versa!

37	83	97	41
53	61	71	73
89	67	59	43
79	47	31	101

41	71	103	61
97	79	47	53
37	67	83	89
101	59	43	73

Examples of 4-by-4 magic squares of primes.

It has long been known that there are n-by-n magic squares for any $n \geq 3$. If the entries are $m_{i,j}$, $1 \leq i, j \leq n$, then the square with (i,j)th entry $a + m_{i,j}b$ is also an n-by-n magic square. The Green-Tao theorem implies that there are infinitely many pairs of integers a, b for which all of the integers $a+\ell b$, $\min_{i,j} m_{i,j} \leq \ell \leq \max_{i,j} m_{i,j}$ are prime, and this yields infinitely many n-by-n magic squares of primes. By the obvious modifications of this argument we can show that if there is a magic cube of a given size then there are infinitely many magic cubes of primes of the same size, and the same is true for higher dimensional objects of this type.

`Exercise E7.1.` Show that there are arbitrarily large sets of primes such that the average of any two elements of the set is also prime.

Beyond primes in arithmetic progressions. To generalize the notion of an arithmetic progression $a+jd$, $0 \leq j \leq k-1$, which is a set of k linear polynomials in $\mathbb{Z}[a,d]$, we consider the k-tuple of linear polynomials $L_1(x_1, \dots, x_n), \dots, L_k(x_1, \dots, x_n) \in \mathbb{Z}[x_1, \dots, x_n]$. We wish to determine whether there are infinitely many sets of integers $\{a_1, a_2, \dots, a_n\}$ for which each $|L_j(a_1, a_2, \dots, a_n)|$ is prime. There are examples for which there are only a finite number of sets $\{a_1, a_2, \dots, a_n\}$ with each $|L_j(a_1, a_2, \dots, a_n)|$ prime; for example, if we have the polynomials $L_j(a) = dj + a$ for $1 \leq j \leq p$, where prime p does not divide integer d, then p always divides the value of one of the linear forms no matter what the choice of a. To exclude this possibility we call the set of linear forms *admissible* if, for all primes p, there are integers $a_1, \dots, a_n$ such that $\prod_{j=1}^{k} L_j(a_1, \dots, a_n)$ is not divisible by p. The *extended prime k-tuplets conjecture* states that if the set of linear forms is admissible then there are infinitely many choices of integers $a_1, \dots, a_n$ for which each $|L_j(a_1, \dots, a_n)|$ is prime.

The most famous examples of this conjecture are

— There are infinitely many pairs of primes $p, p+2$ (*the twin prime conjecture*);

— For any large even integer N there are pairs of primes $p, N-p$ (*the Goldbach conjecture*), and

— There are infinitely many pairs of primes $p, 2p+1$ (*Sophie Germain twins*).

These are all examples of *difficult pairs* of linear forms, L_1, L_2, in that there exist nonzero integers a, b, c for which $aL_1 + bL_2 = c$; and they all seem to be beyond the reach of the methods of Green and Tao. However Green, Tao and Ziegler recently showed that the method extends to all other cases of the prime k-tuplets conjecture:

Theorem E7.1. *If $L_1(x_1, \dots, x_n), \dots, L_k(x_1, \dots, x_n) \in \mathbb{Z}[x_1, \dots, x_n]$ are a set of admissible linear forms, containing no two which form a difficult pair, then there are infinitely many choices of integers $a_1, \dots, a_n$ for which each $|L_j(a_1, \dots, a_n)|$ is prime.*

This great Theorem has many applications.

`Exercise E7.2.` Show that there are infinitely many monic polynomials of degree two whose first k values are all prime.

Pythagorean triples: It is well known that any solution to $x^2 + y^2 = z^2$ in coprime

integers must be of the form

$$x = r^2 - s^2, \quad y = 2rs, \quad z = r^2 + s^2,$$

where r and s are coprime integers with $r + s$ odd. The area of the right-angled triangle with sides x, y, and z is given by

$$A = \frac{xy}{2} = rs(r + s)(r - s),$$

and must be divisible by 6 since one of r and s must be even (as $r + s$ is odd), and since one of $r, s, r^2 - s^2$ must be divisible by 3. Hence we can ask how few prime factors can $A/6$ have? In (5.1) we saw that A is the product of four factors which are linear polynomials in r and s, so there can be only finitely many pairs r, s for which $A/6$ has fewer than three prime factors. Calculations reveal that $A/6 = 1$ only for the $(3, 4, 5)$ triangle, and that $A/6$ has exactly one prime factor only for the $(5, 12, 13)$ triangle. The only Pythagorean triples for which $A/6$ has exactly two prime factors are

$$(8, 15, 17),\ (7, 24, 25),\ (12, 35, 37),\ (20, 21, 29),\ (11, 60, 61),\ \text{and}\ \ (13, 84, 85).$$

We believe that there are infinitely many Pythagorean triples for which $A/6$ has exactly three prime factors, since the prime k-tuplets conjecture predicts that there are infinitely many prime triplets $p - 6, p, p + 6$ and, when we do have such a triplet, we can take $r = p$ and $s = 6$ above. Unfortunately Theorem E7.1 does not apply here since $p, p - 6$ is a difficult pair.

`Exercise E7.3.` Use Theorem E7.1 to show that there are infinitely many Pythagorean triples such that $A/6$ has exactly four prime factors.

E8. How many twin primes are there? We discussed Gauss's heuristic that the density of primes around x is about $1/\log x$. This seems to suggest that, for fixed k, the density of prime pairs $p, p+k$ around x is about $(1/\log x)^2$; at least if one could regard the events of p being prime, and $p+k$ being prime are "independent". Of course they are not independent. In particular if $k=1$ then one of the two is even so they cannot both be prime (except if $p=2, p+1=3$); therefore to make a good guess for the number of primes we should take this into account:

If one has two random integers a, b then the probability that neither of them is divisible by p is $\left(1-\frac{1}{p}\right)^2$.

Given $k \geq 1$, we want to determine the probability that for a random integer a, neither a nor $a+k$ is divisible by p.

Now if p divides k then $a \equiv a+k \pmod p$, so neither is divisible by p if and only if $a \not\equiv 0 \pmod p$. This has probability $1-\frac{1}{p}$.

If p does not divide k then neither a nor $a+k$ is divisible by p if and only if $a \not\equiv 0$ or $-k \pmod p$. This has probability $1-\frac{2}{p}$.

Hence to make the correct adjustment we should multiply $(1/\log x)^2$ by a factor

$$\frac{1-\frac{1}{p}}{\left(1-\frac{1}{p}\right)^2} \quad \text{if } p|k, \quad \text{and} \quad \frac{1-\frac{2}{p}}{\left(1-\frac{1}{p}\right)^2} \quad \text{if } p\not|k.$$

We note that this equals 0 only where $p=2$ does not divide k. It is evident that there is no more than one prime pair $p, p+k$ when k is odd. Our heuristic yields the guess:

$$\#\{p \leq x : \ p, p+2k \text{ are prime}\} \sim \prod_{p|2k} \frac{1-\frac{1}{p}}{\left(1-\frac{1}{p}\right)^2} \prod_{p\not|2k} \frac{1-\frac{2}{p}}{\left(1-\frac{1}{p}\right)^2} \frac{x}{(\log x)^2}.$$

The constant here looks a little daunting, but that can be simplified. Note that 2 always divides $2k$ so $p=2$ contributes a factor 2. We define the *twin prime constant*

$$C := 2\prod_{p\geq 3}\left(1-\frac{2}{p}\right)\left(1-\frac{1}{p}\right)^{-2}.$$

Conjecture. *For any $k \geq 1$ we have*

$$\#\{p \leq x : \ p, p+2k \text{ are prime}\} \sim C\prod_{\substack{p\geq 3\\ p|k}} \frac{p-1}{p-2} \frac{x}{(\log x)^2}.$$

The computational evidence is very compelling that this conjecture is correct.

One can generalize the "heuristic analysis" given in this section to quantify the number of primes expected in every example of the prime k-tuplets conjecture.

E9. Other primes. In the section E7 we discussed when linear polynomials give prime values. But we also are interested in when polynomials give prime values, believing that any admissible polynomial, that is that does not have a fixed prime divisor, takes on infinitely prime values. For example we believe that there are infinitely many primes of the form $x^2 + 1$, but all conjectures of this type are open.

The proof of the prime number theorem extends to show that any admissible, irreducible binary quadratic form takes on infinitely many prime values. That is any degree 2 polynomial in two variables.

We do not know how to prove such a result for degree 3 polynomials in three variables. However Heath-Brown showed that any admissible, irreducible binary ternary form takes on infinitely many prime values; that is polynomials of the form $ax^3 + bx^2y + cxy^2 + dy^3$, as x, y run through all pairs of integers. This is very surprising because the form takes on roughly $N^{2/3}$ integer values up to N, whereas in all the previous cases mentioned one has roughly cN integer values up to N, or at worst $CN/\sqrt{\log N}$.

The first result of this type, was Friedlander and Iwaniec's result that there are infinitely many primes of the form $x^2 + y^4$.

The outstanding question of this type is to show that there are infinitely many cubic polynomials whose discriminant is prime. That is infinitely many primes of the form $4a^3 + 27b^2$.

E10. Conway's prime producing machine. Begin with the integer 2 and multiply it by the first fraction in the list

$$\frac{17}{91}, \frac{78}{85}, \frac{19}{51}, \frac{23}{38}, \frac{29}{33}, \frac{77}{29}, \frac{95}{23}, \frac{77}{19}, \frac{1}{17}, \frac{11}{13}, \frac{13}{11}, \frac{15}{14}, \frac{15}{2}, \frac{55}{1}$$

for which the product is an integer. The repeat the process with the product, and continue over and over again. One obtains the integers

$$2, 15, 825, 725, 1925, 2275, 425, 390, 330, 290, 770, \ldots$$

including only the prime powers of 2, namely $2^2,\ 2^3,\ 2^5,\ 2^7,\ 2^{11},\ 2^{13},\ 2^{17},\ 2^{19}\ldots$. This is an extraordinary way to find the primes. It is a challenge to determine why this works.

F. Analytic number theory

F1. More multiplicative functions.

An integer n is *squarefree* if there does note exist a prime p for which p^2 divides n.

A proof of Theorem 4.2 using the inclusion-exclusion principle. We saw that $\phi(p^a)$ is the total number of integers up to p^a, minus the number of those that are divisible by p.

Similarly one can determine $\phi(p^aq^b)$ as the total number of integers up to p^aq^b, minus the number of those that are divisible by p, minus the number of those that are divisible by q, plus the number of those that are divisible by pq (since they were subtracted out twice). By exercise 4.2 this yields

$$\begin{aligned}\phi(p^aq^b) &= p^aq^b - \frac{p^aq^b}{p} - \frac{p^aq^b}{q} + \frac{p^aq^b}{pq}\\ &= p^aq^b\left(1-\frac{1}{p}-\frac{1}{q}+\frac{1}{pq}\right) = p^aq^b\left(1-\frac{1}{p}\right)\left(1-\frac{1}{q}\right).\end{aligned}$$

More generally suppose that $n = p_1^{n_1}p_2^{n_2}\dots p_k^{n_k}$. The idea is to determine $\phi(n)$ by counting 1 for each integer a in the range $1 \le a \le n$ which is coprime with n, and 0 for those that are not. In other words, if ℓ is the number of distinct prime factors of (a,n) then we count $(1-1)^\ell$, since this equals 1 with $\ell = 0$, and 0 when $\ell \ge 1$. Therefore writing $\omega(m)$ for the number of distinct prime factors of m we have

$$\phi(n) = \sum_{1\le a\le n} (1-1)^{\omega((a,n))}.$$

If we expand $(1-1)^{\omega(m)}$ using the binomial theorem we obtain

$$(1-1)^{\omega(m)} = \sum_{j=0}^{\omega(m)} (-1)^j \binom{\omega(m)}{j}.$$

Now $\binom{\omega(m)}{j}$ denotes the number of subsets of $\omega(m)$ elements of size j, and if we take the set of $\omega(m)$ elements to be the $\omega(m)$ distinct prime factors of m, then the subsets of size j correspond to the squarefree divisors of m with exactly j prime factors. Hence

$$(1-1)^{\omega(m)} = \sum_{j=0}^{\omega(m)} (-1)^j \sum_{\substack{d|m,\ \omega(d)=j\\ d \text{ is squarefree}}} 1 = \sum_{\substack{d|m\\ d \text{ is squarefree}}} (-1)^{\omega(d)}.$$

Noting that $d|(a,n)$ if and only if $d|a$ and $d|n$ we therefore have

$$\begin{aligned}\phi(n) &= \sum_{1\le a\le n} \sum_{\substack{d|a,\ d|n\\ d \text{ is squarefree}}} (-1)^{\omega(d)} = \sum_{\substack{d|n\\ d \text{ is squarefree}}} (-1)^{\omega(d)} \sum_{\substack{1\le a\le n\\ d|a}} 1\\ &= \sum_{\substack{d|n\\ d \text{ is squarefree}}} (-1)^{\omega(d)}\cdot\frac{n}{d} = n\cdot \sum_{\substack{d|n\\ d \text{ is squarefree}}} \frac{(-1)^{\omega(d)}}{d}\end{aligned}$$

by exercise 4.2. To complete the proof:

`Exercise F1.1.` Using induction on the number of prime factors of n, or otherwise, prove that

$$\sum_{\substack{d|n \\ d \text{ is squarefree}}} \frac{(-1)^{\omega(d)}}{d} = \prod_{\substack{p \text{ prime} \\ p|n}} \left(1 - \frac{1}{p}\right).$$

The coefficients $(-1)^{\omega(d)}$ occur in many places in number theory. They are obviously a multiplicative function of d, called the *Mobius function* and denoted $\mu(d)$. That is $\mu(.)$ is multiplicative with $\mu(p) = -1$ and $\mu(p^k) = 0$ for all $k \geq 2$. Hence we may write the last exercise as

$$\prod_{\substack{p \text{ prime} \\ p|n}} \left(1 - \frac{1}{p}\right) = \sum_{d|n} \frac{\mu(d)}{d}.$$

We saw above that

$$\sum_{d|m} \mu(d) = (1-1)^{\omega(m)} = \begin{cases} 1 & \text{for } m = 1 \\ 0 & \text{for } m \geq 2. \end{cases} \tag{F1.1}$$

Now $\ell = m/d$ runs through the divisors of m as d does, so the last sum may also be written as $\sum_{\ell|m} \mu(m/\ell)$.

In our proof of Theorem 4.2 above we saw that

$$\phi(n) = \sum_{d|n} \mu(d)\, \frac{n}{d} = \sum_{m|n} \mu(n/m) \cdot m,$$

taking $m = n/d$. This should be compared to Proposition 4.3 which yields

$$m = \sum_{n|m} \phi(n).$$

It is often easier to write $n = \ell m$ rather than $m|n$; the two identities here then become

$$\phi(n) = \sum_{ab=n} \mu(a) b \quad \text{and} \quad m = \sum_{cd=m} \phi(d).$$

By this we mean that we obtain $\phi(n)$ as a weighted sum of m over the divisors m of n (with weight $\mu(n/m)$), and then that m is the sum of $\phi(n)$ over the divisors n of m. Rather surprisingly this is not a co-incidence but rather holds for all multiplicative functions.

The Mobius inversion formula. *Suppose that f and g are two given multiplicative functions. Then $f(n) = \sum_{ab=n} \mu(a) g(b)$ for all integers $n \geq 1$ if and only if $g(m) = \sum_{cd=m} f(d)$ for all integers $m \geq 1$.*

Proof. If $g(m)=\sum_{cd=m} f(d)$ for all integers $m\ge 1$ then

$$\sum_{ab=n}\mu(a)g(b)=\sum_{ab=n}\mu(a)\sum_{cd=b}f(d)=\sum_{acd=n}\mu(a)f(d)=\sum_{d|n}f(d)\cdot\sum_{ac=n/d}\mu(a)=f(n),$$

since this last sum is 0 unless $n/d=1$. Similarly if $f(n)=\sum_{ab=n}\mu(a)g(b)$ for all integers $n\ge 1$ then

$$\sum_{cd=m}f(d)=\sum_{cd=m}\sum_{ab=d}\mu(a)g(b)=\sum_{abc=m}\mu(a)g(b)=\sum_{b|m}g(b)\sum_{ac=m/b}\mu(a)=g(m).$$

`Exercise F1.2.` By using the Mobius inversion formula, or otherwise, prove that $\phi_n(t)=\prod_{d|n}(t^d-1)^{\mu(n/d)}$. (These cyclotomic polynomials are defined in section 7.9)

Convolutions of Dirichlet series. Given two multiplicative functions $f(a)$ and $g(b)$, we can define another multiplicative function $h(n)$ by

$$h(n):=\sum_{ab=n}f(a)g(b).$$

We have seen quite a few examples of this above. To prove this is multiplicative, suppose that $(m,n)=1$:

`Exercise F1.3.` Prove that if $ab=mn$ then there exist integers r,s,t,u with $a=rs,\ b=tu,\ m=rt,\ n=su$ with $(r,s)=(t,u)=1$.

Hence

$$h(mn)=\sum_{ab=mn}f(a)g(b)=\sum_{\substack{rs=m,\\ tu=n}}f(rs)g(tu)=\sum_{rs=m}f(rs)\sum_{tu=n}g(tu)=h(m)h(n).$$

Now consider the *Dirichlet series*

$$F(s)=\sum_{a\ge1}\frac{f(a)}{a^s},\ G(s)=\sum_{b\ge1}\frac{g(b)}{b^s},\ \text{and } H(s)=\sum_{n\ge1}\frac{h(n)}{n^s}\ .$$

Then, grouping together terms where $ab=n$ we have

$$F(s)G(s)=\sum_{a,b\ge1}\frac{f(a)g(b)}{(ab)^s}=\sum_{n\ge1}\sum_{ab=n}f(a)g(b)\cdot\frac{1}{n^s}=\sum_{n\ge1}\frac{h(n)}{n^s}=H(s).$$

Some interesting examples of Dirichlet series include $\zeta(s)=\sum_{n\ge1}\frac{1}{n^s}$ and $\zeta(s)^{-1}=\sum_{n\ge1}\frac{\mu(n)}{n^s}$. We define $\mathbf{1}(s)=1$. Hence (F1.1) comes from taking the convolution of 1 with μ, and corresponds to the identity $\zeta(s)\cdot\zeta(s)^{-1}=\mathbf{1}(s)$. The Mobius inversion formula can be re-stated as

$$G(s)=\zeta(s)F(s)\ \textit{if and only if}\ F(s)=\zeta(s)^{-1}G(s).$$

`Exercise F1.4.` Describe the identity in Proposition 4.3 in terms of Dirichlet series. What are the Dirichlet series with coefficient n, with coefficient $\tau(n)$, with coefficient $\sigma(n)$? What are the coefficients of $L(s,\chi)\zeta(s)$? When are they non-zero in the case that $\chi(n)=(-4/n)$?

F2. Character Sums. One can ask how many quadratic residues mod p there are up to some given integer x. If there are N residues and R non-residues then $N+R=x$ and $N-R=\sum_{n\le x}(n/p)$, so the difficulty of the problem is equivalent to determining the *character sum* $\sum_{n\le x}(n/p)$. One notices that $\sum_{n\le p}(n/p)=0$, since there are $(p-1)/2$ quadratic residues, and the same number of non-residues. How about up to $p/2$. If $(-1/p)=1$ then $(n/p)=((p-n)/p)$ and so $\sum_{n<p/2}(n/p)=\sum_{p/2<n<p}(n/p)$, and therefore $\sum_{n<p/2}(n/p)=\frac12\sum_{n\le p}(n/p)=0$. If $(-1/p)=-1$ this turns out to be a much deeper problem that we will discuss in section *. One can ask what is the biggest that $\sum_{n\le x}(n/p)$ can be as one varies over x. Going back to our original motivation, one might ask from what point onwards are there roughly the same number of quadratic residues as non-residues, that is for what x is $\sum_{n\le x}(n/p)$ "small" compared to x. These are both questions that are a little deep for our course.... so we should give here the best results known.

Another problem is to ask whether the values of a given polynomial are equally often quadratic residues and non-residues. So suppose that $f(x)\in\mathbb{Z}[x]$ has no repeated roots. Then can we get a good upper bound on $\sum_{1\le n\le p}(f(n)/p)$? This is of particular interest since, by Corollary 8.2

$$\#\{x,y \pmod p:\ y^2\equiv f(x) \pmod p\}=\sum_{x \pmod p} 1+\left(\frac{f(x)}{p}\right)=p+\sum_{n \pmod p}\left(\frac{f(n)}{p}\right).$$

`Exercise F2.1.` Prove that if $f(n)=an+b$ where $p\not| a$ then $\sum_{1\le n\le p}(f(n)/p)=0$.

With quadratic polynomials things are a bit more complicated.

`Exercise F2.2.` Show that if $p\not| a$ then

$$\#\{x \pmod p:\ m\equiv x^2 \pmod p\}-1=\left(\frac{a}{p}\right)(\#\{x \pmod p:\ m\equiv ax^2 \pmod p\}-1).$$

Taking $m=y^2-b$ deduce that

$$\sum_{1\le n\le p}\left(\frac{an^2+b}{p}\right)=\left(\frac{a}{p}\right)\sum_{1\le n\le p}\left(\frac{n^2+b}{p}\right).$$

Show that the solutions x,y mod p to $y^2\equiv x^2+b \pmod p$ are in 1-to-1 correspondence with the solutions r,s mod p to $rs\equiv b \pmod p$. Deduce that $\sum_{1\le n\le p}\left(\frac{n^2+b}{p}\right)=-1$, and $\sum_{1\le n\le p}\left(\frac{an^2+b}{p}\right)=-\left(\frac{a}{p}\right)$.

`Exercise F2.3.` Determine $\#\{x,y \pmod p:\ y^2\equiv ax^2+bx+c \pmod p\}$ in all cases.

We have seen that is is fairly straightforward to compute the number of solutions to $y^2\equiv f(x) \pmod p$ when $f(x)$ has degree 1 or 2. This question gets more considerably more difficult for f of degree 3 (or more). We will discuss here a couple of fairly simple cases that will come in useful later.

Given any equation $y^2+Ey=Ax^3+Bx^2+Cx+D$ we can replace y by $y/A+AE/2$ and x by $x/A-B/3$ to obtain an equation of the form $y^2=x^3+ax+b$. This transformation works in $\mathbb{C}$ and mod p for any prime $p>3$. So in general we are interested in $\#\{x,y \pmod p:\ y^2\equiv x^3+ax+b \pmod p\}$. In the next two parts we focus on the two very special cases $b=0$ and $a=0$.

The equation $y^2 = x^3 + ax$. We define

$$S_a := \sum_{1\le n\le p} \left(\frac{n^3+an}{p}\right).$$

Note that $(-n)^3 + a(-n) = -(n^3 + an)$ so that if $p \equiv 3 \pmod 4$ then $\left(\frac{n^3+an}{p}\right) + \left(\frac{(p-n)^3+a(p-n)}{p}\right) = 0$ and therefore $S_a = 0$. So henceforth assume that $p \equiv 1 \pmod 4$.

`Exercise F2.4.` By making the substitution $n \equiv m/b \pmod p$ show that $S_{ab^2} = \left(\frac{b}{p}\right) S_a$. Show that $S_0 = 0$.

Let $T_+ = S_{-1}$ and $T_- = S_a$ where $(a/p) = -1$. By exercise F2.4 we have

$$\begin{aligned}\frac{p-1}{2}\,(T_+^2 + T_-^2) &= \sum_{a \pmod p} S_a^2 = \sum_{a \pmod p}\ \sum_{m,n \pmod p} \left(\frac{m^3+am}{p}\right)\left(\frac{n^3+an}{p}\right)\\ &= \sum_{m,n \pmod p} \left(\frac{mn}{p}\right) \sum_{a \pmod p} \left(\frac{(a+m^2)(a+n^2)}{p}\right)\\ &= p \sum_{\substack{m,n \pmod p\\ m^2\equiv n^2 \pmod p}} \left(\frac{mn}{p}\right) - \sum_{m,n \pmod p} \left(\frac{mn}{p}\right)\end{aligned}$$

by exercises F2.2 and 3. The second sum here is clearly 0. In the first sum we get 0 when $n = 0$. For all other n we have $m \equiv \pm n \pmod p$ so that $\left(\frac{mn}{p}\right) = \left(\frac{\pm n^2}{p}\right) = 1$, so the sum equals $(p-1)\cdot 2$. Therefore we have

$$T_+^2 + T_-^2 = 4p.$$

Now $\left(1+\left(\frac{n-1}{p}\right)\right)\left(1+\left(\frac{n}{p}\right)\right)\left(1+\left(\frac{n+1}{p}\right)\right) = 0$ or 8 unless $n \equiv -1, 0$ or $1 \pmod p$. For $n = \pm 1$ we get $2\left(1+\left(\frac{2}{p}\right)\right)$ which sum to 0 or 8, and for $n = 0$ we get 4. Hence

$$\sum_{n \pmod p} \left(1+\left(\frac{n-1}{p}\right)\right)\left(1+\left(\frac{n}{p}\right)\right)\left(1+\left(\frac{n+1}{p}\right)\right) \equiv 4 \pmod 8.$$

But each $\sum_{n \pmod p}\left(\frac{n+i}{p}\right) = 0$ and each $\sum_{n \pmod p}\left(\frac{n+i}{p}\right)\left(\frac{n+j}{p}\right) = -1$, so the above expands out to being $p-3+S_{-1} = p-3+T_+$. Hence $T_+ \equiv -(p+1) \pmod 8$. We deduce that $T_+ = 2a,\ T_- = 2b$ where a is odd; and $a^2 + b^2 = p$. We choose the sign of a so that $a \equiv -(\frac{p+1}{2}) \pmod 4$. Summarizing:

Proposition F2.1. *Let p be a prime $\equiv 1 \pmod 4$, and a and b be those unique integers (up to sign) for which $p = a^2 + b^2$ with a odd and b even. Then*

$$\#\{x, y \pmod p :\ y^2 \equiv x^3 - x \pmod p\} = p - 2(-1)^{\frac{a+b+1}{2}} a.$$

We also have $\#\{x, y \pmod p : \ y^2 \equiv x^3 - k^2x \pmod p\} = p - 2(-1)^{\frac{a+b+1}{2}} a\left(\frac{k}{p}\right)$ *for any* $k \not\equiv 0 \pmod p$*; and* $\#\{x, y \pmod p : \ y^2 \equiv x^3 - rk^2x \pmod p\} = p - 2b\left(\frac{k}{p}\right)$ *if* $(r/p) = -1$.

Note also that $S_1 = \sum_{1\le n\le p-1}\left(\frac{n+1/n}{p}\right)$. This helps us to obtain the number or solutions to $a + b \equiv c^2 \pmod p$ where $ab \equiv 1 \pmod p$.

The equation $y^2 = x^3 + b$**.** Since the map $x \to x^3$ is an automorphism if $p \equiv 2 \pmod 3$, we then have that

$$S_b := \sum_{1\le n\le p}\left(\frac{n^3+b}{p}\right)$$

equals 0. Hence we may assume $p \equiv 1 \pmod 3$. The map $n \to n/d$ yields that $S_b = S_{bd^3}\left(\frac{d}{p}\right)$. So define $T_i = S_{g^i}\left(\frac{g^i}{p}\right)$ for $i = 0, 1, 2$, where g is a primitive root mod p. As before we have $S_0 = 0$ and so

$$\begin{aligned}\frac{p-1}{3}\left(T_0^2 + T_1^2 + T_2^2\right) &= \sum_{b \pmod p} S_b^2 = \sum_{m,n \pmod p}\ \sum_{b \pmod p}\left(\frac{(b+m^3)(b+n^3)}{p}\right)\\ &= p\sum_{\substack{m,n \pmod p\\ m^3\equiv n^3 \pmod p}} 1 - \sum_{m,n \pmod p} 1\\ &= p(1 + 3(p-1)) - p^2 = 2p(p-1),\end{aligned}$$

yielding that $T_0^2 + T_1^2 + T_2^2 = 6p$. Note also that

$$\begin{aligned}\frac{p-1}{3}\left(T_0 + T_1 + T_2\right) &= \sum_{b \pmod p}\left(\frac{b}{p}\right) S_b = \sum_{1\le n\le p}\ \sum_{b \pmod p}\left(\frac{b(b+n^3)}{p}\right)\\ &= (p-1) + \sum_{1\le n\le p-1}(-1) = 0.\end{aligned}$$

Now $((n^3 + b)/p) \equiv 1 \pmod 2$ unless $n^3 \equiv -b \pmod p$. There are three solutions to this for $b = 1$ and none for $b = g$ or g^2, and so T_0 is even and T_1, T_2 odd. We also have that if $m \not\equiv 0 \pmod p$ and $n^3 \equiv m \pmod p$ then there are 3 such solutions. Therefore $S_b \equiv (b/p) \pmod 3$, and so each $T_i \equiv 1 \pmod 3$. Hence if we write $T_0 = 2a$ then $a + T_1$ is divisible by 3, so call is $3b$. So $T_1 = -a - 3b$ and therefore $T_2 = -a + 3b$ as $T_0 + T_1 + T_2 = 0$. But then $T_0^2 + T_1^2 + T_2^2 = 6(a^2 + 3b^2)$ and therefore

$$p = a^2 + 3b^2.$$

Finding the sign of a is easy since $a \equiv 2 \pmod 3$.

Proposition F2.2. *Let p be a prime $\equiv 1 \pmod 3$, and a and b be those unique integers (up to sign) for which $p = a^2 + 3b^2$ with $a \equiv 2 \pmod 3$. Then, for any $k \not\equiv 0 \pmod p$,*

$$\#\{x, y \pmod p : \ y^2 \equiv x^3 + k^3 \pmod p\} = p - 2a\left(\frac{k}{p}\right).$$

Moreover $\#\{x, y \pmod p : \ y^2 \equiv x^3 + g^i k^3 \pmod p\} = p + (a + 3(-1)^i b)\left(\frac{k}{p}\right)$ for $i = 1$ or 2.

Chevalley-Warning theorem. Let $f(x_1, x_2, \dots, x_n) \in \mathbb{Z}[x_1, x_2, \dots, x_n]$ and suppose that it has degree $d < n$.[23] Then the number of solutions to $f \equiv 0 \pmod p$ is congruent to

$$\sum_{m_1, \dots, m_n \pmod p} 1 - f(m_1, \dots, m_n)^{p-1} \pmod p.$$

The first term evidently sums to $p^n \equiv 0 \pmod p$. When we expand the second term we get a sum of terms, each of total degree $\le d(p-1)$. For the sum, over the m_i, of the term to be non-zero, the degree in each variable must be $\ge p-1$ (by Corollary 7.9), and so the total degree of the term must be $\ge n(p-1)$. This implies that $n(p-1) \le d(p-1)$ and hence $d \ge n$, a contradiction. We deduce that

$$\#\{m_1, \dots, m_n \pmod p : \ f(m_1, \dots, m_n) \equiv 0 \pmod p\} \equiv 0 \pmod p.$$

Therefore if $f(0, 0, \dots, 0) = 0$, that is f has a zero constant term, then there are $\ge p-1$ distinct non-zero solutions to $f(m_1, \dots, m_n) \equiv 0 \pmod p$.

One example is the equation $ax^2 + by^2 + cz^2 \equiv 0 \pmod p$.

Gauss Sums. For a character $\chi \pmod q$, define the *Gauss sum*

$$g(\chi) := \sum_{a \pmod q} \chi(a) e\left(\frac{a}{q}\right).$$

`Exercise F2.5.` Show that $g(\chi_0) = \mu(q)$.

`Exercise F2.6.` Show that $\overline{g(\chi)} = \chi(-1) g(\overline{\chi})$. (Hint: Look for a simple change of variable.)

For fixed m with $(m, q) = 1$ we change the variable a to mb, as b varies through the residues mod q, coprime to q, so that

$$g(\chi, m) := \sum_{a \pmod q} \chi(a) e\left(\frac{am}{q}\right) = \overline{\chi}(m) g(\chi).$$

[23] The degree of $x_1^{e_1} x_2^{e_2} \dots x_n^{e_n}$ is $e_1 + e_2 + \dots + e_n$.

Therefore, for $q = p$ prime and χ non-principal,

$$\begin{aligned}|g(\chi)|^2 =& \frac{1}{p-1}\sum_{m=1}^{p-1}|g(\chi,m)|^2 = \frac{1}{p-1}\sum_{m=0}^{p-1}|g(\chi,m)|^2\\ =& \frac{1}{p-1}\sum_{m=0}^{p-1}\sum_{a \pmod p}\chi(a)\exp\left(\frac{am}{p}\right)\sum_{b \pmod p}\overline{\chi}(b)\exp\left(-\frac{bm}{p}\right)\\ =& \sum_{a \pmod p}\sum_{b \pmod p}\chi(a)\overline{\chi}(b)\,\frac{1}{p-1}\sum_{m=0}^{p-1}\exp\left(\frac{(a-b)m}{p}\right)\\ =& \sum_{\substack{a,b \pmod p\\ a\equiv b \pmod p}}\chi(a)\overline{\chi}(b)\,\frac{p}{p-1} = p,\end{aligned}$$

and so

$$|g(\chi)| = \sqrt{p}.$$

Now suppose that $q = rs$ where $(r, s) = 1$, so there exist integers u, v such that $us+vr = 1$, and hence

$$\frac{1}{q} = \frac{u}{r} + \frac{v}{s}\ .$$

Write χ as $\rho\sigma$ where ρ has conductor r and σ has conductor s. Now $a = aus + avr \equiv aus \pmod r$ and $\equiv avr \pmod s$, so that $\chi(a) = \rho(a)\sigma(a) = \rho(aus)\sigma(avr)$. Hence, letting $b \equiv au \pmod r$ and $c \equiv av \pmod s$, we have

$$\begin{aligned}g(\chi) =& \sum_{a \pmod q}\rho(aus)\sigma(avr)e\left(\frac{au}{r}+\frac{av}{s}\right)\\ =& \rho(s)\sigma(r)\sum_{b \pmod r}\rho(b)e\left(\frac{b}{r}\right)\cdot\sum_{c \pmod s}\sigma(c)e\left(\frac{c}{s}\right) = \rho(s)\sigma(r)g(\rho)g(\sigma).\end{aligned}$$

Hence $|g(\chi)| = |g(\rho)|\ |g(\sigma)|$.

`Exercise F2.7.` Show that if χ is a non-principal character modulo squarefree q then $|g(\chi)| = \sqrt{q}$.

Now, by changing variables $b = q - a$,

$$\overline{g(\chi)} = \sum_{a \pmod q}\overline{\chi}(a)\exp\left(-\frac{a}{q}\right) = \sum_{b \pmod q}\overline{\chi}(q-b)\exp\left(\frac{b}{q}\right) = \chi(-1)g(\overline{\chi}).$$

`Exercise F2.8.` Deduce that if χ is a non-principal real character then $g(\chi) = \pm\sqrt{\chi(-1)p}$. Deciding which of these two choices gives the value of $g(\chi)$ is a substantially more difficult question, which took Gauss four years to resolve!

Another proof of the law of quadratic reciprocity. Let $\chi = (./q)$ where q is an odd prime. By exercise F2.8 and since χ is real we have $\overline{g(\chi)} = \chi(-1)g(\chi)$, so that $g(\chi)^2 = \chi(-1)|g(\chi)|^2 = \chi(-1)q$, by the above. Let p be any different odd prime so that

$$g(\chi)^p \equiv \sum_{a \pmod q} \left(\frac{a}{q}\right)^p e\left(\frac{ap}{q}\right) = \left(\frac{p}{q}\right) \sum_{b \pmod q} \left(\frac{b}{q}\right) e\left(\frac{b}{q}\right) = \left(\frac{p}{q}\right) g(\chi) \pmod p$$

letting $b \equiv ap \pmod q$. Now multiplying through by $\overline{g(\chi)}$, and dividing through by q, we obtain

$$(\chi(-1)q)^{\frac{p-1}{2}} = g(\chi)^{p-1} \equiv \chi(p) \pmod p.$$

Now $q^{\frac{p-1}{2}} \equiv (q/p) \pmod p$ by Euler's criterion, and $\chi(-1) = (-1)^{\frac{q-1}{2}}$. Putting all this together yields the law of quadratic reciprocity, as desired.

F3. The least quadratic non-residue.

Theorem F3.1. *For every odd prime $p \equiv 3 \pmod 4$ there exists a prime $q < 2\sqrt{p}$ with $\left(\frac{q}{p}\right) = -1$.*

Proof. If $p \equiv 3 \pmod 4$ select $a = [\sqrt{p}]$ so that $1 \le p - a^2 < p - (\sqrt{p}-1)^2 = 2\sqrt{p} - 1$. Now $\left(\frac{p-a^2}{p}\right) = \left(\frac{a}{p}\right)^2 \left(\frac{-1}{p}\right) = -1$ as $p \equiv 3 \pmod 4$, and so there exists a prime factor q of $p - a^2$ for which $\left(\frac{q}{p}\right) = -1$.

Theorem F3.2. *For every odd prime $p \equiv 1 \pmod 4$ there exists a prime $q < \sqrt{p}$ with $\left(\frac{q}{p}\right) = -1$.*

Proof. If $p \equiv 1 \pmod 4$ and $\left(\frac{q}{p}\right) = 1$ for all primes $q \le N$ then $\left(\frac{n}{p}\right) = 1$ for all integers $n \le N$. Let b be any quadratic non-residue $\pmod p$. Then $b, 2b, 3b, \ldots, Nb \pmod p$ are also quadratic non-residues mod p. By the pigeonhole principle there exist $0 \le i < j \le N$ such that the least positive residues of ib and $jb \pmod p$ differ by $< \frac{p}{N}$. Now let $n = j - i$ and $k = |n|$ so that $\left(\frac{k}{p}\right) = \left(\frac{n}{p}\right)$ as $\left(\frac{-1}{p}\right) = 1$. Therefore, if $B \equiv kb \pmod p$ then $0 < B < \frac{p}{N}$, and $\left(\frac{B}{p}\right) = \left(\frac{k}{p}\right)\left(\frac{b}{p}\right) = -1$. This gives a contradiction for $N > \sqrt{p}$.

It is hard to resist giving another result of this type even though it is not strictly on the topic.

Theorem F3.3. *If $p \equiv 1 \pmod 4$, $p > 17$, there exists a prime $q < 4(\sqrt{p}+1)$ with $\left(\frac{-p}{q}\right) = -1$.*

Proof. Let $2a$ be that even integer immediately greater than $\sqrt{p}$, so that $4a^2 - p \equiv 3 \pmod 4$. Let q be a prime divisor of $4a^2 - p$ which is $\equiv 3 \pmod 4$ so that $p \equiv 4a^2 \pmod q$ and hence $\left(\frac{p}{q}\right) = 1$. But then $\left(\frac{-p}{q}\right) = -1$ as $q \equiv 3 \pmod 4$. Also $2a < \sqrt{p} + 2$ and so $q \le 4a^2 - p < (\sqrt{p}+2)^2 - p = 4(\sqrt{p}+1)$.

The proof of Rabinowicz's criterion in section 12 implies, since $\left(\frac{d}{p}\right) = \left(\frac{p}{|d|}\right)$:

Theorem F3.4. *Let q be a prime $\equiv -1 \pmod 4$. Then $\left(\frac{p}{q}\right) = -1$ for all primes $p < \frac{q+1}{4}$ if and only if $h(-q) = 1$.*

Therefore we see that finding a small prime p with $\left(\frac{p}{q}\right) = 1$, can be a deep problem.

To find quadratic non-residues one can appeal to several results. The first is the Polya-Vinogradov Theorem (1919) which shows that for any non-principal character $\chi \bmod q$ one has, for any M and N.

$$\left| \sum_{n=M+1}^{M+N} \chi(n) \right| \le \sqrt{q} \log q.$$

Since $|\chi(n)| \le 1$ for all n, this is a non-trivial bound only if the length of the interval, N, is somewhat larger than $\sqrt{q}\log q$.

`Exercise F3.1.` Deduce that the smallest n for which $\chi(n) \ne 1$ is $< 2\sqrt{q}\log q$.

If χ is a character of order k then the non-zero values of $\chi(n)$ are kth roots of unity. If we imagine that they are distributed much as N random kth roots of unity would be distributed then we might expect that the maximum value of the sum (as we vary over M) is about $\sqrt{N}\log q$; this perhaps indicates why we get this bound. However if N is smaller then we might expect far smaller sums than the bound given by the Polya-Vinogradov Theorem. Indeed what we would really like is to have

$$\sum_{n=1}^{N} \chi(n) = o(N).$$

We believe that this is true if, for instance $N = q^\epsilon$ for any fixed $\epsilon > 0$. The best result known is due to Burgess (1962), that this holds when $N = q^{1/4+\epsilon}$. One can deduce from this that the least quadratic non-residue mod prime q is $< q^{1/4}$; and with some ingenuity that it is $< q^{1/4\sqrt{e}+\epsilon}$. These results have not been significantly improved in a long time, and fall far short of Vinogradov's conjecture that the least quadratic non-residue is $< q^\epsilon$ for all sufficiently large q.

F4. Other ways of counting solutions.

If $p \equiv 1 \pmod 4$ then

$$\binom{\frac{p-1}{2}}{\frac{p-1}{4}} \equiv \left(1+ \frac{2^{p-1}-1}{2}\right)\left(2a- \frac{p}{2a}\right) \pmod{p^2}.$$

This follows from the facts that

$$\binom{\frac{p-1}{2}}{\frac{p-1}{4}} \equiv 2a \pmod p \text{ and } \binom{\frac{p-1}{2}}{\frac{p-1}{4}}^2 \equiv 2^{p+1}a^2 - 2p \pmod{p^2}.$$

We also have

$$\binom{p-1}{\frac{p-1}{2}} \equiv (-1)^{\frac{p-1}{2}}(1+2^p-2) \pmod{p^2} \text{ and } \binom{p-1}{\frac{p-1}{4}} \equiv (-1)^{\frac{p-1}{4}}(1+3(2^{p-1}-1) \pmod{p^2}$$

How many solutions are there to $f(x) = 0$? The trick is to use the fact that

$$1 - f(n)^{p-1} \equiv \begin{cases} 1 \pmod p & \text{if } f(n) \equiv 0 \pmod p; \\ 0 \pmod p & \text{if } f(n) \not\equiv 0 \pmod p, \end{cases}$$

by Fermat's little theorem, so that

$$\#\{n \pmod p : f(n) \equiv 0 \pmod p\} \equiv \sum_{n \pmod p} (1-f(n)^{p-1}) \equiv - \sum_{n \pmod p} f(n)^{p-1} \pmod p.$$

`Exercise F4.1.` Show that the number of solutions $x, y \pmod p$ to $y^2 = f(x) \bmod p$ is congruent to

$$- \sum_{m,n \pmod p} (m^2 - f(n))^{p-1} \pmod p.$$

We expand this using the binomial theorem to obtain

$$\begin{aligned} & - \sum_{n \pmod p} \sum_{j=0}^{p-1} \binom{p-1}{j} (-f(n))^j \sum_{m \pmod p} m^{2(p-1-j)} \\ & \equiv \sum_{n \pmod p} (1 + \binom{p-1}{\frac{p-1}{2}} (-f(n))^{\frac{p-1}{2}}) \pmod p \\ & \equiv \sum_{n \pmod p} f(n)^{\frac{p-1}{2}} \pmod p \end{aligned}$$

using Corollary 7.9, since $\binom{p-1}{\frac{p-1}{2}} = \prod_{i=1} \frac{p-1}{2} \frac{p-i}{i} \equiv (-1)^{\frac{p-1}{2}} \pmod p$.

Now if $f(x) = ax^2 + bx + c$ we have, by the multinomial theorem,

$$(an^2 + bn + c)^{\frac{p-1}{2}} = \sum_{i+j+k=\frac{p-1}{2}} \frac{\frac{p-1}{2}!}{i!j!k!} (an^2)^i (bn)^j c^k;$$

if the exponent on n is divisible by $p-1$ and > 0 then $i = \frac{p-1}{2},\ j = k = 0$, and so

$$\sum_{n \pmod p} f(n)^{\frac{p-1}{2}} \equiv -a^{\frac{p-1}{2}} \pmod p$$

by Corollary 7.9. Now . Collecting up this information yields that the number of solutions to $y^2 = ax^2 + bx + c \bmod p$ is congruent to $-a^{\frac{p-1}{2}} \equiv -(a/p) \pmod p$.

We can ask other questions. Like the number of solutions to $y^2 = x^3 + ax + b \bmod p$ is congruent to

$$\sum_{\substack{i+j+k=\frac{p-1}{2} \\ i,j,k\geq 0}} \sum_{n \pmod p} \frac{\frac{p-1}{2}!}{i!j!k!} (n^3)^i (an)^j b^k \equiv - \sum_{\substack{i+j+k=\frac{p-1}{2} \\ 3i+j=p-1}} \frac{\frac{p-1}{2}!}{i!j!k!} a^j b^k.$$

Now, the conditions $i+j+k = \frac{p-1}{2},\ 3i+j = p-1$ imply that $j = p-1-3i,\ k = 2i - \frac{p-1}{2}$ and so the above becomes

$$\sum_{\frac{p-1}{4} \leq i \leq \frac{p-1}{3}} c_i a^{p-1-3i} b^{2i-\frac{p-1}{2}} \quad \text{where } c_i := - \frac{\frac{p-1}{2}!}{i!(p-1-3i)!(2i-\frac{p-1}{2})!} \ .$$

If $a = 0$ then this is 0 unless $p \equiv 1 \pmod 3$ in which case the $i = \frac{p-1}{3}$ term gives

$$-\binom{\frac{p-1}{2}}{\frac{p-1}{6}} b^{\frac{p-1}{6}}$$

If $b = 0$ then this is 0 unless $p \equiv 1 \pmod 4$ in which case the $i = \frac{p-1}{4}$ term gives

$$-\binom{\frac{p-1}{2}}{\frac{p-1}{4}} a^{\frac{p-1}{4}}$$

Otherwise $\#\{x, y \pmod p :\ y^2 \equiv x^3 + ax + b \pmod p\} \equiv \left(\frac{b}{p}\right) H(b^2/a^3)$ where we have $H(t) := \sum_{\frac{p-1}{4} \leq i \leq \frac{p-1}{3}} c_i t^i$. Hence we see that when it comes to counting points the key variable is b^2/a^3 (and allowing the values here 0 and ∞).

At first sight the reduction from the two variables a, b to one, b^2/a^3, is quite surprising.

`Exercise F4.2.` Show that if $b^2/a^3 \equiv t \pmod p$ where $t \not\equiv 0 \pmod p$ then there exists $m \pmod p$ such that $a \equiv m^2 t,\ b \equiv m^3 t^2 \pmod p$.

Now given the curve $y^2 \equiv x^3+tm^2x+m^3t^2 \pmod p$, let's substitute $y = my,\ x = mx$ to obtain $y^2 \equiv m(x^3 + tx + t^2) \pmod p$. The number of solutions is

$$\sum_{x \pmod p} 1 + \left(\frac{m}{p}\right)\left(\frac{x^3+tx+t^2}{p}\right) \equiv \left(\frac{b}{p}\right) \sum_{x \pmod p} \left(\frac{x^3+tx+t^2}{p}\right),$$

since $b \equiv m(mt)^2 \pmod p$, as expected.

The amazing theorem of Hasse states that

$$|\#\{x, y \pmod p :\ y^2 \equiv x^3 + ax + b \pmod p\} - p| \le 2\sqrt{p},$$

which means we can identify the precise number using the congruence above. This was generalized by Weil, so that if $f(x)$ is a polynomial of degree d that has no repeated factors mod p then

$$|\#\{x, y \pmod p :\ y^2 \equiv f(x) \pmod p\} - p| \le (d-1)\sqrt{p}.$$

Some basic sums. In exercise E3.2 we evaluated the sum $1/n$ to a good level of accuracy. Our goal now is to prove that $\lim_{N\to\infty}(1/1 + 1/2 + 1/3 + \cdots + 1/N - \log N)$ exists — it is usually denoted by γ and called the *Euler-Mascheroni constant*. Now let $x_n = 1/1 + 1/2 + 1/3 + \cdots + 1/n - \log n$ for each integer $n \ge 1$.

`Exercise F4.3.` (a) By the same argument as in exercise E3.2 show that if $n > m$ then $0 \le x_m - x_n \le \log(1 + 1/m) - \log(1 + 1/n) < 1/m$. Thus x_m is a Cauchy sequence and converges to a limit as desired. It can be shown that $\gamma = .5772156649\ldots$.

b) Prove that $0 \le 1/1 + 1/2 + 1/3 + \cdots + 1/N - \log N - \gamma \le 1/N$.

c) Let $\{t\} = t - [t]$ denote the *fractional part* of t. Prove that

$$\gamma = 1 - \int_1^\infty \frac{\{t\}}{t^2}\ dt.$$

The hyperbola trick. *What is the average number of divisors of integers up to x?* The easiest way to do this is to write the appropriate sums out, using exercise A4.3:

$$\sum_{n\le x}\sum_{d|n} 1 = \sum_{d\le x}\sum_{\substack{n\le x\\ d|n}} 1 = \sum_{d\le x}\left[\frac{x}{d}\right] = \sum_{d\le x}\Big(\frac{x}{d} + O(1)\Big) = x\sum_{d\le x}\frac{1}{d} + O(x) = x(\log x + O(1)).$$

Dirichlet, however, noted a nice trick to improve the error term here: The poor error term was caused by summing over the integers d all the way up to x. What Dirichlet noted was that divisors come in pairs $ab = n$ with $a \le b$; so instead of counting 1 for each of a and b, rather count 2 for a (unless it is $= b = \sqrt{n}$ in which case we count 1). Therefore, using exercise A4.3,

$$\begin{aligned}\sum_{n\le x}\sum_{d|n} 1 &= \sum_{n\le x}\sum_{\substack{d|n\\ d<\sqrt{n}}} 2 + \sum_{\substack{a\ge 1\\ a^2=d\le x}} 1 = \sum_{d<\sqrt{x}}\sum_{\substack{d^2<n\le x\\ d|n}} 2 + [\sqrt{x}] = 2\sum_{d<\sqrt{x}}\Big(\Big[\frac{x}{d}\Big] - d\Big) + O(\sqrt{x})\\ &= 2\sum_{d<\sqrt{x}}\Big(\frac{x}{d} - d + O(1)\Big) + O(\sqrt{x}) = 2x\sum_{d<\sqrt{x}}\frac{1}{d} - x + O(\sqrt{x})\\ &= x(\log x + 2\gamma - 1) + O(\sqrt{x}).\end{aligned}$$

The error term improves from a multiple of x, to a multiple of $\sqrt{x}$; a remarkable improvement! Getting as strong an error term as possible is an important challenge.

F5. Sums of two squares, 4 squares and quaternions (see H and W).

Look for solutions to $u^2 + v^2 \equiv 0 \pmod p$, so that $(u + iv, p)(u - iv, p) = (p)$. Now show that these two ideals are principal.

This is the same as the quadratic form proof in disguise.

Let $r(n)$ be the number of ways in which n can be written as the sum of two squares.

We need to prove that there is a unique way to write $p \equiv 1 \pmod 4$, say $p = a^2 + b^2$. Then we have $p = (\pm a)^2 + (\pm b)^2 = (\pm b)^2 + (\pm a)^2$, that is $r(p) = 8$. We also have the unique factorization $p = (a+ib)(a-ib)$ so just two prime factors, and there are four units $1, -1, i, -i$. Let $R(n) = r(n)/4$, so that $R(p) = 2$, corresponding to the two possibilities $a + ib$ and $a - ib$. Now there are the three factors $(a + ib)^2, (a + ib)(a - ib), (a - ib)^2$ of p^2 so that $R(p^2) = 3$, and in general p^k has the factors $(a + ib)^j(a - ib)^{k-j}$ for $0 \le j \le k$, so that $R(p^k) = k + 1$.

Now $2 = i(1 - i)^2$ so that $R(2^k) = 1$. Finally, if $p \equiv 3 \pmod 4$ then $R(p^{\text{odd}}) = 0$ and $R(p^{\text{even}}) = 1$.

Hence $r(n) = 4R(n)$ is a multiplicative function. By Theorem 9.3 we saw that $r(n) \neq 0$ if and only if we can write $n = 2^k m_+ m_-^2$ where if $p|m_\pm$ then $p \equiv \pm 1 \pmod 4$. In that case $R(n) = \tau(m_+)$.

When we write $p = a^2 + b^2$ it would be nice to have an easy way to determine a and b.

In section F2 we saw that if $s_m := \sum_{1\le n\le p} \left(\frac{n^3-mn}{p}\right)$ then $a = s_{-1}/2$ and $b = s_r/2$ where $(r/p) = -1$. In section H2 we will deduce from this that a is the least residue of $\binom{\frac{p-1}{2}}{\frac{p-1}{4}} \pmod p$.

Serret used continued fractions: Suppose that $r^2 \equiv -1 \pmod p$ with $0 < r < p/2$. Let us suppose that $p/r = [a_0, a_1, \dots, a_n]$. By exercise C2.3.1 we know that $a_n \ge 2$, and this starts an induction hypothesis that shows that the entries in the first column of

$$\begin{pmatrix} a_k & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_{k+1} & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix},$$

are at least twice the corresponding entries in the second column. In particular, when $k = 0$ we have $p = p_n \ge 2p_{n-1}$ where $p_n/q_n = p/r$. Taking determinants we know that $-p_{n-1}r \equiv pq_{n-1} - p_{n-1}r = p_nq_{n-1} - p_{n-1}q_n = \pm 1 \pmod p$, so that $p_{n-1} \equiv \pm r \pmod p$. Now $p_{n-1} \equiv \pm r \pmod p$, together with $0 \le p_{n-1} < p/2$, implies that $p_{n-1} = r$. Hence if $r^2 + 1 = p\ell$ then

$$\begin{pmatrix} p & r \\ r & \ell \end{pmatrix} = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix},$$

and we know that $n + 1$ is even taking determinants. Moreover by taking the transpose we see that $a_j = a_{n-j}$ for each j. Hence if $n + 1 = 2m$ and we let

$$\begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_m & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

then

$$\begin{pmatrix} p & r \\ r & \ell \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix},$$

so that $p = a^2 + b^2$.

There is also a method due to Legendre using continued fractions: In section C2.5 we saw that $\sqrt{d} + [\sqrt{d}] = [\overline{2a_0, a_1, \ \dots, \ a_{n-1}}]$ with $a_j = a_{n-j}$ for $j = 1, 2, \dots, n-1$. Now we have that

$$\begin{pmatrix} \sqrt{d} \\ 1 \end{pmatrix} = \begin{pmatrix} p_m & p_{m-1} \\ q_m & q_{m-1} \end{pmatrix} \begin{pmatrix} \alpha_{m+1} \\ 1 \end{pmatrix} \text{ so that } \begin{pmatrix} \alpha_{m+1} \\ 1 \end{pmatrix} = (-1)^{m-1} \begin{pmatrix} q_{m-1} & -p_{m-1} \\ -q_m & p_m \end{pmatrix} \begin{pmatrix} \sqrt{d} \\ 1 \end{pmatrix},$$

and so $\alpha = \frac{p_{m-1} - q_{m-1}\sqrt{d}}{p_m - q_m\sqrt{d}} = \frac{p_{m-1}p_m - dq_{m-1}q_m + (-1)^m\sqrt{d}}{p_m^2 - dq_m^2}$. Write $a = (-1)^m(p_m^2 - dq_m^2)$ and $b = (-1)^m(p_{m-1}p_m - dq_{m-1}q_m)$ so that $(b + \sqrt{d})/a = [a_{m+1}, a_{m+2}, \dots]$, with a and b integers.

If n is even, say $n = 2m$, then $(b + \sqrt{d})/a = [\overline{a_{m+1}, a_{m+2}, \ \dots, \ a_{3m}}]$ is periodic, and the period is symmetric so that $-a/(b - \sqrt{d}) = [\overline{a_{3m}, \ \dots, \ a_{m+1}}] = (b + \sqrt{d})/a$ by the discussion in section C2.4, which implies that $d = a^2 + b^2$.

One can show that if d is a prime $p \equiv 1 \pmod 4$ then n is even and so we obtain integers a and b for which $p = a^2 + b^2$. In fact n is even if and only if the fundamental unit ϵ_d has norm -1, and thus in this case we always get a solution to $d = a^2 + b^2$ from the continued fraction.

More on the number of representations. Given a discriminant $d < 0$ and an integer a we are interested in how many inequivalent primitive representations of n there are by binary quadratic forms of discriminant d. Let $f(x, y)$ be a reduced form and suppose that $n = f(\alpha, \gamma)$ where $(\alpha, \gamma) = 1$. We choose integers β, δ so that $\alpha\delta - \beta\gamma = 1$ and transform f to an equivalent binary quadratic form with leading coefficient n so that our representation becomes $n = f(1, 0)$. Next we transform $x \to x + ky, \ y \to y$ so that f is equivalent to a binary quadratic form $nx^2 + Bxy + Cy^2$ with $B^2 \equiv d \pmod{4n}$ and $-n < B \le n$ (and $C = (B^2 - d)/4n$).

`Exercise F5.1.` Prove that if $f(\alpha x + \beta y, \gamma x + \delta y) = nx^2 + Bxy + Cy^2$ then, no matter what integers β, δ we take satisfying $\alpha\delta - \beta\gamma = 1$, we get the same value of $B \pmod{2n}$.

We need to determine whether any two representations of n by f lead to the same $nx^2 + Bxy + Cy^2$. If so we have $Tf = nx^2 + Bxy + Cy^2 = Uf$ for two different (and invertible) transformations T, U and so $T^{-1}Uf = f$. So we will now find all *automorphisms* of reduced $f = ax^2 + bxy + cy^2$. Now given any such automorphism we must have $a = f(\alpha, \gamma)$ and $c = f(\beta, \delta)$.

`Exercise F5.2.` (0) Prove that the automorphisms form a group, containing $\pm I$ (that is $(x, y) \to \pm(x, y)$).

(1) Use exercise 12.2.2(i) to show that if $0 < |b| < a < c$ then the only automorphisms are $\pm I$.
(2) Show that if $b = 0$ and $a < c$ we also have $(x, y) \to \pm(x, -y)$.
(3) Show that if $|b| < a = c$ we also have $(x, y) \to \pm(y, x)$.
(4) Determine the complete set of automorphisms in all cases.

We deduce then that, in all except certain special cases, the number, $r_d(n)$, of representations of n by all binary quadratic forms of discriminant d is equal to the number of solutions $B \pmod{2n}$ to $B^2 \equiv d \pmod{4n}$. By the Chinese Remainder Theorem we see that $r_d(n)$ is a multiplicative function, so suppose that $n = p^e$ where p is prime, with $(p, d) = 1$. If p is odd then $B^2 \equiv d \pmod{p^e}$ and $B \equiv d \pmod 2$. Hence $r_d(p^e) = 1 + \left(\frac{d}{p}\right)$. If $p = 2$ then $r_d(n) = 0$ unless $d \equiv 1 \pmod 8$ in which case $r_d(n) = 2$. Hence if $(n, d) = 1$ then either $r_d(n) = 0$ or $r_d(n) = 2^{\omega(n)}$.

We can also look at

$$N(x) = \#\{(a, b) \in \mathbb{Z} : \ a^2 + b^2 \le x\} = \sum_{n \le x} r(n).$$

Notice that this should be well approximated by the area of the circle πx with an error proportional to the circumference, that is bounded by a multiple of $\sqrt{x}$.

Let $f(x, y)$ be a binary quadratic form and let $r_f(N)$ be the number of representations of N by f; that is, the number of pairs of integers m, n for which $f(m, n) = N$.

Lagrange's Theorem. *Every positive integer is the sum of four squares*

Proof. We start from the identity

$$\begin{aligned}(a^2 + b^2 + c^2 + d^2)(u^2 + v^2 + w^2 + x^2) &= (au + bv + cw + dx)^2 + (av - bu - cx + dw)^2 \\ &\quad + (aw + bx - cu - dv)^2 + (ax - bw + cv - du)^2, \end{aligned} \tag{F5.1}$$

which is much like what we saw for the sum of two squares. Hence it suffices to show that every prime is the sum of four squares, and we can show any product of primes is the sum of four squares using the above identity. Now $2 = 1^2 + 1^2 + 0^2 + 0^2$ so we focus on odd primes p: We know that there exist non-zero integers a, b, c, d such that $a^2+b^2+c^2+d^2 \equiv 0 \pmod p$; select them so that m is minimal, where $mp = a^2 + b^2 + c^2 + d^2$. Our goal is to show that $m = 1$.

`Exercise F5.3.` Prove that $|a|, |b|, |c|, |d| < p/2$, so that $m < p$.

`Exercise F5.4.` Show that m is odd: Show that if m is even then we can reorder a, b, c, d so that $a - b$ and $c - d$ are both even. But then $\frac{m}{2} \cdot p = (\frac{a-b}{2})^2 + (\frac{a+b}{2})^2 + (\frac{c-d}{2})^2 + (\frac{c+d}{2})^2$ contradicting the minimality of m.

Let u, v, w, x be the least residues, in absolute value, of $a, b, c, d \pmod m$, respectively. Therefore $u^2+v^2+w^2+x^2 \equiv a^2+b^2+c^2+d^2 \equiv 0 \pmod m$. Moreover $|u|, |v|, |w|, |x| < m/2$ (since m is odd), and so $u^2 + v^2 + w^2 + x^2 < 4(m/2)^2 = m^2$. Hence we can write $u^2 + v^2 + w^2 + x^2 = mn$ for some integer $n < m$.

`Exercise F5.5.` Prove that $au+bv+cw+dx \equiv av-bu-cx+dw \equiv aw+bx-cu-dv \equiv ax-bw+cv-du \equiv 0 \pmod m$.

Let $A = \frac{au+bv+cw+dx}{m}$, $B = \frac{av-bu-cx+dw}{m}$, $C = \frac{aw+bx-cu-dv}{m}$, $D = \frac{ax-bw+cv-du}{m}$, which are integers by the last exercise, and so

$$A^2 + B^2 + C^2 + D^2 = \frac{(a^2+b^2+c^2+d^2)}{m} \cdot \frac{(u^2+v^2+w^2+x^2)}{m} = np.$$

This contradicts the minimality of m unless $n = 0$ in which case $u = v = w = x = 0$ so that $a \equiv b \equiv c \equiv d \equiv 0 \pmod{m}$ and so $m^2 | a^2 + b^2 + c^2 + d^2 = mp$. Therefore $m = 1$ as $m < p$, which is what we wished to prove.

`Exercise F5.6.` Use the same approach to prove that every prime $\equiv 1 \pmod 4$ is the sum of two squares.

Quaternions. Just as we saw that a proof of which primes are the sum of two squares can be rephrased in terms of elements of $\mathbb{Z}[i]$, we can make a similar transition for our proof of the sum of four squares, but now in terms of the mysterious quaternions: Here we have three different special elements i, j, k for which $i^2 = j^2 = k^2 = -1$, but which do not commute:

$$ij = -ji = k, \quad jk = -kj = i, \quad ki = -ik = j,$$

so that $(a+bi+cj+dk)(a-bi-cj-dk) = a^2+b^2+c^2+d^2$, when we multiply a quaternion by its conjugate. The key observation is that

$$\begin{aligned}(a-bi-cj-dk)(u+vi+wj+xk) = (au+bv+cw+dx) + (av-bu-cx+dw)i \\ + (aw+bx-cu-dv)j + (ax-bw+cv-du)k\end{aligned}$$

which allows us to recover (F5.1) when we multiply each side by its (quaternionic) conjugate.

`Exercise F5.7.` Rewrite the above proof, that every prime is the sum of four squares, in terms of quaternions.

Universality of quadratic forms. Once one knows that every positive integer can be represented by the sum of four squares, but not as the sum of three squares, one might ask for further positive definite quadratic forms with this property.

It turns out that no quadratic or ternary quadratic form can represent all integers.

In 1916 Ramanujan asserted that the quaternary quadratic forms with the following coefficients represent all integers: $\{1,1,1,k\}$, $\{1,2,2,k\}, : 1 \le k \le 7$; $\{1,1,2,k\}$, $\{1,2,4,k\} : 1 \le k \le 14$; $\{1,1,3,k\} : 1 \le k \le 6$; $\{1,2,3,k\}, \{1,2,5,k\} : 1 \le k \le 10$; though this is not quite true for $\{1,2,5,5\}$ since it represent every positive integer except 15. We deduce

The Fifteen criterion, I. *Suppose that f is a positive definite diagonal quadratic form. Then f represents all positive integers if and only if f represents all positive integers ≤ 15.*

Proof. Suppose that $f = a_1x_1^2 + a_2x_2^2 + \ldots + a_dx_d^2$, with $1 \le a_1 \le a_2 \le \ldots \le a_d$ represents all positive integers. Since f represents 1 we must have $a_1 = 1$. Since f represents 2 we must have $a_2 = 1$ or 2. If $a_1 = a_2 = 1$ then, since f represents 3 we must have $a_3 = 1, 2$

or 3. If $a_1 = 1, a_2 = 2$ then, since f represents 5 we must have $a_3 = 2, 3, 4$ or 5. Now

$$\begin{aligned}
x_1^2 + x_2^2 + x_3^2 &\text{ represents } m,\ 1 \le m \le 6, \text{but not } 7,\ \text{and so } 1 \le a_4 \le 7;\\
x_1^2 + x_2^2 + 2x_3^2 &\text{ represents } m,\ 1 \le m \le 13, \text{but not } 14,\ \text{and so } 1 \le a_4 \le 14;\\
x_1^2 + x_2^2 + 3x_3^2 &\text{ represents } m,\ 1 \le m \le 5, \text{but not } 6,\ \text{and so } 1 \le a_4 \le 6;\\
x_1^2 + 2x_2^2 + 2x_3^2 &\text{ represents } m,\ 1 \le m \le 6, \text{but not } 7,\ \text{and so } 1 \le a_4 \le 7;\\
x_1^2 + 2x_2^2 + 3x_3^2 &\text{ represents } m,\ 1 \le m \le 9, \text{but not } 10,\ \text{and so } 1 \le a_4 \le 10;\\
x_1^2 + 2x_2^2 + 4x_3^2 &\text{ represents } m,\ 1 \le m \le 13, \text{but not } 14,\ \text{and so } 1 \le a_4 \le 14;\\
x_1^2 + 2x_2^2 + 5x_3^2 &\text{ represents } m,\ 1 \le m \le 9, \text{but not } 10,\ \text{and so } 1 \le a_4 \le 10.
\end{aligned}$$

Ramanujan's result implies that $a_1x_1^2 + a_2x_2^2 + a_3x_3^2 + a_4x_4^2$ then represents every positive integer except perhaps 15, and the result follows.

We could look to represent only $1, 2, 3, 5, 6, 7, 10, 14, 15$ rather than all integers ≤ 15.

By the 1940s researchers had come up with complicated criteria to decide whether a quadratic form represented all integers, but it took the genius of John Conway to come up with the following simply checked criterion:

The Fifteen criterion, II. *Suppose that f is a positive definite quadratic form, which is diagonal mod 2. Then f represents all positive integers if and only if f represents all positive integers ≤ 15.*

Notice that this is sharp since $x^2 + 2y^2 + 5z^2 + 5w^2$ represents every positive integer other than 15.

This was extended to all quadratic forms by Bhargava and Hanke:

The 290 criterion. *Suppose that f is a positive definite quadratic form. Then f represents all positive integers if and only if f represents all positive integers ≤ 290.*

Notice that this is sharp since $x^2 + xy + 2y^2 + xz + 4z^2 + 29(a^2 + ab + b^2)$ represents every positive integer other than 290.

The number of representations. In 1834 Jacobi showed that there are $8\sigma(n)$ representations of n as a sum of four squares if n is odd, and $24\sigma(m)$ representations if $n = 2^k m,\ k \ge 1$ is even.

`Exercise F5.8.` Prove that this can re-written as follows:

$$\left(\sum_{n\in\mathbb{Z}} x^{n^2}\right)^4 = 8\sum_{\substack{d\ge 1\\ 4\nmid d}} \frac{dx^d}{1-x^d}\ .$$

Representation by positive definite quadratic forms. Let us suppose that f is any positive definite quadratic form in three or more variables. One might ask which integers can be represented by f. In all of the examples we have seen f represents all integers, or (like $x^2 + y^2 + z^2$) all integers in certain residue classes. Is this true in general? We

saw even in the quadratic form case that if an integer satisfies certain obvious congruence conditions that it is represented by some form of the given discriminant, and this result easily generalizes; however we are interesting in representation by a specific form.

In 1929 Tratowsky showed that for any positive definite quadratic form f of discriminant $D > 0$ in five or more variables, if n is sufficiently large then n is represented by f if and only if n is represented by $f \pmod D$.

For three or four variables it usually makes sense (and is easier) to restrict our attention to the representation of squarefree integers n. In 1926 Kloosterman introduced an analytic method which implies that if f of discriminant $D > 0$ has four variables, and if n is a sufficiently large squarefree integer then n is represented by f if and only if n is represented by $f \pmod D$. It was only in 1990, that Duke and Schulze-Pillot extended this to positive definite quadratic forms f of discriminant $D > 0$ in three variables: If n is a sufficiently large squarefree integer then n is represented by f if and only if n is represented by $f \pmod{D^2}$; or more explicitly:

Theorem F5.1. *There exists an absolute constant $c > 0$ such that if n is a squarefree integer with $n > cD^{337}$ then n can be represented by f if and only if there exist integers a, b, c such that $f(a, b, c) \equiv n \pmod{8D^3}$.*

Waring's problem. We have seen that every positive integer is the sum of four squares. For $n \equiv a \pmod 6$ with $-2 \le a \le 3$ we use the identity

$$n = (x+1)^3 + (x-1)^3 + (-x)^3 + (-x)^3 + a^3 = 6x + a^3$$

to note that every integer is the sum of five cubes. This is rather too easy so let us insist on cubes of non-negative integers.

More generally we can ask whether, for each $k \ge 2$, there exists an integer $g(k)$ such that every integer is the sum of $g(k)$ kth powers of non-negative integers. Hilbert showed that $g(k)$ exists, and we have seen that $g(2) = 4$. Indeed $g(3) = 9$, $g(4) = 19$, $g(5) = 37, g(6) = 73\ldots$ Actually $g(k)$ grows fast, the reason being the large number of kth powers needed to represent small integers. Some candidates for the worst are $2^k - 1$ which needs $2^k - 1$ times 1^k, then $2^k[(3/2)^k] - 1$ which requires $[(3/2)^k] - 1$ times 2^k, plus $2^k - 1$ times 1^k. There is similarly a candidate a little smaller than 4^k. Euler's son thus conjectured that $g(k) = 2^k + [(3/2)^k] - 2$ which is true if $2^k(3/2)^k + [(3/2)^k] \le 2^k$ which holds for all $k < 10^8$ and is probably always true.

It turns out to be a good idea to ignore these small exceptional values. Hence we define $G(k)$ to be the smallest integer such that every sufficiently large integer is the sum of $G(k)$ kth powers of non-negative integers. Evidently $G(2) = 4$ and $G(k)$ is $\le g(k)$, usually far smaller than $g(k)$. We know that $4 \le G(3) \le 7$, and in general $G(k) \le k\log k + k\log\log k + Ck$ for some constant C.

Taxicab numbers and other diagonal surfaces. When Ramanujan lay ill from pneumonia in an English hospital he was visited by G.H. Hardy, his friend and co-author. Struggling for conversation, Hardy remarked that the number, 1729, of the taxicab he had ridden from the train station to the hospital was extremely dull. Ramanujan contradicted him noting that it is the smallest number which is the sum of two cubes in two different ways:

$$1^3 + 12^3 = 9^3 + 10^3 = 1729.$$

(Ramanujan might also have mentioned that it is the third smallest Carmichael number!). There are many other such identities; indeed Euler showed that all solutions to

$$a^3 + b^3 = c^3 + d^3$$

can be obtained by scaling

$$a = r^4 + (p - 3q)(p^2 + 3q^2)r, \;\; b = (p + 3q)r^3 + (p^2 + 3q^2)^2,$$
$$c = r^4 + (p + 3q)(p^2 + 3q^2)r, \;\; d = (p - 3q)r^3 + (p^2 + 3q^2)^2.$$

How about $a^4+b^4+c^4 = d^4$? Euler conjectured that there are no non-trivial solutions, but in 1986 Elkies showed that there are infinitely many, the smallest of which is

$$95800^4 + 217519^4 + 414560^4 = 422481^4.$$

(It was rather lucky that this is just large enough to have avoided direct computer searches to that time, since Elkies was inspired to give his beautiful solution to this problem). Euler had even conjectured that there is no non-trivial solution to the sum of $n - 1$ powers equalling an nth power, but that had already been disproved via the example

$$27^5 + 84^5 + 110^5 + 133^5 = 144^5.$$

G. Combinatorial number theory

G1. Partitions. Let $p(n)$ denote the number of ways of partitioning n into smaller integers. For example $p(7) = 15$ since

$$\begin{aligned} 7 &= 6+1 = 5+2 = 5+1+1 = 4+3 = 4+2+1 = 4+1+1+1 = 3+3+1 \\ &= 3+2+2 = 3+2+1+1 = 3+1+1+1+1 = 2+2+2+1 \\ &= 2+2+1+1+1 = 2+1+1+1+1+1 = 1+1+1+1+1+1+1. \end{aligned}$$

Euler observed that there is a beautiful generating function for $p(n)$: In the generating function $p(n)$ is the coefficient of t^n, and for each partition $n = a_1 + \ldots + a_k$ we can think of t^n as $t^{a_1} \ldots t^{a_k}$. Splitting this product up into the values of the a_i, but taking $(t^a)^j$ if there are j of the a_i's that equal a, we see that

$$\sum_{n\geq 0} p(n)t^n = \prod_{a\geq 1}\left(\sum_{j\geq 0}(t^a)^j\right) = \frac{1}{(1-t)(1-t^2)(1-t^3)\ldots}.$$

Similarly the generating function for the number of partitions into odd parts is $1/(1-t)(1-t^3)(1-t^5)\ldots$, for the number of partitions with no repeated parts is $(1+t)(1+t^2)(1+t^3)\ldots$, etc.

`Exercise G1.1.` Deduce that the number of partitions of n into odd parts is equals to the number of partitions of n with no repeated parts.

Partitions can be represented by rows and columns of dots in a *Ferrers diagram*; for example $27 = 11 + 7 + 3 + 3 + 2 + 1$ is represented by

o o o o o o o o o o o
o o o o o o o
o o o
o o o
o o
o

the first row having 11 dots, the second 7, etc. Now, reading the diagram in the other direction yields the partition $27 = 6+5+4+2+2+2+1+1+1+1$. This bijection between partitions is at the heart of many beautiful theorems about partitions. For example if a partition has m parts then its "conjugate" has largest part m. Using generating functions, we therefore find that the number of partitions with $\leq m$ parts, equals the number of partitions with all parts $\leq m$, which has generating function

$$\frac{1}{(1-t)(1-t^2)(1-t^3)\ldots(1-t^m)}.$$

Using the Ferrers diagram the partitions come in pairs, other than those that are *self-conjugate*; that is the conjugate partition is the same as the original partition. This

implies a symmetry about the diagonal axis of the diagram. Hence a self-conjugate Ferrers diagram looks like

o o o o o o	1 1 1 1 1 1
o o o o o	1 2 2 2 2
o o o	1 2 3
o o	1 2
o o	1 2
o	1

yielding $19 = 6+5+3+2+2+1$. We have constructed another partition of 19, using the same Ferrers diagram. The first element is obtained by peeling off the top row and top column (which have a total of 11 entries), then what's left of the second row and second column (7 remaining entries), and finally the single element left (1 entry), yielding the partition $19 = 11 + 7 + 1$.

`Exercise G1.2.` Prove that there is a bijection between self-conjugate partitions and partitions where all the entries are odd and distinct. Give an elegant form for the generating function for the number of self-conjugate partitions.

The sequence $p(n)$ begins $p(1) = 1$, 2, 3, 5, 7, 11, 15, 22, 30, $p(10) = 42$, 56, 77, 101, 135, 176, 231, 297, 385, 490, $p(20) = 627, \dots$ with $p(100) = 190,569,292$ and $p(1000) \approx 2.4 \times 10^{31}$. Ramanujan was intrigued by these numbers, both their growth (which seems to get quite fast) and their congruence conditions. For $n = 1000$ we see that there are roughly $10^{\sqrt{n}}$ partitions, which is an unusual function in mathematics. Hardy and Ramanujan proved the extraordinary asymptotic

$$p(n) \sim \frac{1}{4n\sqrt{3}}\, e^{\pi\sqrt{2n/3}}, \tag{G1.1}$$

and Rademacher developed their idea to give an exact formula. This is also too difficult for this book, but we will discuss one or two of the main ideas a little later, and see how this one proof gave birth to the circle method, still one of the most important techniques in number theory.

Counting partitions and the circle method. The Dedekind eta function is defined on the *upper half-plane*, that is $\mathcal{H} := \{\tau \in \mathbb{C} : \ \mathrm{Im}(\tau) > 0\}$. For any such $\tau \in \mathcal{H}$, we let $q = e^{2i\pi\tau}$, and define the eta function by

$$\eta(\tau) := q^{\frac{1}{24}} \prod_{n=1}^{\infty} (1 - q^n).$$

Hence the generating function for the $p(n)$, can be written

$$P(q) := \sum_{m \geq 0} p(m) q^m = q^{1/24}/\eta(\tau).$$

The surprise is that η satisfies the equations

$$\eta(\tau+1) = e^{\frac{\pi \mathrm{i}}{12}}\ \eta(\tau) \qquad \text{and} \quad \eta(-1/\tau) = \sqrt{-\mathrm{i}\tau}\ \eta(\tau).$$

It is, incidentally, a bit more natural to define $\Delta(\tau) := q\prod_{n=1}^{\infty}(1-q^n)^{24} = \eta(\tau)^{24}$ so that $\Delta(\tau+1) = \Delta(\tau)$ and $\Delta(-1/\tau) = \tau^{12}\Delta(\tau)$. Now we saw that the maps $\tau \to \tau+1$ and $\tau \to -1/\tau$ generate all of $\mathrm{SL}(2,\mathbb{Z})$ and so for any integers a,b,c,d with $ad-bc=1$ we have the remarkable identity

$$\Delta\left(\frac{a\tau+b}{c\tau+d}\right) = (c\tau+d)^{12}\Delta(\tau).$$

Now, suppose τ is a near to a rational number, say $\tau = -d/c + i\delta$ (taking 1 in place of 0). Then the above yields that $\Delta(-d/c+i\delta) = (ci\delta)^{-12}\Delta\left(a/c + i/c^2\delta\right)$; here we can take a to the inverse of d mod c. Now if N is very large and $\tau \approx iN$ then $\Delta(iN) \approx q$. Therefore $\Delta(-d/c+i\delta) = (ci\delta)^{-12}e^{-2\pi/c^2\delta}e^{2i\pi a/c}$. This implies that $\eta(-d/c+i\delta) = \xi(c\delta)^{-1/2}e^{-\pi/12c^2\delta}$ for some $\xi = \xi_{d/c}$ with $|\xi| = 1$.

We use the fact that if k is an integer then

$$\int_0^1 e^{2i\pi kt}dt \ = \ \begin{cases} 1 & \text{if } k=0 \\ 0; & \text{otherwise,} \end{cases}$$

so that if $r = e^{-2\pi\rho}$ is < 1 then

$$\begin{aligned} p(n) \ &= \ \sum_{m\geq 0} p(m)r^{m-n}\int_0^1 e^{2i\pi(m-n)t}dt \ = \ \int_0^1 P(e^{2i\pi(t+i\rho)})\ e^{-2i\pi n(t+i\rho)}dt \\ &= \int_{\substack{0\leq t\leq 1 \\ \tau = t+i\rho}} \frac{e^{-2i\pi(n-\frac{1}{24})\tau}}{\eta(\tau)}\ dt \end{aligned}$$

We will let $\rho \to 0$ so that q moves towards the unit circle. The zeros of $\eta(\tau)$ will evidently give rise to poles of the integrand, and they occur at $e^{2i\pi a/b}$ for all rationals a/b, of order inversely proportional to b. Hence the largest will be at $0/1$, and this is confirmed by our formula above. Substituting in to the integrand with $\tau = t+i\rho = i\delta$, the integrand becomes a constant of absolute value 1 times

$$(\rho - it)^{1/2}e^{\pi/12(\rho-it)}e^{2\pi(n-\frac{1}{24})\rho}e^{-2i\pi(n-\frac{1}{24})t}$$

This will not vary much as long as t is in an interval of length $\asymp 1/n$. We chose $\rho = 1/\sqrt{24n-1}$. Hence the integrand is about

$$\rho^{1/2}e^{\pi/6\rho} \approx e^{\pi\sqrt{2n/3}}.$$

Getting the precise estimate (G1.1) is considerably more complicated.

Partition congruences. Ramanujan also noted several congruences for the $p(n)$:

$$p(5k+4) \equiv 0 \pmod 5, \quad p(7k+5) \equiv 0 \pmod 7, \quad p(11k+6) \equiv 0 \pmod{11},$$

for all k. Notice that these are all of the form $p(n) \equiv 0 \pmod q$ whenever $q|24n-1$, and these seem to be the only such congruences. However Ono has found many more such congruences, only a little more complicated: For any prime $q \geq 5$ there exist primes ℓ such that $p(n) \equiv 0 \pmod q$ whenever $q\ell^3|24n-1$.

More identities. There are many beautiful identities involving the power series from partitions. One of the most extraordinary is Jacobi's powerful *triple product identity*: If $|x|<1$ then

$$\prod_{n\geq 1}(1-x^{2n})(1+x^{2n-1}z)(1+x^{2n-1}z^{-1}) = \sum_{m\in\mathbb{Z}} x^{m^2}z^m.$$

We shall determine some useful consequences of it:

Letting $x=t^a,\ z=t^b$ and $n=k+1$ in Jacobi's triple product identity we obtain

$$\prod_{k\geq 0}(1-t^{2ak+2a})(1+t^{2ak+a+b})(1+t^{2ak+a-b}) = \sum_{m\in\mathbb{Z}} t^{am^2+bm}.$$

Some special cases include $a=1,\ b=0$, yielding

$$\prod_{n\geq 1}(1-t^{2n})(1+t^{2n-1})^2 = \sum_{m\in\mathbb{Z}} t^{m^2};$$

and $a=b=\pm\frac{1}{2}$, yielding

$$\prod_{k\geq 0}(1+t^k)(1-t^{2k+2}) = \sum_{m\in\mathbb{Z}} t^{\frac{m^2+m}{2}}.$$

`Exercise G1.3.` By writing $1+t^k$ as $(1-t^{2k})/(1-t^k)$ or otherwise, deduce that

$$\sum_{m\geq 1} t^{\frac{m^2+m}{2}} = \frac{(1-t^2)(1-t^4)(1-t^6)\dots}{(1-t)(1-t^3)(1-t^5)\dots}$$

Letting $x=t^a,\ z=-t^b$ and $n=k+1$ in Jacobi's triple product identity we obtain

$$\prod_{k\geq 0}(1-t^{2ak+2a})(1-t^{2ak+a+b})(1-t^{2ak+a-b}) = \sum_{m\in\mathbb{Z}} (-1)^m t^{am^2+bm}.$$

Some special cases include $a=1,\ b=0$, yielding

$$\prod_{n\geq 1}(1-t^{n})(1-t^{2n-1}) = \sum_{m\in\mathbb{Z}} (-1)^m t^{m^2};$$

and $a = \frac{3}{2},\ b = \frac{1}{2}$, yielding Euler's identity,

$$\prod_{n\geq 1}(1-t^n) = \sum_{m\in\mathbb{Z}}(-1)^m t^{\frac{3m^2+m}{2}}.$$

Exercise G1.4. Interpret this combinatorially, in terms of the number of partitions of m into unequal parts.

Exercise G1.5. If $(12/.)$ is the Jacobi symbol, show that

$$t^{1/24}\prod_{n\geq 1}(1-t^n) = \sum_{m\geq 1}\left(\frac{12}{m}\right)\ t^{\frac{m^2}{24}}\ .$$

Letting $x = t^4$ and $z = -1$ in Jacobi's triple product identity we obtain

$$\prod_{n\geq 1}(1-t^{8n})(1-t^{8n-4})^2 = \sum_{\substack{b\in\mathbb{Z}\\ b \text{ even}}}(-1)^{b/2}t^{b^2}.$$

Now letting $x = t^4$ and $z = -t^4u$ in Jacobi's triple product identity we obtain

$$\prod_{n\geq 1}(1-t^{8n})(1-\alpha t^{8n}u)(1-t^{8n-8}/u) = \sum_{m\in\mathbb{Z}} t^{4m^2+4m}(-u)^m.$$

In the product we change the exponent of the last term taking n in place of $n-1$, and hence we have a left over term of $1-1/u$. On the right side we pair up the term for m with the term for $-m-1$. Dividing through by $1-1/u$ then yields

$$\prod_{n\geq 1}(1-t^{8n})(1-t^{8n}u)(1-t^{8n}/u) = \sum_{m\geq 0} t^{4m^2+4m+1}(-1)^m(u^m + u^{m-1} + \ldots + u^{-m}).$$

Taking $u = 1$ we obtain

$$t\prod_{n\geq 1}(1-t^{8n})^3 = \sum_{\substack{a\geq 1\\ a \text{ odd}}}(-1)^{\frac{a-1}{2}}at^{a^2}.$$

Multiplying these together yields

$$t\prod_{n\geq 1}(1-t^{4n})^2(1-t^{8n})^2 = \sum_{\substack{a,b\in\mathbb{Z},\ a\geq 1,\\ a \text{ odd},\ b \text{ even}}}(-1)^{\frac{a+b-1}{2}}at^{a^2+b^2} = \sum_{n\geq 1}a_nt^n, \tag{G1.2}$$

for certain integers a_n. In particular if p is a prime $\equiv 3 \pmod 4$ then $a_p = 0$; if p is a prime $\equiv 3 \pmod 4$ then writing $p = a^2 + b^2$ with a odd, we have $a_p = 2(-1)^{\frac{a+b-1}{2}}a$.

G2. The Freiman-Ruzsa Theorem. For finite sets of integers A, B subsets of an additive group Z, we define $A+B$ to be the *sumset* $\{a+b : a \in A, b \in B\}$.

It is easy to see that $|A+A| \le |A|(|A|+1)/2$, since the distinct elements of $A+A$ are a subset of $\{a_i + a_j : \ 1 \le i \le j \le |A|\}$

`Exercise G2.1.` Give an example of a set A with n elements where $|A+A| = n(n+1)/2$. (Hint: You could try $A = \{1, 2, 2^2, 2^3, \dots, 2^{n-1}\}$ or be more adventurous.)

Typically $|A+A|$ is large and is only small in very special circumstances:

Lemma G2.1. *If A and B are finite subsets of $\mathbb{Z}$ then $|A+B| \ge |A| + |B| - 1$. Equality holds if and only if A and B are each complete finite segments of an arithmetic progression to the same modulus (that is $A = \{a, a+d, a+2d, \dots, a+(r-1)d\}$ and $B = \{b, b+d, b+2d, \dots, b+(s-1)d\}$ for some a, b, r, s and $d \ge 1$).*

Proof. Write the elements of A as $a_1 < a_2 < \dots a_r$, and those of B as $b_1 < b_2 < \dots b_s$. Then $A+B$ contains the $r+s-1$ distinct elements

$$a_1+b_1 < a_1+b_2 < a_1+b_3 < \cdots < a_1+b_s < a_2+b_s < a_3+b_s < \cdots < a_r+b_s.$$

If it contains exactly $r+s-1$ elements then these must be the same, in the same order, as $a_1+b_1 < a_2+b_1 < a_2+b_2 < a_2+b_3 < \cdots < a_2+b_s < a_3+b_s < \cdots < a_r+b_s$. Comparing terms, we have $a_1 + b_{i+1} = a_2 + b_i$ for $1 \le i \le s-1$; that is $b_j = b_1 + (j-1)d$ where $d = a_2 - a_1$. A similar argument with the roles of a and b swapped, reveals our result.

If $A+B$ is small, not as small as $|A| + |B| - 1$ but not much bigger, then we might expect to be able to use a similar proof to prove a similar structure theorem. However the combinatorics of comparing different sums quickly becomes very complicated.

Looking for other examples in which $A+B$ is small, one soon finds the possibility that A and B are both large subsets of complete finite segments of an arithmetic progression to the same modulus. For example, if A contains $2m$ integers from $\{1, 2, 3, \dots, 3m\}$ then $A+A$ is a subset of $\{2, 2, 3, \dots, 6m\}$, and so $|A+A| < 3|A|$. One can find a criterion similar to Lemma G2.1: If $|A| \ge |B|$ and $|A+B| \le |A| + 2|B| - 4$ then A and B are each subsets of arithmetic progressions with the same common difference, of lengths $\le |A+B| - |B| + 1$ and $\le |A+B| - |A| + 1$, respectively.

A further interesting example is given by

$$A = B = \{1, 2, \dots, 10, 101, 102, \dots 110, 201, 202, \dots 210\},$$

or its large subsets. This can be written as $1 + \{0, 1, 2, \dots, 9\} + \{0, 100, 200\}$, a translate of the sum of complete finite segments of two arithmetic progressions. More generally, define a *generalized arithmetic progression* $C = C(a_0, \dots a_k; N_1, \dots, N_k)$ as

$$C := \{a_0 + a_1 n_1 + a_2 n_2 + \cdots + a_k n_k : \ 0 \le n_j \le N_j - 1 \text{ for } 1 \le j \le k\}$$

where $a_0, a_1, \dots a_k$ are given integers, and $N_1, N_2, \dots N_k$ are given positive integers. Note that $C(a_0, a_1, \dots a_k; N_1, N_2, \dots, N_k) = a_0 + \sum_{i=1}^{k} a_i \cdot \{0, 1, \dots, N_i - 1\}$. This generalized arithmetic progression is said to have *dimension* k and *volume* $N_1 N_2 \dots, N_k$. Notice that

$$2C(a_0, a_1, \dots a_k; N_1, N_2, \dots, N_k) = C(2a_0, a_1, \dots a_k; 2N_1 - 1, 2N_2 - 1, \dots, 2N_k - 1).$$

so that $|2C| < 2^k|C|$. We think of C as an image of the k-dimensional lattice segment

$$\{0, 1, \dots, N_1 - 1\} \times \{0, 1, \dots, N_2 - 1\} \times \dots \times \{0, 1, \dots, N_k - 1\}.$$

Indeed the inequality $|2C| < 2^k|C|$ generalizes to the "image" C in $\mathbb{Z}$ of any part of a lattice inside a convex, compact region of $\mathbb{R}^k$.

We can combine our two ideas so that if A is a large subset of a generalized arithmetic progression then $|A + A| < \kappa|A|$ for some smallish constant κ.

Are there any other examples of sets A and B for which $A + B$ is small? Freiman showed that the answer is "no", having the extraordinary insight to suggest and prove that $A+A$ can be "small" if and only if it is a "large" subset of a "low" dimensional generalized arithmetic progression of "not too big" volume.[24]

We have seen that if $A + A$ is small then A has a lot of additive structure, that is, it is a subset of a generalized arithmetic progression. In the prototypical case A is the set of integers $\{1, 2, \dots, N\}$. In that case, we have seen that $|A \cdot A| < \epsilon N^2$ (the multiplication table theorem) but it is not difficult to see, by taking the products of pairs of primes $\leq N$, that $|A \cdot A| \geq \pi(N)^2/2 > N^2/3(\log N)^2$, so that $A \cdot A$ is not much smaller than N^2. One might guess that this happens whenever $A + A$ is small.

`Exercise G2.2.` Explain the bijection between $A \cdot A$ and $\log A + \log A$.

If $A+A$ is small then A has a lot of additive structure by the Freiman-Ruzsa Theorem. If $A+A$ is small then $\log A$ has a lot of additive structure by the Freiman-Ruzsa Theorem; that is A has a lot of multiplicative structure. Can a set have both types of structure at once? Erdős and Szemerédi conjectured that this is impossible, predicting the *sum-product inequality*

$$\max\{|A + A|,\ |A \cdot A|\} \geq c_\epsilon |A|^{2-\epsilon}$$

for some constant $c_\epsilon > 0$ for any $\epsilon > 0$. More daringly one might guess, from the same reasoning that *either* $|A + B| \geq c_\epsilon(|A||B|)^{1-\epsilon}$ *or* $|A \cdot C| \geq c_\epsilon(|A||C|)^{1-\epsilon}$ for any finite sets of integers A, B, C. The best result in this area was given by Solymosi in 2009, we showed that if A and B are two finite sets of real numbers with $|A| \geq |B| > 1$ then

$$|AB||A + A||B + B| \geq c\, \frac{(|A||B|)^2}{\log |B|},$$

for some constant $c > 0$.

`Exercise G2.3.` Deduce the sum-product inequality for any $\epsilon > 2/3$.

More additive number theory. Given a largish subset of the integers up to N one can ask whether it contains certain simple structures, simply because of its large size. For example are there necessarily two consecutive elements of our set? Are there necessarily two elements of the set that add to a third? Are there three different elements of the set in arithmetic progression? Rather than quantify "largish" one might instead partition the

[24]Freiman's 1962 proof is both long and difficult to understand. Ruzsa's 1994 proof of Freiman's result, is extraordinarily elegant and insightful, and heralded an explosion of ideas in this area.

integers into two (or more) sets and ask whether either of them have the given structure. This is a familiar theme from combinatorics, and ideas from that subject will allow us to give a first answer to these questions.

We begin with a well-known result from graph theory:

Lemma G2.2. *If the edges of the complete graph with N vertices are coloured with r colours, with $N \geq N(r)$ then there is a monochromatic triangle.*

Proof. By induction. Evidently $N(1) = 3$. For larger r consider the edges attached to any one vertex v. If $N \geq r(N(r-1)-1)+2$ then there must be some colour c for which there are $\geq N(r-1)$ edges adjacent to v of colour c. Let H be those vertices that share an edge of colour c with v. If there are any two vertices in H that are attached by an edge of colour c, then these two vertices along with v form a monochromatic triangle. Otherwise the edges of H are coloured by just $r-1$ colours and the result follows by induction.

`Exercise G2.4.` Justify that if $N \geq r(N(r-1)-1)+2$ then there must be some colour c for which there are $\geq N(r-1)$ edges adjacent to v of colour c. Show that we may take $N(r) \leq 3\ r$!

This is a typical Ramsey theory proof in that the proof is really just a greedy algorithm, and leads to a bound that is probably far too big. Indeed there are questions in the subject in which the bound cannot be discussed using *primitive recursive functions*.

There is a quite beautiful corollary:

Schur's Theorem. *If the integers up to N are coloured with r colours, with $N \geq N(r)$ then there is a monochromatic solution to $x+y=z$ in positive integers $x, y, z \leq N$*

Proof. We construct the complete graph on N vertices, labeling the vertices $1, 2, \ldots, N$, and joining vertices i and j by the colour of $|j-i|$. Lemma 5.1 tells us that there is a monochromatic triangle, say joining the vertices with labels $i < j < k$, and hence if $x = j-i,\ y = k-j$ and $z = k-i$, we know that these positive integers all have the same colour and indeed satisfy $x+y=z$.

In 1927 van der Waerden [20] answered a conjecture of Schur, by showing that if the positive integers are partitioned into two sets then one set must contain arbitrarily long arithmetic progressions.

van der Waerden's Theorem (1927)**.** *Fix integers $r \geq 2$ and $k \geq 3$. If we colour the integers with r colours then there is a monochromatic k-term arithmetic progression.*

`Exercise G2.5.` Prove van der Waerden's Theorem for $k = 3$ and $r = 2$.

`Exercise G2.6.` Partition the integers into two sets neither of which has an infinitely long arithmetic progression.

Szemerédi's Theorem (1974)**.** *For any $\delta > 0$ and integer $k \geq 3$, there exists an integer $N_{k,\delta}$ such that if $N \geq N_{k,\delta}$ and $A \subset \{1, 2, \ldots, N\}$ with $|A| \geq \delta N$ then A contains an arithmetic progression of length k.*

`Exercise G2.7.` Show that if $N \geq N_{k,\delta}$ and A is a subset of an arithmetic progression of length N, with $|A| \geq \delta N$, then A contains an arithmetic progression of length k .

`Exercise G2.8.` Deduce van der Waerden's Theorem from Szemerédi's Theorem.

The $k = 3$ case was first proved in a cunning proof by Roth in 1952 using Fourier analysis. In 1969 Szemerédi proved the $k = 4$ case by combinatorial methods and extended this to all k in 1974. In 1977 Furstenberg proved Szemerédi's Theorem in a very surprising manner, using ergodic theory. It was not until 19**, that Tim Gowers finally gave an analytic proof of Szemerédi's Theorem, the proof based on the overall plan of Roth, but involving a new kind of higher dimensional analysis (partly based on the Freiman-Ruzsa theorem). Gowers' proof was the starting point for

Green and Tao (2008)**.** *For any integer N one can find (infinitely many different) pairs of integers $a, d \geq 1$, such that $a, a + d, \dots, a + (N - 1)d$ are all primes.*

How much further can one develop Szemerédi's Theorem? Erdős conjectured that any set A of positive integers for which

$$\sum_{a \in A} \frac{1}{a} = \infty$$

must contain arbitrarily long arithmetic progressions, a question that is still very open today. Erdős stated this conjecture as a means to prove that there are arbitrarily long arithmetic progressions of primes (but this is not how Green and Tao proceeded). Can this even be proved in the $k = 3$ case?

How large is the largest subset $S(N)$ of $\{1, 2, \dots, N\}$ that has no three term arithmetic progression? If one could show that $|S(N)| < N/\log N$ then one would know that there are infinitely many three term arithmetic progressions of primes. Recently Tom Sanders has come agonizingly close to this goal by showing that $|S(N)| < c(\log \log N)^5 N/\log N$ for some constant $c > 0$. Is this close to the true size of $S(N)$? The best we know is the far smaller lower bound $S(N) > Ne^{-c\sqrt{\log N}}$ for some $c > 0$ given by a beautiful construction of Behrend:

`Exercise G2.9.` Show that a, b, c are in arithmetic progression if and only if $a + c = 2b$.

`Exercise G2.10.` Write $a = \sum_{i=1}^{k} a_i (2m)^{i-1} \in C := C(0, 1, 2m, (2m)^2, \dots (2m)^{k-1}; m, m, \dots, m)$, etc.

(1) Show that $a, b, c \in C$ are in arithmetic progression if and only if the vector $\mathbf{a}= (a_1, \dots, a_k)$, $\mathbf{b}$ and $\mathbf{c}$ are collinear.
(2) Let $C_r = \{\mathbf{a} \in C : \ |\mathbf{a}| = r\}$. Show no three distinct elements of C_r are collinear.

G3. Bouncing billiard balls and $n\alpha$ mod 1. In chapter 11, Dirichlet's Theorem stated that if α is a real, irrational number then for each $N \geq 1$ there exists a positive integer $n \leq N$ such that $0 < |n\alpha - m| < \frac{1}{N}$ for some integer m. In other words $n\alpha$ mod 1 gets arbitrarily close to 0. One might ask whether $n\alpha$ mod 1 gets arbitrarily close to any given $\theta \in [0, 1)$? Now let us suppose that $n\alpha - m = \delta$ where $0 < \delta < 1/N$ (an analgous argument works if $-1/N < \delta < 0$); and let k be the largest integer $\leq \theta/\delta$. Then $k \leq \theta/\delta < k + 1$ so that

$$0 \leq \theta - k\delta < \delta.$$

Now $k\delta = k(n\alpha - m) = kn\alpha - km$ so that $\{kn\alpha\} = \{k\delta\} = k\delta$ as $0 < k\delta \leq \theta < 1$. Hence

$$|\{kn\alpha\} - \theta| < \delta,$$

where $\delta < 1/N$ and $kn \leq \theta N/\delta < 1/\delta^2$. Hence we have proved:

Theorem G3.1. *If α is a real irrational number then the numbers $\{n\alpha\}$ are dense on $[0, 1)$.*

`Exercise G3.1.` Show that the conclusion of the Theorem is not true is α is rational.

Have you ever played billiards or pool? You play on a rectangular table, hitting your ball along the surface. The sides of the table are cushioned so that the ball bounces off the side at the opposite angle to which it hits. That is if it hits at α^o then it bounces of at $(180 - \alpha)^o$. Sometimes one miscues and the ball carries on around the table, coming to a stop without hitting another ball. Have you ever wondered what would happen if there was no friction, so that the ball never stops? Would your ball eventually hit the ball it is supposed to hit, no matter where that other ball is placed? Or could it go on bouncing for ever without ever getting to the other ball? We could rephrase this question more mathematically by supposing that we play on a table in the complex plane, with two sides along the $x-$ and $y-$ axes. Say the table length is ℓ, and width is w so that it is the square with corners at $(0, 0), (0, \ell), (w, 0), (w, \ell)$. Let us suppose that the ball is hit from the point (u, v) along a line with slope α. As the line continues on indefinitely inside the box, does it get arbitrarily close to every point inside the box?

`Exercise G3.2.` Show that by rescaling with the map $x \to x/\ell,\ y \to y/w$ we can assume, without any loss of generality, that the billiard table is the unit square.

The ball would run along the line $\mathcal{L} := \{(u + t, v + \alpha t),\ t \geq 0\}$ if it did not hit the sides of the table. Notice though that if after each time it hit a side we reflected the true trajectory through the line that represents that side, then indeed the ball's trajectory would be $\mathcal{L}$. Develop this to prove:

`Exercise G3.3.` Show that the billiard ball is at (x, y) after time t, where x and y are given as follows:
Let $m = [u + t]$. If m is even let $x = \{u + t\}$; if m is odd let $x = 1 - \{u + t\}$.
Let $n = [v + \alpha t]$. If n is even let $y = \{v + \alpha t\}$; if n is odd let $y = 1 - \{v + \alpha t\}$.

`Exercise G3.4.` Show that if α is rational then the ball eventually ends up exactly where it started from, and so it does not get arbitrarily close to every point on the table.

So how close does the trajectory get to the point (r, s), where $r, s \in [0, 1)$? Let us consider all of those values of t for which $x = r$, with m and n even (to simplify matters), and see if y is ever close to s.

`Exercise G3.5.` Show that $[z]$ is even if and only if $\{z/2\} \in [0, 1/2)$. Deduce that $[z]$ is even and $\{z\} = r$ if and only if $\{z/2\} = r/2$.

Hence we want that $(u+t)/2 = k+r/2$ for some integer k; that is $t = 2k+(r-u),\ k \in \mathbb{Z}$. In that case $v + \alpha t = 2\alpha k + \alpha(r-u) + v$ so we want $\{\alpha k + (\alpha(r-u)+v)/2\}$ close to $s/2$. That is $k\alpha$ mod 1 should be close to $\theta := \{\frac{(s-v)+\alpha(u-r)}{2}\}$. Now, in the Theorem above, we showed that the values $k\alpha$ mod 1 are dense in $[0, 1)$ when α is irrational, and so in particular there are values of k that allow $k\alpha$ mod 1 to be arbitrarily close to θ. Hence we have proved the difficult part of:

Corollary G3.2. *If α is a real irrational number then any ball moving at angle α (to the co-ordinate axes) will eventually get arbitrarily close to any point on a 1-by-1 billiards table.*

Weyl's criterion

G4. Transcendental numbers.

Give the countable vs. uncountable argument

Prove that π is transcendental or irrational?

`Exercise G4.1.` Let $\alpha := \sqrt{2}^{\sqrt{2}}$. We wish to show that there exist irrational numbers x, y such that x^y is rational. Use either α or $\alpha^{\sqrt{2}}$ to prove this.

H. Elliptic curves and beyond

H1. The group of rational points on elliptic curves. In section C11 we saw that the general form $y^2 + dxy + ey = x^3 + ax^2 + bx + c + k$ can be transformed to an equation of the affine form

$$E: \ y^2 = x^3 + ax + b$$

with $a, b \in \mathbb{Z}$ by linear maps with rational coefficients. This is called an *elliptic curve*. In section 6.1 we saw that two rational points on the unit circle gave rise to a line with rational coefficients and vice-versa; this allowed us to find all the rational points on the unit circle. We extend that idea to elliptic curves. Let $E(\mathbb{Q})$ denote all of the rational points on E (that is (x, y) on E with $x, y \in \mathbb{Q}$).

`Exercise H1.1.` Show that if $(x, y) \in E(\mathbb{Q})$ then there exist integers ℓ, m, n such that $x = m/n^2,\ y = \ell/n^3$ with $(\ell m, n) = 1$.

`Exercise H1.2.` Let $\Delta = 4a^3 + 27b^2$. Show that if $a > 0$ or if $\Delta > 0$ then $x^3 + ax + b = 0$ has just one real root. Show that if $a, \Delta < 0$ then $x^3 + ax + b = 0$ has three real roots. Sketch the shape of the curve $y^2 = x^3 + ax + b$ in the two cases.

Suppose that we are given two points $(x_1, y_1), (x_2, y_2) \in E(\mathbb{Q})$. The line between them, $y = mx + \nu$ has $m, \nu \in \mathbb{Q}$.[25] These two points are both intersections of the line $y = mx + \nu$ with the elliptic curve $y^2 = x^3 + ax + b$, that is x_1, x_2 satisfy

$$(mx + \nu)^2 = y^2 = x^3 + ax + b;$$

in other words x_1 and x_2 are two of the three roots of the cubic polynomial

$$x^3 - m^2x^2 + (a - 2m\nu)x + (b - \nu^2) = 0.$$

If the third root is x_3 then $x_3 = m^2 - x_1 - x_2 \in \mathbb{Q}$ and if we let $y_3 = mx_3 + \nu$ we obtain the third intersection of the line with E, and $(x_3, y_3) \in E(\mathbb{Q})$. This method of generating a third rational point from two given ones goes back to Fermat.

Actually one can do even better and generate a second point from a given one: If $(x_1, y_1) \in E(\mathbb{Q})$ let $y = mx + \nu$ be the equation of the tangent line to $y^2 = x^3 + ax + b$ at (x_1, y_1). To calculate this simply differentiate to obtain $2y_1m = 3x_1^2 + a$ and then $\nu = y_1 - mx_1$. Now our cubic polynomial has a double root at $x = x_1$ and we again compute a third point by taking $x_3 = m^2 - 2x_1,\ y_3 = mx_3 + \nu$ so that $(x_3, y_3) \in E(\mathbb{Q})$.

In these constructions we missed the case when the line is vertical (in the first case $x_1 = x_2$ which implies that $y_2 = -y_1$; in the second case $y = 0$). Where is the third point of intersection? One cannot see another point of intersection on the graph (that is on the real plane), but the line stretches to infinity, and indeed the third point is, rather surprisingly, the point at infinity, which we denote 0. Remember from section C11, in projective co-ordinates the elliptic curve is $y^2z = x^3 + axz^2 + bz^3$ so the point at infinity is $(0, 1, 0)$.

[25] Or is of the form $x = x_1 = x_2$, a situation we will deal with a little later.

`Exercise H1.3.` Prove that there cannot be four points of $E(\mathbb{Q})$ on the same line.

Poincaré made an extraordinary observation: If we take any three points P, Q, R of E on the same line then we can define a group by taking $P + Q + R = 0$. The line at infinity tells us that the point at infinity is indeed the 0 of this group. Moreover we have seen that $(x, y) + (x, -y) = 0$. Note that this implies that, in the notation above, $(x_3, -y_3) = (x_1, y_1) + (x_2, y_2)$

It is clear that the operation is closed under addition (and, most interestingly, closed in the subgroup $E(\mathbb{Q})$). The one thing that is complicated to justify is that Poincaré's operation (of addition) is indeed associative, and that hence we do indeed have a group.

`Exercise H1.4.` Show that the addition law given here is indeed associative

It is also obvious that the addition law is commutative. The question then becomes to identify the structure of the group of rational points, $E(\mathbb{Q})$.

Is $E(\mathbb{Q})$ finite or infinite? Suppose that we have a rational point P. Take the tangent, find the third point of intersection of the tangent line with E to obtain $-2P$, and then reflect in the x-axis to obtain $2P$. Fermat suggested that if we repeat this process over and over again, then we are unlikely to come back again to the same point. If we never return to the same point then we say that P has *infinite order*; otherwise P has finite order, the order being the minimum positive integer n for which $nP = 0$ (points of finite order are known as *torsion points*).

`Exercise H1.5.` Prove that the torsion points form a subgroup.

`Exercise H1.6.` Prove that if $P = (x, y)$ with $y \neq 0$ then $2P = (X, Y)$ where

$$X = \frac{(x^2 - a)^2 - 8bx}{4y^2}, \quad Y = \frac{x^6 + 5ax^4 + 20bx^3 - 5a^2x^2 - 4abx - 8b^2 - a^3}{8y^3}.$$

Notice that the numerator and denominator of X are polynomials of degree four in x implying (roughly) that the co-ordinates of $2P$ are about four times the length of the co-ordinates of P, unless there is an enormous amount of cancelation between numerator and denominator. To express this better it is convenient to define the *height* of P, $h(P) := \max\{|m|, n^2\}$. Our observation is that $h(2P) \approx h(P)^4$.

`Exercise H1.7.` Show that (x, y) has order 2 if and only if $y = 0$. Deduce that the number of points of order 1 or 2 is one plus the number of integer roots of $x^3 + ax + b$; and therefore equals $1, 2$ or 4.

Elizabeth Lutz and Nagell showed that there are only finitely many torsion points in $E(\mathbb{Q})$. By suitably manipulating the above formulae[26] they showed that if $x(P) = m/n^2$ with $(m, n) = 1$ and $n > 1$ then $x(kP) = M/N^2$ with $(M, N) = 1$ and N divisible by n. An immediate consequence of this is that if $P = (x, y)$ is a torsion point then x is an integer, and hence y is an integer.

Now if $P = (x, y)$ is a torsion point, that is $nP = 0$ then $n \cdot (2P) = 2 \cdot nP = 0$ so $2P$ is also a torsion point, and hence x, y, X, Y are all integers provided $P, 2P < ne0$. Now $m^2 = 2x + X \in \mathbb{Z}$ so that $m = (3x^2 + a)/2y \in \mathbb{Z}$; that is $2y$ divides $3x^2 + a$. Hence y divides $9(3b - 2ax)(x^3 + ax + b) + (6ax^2 - 9bx + 4a^2)(3x^2 + a) = \Delta := 4a^3 + 27b^2$. Since

[26]The proof is elementary but complicated.

there are only finitely many divisors y of $4a^3+27b^2$, and each such y gives rise to at most three values of x, hence there are only finitely many torsion points in $E(\mathbb{Q})$.

`Exercise H1.8.` Show that (x,y) is a torsion point that y^2 divides $b(4a^3+27b^2)$.

`Exercise H1.9.` Show that the cancelation between the numerator and denominator in the expression for $2P$ above is bounded by Δ^2. Deduce that there exists a constant c depending only on a and b such that if $h(P) > c\sqrt{|\Delta|}$ then P has infinite order.

Mazur improved this showing that the torsion subgroup of $E(\mathbb{Q})$ contains at most 16 points. In fact this subgroup is either $\mathbb{Z}/N\mathbb{Z}$ for some $1 \le N \le 10$ or $N = 12$, or it is $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2N\mathbb{Z}$ for some $1 \le N \le 4$.

A final word on torsion points: There can be torsion points in fields other than $\mathbb{Q}$. One can ask for them in $\mathbb{C}$; in fact their x-co-ordinates a roots of a polynomial with integer coefficients, so they are algebraic numbers. One can show that the subgroup of torsion points of order N is isomorphic to $\mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$ for each $N \ge 1$, so that there are N^2 points of order dividing N.

`Exercise H1.10.` Prove that there are exactly $N^2 \prod_{p|N}(1-\frac{1}{p^2})$ points of order N on $E(\mathbb{C})$.

Given that $E(\mathbb{Q})$ is abelian we can write it as $T \times \mathbb{Z}^r$. Here T is the torsion subgroup, and r can be a integer ≥ 0, or even infinity. A remarkable theorem of Mordell shows that r is always finite.[27] His proof proceeds by descent: Given a point P on the elliptic curve with large co-ordinates he shows how to find a point R from a finite set S such that $P - R = 2Q$ for some other point $Q \in E(\mathbb{Q})$. This means that the co-ordinates of Q are about a quarter the length of those of P. One repeats this process with Q, and keeps on going until one arrives at a point of small height (call the set of such points H). This is rather like the Euclidean algorithm, and when we reverse the process we find that P can be expressed as linear combination of elements of S and H, and hence $r \le |S| + |H|$.

There were several difficult calculations in Mordell's original proof in finding R. Weil made the astute observation that Mordell's process is tantamount to expressing all of $E(\mathbb{Q})$ as a finite set of cosets of $2E(\mathbb{Q})$, and hence it is enough to show that $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite. Weil came up with an elegant argument which generalizes to many other *algebraic groups* (that is generalizations of $E(\mathbb{Q})$). We will exhibit this argument in the special case that x^3+ax+b has three integer roots.

So suppose that $x^3+ax+b = (x-r_1)(x-r_2)(x-r_3)$ with $r_1, r_2, r_3 \in \mathbb{Z}$. Given a rational point $P = (m/n^2, \ell/n^3)$ with $(\ell m, n) = 1$ we have $\ell^2 = (m-r_1n^2)(m-r_2n^2)(m-r_3n^2)$. Since $(m-r_in^2, m-r_jn^2) = (m-r_in^2, r_i-r_j)$ each $m-r_in^2 = \alpha_i\beta_i^2$ where α_i is a squarefree integer which divides $(r_i-r_j)(r_i-r_k)$. Hence we define a map

$$\phi : E(\mathbb{Q}) \to \{(\alpha_1,\alpha_2,\alpha_3) \in \mathbb{Q}^*/(\mathbb{Q}^*)^2 : \ \alpha_1\alpha_2\alpha_3 \in (\mathbb{Q}^*)^2\}$$
$$\text{given by } \ \phi(P) = (x-r_1, x-r_2, x-r_3),$$

where α_i is a squarefree divisor of $(r_i-r_j)(r_i-r_k)$. If one of the $x-r_j$ equals 0 then we let α_j be the product of the other two α_i; for example if $x = r_1$ then $\phi((r_1,0)) = ((r_1-r_2)(r_1-r_3), r_1-r_2, r_1-r_3)$.

[27] Mordell's argument works in any number field.

We will multiply two such vectors as $(\alpha_1, \alpha_2, \alpha_3)(\beta_1, \beta_2, \beta_3) = (\alpha_1\beta_1, \alpha_2\beta_2, \alpha_3\beta_3)$. Note that $\phi(P) = \phi(-P)$.

Now suppose that we have three points P_1, P_2, P_3 on the line $y = mx + b$. Then their x-co-ordinates are all roots of

$$(x - r_1)(x - r_2)(x - r_3) - (mx + b)^2,$$

and so this polynomial equals $(x - x_1)(x - x_2)(x - x_3)$ where $x_j = x(P_j)$. Taking $x = r_i$ we deduce that $(x_1 - r_i)(x_2 - r_i)(x_3 - r_i) = (mr_i + b)^2 \in (\mathbb{Q}^*)^2$ so that

$$\phi(P_1)\phi(P_2)\phi(P_3) = (1, 1, 1),$$

In particular we deduce that $\phi(2P) = (1, 1, 1)$.

On the other hand suppose that $\phi(Q) = (1, 1, 1)$ where $Q = (U, V)$, so that there exist $t_1, t_2, t_3 \in \mathbb{Q}$ such that $U - r_i = t_i^2$ for each i. Now

$$\begin{aligned}\det \begin{pmatrix} t_1 & r_1 & 1 \\ t_2 & r_2 & 1 \\ t_3 & r_3 & 1 \end{pmatrix} &= \det \begin{pmatrix} t_1 & U - t_1^2 & 1 \\ t_2 & U - t_2^2 & 1 \\ t_3 & U - t_3^2 & 1 \end{pmatrix} \\ &= \det \begin{pmatrix} t_1 & t_1^2 & 1 \\ t_2 & t_2^2 & 1 \\ t_3 & t_3^2 & 1 \end{pmatrix} = \pm(t_1 - t_2)(t_2 - t_3)(t_3 - t_1) \neq 0,\end{aligned}$$

so there exists rational numbers u, m, b such that

$$\begin{pmatrix} t_1 & r_1 & 1 \\ t_2 & r_2 & 1 \\ t_3 & r_3 & 1 \end{pmatrix} \begin{pmatrix} u \\ m \\ b \end{pmatrix} = \begin{pmatrix} t_1 r_1 \\ t_2 r_2 \\ t_3 r_3 \end{pmatrix}.$$

Therefore $mr_i + b = -t_i(u - r_i)$ for each i so that the monic polynomial $(x - r_1)(x - r_2)(x - r_3) - (mx + b)^2$ takes value $-(u - r_i)^2 t_i^2 = (r_i - u)^2(r_i - U)$ at $x = r_i$. Hence $(x - r_1)(x - r_2)(x - r_3) - (mx + b)^2 = (x - u)^2(x - U)$. Taking $x = u$ yields the rational points $\pm P = (u, \pm(mu + b))$ on the curve, and one verifies that $Q = -2P$.

We have proved more than claimed, specifically that the image of ϕ is isomorphic to $E/2E$ (since the kernel of ϕ is $2E$). Now if $E = T \oplus \mathbb{Z}^r$ then $E/2E = T/2T \oplus \mathbb{Z}^r$.

`Exercise H1.11.` Prove that $|T/2T|$ equals the number of points of order 1 or 2. Deduce if there are 2^t points of order 1 or 2 (the possibilities being $t = 0, 1$ or 2), and the image of ϕ contains 2^s elements, then the rank of $E(\mathbb{Q})$ equals $r = s - t$.

In honor of their work the group of points $E(\mathbb{Q})$ is known as the *Mordell-Weil group*.

Example: The elliptic curve $E: \ y^2 = x^3 - x$ has three points of order two, namely $(-1, 0), (0, 0), (1, 0)$ and so $t = 2$ above. In the map ϕ we see that $\alpha_1 | (-1 - 0)(-1 - 1) = 2$, $\alpha_2 | 1$ and $\alpha_3 | 2$. Moreover $x - 1 < x < x + 1$ so that $x + 1$ cannot be negative. Hence the image of ϕ is a subgroup of a group G, which is generated, multiplicatively, by $(2, 1, 2)$

and $(-1,-1,1)$ and hence $s \le 2$. Therefore $r = s - t \le 2 - 2 = 0$, and so $E(\mathbb{Q})$ has rank 0. Note that $\phi(-1,0) = (-2,-1,2)$, $\phi(0,0) = (-1,-1,1)$, $\phi(1,0) = (2,1,2)$ and $\phi(O) = (1,1,1)$ where O is the point at infinity, and so $s = 2$.

Another example: Four squares in an arithmetic progression: Fermat proved, by descent, that there are no four distinct squares in an arithmetic progression. Let's see how we can prove this using the Mordell-Weil group and our map ϕ. If $a-d$, a, $a+d$ and $a+2d$ are all squares, say $u_{-1}^2, u_0^2, u_1^2, u_2^2$ then $(-2d/a,\ 2u_{-1}u_0u_1u_2/a^2)$ is a point on the elliptic curve

$$E:\ y^2 = (x-1)(x-2)(x+2).$$

Here $t = 2$ again but things are a bit more complicated, since $P = (0,2)$ is a point of order 4. Then $2P = (2,0)$ and the other points of order two are $R = (1,0)$ and $R+2P = (-2,0)$. We also have another point of order four namely $R - P = (4,6)$, as well as $-P$ and $P - R$. Now $\phi(P) = (-1,-2,2)$, $\phi(P-R) = (3,2,6)$, $\phi(R) = (-3,-1,3)$ and $\phi(2P) = (1,1,1)$.

`Exercise H1.12.` Show that the image of ϕ is a subgroup of a group G, which is generated, multiplicatively, by $(-1,-1,0), (1,2,2)$ and $(3,1,3)$. Deduce that the image of ϕ either contains all of G or no more than the four elements we have already identified. Prove also that the rank is therefore either 1 or 0, respectively.

In the next paragraph we will show that there is no point $(x,y) \in E(\mathbb{Q})$ for which $\phi(x,y) = (1,2,2)$ and thus the rank is 0 by the previous exercise. Now if S is a point that corresponds to an example where $a-d$, a, $a+d$ and $a+2d$ are all squares, then $\phi(S) = (-1,-2,2)$ and we deduce that $S = \pm P = (0,\pm 2)$ and hence $d = 0$. Therefore there are no four distinct squares in an arithmetic progression.

If $(x,y) \in E(\mathbb{Q})$ with $\phi(x,y) = (1,2,2)$ then we can write $x - 1 = (b/c)^2$ where $(b,c) = 1$ by exercise H1.1, and hence we also have $b^2 - c^2 = 2v^2$, $b^2 + 3c^2 = 2w^2$ for some integers v and w. Now b and c are both odd (since $b + c \equiv 0 \pmod 2$ and they are coprime), but then $2w^2 \equiv 1 + 3 = 4 \pmod 8$ which is impossible.

`Exercise H1.13.` Use the previous result together with Szemerédi's Theorem from section G2 to prove the following: *For any $\delta > 0$ there exists a constant M_δ such that if $N \ge M_\delta$ then any arithmetic progression of length N contains $< \delta N$ squares.* (It is conjectured that the N-term arithmetic progression with the most squares is $1, 1+24, 1+24\cdot 2, \dots, 1+24(N-1)$, which contains about $\sqrt{8N/3}$ squares; the best bound proved to date is at most a little more than $N^{3/5}$ squares.)

`Exercise H1.14.` Let E be the elliptic curve $y^2 = x(x+2)(x+16)$.

(1) Prove that the point $P = (2,12) \in E(\mathbb{Q})$ has infinite order. Deduce that $r \ge 1$
(2) Prove that there are no integer solutions to $u^2 + 2v^2 = 7w^2$. Deduce that $s < 4$.
(3) Show that $t = 2$ and hence deduce that r, the rank of $E(\mathbb{Q})$, is one.

No rational points by descent: Suppose that $x, y \in \mathbb{Q}$ such that $y^2 = x^3 + x$, so that $x = p/t^2, y = q/t^3$ with $(pq,t) = 1$ and

$$q^2 = p(p^2 + t^4).$$

Now $(p, p^2+t^4) = (p, t^4) = 1$ so that both p and $p^2 + t^4$ are squares, say $p = u^2$ and $p^2 + t^4 = w^2$. Therefore

$$t^4 + u^4 = p^2 + t^4 = w^2,$$

and we showed, in section 6.4, that this has no solutions. Hence there are no rational points on $y^2 = x^3 + x$ except with $x = 0$ or $y = 0$; that is, the point $(0, 0)$.

In general one can write down generators of the Mordell-Weil group, say $P_1, P_2, \dots, P_r$ and all points on $E(\mathbb{Q})$ can be written as $a_1P_1 + a_2P_2 + \dots + a_rP_r$ for some integers $a_1, \dots, a_r$. If P_j has infinite order then we take any $a_j \in \mathbb{Z}$; if P_j has order m_j then we take $a_j \pmod{m_j}$. We can add points, by adding the vectors of the a_j componentwise, and according to these rules.

If there are infinitely many points in $E(\mathbb{Q})$, how are they spaced on the curve itself? Are they dense on the curve? This sort of question can be answered but requires methods from beyond this discussion.

How big the rank can get is an open question. Researchers have found elliptic curves for which $E(\mathbb{Q})$ has rank at least 28, and some people believe that ranks of $E(\mathbb{Q})$ can get arbitrarily large as we vary over elliptic curves E, though for now this is more a belief than a conjecture.

We have already seen something similar to the notion of Mordell-Weil groups when we were considering solutions to Pell's equation. There all solutions take the form $\pm\epsilon^a$ so that this group of units is generated by -1 and ϵ_d and has structure $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}$, this ± 1 being torsion. There can be more torsion than just ± 1; for example, in $\mathbb{Z}[i]$ we also have the units $\pm i$ so the unit group structure is $\mathbb{Z}/4\mathbb{Z}$, generated by i.

`Exercise H1.15.` The size of ϵ_d^n grows exponentially in n. How fast does 2^kP grow (as a function of k)? Can you then deduce a result about the growth of nP?

Integral points on elliptic curves. We change models of elliptic curves by linear changes of variables, which allows us to keep consistent our notion of rational points (as long as we are careful about points at infinity). However such transformations do not preserve integer points, making such question a little more ad hoc, in the sense that the question depends on the choice of model for the elliptic curve.

Siegel showed that there are only finitely many integral points on any model of an elliptic curve, and indeed on any model of any curve that is not transformable to a linear equation. The proof is a little beyond us here, but note that if we had $x(x-1)(x+1) = y^2$ in integers, then either $x - 1$ and x are squares, or $(x - 1)/2$ and $(x + 1)/2$ are squares, the only solution to consecutive squares being 0 and 1, and hence $x = 1, y = 0$. This proof generalizes but in an example like $x(2x + 1)(3x + 1) = y^2$ we see that there exist integers u, v, w such that $x = \pm u^2,\ 2x+1 = \pm v^2,\ 3x+1 = \pm w^2$. Hence $v^2-2u^2 = 3v^2-2w^2 = \pm 1$. Squaring the second solution gives $(3v^2 + 2w^2)^2 - 6(2vw)^2 = 1$, and so we get solutions to a *simultaneous Pell equation*. Since the solutions to one Pell equation are so sparse, it seems likely that there are few co-incidences between two.[28]

Sums of two cubes. Suppose that are studying rational solutions of $a^3 + b^3 = k\ (\neq 0)$. Writing $u = a + b,\ v = a - b$ and then $y = 36kv/u,\ x = 12k/u$ we get $y^2 = x^3 - 3(12k)^2$.

[28]Bennett et. al. proved that there are never more than two solutions to $x^2 - az^2 = y^2 - bz^2 = 1$ for given integers $a > b \geq 1$. This cannot be improved since for any z we can select integers x and y such that $x^2 \equiv y^2 \equiv 1 \pmod{z^2}$ and then take $a = (x^2 - 1)/z^2,\ b = (y^2 - 1)/z^2$.

`Exercise H1.16.` Show that from every rational solution x, y to $y^2 = x^3 - 3(12k)^2$ we can obtain a rational solution a, b to $a^3 + b^3 = k$.

Hence we see that studying the sum of two cubes is also a problem about elliptic curves. We have seen that 1729 is the smallest integer that can be represented in two ways. Are there integers that can be represented in three ways, or four ways, or...? Actually this is not difficult to answer: $1^3 + 12^3 = 9^3 + 10^3 = 1729$. Using the doubling process on the cubic curve $a^3 + b^3 = 1729$

$$\text{If } P = (a, b) \text{ then } 2P = (A, B) \text{ then } A = a\,\frac{a^3 - 3458}{1729 - 2a^3} \text{ and } B = b\,\frac{a^3 + 1729}{1729 - 2a^3}\,.$$

So, starting from the solution $(12, 1)$, we get further solutions $(20760/1727, -3457/1727)$, $(184026330892850640/15522982448334911, 61717391872243199/15522982448334911)$, and the next solution is pointless to write down since each ordinate has seventy digits! The main point is that there are infinitely many different solutions, let us write them as $(u_i/w_i, v_i/w_i)$, $i = 1, 2, \ldots$ with $w_1 | w_2 | \ldots$.[29] Hence we have N solutions to $a^3 + b^3 = 1729 w_N^3$ taking $a = u_i(w_N/w_i)$ and $b = v_i(w_N/w_i)$.

This scaling up of rational points seems like a bit of a cheat, so let's ask whether there exists an integer m that can be written in N ways as the sum of two cubes of coprime integers? People have found examples for $N = 3$ and 4 but not beyond, and this remains an open question.

[29] As we saw when discussing the proof of the Lutz-Nagell Theorem.

H2. Elliptic curves and finite fields. Look at the curve $E:\ y^2 = x^3 - x$. Let N_p be the number of pairs $(x, y) \in [0, p-1]^2$ such that $y^2 \equiv x^3 - x \pmod p$. By Corollary 8.2 we have

$$N_p = \sum_{n=0}^{p-1} 1 + \left(\frac{n^3 - n}{p}\right) = p + \sum_{n=0}^{p-1} \left(\frac{n}{p}\right)\left(\frac{n^2-1}{p}\right).$$

Now if $p \equiv 3 \pmod 4$ then $\left(\frac{(-x)^3-(-x)}{p}\right) = \left(\frac{-(x^3-x)}{p}\right) = -\left(\frac{x^3-x}{p}\right)$ and so the sum on the right side is 0. Hence $N_p = p$.

It not so obvious how to easily calculate the sum when $p \equiv 1 \pmod 4$. If we do calculations we see that

$$N_5 = 7,\ N_{13} = 19,\ N_{17} = 19,\ N_{29} = 39,\ N_{37} = 39,\ N_{41} = 51,\ N_{53} = 67,\ N_{61} = 71,\ N_{73} = 79,$$

and hence $N_5 - 5 = 2,\ N_{13} - 13 = 6,\ N_{17} - 17 = 2,\ N_{29} - 29 = 10,\ N_{37} - 37 = 2,\ N_{41} - 41 = 10,\ N_{53} - 53 = 14,\ N_{61} - 61 = 10,\ N_{73} - 73 = 6, \dots$ In Proposition F2.1, we saw that if $p = a^2 + b^2$ with a odd then

$$N_p = p - 2(-1)^{\frac{a+b+1}{2}} a.$$

In section F4 we saw that

$$N_p - p \equiv -\binom{\frac{p-1}{2}}{\frac{p-1}{4}} \pmod p.$$

Hence we deduce that

$$\binom{\frac{p-1}{2}}{\frac{p-1}{4}} \equiv 2(-1)^{\frac{a+b+1}{2}} a \pmod p,$$

a congruence found by Gauss. Beukers' *supercongruence* allows us to determine rather more. Select i such that $i^2 \equiv -1 \pmod p$ with $i \equiv a/b \pmod{p^2}$. Evidently $a - bi \equiv 0 \pmod{p^2}$, but

$$\binom{\frac{p^2-1}{2}}{\frac{p-1}{4}} \equiv (-1)^{\frac{a+b+1}{2}}(a + bi) \pmod{p^2}.$$

In general, for any elliptic curve mod p there exists a complex number $\alpha_p := a + \sqrt{-d}b$ with $a, b, d \in \mathbb{Z}$ for which $p = a^2 + db^2$ and $N_p = p - 2a = p - \alpha_p - \overline{\alpha_p}$. Thus there is a link between the elliptic curve mod p and $-d$, which has the same squarefree part as $t^2 - 4p$ where $t = p - N_p$. Indeed the number of elliptic curves (up to isomorphism) with $N_p = p - t$ is given by the number of equivalence classes of binary quadratic forms of discriminant $t^2 - 4p$ whether or not they are primitive.

Statement of B SW-D. For each extra rank we increase by 1 the average number of points in $\mathbb{F}_p$ by 1.

Heegner points? And so the Taniyama-Shimura conjecture.

What is Complex-Multiplication, and what is Sato-Tate? State Taylor's Theorem.

The congruent number problem?

H3. More L-functions. Let's study how many solutions there are to $y^2 - dx^2 = 1$ with $|x| \le N$, for N large. Now if x, y are a pair of positive integers for which $0 \le y^2 - dx^2 < 2$ then $y^2 - dx^2 = 1$ so we could guess the number of such pairs is the volume of this region. Given x it is more-or-less true that $0 \le y^2 - dx^2 < 2$ is equivalent to $0 \le y - \sqrt{d}x < 1/\sqrt{d}x$, and hence the volume with $x \le N$ is

$$\approx \int_1^N \frac{\mathrm{d}x}{\sqrt{d}x} = \frac{\log N}{\sqrt{d}} .$$

Now this heuristic pre-supposes that any x, y of the right size are going to be solutions, but we should try to take into account what we know from congruences. That is that the proportion of pairs of integers x, y for which $x^2 - dy^2 \equiv 1 \pmod p$ is not $1/p$ but rather $1/p$ times $1 - \frac{1}{p} \left(\frac{d}{p}\right)$ as we saw in section *. Hence we should multiply the above through by

$$\prod_p \left(1 - \frac{1}{p} \left(\frac{d}{p}\right)\right) = \frac{1}{L(1, \left(\frac{d}{\cdot}\right))} .$$

Hence we might predict that the number of solutions is roughly

$$\frac{\log N}{\sqrt{d}\ L(1, \left(\frac{d}{\cdot}\right))} .$$

However we did see earlier that all solutions to $y^2 - dx^2 = 1$ with $x, y \ge 1$ are powers of the fundamental solutions ϵ_d to Pell's equation, and hence the number of solutions is actually $\frac{\log N}{\log \epsilon_d}$. Equating the two we might thus guess that

$$\log \epsilon_d \approx \sqrt{d}\ L(1, \left(\frac{d}{\cdot}\right)).$$

If one calculates one finds one gets equality here whenever $h(d) = 1$ but not otherwise! In fact the classes of the class group form another thing we have to take into account. Basically one can perform the same calculation corresponding to one reduced form from each class and get different solutions to Pell's equation. Hence we adjust our predicted count to $\frac{h(d) \log N}{\sqrt{d}\ L(1, \left(\frac{d}{\cdot}\right))}$, leading to Dirichlet's formula:

$$h(d) \log \epsilon_d = \sqrt{d}\ L(1, \left(\frac{d}{\cdot}\right)).$$

How many rational points are there on an elliptic curve? Birch and Swinnerton-Dyer reasoned that, like in the case of quadratics, if there are more mod p, on average, then there are probably more rational points. Hence although we might expect p solutions to $y^2 = x^3 + ax + b \pmod p$ we actually get $p - a_p$ and so we should adjust accordingly. Thus a factor

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

should come into play. We saw that $a_p = \alpha_p + \overline{\alpha_p}$ for some quadratic complex number for which $\alpha_p\overline{\alpha_p} = p$. This suggest a factor which is more natural than $1 - a_p/p$, since it factors, namely:

$$\left(1- \frac{\alpha_p}{p}\right)\left(1- \frac{\overline{\alpha_p}}{p}\right) = 1- \frac{a_p}{p} + \frac{1}{p^2}\ .$$

Hence it seems natural to define the elliptic curve for an L-function as

$$\prod_p \left(1- \frac{a_p}{p^s} + \frac{1}{p^{2s}}\right)^{-1}$$

though, as with the Dirichlet L-function, we may have to do something slightly different for the primes p that divide the *conductor*; for Dirichlet L-functions that was the modulus q; here it is some divisor of the discriminant, Δ.

This more-or-less defines the *Hasse-Weil* L-function of an elliptic curve $L(E, s)$, and we might ask to what extent it has the same properties as a Dirichlet L-function? We claimed that giving an analytic continuation of $L(s, \chi)$ to the whole complex plane is not too difficult; for $L(E, s)$ this is substantially more difficult though we now know how to do so (as will be explained shortly). One formula for $L(s, chi)$ shows how values at s are related to those at $1 - s$. For $L(E, s)$ this is true but the symmetry is between $1/2 + s$ and $3/2 - s$. The center of that symmetry is the line $\mathrm{Re}(s) = 1/2$ and we believe that all non-trivial zeros of $L(s, \chi)$ lie on this line, a *Riemann Hypothesis*, and we believe that the same is true for $L(E, s)$.

Transformations. It is interesting to make a power series out of the coefficients of a given Dirichlet series. Hence for $\zeta(s)$ we obtain $\sum_{n\geq 1} t^n = t/(1-t)$. For a Dirichlet L-function $L(s,\chi)$ we have, using the periodicity of $\chi \pmod q$, and so writing $n = qr + m$

$$\sum_{n\geq 1} \chi(n)t^n = \sum_{m=1}^{q}\sum_{r\geq 0} \chi(m)t^{qr+m} = \frac{\sum_{m=1}^{q}\chi(m)t^m}{1-t^q},$$

a rational function. This is not quite as ad hoc a procedure as it seems at first sight since by defining[30]

$$\Gamma(s) = \int_0^\infty e^{-t}t^{s-1}dt$$

for $\mathrm{Re}(s) > 0$ we have, by changing variable $t \to nt$, $\Gamma(s) = n^s \int_0^\infty e^{-nt}t^{s-1}dt$ and so

$$\Gamma(s)L(s,\chi) = \sum_{n\geq 1}\chi(n)\int_0^\infty e^{-nt}t^{s-1}dt = \int_0^\infty \frac{\sum_{m=1}^{q}\chi(m)e^{-mt}}{1-e^{-qt}}\, t^{s-1}dt.$$

This provides an analytic continuation for $L(s,\chi)$.

Let E be the elliptic curve $E : y^2 = x^3 - x$. By Proposition F2.1 we have that $L(E,s) = \sum_{n\geq 1} a_n/n^s$ where the a_n are given by (G1.2).[31] In particular this implies that

$$\Gamma(s)L(E,s) == \int_0^\infty e^{-t}\prod_{n\geq 1}(1-e^{-4nt})^2(1-e^{-8nt})^2t^{s-1}dt.$$

So what is special about these integrands that they give rise to L-functions? The formula in (G1.2) is not a rational function, but it is something elegant. Is there something special about it that leads us to L-functions? This is the question that Taniyama asked himself in 1955 and led to one of the most extraordinary chapters in the history of number theory. More on this in section H7.

[30] $\Gamma(s)$ is the function that extrapolates $n!$, so that $\Gamma(n+1) = n!$. Because of this it is involved in many beautiful combinatorial formulas many of which stem from

$$\frac{1}{s\Gamma(s)} = e^{\gamma s}\prod_{n=1}^{\infty}\left(1+\frac{s}{n}\right)e^{-s/n}.$$

[31] Actually we only have this for a_p immediately but the rest can be proved.

H4. FLT and Sophie Germain. The first strong result on FLT was due to Sophie Germain:

Lemma. *Suppose that p is an odd prime for which $q = 2p + 1$ is also prime. If a, b, c are coprime integers for which $a^p + b^p + c^p \equiv 0 \pmod q$ then q divides at least one of a, b, c.*

Proof. Since $p = \frac{q-1}{2}$ we know that $t^p \equiv -1$ or $1 \pmod q$ for any integer t that is not divisible by q, and so if q does not divide abc then $a^p + b^p + c^p \equiv -3, -1, 1$ or $3 \pmod q$. This is impossible as $q = 2p + 1 > 3$.

Sophie Germain's Theorem. *Suppose that p is an odd prime for which $q = 2p + 1$ is also prime. There do not exist integers x, y, z for which p does not divide x, y, z and $x^p + y^p + z^p = 0$.*

Proof. Assume that there is a solution. As we saw in Proposition 6.3 we may assume that x, y, z are pairwise coprime so, by the lemma, exactly one of x, y, z is divisible by q: Let us suppose that q divides x, without loss of generality since we may re-arrange x, y and z as we please.

By exercise 3.1.21 there exist integers a, b, c, d such that

$$z + y = a^p,\ z + x = b^p,\ x + y = c^p \text{ and } \frac{y^p + z^p}{y + z} = d^p \ \ (\text{as } p \nmid xyz), \text{ where } x = -cd.$$

Now $a^p = z + y \equiv (z + x) + (x + y) \equiv b^p + c^p \pmod q$ as $q|x$, and so we see that q divides at least one of a, b, c by the Lemma. However since $(q, b)|(x, z + x) = (x, z) = 1$ as $q|x$ and $b|z + x$ and so q does not divide b, as well as c, analogously. Hence q divides a, that is $-z \equiv y \pmod q$. But then

$$d^p = \frac{y^p + z^p}{y + z} = \sum_{j=0}^{p-1} (-z)^{p-1-j} y^j \equiv \sum_{j=0}^{p-1} y^{p-1-j} y^j = p y^{p-1} \pmod q.$$

Therefore, as $y \equiv x + y = c^p \pmod q$ and $q - 1 = 2p$, we deduce that

$$4 \equiv 4d^{2p} = (2d^p)^2 \equiv (2py^{p-1})^2 = (-1)^2 (c^{2p})^{p-1} \equiv 1 \pmod q,$$

which is impossible as $q > 3$.

Hence if one can show that there are infinitely many pairs of primes $p, q = 2p+1$ then there are infinitely many primes p for which there do not exist integers x, y, z for which p does not divide x, y, z and $x^p + y^p + z^p = 0$.

After Sophie Germain's Theorem, the study of Fermat's Last Theorem was split into two cases:

I) Where $p \nmid xyz$; and II) where $p|xyz$.

One can easily develop Germain's idea to show that if $m \equiv 2$ or $4 \pmod 6$ then there exists a constant $N_m \neq 0$ such that if p and $q = mp + 1$ are primes for which $q \nmid N_m$ then FLTI is true for exponent p. This was used by Adleman, Fouvry and Heath-Brown to show that FLTI is true for infinitely many prime exponents.

There were many early results on the first case of Fermat's Last Theorem, which showed that if there is a solution with $p \not| xyz$, then some extraordinary other things must happen. Here is a list of a few:

If there is a solution to FLTI then

i) We have $2^{p-1} \equiv 1 \pmod{p^2}$. In section B3 we saw that this seems to happen rarely. In fact one also has $3^{p-1} \equiv 1 \pmod{p^2}, 5^{p-1} \equiv 1 \pmod{p^2}, \dots, 113^{p-1} \equiv 1 \pmod{p^2}$. Indeed one can obtain as many criteria like this as one wishes after a finite amount of calculation.

ii) p divides the numerator of $B_{p-3}, B_{p-5}, \dots, B_{p-r}$ for $r \le (\log p)^{1/2-o(1)}$. And p divides the numerator of at least $\sqrt{p} - 2$ non-zero Bernoulli numbers $B_n,\ 2 \le n \le p-3$.

iii)

Let us try to prove Fermat's Last Theorem, ignoring many of the technical issues. Let ζ be a primitive p the root of unity. Then we can factor

$$x^p + y^p = (x+y)(x+\zeta y)(x+\zeta^2 y)\dots(x+\zeta^{p-1}y).$$

Now we are working in the set $\mathbb{Z}[\zeta]$, and we see that $\gcd(x+\zeta^i y, x+\zeta^j y)$ divides $(x+\zeta^i y) - (x+\zeta^j y) = (\zeta^i - \zeta^j)y$ and $\zeta^j(x+\zeta^i y) - \zeta^i(x+\zeta^j y) = (\zeta^j - \zeta^i)x$, so that $\gcd(x+\zeta^i y, x+\zeta^j y)$ divides $(\zeta^i - \zeta^j)(x,y)$. Note that $\zeta^k,\ 1 \le k \le p-1$ are the roots of $x^{p-1} + x^{p-2} + \dots + 1$ and so

$$\prod_{k=1}^{p-1}(1-\zeta^k) = \prod_{k=1}^{p-1}(x-\zeta^k)\Big|_{x=1} = x^{p-1} + x^{p-2} + \dots + 1\Big|_{x=1} = p;$$

therefore if $k = j - i$ then $\zeta^i - \zeta^j = \zeta^i(1-\zeta^k)$ divides p. So now assume that we have a solution to FLTI, that is $x^p + y^p = z^p$ with $\gcd(x,y) = 1$ and $p \not| z$, and so $x+y,\ x+\zeta y,\ x+\zeta^2 y,\ x+\zeta^{p-1}y$ are pairwise coprime elements of $\mathbb{Z}[\zeta]$ whose roduct is a pth power. If this works like the regular integers then each $x+\zeta^j y$ is a pth power. So we have gone from three linearly independent pth powers to p linearly dependent pth powers! In particular if $x + \zeta^j y = u_j^p$ then

$$(\zeta^j-\zeta^k)u_i^p + (\zeta^k-\zeta^i)u_j^p + (\zeta^i-\zeta^j)u_k^p = (\zeta^j-\zeta^k)(x+\zeta^i y) + (\zeta^k-\zeta^i)(x+\zeta^j y) + (\zeta^i-\zeta^j)(x+\zeta^k y) = 0.$$

Ignoring for a moment two technical details: the coefficients and the fact that we are no longer working over the integers, we see that we have found another solution to FLTI, this time with pth powers which are divisors of the previous pth powers and hence are smaller. Thus this seems to have the makings of a plan to prove FLTI by a descent process.

In 1850 Kummer attempted to prove Fermat's Last Theorem, much along the lines of last subsection. He however resolved a lot of the technical issues that we have avoided, creating the theory for ideals for $\mathbb{Z}[\zeta]$, much as we saw it discussed earlier for quadratic fields. That such similar theories evolved for quite different situations suggested that there was probably a theory of ideals that worked in any number field. Indeed such results were proved by Dedekind and became the basis of algebraic number theory, and indeed much of the study of algebra. Kummer's exact criteria was to show that if p does not divide a

certain class number (associated to $\mathbb{Z}[\zeta]$) then Fermat's Last Theorem is true for exponent p. He showed that p does not divide that certain class number, if and only if p does not divide the numerators of $B_2, B_4, B_6, \dots, B_{p-3}$.

In 1994 Wiles finally proved Fermat's Last Theorem based on an extraordinary plan of Frey and Serre to bring in ideas from the theory of elliptic curves. In fact Fermat's Last Theorem falls as a consequence of Wiles' (partial) resolution of a conjecture about the structure of elliptic curves that we will discuss in section *. The elliptic curve associated to a solution $a^p + b^p + c^p = 0$ is $y^2 = x(x + a^p)(x - b^p)$, because then the discriminant Δ, which is the product of the difference of the roots, squared, equals $(abc)^{2p}$.

Another famous problem about powers in Catalan's 1844 conjecture that the only perfect powers that differ by 1 are 8 and 9 (that is if $x^p - y^q = 1$ then either $x = 0$, or $y = 0$, or $x = q = 3$ and $y = p = 2$. After Baker's Theorem it was known that there could be only finitely many such pairs, but the conjecture was only proved in 2004 by Mihailescu with a proof more along the lines of Kummer's work on Fermat's Last Theorem.

Now that the two key conjectures in this field have been resolved, one can ask about other Diophantine equations involving powers. A hybrid is the *Fermat-Catalan equation*

$$x^p + y^q = z^r.$$

However there can be many uninteresting solutions: For example if $q = p$ and $r = p + 1$ one has solutions $(a(a^p + b^p))^p + (b(a^p + b^p))^p = (a^p + b^p)^{p+1}$ for any integers a and b. This kind of solution can be ignored by assuming that x, y and z are pairwise coprime. However it is not hard to find some solutions:

$$1 + 2^3 = 3^2, \quad 2^5 + 7^2 = 3^4, \quad 7^3 + 13^2 = 2^9, \quad 2^7 + 17^3 = 71^2, \quad 3^5 + 11^4 = 122^2.$$

and with a little more searching one finds five surprisingly large solutions:

$$17^7 + 76271^3 = 21063928^2, \quad 1414^3 + 2213459^2 = 65^7, \quad 9262^3 + 15312283^2 = 113^7,$$

$$43^8 + 96222^3 = 30042907^2, \quad 33^8 + 1549034^2 = 15613^3.$$

The Fermat-Catalan conjecture. *There are only finitely many solutions to $x^p + y^q = z^r$ in coprime integers x, y, z, where $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} < 1$.*

A stronger version of the conjecture states that there are only the ten solutions listed above. However this sort of conjecture is always a little feeble since if someone happens to find one more isolated example, then would we not believe that those eleven solution are all?

Several people have observed that all of the solutions above have at least one exponent 2, so that one can conjecture that there are only finitely many solutions to $x^p + y^q = z^r$ in coprime integers x, y, z, where $p, q, r \geq 3$.

The cases where $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} \geq 1$ are fully understood. When $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$ we only have the solution $1^6 + 2^3 = 3^2$.

`Exercise` Show that there are infinitely many coprime solutions to $x^2 + y^2 = z^r$ for any fixed r. (Hint: Use your understanding of what integers are the sum of two squares.)

For the other cases there are infinitely many solutions; for example the parametrization $(a(a^3-8b^3))^3+(4b(a^3+b^3))^3=(a^6+20a^3b^3-8b^6)^2$ (due to Euler, 1756)

A proof of Fermat's Last Theorem? Fermat claimed that there are no solutions to

$$x^p+y^p=z^p \tag{1}$$

for $p\geq 3$, with x,y and z all non zero. If we assume that there are solutions to (1) then we can assume that x,y and z have no common factor else we can divide out by that factor. Our first step will be to differentiate (1) to get

$$px^{p-1}x'+py^{p-1}y'=pz^{p-1}z'$$

and after dividing out the common factor p, this leaves us with

$$x^{p-1}x'+y^{p-1}y'=z^{p-1}z'. \tag{2}$$

We now have two linear equations (1) and (2) (thinking of x^{p-1},y^{p-1} and z^{p-1} as our variables), which suggests we use linear algebra to eliminate a variable: Multiply (1) by y' and (2) by y, and subtract, to get

$$x^{p-1}(xy'-yx')=z^{p-1}(zy'-yz').$$

Therefore x^{p-1} divides $z^{p-1}(zy'-yz')$, but since x and z have no common factors, this implies that

$$x^{p-1} \text{ divides } zy'-yz'. \tag{3}$$

This is a little surprising, for if $zy'-yz'$ is nonzero then a high power of x divides $zy'-yz'$, something that does not seem consistent with (1).

We want to be a little more precise. Since we differentiated, we evidently never were working with integers x,y,z but rather with polynomials. Thus if $zy'-yz'=0$ then $(y/z)'=0$ and so y is a constant multiple of z, contradicting our statement that y and z have no common factor. Therefore (3) implies that

$$(p-1)\ \text{degree}(x)\leq \text{degree}(zy'-yz')\leq \text{degree}(y)+\text{degree}(z)-1,$$

since $\text{degree}(y')=\text{degree}(y)-1$ and $\text{degree}(z')=\text{degree}(z)-1$. Adding $\text{degree}(x)$ to both sides gives

$$p\ \text{degree}(x)<\text{degree}(x)+\text{degree}(y)+\text{degree}(z). \tag{4}$$

The right side of (4) is symmetric in x,y and z. The left side is a function of x simply because of the order in which we chose to do things above. We could just as easily have derived the same statement with y or z in place of x on the left side of (4), so that

$$p\ \text{degree}(y)<\text{degree}(x)+\text{degree}(y)+\text{degree}(z)$$
$$\text{and } p\ \text{degree}(z)<\text{degree}(x)+\text{degree}(y)+\text{degree}(z).$$

Adding these last three equations together and then dividing out by $\text{degree}(x)+\text{degree}(y)+\text{degree}(z)$, implies

$$p<3,$$

and so Fermat's Last Theorem is proved! Well, not quite, but what we have proved (and so simply) is still of great interest:

Proposition 1. *There are no genuine polynomial solutions* $x(t), y(t), z(t) \in \mathbb{C}[t]$ *to* $x(t)^p + y(t)^p = z(t)^p$ *with* $p \geq 3$. *By "genuine" we mean that the triple* $(x(t), y(t), z(t))$ *is not a polynomial multiple of a solution of* (1) *in* $\mathbb{C}$.

That Fermat's Last Theorem is easy to prove for polynomials is an old result, going back certainly as far as Liouville (1851).

Fermat quotients. There are many questions around like whether p^2 divides $2^{p-1} - 1$. We call $q_p(2) := (2^{p-1} - 1)/p$ the *Fermat quotient*. Similarly $w_p = ((p-1)! + 1)/p$ is the *Wilson quotient*. One interesting connection is that

$$\frac{pB_{p-1} - (p-1)}{p} \equiv w_p \pmod{p},$$

where B_{p-1} is the $(p-1)$st Bernoulli number. The von-Staudt Clausen Theorem states that the denominator of B_n is the product of the primes p for which $p-1$ divides n, and $pB_{p-1} \equiv -1 \pmod{p}$.

Another surprising congruence is that

$$\binom{np}{mp} \equiv \binom{n}{m} \pmod{p^3},$$

not just mod p as we get from Lucas' Theorem (the case $n = 2, m = 1$ is known as Wolstenholme's Theorem). This is always divisible by p^4 if and only if p divides B_{p-3}, the $(p-3)$rd Bernoulli number.

In 1894 Morley showed that

$$(-1)^{\frac{p-1}{2}} \binom{p-1}{(p-1)/2} \equiv 4^{p-1} \pmod{p^3},$$

and this holds mod p^4 if and only if p divides B_{p-3}.

H5. Rational points on curves. We are interested when a curve has infinitely many rational points. We will proceed here with some identities: Throughout we will suppose that a, b, c are given non-zero integers for which $a+b+c=0$. One can check that if r, s, t are integers for which $r+s+t=0$ then

$$a(bs^2+ct^2)^2+b(as^2+cr^2)^2+c(at^2+br^2)^2=0$$

This can be applied to show that any quadratic equation $Ax^2+By^2=Cz^2$ with one non-zero integral point has infinitely many given as polynomials in r, s, t. To do so simply select $a=Ax^2,\ b=By^2,\ c=-Cz^2$ above and we obtain $AX^2+BY^2=CZ^2$ with

$$X=x(B(ys)^2-C(zt)^2),\quad Y=y(A(xs)^2-C(zr)^2),\quad Z=z(A(xt)^2+B(yr)^2).$$

For example, for the equation $x^2+y^2=2z^2$ starting from the solution $(1,1,1)$ we obtain the parametrization $X=s^2-2t^2,\ Y=s^2-2r^2,\ Z=t^2+r^2$, which we can re-write, taking $t=-s-r$, as $-X=2r^2+4rs+s^2,\ Y=s^2-2r^2,\ Z=2r^2+2rs+s^2$. Hence we see that any quadratic with one solution has infinitely many given parametrically.

We also have $a+b+c=0$, we have

$$a(b-c)^3+b(c-a)^3+c(a-b)^3=0$$

so from any given solution to $ax^3+by^3=cz^3$ we can find another significantly larger solution, that is $aX^3+bY^3=cZ^3$ with

$$X=x(by^3+cz^3)=x(ax^3+2by^3), Y=-y(ax^3+cz^3)=-y(2ax^3+by^3),\ Z=z(ax^3-by^3).$$

This is really the doubling process that we saw earlier but here it is rather more elegant to give such a formula, if we simply want to show that there are infinitely many solutions. Notice though that these solutions are considerably more rare than in the case of quadratic equations.

In section * we saw a proof of Fermat's Last Theorem for polynomials, though we do we have solutions for quadratic polynomials: Not only $(t^2-1)^2=(2t)^2=(t^2+1)^2$ but in fact we just saw that any diagonal quadratic that has one solution has a polynomial family of solutions. The intent now is to extend the idea in our proof of Fermat's Last Theorem for polynomials to as wide a range of questions as possible. It takes a certain genius to generalize to something far simpler than the original. But what could possibly be more simply stated, yet more general, than Fermat's Last Theorem? It was Richard C. Mason (1983) who gave us that insight: *Look for solutions to*

$$a+b=c. \tag{5}$$

We will just follow through the proof of FLT and see where it leads: Start by assuming, with no loss of generality, that a, b and c are all non-zero polynomials without common factors (else all three share the common factor and we can divide it out). Then we differentiate to get

$$a'+b'=c'.$$

Next we need to do linear algebra. It is not quite so obvious how to proceed analogously, but what we do learn in a linear algebra course is to put our coefficients in a matrix and solutions follow if the determinant is non-zero. This suggests defining

$$\Delta(t) := \begin{vmatrix} a(t) & b(t) \\ a'(t) & b'(t) \end{vmatrix}.$$

Then if we add the first column to the second we get

$$\Delta(t) = \begin{vmatrix} a(t) & c(t) \\ a'(t) & c'(t) \end{vmatrix},$$

and similarly

$$\Delta(t) = \begin{vmatrix} c(t) & b(t) \\ c'(t) & b'(t) \end{vmatrix}$$

by adding the second column to the first, a beautiful symmetry.

We note that $\Delta(t) \neq 0$, else $ab' - a'b = 0$ so b is a scalar multiple of a (with the same argument as above), contradicting hypothesis. To find the appropriate analogy to (3), we interpret that as stating that the factors of x (as well as y and z) divide our determinant to a high power. So now suppose that α is a root of $a(t)$, and that $(t-\alpha)^e$ is the highest power of $(t-\alpha)$ which divides $a(t)$. Evidently $(t-\alpha)^{e-1}$ is the highest power of $(t-\alpha)$ which divides $a'(t)$, and thus it is the highest power of $(t-\alpha)$ which divides $\Delta(t) = a(t)b'(t) - a'(t)b(t)$ (since α is not a root of $b(t)$). Therefore $(t-\alpha)^e$ divides $\Delta(t)(t-\alpha)$. Multiplying all such $(t-\alpha)^e$ together we obtain

$$a(t) \text{ divides } \Delta(t) \prod_{a(\alpha)=0} (t-\alpha).$$

In fact $a(t)$ only appears on the left side of this equation because we studied the linear factors of a; analogous statements for $b(t)$ and $c(t)$ are also true, and since $a(t), b(t), c(t)$ have no common roots, we can combine those statements to read

$$a(t)b(t)c(t) \quad \text{divides} \quad \Delta(t) \prod_{(abc)(\alpha)=0} (t-\alpha). \tag{6}$$

The next step is to take the degrees of both sides and see what that gives. Using the three different representation of Δ above, we have

$$\text{degree}(\Delta) \leq \begin{cases} \text{degree}(a) + \text{degree}(b) - 1, \\ \text{degree}(a) + \text{degree}(c) - 1, \\ \text{degree}(c) + \text{degree}(b) - 1. \end{cases}$$

The degree of $\prod_{(abc)(\alpha)=0}(t-\alpha)$ is precisely the total number of distinct roots of $a(t)b(t)c(t)$. Inserting all this into (6) we find that

$$\max\{\text{degree}(a), \text{degree}(b), \text{degree}(c)\} < \#\{\alpha \in \mathbb{C} : \ (abc)(\alpha) = 0\}.$$

Put another way, this result can be read as:

The *abc* Theorem for Polynomials. *If $a(t), b(t), c(t) \in \mathbb{C}[t]$ do not have any common roots and provide a genuine polynomial solution to $a(t) + b(t) = c(t)$, then the maximum of the degrees of $a(t), b(t), c(t)$ is less than the number of distinct roots of $a(t)b(t)c(t) = 0$.*

This is a "best possible" result in that we can find infinitely many examples where there is exactly one more zero of $a(t)b(t)c(t) = 0$ than the largest of the degrees. For example the familiar identity

$$(2t)^2 + (t^2 - 1)^2 = (t^2 + 1)^2;$$

or the rather less interesting

$$t^n + 1 = (t^n + 1).$$

Back to the above, we wish to know for what n there can be parametric solutions to

$$ax^n + by^n = cz^n.$$

That is for given integers a, b, c, can there be (coprime) polynomials $x, y, z \in \mathbb{Z}[t]$ or even $\mathbb{C}[t]$ satisfying this equation? If so, then if the maximum of the degrees of x, y, z is d, then we have $dn < 3d$ so that $n \leq 2$, by the *abc*-theorem for polynomials. That is the quadratic examples that we found, provide all of the possible examples.

In the cubic case we wrote x, y, z as polynomials in a, b, c, where x, y, z have no common factors given that $a + b = c$. By looking only at the term of highest degree we may assume that x, y, z are homogenous polynomials in a, b, c of degree d, say. Now we may write $a = ct$ and $b = c(1 - t)$ for some t, and then, dividing through x, y, z by c^d, we obtain an identity $tu^n + (1 - t)v^n = w^n$ where u, v, w are polynomials of degree d (or less) in t. Applying the *abc*-theorem for polynomials we have $dn + 1 < 3d + 2$ and so $n \leq 3$.

The *abc*-conjecture. Could there be an analogous result for the integers, which would also imply Fermat's Last Theorem, and perhaps much more? The idea would be to bound the size of the integers involved (in place of the degree) in terms of their distinct prime factors (in place of the number of roots). A first guess at an analogous result might be if $a + b = c$ with a, b, c pairwise coprime positive integers then a, b and c are bounded in terms of the number of prime factors of a, b, c but if, as we believe, there are infinitely many pairs of twin primes $p, p + 2$ then we have just three prime factors involved in $p + 2 = q$, and they get arbitrarily large. It therefore seems sensible to include the size of the prime factors involved in such a bound so we might guess at $c \leq \prod_{p|abc} p$, but again a simple example excludes this possibility: Let $1 + (2^n - 1) = 2^n$. If we take any prime q and then $n = q(q - 1)$ we have $q^2 | 2^n - 1$ and so $\prod_{p|(2^n-1)2^n} p \leq 2(2^n - 1)/q < 2^n$. In this case $q \approx \sqrt{n} \approx \sqrt{\log n}$ so even though our guess was wrong it is not too far out. This suggests that our guess is almost correct and could be made correct by fudging things a little bit:

The *abc*-conjecture. *For any fixed $\epsilon > 0$ there exists a constant κ_ϵ such that if a, b, c are pairwise coprime positive integers for which*

$$a + b = c$$

then

$$c \le \kappa_\epsilon \left(\prod_{\substack{p \text{ prime} \\ p|abc}} p \right)^{1+\epsilon}.$$

In particular one might guess that $\kappa_1 = 1$; that is $c \le \left(\prod_{p|abc} p\right)^2$. We can apply the *abc*-conjecture to FLT: Let $a = x^p, b = y^p, c = z^p$ with $0 < x, y < z$ so that

$$\prod_{p|abc} p = \prod_{p|xyz} p \le xyz < z^3.$$

The *abc*-conjecture implies that $z^p \le \kappa_\epsilon (z^3)^{1+\epsilon}$.

`Exercise` Deduce that if $p > 3$ then z is bounded independently of p.

`Exercise` Suppose that p, q, r are positive integers for which $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} < 1$.

(1) Show that $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} \le \frac{41}{42}$
(2) Show that the *abc*-conjecture implies that if $x^p + y^q = z^r$ with $(x, y, z) = 1$ then x, y, z are bounded independently of p, q, r.
(3) Deduce that if the *abc*-conjecture is true then the Fermat-Catalan conjecture is true.

Faltings' Theorem née Mordell's conjecture. Let $f(x, y) \in \mathbb{Z}[x, y]$ be an irreducible polynomial in two variables with integer coefficients. We are interested in finding rational numbers u and v for which $f(u, v) = 0$.

We have seen how to completely resolve this for f of degree 1 or 2: there are either no solutions, or an infinite of rational solutions , as a rational function of the variable t.

For f of degree 3 and sometimes 4 we can sometimes reduce the problem to an elliptic curve, and given one solution we can find another as a function of that solution, and thus get infinitely many solutions unless we hit on a torsion point (of which there are no more than 16).

Faltings' Theorem tells us that these are the only two ways in which an equation like $f(u, v) = 0$ have infinitely many rational solutions. That is, if put to one side all solutions of $f(x, y) = 0$ that come from the two methods above, then we are left with finitely many solutions. Therefore, for higher degree f, there are only finitely many "sporadic" solutions. It is even feasible that the number of rational points left over is bounded by a function of the degree of f. Faltings' extraordinary theorem has many wonderful consequences ... For any given $p \ge 4$ there are only finitely many positive coprime integer solutions to $x^p + y^p = z^p$. Similarly

$$x^4 + y^4 = 17z^4 \quad \text{and} \quad x^2 + y^3 = z^7$$

each have only finitely many coprime integer solutions. More generally there are only finitely many positive coprime integer solutions x, y, z to the Fermat-Catalan equation

$$ax^p + by^q = cz^r$$

for any positive coprime integers a, b, c whenever $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} < 1$.[32]

One important failing of Faltings' Theorem is that it does not give an upper bound on the size of the solutions, and so no "algorithm" for finding them all, even though we know there are only finitely many.

In 1991 Elkies showed that using an explicit version of the *abc*-conjecture (that is, with a value assigned to κ_ε for each ϵ), one can deduce an explicit version of Faltings' Theorem. The proof revolves around a careful study of the extreme cases in the *abc*-Theorem for polynomials.

Moret-Bailly, building on ideas of Szpiro, went a step further. He showed that if one could get *good* upper bounds for the size of the co-ordinates of the rational points on[33] $y^2 = x^5 - x$ in any number field[34] then the *abc*-conjecture follows. ("Good" bounds, in this case, are bounds that depend explicitly on the discriminant of the number field over which the points are rational). Therefore, in a certain sense, this problem and the *abc*-conjecture are equivalent.

H6. The local-global principle. Currently section 9.3.

[32]Notice that this is not the full Fermat-Catalan conjecture, since here we have proved that there are only finitely many solutions for each *fixed* p, q, r (for which $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} < 1$), rather than there are only finitely many solutions, in total, over all possible p, q, r.

[33]Or, for the initiated, on any other smooth algebraic curve of genus > 1.

[34]That is, a finite field extension of $\mathbb{Q}$.

H7. Modularity and $e^{\pi\sqrt{163}}$. Jacobi's theta function is defined by

$$\theta(s) := \sum_{n\in\mathbb{Z}} e^{i\pi n^2 s} \quad \text{for all } s \quad \text{for which } \mathrm{Im}(s) > 0.$$

Jacobi showed an extraordinary relationship:

$$\theta(-1/s) \;=\; (is)^{1/2}\,\theta(s).$$

This was the basis for Riemann's ability to analytically continue the Riemann zeta-function to the whole complex plane, and to prove that the value of resulting function at $1-s$ can be easily expressed in terms of its value at s.

`Exercise` Prove that we also have $\theta(s+2) = \theta(s)$.

There are other functions that also satisfy such equations, which we now briefly discuss:

What functions of the reals are periodic? That is satisfy $f(x) = f(x+n)$ for all x. The trigonometric functions are examples, and it can be shown that all such functions are rational functions in the basic trigonometric functions.

One way to view this periodicity is that the function stays constant under the map $x \to x+n$.

Jacobi's result shows that $s^{1/4}\theta(s)$ stays constant under the map $s \to 1/s$. Note though that here θ is a function defined on a half plane.

One can ask whether there are any functions that satisfy both? Simply can we find a function f such that $f(s)$ stays fixed under the map $s \to s+1$ and under the map $s \to -1/s$, and hence under all the possible compositions of the two maps. Notice that we have seen this pair of maps before when we were dealing with binary quadratic forms. Together they generate all of $SL(2,\mathbb{Z})$. So we want functions $f(s)$ that stay constant under the action of $SL(2,\mathbb{Z})$. This is a guess in it turns out to not quite be correct. What we really want is something, like $\theta(z)$, which is easily understood under such transformations. Strangely we work with functional equations like

$$f\left(\frac{at+b}{ct+d}\right) = (ct+d)^k f(t).$$

`Exercise` Verify that if this holds for $f(-1/t)$ and $f(t+1)$ then it holds for all $f\left(\frac{at+b}{ct+d}\right)$.