Random Algorithms

Bernd Souvignier

Autumn 2003

Contents

Chapter 1

Random number generators

Literature:

Donald E. Knuth: *The Art of Computer Programming.* Volume 2, Chapter 3. Addison-Wesley, 1998.

1.1 What is randomness

1.2 Linear congruential method

Chapter 2

Primality testing and proving

There is no point in arguing for the importance of prime numbers in mathematics. They play a crucial role in many areas of pure mathematics, but likewise have found their way into applications, e.g. in cryptography.

Of course there are infinitely many prime numbers, so finding one should not be too hard. The *prime number theorem* asserts that the number of primes up to N is approximately $\frac{N}{\log(N)}$, hence the chance that a random number picked is prime is about $\frac{1}{\log(N)}$, which is not bad at all. The only problem is that we need a means to distinguish prime numbers from non-prime numbers. Methods to accomplish this are the issue of this section.

2.1 Elementary methods

2.1.1 Trial division

The most straightforward method to test whether a number N is prime is to try to divide it by smaller numbers. Clearly, if N is not prime it has a divisor $d \leq \sqrt{N}$, hence it is enough to try candidates up to $\sqrt{N}$. This provides us with our first test:

$$N \text{ is prime if and only if } N \not\equiv 0 \bmod d \text{ for all } 2 \leq d \leq \sqrt{N}.$$

Of course we do not have to test all d, we can restrict ourselves to prime numbers up to $\sqrt{N}$. However, a list of primes may not be readily available and computing such a list for each primality test (e.g. using the sieve of Eratosthenes) is fairly expensive. Computer algebra systems like Magma use lists of primes up to certain bounds, but usually not beyond 10^6 (note that from the prime number theorem we know that there are about $5 \cdot 10^8$ prime numbers with less than 10 digits, and that is already a very long list).

A compromise between testing all numbers up to $\sqrt{N}$ and only prime numbers is to test all numbers which are not multiples of 2, 3 or 5 (after testing $2, 3, 5$ themselves). It is very easy to implement running only over these numbers and as a result we only have to test $(1 - 1/2 - 1/3 - 1/5 + 1/6 + 1/10 + 1/15 - 1/30)\sqrt{N} = 4/15\sqrt{N}$ numbers, which gives a speed-up of 3.75. Still, even with this speed-up, trial division does not lead far.

Assume that we can perform 10^6 modulo operations per second. Then we can do $8.64 \cdot 10^{10}$ modulo operations in a day. Thus, testing all numbers up to $\sqrt{N}$ we can deal with $N < 8 \cdot 10^{21}$ in a day. With our speed-up, we can push $\sqrt{N}$ somewhat further, because we only test 4/15 of the numbers. We can handle $\sqrt{N} < 3.24 \cdot 10^{11}$ and thus $N < 10^{23}$. Finally, if we indeed only test prime numbers, there are about $8.64 \cdot 10^{10}$ prime numbers up to $2.5 \cdot 10^{12}$, hence we could reach $N < 6 \cdot 10^{24}$ if we only test true prime numbers up to $\sqrt{N}$.

It is clear that a faster computer (with modulo operations implemented in its hardware) and waiting somewhat longer does not yield any true progress, we can not expect to handle numbers of more than 30 digits by trial division.

2.1.2 Fermat's theorem

The consequence of the 'failure' with trial division is to try to distinguish prime numbers from non-prime numbers by their properties.

One such property is derived from Fermat's theorem:

$$a^{p-1} \equiv 1 \bmod p \text{ for } p \text{ prime and } a \text{ with } \gcd(a,p) = 1.$$

As an immediate consequence of this theorem we get a powerful compositeness test:

$$\text{if } a^{N-1} \not\equiv 1 \bmod N \text{ for some } a \text{ with } \gcd(a,N) = 1, \text{ then } N \text{ is not prime.}$$

Although applying the Fermat test for a couple of values of a will usually detect the compositeness of N, there are composite numbers N which pass the Fermat test for every a. These numbers are called *Carmichael numbers*. It has recently been proved that Carmichael numbers are not as sparse as one might hope, up to a bound B there are more than $B^{2/7}$ Carmichael numbers (for B sufficiently large).

2.1.1 Proposition *A number N is a Carmichael number if and only if $N = \prod_{i=1}^{k} p_i$ with p_i distinct odd prime numbers, $(p_i - 1)|(N-1)$ and $k \geq 3$.*

Proof: $\Leftarrow$: If $\gcd(a,N) = 1$, then clearly $\gcd(a,p_i) = 1$, hence $a^{p_i-1} \equiv 1 \bmod p_i$ and since $(p_i - 1)|(N-1)$ we have $a^{N-1} \equiv 1 \bmod p_i$. Since the p_i are distinct primes, the Chinese remainder theorem implies that $a^{N-1} \equiv 1 \bmod N$.
$\Rightarrow$: If N is even, we have $(-1)^(N-1) = -1 \not\equiv 1 \bmod N$, since $N > 2$ (being composite). Thus, N has to be odd.

Next, assume that $N = pq$ for two primes $p < q$. In this case we can never have $(q-1)|(N-1)$, since $b(q-1) = pq - 1$ implies that $b \equiv 1 \bmod q$ (after taking both sides modulo q). Clearly, $b = 1$ is impossible and for $b = q+1$ we get $b(q-1) = q^2 - 1 > pq - 1$.

Now let p be a prime such that $p^2|N$, then $(p-1)p|\varphi(N)$, hence $(\mathbb{Z}/N\mathbb{Z})^*$ contains an element a of multiplicative order p. We have $a^p \equiv 1 \bmod N$ and by assumption $a^{N-1} \equiv 1 \bmod N$, hence $\gcd(p, N-1) \neq 1$, which is a contradiction.

We have shown that N is a product of at least three distinct odd primes p_i. We now choose a primitive element for each $\mathbb{Z}/p_i\mathbb{Z}$ (i.e. an element of multiplicative order $p_i - 1$)

and via the Chinese remainder theorem find a such that a is a primitive element for all the $\mathbb{Z}/p_i\mathbb{Z}$. By assumption we have $a^{N-1} \equiv 1 \bmod N$ and thus a fortiori $a^{N-1} \equiv 1 \bmod p_i$. But $p_i - 1$ is the multiplicative order of a modulo p_i, hence we have $(p_i - 1)|(N-1)$. □

In the proof of the above theorem we have already used a stronger property of prime numbers than Fermat's theorem, namely that the multiplicative group $(\mathbb{Z}/N\mathbb{Z})^*$ is a cyclic group of order $N-1$ if and only if N is a prime number. What we check in the Fermat test is that all elements in $(\mathbb{Z}/N\mathbb{Z})^*$ have order dividing $N-1$. But that only means that the least common multiple of the orders of elements in $(\mathbb{Z}/N\mathbb{Z})^*$ (also called the exponent of the group) divides $N-1$.

In order to prove that N is prime it is sufficient to prove that the order of the group $(\mathbb{Z}/N\mathbb{Z})^*$ is $N-1$ and that is the case if and only if there exists an element of order $N-1$ (since the group then is cyclic). The problem is to show that the order of an element is precisely $N-1$ and not a proper divisor of $N-1$.

One situation in which this can be efficiently tested is the case that the prime factors of $N-1$ are known. Then we have the following test (due to Lucas, Lehmer, Kraitchik):

2.1.2 Proposition *Let $N-1 = \prod_{i=1}^{k} p_i^{e_i}$ with p_i distinct primes and $e_i \geq 1$. If for every i there is an element a_i such that $a_i^{N-1} \equiv 1 \bmod N$ and $a_i^{(N-1)/p_i} \not\equiv 1 \bmod N$, then N is a prime number.*

Proof: The fact that $a_i^{N-1} \equiv 1 \bmod N$ shows that the order of a_i divides $N-1$ and is thus of the form $\prod_{i=1}^{k} p_i^{f_i}$ with $f_i \leq e_i$. Since $a_i^{(N-1)/p_i} \not\equiv 1 \bmod N$, if follows that $f_i = e_i$. This shows that $p_i^{e_i}$ divides the order of $(\mathbb{Z}/N\mathbb{Z})^*$. We therefore conclude that $N-1$ divides the order of $(\mathbb{Z}/N\mathbb{Z})^*$ which shows that N is prime. □

2.2 Probabilistic primality tests

Literature:

> E. Bach, J. Shallit: *Algorithmic Number Theory. Volume I: Efficient Algorithms* The MIT Press, 1996.

The idea of probabilistic primality tests is to provide a method which guarantees the primality of a number to a chosen level of certainty. Numbers certified by such a test are often called *(strong) pseudo-primes*.

2.2.1 Miller-Rabin test

We still can look a bit more closely at the properties of $\mathbb{Z}/N\mathbb{Z}$ in the case that N is a prime number. We know that its multiplicative group is a cyclic group of order $N-1$ and that therefore all elements have order dividing $N-1$. But $\mathbb{Z}/N\mathbb{Z}$ is also a field and therefore polynomials have at most as many roots as their degree. In particular, the polynomial

$X^2 - 1$ has only the roots $+1$ and -1. Thus, if we have an element of even multiplicative order, then its square root has to be -1.

Checking this property is the idea of the Miller-Rabin test.

2.2.1 Miller-Rabin test Let $N = 2^s d + 1$ with d odd. If N is prime, then any $a \in (\mathbb{Z}/N\mathbb{Z})^*$ has order dividing $2^s d$ and therefore fulfills either:

(i) $a^d \equiv 1 \bmod N$ or

(ii) $(a^d)^{2^i} \equiv -1 \bmod N$ for some $0 \leq i < s$.

If neither of these is fulfilled, a is a witness for the compositeness of N.

The interesting thing about the Miller-Rabin test is that there is a general estimate for the number of witnesses for a composite N. There does not exist anything like Carmichael numbers for which there are no or only few witnesses.

2.2.2 Theorem *Let $N \neq 9$ be an odd composite number. Then the number of $a \in (\mathbb{Z}/N\mathbb{Z})^*$ passing the Miller-Rabin test is at most $\frac{1}{4}\varphi(N)$.*

Proof: We have to estimate the number of $a \in (\mathbb{Z}/N\mathbb{Z})^*$ with $a^d \equiv 1 \bmod N$ or $(a^d)^{2^i} \equiv -1 \bmod N$ for composite N. Let r be the largest i such that there exists an element a with $(a^d)^{2^r} \equiv -1 \bmod N$, then $r \leq s - 1$. Define $m := 2^r d$, then $m|(N-1)$ and $2m|(N-1)$. Write N as $N = \prod_{i=1}^{k} p_i^{e_i}$ with p_i distinct primes. We now define the following subgroups of $(\mathbb{Z}/N\mathbb{Z})^*$:
$U_1 := \{a \in (\mathbb{Z}/N\mathbb{Z})^* \mid a^{N-1} \equiv 1 \bmod N\}$ (elements of order dividing $N-1$)
$U_2 := \{a \in (\mathbb{Z}/N\mathbb{Z})^* \mid a^m \equiv \pm 1 \bmod p_i^{e_i} \text{ for all } p_i\}$ (elements of order dividing m or $2m$ in $(\mathbb{Z}/p_i^{e_i}\mathbb{Z})^*$)
$U_3 := \{a \in (\mathbb{Z}/N\mathbb{Z})^* \mid a^m \equiv \pm 1 \bmod N\}$ (elements of order dividing m or $2m$).
It is clear that $U_3 \leq U_2$, since the order of an image in a factor group divides the order in the full group. Since $2m|(N-1)$ it follows via the Chinese remainder theorem that $U_2 \leq U_1$.

U_3 is defined such that the elements a which pass the Miller-Rabin test (i.e. the non-witnesses for the compositeness of N) lie in U_3.

We now use the correspondence between $\mathbb{Z}/N\mathbb{Z}$ and $\mathbb{Z}/p_1^{e_1}\mathbb{Z} \times \ldots \times \mathbb{Z}/p_k^{e_k}\mathbb{Z}$ to show that $[U_2 : U_3] = 2^{k-1}$. The group U_3 contains those elements a for which a^m corresponds to either $(+1, \ldots, +1)$ or $(-1, \ldots, -1)$. On the other hand U_2 contains those a for which a^m corresponds to an arbitrary $(\pm 1, \ldots, \pm 1)$ combination.

If $k \geq 3$ we are done, since we have $[U_2 : U_3] \geq 4$. If $k = 1$ we have $|U_1| = p - 1$ and $[(\mathbb{Z}/N\mathbb{Z})^* : U_1] = p^{e-1}$. Since N is an odd composite, we have $p > 2$ and $e \geq 2$, hence we are done except for $N = 9$, which we luckily excluded. Finally, if $k = 2$ we know that N is not a Carmichael number, thus there are elements of $(\mathbb{Z}/N\mathbb{Z})^*$ not lying in U_1, hence $[(\mathbb{Z}/N\mathbb{Z})^* : U_1] \geq 2$ and thus $[(\mathbb{Z}/N\mathbb{Z})^* : U_3] \geq 4$. □

The way to apply the Miller-Rabin test is straightforward: Test n randomly chosen $a \in (\mathbb{Z}/N\mathbb{Z})^*$ and check whether they pass the Miller-Rabin test. If so, N is prime with probability $\geq 1 - (1/4)^n$. Note that it is important to choose the elements a randomly, since for a given sequence $a_1, \ldots, a_n$ of test elements it is possible to construct a composite number N for which these a_i pass the Miller-Rabin test.

2.2.2 Quadratic reciprocity

A further test in the flavour of the Miller-Rabin test is the Solovay-Strassen test which uses the notion of quadratic residues. Like the Miller-Rabin test it has a global bound on the number of non-witnesses for a composite number.

We will see that the Miller-Rabin test is strictly stronger than the Solovay-Strassen test, i.e. a composite number passing the Miller-Rabin test for a certain a will also pass the Solovay-Strassen test for the same a. However, historically the Solovay-Strassen test was the first test with a global bound on the number of non-witnesses, so we regard it worthwhile to consider it here.

The Solovay-Strassen test is based on investigating *quadratic residues* modulo p. We call an integer a a quadratic residue modulo p if there exists some integer b such that $b^2 \equiv a \bmod p$. Quadratic residues are most easily handled by the *Legendre symbol* and its generalization to non-prime numbers, the *Jacobi symbol*.

2.2.3 Definition Let p be an odd prime, a an integer and $N = \prod_{i=1}^{k} p_i$ with p_i (not necessarily distinct) primes.

(i) The *Legendre symbol* $\left(\frac{a}{p}\right)$ is defined as

$$\left(\frac{a}{p}\right) := \begin{cases} 0 & \text{if } p \mid a; \\ 1 & \text{if } a \text{ is a quadratic residue modulo } p; \\ -1 & \text{if } a \text{ is not a quadratic residue modulo } p. \end{cases}$$

(ii) The *Jacobi symbol* $\left(\frac{a}{N}\right)$ is defined as

$$\left(\frac{a}{N}\right) := \prod_{i=1}^{k} \left(\frac{a}{p_i}\right)$$

where the product runs over the Legendre symbols. In particular, the Jacobi symbol coincides with the Legendre symbol for N prime (which justifies using the same symbol). Note that the interpretation with respect to quadratic residues does *not* hold for the Jacobi symbol.

The property of the Legendre symbol which we will use as a test for primality is the following:

2.2.4 Proposition *For an integer a and a prime number p we have*

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

PROOF: Let g be a primitive element of $\mathbb{Z}/p\mathbb{Z}$, i.e. an element of order $p-1$ in $(\mathbb{Z}/p\mathbb{Z})^*$. Then the squares modulo p are precisely the even powers of g, hence $\left(\frac{a}{p}\right) = 1 \Leftrightarrow a = g^{2m} \Leftrightarrow a^{(p-1)/2} = (g^{p-1})^m = 1$. Similarly, the non-squares are the odd powers of g, hence we have $\left(\frac{a}{p}\right) = -1 \Leftrightarrow a = g^{2m+1} \Leftrightarrow a^{(p-1)/2} = (g^{p-1})^m \cdot g^{(p-1)/2} = -1$, since $\mathbb{Z}/p\mathbb{Z}$ is a field. □

The most important properties of the Legendre symbol are summarized in the following theorem:

2.2.5 Theorem *Let p and q be odd primes and let a, b be integers.*

(i) $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$ *if* $a \equiv b \bmod p$.

(ii) $\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right) \cdot \left(\frac{b}{p}\right)$.

(iii) $\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}} = \begin{cases} 1 & \text{if } p \equiv \pm 1 \bmod 8 \\ -1 & \text{if } p \equiv \pm 3 \bmod 8. \end{cases}$

(iv) **Quadratic reciprocity law:** $\left(\frac{p}{q}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}\left(\frac{q}{p}\right)$. *This means that* $\left(\frac{p}{q}\right) = \left(\frac{q}{p}\right)$ *if* $p \equiv 1 \bmod 4$ *or* $q \equiv 1 \bmod 4$ *and* $\left(\frac{p}{q}\right) = -\left(\frac{q}{p}\right)$ *if* $p \equiv q \equiv 3 \bmod 4$.

PROOF: (i): This follows immediately from the definition.
(ii): This follows from Proposition 2.2.4.
(iii): Let $\alpha := \zeta_8 + \zeta_8^{-1}$, then $\alpha^2 = 2$, since $\zeta_8^2 + \zeta_8^{-2} = 0$. We have $\alpha^p \equiv \zeta_8^p + \zeta_8^{-p} \bmod p$, hence $\alpha^p \equiv \alpha \bmod p$ if $p \equiv \pm 1 \bmod 8$ and $\alpha^p \equiv -\alpha \bmod p$ if $p \equiv \pm 3 \bmod 8$. By Proposition 2.2.4 we have $\left(\frac{2}{p}\right) = 2^{(p-1)/2} = (\alpha^2)^{(p-1)/2} = \alpha^{p-1}$, hence $\left(\frac{2}{p}\right) = 1$ if $p \equiv \pm 1 \bmod 8$ and $\left(\frac{2}{p}\right) = -1$ if $p \equiv \pm 3 \bmod 8$.
(iv): Let $\zeta := \zeta_p := e^{\frac{2\pi i}{p}}$ be a primitive p-th root of unity and define α as the Gauss sum $\alpha := \sum_{k=1}^{p-1} \left(\frac{k}{p}\right)\zeta^k$. We have

$$\alpha^2 = \sum_{k,l} \left(\frac{k}{p}\right)\left(\frac{l}{p}\right)\zeta^{k+l} = \sum_{k,l} \left(\frac{kl}{p}\right)\zeta^{k+l}.$$

Instead of k we can let kl run from 1 to $p-1$ and substituting kl for k gives

$$\alpha^2 = \sum_{k,l} \left(\frac{kl^2}{p}\right)\zeta^{kl+l} = \sum_{k,l} \left(\frac{k}{p}\right)\zeta^{l(k+1)} = \sum_{l} \left(\frac{-1}{p}\right) + \sum_{k \neq p-1} \left(\frac{k}{p}\right) \underbrace{\left(\sum_{l} \zeta^{l(k+1)}\right)}_{=-1}.$$

Since there are as many squares as non-squares between 1 and $p-1$ we have $\sum_{k=1}^{p-1}\left(\frac{k}{p}\right)=0$, hence $\sum_{k\neq p-1}\left(\frac{k}{p}\right)=-\left(\frac{-1}{p}\right)$ and we thus obtain

$$\alpha^2=\sum_l\left(\frac{-1}{p}\right)-\sum_{k\neq p-1}\left(\frac{k}{p}\right)=(p-1)\left(\frac{-1}{p}\right)-(-\left(\frac{-1}{p}\right))=p\left(\frac{-1}{p}\right).$$

From this we obtain as a side result that $\alpha=\sqrt{p}$ if $p\equiv 1 \bmod 4$ and $\alpha=\sqrt{-p}$ if $p\equiv 3 \bmod 4$.

We have $\alpha^q\equiv\sum_{k=1}p-1\left(\frac{k}{p}\right)\zeta^{qk}\equiv\left(\frac{q}{p}\right)\sum_{k=1}p-1\left(\frac{qk}{p}\right)\zeta^{qk}\equiv\left(\frac{q}{p}\right)\alpha \bmod q$. From this we conclude that

$$\left(\frac{-1}{p}\right)^{(q+1)/2}p^{(q+1)/2}=(\alpha^2)^{(q+1)/2}=\alpha^{q+1}\equiv\left(\frac{q}{p}\right)\alpha^2=\left(\frac{q}{p}\right)\left(\frac{-1}{p}\right)p \bmod q.$$

This shows that

$$\left(\frac{-1}{p}\right)^{(q-1)/2}p^{(q-1)/2}=\left(\frac{-1}{p}\right)^{(q-1)/2}\left(\frac{p}{q}\right)\equiv\left(\frac{q}{p}\right) \bmod q,$$

which finally gives $((-1)^{(p-1)/2})^{(q-1)/2}\left(\frac{p}{q}\right)=\left(\frac{q}{p}\right)$. □

The quadratic reciprocity law allows to compute the Legendre symbol very efficiently in a way similar to the Euclidean algorithm. Here is an example:

$$\left(\frac{76}{131}\right)=\left(\frac{2}{131}\right)^2\left(\frac{19}{131}\right)=-\left(\frac{131}{19}\right)=-\left(\frac{17}{19}\right)=-\left(\frac{19}{17}\right)=-\left(\frac{2}{17}\right)=-1.$$

The nice thing about the Jacobi symbol is that the properties of the Legendre symbol required to compute it still hold. The proof is left as an exercise, the crucial idea is to use the multiplicativity and to observe that $\sum\frac{p_i-1}{2}\equiv\frac{(\prod p_i)-1}{2} \bmod 2$.

2.2.3 Solovay-Strassen test

In contrast to the quadratic reciprocity law, Proposition 2.2.4 does not generalize to the Jacobi symbol in general, and this is the property used in the Solovay-Strassen test to distinguish prime and non-prime numbers.

The Solovay-Strassen test simply checks whether

$$a^{\frac{N-1}{2}}\equiv\left(\frac{a}{N}\right) \bmod N$$

for a randomly chosen integer a.

2.2.6 Theorem *Let N be an odd composite number. Then the number of $a\in(\mathbb{Z}/N\mathbb{Z})^*$ passing the Solovay-Strassen test is at most $\frac{1}{2}\varphi(N)$.*

PROOF: Let $U_1 := \{a \in (\mathbb{Z}/N\mathbb{Z})^* \mid a^{(N-1)/2} \equiv \left(\frac{a}{N}\right) \bmod N\}$. If we assume that $U_1 = (\mathbb{Z}/N\mathbb{Z})^*$, we have in particular that $a^{N-1} \equiv 1 \bmod N$, thus N is a Carmichael number. This means that $N = \prod_{i=1}^{k} p_i$ with $k \geq 3$ and p_i distinct primes. Using the Chinese remainder theorem we can choose an integer a such that a is a primitive element modulo p_1 and $a \equiv 1 \bmod p_i$ for $i \geq 2$. Then we have $\left(\frac{a}{N}\right) = \prod_{i=1}^{k} \left(\frac{a}{p_i}\right) = -1$. By our assumption we have $a^{(N-1)/2} \equiv -1 \bmod N$ and thus in particular $a^{(N-1)/2} \equiv -1 \bmod p_2$ which is a contradiction, since $(p_2 - 1)|(N-1)$ and $a^{(p_2-1)/2} = 1 \bmod p_2$. This shows that even for Carmichael numbers we have $(\mathbb{Z}/N\mathbb{Z})^* \neq U_1$, hence $[(\mathbb{Z}/N\mathbb{Z})^* : U_1] \geq 2$ in all cases. □

We close this section by proving that the Solovay-Strassen test is entirely superseded by the Miller-Rabin test. It is an exercise to show that for composite numbers N with $N \equiv 3 \bmod 4$ the two tests are equally strong, i.e. that the non-witnesses for the two tests are the same.

2.2.7 Theorem *Let N be a composite number and let a be a non-witness for the compositeness of N in the Miller-Rabin test. Then a is also a non-witness in the Solovay-Strassen test.*

PROOF: Let $N = \prod_{i=1}^{k} p_i^{e_i}$ with p_i distinct (odd) primes and let $N - 1 = 2^s d$ with d odd. For the non-witness a we either have $a^d \equiv 1 \bmod N$ or $(a^d)^{2^{r-1}} \equiv -1 \bmod N$ for some $1 \leq r \leq s$. If we set $r = 0$ in the first case we see that 2^r is the 2-part of the order of a in $(\mathbb{Z}/p_i^{e_i}\mathbb{Z})^*$. Since the 2-part of $(\mathbb{Z}/p_i^{e_i}\mathbb{Z})^*$ is already contained in $(\mathbb{Z}/p_i\mathbb{Z})^*$ we see that 2^r is also the 2-part of the order of a in $(\mathbb{Z}/p_i\mathbb{Z})^*$.

We now write $p_i - 1 = 2^{s_i} d_i$ with d_i odd, then $s_i \geq r$. Let m be the sum of the e_i for which $s_i = r$. We have $p_i = 1 + 2^{s_i} d_i \equiv 1 + 2^{s_i} \bmod 2^{s_i+1}$ and thus $N \equiv (1 + 2^r)^m \equiv 1 + m2^r \bmod 2^{r+1}$.

If m is odd, we necessarily have $r = s$, hence $a^{(N-1)/2)} \equiv -1 \bmod p_i^{e_i}$. If m is even, we conclude that $2^{r+1}|(N-1)$, hence $2^r|\frac{N-1}{2}$ and thus $a^{(N-1)/2)} \equiv 1 \bmod p_i^{e_i}$.

But we have $\left(\frac{a}{p_i}\right) = -1 \Leftrightarrow s_i = r$, since a is a non-square modulo p_i if and only if the order of a contains the full 2-part of $p_i - 1$. From this we see that $\left(\frac{a}{N}\right) = (-1)^m$ which shows that a is a non-witness in the Solovay-Strassen test. □

Chapter 3

Primality proving via elliptic curves

Literature:

Joseph H. Silverman: *The Arithmetic of Elliptic Curves.* Springer, 1986.

Henri Cohen: *A Course in Computational Algebraic Number Theory.* Springer, 1993.

3.1 Background on elliptic curves

In this section we will give a very brief introduction into the theory of elliptic curves, tailored towards their application in primality proving.

3.1.1 Definition An *elliptic curve* E over a field K is the set of pairs $(x, y) \in K \times K$ satisfying the cubic equation

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

where the coefficients a_1, a_2, a_3, a_4, a_6 lie in the algebraic closure $\overline{K}$ of K. If the a_i all lie in some field L, then the elliptic curve is said to be *defined over* L.

The above equation is called the *affine form* of the *Weierstrass equation* of the elliptic curve. Elliptic curves are also studied in the context of projective geometry, in which case the homogeneous form

$$Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3$$

of the Weierstrass equation is considered. The transfer between the two forms is given by the correspondence $(X, Y, Z) = (X/Z, Y/Z, 1) \leftrightarrow (x, y)$. The case $Z = 0$ implies $X = 0$, therefore the only point lost in the transfer from the projective to the affine form is the point $(0, 1, 0)$ which will be added in the affine case as the special point at infinity.

It can be seen that the Weierstrass equation of an elliptic curve can be simplified by some linear transformations in x and y. In the case that $\mathrm{char}(K) \neq 2, 3$ one obtains an equation of the form

$$y^2 = x^3 + Ax + B.$$

Two important invariants are used to characterize elliptic curves, which read as follows for the simplified form of the Weierstrass equation:

(i) the *discriminant* $\Delta := -16(4A^3 + 27B^2)$,

(ii) the *j-invariant* $j := -1728(4A)^3/\Delta$.

It can be shown that two elliptic curves are isomorphic over $\overline{K}$ if and only if they have the same j-invariant.

Similarly as the discriminant of a quadratic polynomial describes the set solutions (none, one or two), the discriminant of an elliptic curves expresses the following about the curve E:

(i) E is non-singular if and only if $\Delta \neq 0$. This means that every point on the curve is smooth, i.e. has a well-defined tangent.

(ii) E has a node if $\Delta = 0$ and $A \neq 0$. In this case there are two tangents with different slopes in the node.

(iii) E has a cusp if $\Delta = 0$ and $A = 0$. In this case the limits of the tangent slopes of the two branches running into the cusp are the same, i.e. the cusp has a well-defined direction.

3.1.1 The group structure

One of the many astonishing things about elliptic curves is that this set of solutions to a cubic equation carries a group structure. The basis of this fact is that a line through two points on the curve intersects the curve in a third point, due to the fact that projectively the curve is given as solutions of a homogeneous polynomial of degree 3. Note that tangent points have to be counted with multiplicities.

3.1.2 Theorem *(Group law)*
Let E be an elliptic curve and denote the point at infinity by O. Let P, Q be points on E and let L be the line connecting P and Q (tangent line in P if $P = Q$). Let R be the third point of intersection of L with E, let L' be the line connecting R and O and let S be the third point of intersection of L' with E. Then $P + Q := S$ defines a group law on E.

By the definition of the group law it is clear that E is an abelian group, since interchanging P and Q does not alter the line L. The coordinates of $P + Q$ can be given explicitly, e.g. for $P = (x_1, y_1)$, $Q = (x_2, y_2)$, $P \neq Q$ one has $P + Q = (x_3, y_3)$ with

$$x_3 = \frac{(x_2 - x_1)^2}{(y_2 - y_1)^2} - x_1 - x_2, \qquad y_3 = -y_1 + \frac{y_2 - y_1}{x_2 - x_1}(x_1 - x_3).$$

The hard part of the proof is the associative law. It can be checked (labouriously) using explicit formulae for the coordinates or (more smoothly) using some deeper theory (Riemann-Roch theorem).

3.1.2 Elliptic curves over $\mathbb{C}$

It should not come as a surprise that in the case of complex numbers elliptic curves are strongly related to elliptic functions. Recall that elliptic functions are defined as doubly periodic meromorphic functions, i.e. functions $f(z)$ on $\mathbb{C}$ with $f(z+w) = f(z)$ for all w in a lattice $L = \{aw_1 + bw_2 \mid a, b \in \mathbb{Z}, w_1/w_2 \notin \mathbb{R}\}$. An elliptic function can therefore be regarded as a function on the torus $\mathbb{C}/L$.

The most important example of an elliptic function is the *Weierstrass $\wp$-function* given as

$$\wp(z) := \frac{1}{z^2} + \sum_{w \in L \setminus \{0\}} (\frac{1}{(z+w)^2} - \frac{1}{w^2}$$

It is easily seen that $\wp$ is double periodic and one checks that it has double poles in the lattice points and no other poles.

From the definition it is clear that $\wp$ is an even function, therefore its derivative $\wp'$ is an uneven elliptic function. The reason that the $\wp$-function is the most important elliptic function lies in the fact that all elliptic functions for a given lattice L are rational combinations of $\wp$ and $\wp'$, i.e. of the form

$$\frac{f(\wp, \wp')}{g(\wp, \wp')}$$

for polynomials f, g with complex coefficients.

Having found an elliptic function for a lattice L we can now use the *Eisenstein weights* to obtain an elliptic curve. Let $G_{2k}(L) := \sum_{w \in L, w \neq 0} w^{-2k}$, then this series is absolutely convergent for $k > 1$. One now finds that

$$\wp(z) = z^{-2} + \sum_{k=1}^{\infty} (2k+1) G_{2k+2} z^{2k}.$$

Comparing coefficients in $\wp'(z)^2 - 4\wp(z)^3 + 60G_4\wp(z) + 140G_6$ one sees that this is a holomorphic function which is also elliptic and therefore has to be constant. Since it vanishes for $z = 0$ one has

$$\wp'(z)^2 = 4\wp(z)^3 + 60G_4\wp(z) + 140G_6 \text{ for all } z \in \mathbb{C}$$

We now define $g_2 := 60G_4$ and $g_3 := 140G_6$ and conclude that the points $(\wp(z), \wp'(z))$ satisfy the equation $y^2 = 4x^3 - g_2 x - g_3$.

The fact that we can associate an elliptic curve (over $\mathbb{C}$) with a lattice in $\mathbb{C}$ leads to an important new interpretation of the j-invariant. We start with some complex number $\tau in H$ in the upper halfplane and define the lattice $L := \mathbb{Z} + \tau\mathbb{Z}$. Then the associated elliptic curve E_τ satisfies $y^2 = 4x^3 - g_2 x - g_3$ and has discriminant $\Delta(E_\tau) = g_2^3 - 27g_3^2$ and j-invariant $1728 g_2^3/\Delta(E_\tau)$. We can therefore interpret $j : H \to \mathbb{C}$ as a function from the upper halfplane into $\mathbb{C}$ by defining $j(\tau) := j(E_\tau)$..

It is now a fairly deep result that in the case that τ is an element in an imaginary quadratic number field (i.e. of the form $a + b\sqrt{d}$ with $d < 0$), $j(\tau)$ is an algebraic number and the degree $[\mathbb{Q}(j(\tau)) : \mathbb{Q}]$ equals the ideal class number $h(\mathbb{Q}(\tau))$ of the imaginary quadratic field containing τ.

3.1.3 Endomorphisms

As with most algebraic structures one also studies elliptic curves in connection with maps between them. Since elliptic curves originally are defined as projective varieties satisfying a cubic polynomial, the properties of the mappings in question are motivated by the geometric background. One is interested in *isogenies* which are defined as *morphisms* mapping the origin (point at infinity) of one curve to the origin of the other curve. A morphism, in turn, is a *rational map* between two projective varieties which is regular in every point.

The important fact about elliptic curves is that the group law defines a morphism $+ : E \times E \to E, (P, Q) \mapsto P + Q$ from $E \times E$ to E.

For a single elliptic curve E, the morphisms from E to E are called *endomorphisms* and by pointwise addition and composition the endomorphisms of E form the *endomorphism ring* $End(E)$.

As a consequence of the fact that the group law is a morphism one sees that the *multiplication-by-m* map

$$[m] : P \mapsto \underbrace{P + \ldots + P}_{m}$$

is an endomorphism of E. This shows that the endomorphism ring $End(E)$ contains a subring isomorphic with $\mathbb{Z}$.

It now turns out that in many cases the multiplications by $m \in \mathbb{Z}$ are the only endomorphisms of an elliptic curve. For the precise statement we need some more terminology:

We say that Λ is an *order* in a $\mathbb{Q}$-algebra A if Λ is a subring and lattice in A. The latter means that there exists a $\mathbb{Q}$-basis B of A such that Λ is the $\mathbb{Z}$-span of this basis.

A definite quaternion algebra over $\mathbb{Q}$ is a 4-dimensional vector space $A = \mathbb{Q} + \mathbb{Q}\alpha + \mathbb{Q}\beta + \mathbb{Q}\alpha\beta$ with multiplication rules $\alpha^2 \in \mathbb{Q}$, $\beta^2 \in \mathbb{Q}$, $\alpha^2 < 0$, $\beta^2 < 0$, $\alpha\beta = -\beta\alpha$. The Hamilton-quaternions are an example of such an algebra. Note that quaternion algebras contain several imaginary quadratic number fields.

3.1.3 Theorem *The endomorphism ring of an elliptic curve E is isomorphic to one of:*

(i) $\mathbb{Z}$,

(ii) an order in an imaginary quadratic number field,

(iii) an order in a definite quaternion algebra over $\mathbb{Q}$.

If $End(E)$ is strictly larger than $\mathbb{Z}$ (i.e. in cases (ii) and (iii)) the elliptic curve E is said to have complex multiplication.

Case (iii) of the above theorem can only happen if $\mathrm{char}(K) > 0$, in this case the curve E is said to be *supersingular*. Over a finite fields, the cases (ii) and (iii) of the above theorem can be distinguished by the j-invariant of the curve. In case $j(E) \in \mathbb{F}_{p^2}$ the curve E is supersingular, otherwise, if $j(E) \in \overline{F_p}$, then $End(E)$ is an order in an imaginary quadratic number field.

3.1.4 The order of an elliptic curve

An interesting question about elliptic curves is to determine the structure of the group of the curve or at least the order of its torsion part. Over the complex numbers the elliptic curve is of the form C/L for a lattice L and is thus a torus. Over an algebraic number field K (thus in particular over $\mathbb{Q}$) the situation is more complicated. The Mordell-Weil theorem states that the group $E(K)$ is finitely generated and therefore is of the form $E(K) \cong E_{tors}(K) \times \mathbb{Z}^r$. The order of the torsion part over a fixed field K is bounded, for $K = \mathbb{Q}$ a theorem of Mazur states that $E_{tors}(\mathbb{Q})$ is one of C_m for $1 \leq m \leq 10$ or $m = 12$ or $C_2 \times C_{2m}$ for $1 \leq m \leq 4$. The rank r of a curve is much harder to determine and up to now there has no algorithms been found that computes r in general.

Over finite fields it is clear that the group of the elliptic curve is a finite group and it can be shown that it is either a cyclic group or a direct product of two cyclic groups, i.e. that $E(\mathbb{F}_q) \cong C_{d_1} \times C_{d_2}$ with $d_2 \mid d_1$. In the case that $E(\mathbb{F}_q)$ is not cyclic one furthermore knows that $d_2 \mid (q-1)$.

Even though the structure of the group over a finite field is very simple, it is by no means a simple question to determine its order. To estimate the order, one may argue as follows: For a given value $x \in \mathbb{F}_q$ the question is whether $x^3 + Ax + B$ is a square in $\mathbb{F}_q$ or not. If so, this x-value yields two points on the curve, if not there is no point with this x-value. On average, the values $x^3 + Ax + B$ will be equally distributed over squares and non-squares, thus one expects squares for half of the x-values. In total, this gives q points and adding the point at infinity one estimates $q + 1$ as the order of the elliptic curve.

It turns out that this heuristic argument gives a reasonable estimate, more precisely, Hasse's theorem states the following:

3.1.4 Theorem *(Hasse)*
Let E be an elliptic curve over $\mathbb{F}_q$, then the order of $E(\mathbb{F}_{q^r})$ is given by

$$|E(\mathbb{F}_{q^r})| = q^r + 1 - \pi^r - \overline{\pi}^r$$

where π is an imaginary quadratic integer such that $|\pi| = \sqrt{q}$.
In particular, $|E(\mathbb{F}_q) - (q+1)| \leq 2\sqrt{q}$.

Note that if we write $|E(\mathbb{F}_q)| = q + 1 - a$, then π is a root of the quadratic polynomial $X^2 - aX + q$.

We now have seen that elliptic curves have connections with imaginary quadratic fields in two ways: a curve may have complex multiplication with an order in such a field and

the order of the curve can be described via an element of norm q in an imaginary quadratic number field. The nice thing is that these two connections are interrelated in the following way:

3.1.5 Theorem *Let E be an elliptic curve with complex multiplication by an order in the imaginary quadratic number field of discriminant D.*

(i) If $\left(\frac{D}{p}\right) = -1$, then $|E(\mathbb{F}_p)| = p + 1$.

(ii) If there exists a prime element π in the maximal order A_d of $\mathbb{Q}(\sqrt{D})$ such that $\pi\overline{\pi} = p$, then $|E(\mathbb{F}_p)| = p + 1 - \pi - \overline{\pi}$.

In the second case the ideal $(p) \trianglelefteq A_d$ splits into a product $(p) = P_1 \cdot P_2$ of prime ideals with $P_1 = (\pi)$ and $P_2 = (\overline{\pi})$. Since the decomposition into prime ideals is unique, the element π generating P_1 is determined up to a unit in A_d. The unit group of $\mathbb{Q}(i)$ is generated by i, the unit group of $\mathbb{Q}(\zeta_3)$ by $-\zeta_3$, for all other imaginary quadratic number fields the unit group just consists of ± 1. In this last case the order of $E(\mathbb{F}_p)$ is either $p + 1 - \pi - \overline{\pi}$ or $p + 1 + \pi + \overline{\pi}$ for an element π of norm p.

3.2 Elliptic curves and primality

The basis for a primality test based on elliptic curves will be Hasse's theorem. The idea is to find a point of order too large to exist over all proper divisors of N. This is analogous to Pocklington's test which shows that the multiplicative group of $\mathbb{Z}/N\mathbb{Z}$ contains an element of an order that can only exist in the cyclic group of order $N - 1$.

The precise statement of the theorem is as follows:

3.2.1 Theorem *Let E be an elliptic curve defined over $\mathbb{Q}$ and let $N \in \mathbb{Z}$ not divisible by 2 or 3.*

(i) Assume there exist $m \in \mathbb{N}$, $q \mid m$ prime and a point $P \in E(\mathbb{Z}/N\mathbb{Z})$ such that

(a) $q > (\sqrt[4]{N} + 1)^2$,

(b) $m \cdot P = (0, 1, 0)$,

(c) $\frac{m}{q} \cdot P = (X, Y, Z)$ with $Z \in (\mathbb{Z}/N\mathbb{Z})^*$.

Then N is prime.

(ii) If $m = |E(\mathbb{Z}/N\mathbb{Z})|$ and if $q \mid m$ is a prime with $q > (\sqrt[4]{N} + 1)^2$, then a point P with the properties (b) and (c) above exists on $E(\mathbb{Z}/N\mathbb{Z})$.

PROOF: (i): Assume that N is composite, then N has a prime divisor p with $p \leq \sqrt{N}$. Since $Z \in (\mathbb{Z}/N\mathbb{Z})^*$, the reduction of $\frac{m}{q} \cdot P$ in $\mathbb{Z}/p\mathbb{Z}$ is also not 0, hence the order of P over $\mathbb{Z}/p\mathbb{Z}$ is a multiple of q. In particular we have

$$q \leq |E(\mathbb{Z}/p\mathbb{Z})| \leq p + 1 + 2\sqrt{p} = (\sqrt{p}+1)^2 \leq (\sqrt[4]{N}+1)^2 < q$$

which is a contradiction, hence such a prime divisor p of N can not exist and hence N is prime.
(ii): We use that $E(\mathbb{Z}/N\mathbb{Z}) \cong C_{d_1} \times C_{d_2}$, hence d_1 is the exponent of $E(\mathbb{Z}/N\mathbb{Z})$ (the least common multiple of the element orders) and $m = d_1 d_2$. If we now assume that $\frac{m}{q} \cdot P = (0,1,0)$ for all $P \in E(\mathbb{Z}/N\mathbb{Z})$ we have $d_1 \mid \frac{m}{q}$, hence $m = |E(\mathbb{Z}/N\mathbb{Z})| = d_1 d_2 \leq d_1^2 \leq (\frac{m}{q})^2$, hence $q^2 \leq m <\leq N + 1 + 2\sqrt{N} = (\sqrt{N}+1)^2$ and hence $(\sqrt[4]{N}+1)^2 < q \leq \sqrt{N}+1$, which is a contradiction. □

The above theorem can be slightly improved by replacing q in condition (a) by a composite number s and by changing condition (c) into $\frac{m}{q} \cdot P = (X,Y,Z)$ with $Z \in (\mathbb{Z}/N\mathbb{Z})^*$ for all prime divisors q of s. This condition suffices to show that s divides the order of P in the reduction modulo p, leading to the same contradiction as above. However, in applications m will usually be too big to obtain a complete factorization and q will be the unfactored part of m which is suspected to be prime by a Miller-Rabin test.

3.2.1 Goldwasser-Kilian test

The first practical algorithm to prove primality via elliptic curves was given by Goldwasser and Kilian in 1986. It was later improved by Atkin and Morain and can now deal with primes of more than 200 digits. We will first outline the algorithm and then comment on some crucial points.

3.2.2 Algorithm (Goldwasser-Kilian)

(1) choose $A, B \in \mathbb{Z}/N\mathbb{Z}$ randomly such that the discriminant $\Delta := 4A^3 + 27B^2 \in (\mathbb{Z}/N\mathbb{Z})^*$ and let E be the elliptic curve defined by $y^2 = x^3 + Ax + B$

(2) compute the order $m = |E(\mathbb{Z}/N\mathbb{Z})|$ of E using e.g. Schoof's algorithm

(3) split off small prime factors from m using trial division and (e.g.) the Pollard-ρ and Pollard-$(p-1)$ methods; use the Miller-Rabin test to check whether the remaining unfactored part q of m is prime and $> (\sqrt[4]{N}+1)^2$; if it is not prime, go back to step (1) and choose new coefficients A, B

(4) find a point P on the curve: choose x randomly and check whether $\left(\frac{x^3+Ax+B}{N}\right) \neq -1$; if so, compute a root y of $x^3 + Ax + B$, if not choose a new x

(5) test whether $m \cdot P = (0,1,0)$ and $\frac{m}{q} \cdot P \neq (0,1,0)$; if the first is not the case, N is not prime, if the second does not hold, go back to step (4) and select a new point P

(6) iterate the algorithm to prove that q found in step (3) is prime

Remarks:

(2) This is the main stumbling block in the algorithm. Computing the order of an arbitrary elliptic curve is still a computationally hard task, but progress has been made due to ideas of Atkin, Morain, Elkies and others.

(3) In the (unlikely) case that q is tested to be prime but is smaller than $(\sqrt[4]{N}+1)^2$, we can use the improved variant of theorem 3.2.1 to prove the primality of N.

(4) In the case that $\left(\frac{x^3+Ax+B}{N}\right) = 0$ we have $gcd(x^3+Ax+B, N) \neq 1$, which means that we either found a proper factor of N or that we can use the point $P = (x, 0, 1)$.
For the computation of a square root modulo N there are different possibilities:

(a) Cantor-Zassenhaus: We are looking for a root y of the polynomial $T^2 - (x^3 + Ax + B)$ which we know to exist modulo N by means of the Jacobi-symbol. Hence, the polynomial splits into linear factors and we can find a linear factor with chance 0.5 by computing $gcd(F^{(N-1)/2} - 1, T^2 - (x^3 + Ax + B))$ for random (linear) polynomials $F = T + c$.

(b) Tonelli-Shanks: Let $N - 1 = 2^s d$ with d odd, then $(\mathbb{Z}/N\mathbb{Z})^* \cong C_{2^s} \times C_d$ (note that we assume that N is prime). Let b be a non-square in $\mathbb{Z}/N\mathbb{Z}$, i.e. $\left(\frac{b}{N}\right) = -1$, then $g := b^d$ is a generator for the 2-Sylow subgroup C_{2^s} of $(\mathbb{Z}/N\mathbb{Z})^*$. We now can write $c := x^3 + Ax + B$ as $c = g^e h$ with $0 \leq e < 2^s$ and $h \in C_d$. Note that every element $h \in C_d$ is a square in $\mathbb{Z}/N\mathbb{Z}$, since $(h^{(d+1)/2})^2 = h^{d+1} = h$. Thus, e has to be even for c square. We now determine the 2-adic expansion of e as follows: let $e_i \equiv e \bmod 2^i$, then $e_1 = 0$ since e is even. Now assume that e_i has been determined. Then we check whether $(cg^{-e_i})^{\frac{N-1}{2^{i+1}}} = 1$. If so, set $e_{i+1} := e_i$, if not set $e_{i+1} := e_i + 2^i$ (note that $(cg^{-e_i})^{\frac{N-1}{2^{i+1}}}$ will always lie in the subgroup C_2 of C_{2^s}). We finally set $h := cg^{-e}$ and define $y := g^{e/2} h^{(d+1)/2}$, then $y^2 = g^e h = c$.

(6) The algorithm has to be iterated, since the unfactored part of the order of E is only prime with a certain probability. The sequence of numbers proven prime, together with the curves (given by their coefficients), their orders and the points used form a certificate for the primality of N.

3.2.2 Atkin's test

The bottleneck in the Goldwasser-Kilian test is the determination of the order m of the elliptic curve. To overcome this problem, Atkin suggested to reverse the process: for curves with complex multiplication in an imaginary quadratic field we can determine the order easily by finding a prime element π of norm N. The idea is thus to first select the endomorphism ring of the curve and then to determine a particular curve with that endomorphism ring.

Recall that the discriminant of an algebraic number field K is defined as $disc(K) := \det(tr(b_i b_j))$ where $(b_1, \ldots, b_n)$ is an integral basis of K (i.e. a basis of the maximal order of K). A different formulation is that $disc(K) = \det(\sigma_i(b_j))^2$ where σ_i denotes the n different embeddings of K into $\mathbb{C}$. For squarefree d the maximal order A_d of $\mathbb{Q}(\sqrt{d})$ is $A_d = \mathbb{Z}[\sqrt{d}]$ if $d \equiv 2, 3 \bmod 4$ and it is $A_d = \mathbb{Z}[\frac{1+\sqrt{d}}{2}]$ if $d \equiv 1 \bmod 4$. Therefore, the quadratic number field $\mathbb{Q}(\sqrt{d})$ has discriminant $D = d$ if $d \equiv 1 \bmod 4$ and $D = 4d$ if $d \equiv 2, 3 \bmod 4$.

3.2.3 Algorithm (Atkin)

(1) choose a discriminant $D < 0$ such that $\left(\frac{D}{N}\right) = 1$

(2) find a solution of $x^2 - Dy^2 = 4N$, then $\pi := (x + y\sqrt{D})/2$ is an element of norm N, i.e. we have $\pi\overline{\pi} = N$; if no solution exists, go back to step (1) and choose a new discriminant

(3) computer the order of E as $m = |E(\mathbb{Z}/N\mathbb{Z})| = N + 1 - a$ where $a = \pi + \overline{pi}$; split off small prime factors from m using trial division and (e.g.) the Pollard-ρ and Pollard-$(p-1)$ methods; use the Miller-Rabin test to check whether the remaining unfactored part q of m is prime and $> (\sqrt[4]{N} + 1)^2$; if no q is found, we can replace π by $-\pi$ and try the same for $m = N + 1 + a$; if no suitable q is found, go back to step (1) and choose a new discriminant D

(4) compute the minimal polynomial of the (complex) j-invariant $j(\tau)$ of E_τ for $\tau = \frac{D+\sqrt{D}}{2}$ (which amounts to the same as $\tau = \sqrt{d}$ for $d \equiv 2, 3 \pmod 4$ and $\tau = \frac{1+\sqrt{d}}{2}$ for $d \equiv 1 \pmod 4$) and find a root modulo N of this minimal polynomial (which is guaranteed to exist); let $c := \frac{j}{j-1728}$ and let g be a non-square modulo N; then $E_1 : y^2 = x^3 - 3cx + 2c$ and $E_2 : y^2 = x^3 - 3cg^2 + 2cg^3$ are representatives of the two classes of elliptic curves with complex multiplication in $\mathbb{Q}(\sqrt{d})$ and j-invariant $j(\tau)$

(5) find a point P on E_1 (as in the Goldwasser-Kilian algorithm); if $m \cdot P \neq (0, 1, 0)$ we are on the wrong curve and have to use E_2 instead; if $\frac{m}{q} \cdot P = (0, 1, 0)$, choose a new point, otherwise N is proven to be prime (subject to the primality of q)

(6) iterate the algorithm to prove that q found in step (3) is prime

Remarks:

(1) This condition means that the principal ideal (N) splits into a product of two prime ideals in $\mathbb{Q}(\sqrt{d})$. Note that $4d$ is a square in $\mathbb{Z}/N\mathbb{Z}$ if and only if d is a square. Since the complexity of the algorithm depends on the discriminant D, in particular on the ideal class number of $\mathbb{Q}(\sqrt{d})$, one usually starts with the discriminants belonging to quadratic number fields with class number 1, then goes to class number 2 and so on.

(2) We know that the principal ideal (N) splits into a product of two prime ideals P_1 and P_2. These prime ideals have norm N, since (N) has norm N^2 (the norm of an

ideal is its index in the maximal order A_d). Moreover, P_2 is the ideal containing the complex conjugates of the elements in P_1, and since the decomposition into prime ideals is unique this shows that P_1 and P_2 are the only ideals of norm N.

(a) If $d \equiv 2,3 \pmod 4$ choose x_0 such that $x_0^2 \equiv d \pmod N$. Now let L be the sublattice of A_d generated by N and $x_0 + \sqrt{d}$, then L has index N in A_d. Furthermore we see that $L \trianglelefteq A_d$, since $\sqrt{d} \cdot N = -x_0 \cdot N + N \cdot (x_0 + \sqrt{N}) \in L$ and $\sqrt{d} \cdot (x_0 + \sqrt{d}) = d + x_0\sqrt{d} = \frac{d-x_0^2}{N} \cdot N + x_0 \cdot (x_0 + \sqrt{d}) \in L$ (since $d - x_0^2$ is a multiple of N).

(b) If $d \equiv 1 \pmod 4$ choose x_0 such that x_0 is odd and $x_0^2 \equiv d \pmod N$ (if x_0 is even, then $N - x_0$ is odd and $(N - x_0)^2 \equiv x_0^2 \pmod N$). Then $\frac{x_0^2-d}{4N}$ is an integer.

Now let $b := \frac{x_0+\sqrt{d}}{2}$, then b is an integral element with norm $\frac{x_0^2-d}{4}$. Let L be the sublattice of A_d generated by N and b, then L has index N in A_d. Furthermore, $(1, \overline{b})$ form an integral basis of A_d, since $\overline{b} = -b + x_0$. To check that L is an ideal in A_d, we see that $\overline{b} \cdot N = (x_0 - b) \cdot N = x_0 \cdot N - N \cdot b \in L$ and $\overline{b} \cdot b = \frac{x_0^2-d}{4} = \frac{x_0^2-d}{4N} \cdot N \in L$.

In both cases we have explicitly found an ideal L of index N in A_d. In order to determine whether the L is a principal ideal we have to check whether there exists an element of norm N. If so, it has to be an element of minimal norm in L, since the norm of an element is the index of the principal ideal generated by it, i.e. $N(a) = [A_d : (a)]$. It therefore suffices to compute an element of minimal norm in L. If it has norm N, the ideal L is a principal ideal, otherwise it is not and there exists no element of norm N.

To find an element of minimal norm we apply *Gauss-reduction*. Let L be a 2-dimensional lattice generated by the vectors v and w with $\|v\| \leq \|w\|$. Choose $\lambda \in \mathbb{Z}$ such that $\|w - \lambda v\|$ is minimal. Then we can replace the basis (v, w) by the improved basis $(v, w - \lambda v)$. The value of λ is the integer closest to $\frac{\langle v,w \rangle}{\|v\|^2}$, where $\langle \cdot, \cdot \rangle$ denotes the inner product derived from the norm $\| \cdot \|$. The crucial observation is now:

Theorem: The vector v is a vector of minimal length if w can not be improved by this procedure.

Proof: Let $u = av + bw \in L$. Let $a = qb + r$ with $0 \leq r < |b|$ (if $b = 0$ then $\|u\| \geq \|v\|$ and we are done). Then $\|u\| = \|b(qv + w) + rv\| \geq |b|\|qv + w\| - r\|w\| \geq (|b| - r)\|v\| \geq \|v\|$, since by assumption $\|qv + w\| \geq \|w\| \geq \|v\|$ and $|b| - r \geq 1$.

(3) For the discriminants $D = -4$ and $D = -3$ we have even more possibilities for m, since the unit groups of these fields are bigger. For $D = -4$ we can try m of the form $m = N + 1 - (i^k\pi + (-i)^k\overline{\pi})$ for $0 \leq k < 4$ and for $D = -3$ we can try $m = N + 1 - (\zeta_6^k\pi + \zeta_6^{-k}\overline{\pi})$ for $0 \leq k < 6$.

(4) The minimal polynomial of $j(\tau)$ is called the *Hilbert class polynomial* and is known to have degree $h(D)$, where $h(D)$ is the ideal class number of $\mathbb{Q}(\sqrt{d})$. There are direct methods to obtain this Hilbert class polynomial without actually computing $j(\tau)$, but instead enumerating the reduced forms of discriminant D.

However, one can also approximate $j(\tau)$ numerically as follows: writing $q := e^{2\pi i\tau}$ one has
$g_2 = 60 \sum_{w\in L\setminus\{0\}} w^{-4} = \frac{1}{12}(2\pi)^4(1 + 240 \sum_{n\geq 1} \frac{n^3q^n}{1-q^n})$ and
$g_3 = 140 \sum_{w\in L\setminus\{0\}} w^{-6} = \frac{1}{216}(2\pi)^6(1 - 504 \sum_{n\geq 1} \frac{n^5q^n}{1-q^n})$.
From these values one computes $\Delta_\tau = g_2^3 - 27g_3^2$ and $j(\tau) = 1728g_2^3/\Delta_\tau$.

Having obtained a numerical approximation of $j(\tau)$ and knowing the degree $h(D)$ of the minimal polynomial of $j(\tau)$ one can find the minimal polynomial by lattice reduction methods. The idea is to look at the lattice of linear combinations of the powers of $j(\tau)$ which are close to 0. More precisely, one takes the lattice L_k of those $(c_0, \ldots, c_n) \in \mathbb{Z}^{n+1}$ (for $n = h(D)$) such that $\sum_{i=0}^n c_i \lfloor j(\tau)^i \cdot 10^k \rfloor \equiv 0 \pmod{10^k}$. This lattice has index 10^k in $\mathbb{Z}^{n+1}$ and a basis is $v_0 = (10^k, 0, \ldots, 0)$ and $v_i = (\lfloor j(\tau)^i \cdot 10^k \rfloor, 0, \ldots, -1, 0, \ldots)$, with the -1 in the $(i+1)$st component. Clearly, L_k contains as a 1-dimensional sublattice the multiples of the minimal polynomial of $j(\tau)$. The other elements, however, are *virtual* relations between the powers of $j(\tau)$ and will vanish on increasing k. Therefore, the minimal polynomial can be found as a vector of minimal length in L_k once k is large enough in order to make the norms of each virtual dependency several orders of magnitude larger than the norm of the minimal polynomial.

Chapter 4

Galois groups

Literature:

B.L. van der Waerden *Algebra I.* Springer, 1971.

Emil Artin: *Galois Theory.* Dover, 1998.

This chapter is devoted to certain aspects of Galois theory, in particular the determination of Galois groups. Galois theory describes the intermediate fields in a field extension in terms of the groups of field automorphisms in a most elegant manner. This topic stands at the crib of (abstract) group theory and has connections to classical problems like the construction of regular n-gons and the solution of equations by radicals.

4.1 Field extensions

In this section we will recapitulate some basic facts about field extensions.

A *field extension* L/K is a pair of fields $L \supseteq K$ and the larger field L can be viewed as a K-vector space (since the field axioms imply the vector space axioms). The dimension of this vector space is called the *degree* of the extension, denoted by $[L : K]$ and the extension is called finite if this dimension is finite.

The *degree theorem* says that for finite field extensions L/M and M/K the extension L/K is also finite with degree $[L : K] = [L : M] \cdot [M : K]$.

We will only be concerned with *algebraic* field extensions, i.e. extensions in which every element is a root of a monic polynomial with integral coefficients. Whereas finite extensions are always algebraic, the opposite is not necessarily true, for example the algebraic closure $\overline{\mathbb{Q}}$ of the rationals (consisting of all $\alpha \in \mathbb{C}$ which are algebraic over $\mathbb{Q}$) has infinite degree.

If a field extension L/K is of the form $L = K(\alpha)$ we call it a *simple field extension.*

A simple field extension $L = K(\alpha)$ is characterized by the minimal polynomial of α. The other way round, we can also start with an irreducible polynomial f and construct a simple extension from it: By *Kronecker's theorem* we know that $L := K[X]/(f)$ is a simple field extension of degree $\deg(f)$ generated by the root $X + (f)$ of f. Iteration of

this construction shows the existence of a splitting field for an arbitrary polynomial f over K (recall that a splitting field of a polynomial f is defined to be a minimal field in which f splits into linear factors, i.e. a field obtained by adjoining only roots of f). In order to establish uniqueness (up to isomorphism) of the splitting field we need some additional results which are interesting in their own right.

4.1.1 Theorem *Let $\varphi : K \to K'$ be a field isomorphism, let $f \in K[X]$ be irreducible and let $\alpha \in L \supseteq K$ be a root of f in some field extension. If $\alpha' \in L' \supseteq K'$ is a root of $\varphi(f)$ (where φ is applied to the coefficients of f), then there exists an isomorphism $\tilde{\varphi} : K(\alpha) \to K'(\alpha')$ which extends φ and maps α to α'.*

PROOF: Since $K(\alpha)$ is generated by K and α there is actually no choice in the definition of $\tilde{\varphi}$: we require $\tilde{\varphi}(c) = \varphi(c)$ for $c \in K$ and $\tilde{\varphi}(\alpha) = \alpha'$. To show that this defines an isomorphism note that the mapping $\sum c_i \alpha^i \mapsto \sum \varphi(c_i)\alpha'^i$ induces an isomorphism $K(\alpha) \cong K[X]/(f) \to K'(\alpha') \cong K'[X]/(\varphi(f))$, which is precisely $\tilde{\varphi}$. □

If we apply this theorem with $\varphi = id_K$ we see that two roots α, β of an irreducible polynomial f generate isomorphic extensions $K(\alpha) \cong K(\beta)$.

4.1.2 Theorem *Let $\varphi : K \to K'$ be a field isomorphism, let $f \in K[X]$ and let $L = K(\alpha_1, \ldots, \alpha_n)$ be a splitting field of f over K. If $L' = K'(\beta_1, \ldots, \beta_n)$ is a splitting field of $\varphi(f)$ over K', then there exists an extension $\tilde{\varphi}$ of φ such that $\tilde{\varphi} : L \to L'$ is an isomorphism and $\tilde{\varphi}(\alpha_i) = \beta_{\pi(i)}$ for a suitable permutation π.*

PROOF: Let f_1 be an irreducible factor of f, then we can assume the α_i to be numbered such that $f_1(\alpha_1) = 0$. We have $\varphi(f)(\beta_j) = 0$ for a suitable j and by theorem 4.1.1 we can extend φ to $\hat{\varphi} : K(\alpha_1) \to K'(\beta_j)$. By induction over the number of roots of f not contained in K we can extend $\hat{\varphi}$ to an isomorphism $\tilde{\varphi}$ of L. □

Applying this theorem with $\varphi = id_K$ shows the uniqueness of the splitting field. Note, however, that intermediate fields between K and the splitting field L may well be non-isomorphic, only adjunction of all roots makes L unique.

We close this section with the important result that under mild assumptions all finite field extensions are actually simple extensions. For that we need the notion of *separable* elements: An algebraic element α is called *separable over K* if its minimal polynomial does not have double roots. It is easy to see that the minimal polynomial of an inseparable element has derivative 0, which shows that in case that $\mathrm{char}(K) = 0$ or $|K| < \infty$ every element is separable. For an infinite field K of characteristic p it is sufficient that K contains a p-th root of each element in order to conclude that every finite extension is separable.

An example of an inseparable element is a root of the Eisenstein polynomial $X^p - t$ over the field $\mathbb{F}_p(t)$ of rational functions over $\mathbb{F}_p$.

4.1.3 Theorem *(Primitive element)*
Let $L = K(\alpha_1, \ldots, \alpha_n)$ with $[L : K] < \infty$ and assume that $\alpha_2, \ldots, \alpha_n$ are separable over K. Then there exists an element $\theta \in L$ such that $L = K(\theta)$.

PROOF: By an obvious induction it is sufficient to prove the theorem for the case $L = K(\alpha, \beta)$ with β separable, since $K(\alpha_1, \ldots, \alpha_{n-1}) = K(\theta)$ implies $L = K(\theta, \alpha_n)$. Furthermore we can assume that $|K| = \infty$, since the theorem clearly holds for finite fields (being generated by a $(q-1)$st root of unity).

Now let f_α and f_β be the minimal polynomials of α and β, respectively, and denote the distinct roots of these polynomials (in some splitting field) by $\alpha = \alpha_1, \alpha_2, \ldots, \alpha_r$ and $\beta = \beta_1, \beta_2, \ldots, \beta_s$. Since β is separable we have $\beta_k \neq \beta$ for $k \neq 1$. Therefore, each of the equations $\alpha_i - \alpha_1 = (\beta_1 - \beta_k)x$ has at most one solution in K (the solution possibly lying in an extension). Since K is infinite we can choose $c \in K$ which does not solve any of these equations and define $\theta := \alpha_1 + c\beta_1 = \alpha + c\beta$. We have $f_\alpha(\theta - c\beta) = 0 = f_\beta(\beta)$, therefore β is a common root of the polynomials $f_\alpha(\theta - cX)$ and $f_\beta(X)$. Since $\theta - c\beta_k \neq \alpha_i$ for all $k \neq 1$ we know that β is the only common root of the two polynomials. This shows that $\gcd(f_\alpha(\theta - cX), f_\beta(X)) = X - \beta$, on the other hand the *gcd* lies in $K(\theta)[X]$, hence $\beta \in K(\theta)$. But this implies that $\alpha = \theta - c\beta \in K(\theta)$ and hence $K(\alpha, \beta) \subseteq K(\theta)$. □

Note that the proof actually gives a construction for a primitive element: Take a linear combination $\theta = \alpha + c\beta$, then with a finite number of exceptions θ generates the same extension as α and β together.

4.2 Galois theory

In the construction of a splitting field L of a polynomial f over K we make several choices, since every root of an irreducible factor yields an isomorphic extension. Keeping track of these choices we actually construct various automorphisms of the splitting field. Galois theory deals with the connection between these automorphisms and the intermediate fields between K and L.

In the case of finite fields we can easily describe the situation: The field $\mathbb{F}_q$ with $q = p^n$ is the splitting field of $X^q - X$ over $\mathbb{F}_p$ and the intermediate fields are of the form $\mathbb{F}_{p^d}$ with $d|n$. Each automorphism of $\mathbb{F}_q$ is a power of the Frobenius-automorphism $\varphi : a \mapsto a^p$ and the intermediate fields can be described as $\mathbb{F}_{p^d} = \{a \in \mathbb{F}_q \mid \varphi^d(a) = a\}$. We therefore have a correspondence between subfields of $\mathbb{F}_q$ and subgroups of $Aut(\mathbb{F}_q)$, given by $\mathbb{F}_q \supseteq M \leftrightarrow U \leq Aut(\mathbb{F}_q)$ where $M = \{a \in \mathbb{F}_q \mid \sigma(a) = a \text{ for all } \sigma \in U\}$.

The goal is now to generalize this correspondence to arbitrary fields.

4.2.1 Definition Let L/K be a field extension and $f \in K[X]$.

(i) The group $Gal(L, K) := \{\sigma \in Aut(L) \mid \sigma_{|K} = id_K\}$ is called the *Galois group* of the field extension L/K.

(ii) If L is the splitting field of f over K, then the group $Gal(f,K) := Gal(L,K)$ is called the *Galois group of the equation* f.

Note that every $\sigma \in Aut(L)$ automatically fixes the prime field $P(L)$ elementwise, since $\sigma(1) = 1$, hence $Aut(L) = Gal(L, P(L))$.

The action of $\sigma \in Gal(f,K)$ is determined by the action on the roots of f. In particular this implies that for simple extensions we have

$$|Gal(K(\alpha),K)| \leq [K(\alpha) : K].$$

We will see that equality in this relation will yield the desired correspondence between subfields and subgroups of the automorphism group.

The two extreme cases of the above inequality are illustrated by the following two examples:

(1) Let α be the real root of $f := X^3 - 2$. Then $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 3$, but $Gal(\mathbb{Q}(\alpha), \mathbb{Q}) = \{1\}$, since the other roots of f are not contained in $\mathbb{Q}(\alpha)$. The splitting field $\mathbb{Q}(\alpha, \zeta_3)$ of f has degree 6 and the Galois group of f over $\mathbb{Q}(\zeta_3)$ is indeed a cyclic group of order 3.

(2) Let ζ_n be a primitive n-th root of unity, then $\mathbb{Q}(\zeta_n)$ is the splitting field of $f := X^n - 1$ and $[\mathbb{Q}(\zeta_n) : \mathbb{Q}] = \varphi(n)$ (where φ is the Euler-φ function). The other primitive n-th roots of unity are of the form ζ_n^i with $\gcd(i,n) = 1$ and mapping ζ_n to ζ_n^i for such an i defines an automorphism of $\mathbb{Q}(\zeta_n)$. Hence, $Gal(f, \mathbb{Q}) \cong \mathbb{Z}_n^*$ is of order $\varphi(n)$.

We will now show that the above inequality for simple extensions even holds for arbitrary extensions. The key for this (and several other results) is Dedekind's theorem:

4.2.2 Theorem *(Dedekind)*
Let $\sigma_1, \ldots, \sigma_n \in Aut(L)$ be distinct automorphisms. Then $\sigma_1, \ldots, \sigma_n$ are linearly independent, i.e. $a_1\sigma_1(x) + \ldots + a_n\sigma_n(x) = 0$ for all $x \in L$ implies $a_1 = \ldots = a_n = 0$.

Proof: Assume that $a_1\sigma_1(x) + \ldots + a_n\sigma_n(x) = 0$ for all $x \in L$. For $n = 1$ this implies $a_1 = 0$ and we are done. By induction we can assume that $n-1$ automorphisms are linearly independent, hence none of the coefficients a_i is 0, since otherwise the remaining $n-1$ automorphisms would be dependent. We therefore can divide by a_n and obtain

$$\frac{a_1}{a_n}\sigma_1(x) + \ldots + \sigma_n(x) = 0.$$

Furthermore, we have $\sigma_1(\alpha) \neq \sigma_n(\alpha)$ for some $\alpha \in L$, since the σ_i are distinct. In the above equation we now replace x by αx and divide by $\sigma_n(\alpha)$, this gives

$$\frac{a_1}{a_n}\frac{\sigma_1(\alpha)}{\sigma_n(\alpha)}\sigma_1(x) + \ldots + \sigma_n(x) = 0.$$

Subtracting the two equations yields a dependency between $n-1$ automorphisms, hence all coefficients have to be 0. In particular we have $\frac{a_1}{a_n} = \frac{a_1}{a_n}\frac{\sigma_1(\alpha)}{\sigma_n(\alpha)}$ and since $a_1 \neq 0$ this gives $\sigma_1(\alpha) = \sigma_n(\alpha)$, which is a contradiction. □

4.2.3 Corollary *Let L/K be a field extension, then $|Gal(L,K)| \leq [L:K]$.*

PROOF: Let $Gal(L,K) = \{\sigma_1, \ldots, \sigma_n\}$ and assume that $[L:K] = r < n$. Let $(v_1, \ldots, v_r)$ be a basis of L over K, then the equations $\sum_{i=1}^{n} \sigma_i(v_j)x_i = 0$ for $1 \leq j \leq r$ form a homogeneous linear system with r equations and $n > r$ indeterminates which therefore has a non-trivial solution $(c_1, \ldots, c_n)$.

For $\alpha = \sum_{j=1}^{r} a_j v_j$ we have $\sigma_i(\alpha) = \sum_{j=1}^{r} \sigma_i(a_j)\sigma_i(v_j) = \sum_{j=1}^{r} \sigma_j(a_j)\sigma_i(v_j)$ (note that $\sigma_i(a_j) = a_j$ for all i, j, since $a_j \in K$) and thus $\sum_{i=1}^{n} c_i\sigma_i(\alpha) = \sum_{i=1}^{n} c_i(\sum_{j=1}^{r} \sigma_j(a_j)\sigma_i(v_j)) = \sum_{j=1}^{r} \sigma_j(a_j)(\sum_{i=1}^{n} \sigma_i(v_j)c_i) = 0$. This shows that $\sum_{i=1}^{n} c_i\sigma_i(\alpha) = 0$ for all $\alpha \in L$, which is a contradiction to the linear independence of the σ_i. □

4.2.4 Theorem *Let $G = \{\sigma_1, \ldots, \sigma_n\} \leq Aut(L)$ and $K := Fix_L(G) := \{a \in L \mid \sigma(a) = a \text{ for all } \sigma \in G\}$. Then $[L:K] = n$.*

PROOF: By the above it is clear that $[L:K] \geq n$. Assume that $v_1, \ldots, v_{n+1}$ are linearly independent over K. The system of n homogeneous equations $\sum_{i=1}^{n+1} \sigma_j(v_i)x_i = 0$ $(1 \leq j \leq n)$ has a non-trivial solution $c_1, \ldots, c_{n+1}$. The c_i can not all lie in K, since otherwise the equation with $\sigma_j = id_L$ would be a dependency of the v_i. We choose the solution $c_1, \ldots, c_{n+1}$ such that the number r of $c_i \neq 0$ is minimal and we assume that $c_{r+1} = \ldots = c_{n+1} = 0$. Note that $r > 1$, since otherwise $c_1\sigma_1(v_1) = 0$ and thus $c_1 = 0$.

By dividing by c_r and rearranging if necessary we can assume that $c_r = 1$ and $c_1 \notin K$. therefore there exists σ_k with $\sigma_k(c_1) \neq c_1$. Applying σ_k to each of the equations $\sum_{i=1}^{r-1} \sigma_j(v_i)c_i + \sigma_j(v_r) = 0$ gives equations $\sum_{i=1}^{r-1} \sigma_j(v_i)\sigma_k(c_i) + \sigma_j(v_r) = 0$, since the products $\sigma_k\sigma_j$ are just a permutation of the σ_j. Subtracting the latter from the former gives $\sum_{i=1}^{r-1} \sigma_j(v_i)(c_i - \sigma_k(c_i)) = 0$, Since $c_1 \neq \sigma_k(c_1)$ this gives a non-trivial solution to the homogeneous system of equations, contradicting the minimality of r. □

4.2.5 Corollary

(i) Let $G \leq Aut(L)$, $K = Fix_L(G)$, $\sigma \in Aut(L)$ with $\sigma_{|K} = id_K$. Then $\sigma \in G$.

(ii) Let $G_1, G_2 \leq Aut(L)$ with $G_1 \neq G_2$. Then $Fix_L(G_1) \neq Fix_L(G_2)$.

We have thus seen that a subgroup of automorphisms characterizes a subfield and that two different subgroups correspond to different subfields. The special case of $G = Gal(L,K)$ gives rise to a crucial definition:

4.2.6 Definition A field extension L/K is called *normal* if $Fix_L(Gal(L,K)) = K$.

It is clear that the fixed field of $Gal(L,K)$ contains K. For a normal field extension we therefore have $|Gal(L,K)| = [L:K]$ by theorem 4.2.4. Note that sometimes this property is chosen as the definition of normal extensions. A third equivalent characterization is given by the following theorem:

4.2.7 Theorem *A field extension L/K is normal if and only if L is the splitting field of a separable polynomial over K.*

PROOF: $\Rightarrow$: We will first show something stronger, namely that any $\alpha \in L$ is a root of a separable polynomial over K which splits in L. Assume that $K = Fix_L(G)$ with $G = \{\sigma_1, \ldots, \sigma_n\}$ and let $\alpha \in L$. If we denote the distinct images $\sigma_j(\alpha)$ by $\alpha = \alpha_1, \ldots, \alpha_r$, then the α_i are permuted by the σ_j, hence the polynomial $f := \prod_{i=1}^{r}(X - \alpha_i)$ is invariant under each σ_j and thus $f \in K[X]$. On the other hand, if we have $g(\alpha) = 0$ for some $g \in K[X]$, then $g(\alpha_i) = g(\sigma_j(\alpha)) = \sigma_j(g(\alpha)) = 0$ for all $1 \leq i \leq r$, hence $\deg(g) \geq r$. This shows that f is the minimal polynomial of α over K and is thus irreducible. Furthermore, f has no double roots and splits in L.

The claim of the theorem now follows if we let $(v_1, \ldots, v_n)$ be a basis of L over K and let f_i be the minimal polynomial of v_i over K. Then L is the splitting field of the separable polynomial $f := \prod_{i=1}^{n} f_i$.
$\Leftarrow$: We proceed by induction over the number of roots of f lying outside K. If this number is 0 we are done, since $L = K$. Now let f_1 be an irreducible factor of f with $\deg(f_1) > 1$ and let α_1 be a root of f_1. From theorem 4.1.2 we know that there exist automorphisms $\sigma_i \in Gal(L, K)$ mapping α_1 to the distinct roots $\alpha_1, \ldots, \alpha_s$ of f_1. Now let θ be an element fixed under $Gal(L, K)$, then $\theta \in K(\alpha_1)$, since we know by induction that $L/K(\alpha_1)$ is normal and hence for any $\beta \in L \setminus K(\alpha_1)$ there exists an automorphism σ with $\sigma(\beta) \neq \beta$. Since $\deg(f_1) = s$ we can write θ as $\theta = c_0 + c_1\alpha_1 + \ldots + c_{s-1}\alpha_1^{s-1}$ with $c_i \in K$ and we have $\theta = \sigma_i(\theta) = c_0 + c_1\alpha_i + \ldots + c_{s-1}\alpha_i^{s-1}$. This shows that the polynomial $c_{s-1}X^{s-1} + \ldots + c_1X + (c_0 - \theta)$ has the s roots $\alpha_1, \ldots, \alpha_s$, hence it is 0 and we have $\theta = c_0 \in K$. □

We now return to examples of extensions to illustrate the results obtained.

(1) For $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$ we have seen that the Galois group consists only of the trivial automorphism. Thus, the fixed field is $\mathbb{Q}(\sqrt[3]{2})$, and hence the extension is not normal. On the other hand, the splitting field $\mathbb{Q}(\sqrt[3]{2}, \zeta_3)$ has degree 6 over $\mathbb{Q}$ and one checks that the three roots of $X^3 - 2$ and the two roots of $X^2 + X + 1$ can be mapped independently by automorphisms. The group of automorphisms is isomorphic to the symmetric group S_3 and the subgroup isomorphic to C_3 interchanges the three roots of $X^3 - 2$. The fixed field of this subgroup is $\mathbb{Q}(\zeta_3)$.

(2) For a primitive n-th root of unity the cyclotomic field $\mathbb{Q}(\zeta_n)$ is the splitting field of the (separable) polynomial $X^n - 1$, hence the extension is normal, and so the fixed field of all automorphisms $\zeta_n \mapsto \zeta_n^i$ with $\gcd(i, n) = 1$ is just $\mathbb{Q}$.

(3) For distinct squarefree $a, b \in \mathbb{Z}$ the biquadratic extension $\mathbb{Q}(\sqrt{a}, \sqrt{b})$ has degree 4 over $\mathbb{Q}$ and is generated by $\sqrt{a} + \sqrt{b}$ (as can be seen from the proof of the theorem on the primitive element). The automorphisms map $\sqrt{a} + \sqrt{b}$ to $\pm\sqrt{a} \pm \sqrt{b}$ and the Galois group is seen to be a Klein four group V_4. This shows that the extension is normal. The three subgroups of V_4 isomorphic to C_2 have fixed fields $\mathbb{Q}(\sqrt{a})$, $\mathbb{Q}(\sqrt{b})$ and $\mathbb{Q}(\sqrt{ab})$, respectively.

(4) The field $\mathbb{Q}(\sqrt[4]{2})$ contains only one further root of $X^4 - 2$, namely $-\sqrt[4]{2}$. The fixed field of the automorphism interchanging these two roots is $\mathbb{Q}(\sqrt{2})$, hence the extension is

not normal. The splitting field of X^4-2 is $\mathbb{Q}(\sqrt[4]{2}, i)$ (where i is a root of X^2+1) and the Galois group is isomorphic to the dihedral group D_8 (note that any automorphism maps $\sqrt[4]{2}$ to $i^k\sqrt[4]{2}$ with $0 \leq k \leq 3$ and i to $\pm i$).

We now come to the main theorem of Galois theory which establishes the correspondence between intermediate fields in a normal extension and subgroups of the Galois group.

4.2.8 Theorem *(Fundamental theorem of Galois theory)*
Let L/K be a normal field extension.

(i) There exists an inclusion-reversing bijective correspondence between intermediate fields $K \subseteq M \subseteq L$ and subgroups $U \leq Gal(L,K)$, given by $M \leftrightarrow Gal(L,M)$ and $U \leftrightarrow M = Fix_L(U)$.

(ii) For corresponding $K \subseteq M \subseteq L$ and $U \leq Gal(L,K)$ one has $[Gal(L,K) : U] = [M : K]$ and $|U| = [L : M]$.

(iii) The extension M/K is normal if and only if $Gal(L,M) \trianglelefteq Gal(L,K)$ and in this case one has $Gal(M,K) \cong Gal(L,K)/Gal(L,M)$.

Proof: (i)+(ii): If $K \subseteq M \subseteq L$, then L/M is normal, since L is also the splitting field over M. If we define $U := Gal(L,M)$ we have $U \leq G := Gal(L,K)$ and $|U| = [L : M]$, since L/M is normal. Conversely, for $U \leq G$ we know that $M := Fix_L(U)$ is uniquely determined by U. We have $[L : M] \cdot [M : K] = [L : K] = |G| = [G : U] \cdot |U|$ and hence $[L : K] = [G : U]$.
(iii): It is clear that any $\sigma \in G$ maps M to an isomorphic field and the cosets of G/U correspond to the different isomorphisms. Conversely, any isomorphism of M fixing K can be extended to an isomorphism of L and is thus the restriction of some $\sigma \in G$. Therefore, the number of isomorphisms of M fixing K equals $[G : U] = [M : K]$. Now, M/K is a normal extension if and only if the number of automorphisms of M fixing K equals $[M : K]$, thus, M/K is normal if and only if all $\sigma \in G$ map M onto itself. But this is equivalent with $Gal(L,M) \trianglelefteq G$, since $\sigma^{-1} \circ \tau \circ \sigma(a) = a$ for $\tau \in Gal(L,M)$ if and only if $\sigma(a) \in M$ for all $a \in M$ (otherwise, if $\sigma(a) \notin M$ there would be τ not fixing $\sigma(a)$). Finally, $Gal(M,K) \cong G/U$ is clear since the cosets represent the different isomorphisms of M which we have seen to be automorphisms in the case that $U \trianglelefteq G$. □

The correspondence of terminology for normal extensions and normal subgroups is no coincidence. Note that the fact that subgroups of index 2 are always normal corresponds to the fact that field extensions of degree 2 are always normal.

Having established the correspondence, an obvious question is how the intermediate fields are actually constructed, given a subgroup of the Galois group. This problem is solved by the *trace map*, which maps an element to the sum of its Galois conjugates. More precisely, if $\alpha \in L$ and $U \leq Gal(L,K)$, we define

$$\mathrm{tr}_U(\alpha) := \sum_{\sigma \in U} \sigma(\alpha).$$

It is clear that $\mathrm{tr}_U(\alpha)$ is fixed under all $\sigma \in U$. On the other hand, any element $\beta \in Fix_L(U)$ is obtained as a trace: Since according to Dedekind's theorem the $\sigma \in U$ are linearly independent, there exists $z \in L$ such that $z' := \sum_{\sigma \in U} \sigma(z) \neq 0$. Then $\mathrm{tr}_U(\frac{z}{z'}\beta) = \sum_{\sigma \in U} \frac{\sigma(z)}{\sigma z'}\sigma(\beta) = \frac{\sum_{\sigma \in U}\sigma(z)}{z'}\beta = \beta$.

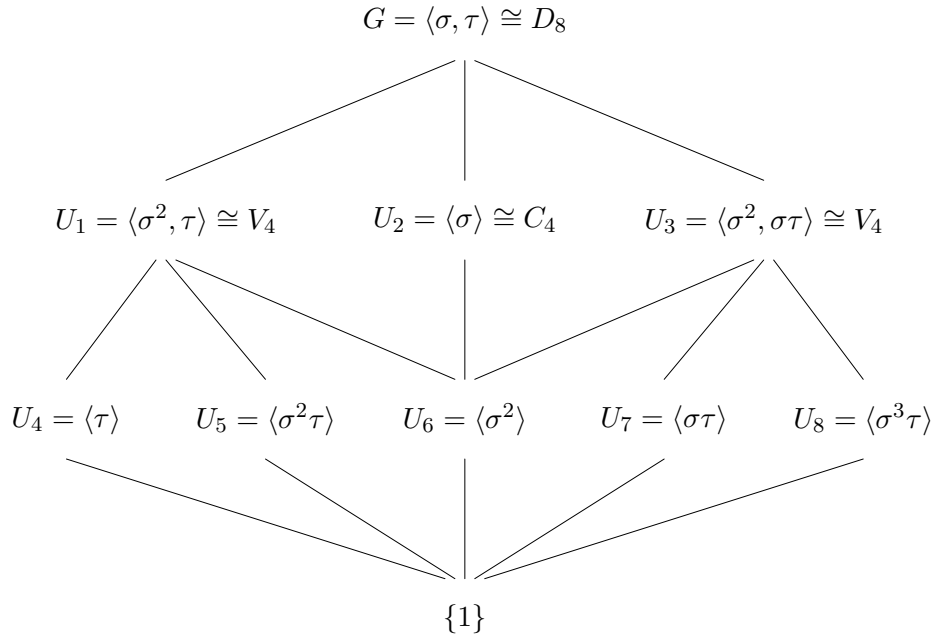
Running over a basis of L over K thus will produce generators for the fixed field of U, but usually one can stop early because the required degree is known in advance.

As an example illustrating the fundamental theorem we will analyze the splitting field $\mathbb{Q}(\sqrt[4]{2}, i)$ of $X^4 - 2$. Defining automorphisms σ, τ by

$$\sigma(\sqrt[4]{2}) := i\sqrt[4]{2},\ \sigma(i) = i,$$

$$\tau(\sqrt[4]{2}) := \sqrt[4]{2},\ \tau(i) = -i$$

we get the Galois group $G = \langle \sigma, \tau \rangle \cong D_8$. The subgroup diagram of G looks as follows:



The normal subgroups are U_1, U_2, U_3 and U_6, so these will correspond to normal extensions of $\mathbb{Q}$. Note that the intersection of two subfields corresponds to the group generated by the two subgroups and the intersections of two subgroups corresponds to the field generated by the two subfields. We easily identify $\mathbb{Q}(\sqrt[4]{2})$ as the fixed field of U_4 and $\mathbb{Q}(i)$ as the fixed field of U_2. Furthermore, $\mathbb{Q}(\sqrt{2})$ corresponds to a subgroup contained in U_4, thus to U_1. The third quadratic subfield $\mathbb{Q}(\sqrt{-2})$ therefore has to correspond to U_3. The three quadratic subfields generate the fixed field of U_6 and one sees that this is the $8th$ cyclotomic field $\mathbb{Q}(\zeta_8)$. We have not found the field $\mathbb{Q}(i\sqrt[4]{2})$ so far, and this is seen to be the fixed field of U_5. Finally, we can use the trace map to see that the fixed field of U_7 is $\mathbb{Q}((1+i)\sqrt[4]{2})$ and that U_8 corresponds to $\mathbb{Q}((1-i)\sqrt[4]{2})$.

4.3 Computing Galois groups

In principle we have seen how the Galois group of a polynomial f can be computed: One constructs the splitting field L of the f, computes a primitive element θ of L and its minimal polynomial f_θ and defines automorphisms mapping θ to the different roots of f_θ. The problem is, that except for very small (or fortunate) cases, the degree of L will be to large to be able to construct the extension effectively. It can be shown that in a certain sense polynomials of degree n have Galois group S_n with probability 1: if $P(N)$ denotes the probability that a polynomial with coefficients bounded by N has Galois group S_n, then $P(N)$ converges to 1 for $N \to \infty$. Hence in many cases the splitting field has degree $n!$ over $\mathbb{Q}$. Except for very small (or fortunate) cases it is therefore impractical to construct the splitting field L explicitly, since field extensions of degree much larger than 100 are not feasible.

One might think of approximating the roots of f numerically (either complex or p-adic) and define the automorphisms on the approximations. This, however, opens a new problem, since one root has to be expressed as a rational linear combination of the powers of another root and this leads back to algebraic reconstruction of elements (as was considered in the context of elliptic curves to find the minimal polynomial of $j(\tau)$) which is also unrealistic in high dimensions.

In practice, one uses a different approach to compute the Galois group of a polynomial f. Note that the splitting field L is obtained by adjoining the root of f. Since the Galois group of f permutes these roots, it is naturally represented as a permutation group on n points, namely on the roots of f. Moreover, every root can be mapped to any other root (by the theorem on extending automorphisms), hence we know that $Gal(f, K)$ is a transitive subgroup of S_n. The idea is now to find enough information about this transitive subgroup without actually constructing the splitting field L. The crucial observation is, that the reduction of f modulo p provides in many cases enough information, in particular in order to conclude that the Galois group is in fact the full symmetric group S_n.

The first question we want to answer is whether $G := Gal(f, \mathbb{Q})$ is a subgroup of the alternating group A_n. This can actually be read off the discriminant of f: Recall that if $\alpha_1, \ldots, \alpha_n$ are the roots of f, then the discriminant of f can be computed as $\operatorname{disc}(f) = \prod_{i<j}(\alpha_i - \alpha_j)^2$. If we define $\theta := \prod_{i<j}(\alpha_i - \alpha_j)$, we have $\sigma(\theta) = \operatorname{sgn}(\sigma)\theta$, since we can think of σ as a product of transpositions. Now we see that θ is fixed under all automorphisms if and only if $\theta \in \mathbb{Z}$ (note that θ is an algebraic integer) which proves:

4.3.1 Theorem *$Gal(f, \mathbb{Q}) \leq A_n$ if and only if $\operatorname{disc}(f)$ is a square in $\mathbb{Z}$.*

The crucial result which allows us to obtain information about $Gal(f, \mathbb{Q})$ by reductions of f modulo p is the following:

4.3.2 Theorem *If $p \nmid \operatorname{disc}(f)$, then each automorphism of the splitting field of f over $\mathbb{F}_p$ is induced by an automorphism of the splitting field of f over $\mathbb{Q}$. In particular, the Galois group of f over $\mathbb{F}_p$ embeds into the Galois group of f over $\mathbb{Q}$.*

Before we give two proofs of this theorem we show its consequence:

4.3.3 Theorem *If f is irreducible over $\mathbb{Q}$ and $f \equiv f_1 \dots f_s \bmod p$ for $p \nmid \operatorname{disc}(f)$, such that f_i is irreducible over $\mathbb{F}_p$ and $\deg(f_i) = n_i$, then $Gal(f, \mathbb{Q})$ contains an element with cycle structure $n_1, \dots, n_s$.*

PROOF: The splitting field of f over $\mathbb{F}_p$ is $\mathbb{F}_q = \mathbb{F}_{p^d}$ with $d = \operatorname{lcm}(n_1, \dots, n_s)$. The Galois group of $\mathbb{F}_q$ is generated by the Frobenius automorphism φ and if α is a root of f_i then the other roots of f_i are $\varphi^j(\alpha)$ with $0 \le j < n_i$. Hence, after suitable renumbering of the roots, φ acts as $(1, \dots, n_1)(n_1 + 1, \dots, n_1 + n_2) \dots (n - n_s + 1, \dots, n)$. □

PROOF: First proof of theorem 4.3.2:
Let $\alpha_1, \dots, \alpha_n$ be the roots of f in a splitting field L and let $x_1, \dots, x_n$ be indeterminates. For $\pi \in S_n$ define $u_\pi := \sum_{i=1}^n \alpha_{\pi(i)} x_i$ and let $F(X) := \prod_{\pi \in S_n} (X - u_\pi)$. Since $u_{\pi_1} \neq u_{\pi_2}$ for $\pi_1 \neq \pi_2$ the orbit of u_π under $G := Gal(f, \mathbb{Q}) \le S_n$ has length $|G|$. Now let $F(X) = F_1(X) \dots F_s(X)$ be the factorization of $F(X)$ into irreducible elements of $\mathbb{Q}[x_1, \dots, x_n][X]$. We claim that G is precisely the group of permutations leaving $F_j(X)$ invariant, i.e. that $F_j(X)$ is of the form

$$\prod_{\sigma \in G} (X - \sum_i \alpha_{\sigma\pi(i)} x_i).$$

To see this, let $\theta := \sum_i \alpha_i x_i$ and assume that the F_j are numbered such that $(X - \theta) \mid F_1(X)$. The elements of S_n fixing $F_1(X)$ are precisely those permutations mapping $X - \theta$ on some linear factor of $F_1(X)$ (note that $F(X)$ has no multiple factors). On the other hand, $F_1(X)$ is the minimal polynomial of θ, and thus also the minimal polynomial of the conjugates $\sigma(\theta)$ for $\sigma \in G$, hence each $X - \sigma(\theta)$ divides $F_1(X)$. The other $F_j(X)$ are obtained in the same manner by replacing θ by $\sum_i \alpha_{\pi(i)} x_i$ where π runs over a transversal of G in S_n. Thus, $F_1(X)$ is mapped to $F_1(X)$ under G and to the other $F_j(X)$ under the other cosets of S_n/G.

If we reduce the factorization $F(X) = F_1(X) \dots F_s(X)$ modulo p (and call the reductions again F_j), the $F_j(X)$ do not necessarily remain irreducible. Since we assume that $p \nmid \operatorname{disc}(f)$, also the reduction modulo p does not have multiple factors. Now let $\overline{G}$ be the Galois group of f over $\mathbb{F}_p$, then as before each irreducible factor of $F_1(X)$ is invariant under $\overline{G}$ and hence $F_1(X)$ is mapped to $F_1(X)$. Since the automorphisms leaving $F_1(X)$ invariant are precisely those of G, this shows that $\overline{G} \le G$. □

PROOF: Second proof of theorem 4.3.2:
We look at the Galois group of f over the p-adic completion $\mathbb{Q}_p$ of $\mathbb{Q}$. Let $L_p := \mathbb{Q}_p(\alpha_1, \dots, \alpha_n)$, where the α_i are the roots of f in the splitting field L over $\mathbb{Q}$. Note that every $\sigma \in Gal(L, \mathbb{Q})$ induces an automorphism $\sigma \in Aut(L_p)$, but since $\mathbb{Q}_p$ may contain some intermediate field of $L/\mathbb{Q}$, $\mathbb{Q}_p$ is not necessarily fixed pointwise by σ. In any case, we have $Gal(L_p, \mathbb{Q}_p) \le Gal(L, \mathbb{Q})$. The field L_p has a valuation ν which is an extension of the usual p-adic valuation on $\mathbb{Q}_p$ and the valuation ring $O(L_p)$ is given by $O(L_p) = \{a \in L_p \mid \nu(a) \ge 0\}$. The unique maximal ideal P of $O(L_p)$ is

$P = \{a \in L_p \mid \nu(a) > 0\}$ and $O(L_p)/P \cong \mathbb{F}_q$. In particular, $\mathbb{F}_q$ is the splitting field of the reduction of f modulo p.

Since P is the unique maximal ideal in L_p, it has to be fixed by any automorphism of L_p, hence reduction modulo P yields an automorphism of $\mathbb{F}_q = O(L_p)/P$. Furthermore, restricting the action of $\sigma \in Aut(L_p)$ to $\mathbb{Q}_p$, we reduce modulo $P \cap \mathbb{Q}_p = (p)$, hence reducing modulo P gives an automorphism of $\mathbb{F}_q$ fixing $\mathbb{F}_p = \mathbb{Z}_p/(p)$ pointwise. This shows that reduction modulo P gives a natural homomorphism $Gal(L_p, \mathbb{Q}_p) \to Gal(\mathbb{F}_q, \mathbb{F}_p)$ which is an epimorphism, since each automorphism modulo P can be lifted to an automorphism of L_p via Hensel lifting. The kernel of this homomorphism is called the ramification subgroup of $Gal(L_p, \mathbb{Q}_p)$ and it is trivial if $p \nmid \mathrm{disc}(f)$ (as we will see below). This shows that $Gal(L_p, \mathbb{Q}_p) \cong Gal(\mathbb{F}_q, \mathbb{F}_p)$ and in particular the Frobenius automorphism generating $Gal(\mathbb{F}_q, \mathbb{F}_p)$ is induced by an element of $Gal(L_p, \mathbb{Q}_p)$ which we have already seen to be a subgroup of $Gal(L, \mathbb{Q})$.

In order to show that the ramification subgroup is trivial in case $p \nmid \mathrm{disc}(f)$, we first assume that $P = p \cdot O(L_p)$, i.e. that the uniformizing element p of $\mathbb{Q}_p$ remains a uniformizing element in L_p. In this case the action of $\sigma \in Aut(L_p)$ is already determined on $O(L_p)/p \cdot O(L_p)$, since p remains fixed under every automorphism. The question is thus, under which condition $P = (\pi)$ with $\pi^e = p$ and $e > 1$. It can be seen that there is a unique maximal subfield I_p of L_p in which p remains the uniformizing element and that π is the root of an Eisenstein polynomial of degree e over I_p. From the formula for the discriminant one now concludes that this can only happen for $e > 1$ if $p \mid \mathrm{disc}(f)$. □

As a first application of theorem 4.3.2 we can now show how a polynomial with Galois group S_n for arbitrary n is constructed. For that we use the fact that a transitive subgroup of S_n containing an $(n-1)$-cycle and a transposition is S_n itself. Let f_1, f_2, f_3 be polynomials of degree n such that f_1 is irreducible modulo 2, f_2 splits modulo 3 into an irreducible factor of degree $n-1$ and a linear factor and f_3 splits modulo 5 into a quadratic factor and one or two irreducible factors of odd degree. Then $f := -15f_1 + 10f_2 + 6f_3$ is a monic polynomial of degree n and none of $2, 3, 5$ divides $\mathrm{disc}(f)$. Since the Galois group of f contains an $(n-1)$-cycle and (via an odd power of the permutation corresponding to the reduction modulo 5) a transposition, it is S_n.

We can actually exploit theorem 4.3.2 to show that 'almost all' polynomials of degree have Galois group S_n. We denote by

$$P(N) := \{f = X^n + a_1 X^{n-1} + \ldots + a_{n-1}X + a_0 \in \mathbb{Z}[X] \mid |a_i| \leq N\}$$

the set of monic integral polynomials with coefficients bounded by N. Furthermore we define $S(N) := \{f \in P(N) \mid Gal(f, \mathbb{Q}) = S_n\}$. Then we will show that

$$\frac{|S(N)|}{|P(N)|} \to 1 \text{ as } N \to \infty.$$

The idea is to count the polynomials f which have modulo three primes the following types of factorization:

I: f is irreducible modulo p,

II: f splits modulo p' into a linear factor and an irreducible factor of degree $n-1$,

III: f splits modulo p'' into an irreducible quadratic factor and one or two irreducible factors of odd degree.

Using the same argument as above we can conclude that the Galois group contains an $(n-1)$-cycle and a transposition and is thus S_n itself.

We proceed in three steps.

(1) Analysis modulo p:
To estimate the number of irreducible monic polynomials modulo p we count the elements in $\mathbb{F}_{p^n}$ not lying in any proper subfield. This number is at least

$$p^n - \sum_{d|n, d\neq n} p^d \geq p^n - \sum_{d=1}^{n-1} p^d > p^n - \frac{p^n}{p-1} \geq \frac{p^n}{2} \text{ for } p > 2.$$

Since every irreducible polynomial of degree n has n distinct roots, there are $\geq \frac{p^n}{2n}$ irreducible monic polynomials of degree n.

For factorization of type II we get p linear and $\geq \frac{p^{n-1}}{2(n-1)}$ irreducible polynomials of degree $n-1$ and therefore have $\geq \frac{p^n}{2(n-1)}$ distinct monic polynomials with type II factorization.

For factorization of type III we have $\geq \frac{p^2}{4}$ irreducible quadratic polynomials and for n odd $\geq \frac{p^{n-2}}{2(n-2)}$ irreducible polynomials of degree $n-2$ or for n even $\geq p\frac{p^{n-3}}{2(n-3)}$ products of a linear and an irreducible polynomial of degree $n-3$. In any case we have $\geq \frac{p^n}{8(n-2)}$ distinct monic polynomials with type III factorization.

We conclude that for $k = 8(n-2)$ for any chosen type of factorization the proportion of polynomials modulo p with this type is $\geq \frac{1}{k}$ (the case $n = 2$ of course has to be handled slightly differently, but we can choose $k = 4$ here).

(2) Combining primes:
By the Chinese remainder theorem we know that for different primes $p_1, p_2, \ldots, p_m$ the reductions modulo p_i determine a unique polynomial modulo $P := \prod_{i=1}^m p_i$. For a chosen type X we know from step (1) that there are at most $\frac{k-1}{k} p_i^n$ polynomials modulo p_i with factorization not of type X. Combining these for the different primes gives at most $(\frac{k-1}{k})^m P^n$ polynomials which do not have a factorization of type X modulo any of the primes p_i.

For $\varepsilon > 0$ we now choose m such that $(\frac{k-1}{k})^m < \varepsilon$ and let $P := p_1 \ldots p_m$ be the product of the first m odd primes. We then know that of the P^n monic polynomials modulo P at most εP^n do not have a factorization of type I (II, III) for any of the p_i, hence there are at most $3\varepsilon P^n$ polynomials with no such factorization modulo any of the p_i and we call these polynomials with bad reduction. The remaining $\geq (1-3\varepsilon)P^n$ polynomials have type I, II and III factorizations modulo three suitably chosen primes.

(3) Bounding the coefficients:
Let N be the bound on the coefficients a_i then of the possible values of a_i there are at most $\lfloor \frac{2N}{P} \rfloor + 1$ congruent modulo P. Therefore, at most $3\varepsilon P^n(\frac{2N}{P}+1)^n = 3\varepsilon(2N+P)^n$ polynomials with coefficients bounded by N have bad reduction. If we restrict ourselves to N with $2N+2 \geq P$, these are $\leq 3 \cdot 2^n \varepsilon (2N+1)^n$ polynomials and hence the proportion of polynomials with coefficients bounded by N and Galois group a proper subgroup of S_n is at most $3 \cdot 2^n \varepsilon$.

We now see that in step (2) we should rather have chosen m such that $(\frac{k-1}{k})^m < 3 \cdot 2^n \varepsilon$ in order to obtain a proportion of $\geq 1 - \varepsilon$ polynomials with Galois group S_n.

4.4 Density theorems

Literature:

F.G. Frobenius: *Über Beziehungen zwischen den Primidealen eines algebraischen Körpers und den Substitutionen einer Gruppe.* Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin, 1896. (Gesammelte Abhandlungen II, 719-733, Springer, 1968.)

Helmut Hasse: *Bericht über neuere Untersuchungen und Probleme aus der Theorie der algebraischen Zahlkörper. Teil II: Reziprozitätsgesetz.* Physica-Verlag, 1965.

S. Lang: *Algebraic Number Theory.* Addison-Wesley, 1970.

P. Stevenhagen, H.W. Lenstra, Jr: *Chebotarëv and his Density Theorem.* Mathematical Intelligencer, vol. 18, no. 2, 26-37, 1996.

When we talk about relative frequencies of primes with certain properties we have to define what we call the *density* of a set of primes. A natural (or naive) definition is the following:

If $V \subseteq \mathbb{P}$ is a set of primes we say that V has *natural density* δ if the limit

$$\lim_{N \to \infty} \frac{|p \in V \mid p \leq N|}{|p \in \mathbb{P} \mid p \leq N|}$$

exists and has value δ.

From the point of view of analytic number theory, a different definition is often more suitable, namely the Dirichlet or analytic density:

We say that V has *analytic density* δ if the limit

$$\lim_{s \to 1_+} \frac{\sum_{p \in V} p^{-s}}{\sum_{p \in \mathbb{P}} p^{-s}}$$

exists and has value δ.

It can be shown that the existence of the natural density implies the existence of the analytic density and in this case they are equal. Unfortunately, the opposite implication is not true.

We will exclusively deal with the analytic density and from now on omit the attribute 'analytic'.

A beautiful (and important) result about densities of primes is Dirichlet's theorem about *primes in arithmetic progressions*:

4.4.1 Theorem *(Dirichlet, 1837)*
Let $m \in \mathbb{N}$ and $a \in \mathbb{Z}$ such that $\gcd(a, m) = 1$. Then the set $\{p \in \mathbb{P} \mid p \equiv a \bmod m\}$ of primes in the arithmetic progression $a + m\mathbb{Z}$ has density $1/\varphi(m)$.

This theorem not only says that every arithmetic progression contains infinitely many primes, but also that each of them contains the same amount.

The density theorems we are going to discuss rely on the following result of Kronecker, describing the density of primes over which a polynomial has a root:

4.4.2 Theorem *(Kronecker, 1880)*
Let $f \in \mathbb{Z}[X]$ be a polynomial with m irreducible factors of degree ≥ 1. If we denote the number of roots of the reduction $f \bmod p$ by a_p, then

$$\lim_{s \to 1_+} \frac{\sum_{p \in \mathbb{P}} a_p p^{-s}}{\log(\frac{1}{s-1})} = m.$$

What we have seen in the last section is that the Galois group of the reduction of a polynomial $f \bmod p$ embeds into the Galois group over $\mathbb{Q}$. This means that from a factorization modulo p we can conclude that an element with a certain cycle structure exists in the Galois group. However, in order to identify Galois groups from this information we need a kind of converse of this property, namely that we actually get *every* cycle type from the factorization modulo a suitable prime.

We will see that the situation is as good as we could possibly hope: Frobenius' density theorem states that every cycle structure which exists in the Galois group does occur for some prime p and we even know how often we find it on average, namely as often as we find elements with this cycle structure in the group. For example, the density of primes over which f splits completely is the reciprocal of the order of the Galois group.

4.4.3 Theorem *(Frobenius, 1880/1896)*
Let $f \in \mathbb{Z}[X]$ be a monic polynomial of degree n. Then the density of primes $p \in \mathbb{P}$ such that $f \bmod p$ has decomposition type $n_1, \ldots, n_s$ equals the relative frequency of $\sigma \in Gal(f, \mathbb{Q})$ with cycle structure $n_1, \ldots, n_s$.

If we do not know the Galois group yet, Frobenius' theorem is the best we can expect, since we can only identify elements up to conjugacy in the full symmetric group S_n, i.e. up to cycle structure. However, if we know the Galois group, we could look at a finer

classification of the group elements, namely the conjugacy classes in the Galois group. For this situation, Chebotarëv's generalization of Frobenius' theorem states that also for the conjugacy classes the density of the primes for which their Frobenius automorphism embeds into a fixed conjugacy class is the size of this class divided by the group order.

4.4.4 Theorem *(Chebotarëv, 1922)*
Let $f \in \mathbb{Z}[X]$ be a monic polynomial of degree n. For a prime p let $\mathbb{F}_q$ be the splitting field of $f \bmod p$ over $\mathbb{F}_p$ and denote the embedding of the Frobenius automorphism of $\mathbb{F}_q$ into $Gal(f,\mathbb{Q})$ by σ_p. Then the density of primes $p \in \mathbb{P}$ such that σ_p lies in a conjugacy class C of $Gal(f,\mathbb{Q})$ equals $|C|/|Gal(f,\mathbb{Q})|$.

We will only prove Frobenius' theorem here, following the exposition of Hasse. A nice sketch of a proof of Chebotarëv's theorem can be found in the article of Stevenhagen/Lenstra, a full proof (using class field theory) in Lang's book.

We start by fixing some notation.

Let $f \in \mathbb{Z}[X]$ be an irreducible monic polynomial, let L be the splitting field of f over $\mathbb{Q}$ and let $G := Gal(f,\mathbb{Q}) = Gal(L,\mathbb{Q})$ be the Galois group of f. Assume that $f \equiv f_1 \cdot \ldots \cdot f_s \bmod p$, where f_i is irreducible over $\mathbb{F}_p$ and $\deg(f_i) = n_i$. Then the splitting field of f over $\mathbb{F}_p$ is $\mathbb{F}_q$ with $q = p^{\mathrm{lcm}(n_1,\ldots,n_s)}$. If we denote the ring of integers of L by Λ and let P be a prime ideal of Λ containing (p), then $\Lambda/P \cong \mathbb{F}_q$. As usual, we define the degree of P as the degree of $\mathbb{F}_q$ over $\mathbb{F}_p$ (which equals $\mathrm{lcm}(n_1,\ldots,n_s)$).

By 4.3.2 we know that the Galois group of Λ/P embeds into G. This means that we can associate the Frobenius automorphism of Λ/P with an element σ_P of G such that

$$a^p \equiv \sigma_P(a) \bmod P \text{ for all } a \in L.$$

By definition, the order of σ_P equals the degree of P.

Now let τ be an arbitrary element of G, then $\tau(a^p) = \tau\sigma_P(a) \bmod \tau(P)$, hence $\tau(a)^p = \tau\sigma_P\tau^{-1}(\tau(a)) \bmod \tau(P)$. This shows that the prime ideal $\tau(P)$ corresponds to the conjugate element $\tau\sigma_P\tau^{-1}$. Since the Galois group G is transitive on the prime ideals lying over (p), we can associate the conjugacy class $(\sigma_P)^G$ of σ_P with the prime p.

These correspondences are often denoted by the *Frobenius symbol* and the *Artin symbol*: $\sigma_P \leftrightarrow [\frac{L}{P}]$ and $(\sigma_P)^G \leftrightarrow (\frac{L}{p})$.

As an auxiliary result we need an analysis of how prime ideals in intermediate fields correspond with the prime ideals in L over p. We keep the above notation.

4.4.5 Lemma *Let $L \supseteq M \supseteq \mathbb{Q}$, let $G := Gal(L,\mathbb{Q})$, $U := Gal(L,M) \leq G$ and define $S := \langle\sigma_P\rangle$. Denote by $\tau_1,\ldots,\tau_r$ a transversal for the double cosets of G by U and S, i.e. $G = \dot\cup_{i=1}^r U\tau_i S$.*

(i) The decomposition of (p) into prime ideals Q_i of M is $(p) = \prod_{i=1}^r Q_i$ where $Q_i = \prod_{\tau \in U\tau_i S} \tau(P)$.

(ii) The degree f_i of Q_i equals the order of $\tau_i\sigma_P\tau_i^{-1}$ relative to U, i.e. f_i is the smallest positive integer such that $(\tau_i\sigma_P\tau_i^{-1})^{f_i} \in U$. This means that $U\tau_i S = \dot\cup_{j=0}^{f_i-1} U\tau_i\sigma_P^j$.

PROOF: (i): Let Q_i be a prime ideal in M containing (p), then we can assume that the τ_i are numbered such that $Q_i \subseteq \tau_i(P)$. Now let $h \in U$, then $Q_i \subseteq h\tau_i\sigma_P(P)$, since $\sigma_P(P) = P$ and $h(a) = a$ for all $a \in M$. This shows that $Q_i \subseteq \tau(P)$ for all $\tau \in U\tau_i S$. On the other hand, all conjugate ideals of P lying over Q_i lie in one orbit under $U = Gal(L, M)$, hence every $\tau(P)$ with $Q_i \subseteq \tau(P)$ is of the form $\tau(P) = h\tau_i(P)$ for some $h \in U$. Since $\tau^{-1}h\tau_i(P) = P$ we have $\tau^{-1}h\tau_i(P) \in S$, hence $\tau \in U\tau_i S$.
(ii): Denote the degree of $\tau_i(P)$ by f, the degree of Q_i (in M) by f_i and the relative degree of $\tau_i(P)$ over M by $\overline{f_i}$, then clearly $f = f_i \cdot \overline{f_i}$. We have seen that the Galois group of $\Lambda/\tau_i(P)$ corresponds to $\tau_i S\tau_i^{-1}$ and by the Fundamental theorem of Galois theory we conclude that over M the Galois group of $\Lambda/\tau_i(P)$ corresponds to $\tau_i S\tau_i^{-1} \cap U$. In particular we have $f = |S| = |\tau_i S\tau_i^{-1}|$ and $\overline{f_i} = |\tau_i S\tau_i^{-1} \cap U|$ and therefore $f_i = \frac{f}{\overline{f_i}} = \min(j \geq 1 \mid \tau_i\sigma_P^j\tau_i^{-1} \in U)$. □

As a consequence of this lemma we can express the number of prime ideals of degree 1 in L in purely group theoretic terms.

4.4.6 Theorem *Let $L \supseteq M \supseteq \mathbb{Q}$, $G := Gal(L, \mathbb{Q})$, $U := Gal(L, M)$ and $\sigma \in (\sigma_P)^G$. Then the number a_p of prime ideals $Q \supseteq (p)$ of degree 1 in M is*

$$a_p = \frac{1}{|U|}|\{\tau \in G \mid \tau\sigma\tau^{-1} \in U\}|.$$

In particular, a_p is independent of the choice of σ.

PROOF: Keeping the notation of the above lemma we know that a prime ideal $Q_i \supseteq (p)$ has degree 1 if and only if $f_i = 1$, i.e. if and only if $\tau_i\sigma_P\tau_i^{-1} \in U$. Therefore, the number of Q_i of degree 1 equals the number of indices i such that $\tau_i\sigma_P\tau_i^{-1} \in U$ or, equivalently, such that $\tau_i S\tau_i^{-1} \subseteq U$.

Next we see that τ_i satisfies $\tau_i S\tau_i^{-1} \subseteq U$ if and only if any element τ in the double coset $U\tau_i S$ satisfies $\tau S\tau^{-1} \subseteq U$. Hence, for these i every such double coset consists of elements satisfying $\tau S\tau^{-1} \subseteq U$ and since $U\tau_i S = U\tau_i$ every such double coset gives $|U|$ elements τ with this property.

Finally we remark that the cardinality of the set $\{\tau \in G \mid \tau\sigma\tau^{-1} \in U\}$ only depends on the conjugacy class of σ, since for a conjugate element $\varphi\sigma\varphi^{-1}$ we can let $\tau\varphi^{-1}$ instead of τ run over G. □

PROOF: of Theorem 4.4.3 (Frobenius)
Let $\sigma \in G$, let $L \subseteq M \subseteq \mathbb{Q}$ be an intermediate field and let $U := Gal(L, M) \leq G$ be the Galois group of L over M. We denote by $a_p(M)$ the number of prime ideals $Q \supseteq (p)$ of degree 1 in M and define $b_p := |\{\tau \in G \mid \tau\sigma_P\tau^{-1} \in U\}|$. By Theorem 4.4.6 we have $a_p(M) = \frac{b_p}{|U|}$ and thus $\frac{a_p(M)}{[G:U]} = \frac{b_p}{|G|}$.

For a subset $V \subseteq G$ we denote by $f_G(V)$ the relative frequency $f_G(V) := \frac{|V|}{|G|}$ of V in G. Then we have $f_G((\sigma_P)^G \cap U) = \frac{b_p}{|G|} = \frac{a_p}{[G:U]}$ and $f_G(U) = \frac{1}{[G:U]}$. By dividing the equation in

Theorem 4.4.2 by $[G:U]$ we get

$$\lim_{s\to 1_+} \frac{\sum_{p\in\mathbb{P}} \frac{a_p}{[G:U]} p^{-s}}{\log(\frac{1}{s-1})} = \lim_{s\to 1_+} \frac{\sum_{p\in\mathbb{P}} f_G((\sigma_P)^G \cap U) p^{-s}}{\log(\frac{1}{s-1})} = f_G(U). \qquad (*)$$

This equation holds in particular for the cyclic group $U = \langle\sigma\rangle$, but also for its subgroups $U_d = \langle\sigma^d\rangle$ where $d \mid |U|$. Since the statement relates purely to the relative frequencies of subsets of G and does not use the group structure of U, we can conclude that it also holds for U replaced by $U_0 := U \setminus \bigcup_{d||U|} U_d$, i.e. for the set of generators of U.

Since the set U_0 of generators of $U = \langle\sigma\rangle$ is precisely the set of elements of the same order as σ, we have $\tau\sigma\tau^{-1} \in U_0 \Leftrightarrow \tau U \tau^{-1} = U$. This shows that the set $\{\tau \in G \mid \tau\sigma\tau^{-1} \in U_0\}$ is precisely the normalizer $N_G(U)$ of U in G, thus we have $f_G(\sigma^G \cap U_0) = \frac{N_G(U)}{|G|}$.

We call the union of the conjugacy classes of σ^k for $\gcd(k, |U|) = 1$ the *division* of σ and denote this by $\mathcal{A}(\sigma)$ (from the German 'Abteilung'). The division consists thus of the conjugacy classes of the powers of σ having the same cycle structure as σ. It is clear that all the σ^k contained in $\mathcal{A}(\sigma)$ generate U and since conjugate elements have conjugate normalizers all elements in the division of σ have normalizers of the same order.

We now can restrict the sum over all primes in $(*)$ to those primes for which σ_P belongs to the division $\mathcal{A}(\sigma)$ of σ, since we have $f_G((\sigma_P)^G \cap U_0) > 0$ only for these primes. Using the Artin symbol, we thus only have to sum over the primes for which $(\frac{\mathbb{Q}}{p}) \subseteq \mathcal{A}(\sigma)$. If we divide equation $(*)$ (for U_0 instead of U) by $f_G(\sigma^G \cap U_0) = \frac{N_G(U)}{|G|}$ we obtain:

$$\lim_{s\to 1_+} \frac{\sum_{(\frac{\mathbb{Q}}{p})\subseteq\mathcal{A}(\sigma)} p^{-s}}{\log(\frac{1}{s-1})} = f_G(U_0) \cdot \frac{|G|}{|N_G(U)|} = f_G(\mathcal{A}(\sigma)),$$

since $[G : N_G(U)]$ is the length of the orbit of U_0 under conjugation by G.

We have actually proved something slightly stronger than what was stated, namely that the density of primes such that the automorphism σ_P over p belongs to the division of σ equals the relative frequency of this division, which was also already proved by Frobenius. The claim as given in the theorem follows immediately, since the set of elements of a given cycle structure is a union of divisions. □

Note that in the proof we made heavy use of the subgroup $U = \langle\sigma\rangle \leq G$ and its corresponding subfield M of L. The splitting field L is thus viewed as a cyclic extension of the intermediate field M. This idea of a cyclic relative extension is also a crucial idea in the proof of Chebotarëv's theorem, both in the original proof by Chebotarev (using Galois theory) and also in the shorter proof given by Deuring in 1935 (using class field theory).

4.5 Recognizing S_n and A_n

We have seen that 'most' polynomials have Galois group S_n and that we can easily determine whether a Galois group is contained in A_n. It is therefore desirable to have a fast method to recognize that a given polynomial does in fact have Galois group S_n or A_n.

The key to such a method is that S_n is n-fold transitive and A_n is sharply $(n-2)$-fold transitive and that there are no other 'highly' transitive groups. To be precise: Using the classification of finite simple groups one can show that there are no 6-fold transitive groups which are not symmetric or alternating, that the Mathieu groups M_{12} and M_{24} are the only other sharply 5-fold transitive groups and that the Mathieu groups M_{11} and M_{23} are the only other sharply 4-fold transitive groups.

The standard argument now uses the following results:

4.5.1 Lemma *Let G be a transitive permutation group on n points. If G contains a p-cycle for a prime p with $p > \frac{n}{2}$, then G is primitive.*

PROOF: If G is imprimitive, it has a block system of k blocks of size m with $2 \leq k, m \leq \frac{n}{2}$. But then G would be a subgroup of $S_m \wr S_k$ which has order $(m!)^k \cdot k!$ and p would not divide the order of G, which is a contradiction. □

The following theorems are cited without proof and can e.g. be found in H. Wielandt: *Finite permutation groups*, Academic Press, 1964.

4.5.2 Theorem *(Marggraf, 1892)*
Let G be a primitive permutation group on n points. If G contains an m-cycle with $1 < m < n$, then G is $(n-m+1)$-fold transitive.

4.5.3 Theorem *(Jordan, 1873)*
Let G be a primitive permutation group on n points and let p be a prime such that $n = p+k$ with $k \geq 3$. Then G is either S_n or A_n.

To apply these theorems one first has to find a p-cycle for a prime $p > \frac{n}{2}$ which proves the primitivity of G. If $p < n-2$, one applies Jordan's theorem and is done. Otherwise, one has to find some m-cycle for $m < n-2$ not necessarily prime. The exceptional cases of the Mathieu groups are easily excluded.

We now estimate the efficiency of this method.

4.5.4 Proposition *Let $n \in \mathbb{N}$ be the degree of S_n and A_n.*

(i) Let $\frac{n}{2} < m \leq n$. Then the proportion of elements of S_n containing a cycle of length m is $\frac{1}{m}$. The proportion of these elements in A_n is $\frac{1}{m}$ if $m \leq n-2$ and it is 0 or $\frac{2}{m}$ if $m \in \{n-1, n\}$.

(ii) The proportion of elements in S_n or A_n containing a p-cycle for a prime p with $\frac{n}{2} < p < n-2$ is asymptotically $\frac{\log(2)}{\log(n)}$.

PROOF: (i): The number of elements of S_n containing an m-cycle is $\binom{n}{m}(m-1)!(n-m)!$, since we have $\binom{n}{m}$ choices for the m points in the m-cycle, $(m-1)!$ distinct m-cycles on a chosen set of m points and $(n-m)!$ permutations on the complement of these points.

Since $m > \frac{n}{2}$, there is at most one m-cycle in an element, hence we do not count elements twice, and hence the proportion of elements is obtained by dividing by $n!$, which gives $\frac{1}{m}$.

For A_n and $m \leq n-2$, we have to supplement an m-cycle by an even element of S_{n-m} if m is odd and by an uneven element if m is even. In both cases we get $(n-m)!/2$ permutations on the complement of the m points of the m-cycle. Dividing by $n!/2$ gives again the proportion $\frac{1}{m}$.

Finally, if n is even, there are no n-cycles in A_n and for $m = n-1$ there are $n(n-2)!$ m-cycles, giving a proportion of $\frac{2}{m}$. If n is odd, there are no $(n-1)$-cycles, but $(n-1)!$ n-cycles, hence the proportion is again $\frac{2}{m}$.
(ii): Since no element can have cycles of lengths p and q for distinct primes $> \frac{n}{2}$, the proportion of elements with a suitable p-cycle is

$$S(n) := \sum_{\substack{\frac{n}{2}<p<n-2 \\ p \text{ prime}}} \frac{1}{p}.$$

By the prime number density theorem, the sum over $\frac{1}{p}$ up to a bound N is approximately $\log(\log(N)) = \int_e^N \frac{1}{x\log(x)}\,dx$, hence the sum can be estimated as

$$\log(\log(n-2)) - \log(\log(\frac{n}{2})) = \log(\frac{\log(n-2)}{\log(\frac{n}{2})} \approx \log(1 + \frac{\log(2)}{\log(n)}) \approx \frac{\log(2)}{\log(n)}.$$

□

4.5.5 Example We take $f = X^{100} + X^2 + X + 1$, trusting or checking that it is irreducible over $\mathbb{Q}$ and try to prove that it has Galois group S_{100}. We have $\sum_{p=53}^{97} \frac{1}{p} \approx 0.14$, whereas $\frac{\log(2)}{\log(100)} \approx 0.15$. We therefore expect one in every 7 primes to give a factorization with a suitable p-cycle. One checks that none of the primes below 100 divides the discriminant of f and we get the following cycle structures:
$p = 2 : 1, 4, 9, 40, 46$, $p = 3 : 8, 14, 78$, $p = 5 : 3, 7, 12, 16, 62$, $p = 7 : 3, 22, 75$,
$p = 11 : 1, 1, 4, 26, 68$, $p = 13 : 1, 6, 9, 25, 26, 33$, $p = 17 : 1, 3, 9, 10, 77$,
$p = 19 : 2, 98$, $p = 23 : 3, 5, 6, 86$, $p = 29 : 1, 3, 4, 6, 86$, $\mathbf{p = 31} : 3, 3, 12, 29, \mathbf{53}$,
$p = 37 : 3, 13, 84$, $\mathbf{p = 41} : \mathbf{100}$, $p = 43 : 1, 1, 5, 10, 38, 45$, $p = 47 : 3, 45, 52$,
$p = 53 : 1, 1, 6, 9, 11, 13, 19, 40$, $p = 59 : 1, 2, 2, 4, 36, 55$, $p = 61 : 24, 27, 49$,
$p = 67 : 4, 8, 88$, $p = 71 : 1, 3, 3, 6, 7, 12, 29, 39$, $p = 73 : 1, 2, 2, 3, 11, 13, 68$,
$\mathbf{p = 79} : 1, 32, \mathbf{67}$, $p = 83 : 1, 1, 2, 6, 17, 34, 39$, $\mathbf{p = 89} : 4, 43, \mathbf{53}$,
$p = 97 : 1, 1, 5, 12, 16, 30, 35$.

Thus, the prime $p = 41$ shows that f is in fact irreducible and the three (out of 25) primes $p = 31$, $p = 79$ and $p = 89$ show that the Galois group is indeed S_{100}.

Chapter 5

Permutation groups

Literature:

A. Seress: *Permutation Group Algorithms.* Cambridge University Press, 2003.

G. Butler: *Fundamental Algorithms for Permutation Groups.* Springer, 1991.

Since groups naturally occur as operations acting on some set or structure, they are usually represented in one of the following forms:

- as permutation groups
- as matrix groups (over some ring or field)
- as finitely presented groups (with generators and relations).

In this chapter we will focus on groups given by a set of generating permutations. We will describe the standard way to deal with these groups algorithmically and we will see how random methods speed up the standard techniques drastically.

One way of proving a probabilistically obtained result correct forms a nice connection to the third way of representing groups, hence we will also touch the subject of finitely presented groups.

Finally, we will see how an elementary result from linear algebra (or probability theory) leads to a very simple method to generate random group elements from a set of generators.

5.1 Stabilizer chains

In this chapter we will exclusively deal with permutation groups of finite sets, thus with finite permutation groups.

If Ω is a finite set, we denote the group of all permutations of Ω by S_Ω. Identifying Ω with the set $\{1, \ldots, n\}$ we can regard every subgroup of S_Ω as a subgroup $G \leq S_n$.

5.1.1 Definition Let Ω be a finite set and let $G \leq S_\Omega$:

(i) For $x \in \Omega$ the set $G_x := \{g \in G \mid xg = x\}$ is called the *stabilizer* of x in G. It is clear that G_x is a subgroup of G.

(ii) For $x_1, \ldots, x_r \in \Omega$ we define $G_{x_1,\ldots,x_r} := G_{x_1} \cap \ldots \cap G_{x_r}$ to be the *pointwise stabilizer* of $x_1, \ldots, x_r$.

(iii) A sequence $B = (b_1, \ldots, b_m)$ is called a *base* for G if $G_{b_1,\ldots,b_m} = \{1\}$. An element $g \in G$ is uniquely determined by its images on a base.

(iv) For $x \in \Omega$, the set $\{xg \mid g \in G\} \subseteq \Omega$ is called the *orbit* of x in G and is denoted by x^G.

For a base $B = (b_1, \ldots, b_m)$ we define subgroups $G^{(i)}$ by $G^{(i)} := G_{b_1,\ldots,b_i}$. Then the groups $G^{(i)}$ form a chain

$$G = G^{(0)} \geq G^{(1)} \geq \ldots \geq G^{(m)} = \{1\}$$

of subgroups which is called a *stabilizer chain*.

A base B is called *non-redundant* if $G^{(i-1)} \neq G^{(i)}$ for all $1 \leq i \leq m$. Redundancy of a basis is easily determined, since since $G^{(i)} = G^{(i-1)} \cap G_{b_i}$ and therefore $G^{(i-1)} = G^{(i)} \Leftrightarrow G^{(i-1)} \leq G_{b_i}$.

Note that a base is a *sequence* of points, thus the order of the points is important. The following almost trivial example shows that groups may have bases of different lengths and that rearranging the points may result in redundant or non-redundant bases.

The group $G = \langle (1,2,3,4)(5,6) \rangle$ is isomorphic to a cyclic group of order 4. Clearly, (1) is a base, but also $(5,1)$ is a non-redundant base, since $G_5 = \langle (1,3)(2,4) \rangle$ and $G \gneqq G_5 \gneqq \{1\}$. On the other hand, $(1,5)$ is obviously a redundant base.

A redundant base usually does not cause (non-trivial) problems, but to avoid pathologies we will always assume that bases are non-redundant.

Having obtained a stabilizer chain, a lot of properties of a group can be determined inductively by walking down (or up) the stabilizer chain.

For example, the order of the group is obtained as follows:

5.1.2 Proposition *Let $B = (b_1, \ldots, b_m)$ be a base for G and let $G \geq G^{(1)} \geq \ldots \geq G^{(m)}$ be the corresponding stabilizer chain.*

Then $|G| = \prod_{i=1}^{m} [G^{(i-1)} : G^{(i)}] = \prod_{i=1}^{m} |b_i^{G^{(i-1)}}|$.

Proof: By Lagrange's theorem we know that the index of the stabilizer of a point equals the length of the orbit of that point, hence we have $[G^{(i-1)} : G^{(i)}] = |b_i^{G^{(i-1)}}|$. Since $|G| = [G^{(0)} : G^{(1)}] \cdot |G^{(1)}| = \ldots = [G^{(0)} : G^{(1)}] \cdot \ldots \cdot [G^{(m-1)} : G^{(m)}]$ the claim follows. □

The first computational task is to compute the orbit of a point under a group given by generators. This is done by simply applying the generators to the points found so far until no new points are found. By keeping track of the first element by which a new point is reached, we obtain at the same time and at almost no extra cost a transversal for the

stabilizer in the group. This is based on the fact that $g' \in G_x g \Leftrightarrow xg' = xg$ which shows that $y \in x^G \leftrightarrow G_x g, xg = y$ is a bijective correspondence between the points in the orbit and the cosets by the stabilizer.

5.1.3 Orbit Algorithm The cost of computing the orbit x^G is proportional to $|x^G| \cdot |S|$, where S is the set of generators for G. The following pseudocode describes an algorithm for computing the orbit of x in $G = \langle S \rangle$ and a transversal T for G_x in G.

```
orb := [x]; T := [id]; ind := 0;
while #orb > ind do
  ind := ind + 1;
  for s in S do
    y := orb[ind] * s;
    if not y in orb then
      orb := orb cat [y]; T := T cat [T[ind]*s];
    end if;
  end for;
end while;
```

At the end of this algorithm, `orb` contains the orbit of x in G and the element `T[ind]` is a representative for the coset of elements mapping x to `orb[ind]`.

The only obscure point in this algorithm is that it has to be checked whether a point y is actually new or is already contained in the partial orbit. If the size of Ω is not too large (e.g. $\leq 10^6$) this is most efficiently done with a bit array of flags for every point of Ω. Once a point is found in the orbit, its flag is put to true and the inclusion check is only a lookup in the bit array. For the general case, it pays to store the orbit in a binary tree subject to some order on the elements of Ω. Then the inclusion check only requires about $\log_2(|x^G|)$ comparisons.

For a permutation group G the tasks to determine a base and to find a subgroup chain are therefore solved by some simple orbit computations. In this process we also obtain transversals for the steps in the stabilizer chain, which actually allows to put elements of G into a normal form.

5.1.4 Proposition *Let $B = (b_1, \ldots, b_m)$ be a base for G and let $G \geq G^{(1)} \geq \ldots \geq G^{(m)}$ be the corresponding stabilizer chain. Let $[G^{(i-1)} : G^{(i)}] = r_i$ (i.e. $r_i = |b_i^{G^{(i-1)}}|$) and let $T_i := \{t_{i1}, \ldots, t_{ir_i}\}$ be a transversal for $G^{(i)}$ in $G^{(i-1)}$ (as obtained from the orbit algorithm).*

Then every $g \in G$ can uniquely be written as $g = t_m t_{m-1} \ldots t_2 t_1$ with $t_i \in T_i$.

Proof: Let $g \in G$, then there exists a unique $t_1 \in T_1$ such that $b_1 g = b_1 t_1$. Then $gt_1^{-1} \in G^{(1)}$. By repeating this argument we obtain a unique $t_i \in T_i$ such that $b_i g t_1^{-1} \ldots t_{i-1}^{-1} = b_i t_i$ for every $1 \leq i \leq m$. We finally arrive at $gt_1^{-1} \ldots t_m^{-1} \in G^{(m)} = \{1\}$ and hence $g = t_m \ldots t_1$. □

The above process of writing an element $g \in G$ as a product of transversal elements by looking at the action on the base points is called *sifting* or *stripping*. It can also be used to test membership of elements $g \in S_n$ in a subgroup $G \leq S_n$. For that, we assume that we have a stabilizer chain and transversals for G as described above. We now try to sift g through this stabilizer chain, i.e. we try to find transversal elements such that $b_i g t_1^{-1} \dots t_{i-1}^{-1} = b_i t_i$. Then there are three possibilities:

(1) The sifting succeeds with $g t_1^{-1} \dots t_m^{-1} = 1$. Then $g \in G$.

(2) The sifting succeeds with $g t_1^{-1} \dots t_m^{-1} \neq 1$. Then g is an element fixing all the base elements b_i but not being the identity element, hence $g \notin G$.

(3) On some level we have $b_i g t_1^{-1} \dots t_{i-1}^{-1} \notin b_i^{G^{(i-1)}}$, hence there exists no transversal element on this level and hence $g \notin G$.

5.2 Strong generating sets

We have seen how to compute a stabilizer chain for a permutation group by iteratively constructing stabilizers of base points. But so far we have cheated in one important point: We start with a generating set S for G and can thus compute the orbit and transversal for the first level. But to move on we have require generators for the stabilizer of the first base point, and we haven't got these yet.

The solution for this problem are the *Schreier generators* which are a more general construction.

5.2.1 Lemma *Let $G = \langle S \rangle$ be a finite group, let $H \leq G$ and let T be a transversal of H in G such that $1 \in T$. For an element $g \in G$ we denote the element in T representing the coset Hg by $\overline{g}$.*

Then the set $R = \{ts(\overline{ts})^{-1} \mid t \in T, s \in S\}$ generates H. The elements of R are called Schreier generators.

Proof: Since we have $Hts = H\overline{ts}$ it is clear that all the elements of R actually lie in H, hence $\langle R \rangle \leq H$.
Now let $h \in H$ and let h be written as a product of the generators of G, i.e. $h = s_1 \dots s_k$ with $s_i \in S$. Since $1 \in T$ we can rewrite h as $1s_1(\overline{1s_1})^{-1}\overline{1s_1}s_2 \dots s_k = r_1 t_1 s_2 \dots s_k$ with $r_1 = 1s_1(\overline{1s_1})^{-1} \in R$ and $t_1 = \overline{1s_1} \in T$. In the next step we replace $t_1 s_2$ by $t_1 s_2 (\overline{t_1 s_2})^{-1} \overline{t_1 s_2}$ and define $r_2 := t_1 s_2 (\overline{t_1 s_2})^{-1} \in R$ and $t_2 := \overline{t_1 s_2}$. This yields $h = r_1 r_2 t_2 s_3 \dots s_k$. Iterating this process we end up with $h = r_1 \dots r_k t_k$ with $r_i \in R$ and $t_k \in T$. But since $r_i \in H$, we require that $h \in H t_k$, hence $t_k = 1$ and we have $h \in \langle R \rangle$. □

5.2.2 Remark Lemma 5.2.1 actually not only holds for finite groups but for finitely generated groups and subgroups of finite index. Note that a finitely generated group consists of the finite products of the generators and their inverses, hence the elements of H can be written as $h = s_1 \dots s_k$ with $s_i \in S \cup S^{-1}$. We have to show that $H = \langle R \cup R^{-1} \rangle$.

Looking at the proof, we require the elements $ts^{-1}(\overline{ts^{-1}})^{-1}$ for $t \in T$ and $s \in S$. Now let $t' := \overline{ts^{-1}} \in T$, then we have $ts^{-1}(\overline{ts^{-1}})^{-1}t's(\overline{t's})^{-1} = t(\overline{t's})^{-1}$. But we have $t' = \overline{ts^{-1}} \Leftrightarrow Ht' = Hts^{-1} \Leftrightarrow Ht's = Ht \Leftrightarrow \overline{t's} = t$, hence $t(\overline{t's})^{-1} = 1$ and thus $ts^{-1}(\overline{ts^{-1}})^{-1} \in R^{-1}$.

Using the Schreier generators we can now compute a base and a stabilizer chain for a permutation group G. We thus obtain sets of generators for every level of the stabilizer chain. Since generating sets which are adapted to a base are extremely useful, they get a special name.

5.2.3 Definition Let $B = (b_1, \ldots, b_m)$ be a base for G and let $G \geq G^{(1)} \geq \ldots \geq G^{(m)}$ be the corresponding stabilizer chain. Then a set $S \subseteq G$ with $\langle S \rangle = G$ and $\langle S \cap G^{(i)} \rangle = G^{(i)}$ is called a *strong generating set*, often abbreviated as SGS. Thus, in an SGS the elements which stabilize the first i base points actually generate the full stabilizer of these base points.

If we have generators S_0 for G and generators S_i for each level of a stabilizer chain, then the union $S := \cup_{i=0}^{m} S_i$ clearly forms a strong generating set for G.

The following lemma gives a criterion to test whether a given set of generators is a strong generating set with respect to a given base.

5.2.4 Lemma *Let $\{b_1, \ldots, b_k\} \subseteq \Omega$ and let $G \leq S_\Omega$. Assume there are $S_i \subseteq G_{b_1,\ldots,b_i}$ for $0 \leq i \leq k$ such that $G = \langle S_0 \rangle$ and $S_k = \emptyset$.*

If $\langle S_{i-1} \rangle_{b_i} = \langle S_i \rangle$, then $B = (b_1, \ldots, b_k)$ is a base for G and $S := \cup_{i=1}^{k} S_{i-1}$ is a strong generating set of G with respect to the base B.

PROOF: We proceed by induction on k. For $k = 1$ there is nothing to prove.
Let $k \geq 2$, then we can assume by induction that $S' := \cup_{i=2}^{k} S_{i-1}$ is an SGS for $\langle S_1 \rangle$ with respect to the base $B' = (b_2, \ldots, b_k)$. We have $G_{b_1} = \langle S_0 \rangle_{b_1} = \langle S_1 \rangle \leq \langle S \cap G_{b_1} \rangle$, hence $G_{b_1} = \langle S \cap G_{b_1} \rangle$ which shows that the SGS-condition is fulfilled for $i = 1$.
Now let $i \geq 2$. By induction we know that $\langle S' \cap G_{b_1,\ldots,b_i} \rangle = \langle S_1 \rangle_{b_2,\ldots,b_i}$. We therefore have $G_{b_1,\ldots,b_i} \geq \langle S \cap G_{b_1,\ldots,b_i} \rangle \geq \langle S' \cap G_{b_1,\ldots,b_i} \rangle = \langle S_1 \rangle_{b_2,\ldots,b_i} = (G_{b_1})_{b_2,\ldots,b_i} = G_{b_1,\ldots,b_i}$. Thus, we have equality everywhere and hence $G_{b_1,\ldots,b_i} = \langle S \cap G_{b_1,\ldots,b_i} \rangle$ which is the SGS-condition for $i \geq 2$. □

Using this lemma we can construct a base and SGS by the following method which improves a preliminary base and SGS until the condition $\langle S_{i-1} \rangle_{b_i} = \langle S_i \rangle$ is fulfilled on all levels. The algorithm is known as *Schreier-Sims algorithm*.

5.2.5 Schreier-Sims algorithm We assume that $G = \langle S_0 \rangle$, that $B = (b_1, \ldots, b_m)$ with $b_i \in \Omega$ and that $S_i \subseteq G_{b_1,\ldots,b_i}$ for $1 \leq i \leq m$. We say that $S := S_0 \cup \ldots \cup S_m$ is up-to-date below level i if $S_m = \emptyset$ and $\langle S_{j-1} \rangle_{b_j} = \langle S_j \rangle$ for all $i < j \leq m$. A proper base and SGS are thus found if S is up-to-date below level 0.

We start with S_0 such that $\langle S_0 \rangle = G$ and $S_1 = \emptyset$. Choose $b_1 \in \Omega$ such that $b_1 s \neq b_1$ for some $s \in S_0$ and set $B = (b_1)$, $m = 1$. Then S is up-to-date below level 1.

Now assume that S is up-to-date below level $i \geq 1$, then we check whether S is in fact up-to-date below level $i-1$. For that we compute the orbit of b_i under $\langle S_{i-1} \rangle$ and a transversal T for $\langle S_{i-1} \rangle_{b_i}$ in $\langle S_{i-1} \rangle$. Since S is up-to-date below level i we know that we have a SGS for $\langle S_i \rangle$, hence we can sift the Schreier generators $ts(\overline{ts})^{-1}$ with $t \in T$, $s \in S_{i-1}$ which we obtain for $\langle S_{i-1} \rangle_{b_i}$.

If all the Schreier generators can be sifted we conclude that $\langle S_{i-1} \rangle_{b_i} \leq \langle S_i \rangle$. Since new elements in S_i are only obtained from elements in S_{i-1} (see below) we know that $\langle S_{i-1} \rangle \geq \langle S_i \rangle$, hence we can conclude that S is now up-to-date below level $i-1$.

If an element $r = ts(\overline{ts})^{-1}$ does not sift through the SGS for $\langle S_i \rangle$, we have found an element of $\langle S_{i-1} \rangle_{b_i} \setminus \langle S_i \rangle$ and we replace S_i by $S_i \cup \{r\}$. Then S is only up-to-date below level $i+1$. If $i = m$, we have arrived at the bottom of the stabilizer chain and we define a new base point b_{m+1} which is not fixed by r, we set $S_{m+1} := \emptyset$ and adjust $m := m+1$.

Although the Schreier-Sims algorithm performs much better than producing iteratively Schreier generators for each level of the stabilizer chain, it still spends most of its time on checking Schreier generators which actually do sift through the preliminary SGS. One usually observes that if the SGS is not correct yet, then one of the first Schreier generators (not using the identity element in the transversal) will not sift. This leads to the idea to test only some randomly produced Schreier generators and accept the SGS as up-to-date on this level if they can all be sifted.

5.3 Randomized methods

The idea behind a randomized Schreier-Sims algorithm is that in an incorrect SGS many Schreier generators will no sift. One therefore tests only some Schreier generators $ts(\overline{ts})^{-1}$ with randomly chosen $t \in T$ and $s \in \langle S_{i-1} \rangle$ and concludes that S is up-to-date below level $i-1$ if all these elements sift.

By this method one arrives at a probable SGS. Of course, one will try to prove that this probable SGS is actually correct. There are (at least) three approaches to this question:

(1) Use additional information about the group, e.g. the group order.

(2) Accept a small uncertainty about the correctness and show that the probability that the SGS is incorrect is smaller than some chosen bound.

(3) Use further techniques (more efficient than testing all Schreier generators on each level) to rigorously prove the SGS correct.

The first point looks like cheating, but in many problems the group order is actually known in advance, for example if only a new base has to be constructed. If the group order is known, the SGS is clearly proved correct if the product of the orbit lengths on the different levels equals the group order.

The second approach may be worthwhile if in some intermediate computations a possibly incorrect SGS may only result in extra work but not in erroneous results.

The crucial idea here is to estimate the probability that a randomly chosen group element will not sift if the SGS is incorrect. Note that we have not seen a method to produce randomly distributed elements of a permutation group without already having a correct stabilizer chain, but we will come back to this problem in section 5.5.

5.3.1 Lemma *If S is an incorrect SGS obtained from the randomized Schreier-Sims algorithm, then an element of $g \in G$ does not sift through this system with probability $\geq \frac{1}{2}$.*

PROOF: Assume that i is the smallest index where S is incorrect, i.e. such that $\langle S_{i-1}\rangle_{b_i} \gneqq \langle S_i\rangle$. Then we have $[\langle S_{i-1}\rangle_{b_i} : \langle S_i\rangle] = d \geq 2$. Because of the minimality of i, every $g \in G$ can uniquely be written as $g = ht_{i-1}\dots t_1$ with $t_i \in T_i$. But if g is uniformly distributed over G, then h is uniformly distributed over $\langle S_{i-1}\rangle_{b_i}$, hence $h \in \langle S_i\rangle$ with chance $\frac{1}{d}$ and these are the only elements that will sift through S. Thus, the elements of G do not sift through S with probability $1 - \frac{1}{d} \geq \frac{1}{2}$. □

Thus, if we successfully test 20 randomly generated elements of G, the SGS is correct with an error probability smaller than 10^{-6}.

The following section addresses the third approach and gives a rigorous verification method for a SGS.

5.4 Verifying strong generators via presentations

Literature:

> Charles C. Sims: *Computation with finitely presented groups.* Cambridge University Press, 1994.

In this section we will see that a SGS actually provides us with a finite presentation of a permutation group. Furthermore, we will see how we can use techniques for finitely presented groups to prove a probable SGS correct.

5.4.1 Finitely presented groups

The *free group* F_n on n generators is defined as the most general group generated by n elements. It has the property that every group G generated by n elements is a factor group of F_n. Starting from this property, it is by no means obvious that such a group exists at all (apart from the infinite cyclic group $F_1 \cong \mathbb{Z}$). It is clear that free groups (if they exist) have to be enormously big, since for example every non-abelian finite simple group is now known to be generated by two elements and is thus a factor group of F_2.

The actual construction of the free group F_n is based on forming finite words in abstract generators $x_1, \dots, x_n$ and their inverses $x_1^{-1}, \dots, x_n^{-1}$ and defining multiplication as

concatenation of these words. This gives a monoid with the empty word as identity element. In order to obtain a group, we form equivalence classes of words by identifying words which can be transformed into each other by removing or inserting instances of $x_i x_i^{-1}$ or $x_i^{-1} x_i$. It is elementary but tedious to show that concatenation is well-defined on these classes and that the multiplication thus obtained is associative.

5.4.1 Definition Let $X = \{x_1, \ldots, x_n\}$ be the generators of the free group F_n, let $R \subset F_n$ be a finite subset of F_n and let R^{F_n} be the normal closure of R in F_n (i.e. the set of all conjugates of R).

Then $G := \langle X|R\rangle := F_n/R^{F_n}$ is called a *finitely presented group* with presentation $\langle X|R\rangle$. The set R is called the set of *defining relations* for G.

The presentation $\langle X|R\rangle$ defines the most general group with n generators in which the elements in R are the identity. Note that in order to obtain a quotient group we have to factor out the normal closure of R. This corresponds to the fact that we want to remove instances of $r \in R$ in arbitrary positions of words in F_n.

The free groups are constructed such that the only operations on words not changing the group elements are insertion or removal of pairs of generators and their inverses. A word therefore represents the identity element if and only if it can be collapsed to the empty word by these operations. For finitely presented groups, this is a much harder task, known as the *word problem*. Since the 1950's, we actually know that deciding whether a word represents the identity element in a finitely presented group is in general an impossible task. The famous Novikov-Boone-Britton theorem states that there exist finitely presented groups for which the word problem is unsolvable. This means that there is not only no universal algorithm which could solve the problem for every finitely presented group, but there are single groups for which no algorithm exists.

However, the Novikov-Boone-Britton theorem was not the end of dealing with finitely presented groups algorithmically, since there are important classes of groups for which the word problem is solvable, for example Coxeter groups (groups with a presentation of the form $\langle x_1, \ldots, x_n \mid (x_i x_j)^{m_{ij}}, 1 \le i \le j \le n\rangle$ with $m_{ii} = 1$ and $m_{ij} \ge 2$ for $i < j$) or soluble groups.

In the context of permutation groups it is often interesting to construct a presentation of a given group, since this allows to express group elements easily as words in the generators.

One of the nice side-issues of a SGS is that it actually provides a presentation of the permutation group almost for free.

5.4.2 Proposition *For $G = \langle S\rangle$ let $G = G^{(0)} \ge G^{(1)} \ge \ldots \ge G^{(m)} = \{1\}$ be a subgroup chain, let T_i be transversals for $G^{(i)}$ in $G^{(i-1)}$ and let $T := \cup_{i=1}^{m} T_i$. Define $X = \{x_s | s \in S\}$ to be a set of abstract generators and let φ be an epimorphism from the free group F on X onto G. For every $t \in T$ let $w(t) \in F$ be a word such that $\varphi(w(t)) = t$, i.e. a preimage of t under φ.*

Then the set $R := \{w(t)x_s w(t_1)^{-1} \ldots w(t_m)^{-1} \mid t \in T, s \in S$ with $ts = t_m \ldots t_1$ for $t_i \in T_i\}$ is a set of defining relations for G, i.e. $G \cong \langle X|R\rangle$.

PROOF: It is clear that G is a factor group of $\langle X|R\rangle$, since the generators of G fulfill the relations in R. We thus have to prove that a word $w \in F$ such that $\varphi(w) = 1$ already lies in R^F, i.e. that the kernel of φ is the smallest normal subgroup of F containing R.

Let $F^{(i)}$ be the full preimage $\varphi^{-1}(G^{(i)})$ of $G^{(i)}$. Clearly, $F^{(0)}$ is generated by S and since $\{w(t) \mid t \in T_1\}$ is a transversal of $F^{(1)}$ in $F^{(0)}$ we can conclude that $F^{(1)}$ is generated by the elements $w(t)x_s w(t_1)^{-1}$ for $t \in T_1$, $x_s \in X$ such that $b_1 ts = b_1 t_1$. Writing $ts = t_m \dots t_1$, we see that $R^F w(t)x_s w(t_1)^{-1} = R^F w(t_m) \dots w(t_2)$, hence $F^{(1)}$ is generated by R^F and the $w(t)$ with $t \in \cup_{i=2}^{m} T_i$, thus $F^{(1)}/R^F$ is generated by $\{x_s \mid s \in S \cap G^{(1)}\}$.

Repeating this argument for the levels of the stabilizer chain we see that $F^{(m)}/R^F$ is generated by $\{x_s \mid s \in S \cap G^{(m)}\} = \emptyset$. Since $F^{(m)}$ is the kernel of φ this shows that $\ker(\varphi) = R^F$. □

5.4.2 Todd-Coxeter coset enumeration

The algorithms dealing with finitely presented groups can roughly be split into two classes: The first deals with the generators and relations directly by manipulating expressions. Most of these methods are related to Knuth-Bendix methods and Tietze transformations. The second class deals with cosets and factor groups, e.g. the determination of the commutator factor group. One of the most important methods is the Todd-Coxeter coset enumeration which, given a finitely presented group and a subgroup of it, explicitly enumerates the cosets of the group by the subgroup. In particular, this algorithm allows to determine the index of a subgroup (if it is finite) and to assign each element of the group to a coset.

The set-up for Todd-Coxeter coset enumeration is as follows: Let $G = \langle X|R\rangle$ be a finitely presented group with generators $X = \{x_1, \dots, x_n\}$ and let $H = \langle X'\rangle \leq G$ be a subgroup generated by words in the x_i.

Each of the generators x_i acts as a permutation on the cosets of G/H and the goal of the algorithm is to construct these permutations systematically. The crucial (while trivial) point is that the relations in R and the generators of H (elements in X') act trivially on the cosets.

The way the algorithm proceeds resembles a game of solitaire: One sets up a couple of tables for the cosets and tries to fill these in a consistent way. In a standard set, a new coset is defined resulting in adding a new row to the tables. Sometimes we can draw conclusions from an entry in a table and identify cosets as identical which results in collapsing the tables. The goal is to fill the tables completely in a consistent way. One this is accomplished, the game is finished and we have constructed the permutation action of the generators on the cosets.

There are three types of tables in the game:

- coset table: this table records how the generators x_i permute the cosets;
- relation tables: for every relation one table is used to make sure that the relation acts trivially on the cosets;

- subgroup tables: for every generator in X' one table is used to ensure that the generator acts trivially on the cosets.

5.4.3 Example Let $G = \langle a, b \mid a^2, b^2, (ab)^3 \rangle$ and $H = \langle (ab)^2 \rangle$.

We denote the trivial coset H by 1 and start with the following tables:

coset table:

	a	b
1		

relation tables:

	a	a
1		1

	b	b
1		1

	a	b	a	b	a	b
1						1

In this case the relation tables for $a^2 = 1$ and $b^2 = 1$ can be omitted if one adds with every definition $i \cdot a = j$ in the coset table also the definition $j \cdot a = i$ (and analogously for b).

subgroup tables:

	a	b	a	b
1				1

We now proceed by defining new cosets in the coset table, for example by defining that $1 \cdot a = 2$. This means that the first coset is mapped under the first generator to the second coset. In the relation and subgroup tables, at every instance where coset 1 is mapped by a we now insert coset 2 as the image. Noting also that coset 2 is mapped under a to coset 1 we obtain the following new tables:

	a	b
1	2	
2	1	

	a	b	a	b	a	b
1	2					1
2	1					2

	a	b	a	b
1	2			1
2	1			2

Next, we define that $1 \cdot b = 3$. Using that $3 \cdot b = 1$ and the definitions already made, we obtain:

	a	b
1	2	3
2	1	
3		1

	a	b	a	b	a	b
1	2				3	1
2	1	3				2
3				2	1	3

	a	b	a	b
1	2		3	1
2	1	3		2
3		2	1	3

We now have two obvious choices for the next definition: we could define $2 \cdot b = 4$ or $3 \cdot a = 4$. Both of these definitions have the effect that the first row of the subgroup table will be closed. Closing a row (or a part of a row) in a relation or subgroup table provides us with a consequence, since we can read from the inserted element to the left and to the right, and only one of these was given as a definition. In the situation here we see that either of the two possible definitions has the other as a consequence by means of closing the first row in the subgroup table. For the first time we now have a complete coset table:

	a	b
1	2	3
2	1	4
3	4	1
4	3	2

	a	b	a	b	a	b
1	2	4	3	?	3	1
2	1	3	4	2	?	2
3	4	2	?	2	1	3
4	3	1	2	4	?	4

	a	b	a	b
1	2	4	3	1
2	1	3	4	2
3	4	2	1	3
4	3	1	2	4

We can actually fill in the subgroup table consistently, but in the relation table the last entry to be inserted is ambiguous. For example in the first row, looking from the left we should insert 1, since 3 is mapped to 1 under b. Looking from the right, we should insert

4, since 4 is mapped to 3 under a. This means that we have to identify the cosets 1 and 4. From the last row of this table (or from the coset table) we see that we also have to identify the cosets 2 and 3.

We now collapse the tables by replacing 4 by 1 and 3 by 2, which yields:

	a	b
1	2	2
2	1	1

	a	b	a	b	a	b
1	2	1	2	1	2	1
2	1	2	1	2	1	2

	a	b	a	b
1	2	1	2	1
2	1	2	1	2

These tables are consistent and show that H is a subgroup of index 2 in G and that both generators interchange the two cosets of H in G. Furthermore, we know that G is either of order 2 or 6, since H is generated by an element c satisfying the relation $c^3 = 1$ and is thus either trivial or a cyclic group of order 3. It is not hard to see, that the symmetric group S_3 has generators satisfying the relations of G, hence we have $G \cong S_3$.

In many cases it is useful to apply coset enumeration with respect to a cyclic subgroup generated by one of the generators. This often yields information about other generators with unknown order as is indicated in the following example.

5.4.4 Example Let $G = \langle a, b \mid a^3, ab^2, (ab)^3 \rangle$ and $H = \langle a \rangle$.

In this case we know that the first generator a acts trivially on the cosets and we can omit the relation table for a^3 and the subgroup table. We start by defining $1 \cdot b = 2$ and get as a consequence from the first row of the relation table for ab^2 that $2 \cdot b = 1$. This allows to complete the relation tables:

	a	b
1	1	2
2	2	1

	a	b	b
1	1	2	1
2	2	1	2

	a	b	a	b	a	b
1	1	2	2	1	?	1
2	2	1	1	2	?	2

We get a conflict in the relation table for $(ab)^3$ which shows that we have to identify the cosets 1 and 2. We conclude that $G = H = \langle a \rangle$, hence G is a cyclic group of order 3.

We could arrive at the same conclusion by enumerating with respect to the trivial subgroup $H = \langle\rangle$, but we would have to define 5 cosets before the tables collapse to 3 consistent cosets.

There are a number of interesting aspects to the Todd-Coxeter coset enumeration algorithm:

- The algorithm is known to terminate if the index $[G : H]$ is finite. However, there is no bound on the number of cosets that have to be defined intermediately before the tables collapse to the correct number (such a bound would provide a means to attack the word problem). Therefore, nothing can be concluded if the algorithm has not terminated after some time. In practical implementations, a bound is set to the number of cosets that are defined. Once this bound is reached, the algorithm stops indicating that the enumeration process was unsuccessful.

- There are different possible strategies to fill the tables. One is, to draw conclusions as early as possible, but it is sometimes more efficient to postpone conclusions to a later

stage. A lot of (heuristic) research has been done to develop an optimal strategy but it appears that different types of groups require different strategies. Current implementations use combinations of the various strategies, moving from one to the other following certain heuristic rules.

- The permutation group obtained from the action on the cosets is isomorphic to the factor group of G by the intersection of the conjugates of H (the *core* of H in G). Even for a small index of H in G this can be a group of considerable size which is then known to be a factor group of G.

5.4.3 Verification of a base and strong generating set

We can use Todd-Coxeter coset enumeration to prove that a strong generating set is correct. This procedure at the same time provides a new method to obtain a presentation of G (which is usually shorter than the one obtained directly from the SGS).

We have already seen that it is enough to show that $G_{b_i}^{(i-1)} = G^{(i)}$ on all levels $1 \leq i \leq k$ of the stabilizer chain. Since it is easy to compute the length n of the orbit $b_i^{G^{(i-1)}}$ of b_i under $G^{(i-1)}$ it is sufficient to show that $[G^{(i-1)} : G^{(i)}] = n$. Note that the orbit computation shows that the index is at least n.

The idea is now to iteratively perform Todd-Coxeter coset enumerations for $G^{(i-1)}/G^{(i)}$, starting from the bottom of the stabilizer chain. If an enumeration succeeds with n cosets, we can move one level up, otherwise we either find an element indicating that the SGS is incorrect or we improve the intermediate presentation for G.

5.4.5 Schreier-Todd-Coxeter-Sims algorithm We assume that $\langle X|R\rangle$ is a presentation for $H = G^{(i)}$ and try to prove that $[G : H] = n$ for $G = G^{(i-1)}$.

We start with a preliminary presentation $\langle \tilde{X}|R\rangle$ for G where $\tilde{X} := X \cup \{x_s | s \in S_{i-1}\}$, i.e. we add the generators of G not contained in H as abstract generators to X.

Next, we choose a bound $m > n$ for the number of cosets we are prepared to define intermediately.

If we now perform Todd-Coxeter coset enumeration until either the tables are consistently completed or we have defined m cosets, we are in one of the following three situations:

(1) The tables are completed with $n = [G : H]$ cosets. Then we have shown that $G_{b_i}^{(i-1)} = G^{(i)}$ and we can move up one level.

(2) The tables are completed with $r > n$ cosets. Then there are two elements $x, y \in G$ with $b_i x = b_i y$ but $Hx \neq Hy$. Therefore, we have $xy^{-1} \in G_{b_i} \setminus H$, hence the SGS is incorrect and we can use the element xy^{-1} to improve it.

(3) We have defined $m > n$ cosets and the tables are not complete. As in case (2) we have two elements $x, y \in G$ such that $xy^{-1} \in G_{b_i}$. We now try to sift xy^{-1} through H to check whether it is contained in H. If the sifting does not succeed, the SGS

is incorrect and we can use xy^{-1} to improve it. If the sifting does succeed, we have expressed xy^{-1} as a word w in the generators of H. This yields a new relation $xy^{-1}w^{-1}$ which we add to R and we collapse the enumeration tables by identifying Hx and Hy.

Repeating this process if necessary we arrive at less that m cosets and we continue the coset enumeration from there.

Iterating this procedure up the stabilizer chain finally yields a proof that all the subgroups have the correct indices and at the same time produces a presentation for G.

Note that in the case that we are checking a SGS obtained by a randomized method we are morally certain that it is correct. In this case we will never encounter case 2. Also, if we choose $m < 2n$ case 2 can not occur, since H could not be contained in G_{b_i} by an index > 1. A typical value used for m is $1.2 \cdot n$.

5.5 Random elements from Markov processes

We have seen that we can use a stabilizer chain to produce uniformly distributed elements in a group by multiplying together randomly chosen transversal elements from the different levels. However, we sometimes require random group elements *before* we have computed a stabilizer chain, for example if we want to check whether a random SGS is correct. It is therefore desirable to have a method that produces randomly distributed group elements just from the generators.

First we make the notion of randomly distributed group elements, by which we mean uniformly distributed group elements, precise.

5.5.1 Definition Assume that a process X produces group elements $x \in G$.

(i) The process X is said to produce uniformly distributed elements of G if $P(x = g) = |G|^{-1}$ for all $g \in G$.

(ii) For $\varepsilon > 0$ the process X is said to produce ε-uniformly distributed elements of G if $|P(x = g) - |G|^{-1}| < \varepsilon$ for all $g \in G$.

The surprisingly easy result is now that from a certain length of the products onwards, the products in the generators of G are ε-uniformly distributed.

5.5.2 Proposition *Let G be a finite group with $G = \langle S \rangle$ and assume that $1 \in S$. For $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for $n \geq n_0$ the words of length n in S are ε-uniformly distributed.*

Proof: We define a matrix $M \in \mathbb{R}^{|G| \times |G|}$ with rows and columns indexed by the elements of G such that $M_{g,h} = |S|^{-1}$ if there exists $s \in S$ with $gs = h$ and 0 otherwise. In other words, M is the incidence matrix of the Cayley graph of G with respect to S (divided by

the size of S). In particular, M is a doubly stochastic matrix, i.e. the sums of the rows and columns are all 1. Moreover, since G is a finite group, there is a bound d such that every $g \in G$ can be written as a product of at most d elements in S (we can take d to be the diameter of the Cayley graph). This means that M^d has only non-zero entries and thus M is what is called an irreducible positive matrix.

The Perron-Frobenius theorem now implies the following for M:

(1) M has eigenvalue 1 with multiplicity 1 and eigenvector $(1, 1, \ldots, 1)$.

(2) For all other eigenvalues λ of M we have $|\lambda| < 1$.

Even without applying the Perron-Frobenius theorem this can easily be seen: It is clear that $e = (1, 1, \ldots, 1)$ is an eigenvector with eigenvalue 1. Now assume there is an eigenvector $v = (v_1, \ldots, v_{|G|})$ with eigenvalue 1. We can assume that $v_1 = 1$ and $|v_i| \leq 1$. But if any $v_i < 1$, then the first component of vM is smaller than 1, hence we have $v = e$. The same argument actually shows that e is the only eigenvector with eigenvalue λ such that $|\lambda| = 1$.

We also can not have Jordan blocks for $\lambda = 1$, since that would require a vector v with $vM = v + e$ and hence $vM^n = v + n \cdot e$ which would have unbounded norm. But the elements of M^n are bounded by 1, hence the norm of vM^n is also bounded.

Now let v be an arbitrary eigenvector with eigenvalue λ. We can assume that $\|v\| = 1$, hence $\|vM^n\| = |\lambda|^n$. Again, since the norm of vM^n is bounded, we have $|\lambda| \leq 1$.

Transforming M to a matrix M' with respect to a basis of eigenvectors we see that M'^n converges towards a matrix with a 1 in position $(1, 1)$ and zeros anywhere else (note that powers of Jordan blocks for eigenvalues λ with $|\lambda| < 1$ converge to 0). From this we conclude that M^n converges to the matrix with all entries being $|S|^{-1}$. □

The way this proposition is applied in practice is as follows: In a preprocessing stage a reasonably long random word in the generators of G is produced. After that, this element is replaced by the product with one of the generators and every new element thus obtained is output as a random element.

After the preprocessing stage this method produces a random element at the cost of just one group multiplication.

Although there is no reasonable way to actually prove that the elements obtained this way are ε-uniformly distributed, the method turns out to perform extremely well in practice.

An alternative method to produce random elements is the *product replacement algorithm*. For that, assume that $G = \langle S \rangle$ with $S = \{s_1, \ldots, s_k\}$. In every step, choose $1 \leq i, j \leq k$ randomly with $i \neq j$ and replace s_i by $s_i s_j$ or $s_i s_j^{-1}$ (randomly). This produces (after a number of preprocessing steps) a random walk over the generating sets of size k and the sequence of new generators inserted can be used as a sequence of random elements.

Again, the performance of the method is much better than what can be rigorously proved about it.