Master in Mathematics

University of Luxembourg

Student Project

Computing Galois Group via Chebotarev Density

Author : **GASPERINI David**

Supervisors : **Dr. WIESE Gabor - Dr. TSAKNIAS Panagiotis**

Contents

1 Introduction

Let f be a monic irreducible polynomial of degree n with integer coefficients. The aim of this subject is to compute the Galois Group G of f, using the Chebotarev Density Theorem.

Some precision are needed. First consider the factorization of f modulo a prime p. Let F be the factorization type of f, *i.e.* the list of degrees of the irreducible factors.
Moreover G acts transitively on the set of n roots of f. Each element g of G is an automorphism which has a cycle type T.

We will state in the following a theorem which links T and F. Futhermore the Chebotarev Density Theorem claims that the size of the subset of G containing the elements whith cycle type T over the order of G equals the density of prime numbers p for which $f \ mod \ p$ has factorization type $F = T$.

Then our goal is to write an algorithm which finds the size of G and shows its structure in order to identify which subgroup of $\mathbb{S}_n$ is isomorphic to G.

First, necessary notions and theory about the objects we work with will be treated. Then settings in which the Chebotarev density theorem should be stated will be explained. Afterwards our reasoning will be presented step by step. The following section will contain a description of the algorithm together with some commented examples. To estimate the efficiency of this algorithm, some runtime tests will be presented.

2 Setting

Let us have a general overview of Galois theory.
We first take $f \in \mathbb{Z}[X]$ an irreducible monic polynomial of degree n and consider the ring $\mathbb{Q}[X]$. f is irreducible in $\mathbb{Z}[X]$, hence it is irreducible in the ring $\mathbb{Q}[X]$ which is a Principal Ideal Domain. Since f is irreducible, (f) is a maximal ideal of $\mathbb{Q}[X]$ and $K := \mathbb{Q}[X]/(f)$ is a field called number field as it is a finite extension of $\mathbb{Q}$.

Definition 2.1: Let K/k be a minimal field extension (namely there are no intermediate extensions of k), and $P \in k[X]$. K is called a break field of P if P has at least one root in K.

Theorem 2.2: *Let k be a field and P an irreducible polynomial with coefficients in k. Then $K := k[X]/(P)$ is a field extension of k inside which P has at least one root. It is also a break field of P over k.*

Proof: The fact $k[X]/(P)$ is a field containing k has been proved upside. Let us prove the second statement. Denote by x the image of X in K, so we have $P(x) = 0$ and then x is a root of P in K as well. □

Definition 2.3: α is algebraic over k if there exists a non-zero polynomial Q in $k[X]$ such that $Q(\alpha) = 0$.

Moreover we know that if α is algebraic over $\mathbb{Q}$ such that $f(\alpha) = 0$ with f irreducible (α is a primitive element of K), then $K \simeq Q[\alpha]$.
We know that K is a break field of f, and in general not its splitting field.
For instance, let us take $P(X) = X^3 - 2$. $\sqrt[3]{2}$ is the only real root of P. $\mathbb{Q}[X]/(P) = \mathbb{Q}(\sqrt[3]{2})$ is a break field of P, but not its splitting field, since $e^{i\frac{2\Pi}{3}}\sqrt[3]{2}$ does not belong to it. That is: $\mathbb{Q}(\sqrt[3]{2}) \neq \mathbb{Q}(e^{i\frac{2\Pi}{3}}\sqrt[3]{2})$...

Now let us introduce some notions relative to Galois theory, and remain the following definitions.

Let K/k be a field extension.

Definition 2.4: K/k is an algebraic extension if all elements of K are algebraic over k.

Definition 2.5: Let k be a field and α an algebraic element over k. The minimal polynomial of α is the unique monic polynomial of minimal degree annihilating α.

Definition 2.6: $\alpha \in K$ is separable over k if its minimal polynomial on k called X_α is separable over k.

Definition 2.7: Let $P \in k[X]$ irreducible. P is separable over k if all its roots in a splitting field K/k are of multiplicity 1.

Definition 2.8: Let K/k be an algebraic extension. K is separable over k if all elements of K are separable over k.

Definition 2.9: K is an algebraic extension of k. K is normal if every irreducible polynomials of $k[X]$ which has a root in K splits into linear factors in K.

Definition 2.10: A k-automorphism of K is a field automorphism $K \to K$ which is identity on k. The group $Gal(K/k)$ is the group of the k-automorphisms of K, it is called the Galois group of K over k.

Definition 2.11: Let P be a polynomial with coefficents in k and K a splitting field of P over k. The Galois group of P is $Gal(K/k)$.

Definition 2.12: K/k is a Galois extension if it is algebraic, normal and separable.

The concepts introduced by Galois allow to understand finely the structure of the different sub-extensions of K, by meaning of the correspondence theorem:

Galois's correspondence 2.13: *For a group H of automorphisms of a field L, let us define the fixed field $L^H := (x \in L / \forall \sigma \in H, \sigma(x) = x) = Fix_L(H)$.*
1. Let L/K be a Galois extension, whose Galois group is $Gal(L/K)$. There is a one to one correspondence between subgroups H of $Gal(L/K)$ and subfields J of L that contain K given by $J = L^H$.
2. L^H is a Galois extension of $K \Leftrightarrow H$ is a normal subgroup of $Gal(L/K)$, and then its Galois group is the quotient group $Gal(L/K)/H$.

Let us go back to our topic. In the situation we are considering, K is in general not a Galois extension.

Definition 2.14: A field K is algebraically closed if every non-constant polynomial $P \in K[X]$ has at least one root in K.

Definition 2.15: Let K/k be a field extension. K is an algebraic closure of k if:
1. K is algebraically closed.
2. K is algebraic over k.

Let N be the Galois closure of K over $\mathbb{Q}$. That means that N is the intersection of all Galois extensions of $\mathbb{Q}$ containing K. By **Definitions 2.9** and **2.15**, since N is an algebraic closure and a normal extension, f splits into linear factors in N.

Remark: *N is a field in which f splits. Let $S = \{\alpha / f(\alpha) = 0\}$. We obviously have $S \subset N$.*

Now consider $G := G(N/\mathbb{Q})$. For any $\sigma \in G$, for any $\alpha \in S$, $\sigma(\alpha)$ is a root of $\sigma(f) = f$ since the coefficients of f are in $\mathbb{Z}$, so stay the same.
Hence G will always send an element of $S = \{\alpha_1, ..., \alpha_n\}$ to another element of S. It naturally follows the analyse of how G behaves with the elements of S. Fix a bijection between the subset S of N and the set $\{1, ..., n\}$.

Now consider the action of G on S seen as a partition of $[n]$. That is why we will look at G as a subgroup of $\mathbb{S}_n$, via the previous bijection.

Definition 2.16: Let K be a field and $P = X^n + \sum_{i=1}^{n-1} a_i X^i$ a monic polynomial of degree n, with coefficients in K. Let L be a field extension of K in which P is separable and denote by $x_1, ..., x_n$ its roots. Then $\prod_{1 \leqslant i,j \leqslant n}(x_i - x_j)^2 = \triangle_P$

Remark: *In particular, f is irreducible $\Rightarrow$ $\{f$ has no multiple root $\Leftrightarrow \triangle_f \neq 0.\}$*

This statement is clear, by definition of $\triangle_f$.
Furthermore it implies that G acts transitively on S, as will state the following theorem.

Proposition 2.17: *Let L/K be a field extension. Then L/K is finite and normal $\Rightarrow$ L is the splitting field of a polynomial with coefficients in K.*

Proof: Let L/K be a finite normal field extension. Take a basis $(x_1, ..., x_d)$ of the K-vector space L. Let $P_i \in K[X]$ be the minimal polynomial of x_i. Since L/K is normal, each P_i splits in L and $Q := \prod_1^d P_i$ splits too. Then since L is generated over K by all the x_i which are roots of Q, the field L is the splitting field of Q on $K[X]$. $\square$

Remark: *We know that every finite field extension is algebraic. Hence if we assume that L is in addition separable, then L is a Galois extension and we have the equivalence.*

Proposition 2.18: *Let F and L be two fields, $\varphi : F \to L$ a field isomorphism. Let $P \in F[X]$ be a non-zero polynomial. Consider K a splitting field of P over F and M a splitting field of $\varphi(P)$ over L. Then φ extends to an isomorphism $\phi : K \to M$ such that $\phi \mid_F = \varphi$.*

Proof: We prove the statement by induction on $[K : F]$.
-If $[K : F] = 1$ then $\varphi = \phi$ works.
-If $[K : F] > 1$, consider $R \in F[X]$ an irreducible factor of P such that $deg(R) > 1$. This one must exist, otherwise P would split in F. Let $\varphi(R)$ be the corresponding factor of $\varphi(P)$. Let $x \in K$ be a root of R and $y \in M$ a root of $\varphi(R)$. Since $F(x) \simeq F[X]/(R) \simeq L[X]/(\varphi(R)) \simeq L(y)$, then φ extends to an isomorphism $\varphi' : F(x) \to L(y)$ such that $\varphi'(x) = y$.
Hence, by induction, φ' extends to an isomorphism $\phi : K \to M$ which restricts to φ' on $F(x)$ and then to φ on F. $\square$

Theorem 2.19: *Let $P \in K[X]$ be a separable polynomial splitting in L. The action of $Gal(L/K)$ on the set of roots of P in L is transitive $\Leftrightarrow$ P is irreducible in $K[X]$.*

Proof:
($\Rightarrow$) Let us suppose that P is not irreducible. Then we can write $P = Q.R$ with Q, $R \in K[X]$, both non-constant. Since P is separable, neither Q nor R have common roots. Each element of $Gal(L/K)$ $(= Aut_K(L))$ sends each root of Q (inside L) on a root of Q. Hence the roots of R are not in the $Gal(L/K)$-orbits of the roots of Q and then, the action on the roots of P is not transitive.
($\Leftarrow$) Assume now that P is irreducible. Let x and y be two roots of P in L. By **Definition 2.1**, $K(x)$ and $K(y)$ are subfield of L and in particular are break fields of P. Since P is

irreducible, $K(x) \simeq K[X]/(P) \simeq K(y)$. In other words they are K-isomorphic: there exists a K-isomorphism $\sigma : K \to K$ which is identity on K and such that $\sigma(x) = y$. Now consider L as a splitting field of P over both $K(x)$ and $K(y)$ (we see P respectively as a polynomial with coefficients in $K(x)$ and in $K(y)$). By **Proposition 2.18**, we obtain an extension of σ to all of L, *i.e.* an automorphism that fixes K and maps x to y. Hence $Gal(L/K)$ acts transitively on the set of roots of P in L. □

To finish this section, we will take an example of theoretical computation. Consider the following polynomial: $P(X) = X^2 - 2X - 1$. It has two real roots, namely $x_1 = 1 + \sqrt{2}$ and $x_2 = 1 - \sqrt{2}$. Now let L be the set of the polynomials in two variables for which (x_1, x_2) is a root. We can see that $P_1(X, Y) = X^2 - 2X - 1$ and $P_2(X, Y) = X + Y - 2$ are some polynomials of L which verify this property.
Remark that for each polynomial of L, (x_2, x_1) is also a root. Then $id : (x_1, x_2) \mapsto (x_1, x_2)$ and $\sigma : (x_1, x_2) \mapsto (x_2, x_1)$ fix L, they are the two automorphisms of the Galois group of P which is isomorphic to $\mathbb{S}_2$.

3 Reasoning

We will state the reasoning and some results representing the theoretical side of the computing.

3.1 Dedekind's theorem

Definition 3.1: Let K be a field of characteristic a prime p (that means the order for the addition law of the unit for the multiplication law). The field morphism:
$Fr : K \to K$
$x \mapsto x^p$
is called the Frobenius morphism Fr.

Theorem 3.2: *Let H be a finite group of which elements are automorphisms of a field L, Let $L^H = Fix_L(H)$ the corresponding fixed field (as seen above in Galois's Correspondence). Then L/L^H is a Galois extension the Galois group of which is H*

Corollary 3.3: *A finite extension L/K is a Galois extension $\Leftrightarrow K = L^{Gal(L/K)}$.*

Lemma 3.4: *Let p be a prime, $q = p^m$ with $m \in \mathbb{N}$. Then $\mathbb{F}_q/\mathbb{F}_p$ is a Galois extension and $Gal(\mathbb{F}_q/\mathbb{F}_p)$ is cyclic, generated by Fr: $Gal(\mathbb{F}_q/\mathbb{F}_p) = \langle Fr \rangle$.*

Proof: For a finite field, Fr is an automorphism: $Fr \in Gal(\mathbb{F}_q/\mathbb{F}_p)$. We will show that the order of Fr is $m = [\mathbb{F}_q/\mathbb{F}_p] = \mid Gal(\mathbb{F}_q/\mathbb{F}_p) \mid$. This implies that $\mathbb{F}_q/\mathbb{F}_p$ is a Galois Extension (**Corollary 3.3**) and since the order of a finite group equals the order of one of its generators, that Fr generates the Galois group.
The multiplicative group $\mathbb{F}_q^*$ is cyclic, its order is $q - 1$. Let ω be a generator. Since $q = p^m$, for all $\alpha \in]0; m[$, $\frac{Fr^\alpha(\omega)}{\omega} = \omega^{p^\alpha - 1}$, hence $Fr^m = Id$ and $Fr^\alpha \neq Id$. □

Even if f is irreducible over $\mathbb{Q}$, it does not mean that the reduction of f modulo a prime p is irreducible.

However it can be factored into a product (called k in the algorithm) of irreducible polynomials over the finite field $\mathbb{F}_p$. Let us name the degrees of the irreducible factors the factorisation type of f over $\mathbb{F}_p$.
We should find a link between the cycle types that occur in G and the factorization types of f (for some different primes p) in order to match f to its Galois group, and then we use the following Dedekind's theorem.

Now let us define the notion of a transitive subgroup.

Definition 3.5: Let H be a subgroup of $\mathbb{S}_n$. H is called transitive if it action over $[n]$ is transitive, *i.e.* : if $\forall i, j \in \{1, ..., n\}, \exists \sigma \in H / \sigma(i) = j$.

Proposition 3.6: *Let p be a prime, $f \in \mathbb{Z}[X]$ monic. For $x \in \mathbb{Z}$, denote by x_i the simple roots of f on its splitting field. Assume that $\mathbb{Q} \subset E \simeq \mathbb{Q}(x_1, x_2...)$ a splitting field of f and $\mathbb{F}_p \subset L$ a splitting field of $\overline{f}$, the factorization of f modulo p. Suppose $\overline{f}$ is separable. There exists an injective group morphism*

$$\begin{array}{ccc} \Psi : Gal(L/\mathbb{F}_p) & \to & Gal(E/\mathbb{Q}) \\ \sigma & \mapsto & \Psi(\sigma) \end{array}$$

such that $\rho(\Psi(\sigma)) = \sigma(\rho)$ for all $\rho \in Hom(\mathbb{Z}[x_1, x_2...], L)$

Dedekind's theorem: *Let f in $\mathbb{Z}[X]$ be a monic irreducible polynomial of degree n, and let G_f be its Galois group. Put $\overline{f}(x)$ the factorization of $f(x)$ modulo a prime p which does not divide $\triangle_f$, such that $\overline{f}(x)$ is a product of irreducible polynomials in $\mathbb{F}_p[X]$ of degrees $n_1, n_2...n_r$ with $\sum n_i = n$. Then G_f contains a permutation which has the cycle type $(n_1, n_2, ..., n_r)$.*

Proof of Dedekind's theorem: Consider the factorisation of f into irreducible polynomials $\overline{f}_i$ of $\mathbb{F}_p[X]$. Since $p \nmid \triangle(f)$, f has simple roots $x_1, ...x_n$ and then each of the factors only appears once. Let $\overline{x_i}$ be the residue class of x_i modulo p. Let us denote $\overline{f} = \overline{f}_1 \overline{f}_2 ... \overline{f}_r$ the factorization.
$\mathbb{F}_p[\overline{x_1}, \overline{x_2}...]$ is a splitting field of $\overline{f}$ and $\overline{G} = Gal(\mathbb{F}_p[\overline{x_1}, \overline{x_2}...]/\mathbb{F}_p)$ is cyclic (**Lemma 3.4**). Consider $\overline{\sigma}$ a generator of $\overline{G}$. σ can be written as a product of cycles, as follows: $\sigma = (12...l)(l+1...)...$
Since the irreducible factors $\overline{f}_i$ of $\overline{f}$ are in bijection with the subgroups of $\overline{G}$ which act transitively on the set of roots of $\overline{f}_i$, the numbers occuring in the cycles $(12...l)(l+1...)...$ must exactly represent each roots of $\overline{f}_1 \overline{f}_2 ... \overline{f}_r$
Then if we know the degrees of each $\overline{f}_i$, the length of each cycles of which σ consists is known, *i.e.* we have the cycle type of σ.
In conclusion, since $\overline{G} \subseteq G_f$ (**Proposition 3.6**), $\sigma \in G_f$. $\square$

3.2 Chebotarev density theorem

Natural density is used in the Chebotarev density theorem. Let us define this notion.

Definition 3.7: Let $\Gamma \subset \mathbb{Z}$ be the set of prime numbers. Let $\mu \subset \Gamma$ be a subset and suppose that the limit $d(\mu) := \lim_{y \to \infty} \frac{\sharp\{p \in \mu : p \leqslant y\}}{\sharp\{p \in \Gamma : p \leqslant y\}}$ exists. Then $d(\mu)$ is called the natural density of μ.

Chebotarev density theorem: *Let $f \in \mathbb{Z}[X]$ be a monic irreducible polynomial. Assume that $\triangle_f$ does not vanish. Let T be a cycle type of some elements of the Galois group G of f. Then the set $\{p$ prime, $p \nmid \triangle_f \mid \overline{f}$ has factorisation type $F = T\}$ has a density, and this density equals $\frac{n(T)}{|G|}$ where $n(T) = \sharp\{\sigma \in G \mid \sigma$ has cycle type $T\}$.*

Thus we have all the steps we need to build our reasoning.
First of all, we know how to characterize N, and that f which is irreducible, totally splits in N. In addition, since G acts transitively on the roots of f, we know with the results above, that G is isomorphic to a transitive subgroup of $\mathbb{S}_n$ which we will describe forward.
Dedekind's theorem allows us to link in a bijective way the element of G to the factorization types of the factorizations f *mod* p_i with $(p_i)_i$ a family of primes such that p_i do not divide $\triangle_f$. Then, knowing all the factorization types of f (infinitely many since density implies the use of infinitely many p_i), we get all kinds of factorization type (finitely many) that exist, and then we have the different kinds of cycle type that G contains, with their relative proportions.
The problem is that to know all these factorization types we have to compute f *mod* p_i for infinitely many primes p_i, since we deal with density. Moreover, even if we know what are the different kinds of cycle type of the Galois group, for a considered cycle type T, we don't know the number of elements in G which have cycle type T.
This number is given by the Chebotarev theorem, which allows us to match precisely and in a formalised way G to a transitive subgroup of $\mathbb{S}_n$.

4 Computing Galois group

The following will describe how to apply on a computer all the previous reasoning and notions.

4.1 Algorithm construction

SAGE is the software on which we will work to write the algorithm, presented in the Appendix. It has be choosen because it is free and really adapted to theoretical reasoning.

The input of the algorithm is a monic irreducible polynomial f of degree n with integer coefficients, and an integer bound y for the number of primes considered.
The output is the order of G if y is choosen large enough and the frequencies of all kinds of cycle type that occur in G. They will allow us to match G with a transitive subgroup of $\mathbb{S}_n$ to which G is isomorphic.

The algorithm proceeds as follows:
First it calculates $\triangle_f$. Then it determines the set S of all primes p less than or equal to y which do not divide $\triangle_f$.
For every $p \in S$, it computes the factorization of f *mod* p, and then its factorization type F. In fact we obtain an array of factorization types. We saw that for every such prime p there exists σ of G with cycle type F. Then we can compute how many times each cycle type occurs in G. Moreover, since identity (with cycle type $(1, 1, 1, ...)$) occurs only once in a group, we can calculate the size of G (still providing that y is large enough).
To identify G, we will make use of a list of transitive subgroups of $\mathbb{S}_n$.

4.2 Matching with group of permutations $\mathbb{S}_d$

In this section will appear the concrete results of this subject. At first time we recall some basis about the symetric group, then we will have the necessary tools to compute $Gal(N/\mathbb{Q})$.

4.2.1 The symetric group

Let n be a natural number. We define $\mathbb{S}_n$ as the set of all the bijections from $[n]$ to $[n]$. Together with the composition law $\circ$, $\mathbb{S}_n$ is a group of order $n!$.

Let us describe the transitive subgroups of $\mathbb{S}_n$ for $n \leqslant 4$. Actually we should know these subgroups for all n. We can find them on softwares as GAP or Magma.

$\mathbb{S}_1 = \{id\}$

$\mathbb{S}_2 = \{id, (12)\}$
Transitive subgroups:
$-\mathbb{S}_1$
$-\mathbb{S}_2$

$\mathbb{S}_3 = \{id, (12), (13), (23), (123), (132)\}$
Transitive subgroups:
$-\mathbb{S}_1$
$-\mathbb{A}_3 =< (123) >= \{id, (123), (132)\}$
$-\mathbb{S}_3$

$\mathbb{S}_4 = \{id, (12), (13), (14), (23), (24), (34), (123), (132), (124), (142), (134), (143), (234), (243), (12)(34), (13)(24), (14)(23), (1234), (1324), (1243), (1342), (1432), (1423)\}$
Transitive subgroups:
$-\mathbb{S}_1$
$-\mathbb{V}_4 = \{id, (12)(34), (13)(24), (14)(23)\}$
$-\mathbb{C}_4 = \{id, (31)(24), (1234), (1432)\}$ (3 such subgroups)
$-\mathbb{D}_4 = \{id, (12), (34), (12)(34), (13)(42), (14)(32), (1324), (1423)\}$ (3 such subgroups)
$-\mathbb{A}_4 = \{id, (123), (132), (124), (142), (134), (143), (234), (243), (12)(34), (13)(24), (14)(23)\}$
$-\mathbb{S}_4$

Here we present some results which characterize transitive subgroups of $\mathbb{S}_n$.

Lemma 4.1: *Let H be a transitive subgroup of $\mathbb{S}_n$. Suppose H contains a 2-cycle and a $(n-1)$-cycle. Then $H = \mathbb{S}_n$.*

Proof: After a suitable reordering, let $(1\ 2...n-1)$ be the $(n-1)$-cycle and consider the 2-cycle $(i\ j)$ with $i, j \in [n]$. By transitivity of H there exists $\sigma \in H$ such that $\sigma(i\ j)\sigma^{-1} = (k\ n)$ with $k \in \{1, ..., n-1\}$. If $k \leq n-2$, $(1\ 2...n-1)(k\ n)(1\ 2...n-1)^{-1} = (k+1\ n)$. If $k = n-1$, then the conjugation by the $(n-1)$-cycle gives $(1\ n)$. Hence we obtain all cycles $(1\ n)$, $(2\ n)$,... $(n-1\ n)$ which generate $\mathbb{S}_n$. $\square$

Lemma 4.2: *Let p be a prime and H a subgroup of $\mathbb{S}_p$. If H contains a 2-cycle and a p-cycle, then $H = \mathbb{S}_n$.*

Proof: Let τ be a transposition of H. Without loose of generality, assume $\tau = (1\ 2)$. Then a suitable power of a p-cycle in H has the form $\sigma = (1\ 2...)$. After relabelling the objects other than $1, 2..$, we can assume that $\sigma = (1\ 2\ 3...p)$. But then $\forall i \in \{0, 1, ...p-2\}$, $\sigma^i \tau \sigma^{-i} = (i+1\ i+2) \in H$, and these elements generate $\mathbb{S}_p$. $\square$

Remark: *The two last results are part of another reasoning which aims at computing G without making use of Chebotarev density theorem. Actually they allow to reconstruct the wanted transitive subgroup (if this one is for example $\mathbb{S}_n$, as soon as we get (by Dedekind's theorem) a 2-cycle and a $(n-1)$-cycle).*

4.2.2 Matching Galois groups

Here we concretely compute the Galois group G of a polynomial f by using the previously described algorithm. Denote by $n(T)$ the number of cycle type T appearance. Some examples of matching will then illustrate our reasoning:

- $f(x) = x^2 - 2x - 1$, $y = 10^4$

```
evaluate
    f = x^2 - 2*x - 1
    There are 1228 primes less than 10000 which don't divide disc(f)
    The order of G is 2.0
    Identity occurs 603 times, ie: 48. %
    cycle type (2,) occurs 625 times,ie: 51. % n (2,) = 1.0
```

  $| G | = 2$
  $n((2)) = 1$
  $\Rightarrow G \approx \mathbb{S}_2$, as we saw in the end of section 2.

- $f(x) = x^2 + 47x - 10$, $y = 10^4$

```
    f = x^2 + 47*x - 10
    There are 1227 primes less than 10000 which don't divide disc(f)
    The order of G is 2.0
    Identity occurs 604 times, ie: 48. %
    cycle type (2,) occurs 623 times,ie: 51. % n (2,) = 1.0
```

```
evaluate
```

  $| G | = 2$
  $n((2)) = 1$
  $\Rightarrow G \approx \mathbb{S}_2$

- $f(x) = x^3 + 19$, $y = 10^4$

```
f = x^3 + 19
There are 1227 primes less than 10000 which don't divide disc(f)
The order of G is 6.0
Identity occurs 198 times, ie: 16. %
cycle type (2,) occurs 617 times,ie: 50. % n (2,) = 3.0
cycle type (3,) occurs 412 times,ie: 34. % n (3,) = 2.0
```

evaluate

$| G | = 6$
$n((2)) = 3$, $n((3)) = 2$
$\Rightarrow G \approx \mathbb{S}_3$

- $f(x) = x^3 - 3x^2 + 1$, $y = 10^4$

```
f = x^3 - 3*x^2 + 1
There are 1228 primes less than 10000 which don't divide disc(f)
The order of G is 3.0
Identity occurs 404 times, ie: 32. %
cycle type (3,) occurs 824 times,ie: 67. % n (3,) = 2.0
```

$| G | = 3$
$n((3)) = 2$
$\Rightarrow G \approx \mathbb{A}_3$

- $f(x) = x^3 + 24x^2 - 31x + 2$, $y = 10^5$

```
f = x^3 + 24*x^2 - 31*x + 2
There are 9589 primes less than 100000 which don't divide disc(f)
The order of G is 6.0
Identity occurs 1540 times, ie: 16. %
cycle type (2,) occurs 4852 times,ie: 51. % n (2,) = 3.0
cycle type (3,) occurs 3197 times,ie: 33. % n (3,) = 2.0
```

evaluate

$| G | = 6$
$n((2)) = 3$, $n((3)) = 2$
$\Rightarrow G \approx \mathbb{S}_3$

- $f(x) = x^4 + 5x^3 + x^2 - 2$, $y = 10^5$

```
evaluate
    f = x^4 + 5*x^3 + x^2 - 2
    There are 9590 primes less than 100000 which don't divide disc(f)
    The order of G is 24.
    Identity occurs 383 times, ie: 4.0 %
    cycle type (2,) occurs 2324 times,ie: 24. % n (2,) = 6.0
    cycle type (3,) occurs 3238 times,ie: 34. % n (3,) = 8.0
    cycle type (4,) occurs 2443 times,ie: 25. % n (4,) = 6.0
    cycle type (2, 2) occurs 1202 times,ie: 13. % n (2, 2) = 3.0
```

$| G | = 24$
$n((2)) = 6$, $n((3)) = 8$, $n((4)) = 6$, $n((2,2)) = 3$
$\Rightarrow G \approx \mathbb{S}_4$

- $f(x) = x^4 + 3x^2 - 1$, $y = 10^4$

```
    f = x^4 + 3*x^2 - 1
    There are 1227 primes less than 10000 which don't divide disc(f)
    The order of G is 8.0
    Identity occurs 150 times, ie: 12. %
    cycle type (2,) occurs 305 times,ie: 25. % n (2,) = 2.0
    cycle type (2, 2) occurs 458 times,ie: 37. % n (2, 2) = 3.0
    cycle type (4,) occurs 314 times,ie: 26. % n (4,) = 2.0
```

$| G | = 8$
$n((2)) = 2$, $n((2,2)) = 3$, $n((4)) = 2$
$\Rightarrow G \approx \mathbb{D}_4$

- $f(x) = x^4 + 3x^2 + 1$, $y = 10^4$

```
    f = x^4 + 3*x^2 + 1
    There are 1227 primes less than 10000 which don't divide disc(f)
    The order of G is 4.0
    Identity occurs 302 times, ie: 24. %
    cycle type (2, 2) occurs 925 times,ie: 75. % n (2, 2) = 3.0
```

$| G | = 4$
$n((2,2)) = 3$
$\Rightarrow G \approx \mathbb{V}_4$

- $f(x) = x^5 + 10x^3 - 10x^2 + 35x - 18$, $y = 10^6$

```
evaluate
    f = x^5 + 10*x^3 - 10*x^2 + 35*x - 18
    There are 78495 primes less than 1000000 which don't divide disc(f)
    The order of G is 64.
    Identity occurs 1257 times, ie: 1.5 %
    cycle type (5,) occurs 31314 times,ie: 40. % n (5,) = 24.
    cycle type (3,) occurs 26404 times,ie: 34. % n (3,) = 21.
    cycle type (2, 2) occurs 19520 times,ie: 25. % n (2, 2) = 15.
```

  As shown by computing:
  $\mid G \mid = 64$, $n((3)) = 21$, $n((5)) = 24$, $n((2,2)) = 15$
  Considering that $\mathbb{A}_5$ is the transitive subgroup of $\mathbb{S}_5$ of degree 60 we might conclude that$\Rightarrow G \approx \mathbb{A}_5$

- $f(x) = x^5 - 3x - 1$, $y = 10^6$

```
evaluate
    f = x^5 - 3*x - 1
    There are 78497 primes less than 1000000 which don't divide disc(f)
    The order of G is 120.
    Identity occurs 664 times, ie: 0.75 %
    cycle type (5,) occurs 15632 times,ie: 20. % n (5,) = 23.
    cycle type (2, 3) occurs 13146 times,ie: 17. % n (2, 3) = 19.
    cycle type (2, 2) occurs 9787 times,ie: 12. % n (2, 2) = 14.
    cycle type (2,) occurs 6436 times,ie: 8.2 % n (2,) = 9.0
    cycle type (3,) occurs 13192 times,ie: 17. % n (3,) = 19.
    cycle type (4,) occurs 19640 times,ie: 25. % n (4,) = 29.
```

  $\mid G \mid = 120$
  $n((2)) = 9$, $n((3)) = 19$, $n((4)) = 29$, $n((5)) = 23$, $n((2,2)) = 14$, $n((2,3)) = 19$
  $\Rightarrow G \approx \mathbb{S}_5$

  We can also note that if we compute G choosing $y = 10^5$, our output starts to be unreliable:

```
evaluate
    f = x^5 - 3*x - 1
    There are 9591 primes less than 100000 which don't divide disc(f)
    The order of G is 110.
    Identity occurs 84 times, ie: 1.0 %
    cycle type (5,) occurs 1902 times,ie: 20. % n (5,) = 22.
    cycle type (2, 3) occurs 1590 times,ie: 17. % n (2, 3) = 18.
    cycle type (2, 2) occurs 1242 times,ie: 13. % n (2, 2) = 14.
    cycle type (2,) occurs 742 times,ie: 7.7 % n (2,) = 8.0
    cycle type (3,) occurs 1615 times,ie: 17. % n (3,) = 19.
    cycle type (4,) occurs 2416 times,ie: 25. % n (4,) = 28.
```

- $f(x) = x^7 - 3x^3 + 3$, $y = 10^6$

```
f = x^7 - 3*x^3 + 3
There are 78496 primes less than 1000000 which don't divide disc(f)
The order of G is 5100.
Identity occurs 15 times, ie: 0.016 %
cycle type (2, 2, 2) occurs 1591 times,ie: 2.0 % n (2, 2, 2) = 110.
cycle type (3, 3) occurs 4371 times,ie: 5.6 % n (3, 3) = 290.
cycle type (2,) occurs 311 times,ie: 0.40 % n (2,) = 20.
cycle type (3,) occurs 1062 times,ie: 1.4 % n (3,) = 70.
cycle type (4,) occurs 3284 times,ie: 4.2 % n (4,) = 220.
cycle type (5,) occurs 7708 times,ie: 9.8 % n (5,) = 510.
cycle type (2, 3) occurs 6485 times,ie: 8.3 % n (2, 3) = 430.
cycle type (6,) occurs 13177 times,ie: 17. % n (6,) = 880.
cycle type (2, 2) occurs 1651 times,ie: 2.1 % n (2, 2) = 110.
cycle type (7,) occurs 11023 times,ie: 14. % n (7,) = 740.
cycle type (2, 4) occurs 9801 times,ie: 12. % n (2, 4) = 660.
cycle type (2, 2, 3) occurs 3312 times,ie: 4.2 % n (2, 2, 3) = 220.
cycle type (3, 4) occurs 6729 times,ie: 8.6 % n (3, 4) = 450.
cycle type (2, 5) occurs 7976 times,ie: 10. % n (2, 5) = 530.
```

Knowing that $|\ \mathbb{S}_7\ | = 5040$, we may match G with $\mathbb{S}_7$.

Moreover it can appear that y is not large enough to fall on the identity cycle type, and then computing the size of G is not possible in this way. Therefore only the occurence frequency of all kinds of cycle type allows us to identify G, through a probabilistic analyze. Here we illustrate this phenomena:

- $f(x) = x^8 + x^7 + 123x^6 - 23x^4 + 82x - 1$, $y = 10^4$

```
evaluate
    f = x^8 + x^7 + 123*x^6 - 23*x^4 + 82*x - 1
    There are 1227 primes less than 10000 which don't divide disc(f)
    cycle type (2, 6) occurs 105 times,ie: 8.6 %
    cycle type (2, 3, 3) occurs 39 times,ie: 3.2 %
    cycle type (2, 2, 4) occurs 28 times,ie: 2.3 %
    cycle type (2, 3) occurs 25 times,ie: 2.0 %
    cycle type (2, 5) occurs 120 times,ie: 9.8 %
    cycle type (2, 2, 3) occurs 52 times,ie: 4.2 %
    cycle type (3, 5) occurs 80 times,ie: 6.5 %
    cycle type (2, 2, 2) occurs 9 times,ie: 0.73 %
    cycle type (3, 3) occurs 38 times,ie: 3.1 %
    cycle type (2, 2, 2, 2) occurs 5 times,ie: 0.41 %
    cycle type (2,) occurs 1 times,ie: 0.082 %
    cycle type (4, 4) occurs 34 times,ie: 2.8 %
    cycle type (3,) occurs 7 times,ie: 0.57 %
    cycle type (5,) occurs 39 times,ie: 3.2 %
    cycle type (4,) occurs 17 times,ie: 1.4 %
    cycle type (8,) occurs 155 times,ie: 13. %
    cycle type (6,) occurs 104 times,ie: 8.5 %
    cycle type (2, 2) occurs 7 times,ie: 0.57 %
    cycle type (7,) occurs 199 times,ie: 16. %
    cycle type (3, 4) occurs 97 times,ie: 7.9 %
    cycle type (2, 4) occurs 66 times,ie: 5.4 %
```

The question of the compromise between precision and runtime appears, notably with higher degree polynomials as those which follow.

- $f(x) = x^7 + 2x^5 - 7x + 1$ $y = 10^6$

```
f = x^7 + 2*x^5 - 7*x + 1
There are 78496 primes less than 1000000 which don't divide disc(f)
The order of G is 10000.
Identity occurs 8 times, ie: 0.012 %
cycle type (2, 2, 2) occurs 1667 times,ie: 2.1 % n (2, 2, 2) = 210.
cycle type (3, 3) occurs 4276 times,ie: 5.4 % n (3, 3) = 540.
cycle type (2,) occurs 312 times,ie: 0.40 % n (2,) = 39.
cycle type (3,) occurs 1117 times,ie: 1.4 % n (3,) = 140.
cycle type (4,) occurs 3241 times,ie: 4.1 % n (4,) = 400.
cycle type (5,) occurs 7808 times,ie: 10. % n (5,) = 980.
cycle type (2, 3) occurs 6507 times,ie: 8.3 % n (2, 3) = 820.
cycle type (6,) occurs 13086 times,ie: 17. % n (6,) = 1600.
cycle type (2, 2) occurs 1673 times,ie: 2.1 % n (2, 2) = 210.
cycle type (7,) occurs 11267 times,ie: 14. % n (7,) = 1400.
cycle type (2, 4) occurs 9814 times,ie: 12. % n (2, 4) = 1200.
cycle type (2, 2, 3) occurs 3223 times,ie: 4.1 % n (2, 2, 3) = 400.
cycle type (3, 4) occurs 6503 times,ie: 8.3 % n (3, 4) = 820.
cycle type (2, 5) occurs 7994 times,ie: 10. % n (2, 5) = 1000.
```

$deg(f) = 7$ then $\mid G \mid$ should not be greater than $7! = 5040$. We cannot conclude in this way.

- $f(x) = x^9 - x^8 - x^7 - 2x^5 + 4x^4 - 5x^2 + 1$, $y = 10^6$
By [4], the Galois group of f is $\mathbb{A}_9$, which order is $\frac{9!}{2} = 181440$. Here is another example of unreliability, in spite of the choice $y = 10^6$.

```
f = x^9 - x^8 - x^7 - 2*x^5 + 4*x^4 - 5*x^2 + 1
There are 78494 primes less than 1000000 which don't divide disc(f)
The order of G is 82000.
Identity occurs 1 times, ie: 0.0015 %
cycle type (2, 6) occurs 13039 times,ie: 17. % n (2, 6) = 13000.
cycle type (2, 2, 5) occurs 3920 times,ie: 5.0 % n (2, 2, 5) = 3900.
cycle type (9,) occurs 17497 times,ie: 22. % n (9,) = 17000.
cycle type (2, 2, 3) occurs 3263 times,ie: 4.2 % n (2, 2, 3) = 3300.
cycle type (3, 5) occurs 10588 times,ie: 13. % n (3, 5) = 11000.
cycle type (3, 3, 3) occurs 973 times,ie: 1.2 % n (3, 3, 3) = 980.
cycle type (3, 3) occurs 1463 times,ie: 1.9 % n (3, 3) = 1500.
cycle type (2, 2, 2, 2) occurs 382 times,ie: 0.49 % n (2, 2, 2, 2) =
380.
cycle type (4, 4) occurs 4990 times,ie: 6.4 % n (4, 4) = 5000.
cycle type (3,) occurs 63 times,ie: 0.080 % n (3,) = 63.
cycle type (5,) occurs 1282 times,ie: 1.6 % n (5,) = 1300.
cycle type (2, 2) occurs 165 times,ie: 0.21 % n (2, 2) = 160.
cycle type (7,) occurs 11255 times,ie: 14. % n (7,) = 11000.
cycle type (2, 3, 4) occurs 6360 times,ie: 8.1 % n (2, 3, 4) = 6300.
cycle type (2, 4) occurs 3253 times,ie: 4.1 % n (2, 4) = 3300.
```

Remark that the choice $y = 10^7$ will output a more precise computation, but will make the runtime explode.

4.3 Runtime study

As stated by Chebotarev density theorem, all data about G are available via the density previously mentionned, which cannot be exactly obtained with a computer. That is why we should look at a way to compute this result efficiently, finding compromise between runtime and reliability.

The greater y will be chosen, the higher the reliability of the output (*i.e.* the size and structure of G) will be.
Indeed, if y is large enough, the number of prime p which do not divide $\triangle(f)$ will be large enough to find and compute all kinds of cycle type in G and then well describe this group.
Concerning any group H, we know that identity represents $\frac{1}{|H|}$ of all the elements of H.
Then, after getting the frequency of the identity cycle type, we can immediatly find $\mid G \mid$. The only problem is that this frequency has to be reliable (in case it is found! See the example above with $f(x) = x^8 + x^7 + 123x^6 - 23x^4 + 82x - 1$).
Hence, for high degree polynomials, y should be taken very large. And our computer has to run very fast... Obviously there must exist some optimized algorithms which allow in an efficient way to work with big polynomials. However we will limit ourselves to an estimation of the runtime of the presented algorithm.

The calculation of $\triangle_f$ does not depend on y, and since the coefficients of f (of degree d) are bounded, we can assume that this task is done in $O(d)$. Then we ask to find all primes p less than y which take $O(y)$ operations, using a prime number sieve. To test if p divide $\triangle_f$ requires $O(1)$. For each such prime, we do a factorization of f modulo p which takes $O(log(p)d^2)$ operations. Hence the number of operations is here equal to $O(d) + O(y) + p * O(1) + O(log(p)d^2) = O(d + y + p + log(p)d^2) = O(y)$ for y large enough.
Finally to manipulate these data between lists and sequences in order to create all factorization types and gather them requires the utilisation of two for loops of size $O(y)$, that means in average $O(y^2)$.
Hence we have a runtime corresponding to the computation of $O(y^2)$ operations.

Furthermore as shown in the following, $y = 10^6$ begins to slow the machine and takes around $5''$, however it allows reliability on polynomial until degree 7, provided that it is a "simple polynomial", with less than 5 non-zero coefficients. To evaluate runtime with SAGE, we dispose of the following command: "$^0/_0 time$".

For instance: take $f(x) = x^3 + 3x^2 - 1$, put $y = 10^3$. We get $\mid G \mid = 3$, the correct factorization types and how many times they occur, with a runtime t equal to $5s$. Now take $y = 10^5$, the output is also exact, but $t = 1''41s$. Such a y is here not needed. On the other hand take $f(x) = x^7 - 3x^3 + 3$, put $y = 10^3$. Even if $t = 6s$ is "small", the output is not reliable, since the number of p for which we get a factorization was too small to find the identity type (however we know that if y is infinitely large, we are sure to find all factorization types). Now put $y = 10^6$, we get $\mid G \mid \simeq 5040$ which allows a correct matching in $t = 1''55s$.

In conclusion, it seems to be an efficient algorithm if $d \leqslant 7$, *i.e.* as long as $y = 10^6$ allows reliability. For higher y, on the standard laptop we work with, computing is very time consuming.

5 Appendix

The SAGE code of our algorithm is the following:

```
R.<x>=ZZ[x];                                                    #f belongs to Z[x]
f=x^5-7*x+1;                                                    #inputs
y=10000;
print "f =", f;
if f.is_irreducible()==true:                                    #test if f is irreducible
    L=[];                                                       #list L of length c with all obtained factorization types
    for i in range (y):
        if is_prime(i)==true:
            if disc(f)%i!=0:
                    k=f.change_ring(GF(i)).factor();            #reduction of f modulo i into irreducible factors
                    l=[];                                       #list l which become the factorization type
                    for j in range(len(k)):                     #filling in l with degrees: writting of the factorization type
                        if k[j][1]*k[j][0].degree()>=2:
                            l.append(k[j][1]*k[j][0].degree());
                    L.append(tuple(l));                         #filling in L with the factorization types
    c=len(L);
    print "There are",c,"primes less than",y, "which don't divide disc(f)";
    S=Set(L);                                                   #S is the set of all factorization classes
    v=0;
    for r in range(len(S)):
         if S[r]==():
            v=1;
            h=L.count(S[r]);                                    #count occuracy of identity
            o=numerical_approx(c/h,digits=0);
            print "The order of G is",o;
            print "Identity occurs",L.count(S[r]),"times, ie:",RealField(2)(L.count(S[r])*100/c),"%";
    for r in range(len(S)):
         if S[r]!=():
            if v==1:
                print "cycle
type",S[r],"occurs",L.count(S[r]),"times,ie:",numerical_approx(L.count(S[r])*100/c,digits=2),"%","n",S[r],"=",numerical_approx(
L.count(S[r])/h,digits=1);
            else:print "cycle type",S[r],"occurs",L.count(S[r]),"times,ie:",numerical_approx(L.count(S[r])*100/c,digits=2),"%"
else:  print "is not irreducible in ZZ[x]";
```

6 Acknowledgement

I would like to express my thanks to the following people who read this report and offered precious advice and comments.
I'm thinking particularly to Dr. Gabor Wiese and Dr. Panagiotis Tsaknias, who are professors at the University of Luxembourg, for guiding me in my work.

7 Bibliography

[1] Bartel Leendert Van Der Waerden: *Modern Algebra*

[2] David Hernandez: *Yves Laszlo: Introduction à la théorie de Galois*

[3] Marc Sage: *Théorie de Galois*

[4] Jan Vonk: *Calculations with modular Galois representations*

[5] Olivier Debarre: *Algebre 2 - Ecole normale supérieure*

[6] Paul Zimmermann: *Calcul mathématique avec Sage (SageBook)*