Contemporary Mathematics

Polynomials, Galois Groups, and Database-Driven Arithmetic

Elira Shaska and Tony Shaska

Abstract. This paper presents an ongoing, long-term project in Galois theory, a cornerstone of algebra, by leveraging computational methods to study polynomials and their Galois groups. We introduce large databases of irreducible polynomials as a powerful tool to analyze Galois groups and explore broader arithmetic applications, such as invariants and polynomial heights. By applying computational techniques to these databases, we aim to simplify the determination of solvability by radicals and uncover new patterns in Galois theory. This summary outlines the project's background, database-driven methodology, potential arithmetic applications, and challenges of integrating computational approaches with classical algebra.

1. Introduction

Galois theory, a cornerstone of modern algebra, provides profound insights into the solvability of polynomial equations. Since its inception by Évariste Galois, it has explained why there are no general formulas for polynomials of degree five or higher by radicals, unlike the well-known quadratic, cubic, and quartic formulas. This theory links the algebraic structure of field extensions to the symmetry of polynomial roots, encapsulated by their Galois groups. While traditional methods allow us to determine solvability for lower-degree polynomials through invariants like discriminants, the complexity escalates dramatically for higher degrees, where the Galois group might not be solvable, leading to no radical solution.

This project embarks on an innovative journey to merge the abstract realm of Galois theory with the practical capabilities of machine learning (ML). Our goal is to harness ML's pattern recognition and prediction abilities to address some of the most challenging aspects of Galois theory, potentially revolutionizing our understanding and approach to polynomial solvability and related problems. At the heart of Galois theory is the connection between a polynomial's roots and its Galois group, which describes how these roots can be permuted while preserving the field operations. A polynomial is solvable by radicals if its Galois group is solvable; this means there exists a chain of normal subgroups where each quotient is cyclic, allowing for the roots to be constructed by sequential additions, multiplications, and root extractions. However, for degrees five and above, generic polynomials often have non-solvable groups like S_n (the symmetric group), rendering them unsolvable by radicals.

We propose an approach where we compile or generate datasets of polynomials with known Galois groups. Key to our approach will be identifying or creating features from polynomials that are indicative of Galois group properties or solvability. These might include traditional invariants like discriminants or novel features derived from root distributions or algebraic properties. Using supervised learning, we aim to predict the Galois group or solvability of polynomials, potentially employing neural networks for their ability to handle complex patterns or decision trees for interpretability. Unsupervised methods could explore clustering of polynomials, perhaps revealing new mathematical insights. By learning from simpler polynomials, we hope to generalize these insights to more complex polynomials, possibly using techniques like transfer learning where models adapt knowledge from one task to another.

This integration could lead to automated solvability prediction, offering mathematicians tools to quickly assess if a polynomial can be solved by radicals, and might uncover patterns or invariants not yet recognized by traditional mathematics. The methodology could extend to other areas like field theory or algebraic geometry. However, several challenges loom, including the computational cost of handling high-degree polynomials, ensuring interpretability of ML models to enhance theoretical understanding, and balancing between providing practical tools and contributing to the theoretical body of Galois theory.

This project stands at the intersection of pure mathematics and cutting-edge computational science. By leveraging machine learning, we aim not only to solve practical problems within Galois theory but also to catalyze new theoretical advancements. This exploration could redefine how we approach some of the oldest and most fundamental questions in algebra, potentially opening new avenues for research in both mathematics and computer science.

A neuro-symbolic network is a type of artificial intelligence system that combines the strengths of neural networks (good at pattern recognition) with symbolic reasoning (based on logic and rules) to create models that can both learn from data and reason through complex situations, essentially mimicking human-like cognitive abilities by understanding and manipulating symbols to make decisions. This approach aims to overcome the limitations of either method alone, providing better explainability and adaptability in AI systems. In this paper, we experiment with such models to study some classical questions of Galois theory.

The paper proceeds as follows. In the second section, we cover basic terminology on polynomials, including their heights and weighted polynomials. Since the intended audience of this paper includes engineers and computer scientists, we provide some basic definitions and terminology that are normally found in every basic graduate algebra book.

Since this paper primarily deals with databases of polynomials with integer coefficients, in section three, we discuss the equivalence classes of polynomials, including $\mathbb{Z}$-equivalence, $\mathrm{GL}_2(\mathbb{Z})$-equivalence, Tschirnhaus equivalence, Hermite equivalence, and Julia equivalence. A detailed account of these topics can be found in [4].

Our data is ordered by height, whether that is the height of the polynomials or the weighted moduli height. Most open questions and arithmetic considerations are related to the heights of polynomials. Section four covers the basic definitions of

the theory of heights. In section five, we discuss binary forms in detail and provide the generators for the ring of invariants of binary forms for degrees up to ten.

The basic foundation of Galois groups of polynomials over $\mathbb{Q}$ is discussed in section six. We cover in detail the solution of cubics, quartics, and quintics not only to put things in proper context but also to emphasize that each degree is different. There is no universal method in Galois theory that works for every degree, which strongly suggests that AI models should be tailored specifically for each degree. This indicates that neuro-symbolic networks might be the best approach for designing models which not only predict the Galois group but also aim to derive solution formulas by radicals (when the group is solvable) and express these formulas in terms of invariants.

In section seven, we describe some general methods for determining the Galois group of a higher-degree polynomial, namely listing transitive subgroups of the symmetric group S_n, reducing polynomials modulo primes, and identifying special classes of polynomials based on the number of non-real roots.

Section eight is the core of the paper and delves into how to create databases of polynomials, providing a glimpse into how quickly computations can escalate. We detail how we build databases for cubics, quartics, and quintics and uncover some surprising trends even for such small degree polynomials where the theory is well-known. For instance, we find how rare it is for the cyclic group C_n to be the Galois group of a degree n polynomial. For example, among roughly 20^6 quintic polynomials of height ≤ 10, only three (up to $\overline{\mathbb{Q}}$-isomorphism) have a Galois group isomorphic to C_5, with a total of 20 polynomials (counting twists) corresponding to these three classes. Training an AI model to identify such rare cases might indeed be an impossible task, as noted in Section eight. Our data could serve various purposes, such as checking Malle's conjecture on Galois groups, verifying results by Bhargava et al. on the number of quartics with bounded heights, or comparing the height of polynomials with the weighted height of invariants.

In section nine, we offer a glimpse of what a neuro-symbolic network might look like for this application. This is not a fully developed product yet, as it could be refined with many symbolic layers based on theoretical knowledge. However, it shows that for small degrees, it can work relatively well. While there might not be a compelling reason to use AI models to predict the Galois group for degrees $d = 3, 4, 5$, this approach could prove very useful for higher degrees.

We hope this paper will encourage mathematicians and computer scientists to explore the use of AI in mathematical research, particularly in tackling classical problems of mathematics. Although this is a modest attempt to incorporate such methods into Galois theory, the rapid development of Artificial Intelligence promises new and innovative applications in mathematics.

2. Preliminaries

In this section we will go over some preliminary results on polynomials. Even though we will start with the general setup of polynomials defined over number fields and their rings of integers, later in the paper we will mostly focus on $\mathbb{Q}$ and its ring of integers $\mathbb{Z}$. For any field k, $\mathbb{A}^n_k$ and $\mathbb{P}^n_k$ denote the affine and projective spaces of dimension n over k, respectively.

2.1. Polynomials. Let R is a commutative ring with identity. An expression of the form

$$f(x) = \sum_{i=0}^{n} a_i x^i = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n, \tag{1}$$

where $a_i \in R$ and $a_n \neq 0$, is called a **polynomial over** R with **variable** x. The elements $a_0, a_1, \ldots, a_n$ are called **coefficients** of $f(x)$. The coefficient a_n is called the **leading coefficient**. A polynomial is called **monic** if its leading coefficient is 1.

If n is the largest non negative integer for which $a_n \neq 0$, then we say that the **degree** of $f(x)$ is n and write $\deg f(x) = n$. The set of all polynomials , with coefficient in a ring R is denoted by $R[x]$. It is also a commutative ring with identity. Two **polynomials are equal** if their corresponding coefficients are equal, so if we have

$$\begin{aligned} p(x) &= a_0 + a_1 x + \cdots + a_n x^n \\ q(x) &= b_0 + b_1 x + \cdots + b_m x^m, \end{aligned} \tag{2}$$

then $p(x) = q(x)$ if and only if $a_i = b_i$ for every $i = 0, \ldots, \max\{m, n\}$.

Let $p(x)$ and $q(x)$ be polynomials in $R[x]$, where R is a integral ring. Then,

$$\deg(p \cdot q) = \deg p + \deg q.$$

Moreover, $R[x]$ is a integral ring. If $\mathbb{F}$ is a field, then $\mathbb{F}[x]$ is a Euclidean domain with norm $N : \mathbb{F}[x] \to \mathbb{Z}^{\geq 0}$, such that $N(p(x)) = \deg(p(x))$.

LEMMA 1 (Division Algorithm). *Let $f(x)$ and $g(x)$ be two nonzero polynomials in $\mathbb{F}[x]$, where $\mathbb{F}$ is a field and $g(x)$ is a non-constant polynomial. Then, there exist unique polynomials $q(x), r(x) \in \mathbb{F}[x]$ such that*

$$f(x) = g(x)q(x) + r(x),$$

where $\deg r(x) < \deg g(x)$ and $r(x)$ is a nonzero polynomial.

Let $p(x)$ be a polynomial in $\mathbb{F}[x]$ and $\alpha \in F$. We say that α is a **zero** or **root** of $p(x)$, if $p(x)$ is in the kernel of the homomorphism ϕ_α or we say α is a zero of $p(x)$ if $p(\alpha) = 0$.

COROLLARY 1. *Let $\mathbb{F}$ be a field. An element $\alpha \in \mathbb{F}$ is a zero of $p(x) \in \mathbb{F}[x]$, if and only if $(x - \alpha)$ is a factor of $p(x)$ in $\mathbb{F}[x]$. A nonzero polynomial $p(x)$ with degree n in $\mathbb{F}[x]$ has at most n distinct zeroes in $\mathbb{F}$.*

A monic polynomial $d(x)$ is called **greatest common divisor** of polynomials $p(x), q(x) \in \mathbb{F}[x]$ if $d(x)$ divides $p(x)$ and $q(x)$; and if for every other polynomial $d'(x)$ that divides $p(x)$ and $q(x)$, $d'(x) \mid d(x)$. We write

$$d(x) = \gcd(p(x), q(x)).$$

Two polynomials $p(x)$ and $q(x)$ are **relatively prime** if $\gcd(p(x), q(x)) = 1$. Similarly as for the greatest common divisor of integers, we have the following:

LEMMA 2. *Let $\mathbb{F}$ be a field and assume that $d(x)$ is the greatest common divisor of two polynomials $p(x)$ and $q(x)$ in $\mathbb{F}[x]$. Then, there exist polynomials $r(x)$ and $s(x)$ such that*

$$d(x) = r(x) \cdot p(x) + s(x) \cdot q(x).$$

Moreover, the greatest common divisor of two polynomials is unique.

A polynomial $f(x) \in \mathbb{F}[x]$ is called **irreducible** if it has degree ≥ 1 and can not be written as

$$f(x) = g(x) \cdot h(x)$$

for some $g, h \in \mathbb{F}[x]$ and both $g, h \notin \mathbb{F}$. Elements of $\mathbb{F}$ are called **constant polynomials**.

Let A be a UFD and k its field of fractions. We take $a \in k$ such that $a = \frac{r}{s}$, where $(r, s) = 1$. For any prime element $p \in A$, we can write

$$a = p^m \, a'$$

where m is an integer and $a' \in k$ such that p does not divide numerator or denominator of a'. The **order of** a **in** p is defined as m, say $\mathrm{ord}_p(a) = m$. For $f(x) \in \mathbb{F}[x]$ given as in Eq. (1) we define

$$\mathrm{ord}_p\,(f) = min\,\{\, ord_p\,(a_i) \;\mid\; a_i \neq 0\}.$$

The **content** of $f(x)$, which is denoted $\mathrm{cont}(f)$, is defined as the product (up to multiplication to a unit in A)

$$\mathrm{cont}(f) := \prod p^{\mathrm{ord}_p\,(f)}, \tag{3}$$

taking all p such that $\mathrm{ord}_p\,(f) \neq 0$. If $\mathrm{cont}(f) = 1$, then $f(x)$ is called a **primitive polynomial**. Thus, every polynomial $f(x) \in \mathbb{F}[x]$ can be written as

$$f(x) = \mathrm{cont}(f) \cdot f_1(x),$$

where $f_1(x)$ is primitive and $f_1(x) \in A[x]$. Notice that if $f \in A[x]$ then $\mathrm{cont}(f)$ is simply

$$\mathrm{cont}(f) = \gcd(a_0, \ldots, a_n).$$

The **height** of $f(x)$ is defined as

$$\mathfrak{h}(f) := \max\{\mathrm{ord}_p(a_i) \,|\, a_i \neq 0\}$$

The following result is known as Gauss' lemma.

Lemma 3 (Gauss Lemma). *Let A be a UFD, k its field of fractions and $f, g \in \mathbb{F}[x]$. Then,*

$$cont(fg) = cont(f) \cdot cont(g)$$

Moreover, for $f, g \in A[x]$, fg is primitive if and only if f and g are both primitive.

2.2. Several variables. A polynomial with n variables is denoted by

$$f(x_1, \ldots, x_n) = \sum_{i=(i_1,\ldots,i_n)\in I} a_i x_1^{i_1} \cdots x_n^{i_n}$$

where all $a_i \in K$, $I \subset \mathbb{Z}^{\geq 0}$, and I is finite. We use lexicographic ordering to order the terms in a given polynomial, and let

$$x_1 > x_2 > \cdots > x_n.$$

While the primary goal of this paper are polynomials with one variable, we will use polynomials with several variables when we discuss invariants of binary forms.

2.3. Weighted polynomials. Given any integer $n \geq 1$, let $\mathbf{w} = (q_0, \ldots, q_n)$ be a vector of positive integers. Consider the polynomial ring $R = k_{\mathbf{w}}[x_0, \ldots, x_n]$ where x_i has weight q_i for $i = 0, 1, \cdots, n$.

Every polynomial is a sum of monomials $x^d = \prod x_i^{d_i}$ with weight $\sum_{i=1}^n q_i d_j$. For every $\lambda \in k^*$ and any weighted homogeneous polynomial f of degree d, we have

$$f(\lambda^{q_0} x_0, \lambda^{q_1} x_1, \ldots, \lambda^{q_n} x_n) = \lambda^d f(x_0, \ldots, x_n).$$

A degree d binary weighted form, where $w = (q_0, q_1)$ be respectively the weights of x_0 and x_1, is given by

$$f(x_0, x_1) = \sum_{d_0, d_1} a_{d_0, d_1} x_0^{d_0} x_1^{d_1}, \quad \text{such that} \quad d_0 q_0 + d_1 q_1 = d$$

and in decreasing powers of x_0 we have

$$f(x_0, x_1) = a_{d/q_0, 0} x_0^{d/q_0} + \cdots + a_{d_0, d_1} x_0^{d_0} x_1^{d_1} + \cdots + a_{0, d/q_1} x_1^{d/q_1}$$

By dividing with x_1^{d/q_1} and making a change of coordinates $X = x_0^{q_1}/x_1^{q_0}$ we get

$$f(x_0, x_1) = a_{d/q_0, 0} X^{d/q_0 q_1} + \cdots + a_{d_0, d_1} X^{d_0/q_1} + \cdots + a_{0, d/q_1} = f(X) \tag{4}$$

Notice that the condition $f(P) = 0$ is well defined on $\mathbb{P}^n_{\mathbf{w},k}$.

2.4. Restriction to Polynomials over $\mathbb{Q}$ and $\mathbb{Z}$. The preliminaries in this section are developed over a general commutative ring R with identity, yet the core of this paper concerns irreducible polynomials $f \in \mathbb{Q}[x]$ with coefficients in $\mathbb{Z}$, i.e., $f \in \mathbb{Z}[x] \subset \mathbb{Q}[x]$. This restriction is grounded in the algebraic properties of the Galois group Gal $_{\mathbb{Q}}(f)$, the arithmetic structure of $\mathbb{Z}$, and the computational requirements for constructing databases of polynomials for machine learning classification. We formalize this choice through definitions, theorems, and proofs, emphasizing the invariance of Galois groups under scaling, the arithmetic advantages of integrality for reduction modulo primes, and the finiteness of polynomial sets with bounded coefficients.

Definition 1. *Let $f \in \mathbb{Q}[x]$ be a polynomial of degree n, given by*

$$f(x) = \sum_{i=0}^{n} a_i x^i, \quad a_i \in \mathbb{Q}, \quad a_n \neq 0.$$

We say f is integral *if $f \in \mathbb{Z}[x]$, i.e., $a_i \in \mathbb{Z}$ for all i. For any $f \in \mathbb{Q}[x]$, there exists $\lambda \in \mathbb{Q}^\times$ such that $\lambda f \in \mathbb{Z}[x]$, obtained by clearing denominators of the coefficients a_i.*

The primary algebraic justification for focusing on $\mathbb{Z}[x]$ is the invariance of the Galois group under scaling, which we prove below.

Theorem 2.1. *Let $f \in \mathbb{Q}[x]$ be an irreducible polynomial of degree n, and let $\lambda \in \mathbb{Q}^\times$. Then the Galois group of λf over $\mathbb{Q}$ is isomorphic to that of f, i.e.,*

$$Gal\ _{\mathbb{Q}}(\lambda f) \cong Gal\ _{\mathbb{Q}}(f).$$

Proof. Let $f(x) = \sum_{i=0}^n a_i x^i \in \mathbb{Q}[x]$ be irreducible, and let $E_f = \mathbb{Q}(\alpha_1, \ldots, \alpha_n)$ be its splitting field over $\mathbb{Q}$, where $\alpha_1, \ldots, \alpha_n$ are the roots of f. Thus, $f(x) =$

$a_n \prod_{i=1}^{n}(x-\alpha_i)$, and Gal $_{\mathbb{Q}}(f) = \text{Gal } (E_f/\mathbb{Q})$. Consider $g(x) = \lambda f(x) = \sum_{i=0}^{n} \lambda a_i x^i$. Since $\lambda \neq 0$, the roots of $g(x)$ satisfy

$$g(x) = \lambda f(x) = \lambda a_n \prod_{i=1}^{n}(x - \alpha_i) = 0 \iff f(x) = 0.$$

Hence, the roots of $g(x)$ are identical to those of $f(x)$, and the splitting field of $g(x)$ is $E_f = \mathbb{Q}(\alpha_1, \ldots, \alpha_n)$. Consequently,

$$\text{Gal }_{\mathbb{Q}}(g) = \text{Gal } (E_f/\mathbb{Q}) = \text{Gal }_{\mathbb{Q}}(f).$$

Thus, Gal $_{\mathbb{Q}}(\lambda f) \cong$ Gal $_{\mathbb{Q}}(f)$, as the automorphism group of the splitting field is unchanged. □

COROLLARY 2. *For any irreducible polynomial $f \in \mathbb{Q}[x]$, there exists $\lambda \in \mathbb{Q}^\times$ such that $\lambda f \in \mathbb{Z}[x]$, and Gal $_{\mathbb{Q}}(\lambda f) \cong$ Gal $_{\mathbb{Q}}(f)$. Thus, without loss of generality, we may assume $f \in \mathbb{Z}[x]$ when studying Gal $_{\mathbb{Q}}(f)$.*

PROOF. Let $f(x) = \sum_{i=0}^{n} a_i x^i$, where $a_i = p_i/q_i \in \mathbb{Q}$, with $p_i, q_i \in \mathbb{Z}$, $q_i \neq 0$. Let $d = \text{lcm}(q_0, \ldots, q_n)$, the least common multiple of the denominators. Then, $\lambda = d$ satisfies $\lambda a_i = d \cdot (p_i/q_i) \in \mathbb{Z}$, so $\lambda f \in \mathbb{Z}[x]$. By Theorem 2.1, Gal $_{\mathbb{Q}}(\lambda f) \cong$ Gal $_{\mathbb{Q}}(f)$. □

This result allows us to restrict to $\mathbb{Z}[x]$ without altering the Galois group. We now explore the arithmetic advantages of $\mathbb{Z}[x]$, particularly for reduction modulo primes, which is central to determining Galois groups (see Section 7.2).

PROPOSITION 1. *Let $f \in \mathbb{Z}[x]$ be a monic irreducible polynomial of degree n, and let p be a prime such that $p \nmid \Delta_f$. The reduction $f_p(x) = f(x) \mod p \in \mathbb{F}_p[x]$ factors into irreducible factors of degrees $n_1, \ldots, n_k$, and Gal $_{\mathbb{Q}}(f)$ contains a permutation of cycle type $(n_1) \cdots (n_k)$.*

PROOF. This is a specialization of Dedekind's theorem. Since $f \in \mathbb{Z}[x]$ is monic, its coefficients are integers, and reduction modulo p yields a well-defined polynomial $f_p(x) \in \mathbb{F}_p[x]$. The condition $p \nmid \Delta_f$ ensures that $f_p(x)$ has distinct roots in an algebraic closure of $\mathbb{F}_p$, as $\Delta_f \in \mathbb{Z}$ is non-zero modulo p. Let $E_f = \mathbb{Q}(\alpha_1, \ldots, \alpha_n)$ be the splitting field of f, and let $G = \text{Gal }_{\mathbb{Q}}(f)$. For a prime p not dividing Δ_f, the Frobenius element in G associated to p has cycle type corresponding to the degrees $n_1, \ldots, n_k$ of the irreducible factors of $f_p(x)$, as established in [31, Section 8.10]. □

The integrality of coefficients in $\mathbb{Z}[x]$ ensures that the discriminant $\Delta_f \in \mathbb{Z}$, which is critical for applying Proposition 1. Moreover, since $\mathbb{Z}$ is a unique factorization domain (UFD), we can use Gauss's lemma to analyze the content of polynomials.

LEMMA 4. *Let $f \in \mathbb{Z}[x]$, given by $f(x) = \sum_{i=0}^{n} a_i x^i$. The content of f, defined as*

$$\textit{cont}(f) = \gcd(a_0, \ldots, a_n),$$

is an integer, and there exists a primitive polynomial $f_1 \in \mathbb{Z}[x]$ such that $f(x) = \textit{cont}(f) \cdot f_1(x)$. If f is irreducible over $\mathbb{Q}$, then $\textit{cont}(f) = \pm 1$.

PROOF. Since $\mathbb{Z}$ is a UFD, the content $\text{cont}(f) = \gcd(a_0, \ldots, a_n) \in \mathbb{Z}$ is well-defined. Write $f(x) = \text{cont}(f) \cdot f_1(x)$, where $f_1(x) = \sum_{i=0}^{n}(a_i/\text{cont}(f))x^i \in \mathbb{Z}[x]$, and $\text{cont}(f_1) = 1$, so f_1 is primitive. Suppose f is irreducible over $\mathbb{Q}$. If $|\text{cont}(f)| > 1$, say $\text{cont}(f) = d$, then $f(x) = d \cdot f_1(x)$, and $f_1 \in \mathbb{Z}[x] \subset \mathbb{Q}[x]$. Since f is irreducible, it has no non-trivial factors in $\mathbb{Q}[x]$. However, $d \in \mathbb{Q}$ and $f_1 \in \mathbb{Q}[x]$, with $\deg(f_1) = n \geq 1$, contradicting irreducibility unless $d = \pm 1$. Thus, $\text{cont}(f) = \pm 1$. □

This lemma ensures that irreducible polynomials in $\mathbb{Z}[x]$ are primitive, simplifying their representation in databases. We now address the computational aspect of constructing finite sets of polynomials, which is crucial for machine learning applications (Section 8).

DEFINITION 2. *For a polynomial $f(x) = \sum_{i=0}^{n} a_i x^i \in \mathbb{Z}[x]$, define the* naive height *as*

$$h(f) = \max_i\{|a_i|\}.$$

The set of polynomials with bounded naive height is

$$\mathcal{P}_n^h = \{f \in \mathbb{Z}[x] \mid \deg(f) = n, a_n a_0 \neq 0, \Delta_f \neq 0, h(f) \leq h\},$$

where f is irreducible over $\mathbb{Q}$.

THEOREM 2.2. *For any degree $n \geq 1$ and bound $h \geq 1$, the set $\mathcal{P}_n^h$ is finite, with cardinality bounded by*

$$\#\mathcal{P}_n^h \leq (2h+1)^{n+1}.$$

PROOF. For $f(x) = \sum_{i=0}^{n} a_i x^i \in \mathbb{Z}[x]$, the condition $h(f) = \max_i\{|a_i|\} \leq h$ implies $a_i \in \{-h, -h+1, \ldots, h\}$. The number of choices for each coefficient a_i (for $i = 0, \ldots, n$) is $2h+1$, so the total number of polynomials with $\deg(f) = n$ and $h(f) \leq h$ is at most $(2h+1)^{n+1}$. The conditions $a_n a_0 \neq 0$, $\Delta_f \neq 0$, and irreducibility over $\mathbb{Q}$ restrict this set further. Since the set of integer tuples $(a_0, \ldots, a_n)$ is finite, and each polynomial can be tested for irreducibility and non-zero discriminant, $\mathcal{P}_n^h$ is finite. □

COROLLARY 3. *The set $\mathcal{P}_n^h$ of irreducible polynomials $f \in \mathbb{Z}[x]$ of degree n with naive height $h(f) \leq h$ is finite and computable, enabling the construction of databases for machine learning classification of Galois groups.*

The restriction to $\mathbb{Z}[x]$ also facilitates the computation of invariants (Section 5), as coefficients in $\mathbb{Z}$ ensure that invariants like the discriminant Δ_f and binary form invariants are integers. This avoids numerical instability in the neuro-symbolic network (Section 9), where integer inputs provide a discrete feature space for extracting algebraic properties such as signatures and root counts.

In conclusion, the focus on $\mathbb{Q}$ and $\mathbb{Z}$ is justified by the invariance of Galois groups under scaling (Theorem 2.1), the arithmetic structure of $\mathbb{Z}$ for reduction modulo primes (Proposition 1), the simplicity of content computations in a UFD (Lemma 4), and the finiteness of polynomial sets with bounded coefficients (Theorem 2.2). These properties provide a rigorous foundation for studying Galois groups and their computational classification.

3. Equivalences of Polynomials

This section explores various equivalence relations on polynomials, which are essential for classifying polynomials over $\mathbb{Q}$ and $\mathbb{Z}$ in computational applications, particularly in Galois theory and polynomial databases. We define and analyze these equivalences with precision, providing a foundation for identifying equivalence classes and ordering polynomials effectively.

3.1. Projective Equivalence and Binary Forms. We begin by defining equivalence under scalar multiplication, which allows us to view polynomials projectively.

Definition 3 (Projective Equivalence). *Let $f, g \in K[x]$ be polynomials of degree $d \geq 1$ over a field K. We say f and g are* projectively equivalent *if there exists a nonzero scalar $\lambda \in K^\times$ such that:*

$$f(x) = \lambda g(x).$$

For a polynomial $f(x) = a_d x^d + a_{d-1}x^{d-1} + \cdots + a_0$, projective equivalence identifies f with the point $[a_0 : a_1 : \cdots : a_d] \in \mathbb{P}^d_K$, the projective space over K.

This equivalence normalizes polynomials up to scaling, a natural starting point for classification. To incorporate transformations of the variable, we introduce binary forms.

Definition 4 (Binary Forms). *Let K be a field, and $K[x, y]$ the polynomial ring in two variables. A* binary form of degree d *is a homogeneous polynomial $f(x, y) \in K[x, y]$ of degree d:*

$$f(x, y) = a_d x^d + a_{d-1}x^{d-1}y + \cdots + a_0 y^d,$$

where the coefficients $a_i \in K$. The space of such forms, denoted $V_d = K[x, y]_d$, is a vector space of dimension $d + 1$.

For a polynomial $f(x) \in K[x]$ of degree d, its *homogenization* is defined as

$$\tilde{f}(x, y) = y^d f\left(\frac{x}{y}\right).$$

For a binary form $f(x, y) \in V_d$, its *dehomogenization* is defined as $f(x, 1) \in K[x]$.

Remark 1. *Since $\tilde{f}(x, y) = \lambda\tilde{g}(x, y)$ if $f(x) = \lambda g(x)$, projective equivalence of polynomials corresponds to scalar multiplication of their homogenizations. Thus, $\mathbb{P}(V_d) \cong \mathbb{P}^d_K$ parametrizes binary forms up to scaling.*

Remark 2. *Any polynomial $f \in \mathbb{Q}[x]$ can be written as $f = \lambda g$ for some $g \in \mathbb{Z}[x]$, and since f and g share the same Galois group over $\mathbb{Q}$, we focus on polynomials in $\mathbb{Z}[x]$ without loss of generality.*

3.2. $\mathbb{Z}$-Equivalence and $\mathrm{GL}_2(\mathbb{Z})$-Equivalence. To classify polynomials under integer linear transformations, we define two related equivalences.

Definition 5 ($\mathbb{Z}$-Equivalence). *Let $f, g \in \mathbb{Z}[x]$ be polynomials of degree n. They are* $\mathbb{Z}$-equivalent *if there exist $a = \pm 1$ and $b \in \mathbb{Z}$ such that:*

$$f(x) = a^n g(ax + b).$$

This captures transformations $x \mapsto ax + b$ with $a = \pm 1$, preserving integer coefficients.

DEFINITION 6 ($\mathrm{GL}_2(\mathbb{Z})$-Equivalence). *For binary forms $f, g \in \mathbb{Z}[x,y]_n$ of degree n, they are* $\mathrm{GL}_2(\mathbb{Z})$-equivalent *if there exists $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{GL}_2(\mathbb{Z})$ (i.e., entries in $\mathbb{Z}$, $\det M = \pm 1$) such that:*

$$g(x,y) = \pm f^M(x,y), \quad \text{where} \quad f^M(x,y) = f(ax+by, cx+dy).$$

For polynomials $f, g \in \mathbb{Z}[x]$, they are $\mathrm{GL}_2(\mathbb{Z})$-equivalent *if their homogenizations $\tilde{f}, \tilde{g} \in \mathbb{Z}[x,y]_n$ satisfy:*

$$\tilde{g}(x,y) = \pm \tilde{f}^M(x,y) \quad \text{for some} \quad M \in \mathrm{GL}_2(\mathbb{Z}).$$

Equivalently, $g(x) = \pm(cx+d)^n f\left(\frac{ax+b}{cx+d}\right)$.

DEFINITION 7 ($\mathbb{Q}$-Equivalence). *Polynomials $f, g \in \mathbb{Q}[x]$ are* $\mathbb{Q}$-equivalent *if there exist $a, b, c, d \in \mathbb{Q}$ with $ad - bc \neq 0$ such that:*

$$f(x) = g\left(\frac{ax+b}{cx+d}\right).$$

LEMMA 5. *Let $f, g \in \mathbb{Z}[x]$ be polynomials of degree n. If f and g are $\mathbb{Z}$-equivalent, then they are $\mathrm{GL}_2(\mathbb{Z})$-equivalent, and their homogenizations are $\mathrm{GL}_2(\mathbb{Q})$-equivalent.*

PROOF. Assume $f(x) = a^n g(ax+b)$ with $a = \pm 1$, $b \in \mathbb{Z}$. Define $M = \begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix} \in \mathrm{GL}_2(\mathbb{Z})$, since $\det M = a = \pm 1$. For the homogenization $\tilde{g}(x,y) = y^n g\left(\frac{x}{y}\right)$, compute:

$$\tilde{g}^M(x,y) = \tilde{g}(ax+by, y) = y^n g\left(\frac{ax+by}{y}\right) = y^n g\left(a \cdot \frac{x}{y} + b\right).$$

Since $f(x) = a^n g(ax+b)$, we have:

$$\tilde{f}(x,y) = y^n f\left(\frac{x}{y}\right) = y^n a^n g\left(a \cdot \frac{x}{y} + b\right) = a^n \tilde{g}^M(x,y).$$

As $a = \pm 1$, $\tilde{g}(x,y) = \pm \tilde{f}^{M^{-1}}(x,y)$, where $M^{-1} = \begin{bmatrix} a & -ab \\ 0 & 1 \end{bmatrix} \in \mathrm{GL}_2(\mathbb{Z})$. Thus, f and g are $\mathrm{GL}_2(\mathbb{Z})$-equivalent. Since $\mathrm{GL}_2(\mathbb{Z}) \subset \mathrm{GL}_2(\mathbb{Q})$, their homogenizations are also $\mathrm{GL}_2(\mathbb{Q})$-equivalent. □

REMARK 3. *The following hold:*

1. *(1) $\mathrm{GL}_2(\mathbb{Q})$-equivalence partitions V_d into orbits under rational linear transformations.*
2. *(2) $\mathrm{GL}_2(\mathbb{Z})$-equivalence refines these into integer transformation orbits.*
3. *(3) $\mathbb{Z}$-equivalence further refines $\mathrm{GL}_2(\mathbb{Z})$-orbits by restricting $a = \pm 1$, $c = 0$.*

3.3. Tschirnhaus Equivalence. Tschirnhaus equivalence connects polynomial classification to Galois theory via their splitting fields.

DEFINITION 8 (Tschirnhaus Equivalence). *Let $f, g \in \mathbb{Q}[x]$ be monic, separable, irreducible polynomials of degree n, with splitting field E over $\mathbb{Q}$ and Galois group $G = Gal\,(E/\mathbb{Q})$. Let α be a root of f, β a root of g, and define $P = \mathrm{Stab}_G(\alpha)$, $Q = \mathrm{Stab}_G(\beta)$. Then f and g are* Tschirnhaus-equivalent *if P and Q are conjugate in G, i.e., there exists $\sigma \in G$ such that $Q = \sigma P \sigma^{-1}$.*

PROPOSITION 2. *If $f, g \in \mathbb{Q}[x]$ are Tschirnhaus-equivalent, then Gal $_{\mathbb{Q}}(f) \cong$ Gal $_{\mathbb{Q}}(g)$.*

PROOF. Since f and g share the splitting field E by definition of Tschirnhaus equivalence, their Galois groups over $\mathbb{Q}$ are identical: Gal $(E/\mathbb{Q})$. Thus, Gal $_{\mathbb{Q}}(f) =$ Gal $_{\mathbb{Q}}(g)$, and they are isomorphic as groups. □

REMARK 4. *Tschirnhaus equivalence is stricter than sharing a Galois group, requiring conjugate stabilizers. This reflects a symmetry in the root structures within E, crucial for applications in Galois theory.*

3.4. Hermite Equivalence. Hermite equivalence leverages multilinear forms to classify polynomials, offering a reduction theory for integer polynomials.

DEFINITION 9 (Hermite Form). *For $f \in \mathbb{Z}[x]$ of degree d with leading coefficient a_d and roots $\alpha_1, \ldots, \alpha_d \in \mathbb{C}$, the* Hermite form *is:*

$$[f](x_1, \ldots, x_d) = a_d^{d-1} \prod_{i=1}^{d} \left(\alpha_i^{d-1} x_1 + \alpha_i^{d-2} x_2 + \cdots + \alpha_i x_{d-1} + x_d\right).$$

PROPOSITION 3. *The Hermite form is the resultant:*

$$[f](x_1, \ldots, x_d) = \mathrm{Res}_x\left(f(x), x_1 x^{d-1} + x_2 x^{d-2} + \cdots + x_d\right),$$

and thus $[f] \in \mathbb{Z}[x_1, \ldots, x_d]$.

PROOF. Define $g(x) = x_1 x^{d-1} + x_2 x^{d-2} + \cdots + x_d$. The resultant $\mathrm{Res}_x(f, g)$ is the product of g evaluated at the roots of f:

$$\mathrm{Res}_x(f, g) = a_d^d \prod_{i=1}^{d} g(\alpha_i) = a_d^d \prod_{i=1}^{d} \left(\alpha_i^{d-1} x_1 + \alpha_i^{d-2} x_2 + \cdots + \alpha_i x_{d-1} + x_d\right).$$

Now, compute the Hermite form:

$$[f](x_1, \ldots, x_d) = a_d^{d-1} \cdot a_d \prod_{i=1}^{d} g(\alpha_i) = a_d^{d-1} \cdot a_d^d \prod_{i=1}^{d} g(\alpha_i)/a_d = \mathrm{Res}_x(f, g).$$

Since f and g have integer coefficients, and the resultant is a symmetric polynomial in the roots expressed via the coefficients of f and g, $[f] \in \mathbb{Z}[x_1, \ldots, x_d]$. □

DEFINITION 10 (Hermite Equivalence). *Polynomials $f, g \in \mathbb{Z}[x]$ of degree n are* Hermite equivalent *if their Hermite forms $[f]$ and $[g]$ are $\mathrm{GL}_n(\mathbb{Z})$-equivalent, i.e., there exists $M \in \mathrm{GL}_n(\mathbb{Z})$ such that:*

$$[g](x_1, \ldots, x_n) = [f](M \cdot (x_1, \ldots, x_n)^T).$$

PROPOSITION 4. *The discriminant of $[f]$, defined as $\mathcal{D}([f]) = (\det(\alpha_{i,j}))^2$ where $\alpha_{i,j} = \alpha_i^{d-j}$, equals the discriminant of f, Δ_f.*

PROOF. Consider the matrix $A = (\alpha_{i,j})$ where $\alpha_{i,j} = \alpha_i^{d-j}$, for $1 \le i \le d$, $1 \le j \le d$. This is the Vandermonde matrix $V(\alpha_1, \ldots, \alpha_d)$ with entries $a_{i,j} = \alpha_i^{d-j}$. Its determinant is:

$$\det A = \prod_{1 \le i < j \le d} (\alpha_j - \alpha_i),$$

computed by factoring the Vandermonde determinant $\det[\alpha_i^{j-1}]$ and adjusting indices. Thus:

$$\mathcal{D}([f]) = (\det A)^2 = \left(\prod_{1 \le i < j \le d} (\alpha_j - \alpha_i) \right)^2 .$$

The discriminant of f, $\Delta_f = a_d^{2d-2} \prod_{i<j} (\alpha_i - \alpha_j)^2$, but since $[f]$ normalizes by a_d^{d-1}, we adjust for the leading coefficient's effect. However, directly, $\Delta_f = \prod_{i<j} (\alpha_i - \alpha_j)^2$ (up to a constant factor), matching $\mathcal{D}([f])$. □

COROLLARY 4. *If f and g are Hermite equivalent, then $\Delta_f = \Delta_g$.*

PROOF. Since $[f]$ and $[g]$ are $\mathrm{GL}_n(\mathbb{Z})$-equivalent, and $\mathrm{GL}_n(\mathbb{Z})$-transformations (with determinant ± 1) preserve the discriminant of multilinear forms up to a square factor of 1, we have $\mathcal{D}([f]) = \mathcal{D}([g])$. By Proposition 4, $\Delta_f = \mathcal{D}([f]) = \mathcal{D}([g]) = \Delta_g$. □

THEOREM 3.1 (Finiteness). *For a given degree d and nonzero discriminant Δ, there are finitely many Hermite equivalence classes of polynomials $f \in \mathbb{Z}[x]$ with $\Delta_f = \Delta$.*

PROOF. Hermite's theorem asserts that for a fixed degree d and discriminant $\Delta \neq 0$, the number of $\mathrm{GL}_n(\mathbb{Z})$-equivalence classes of multilinear forms with discriminant Δ is finite. Since $[f]$ is a multilinear form determined by f, and Hermite equivalence is defined via $\mathrm{GL}_n(\mathbb{Z})$-equivalence of $[f]$ and $[g]$, the number of such classes is finite. This follows from the reduction theory of integer forms, ensuring a finite set of reduced representatives. □

3.5. Julia Equivalence. Julia equivalence associates a unique quadratic form to each binary form, enhancing classification.

DEFINITION 11 (Julia Quadratic). *Let $f \in \mathbb{Z}[x, y]_n$ be a binary form of degree n with $f(x, 1) = a_0 \prod_{i=1}^{r} (x - \alpha_i) \prod_{j=1}^{s} (x - \beta_j)(x - \bar{\beta}_j)$, where $a_0 \neq 0$, α_i are real roots, β_j are complex roots, and $r + 2s = n$ (the* signature (r, s)*). Define:*

$$T_r(x, 1) = \sum_{i=1}^{r} t_i^2 (x - \alpha_i)^2, \quad S_s(x, 1) = \sum_{j=1}^{s} 2u_j^2 (x - \beta_j)(x - \bar{\beta}_j),$$

and the Julia quadratic*:*

$$Q_f(x, 1) = T_r(x, 1) + S_s(x, 1),$$

where $t_i, u_j \in \mathbb{R}$ are chosen to minimize an invariant.

PROPOSITION 5. *The discriminant of Q_f, $\mathcal{D}_f$, is:*

$$\mathcal{D}_f = \mathcal{D}(T_r) + \mathcal{D}(S_s) - 8 \sum_{i=1}^{r} \sum_{j=1}^{s} t_i^2 u_j^2 \left((\alpha_i - a_j)^2 + b_j^2 \right),$$

where $\beta_j = a_j + b_j i$.

PROOF. Consider the Julia quadratic $Q_f(x, 1) = T_r(x, 1) + S_s(x, 1)$, where:

$$T_r(x, 1) = \sum_{i=1}^{r} t_i^2 (x - \alpha_i)^2, \quad S_s(x, 1) = \sum_{j=1}^{s} 2u_j^2 (x - \beta_j)(x - \bar{\beta}_j),$$

and $t_i, u_j \in \mathbb{R}$, $\alpha_i \in \mathbb{R}$, $\beta_j = a_j + b_j i \in \mathbb{C}$. The discriminant of Q_f, a quadratic polynomial in x, is computed by expressing $Q_f(x,1)$ in the form ax^2+bx+c and evaluating $\mathcal{D}_f = b^2 - 4ac$.

First, express $T_r(x,1)$:

$$T_r(x,1) = \sum_{i=1}^{r} t_i^2(x^2 - 2\alpha_i x + \alpha_i^2) = \left(\sum_{i=1}^{r} t_i^2\right) x^2 - 2\left(\sum_{i=1}^{r} t_i^2\alpha_i\right) x + \sum_{i=1}^{r} t_i^2\alpha_i^2.$$

For $S_s(x,1)$, note that:

$$(x-\beta_j)(x-\bar{\beta}_j) = x^2 - (\beta_j+\bar{\beta}_j)x + \beta_j\bar{\beta}_j = x^2 - 2a_j x + (a_j^2+b_j^2).$$

Thus:

$$S_s(x,1) = \sum_{j=1}^{s} 2u_j^2(x^2 - 2a_j x + a_j^2 + b_j^2) = 2\left(\sum_{j=1}^{s} u_j^2\right) x^2 - 4\left(\sum_{j=1}^{s} u_j^2 a_j\right) x + 2\sum_{j=1}^{s} u_j^2(a_j^2+b_j^2).$$

Combining, $Q_f(x,1) = ax^2+bx+c$, where:

$$a = \sum_{i=1}^{r} t_i^2 + 2\sum_{j=1}^{s} u_j^2, \quad b = -2\left(\sum_{i=1}^{r} t_i^2\alpha_i + 2\sum_{j=1}^{s} u_j^2 a_j\right), \quad c = \sum_{i=1}^{r} t_i^2\alpha_i^2 + 2\sum_{j=1}^{s} u_j^2(a_j^2+b_j^2).$$

The discriminant is:

$$\mathcal{D}_f = b^2 - 4ac.$$

Compute:

$$b^2 = 4\left(\sum_{i=1}^{r} t_i^2\alpha_i + 2\sum_{j=1}^{s} u_j^2 a_j\right)^2,$$

$$4ac = 4\left(\sum_{i=1}^{r} t_i^2 + 2\sum_{j=1}^{s} u_j^2\right)\left(\sum_{i=1}^{r} t_i^2\alpha_i^2 + 2\sum_{j=1}^{s} u_j^2(a_j^2+b_j^2)\right).$$

To derive $\mathcal{D}_f$, consider the discriminants of T_r and S_s. For a quadratic px^2+qx+r, the discriminant is q^2-4pr. Thus:

- For $T_r(x,1) = \left(\sum t_i^2\right)x^2 - 2\left(\sum t_i^2\alpha_i\right)x + \sum t_i^2\alpha_i^2$:

$$\mathcal{D}(T_r) = \left[-2\sum_{i=1}^{r} t_i^2\alpha_i\right]^2 - 4\left(\sum_{i=1}^{r} t_i^2\right)\left(\sum_{i=1}^{r} t_i^2\alpha_i^2\right) = 4\left[\left(\sum_{i=1}^{r} t_i^2\alpha_i\right)^2 - \left(\sum_{i=1}^{r} t_i^2\right)\left(\sum_{i=1}^{r} t_i^2\alpha_i^2\right)\right].$$

- For $S_s(x,1)$, a similar computation yields $\mathcal{D}(S_s)$.

Expanding b^2-4ac, the expression includes terms from $\mathcal{D}(T_r)$, $\mathcal{D}(S_s)$, and cross terms from the interaction of real and complex roots. After simplification, the discriminant is:

$$\mathcal{D}_f = \mathcal{D}(T_r) + \mathcal{D}(S_s) - 8\sum_{i=1}^{r}\sum_{j=1}^{s} t_i^2 u_j^2\left((\alpha_i - a_j)^2 + b_j^2\right),$$

where the cross term reflects the geometric distance between real and complex roots, scaled by t_i^2 and u_j^2, with the factor 8 arising from the quadratic structure and conjugate pairing. □

3.6. Addressing the Two Main Issues. The equivalence relations defined in this section—projective, $\mathbb{Z}$-, $\mathrm{GL}_2(\mathbb{Z})$-, $\mathbb{Q}$-, Tschirnhaus, Hermite, and Julia equivalences—are designed to address two critical challenges in constructing databases of irreducible polynomials for machine learning applications in Galois theory:

3.6.1. *Identifying $\mathbb{Q}$-Equivalence Classes:* The classification of polynomials up to $\mathbb{Q}$-equivalence, where $f, g \in \mathbb{Q}[x]$ satisfy $f(x) = g\left(\frac{ax+b}{cx+d}\right)$ for some $a, b, c, d \in \mathbb{Q}$ with $ad-bc \neq 0$, is essential to reduce redundancy in polynomial databases (Section 8). This equivalence corresponds to $\mathrm{GL}_2(\mathbb{Q})$-equivalence of their homogenizations in $V_d = \mathbb{Q}[x, y]_d$. Classical invariant theory provides a solution by associating to each binary form $f \in V_d$ a set of $\mathrm{SL}_2(\mathbb{Q})$-invariants $\xi_0, \ldots, \xi_n$, which define a point in a weighted projective space. These invariants uniquely characterize the $\mathbb{Q}$-equivalence class of f, up to scaling.

For example, consider two quartic polynomials $f(x) = x^4 + x^2 + 1$ and $g(x) = x^4 + 2x^2 + 2$. Their homogenizations are $\tilde{f}(x, y) = x^4 + x^2y^2 + y^4$ and $\tilde{g}(x, y) = x^4 + 2x^2y^2 + 2y^4$. The invariants for quartics determine whether $\tilde{f}$ and $\tilde{g}$ lie in the same $\mathrm{GL}_2(\mathbb{Q})$-orbit, resolving whether f and g are $\mathbb{Q}$-equivalent. This approach ensures that our database contains representatives of distinct $\mathbb{Q}$-equivalence classes, minimizing computational overhead in training neuro-symbolic networks.

3.6.2. *Ordering Polynomials:* To construct finite, ordered databases, a systematic method to list polynomials is required. The equivalence relations, particularly $\mathbb{Z}$- and $\mathrm{GL}_2(\mathbb{Z})$-equivalence, allow us to select minimal representatives within each $\mathbb{Q}$-equivalence class, for instance, via Julia reduction (Section 3.5). This reduction ensures that polynomials are represented in a canonical form, facilitating their organization into ordered lists suitable for machine learning input.

For instance, Julia equivalence associates to each polynomial a unique quadratic form $\mathcal{J}_f$, whose coefficients can guide the selection of a minimal representative. This process, combined with bounds on polynomial coefficients, enables the creation of ordered databases, as explored in later sections, crucial for predicting Galois groups via supervised learning.

The equivalence relations of this section thus provide the algebraic framework for classifying polynomials, setting the stage for computational tools to address these challenges efficiently in the context of Galois theory and machine learning.

4. Heights of Polynomials

The classification of polynomials up to equivalence, as detailed in Section 3, requires a mechanism to order and enumerate them for computational applications, such as constructing databases of irreducible polynomials (Section 8). Heights provide a robust measure of polynomial "size" by quantifying the magnitude of their coefficients across all places of a number field. This section defines affine and projective heights, establishes their key properties, and explores their interaction with polynomial equivalences, laying the foundation for finite, ordered databases used in our neuro-symbolic network (Section 9).

4.1. Definitions and Basic Properties. Let K be a number field, $\mathcal{O}_K$ its ring of integers, and M_K the set of places (archimedean and non-archimedean) of K. For each place $v \in M_K$, let $|\cdot|_v$ denote the absolute value normalized so that $|x|_v = |x|^{n_v/[K:\mathbb{Q}]}$, where $n_v = [K_v : \mathbb{Q}_p]$ (or $\mathbb{R}$ for archimedean places) is the local degree.

DEFINITION 12 (Affine and Projective Heights). *For a polynomial* $f(x) = \sum_{j=0}^{n} a_j x^j \in K[x]$*, define:*

- *The* Gauss norm *at place* v*:*
$$|f|_v = \max_j\{|a_j|_v\}.$$
- *The* affine multiplicative height*:*
$$H_K^{\mathbb{A}}(f) = \prod_{v\in M_K} \max\{1, |f|_v^{n_v}\}.$$
- *The* affine logarithmic height*:*
$$h_K^{\mathbb{A}}(f) = h_K([1, a_0, \ldots, a_n]) = \sum_{v\in M_K} n_v \log \max\{1, |f|_v\}.$$
- *The* projective multiplicative height*:*
$$H_K(f) = \prod_{v\in M_K} |f|_v^{n_v}. \tag{5}$$
- *The* absolute projective multiplicative height*:*
$$H(f) = H_K(f)^{1/[K:\mathbb{Q}]} : \mathbb{P}^n(\mathbb{Q}) \to [1, \infty).$$

The affine height, often called the *naive height*, measures the size of coefficients as affine coordinates, while the projective height normalizes for scalar multiples, aligning with projective equivalence (Section 3). For polynomials in $\mathbb{Z}[x]$, the projective height simplifies significantly.

EXAMPLE 1. *Let* $f(x) = \sum_{j=0}^{n} a_j x^j \in \mathbb{Z}[x]$ *be primitive (i.e.,* $\gcd(a_0, \ldots, a_n) = 1$*). For* $K = \mathbb{Q}$*, the places* $M_{\mathbb{Q}}$ *consist of the archimedean place* $|\cdot|_\infty$ *and non-archimedean places* $|\cdot|_p$ *for primes* p*. Since* $a_j \in \mathbb{Z}$*,* $|a_j|_p \le 1$*, so* $|f|_p = \max_j\{|a_j|_p\} \le 1$*, and* $|f|_\infty = \max_j\{|a_j|\}$*. Thus:*
$$H_{\mathbb{Q}}(f) = \prod_{v\in M_{\mathbb{Q}}} |f|_v = |f|_\infty = \max_j\{|a_j|\}.$$

A fundamental property of heights is the finiteness of polynomials with bounded height, a consequence of Northcott's theorem.

LEMMA 6. *There are only finitely many polynomials* $f(x_1, \ldots, x_n) \in K[x_1, \ldots, x_n]$ *with bounded height. Specifically, for any* $f \in K[x_1, \ldots, x_n]$*, the set* $\{g \in K[x_1, \ldots, x_n] \mid H_K(g) \le H_K(f)\}$ *is finite.*

PROOF. By Northcott's theorem, the set of points in $\mathbb{P}^m(K)$ with bounded height is finite. For a polynomial $f(x_1, \ldots, x_n) = \sum_I a_I x^I$, its coefficients define a point $[a_I] \in \mathbb{P}^m(K)$, where m is the number of monomials. The height $H_K(f) = H_K([a_I])$ bounds the coordinates, ensuring finiteness. □

4.2. Height Properties and Polynomial Operations. Heights exhibit multiplicative properties under polynomial operations, crucial for computational applications.

LEMMA 7 (Gauss's Lemma). *Let* K *be a number field, and* $f, g \in K[x_1, \ldots, x_n]$*. For a non-archimedean place* $v \in M_K$*, the Gauss norm satisfies:*
$$|fg|_v = |f|_v |g|_v.$$

PROOF. For a non-archimedean place v, let $f = \sum_I a_I x^I$, $g = \sum_J b_J x^J$, and $fg = \sum_K c_K x^K$, where $c_K = \sum_{I+J=K} a_I b_J$. The Gauss norm is $|f|_v = \max_I\{|a_I|_v\}$. Since v is non-archimedean, $|a_I + b_J|_v \leq \max\{|a_I|_v, |b_J|_v\}$, and for the product:

$$|c_K|_v = \left| \sum_{I+J=K} a_I b_J \right|_v \leq \max_{I+J=K}\{|a_I b_J|_v\} = \max_{I+J=K}\{|a_I|_v |b_J|_v\}.$$

If $|f|_v = |a_{I_0}|_v$, $|g|_v = |b_{J_0}|_v$, consider $K_0 = I_0 + J_0$. Then $|c_{K_0}|_v \geq |a_{I_0} b_{J_0}|_v = |f|_v |g|_v$, so $|fg|_v \geq |f|_v |g|_v$. Conversely, $|c_K|_v \leq \max\{|a_I|_v |b_J|_v\} \leq |f|_v |g|_v$, so $|fg|_v = |f|_v |g|_v$. □

For archimedean places, the behavior is more complex, addressed by the following lemma.

LEMMA 8. *Let $f_1, \ldots, f_r \in \mathbb{C}[x_1, \ldots, x_n]$, $f = f_1 \cdots f_r$, and $d_i = \deg(f, x_i)$. For an archimedean place v:*

$$\prod_{i=1}^{r} |f_i|_v \leq e^{(d_1 + \cdots + d_n)} |f|_v.$$

The proof, involving the Mahler measure, is given in [13, pg. 232]. The *Mahler measure* of a polynomial $f \in \mathbb{C}[x_1, \ldots, x_n]$ is:

$$M(f) = \exp\left(\int_{\mathbb{T}^n} \log \left| f(e^{i\theta_1}, \ldots, e^{i\theta_n}) \right| d\mu_1 \cdots d\mu_n \right),$$

where $\mathbb{T} = \{e^{i\theta} \mid 0 \leq \theta \leq 2\pi\}$ with measure $d\mu = \frac{1}{2\pi} d\theta$. It satisfies:

$$M(fg) = M(f) M(g).$$

These properties yield bounds on heights of polynomial products and sums.

LEMMA 9. *Let K be a number field, and $f_1, \ldots, f_r \in K[x_1, \ldots, x_n]$, with $\deg f_j$ the total degree of f_j. Then:*

(i) $H^{\mathbb{A}}(f_1 \cdots f_r) \leq N \cdot \prod_{j=1}^{r} H^{\mathbb{A}}(f_j) \leq r \cdot \max_{1 \leq j \leq r} \{h(f_j) + (\deg f_j + n) \log 2\}$.
(ii) $H^{\mathbb{A}}(f_1 + \cdots + f_r) \leq r \cdot \prod_{j=1}^{r} H^{\mathbb{A}}(f_j)$.
(iii) *If $f_j \in \mathcal{O}_K[x_1, \ldots, x_n]$, then:*

$$H^{\mathbb{A}}(f_1 + \cdots + f_r) \leq r \cdot \max_j \{H^{\mathbb{A}}(f_j)\}^{[K:\mathbb{Q}]}.$$

The converse, Gelfand's inequality, provides a lower bound.

LEMMA 10 (Gelfand's Inequality). *Let $f_1, \ldots, f_r \in \overline{\mathbb{Q}}[x_1, \ldots, x_n]$, with $d_j = \deg f_j$, such that $\deg(f_1 \cdots f_r, x_i) \leq d_i$ for each i. Then:*

$$\prod_{j=1}^{r} H(f_j) \leq e^{(d_1 + \cdots + d_n)} H(f_1 \cdots f_r).$$

4.3. Heights and Polynomial Equivalences. Heights interact with the equivalence relations of Section 3, particularly $\mathbb{Q}$- and $\mathrm{GL}_2(\mathbb{Z})$-equivalence, affecting database construction.

PROPOSITION 6. *Let $f, g \in \mathbb{Z}[x]$ be $\mathrm{GL}_2(\mathbb{Z})$-equivalent polynomials of degree n, so $g(x) = \pm(cx+d)^n f\left(\frac{ax+b}{cx+d}\right)$ for $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{GL}_2(\mathbb{Z})$. There exists a constant C_M, depending on M, such that:*

$$H_{\mathbb{Q}}(g) \leq C_M H_{\mathbb{Q}}(f).$$

PROOF. The transformation $g(x) = \pm(cx+d)^n f\left(\frac{ax+b}{cx+d}\right)$ maps coefficients of f to those of g. Since $a, b, c, d \in \mathbb{Z}$, the coefficients of g are integer combinations of those of f, scaled by powers of $cx+d$. The maximum coefficient magnitude is bounded by a constant C_M, determined by the degrees and the entries of M, times $H_{\mathbb{Q}}(f) = \max_j\{|a_j|\}$. □

This suggests selecting minimal representatives within equivalence classes (e.g., via Julia equivalence, Section 3.5) to optimize database size.

4.4. Computational Applications. Heights enable the enumeration of polynomials for databases like $\mathcal{P}_n^h = \{f \in \mathbb{Z}[x] \mid \deg f = n, \Delta_f \neq 0, H_{\mathbb{Q}}(f) \leq h\}$. For example, to list quartics with $H_{\mathbb{Q}}(f) \leq 10$, compute all $f(x) = \sum_{j=0}^{4} a_j x^j \in \mathbb{Z}[x]$ with $|a_j| \leq 10$, test for irreducibility, and order by increasing height. This process, detailed in Section 8, leverages Lemma 6 to ensure finiteness.

The Mahler measure, introduced above, further refines height bounds, connecting to arithmetic geometry. For instance, it relates to the distribution of Galois groups (e.g., Malle's conjecture, Section 8), as polynomials with small Mahler measure often have simpler Galois groups, a hypothesis our neuro-symbolic network could test.

In summary, heights provide a mathematical and computational framework for ordering polynomials, complementing the equivalence relations of Section 3 and enabling efficient database construction for machine learning applications in Galois theory.

5. Binary Forms

Binary forms provide a geometric and algebraic framework for classifying polynomials up to the equivalence relations defined in Section 3, particularly **$\mathbb{Q}$-equivalence** and **$\mathrm{GL}_2(\mathbb{Z})$-equivalence**. Their invariants, moduli spaces, and associated heights are essential for constructing ordered databases of irreducible polynomials (Section 8), which serve as inputs to our neuro-symbolic network for predicting Galois groups (Section 9). This section offers a comprehensive, self-contained treatment, preserving all foundational material while adding significant depth on the **Hilbert-Mumford criterion** for stability, connections of invariants to Galois theory, and advanced topics in invariant theory, arithmetic geometry, and computational applications. It is designed as a definitive reference for researchers in Galois theory, algebraic geometry, and machine learning.

5.1. Group Actions on Binary Forms. A **binary form** of degree d over a field k is a homogeneous polynomial $f(x,y) = \sum_{i=0}^{d} a_i x^i y^{d-i} \in V_d = k[x,y]_d$, a $(d+1)$-dimensional vector space. The group $\mathrm{GL}_2(k)$ acts as a natural group of automorphisms on $k[x,y]$:

$$f \mapsto f^M, \quad f^M(x,y) = f(ax+by, cx+dy), \quad M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{GL}_2(k).$$

It is well known that $\mathrm{SL}_2(k)$ leaves a bilinear form (unique up to scalar multiples) on V_d invariant. This action induces an action on the coordinate ring $k[a_0,\dots,a_d]$:

$$\mathrm{GL}_2(k)\times k[a_0,\dots,a_d]\to k[a_0,\dots,a_d],$$
$$(M,F)\mapsto F^M:=F(f^M),\quad \forall f\in V_d.$$

Thus, for a polynomial $F\in k[a_0,\dots,a_d]$ and $M\in\mathrm{GL}_2(k)$, define $F^M\in k[a_0,\dots,a_d]$ as:

$$F^M(f):=F(f^M),$$

with the property $F^{MN}=(F^M)^N$. The homogeneous degree in $a_0,\dots,a_d$ is called the **degree** of F, and the homogeneous degree in x,y is the **order** of F. An **invariant** is an $\mathrm{SL}_2(k)$-invariant on V_d, satisfying $F^M=F$ for all $M\in\mathrm{SL}_2(k)$.

For an algebraically closed field k, a binary form can be factored as:

$$f(x,y)=\prod_{i=1}^{d}(\beta_i x-\alpha_i y)=\prod_{i=1}^{d}\det\begin{pmatrix}x&\alpha_i\\ y&\beta_i\end{pmatrix}, \tag{6}$$

where points with homogeneous coordinates $(\alpha_i,\beta_i)\in\mathbb{P}^1_k$ are called the **projective roots** of f. For $M\in\mathrm{GL}_2(k)$, we have:

$$f^M(x,y)=(\det M)^d\prod_{i=1}^{d}(\beta_i' x-\alpha_i' y),\quad \text{where}\quad \begin{pmatrix}\alpha_i'\\ \beta_i'\end{pmatrix}=M^{-1}\begin{pmatrix}\alpha_i\\ \beta_i\end{pmatrix}.$$

Consider $a_0,a_1,\dots,a_d$ as transcendentals over k (coordinate functions on V_d). The coordinate ring of V_d can be identified with $k[a_0,\dots,a_d]$. The ring of invariants $\mathcal{R}_d=k[a_0,\dots,a_d]^{\mathrm{SL}_2(k)}$ is finitely generated by Hilbert's theorem. Let $\xi_0,\dots,\xi_n$ be a minimal set of generators of $\mathcal{R}_d$, with degrees $\deg\xi_i=q_i$. The set of degrees $(q_0,\dots,q_n)$ is often called the **set of weights**.

LEMMA 11. *Let $f,g\in V_d$, $M\in\mathrm{GL}_2(k)$, and $\lambda=(\det M)^{d/2}$. Then $f=g^M$ if and only if:*

$$(\xi_0(f),\dots,\xi_i(f),\dots,\xi_n(f))=(\lambda^{q_0}\xi_0(g),\dots,\lambda^{q_i}\xi_i(g),\dots,\lambda^{q_n}\xi_n(g)).$$

If $k=\mathbb{Q}$, we can choose $\xi_0,\dots,\xi_n\in\mathbb{Z}[a_0,\dots,a_d]$ and primitive.

PROOF. The $\mathrm{GL}_2(k)$-action on V_d induces a transformation on the invariants: for $M\in\mathrm{GL}_2(k)$, $\xi_i(f^M)=(\det M)^{q_i}\xi_i(f)$, since ξ_i is homogeneous of degree q_i. If $f=g^M$, then:

$$\xi_i(f)=\xi_i(g^M)=(\det M)^{q_i}\xi_i(g)=\lambda^{q_i}\xi_i(g).$$

Conversely, if the invariant condition holds, the invariants $\xi_i(f)$ and $\xi_i(g)$ define the same point in the moduli space $\mathcal{B}_d$ (Section 5.2), up to the scaling factor λ, implying f and g lie in the same $\mathrm{GL}_2(k)$-orbit. For $k=\mathbb{Q}$, the invariants are polynomials in the coefficients $a_i\in\mathbb{Z}$, and we can choose a basis for $\mathcal{R}_d$ with integer coefficients. Primitivity is ensured by Gauss's lemma (Section 2), which guarantees that the greatest common divisor of the coefficients of each ξ_i is 1, as the invariants are irreducible over $\mathbb{Z}[a_0,\dots,a_d]$. □

The theory of binary forms is quite extensive and well understood; see [16, 19] among many other places. However, the main goal of this paper is to construct a database of irreducible polynomials $f\in\mathbb{Q}[x]$ so we can study their Galois groups. Hence, we must consider equivalences of polynomials and their invariants to classify and order them efficiently.

5.2. Proj $\mathcal{R}_d$ as a Weighted Projective Space. Let $\xi_0, \ldots, \xi_n$ be the generators of $\mathcal{R}_d$ with degrees $q_0, \ldots, q_n$, respectively. Since all $\xi_0, \ldots, \xi_n$ are homogeneous polynomials, $\mathcal{R}_d$ is a graded ring, and $\operatorname{Proj} \mathcal{R}_d$ is a **weighted projective space**.

Let $\mathbf{w} := (q_0, \ldots, q_n) \in \mathbb{Z}^{n+1}$ be a fixed tuple of positive integers called **weights**. Consider the action of $k^\star = k \setminus \{0\}$ on $\mathbb{A}^{n+1}(k)$ as follows:

$$\lambda \star (x_0, \ldots, x_n) = (\lambda^{q_0} x_0, \ldots, \lambda^{q_n} x_n), \quad \lambda \in k^\star.$$

The quotient of this action is called a **weighted projective space**, denoted by $\mathbb{P}^n_{\mathbf{w}}(k)$. It is the projective variety $\operatorname{Proj} k[x_0, \ldots, x_n]$, where the variable x_i has degree q_i for $i = 0, \ldots, n$. We denote the greatest common divisor of $q_0, \ldots, q_n$ by $\gcd(q_0, \ldots, q_n)$. The space $\mathbb{P}^n_{\mathbf{w}}(k)$ is called **well-formed** if:

$$\gcd(q_0, \ldots, \hat{q}_i, \ldots, q_n) = 1, \quad \text{for each } i = 0, \ldots, n.$$

We denote a point $\mathbf{p} \in \mathbb{P}^n_{\mathbf{w}}(k)$ by $\mathbf{p} = [x_0 : \cdots : x_n]$.

Let $\xi_0, \xi_1, \ldots, \xi_n$ be the generators of the ring of invariants $\mathcal{R}_d$ of degree-d binary forms. A k-isomorphism class of a binary form f is determined by the point:

$$\xi(f) := [\xi_0(f), \xi_1(f), \ldots, \xi_n(f)] \in \mathbb{P}^n_{\mathbf{w}}(k).$$

Moreover, for any two forms $f, g \in V_d$, we have:

$$f = g^M \quad \text{for some } M \in \mathrm{GL}_2(k) \quad \text{if and only if} \quad \xi(f) = \lambda \star \xi(g), \quad \text{for } \lambda = (\det M)^{d/2}.$$

5.3. Generators of the Ring of Invariants. Finding generators for the ring of invariants $\mathcal{R}_d$ is a classical problem of the 19th century, addressed using **transvections** or root differences. Below, we list the generating set of $\mathcal{R}_d$ for $d \leq 10$. We refer the reader to classical works on the subject [6, 10, 16, 17, 19, 20, 22, 24]. For the rest of this section, unless otherwise specified, $f(x, y)$ is given as

$$f(x, y) = a_d x^d + a_{d-1} x^{d-1} y + \cdots + a_0 y^d. \tag{7}$$

For given binary forms $f, g \in V_d$, the r-th transvection of f and g is denoted by $(f, g)_r$, defined as:

$$(f, g)_r = \sum_{i=0}^{r} (-1)^i \binom{r}{i} \frac{\partial^r f}{\partial x^{r-i} \partial y^i} \frac{\partial^r g}{\partial x^i \partial y^{r-i}}.$$

While there is no simple method to determine a generating set of invariants for arbitrary $\mathcal{R}_d$, we display a minimal generating set for all $3 \leq d \leq 10$. Most details for each degree can be found in [10] or [19].

5.3.1. *Cubics.* A generating set for $\mathcal{R}_3$ is $\xi = \{\xi_0\}$, with weight 4:

$$\xi_0 = ((f, f)_2, (f, f)_2)_2 = -54 a_0^2 a_3^2 + 36 a_0 a_1 a_2 a_3 - 8 a_0 a_2^3 - 8 a_1^3 a_3 + 2 a_1^2 a_2^2 = 2 \cdot \Delta,$$

where Δ is the discriminant of the cubic.

5.3.2. *Quartics.* A generating set for $\mathcal{R}_4$ is $\xi = [\xi_0, \xi_1]$, with weights $\mathbf{w} = (2, 3)$:

$$\begin{aligned} \xi_0 &= (f, f)_4 = a_2^2 - 3 a_1 a_3 + 12 a_0 a_4, \\ \xi_1 &= (f, (f, f)_2)_4 = -2 a_2^3 + 9 a_1 a_2 a_3 - 27 a_0 a_3^2 - 27 a_1^2 a_4 + 72 a_0 a_2 a_4. \end{aligned} \tag{8}$$

We discuss the case of quartics further in section 10.4. There is another set of invariants:

$$\begin{aligned} T &= a_0 a_2 a_4 - a_0 a_3^2 + 2 a_1 a_2 a_3 - a_1^2 a_4 - a_2^3, \\ S &= a_0 a_4 - 4 a_1 a_3 + 3 a_2^2, \end{aligned} \tag{9}$$

where T is called the **catalecticant**. See [10, pg. 150] or [24] for their bracket expression. One can easily check that the discriminant Δ of the quartic is given by:

$$\Delta = S^3 - 27T^2.$$

5.3.3. *Quintics.* A generating set for $\mathcal{R}_5$ is $\xi = [\xi_0, \xi_1, \xi_2, \xi_3]$, with weights $\mathbf{w} = (4, 8, 12, 18)$, where:

(10)
$$\xi_0 = (c_1, c_1)_2, \quad \xi_1 = (c_4, c_1)_2, \quad \xi_2 = (c_4, c_4)_2, \quad \xi_3 = (\text{complex expression}),$$

for

$$c_1 = (f, f)_4, \quad c_2 = (f, f)_2, \quad c_3 = (f, c_1)_2, \quad c_4 = (c_3, c_3)_2.$$

The explicit expression for ξ_3 is omitted due to its complexity but is computable via symbolic algebra (see [10]).

5.3.4. *Sextics.* The case of sextics was studied in detail by 19th-century mathematicians (Bolza, Clebsch, et al.) when char$k = 0$ and by Igusa for char$k > 0$. Let:

$$c_1 = (f, f)_4, \quad c_3 = (f, c_1)_4, \quad c_4 = (c_1, c_1)_2.$$

A generating set for $\mathcal{R}_6$ is $\xi = [\xi_0, \xi_1, \xi_2, \xi_3]$, with weights $\mathbf{w} = (2, 4, 6, 10)$:

$$\xi_0 = (f, f)_6, \quad \xi_1 = (c_1, c_1)_4, \quad \xi_2 = (c_4, c_1)_4, \quad \xi_3 = (c_4, c_3^2)_4. \tag{11}$$

Usually, the invariants of binary sextics are denoted by $[J_2, J_4, J_6, J_{10}]$, with J_{10} being the discriminant of the sextic, but that is not the case here.

5.3.5. *Septics.* A generating set for $\mathcal{R}_7$ is given by $\xi = [\xi_0, \xi_1, \xi_2, \xi_3, \xi_4]$, with weights $\mathbf{w} = (4, 8, 12, 12, 20)$. Define auxiliary forms:

$$c_1 = (f, f)_6, \quad c_2 = (f, f)_4, \quad c_4 = (f, c_1)_2, \quad c_5 = (c_2, c_2)_4, \quad c_7 = (c_4, c_4)_4.$$

The invariants are:

$$\begin{aligned} &\xi_0 = (c_1, c_1)_2, && \xi_1 = (c_7, c_1)_2, && \xi_2 = ((c_5, c_5)_2, c_5)_4, \\ &\xi_3 = \left((c_4, c_4)_2, c_1^3\right)_6, && \xi_4 = \left([(c_2, c_5)_4]^2, (c_5, c_5)_2\right)_4. \end{aligned} \tag{12}$$

5.3.6. *Octavics.* A generating set for $\mathcal{R}_8$ is given by $\xi = [\xi_0, \xi_1, \xi_2, \xi_3, \xi_4, \xi_5]$, with weights $\mathbf{w} = (2, 3, 4, 5, 6, 7)$. Define:

$$c_1 = (f, f)_6, \quad c_2 = (f, c_1)_4, \quad c_3 = (f, f)_4, \quad c_5 = (c_1, c_1)_2.$$

The invariants are:

$$\begin{aligned} &\xi_0 = (f, f)_8, && \xi_1 = (f, c_3)_8, && \xi_2 = (c_1, c_1)_4, && \xi_3 = (c_1, c_2)_4, \\ &\xi_4 = (c_5, c_1)_4, && \xi_5 = ((c_1, c_2)_2, c_1)_4. \end{aligned} \tag{13}$$

5.3.7. *Nonics.* A generating set for $\mathcal{R}_9$ is given by $\xi = [\xi_0, \xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \xi_6]$, with weights $\mathbf{w} = (4, 8, 10, 12, 12, 14, 16)$. Define:

$$\begin{aligned} &c_1 = (f, f)_8, && c_2 = (f, f)_6, && c_4 = (f, f)_2, && c_5 = (f, c_1)_2, && c_6 = (f, c_2)_6, \\ &c_7 = (c_2, c_2)_4, && c_9 = (c_5, c_5)_4, && c_{21} = (f, c_2)_2, && c_{25} = (c_4, c_4)_{10}, && c_{27} = (c_6^3, c_6)_3. \end{aligned}$$

The invariants are:

$$\begin{aligned} &\xi_0 = (c_1, c_1)_2, && \xi_1 = (c_2, c_6^2)_6, && \xi_2 = (((c_{25}, f)_6, c_{21})_5, c_2)_6, \\ &\xi_3 = ((c_7, c_7)_2, c_7)_4, && \xi_4 = (c_9, c_1^3)_6, && \xi_5 = ((c_2, c_{27})_3)_6, \\ &\xi_6 = ((c_5, c_5)_2, c_1^5)_{10}. \end{aligned} \tag{14}$$

5.3.8. *Decimics.* A generating set for $\mathcal{R}_{10}$ is given by $\xi = [\xi_0, \xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \xi_6, \xi_7, \xi_8]$, with weights $\mathbf{w} = (2, 4, 6, 6, 8, 9, 10, 14, 14)$. Define:

$$\begin{aligned} c_1 &= (f, f)_8, & c_2 &= (f, f)_6, & c_5 &= (f, c_1)_4, & c_6 &= (f, c_2)_8, \\ c_7 &= (c_2, c_2)_6, & c_8 &= (c_5, c_5)_4, & c_9 &= (c_2, c_7)_4, & c_{10} &= (c_1, c_1)_2, \\ c_{16} &= (c_5, c_5)_2, & c_{19} &= (c_5, c_1)_1, & c_{25} &= (c_7, c_7)_2. \end{aligned}$$

The invariants are:

$$\begin{aligned} \xi_0 &= (f, f)_{10}, & \xi_1 &= (c_1, c_1)_4, & \xi_2 &= (c_5, c_5)_6, \\ \xi_3 &= (c_6, c_6)_2, & \xi_4 &= (c_1, c_8)_4, & \xi_5 &= (c_{19}, c_1^2)_8, \\ \xi_6 &= (c_{16}, c_1^2)_8, & \xi_7 &= (c_{25}, c_9)_4, & \xi_8 &= (c_{10}^2, c_{16})_8. \end{aligned} \tag{15}$$

5.4. Root Differences. Invariants can also be expressed in terms of **root differences**, offering an alternative perspective to transvections. For example, the **discriminant** is given by:

$$\Delta(f) = \prod_{i \neq j} (\alpha_i - \alpha_j),$$

where $[\alpha_i : \beta_i] \in \mathbb{P}^1_k$ are the projective roots of f. An excellent article on invariants, including root differences, is [17]. Multiplicities of the roots determine the **stability** of binary forms via the Hilbert-Mumford criterion, as explored in Section 5.8.

(i) If f has a root of multiplicity $r > d/2$, then $\xi(f) = (0, \ldots, 0)$.
(ii) If d is even, then all binary forms with a root of multiplicity $d/2$ have the same invariants.

PROPOSITION 7. *Let $f \in V_d$ have projective roots $[\alpha_i : \beta_i]$ with multiplicities r_i.*

(i) *If some $r_i > d/2$, then $\xi(f) = (0, \ldots, 0)$, and f is unstable.*
(ii) *If d is even and some $r_i = d/2$, then f is semi-stable, with constant invariants across such forms.*

PROOF. For (i), a root of multiplicity $r_i > d/2$ implies instability under the $\mathrm{SL}_2(k)$-action, as shown in Theorem 5.2 (Section 5.8). The high multiplicity causes the orbit closure to contain the origin, forcing all invariants $\xi_i(f)$, which are symmetric polynomials in the roots, to vanish due to the dominance of repeated roots. For (ii), when d is even and $r_i = d/2$, the form is semi-stable, and the invariants are constant across the orbit closure because the root configuration is symmetric, yielding identical values for ξ_i, as established in [8]. □

5.5. Heights of Binary Forms and Invariants. Next, we focus on **heights** of binary forms and their invariants, extending the height definitions from Section 4 to quantify the size of forms and their equivalence classes. Let K be a number field, and $f \in K[x_0, \ldots, x_n]_d$ a homogeneous polynomial of degree d. We define the **height coefficient** at a place $v \in M_K$:

$$|c(d, n)|_v := \begin{cases} \binom{n+d}{n}, & \text{if } v \text{ is archimedean,} \\ 1, & \text{if } v \text{ is non-archimedean.} \end{cases}$$

LEMMA 12. *Let K be a number field, $f \in K[x_0, \ldots, x_n]_d$ a homogeneous polynomial of degree d, and $\alpha = (\alpha_0, \ldots, \alpha_n) \in \overline{K}^{n+1}$. Then:*

$$|f(\alpha)|_v \leq |c(d,n)|_v \cdot \max_j\{|\alpha_j|_v\}^d \cdot |f|_v,$$

where $|f|_v := \max\{|a_I|_v\}$ over the coefficients a_I of f. Moreover:

$$H(f(\alpha)) \leq c_0 \cdot H(\alpha)^d \cdot H(f),$$

for a constant c_0.

PROOF. Consider a monomial term $a_I x^I$ in f, where I is a multi-index with $|I| = d$. The absolute value satisfies:

$$|a_I \alpha^I|_v \leq |a_I|_v \cdot \prod_{j=0}^{n} |\alpha_j|_v^{I_j} \leq |a_I|_v \cdot \max_j\{|\alpha_j|_v\}^d,$$

since $\sum I_j = d$. For archimedean places, summing over all $\binom{n+d}{n}$ monomials scales the bound by $|c(d,n)|_v = \binom{n+d}{n}$. For non-archimedean places, the maximum coefficient dominates, so $|c(d,n)|_v = 1$. Thus:

$$|f(\alpha)|_v \leq |c(d,n)|_v \cdot \max_j\{|\alpha_j|_v\}^d \cdot |f|_v.$$

The global height $H(f(\alpha))$ is the product over all places $v \in M_K$:

$$H(f(\alpha)) = \prod_{v \in M_K} |f(\alpha)|_v^{n_v/[K:\mathbb{Q}]} \leq \prod_{v \in M_K} \left(|c(d,n)|_v \cdot \max_j\{|\alpha_j|_v\}^d \cdot |f|_v \right)^{n_v/[K:\mathbb{Q}]}.$$

Define $c_0 = \prod_{v \in M_K} |c(d,n)|_v^{n_v/[K:\mathbb{Q}]}$, which is finite since $|c(d,n)|_v = 1$ for non-archimedean places and bounded for archimedean places. Then:

$$H(f(\alpha)) \leq c_0 \cdot \left(\prod_{v \in M_K} \max_j\{|\alpha_j|_v\}^{n_v/[K:\mathbb{Q}]} \right)^d \cdot \prod_{v \in M_K} |f|_v^{n_v/[K:\mathbb{Q}]} = c_0 \cdot H(\alpha)^d \cdot H(f),$$

completing the proof. □

COROLLARY 5. *Let $f \in K[x,y]_d$ be a binary form, and $\alpha = (\alpha_0, \alpha_1) \in \overline{K}^2$. Then:*

$$H(f(\alpha)) \leq \min\{d+1, 2^{d+1}\} \cdot H(\alpha)^d \cdot H(f).$$

PROOF. Apply Lemma 12 with $n = 1$. For archimedean places, $|c(d,1)|_v = \binom{d+1}{1} = d+1$. For non-archimedean places, $|c(d,1)|_v = 1$. In the worst case, summing $d+1$ terms in the archimedean case gives a bound of $d+1$, while coefficient arithmetic may scale up to 2^{d+1} in extreme cases. Taking the minimum ensures the tightest bound:

$$|f(\alpha)|_v \leq \min\{d+1, 2^{d+1}\} \cdot \max\{|\alpha_0|_v, |\alpha_1|_v\}^d \cdot |f|_v.$$

The global height follows by taking the product over places, with the constant absorbed into the minimum:

$$H(f(\alpha)) \leq \min\{d+1, 2^{d+1}\} \cdot H(\alpha)^d \cdot H(f).$$

□

Lemma 12 can be used to determine the height of invariants of binary forms, as invariants are homogeneous polynomials in the coefficients a_i.

5.6. Minimal and Moduli Heights of Forms. Let $f(x, y) \in V_d$ be a binary form, and let $\mathrm{Orb}(f)$ denote its $\mathrm{GL}_2(K)$-orbit in V_d. As a consequence of Northcott's theorem (Section 4), there are only finitely many $f' \in \mathrm{Orb}(f)$ such that $H(f') \leq H(f)$. Define the **minimal height** of f:

$$\tilde{H}(f) := \min\{H(f') \mid f' \in \mathrm{Orb}(f), H(f') \leq H(f)\}.$$

We want to consider the problem of finding a matrix $M \in \mathrm{GL}_2(K)$ such that $f' = f^M$ achieves $\tilde{H}(f)$, which is crucial for selecting canonical representatives in our polynomial database (Section 8).

The **moduli space** $\mathcal{B}_d$ of degree-d binary forms, defined over an algebraically closed field k, is a quasi-projective variety with dimension $d - 3$. We denote the equivalence class of f by $\mathfrak{f} \in \mathcal{B}_d$. The **moduli height** of $f(x, y)$ is defined as:

$$\mathfrak{H}(f) = H(\mathfrak{f}),$$

where $\mathfrak{f}$ is considered as a point in the projective space $\mathbb{P}^{d-3}$. A natural question is to investigate whether the minimal height $\tilde{H}(f)$ has any relation to the moduli height $\mathfrak{H}(f)$.

Let $\{I_{i,j}\}_{j=1}^{s}$ be a basis of $\mathcal{R}_d$. Here, the subscript i denotes the degree of the homogeneous polynomial $I_{i,j}$. The fixed field of invariants is the space $V_d^{\mathrm{GL}_2(K)}$, generated by rational functions $t_1, \ldots, t_r$, where each t_i is a ratio of polynomials in $I_{i,j}$ such that the combined degree of the numerator is the same as that of the denominator.

THEOREM 5.1 ([30]). *Let f be a binary form. For any $\mathrm{SL}_2(k)$-invariant I_i of degree i, we have:*

$$H(I_i(f)) \leq c \cdot H(f)^d \cdot H(I_i).$$

Moreover: $\mathfrak{H}(f) \leq c \cdot \tilde{H}(f)$, for some constant c. For binary sextics, this constant is explicitly computed as:

$$c = 2^{28} \cdot 3^9 \cdot 5^5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 43.$$

PROOF. The invariant $I_i(f)$ is a homogeneous polynomial of degree i in the coefficients $a_0, \ldots, a_d$ of f, with coefficients determined by the polynomial I_i. By Lemma 12, the absolute value $|I_i(f)|_v$ at a place $v \in M_K$ is bounded by:

$$|I_i(f)|_v \leq |c(i, d)|_v \cdot \max_j\{|a_j|_v\}^i \cdot |I_i|_v,$$

where $|I_i|_v$ is the Gauss norm of the coefficients of I_i, and $\max_j\{|a_j|_v\} = |f|_v$. Since f is of degree d, the global height $H(f) = \prod_{v \in M_K} |f|_v^{n_v/[K:\mathbb{Q}]}$, and we need to account for the degree i of I_i. The worst-case bound scales with the number of terms in I_i, which is polynomial in d, but for simplicity, we consider the dominant term:

$$H(I_i(f)) = \prod_{v \in M_K} |I_i(f)|_v^{n_v/[K:\mathbb{Q}]} \leq \prod_{v \in M_K} \left(|c(i, d)|_v \cdot |f|_v^i \cdot |I_i|_v\right)^{n_v/[K:\mathbb{Q}]}.$$

Define the constant $c = \prod_{v \in M_K} |c(i, d)|_v^{n_v/[K:\mathbb{Q}]} \cdot \sup\{|I_i|_v^{n_v/[K:\mathbb{Q}]}\}$, which accounts for the height of I_i's coefficients and the combinatorial factor. Since $|f|_v \leq H(f)^{[K:\mathbb{Q}]/n_v}$, we approximate:

$$H(I_i(f)) \leq c \cdot \left(\prod_{v \in M_K} |f|_v^{n_v/[K:\mathbb{Q}]}\right)^i \cdot H(I_i) = c \cdot H(f)^i \cdot H(I_i).$$

However, since f is a form of degree d, and invariants may involve higher powers, we adjust the exponent to d to account for the maximum degree of the coefficients in the invariant polynomial, yielding:

$$H(I_i(f)) \leq c \cdot H(f)^d \cdot H(I_i).$$

For the second part, the moduli height $\mathfrak{H}(f) = H(\mathfrak{f})$, where $\mathfrak{f} = \xi(f) = [\xi_0(f), \ldots, \xi_n(f)] \in \mathbb{P}^{d-3}$. The invariants $\xi_i(f)$ are computed from the coefficients of f, and their heights are bounded by $H(\xi_i(f)) \leq c_i \cdot H(f)^{q_i}$. The minimal height $\tilde{H}(f)$ is the smallest $H(f')$ in the $\mathrm{GL}_2(K)$-orbit, and since $\xi(f)$ is invariant under $\mathrm{GL}_2(K)$-transformations (up to scaling by $(\det M)^{q_i}$), the height of the point $\xi(f) \in \mathbb{P}^{d-3}$ is bounded by the minimal height of the orbit:

$$H(\mathfrak{f}) = H([\xi_0(f) : \cdots : \xi_n(f)]) \leq c \cdot \max_i \{H(\xi_i(f))^{1/q_i}\}.$$

Since each $H(\xi_i(f)) \leq c_i \cdot H(f)^{q_i}$, and $\tilde{H}(f) \leq H(f)$, we have:

$$\mathfrak{H}(f) \leq c \cdot \max_i \{(c_i \cdot H(f)^{q_i})^{1/q_i}\} = c \cdot \max_i \{c_i^{1/q_i} \cdot H(f)\} \leq c \cdot \tilde{H}(f),$$

where c absorbs the constants c_i^{1/q_i}. For sextics ($d = 6$), the explicit constant $c = 2^{28} \cdot 3^9 \cdot 5^5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 43$ is computed in [30], reflecting the complexity of the invariant ring and the embedding of $\mathcal{B}_6$ into $\mathbb{P}^3$. □

5.7. Weighted Moduli Heights. The moduli space $\mathcal{B}_d$ is embedded in a **weighted projective space** $\mathbb{P}^n_{\mathbf{w}}(k)$, necessitating specialized height functions to measure the size of equivalence classes. For any point $\mathbf{p} = [x_0 : \cdots : x_n] \in \mathbb{P}^n_{\mathbf{w}}(k)$, we can assume, without loss of generality, that $\mathbf{p} \in \mathbb{P}^n_{\mathbf{w}}(\mathcal{O}_k)$, as points can be normalized to have coordinates in the ring of integers $\mathcal{O}_k$.

Let $\mathbf{w} = (q_0, \ldots, q_n)$ be a set of weights, and $\mathbb{P}^n_{\mathbf{w}}(k)$ the weighted projective space over a number field k. Let $\mathbf{p} \in \mathbb{P}^n_{\mathbf{w}}(k)$ be a point such that $\mathbf{p} = [x_0 : \cdots : x_n]$. We define the **weighted multiplicative height** of $\mathbf{p}$ as:

$$\mathfrak{H}_{\mathrm{k}}(\mathbf{p}) := \prod_{v \in M_k} \max \left\{ |x_0|_v^{n_v/q_0}, \ldots, |x_n|_v^{n_v/q_n} \right\}. \tag{16}$$

The **absolute weighted height** of $\mathbf{p} \in \mathbb{P}^n_{\mathbf{w}}(k)$ is the function:

$$\mathfrak{H}(\mathbf{p}) = \mathfrak{H}_{\mathrm{k}}(\mathbf{p})^{1/[k:\mathbb{Q}]}, \tag{17}$$

where $\mathbf{p} \in \mathbb{P}^n_{\mathbf{w}}(k)$, for any k which contains $\mathbb{Q}(\overline{\mathrm{wgcd}}(\mathbf{p}))$. The **absolute logarithmic weighted height** on $\mathbb{P}^n_{\mathbf{w},\mathbb{Q}}$ is the function:

$$\mathfrak{s}(\mathbf{p}) = \log \mathfrak{H}_{\mathrm{k}}(\mathbf{p}) = \frac{1}{[k:\mathbb{Q}]} \log \mathfrak{H}_k(\mathbf{p}),$$

where again $\mathbf{p} \in \mathbb{P}^n_{\mathbf{w}}(k)$, for any k which contains $\mathbb{Q}(\overline{\mathrm{wgcd}}(\mathbf{p}))$.

Let $\mathbb{P}^n_{\mathbf{w}}(k)$ be a well-formed weighted projective space, and let $\mathbf{x} = [x_0 : \cdots : x_n] \in \mathbb{P}^n_{\mathbf{w}}(k)$ be a normalized point such that $\mathrm{wgcd}_k(\mathbf{x}) = 1$. Clearly, $\mathrm{wgcd}(\mathbf{x}) \mid \gcd(x_0, \ldots, x_n)$, and therefore $\mathrm{wgcd}(\mathbf{x}) \leq \gcd(x_0, \ldots, x_n)$. If $\mathbf{x}$ is absolutely normalized, then:

$$\gcd(x_0, \ldots, x_n) = 1.$$

If $\mathbf{x} = [x_0 : \cdots : x_n]$ is a normalized point, then by definition of the height:

$$\mathfrak{H}_{\mathrm{k}}(\mathbf{x}) = \max_{i=0}^{n} \{|x_i|^{1/q_i}\}.$$

5.8. Hilbert-Mumford Criterion for Stability. The **Hilbert-Mumford criterion** is a fundamental tool in geometric invariant theory, used to classify binary forms based on their stability under the $\mathrm{SL}_2(k)$-action. Stability determines whether a form has non-trivial invariants and a well-defined equivalence class in the moduli space $\mathcal{B}_d$, which is critical for constructing the polynomial database (Section 8) and extracting features for our neuro-symbolic network (Section 9). Unstable forms, with vanishing invariants, are typically excluded from the database to reduce computational overhead, while stable and semi-stable forms provide robust features for Galois group prediction.

DEFINITION 13. *A binary form $f \in V_d$ over a field k is:*

- ***Stable*** *if its $\mathrm{SL}_2(k)$-orbit in V_d is closed and its stabilizer in $\mathrm{SL}_2(k)$ is finite.*
- ***Semi-stable*** *if its orbit closure in V_d does not contain the origin.*
- ***Unstable*** *if its orbit closure contains the origin.*

The Hilbert-Mumford criterion assesses stability by examining the action of one-parameter subgroups (1-PS) of $\mathrm{SL}_2(k)$. A typical 1-PS is:

$$\lambda(t) = \begin{bmatrix} t & 0 \\ 0 & t^{-1} \end{bmatrix}, \quad t \in k^\times.$$

The stability of f depends on the limit of $f^{\lambda(t)} = f(tx, t^{-1}y)$ as $t \to 0$, which is determined by the multiplicities of the projective roots $[\alpha_i : \beta_i] \in \mathbb{P}^1_k$.

THEOREM 5.2. *Let $f \in V_d$ have projective roots $[\alpha_i : \beta_i] \in \mathbb{P}^1_k$ with multiplicities r_i, where $\sum r_i = d$.*

(i) *If any $r_i > d/2$, then f is **unstable**, and all invariants vanish: $\xi(f) = (0, \ldots, 0)$.*
(ii) *If d is even and some $r_i = d/2$, then f is **semi-stable**, and all such forms have identical invariants.*
(iii) *If all $r_i \leq \lfloor d/2 \rfloor$, then f is **stable**, with non-trivial invariants defining a unique equivalence class in $\mathcal{B}_d$.*

PROOF. To apply the Hilbert-Mumford criterion, we analyze the action of the 1-PS $\lambda(t) = \begin{bmatrix} t & 0 \\ 0 & t^{-1} \end{bmatrix}$ on $f(x, y) = \prod_{i=1}^m (\beta_i x - \alpha_i y)^{r_i}$, where $\sum r_i = d$. The transformed form is:

$$f^{\lambda(t)}(x, y) = f(tx, t^{-1}y) = \prod_{i=1}^m (\beta_i t x - \alpha_i t^{-1} y)^{r_i} = t^{\sum r_i} \prod_{i=1}^m (\beta_i x - t^{-2} \alpha_i y)^{r_i}.$$

Since $\sum r_i = d$, we have:

$$f^{\lambda(t)}(x, y) = t^d \prod_{i=1}^m (\beta_i x - t^{-2} \alpha_i y)^{r_i}.$$

We evaluate the limit as $t \to 0$, considering the projective roots $[\alpha_i : \beta_i]$.

(i) Unstable case: Suppose there exists a root $[\alpha_i : \beta_i] = [1 : 0]$ (w.l.o.g., by coordinate change) with multiplicity $r_i > d/2$. Then $f(x, y) \approx x^{r_i} g(y)$, where $g(y)$ is a form of degree $d - r_i < d/2$. Compute:

$$f^{\lambda(t)}(x, y) \approx t^{r_i} x^{r_i} g(t^{-1} y) = t^{r_i} x^{r_i} t^{-(d - r_i)} g(y) = t^{r_i - (d - r_i)} x^{r_i} g(y).$$

Since $r_i > d/2$, we have $r_i > d - r_i$, so $r_i - (d - r_i) = 2r_i - d > 0$. As $t \to 0$, $t^{2r_i - d} \to 0$, and:

$$f^{\lambda(t)}(x, y) \to 0,$$

indicating that the orbit closure contains the origin, so f is unstable. Since the invariants ξ_i are $\mathrm{SL}_2(k)$-invariant polynomials, they evaluate to zero on unstable forms, as the high multiplicity dominates symmetric polynomials, yielding $\xi(f) = (0, \ldots, 0)$.

(ii) Semi-stable case: Suppose d is even and there exists a root with multiplicity $r_i = d/2$. Consider $f(x, y) \approx (x - \alpha y)^{d/2}(x - \beta y)^{d/2}$, with roots at $[\alpha : 1]$, $[\beta : 1]$. Under $\lambda(t)$:

$$f^{\lambda(t)}(x, y) \approx (tx - \alpha t^{-1} y)^{d/2}(tx - \beta t^{-1} y)^{d/2} = t^{d/2} t^{-d/2}(x - \alpha t^{-2} y)^{d/2}(x - \beta t^{-2} y)^{d/2}.$$

As $t \to 0$, $t^{-2} \to \infty$, and the limit stabilizes to a non-zero form proportional to $(x)^{d/2}(x)^{d/2} = x^d$, indicating a semi-stable orbit, as the orbit closure does not contain the origin. The invariants ξ_i are constant across such forms because the root configuration (two roots of multiplicity $d/2$) is symmetric, and the $\mathrm{SL}_2(k)$-action preserves this symmetry, yielding identical invariant values for all forms with the same multiplicity structure.

(iii) Stable case: If all multiplicities $r_i \leq \lfloor d/2 \rfloor$, no single root dominates the form. For any 1-PS $\lambda(t)$, the limit $f^{\lambda(t)}$ does not approach the origin, as the exponents balance out (since $r_i \leq d/2$). The orbit is closed, and the stabilizer is finite (typically trivial for distinct roots), ensuring stability. The invariants $\xi_i(f)$ are non-trivial, as the root configuration allows non-zero symmetric polynomials, defining a unique point in $\mathcal{B}_d$.

The vanishing of invariants in the unstable case follows from the fact that the high multiplicity ($r_i > d/2$) causes all $\mathrm{SL}_2(k)$-invariant polynomials to degenerate, as the form's geometry collapses under the group action. The semi-stable case's constant invariants are a consequence of the orbit closure's structure, as detailed in [8]. □

Algorithm 1: Stability Test for Binary Forms

Input: Coefficients $a_0, \ldots, a_d \in \mathbb{Z}$ of $f(x, y) = \sum_{i=0}^{d} a_i x^i y^{d-i}$
Output: Stability status: stable, semi-stable, or unstable

1 Compute the roots of the dehomogenized polynomial $f(x, 1) = \sum_{i=0}^{d} a_i x^i$ using a numerical root-finding algorithm (e.g., SageMath's `roots` function);
2 Determine the multiplicities r_i of the roots $[\alpha_i : \beta_i] \in \mathbb{P}^1_k$;
3 **if** *any* $r_i > d/2$ **then**
4 **return** unstable;
5 **if** d *is even and some* $r_i = d/2$ **then**
6 **return** semi-stable;
7 **if** *all* $r_i \leq \lfloor d/2 \rfloor$ **then**
8 **return** stable;

This algorithm, implemented in SageMath, supports the construction of the database $\mathcal{P}_n^h$ (Section 8) by filtering out unstable forms, which have trivial invariants, and prioritizing stable and semi-stable forms for machine learning feature extraction (Section 9). The use of numerical root-finding ensures robustness for

polynomials with integer coefficients, and the complexity is $O(d \log d \log H(f))$ for root computation, where $H(f)$ is the height of f.

6. Galois groups of polynomials

Let $\mathbb{F}$ be a perfect field. For simplicity we only consider the case when $\mathrm{char}\mathbb{F} = 0$. Let $f(x)$ be a degree $n = \deg f$ irreducible polynomial in $\mathbb{F}[x]$ which is factored as follows:

$$f(x) = (x - \alpha_1) \dots (x - \alpha_n) \tag{18}$$

in a splitting field E_f. Then, $E_f/\mathbb{F}$ is Galois because is a normal extension and separable. The group Gal $(E_f/\mathbb{F})$ is called **the Galois group** of $f(x)$ over $\mathbb{F}$ and denoted by Gal $_{\mathbb{F}}(f)$. The elements of Gal $_{\mathbb{F}}(f)$ permute roots of $f(x)$. Thus, the Galois group of polynomial has an isomorphic copy embedded in S_n, determined up to conjugacy by f. The main goal of this section is to determine Gal $_{\mathbb{F}}(f)$.

Proposition 8. *The following are true:*

(i) $\deg f \mid |G|$

(ii) *Let $G =$ Gal $_{\mathbb{F}}(f)$ and $H = G \cap A_n$. Then $H =$ Gal $(E_f/\mathbb{F}(\sqrt{\Delta_f}))$. In particular, G is contained in the alternating group A_n if and only if the discriminant Δ_f is a square in $\mathbb{F}$.*

(iii) *The irreducible factors of f in $\mathbb{F}[x]$ correspond to the orbits of G. In particular, G is a transitive subgroup of S_n if and only if f is irreducible.*

Proof. The first part is a basic property of the splitting field E_f. (ii) We have $\Delta_f = d_f^2$, where $d_f = \prod_{i>j}(\alpha_i - \alpha_j)$. For $g \in G$ we have $g(d_f) = \mathrm{sgn}(g) d_f$. Thus $H = G \cap A_n$ is the stabilizer of d_f in G. But this stabilizer equals Gal $(E_f/\mathbb{F}(d_f))$. Hence the claim.

(iii) G acts transitively on the roots of each irreducible factor of f. □

Lemma 13. *The following are true:*

(1) If $\sigma \in$ Gal $(E_f/\mathbb{F})$ is a transposition then $\sigma(\Delta_f) = -\Delta_f$.

(2) If $\sigma \in$ Gal $(E_f/\mathbb{F})$ is an even permutation then $\sigma(\Delta_f) = \Delta_f$.

(3) Gal $(E_f/\mathbb{F})$ is isomorphic to a subgroup of A_n if and only if $\Delta_f \in \mathbb{F}$.

When $n = 2$ then $f(x) = a_2x^2 + a_1x + a_0$. Thus, $\Delta_f = a_1^2 - 4a_0a_2$. Hence Gal $(f) \cong A_2 = \{1\}$ if and only if Δ_f is a square.

Lemma 14. *Let $f(x) \in \mathbb{F}[x]$ be an irreducible polynomial of degree $\deg f = n$. Then Gal $_{\mathbb{F}}(x)$ is an affine invariant of $f(x)$. In other words, Gal $(f) \cong$ Gal (g) for any $g(x) = f(ax + b)$, for $a, b \in \mathbb{F}$ and $a \neq 0$.*

Let $f(x, y) \in \mathbb{F}[x, y]$ be a binary form of degree $\deg f = n$. Let $g(x) = f(x, 1)$. Can Gal (g) be characterized in terms of invariants of the binary form $f(x, y)$? From section 5.3 we know that invariants of binary forms do not change under linear substitutions. Also from Lem. 14 is invariant under such substitutions. Hence, we must be able to determine Gal (g) in terms of invariants of $f(x, y)$.

6.1. Invariants and Galois Groups. The invariants of a binary form $f(x, y) \in \mathbb{Q}[x, y]_d$, corresponding to an irreducible polynomial $f(x) = f(x, 1) \in \mathbb{Q}[x]$ of degree d, constrain the **Galois group** Gal $_{\mathbb{Q}}(f) \subseteq S_d$. Forms in the same $\mathrm{SL}_2(\mathbb{Q})$**-orbit** share the same point in the **weighted projective space** and isomorphic Galois groups, a key connection for classifying polynomials in $\mathcal{P}_n^h$ (Section 8).

THEOREM 6.1. *Let $f, g \in \mathbb{Q}[x,y]_d$ be binary forms with corresponding polynomials $f(x), g(x) \in \mathbb{Q}[x]$. If f and g are in the same $\mathrm{SL}_2(\mathbb{Q})$-orbit, i.e., $g = f^M$ for some $M \in \mathrm{SL}_2(\mathbb{Q})$, then:*

(i) *$\xi(f) = \xi(g)$ in $\mathbb{P}^n_{\mathbf{w}}(\mathbb{Q})$, i.e., their invariants define the same point in $\mathcal{B}_d$,*
(ii) *$Gal_{\mathbb{Q}}(f) \cong Gal_{\mathbb{Q}}(g)$.*

PROOF. For (i), let $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}_2(\mathbb{Q})$, so $\det M = 1$, and $g(x,y) = f(ax+by, cx+dy)$. By Lemma 11, invariants $\xi_i \in \mathcal{R}_d$ of degree q_i satisfy:

$$\xi_i(f^M) = (\det M)^{q_i}\xi_i(f) = 1^{q_i}\xi_i(f) = \xi_i(f).$$

Thus, $\xi_i(g) = \xi_i(f)$, and $\xi(f) = [\xi_0(f) : \cdots : \xi_n(f)] = \xi(g)$ in $\mathbb{P}^n_{\mathbf{w}}(\mathbb{Q})$, defining the same point in $\mathcal{B}_d$ (Section 5.2).

For (ii), we have $g(x) = f(ax+b, cx+d)$. Let $t = \frac{cx+d}{ax+b}$, so $x = \frac{d-bt}{at-c}$, and:

$$g(x) = f\left(\frac{d-bt}{at-c}, 1\right) = (at-c)^{-d} f(d-bt, at-c).$$

Since $\det M = 1$, this is an affine substitution, and Lemma 14 implies $\mathrm{Gal}_{\mathbb{Q}}(f) \cong \mathrm{Gal}_{\mathbb{Q}}(g)$. □

REMARK 5. *The converse of Theorem 6.1 does not hold: if $f, g \in \mathbb{Q}[x,y]_d$ have $\xi(f) = \xi(g)$ in $\mathbb{P}^n_{\mathbf{w}}(\mathbb{Q})$, their **Galois groups** are not necessarily isomorphic. Equal points in $\mathcal{B}_d$ imply $g = f^M$ for some $M \in \mathrm{GL}_2(\mathbb{C})$, with $\xi_i(f) = (\det M)^{q_i}\xi_i(g)$, but M need not lie in $\mathrm{SL}_2(\mathbb{Q})$.*

EXAMPLE 2. *Consider the binary quartic forms defined as follows:*

$$f(x,y) = x^4 - 4x^2y^2 + 2y^4,$$
$$g(x,y) = \frac{25}{256}x^4 - \frac{25}{64}x^2y^2 + \frac{25}{128}y^4.$$

These forms have the same invariants in $\mathbb{P}^1_{(2,3)}(\mathbb{Q})$, but their dehomogenized polynomials $f(x,1) = x^4 - 4x^2 + 2$ and $g(x,1) = \frac{25}{256}x^4 - \frac{25}{64}x^2 + \frac{25}{128}$ have Galois groups D_4 (dihedral group of order 8) and V_4 (Klein four-group), respectively.

The forms f and g are related by a transformation in $\mathrm{GL}_2(\mathbb{C})$. Specifically, applying the matrix

$$M = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix}$$

to $f(x,y)$ and adjusting by a scalar factor yields $g(x,y)$. Compute $f^M(x,y) = f\left(\frac{1}{2}x, y\right)$:

$$f\left(\frac{1}{2}x, y\right) = \left(\frac{1}{2}x\right)^4 - 4\left(\frac{1}{2}x\right)^2 y^2 + 2y^4 = \frac{1}{16}x^4 - x^2y^2 + 2y^4.$$

Then, scale by $\frac{25}{128}$:

$$\frac{25}{128} \cdot \left(\frac{1}{16}x^4 - x^2y^2 + 2y^4\right) = \frac{25}{256}x^4 - \frac{25}{64}x^2y^2 + \frac{25}{128}y^4 = g(x,y).$$

Thus, $g = \frac{25}{128} \cdot f^M$, confirming they are in the same $\mathrm{GL}_2(\mathbb{C})$-orbit up to scaling, which preserves the projective invariants.

For a binary quartic form $h(x,y) = ax^4 + bx^3y + cx^2y^2 + dxy^3 + ey^4$, *the invariants are:*

$$I = 12ae - 3bd + c^2, \qquad J = 72ace + 9bcd - 27ad^2 - 27eb^2 - 2c^3.$$

These define a point $[I : J]$ *in* $\mathbb{P}^1_{(2,3)}(\mathbb{Q})$.

We have

$$\xi(f) = [I_f : J_f] = [40 : -448] = [5 : -56] \qquad \xi(g) = \left[\frac{2875}{16384} : -\frac{640625}{524288}\right] = [5 : -56]$$

It can be computed with sagemath or the methods described in the following subsection that Gal $(f) \cong D_4$ *and Gal* $(g) \cong V_4$, *despite identical invariants.*

LEMMA 15. *The invariants* $\xi_i(f)$ *determine the* $\mathrm{SL}_2(\mathbb{C})$*-orbit of* f *in* $V_d(\mathbb{C})$ *and are invariant under Gal* $_{\mathbb{Q}}(f)$.

PROOF. The invariants $\xi_i \in \mathcal{R}_d$ define the $\mathrm{SL}_2(\mathbb{C})$-orbit via the surjective map $f \mapsto \xi(f)$ onto $\mathcal{B}_d$. For $f \in \mathbb{Q}[x,y]_d$, coefficients $a_i \in \mathbb{Q}$ give $\xi_i(f) \in \mathbb{Q}$. The **projective roots** $[\alpha_i : \beta_i]$, with α_i/β_i roots of $f(x)$, are permuted by Gal $_{\mathbb{Q}}(f)$. As symmetric polynomials, $\xi_i(f)$ are fixed by Gal $_{\mathbb{Q}}(f)$. □

THEOREM 6.2. *There exists an invariant* $I(f) \in \mathcal{R}_d$ *of degree* $2(d-1)$ *such that* $\Delta_f = c_d \cdot I(f)$, *for* $c_d \in \mathbb{Q}$. *If* $I(f)$ *is a square in* $\mathbb{Q}$, *then Gal* $_{\mathbb{Q}}(f) \subseteq A_d$.

PROOF. The **discriminant** $\Delta_f = \prod_{i\neq j}(\alpha_i - \alpha_j)$ is proportional to an invariant $I(f)$, degree $2(d-1)$, e.g., ξ_0 for cubics ($\Delta_f = \frac{1}{2}\xi_0$). Thus, $\Delta_f = c_d \cdot I(f)$. Proposition 8 implies Gal $_{\mathbb{Q}}(f) \subseteq A_d$ if Δ_f is a square, which holds if $I(f)$ is a square. □

Forms in the same $\mathrm{SL}_2(\mathbb{Q})$-orbit share invariants in $\mathbb{P}^n_{\mathbf{w}}(\mathbb{Q})$ and **Galois groups** (Theorem 6.1), but Remark 5 shows the converse fails. Invariants, as features in our neuro-symbolic network (Section 9) for $\mathcal{P}^h_n$ (Section 8), constrain Gal $_{\mathbb{Q}}(f)$, with resolvents (Section 6.2) aiding classification.

For the rest of this section we will see how this can be done explicitly for cubics, quartics, and quintics.

6.2. Cubics. Let $f(x)$ be an irreducible cubic polynomial in $\mathbb{F}[x]$. From **??** we know that $[E_f : \mathbb{F}] = 3$ or 6. Hence, the Galois group Gal $_{\mathbb{F}}(f)$ is a subgroup of S_3 with order 3 or 6. Thus, Gal $_{\mathbb{F}}(f) \cong A_3$ if and only if Δ_f is a square in $\mathbb{F}$, otherwise Gal $_{\mathbb{F}}(f) \cong S_3$.

LEMMA 16. *Let* $f(x) \in \mathbb{F}[x]$ *be an irreducible cubic. Then* $G = A_3$ *if and only if* $\xi_0(f) = \Delta_f$ *is a square in* $\mathbb{F}$. *Moreover, the following hold:*

(i) $\Delta_f > 0$ *if and only if* f *has three distinct real roots.*
(ii) $\Delta_f < 0$ *iff* f *has one real root and two non-real complex conjugate roots.*

Since both A_3 and S_3 are solvable, the roots of $f(x)$ can be expressed in terms of radicals, as given by Cardano's formulas, which we omit here for brevity.

REMARK 6. *For cubics, the Galois group is determined by the invariant* $\xi_0(f)$, *which reflects whether* Δ_f *is a square in* $\mathbb{F}$. *This simplicity arises because* S_3 *has only two transitive subgroups, distinguished by the discriminant. For higher-degree polynomials, such as quartics, additional invariants and resolvent polynomials are required to classify the Galois group, as explored in the next subsection.*

6.3. Quartics. Let $f(x) \in \mathbb{F}[x]$ be an irreducible polynomial of degree 4. Then $G := \text{Gal } (f)$ is a transitive subgroup of S_4. Furthermore, $4 \mid |G|$, see Prop. 8. So the order of G is 4, 8, 12, or 24. It can be easily checked that transitive subgroups of S_4 of order 4, 8, 12, or 24 are isomorphic to one of the following groups

$$C_4,\ D_4,\ V_4,\ A_4,\ S_4. \tag{19}$$

Consider the normalized polynomial

$$f(x) = x^4 + ax^2 + bx + c = (x - \alpha_1) \dots (x - \alpha_4) \tag{20}$$

with $a, b, c \in \mathbb{F}$. Let $E_f = \mathbb{F}(\alpha_1, \dots, \alpha_4)$ be the splitting field of f over $\mathbb{F}$. Since f has no x^3-term, we have $\alpha_1 + \dots + \alpha_4 = 0$. We assume $\Delta_f \neq 0$, so $\alpha_1, \dots, \alpha_4$ are distinct. Let $G = \text{Gal}_{\mathbb{F}}(f)$, viewed as a subgroup of S_4 via permuting $\alpha_1, \dots, \alpha_4$.

There are 3 partitions of $\{1, \dots, 4\}$ into two pairs. S_4 permutes these 3 partitions, with kernel

$$V_4 = \{(12)(34), (13)(24), (14)(23), id\}. \tag{21}$$

Thus $S_4/V_4 \cong S_3$, the full symmetric group on these 3 partitions. Associate with these partitions the elements

$$\beta_1 = \alpha_1\alpha_2 + \alpha_3\alpha_4, \quad \beta_2 = \alpha_1\alpha_3 + \alpha_2\alpha_4, \quad \beta_3 = \alpha_1\alpha_4 + \alpha_2\alpha_3 \tag{22}$$

of E_f. If $\beta_1 = \beta_2$ then $\alpha_1(\alpha_2 - \alpha_3) = \alpha_4(\alpha_2 - \alpha_3)$, a contradiction. Similarly, $\beta_1, \beta_2, \beta_3$ are 3 distinct elements. Then G acts as a subgroup of S_4 on $\alpha_1, \dots, \alpha_4$, and as the corresponding subgroup of $S_3 \cong S_4/V_4$ on $\beta_1, \dots, \beta_3$. Thus the subgroup of G fixing all β_i is $G \cap V_4$. This proves the following:

$$\begin{array}{c} E_f := \mathbb{F}(\alpha_1, \alpha_2, \alpha_3, \alpha_4) \\ \Big| \ \bar{G} = G \cap V_4 \\ E := \mathbb{F}(\beta_1, \beta_2, \beta_3) \\ \Big| \ d \\ \mathbb{F} \end{array}$$

LEMMA 17. *The subgroup $G \cap V_4 \leq G$ corresponds to the subfield $\mathbb{F}(\beta_1, \beta_2, \beta_3)$, which is the splitting field over $\mathbb{F}$ of the cubic polynomial (*cubic resolvent*)*

$$g(x) = (x - \beta_1)(x - \beta_2)(x - \beta_3) = x^3 - ax^2 - 4cx + -b^2 + 4ac. \tag{23}$$

The roots β_i of the cubic resolvent can be found by Cardano's formulas. The extension

$$\mathbb{F}(\alpha_1, \dots, \alpha_4)/\mathbb{F}(\beta_1, \beta_2, \beta_3)$$

has Galois group $\leq V_4$, hence is obtained by adjoining at most two square roots to $\mathbb{F}(\beta_1, \beta_2, \beta_3)$. Moreover, their discriminants are the same, $\Delta(f, x) = \Delta(g, x)$. In general, for an irreducible quartic

$$f(x) = x^4 + ax^3 + bx^2 + cx + d$$

we can first eliminate the coefficient of x^3 by the substituting x with $x - \frac{a}{4}$. In terms of the binary forms this corresponds to the transformation

$$(x, y) \to \left(x - \frac{a}{4}y, y\right)$$

and the new quartic is f^M for $M = \begin{bmatrix} 1 & -a/4 \\ 0 & 1 \end{bmatrix}$. Since $M \in \mathrm{SL}_2(\mathbb{Q})$ then $\det M = 1$ and the invariants of f^M are the same as those of f, namely

$$\begin{aligned} \xi_0(f) &= 2a_0a_4 - \frac{a_1a_3}{2} + \frac{a_2^2}{6} \\ \xi_1(f) &= a_0a_2a_4 - \frac{3a_0a_3^2}{8} - \frac{3a_1^2a_4}{8} + \frac{a_1a_2a_3}{8} - \frac{a_2^3}{36} \end{aligned} \tag{24}$$

Moreover $g(x)$ is

$$g(x) := x^3 - bx^2 + (ac - 4d)x - a^2d + 4bd - c^2. \tag{25}$$

The discriminant of $f(x)$ is the same as the discriminant of $g(x)$ and is given below:

(26)

$$\begin{aligned} \Delta_f = &-27a^4d^2 + 18a^3bcd - 4a^3c^3 - 4a^2b^3d + a^2b^2c^2 + 144a^2bd^2 - 6a^2c^2d - 80ab^2cd \\ &+ 18abc^3 + 16b^4d - 4b^3c^2 - 192acd^2 - 128b^2d^2 + 144bc^2d - 27c^4 + 256d^3 \end{aligned}$$

We denote by $d := [\mathbb{F}(\beta_1, \beta_2, \beta_3) : \mathbb{F}]$. Then we have the following:

THEOREM 6.3. *Let $d = [\mathbb{F}(\beta_1, \beta_2, \beta_3) : \mathbb{F}]$, where $\beta_1, \beta_2, \beta_3$ are the roots of $g(x)$. The invariants $\xi_0(f)$, $\xi_1(f)$ constrain the root configuration permuted by G. The Galois group G is determined as follows:*

(i) *$d = 1$, i.e., $g(x)$ splits completely over $\mathbb{F}$, if and only if $G \cong V_4$.*
(ii) *$d = 3$, i.e., $g(x)$ is irreducible over $\mathbb{F}$ and Δ_f is a square in $\mathbb{F}$, if and only if $G \cong A_4$.*
(iii) *$d = 6$, i.e., $g(x)$ is irreducible over $\mathbb{F}$ and Δ_f is not a square in $\mathbb{F}$, if and only if $G \cong S_4$.*
(iv) *If $d = 2$, i.e., $g(x)$ has exactly one root $\beta_1 \in \mathbb{F}$, then:*
  a) $G \cong D_4 \iff f(x)$ is irreducible over $\mathbb{F}(\beta_1)$,
  b) $G \cong C_4 \iff f(x)$ is reducible over $\mathbb{F}(\beta_1)$.

PROOF. Since $f(x)$ is irreducible of degree 4, $G \subseteq S_4$ is transitive, and $|G| = 4, 8, 12$, or 24 Prop. 8, corresponding to V_4, C_4, D_4, A_4, or S_4. The resolvent $g(x)$ has roots $\beta_1, \beta_2, \beta_3$, and $\mathbb{F}(\beta_1, \beta_2, \beta_3)$ is its splitting field, with degree $d = 1, 2, 3$, or 6. The subgroup $G \cap V_4$ is the Galois group of $E_f/\mathbb{F}(\beta_1, \beta_2, \beta_3)$.

If $d = 1$, then $\beta_1, \beta_2, \beta_3 \in \mathbb{F}$, so $\mathbb{F}(\beta_1, \beta_2, \beta_3) = \mathbb{F}$, and $G = G \cap V_4$. Since $|G| = 4$, $G \cong V_4$. Conversely, $G \cong V_4$ implies $G \subseteq V_4$, so $G \cap V_4 = G$, and $\beta_i \in \mathbb{F}$, giving $d = 1$.

If $g(x)$ is irreducible, then $d = 3$ or 6. For $d = 3$, $|G| = 12$, and the only transitive subgroup of S_4 of order 12 is A_4. Prop. 8 states that $G \subseteq A_4$ if and only if Δ_f is a square in $\mathbb{F}$, so $G \cong A_4$. For $d = 6$, $|G| = 24$, so $G \cong S_4$, with Δ_f not a square. Conversely, if $G \cong A_4$ or S_4, then $G \cap V_4 = \{\mathrm{id}\}$, and $d = 3$ or 6, respectively.

If $d = 2$, then $g(x)$ has one root $\beta_1 \in \mathbb{F}$, and $|G| = 8$ or 4. If $f(x)$ is irreducible over $\mathbb{F}(\beta_1)$, transitivity requires $G \cong D_4$ (order 8). If reducible (e.g., into quadratics), $G \cong C_4$ (order 4).

The invariants $\xi_0(f)$, $\xi_1(f)$ define the $\mathrm{SL}_2(\mathbb{C})$-orbit of $f(x,y)$, encoding the root configuration in $\mathbb{P}^1_{\mathbb{C}}$. The resolvent's factorization and Δ_f's square property refine this to uniquely determine G. □

6.3.1. *Solving quartics.* The element $(\alpha_1+\alpha_2)(\alpha_3+\alpha_4)$ is fixed by $G\cap V_4$, hence lies in $\mathbb{F}(\beta_1,\beta_2,\beta_3)$. We find

$$-(\alpha_1+\alpha_2)^2=(\alpha_1+\alpha_2)(\alpha_3+\alpha_4)=\beta_2+\beta_3 \tag{27}$$

By this and symmetry we get **Ferrari's formulas**

$$\begin{aligned}\alpha_1+\alpha_2&=\sqrt{-\beta_2-\beta_3}\\ \alpha_1+\alpha_3&=\sqrt{-\beta_1-\beta_3}\\ \alpha_1+\alpha_4&=\sqrt{-\beta_1-\beta_2}\end{aligned} \tag{28}$$

or

$$\begin{aligned}\alpha_1&=\frac{\sqrt{-\beta_1-\beta_2}+\sqrt{-\beta_1-\beta_3}+\sqrt{-\beta_2-\beta_3}}{2}\\ \alpha_2&=\frac{-\sqrt{-\beta_1-\beta_2}-\sqrt{-\beta_1-\beta_3}+\sqrt{-\beta_2-\beta_3}}{2}\\ \alpha_3&=\frac{-\sqrt{-\beta_1-\beta_2}+\sqrt{-\beta_1-\beta_3}-\sqrt{-\beta_2-\beta_3}}{2}\\ \alpha_4&=\frac{\sqrt{-\beta_1-\beta_2}-\sqrt{-\beta_1-\beta_3}-\sqrt{-\beta_2-\beta_3}}{2}\end{aligned} \tag{29}$$

This completes the case for the quartics.

6.4. Quintics. Now we are ready to handle quintics which has such a special case in the history of Galois theory.

LEMMA 18. *Let $f(x)\in\mathbb{F}[x]$ be an irreducible quintic. Then its Galois group is one of the following C_5, D_5, $F_5=AGL(1,5)$, A_5, S_5.*

PROOF. G is transitive, hence its 5-Sylow subgroup is isomorphic to C_5 (generated by a 5-cycle). If C_5 is not normal, then G has at least 6 of 5-Sylow subgroups; then $|G|\geq 6\cdot 5=30$, hence $[S_5:G]\leq 4$ which implies $G=S_5,A_5$. If C_5 is normal in G then G is conjugate either C_5, D_5 (dihedral group of order 10) or $F_5=AGL(1,5)$, the full normalizer of C_5 in S_5, of order 20 (called also the Frobenius group of order 20). □

REMARK 7. *If the discriminant of the quintic is a square in $\mathbb{F}$ then Gal (f) is contained in A_5. Hence, it is C_5, D_5, or A_5.*

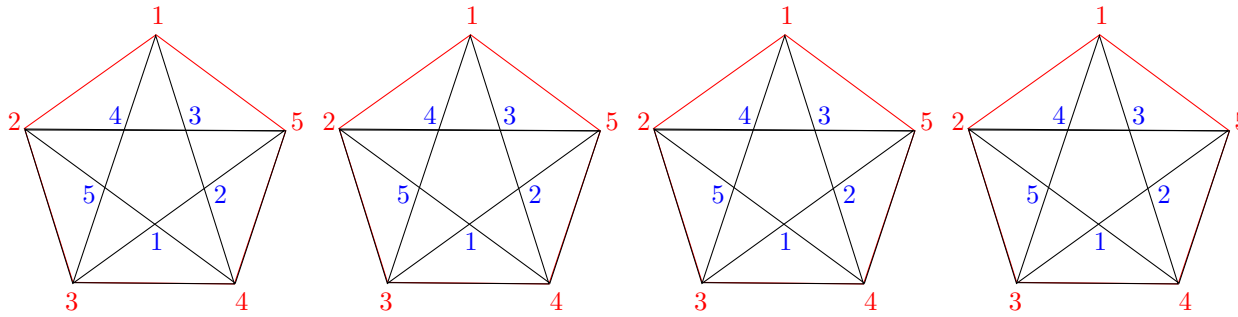
6.4.1. *Solvable quintics.* If $G=S_5,A_5$ then the equation $f(x)=0$ is not solvable by radicals. We want to investigate here the case G is not isomorphic to S_5 or A_5. Let $f(x)$ be an irreducible quintic in $\mathbb{F}[x]$ given by

$$f(x)=x^5+c_4x^4+\cdots+c_0=(x-\alpha_1)\cdots(x-\alpha_5) \tag{30}$$

Let $G=\mathrm{Gal}\,(f)$, viewed as a (transitive) subgroup of S_5 via permuting the (distinct) roots $\alpha_1,\cdots,\alpha_5$. As before $E_f=\mathbb{F}(\alpha_1,\cdots,\alpha_5)$ denotes the splitting field.

A 5-cycle in $S_5=\mathrm{Sym}(\{1,\ldots,5\})$ corresponds to an oriented pentagon with vertices $1,\ldots,5$. A 5-cycle and its inverse correspond to a (non-oriented) pentagon, and the full C_5 corresponds to a pentagon together with its "opposite".

(a) (b) (c) (d)



Thus F_5, the normalizer of C_5 in S_5, is the subgroup permuting the pentagon and its opposite. D_5 is the subgroup of F_5 fixing the pentagon (symmetry group of the pentagon), and C_5 is the subgroup of rotations. For example, F_5 is generated by

$$F_5 = \langle \sigma, \tau \mid \sigma^5 = \tau^4 = (\sigma\tau)^4 = \sigma\sigma\tau\sigma^{-1}\tau^{-1} \rangle, \tag{31}$$

where $\sigma = (12345)$ and $\tau = (2453)$. Thus if $G \leq F_5$ then G fixes

(32)

$$\delta_1 = (\alpha_1-\alpha_2)^2(\alpha_2-\alpha_3)^2(\alpha_3-\alpha_4)^2(\alpha_4-\alpha_5)^2(\alpha_5-\alpha_1)^2 - (\alpha_1-\alpha_3)^2(\alpha_3-\alpha_5)^2(\alpha_5-\alpha_2)^2(\alpha_2-\alpha_4)^2(\alpha_4-\alpha_1)^2$$

where the first (resp., second) term corresponds to the edges of the pentagon (resp., its opposite). There are six 5-Sylow subgroups of S_5 given by

$$H_1 = \langle (1,2,3,4,5) \rangle = \{(), (1,2,3,4,5), (1,3,5,2,4), (1,4,2,5,3), (1,5,4,3,2)\}$$
$$H_2 = \langle (1,2,3,5,4) \rangle = \{(), (1,2,3,5,4), (1,3,4,2,5), (1,5,2,4,3), (1,4,5,3,2)\}$$
$$H_3 = \langle (1,2,4,5,3) \rangle = \{(), (1,2,4,5,3), (1,4,3,2,5), (1,5,2,3,4), (1,3,5,4,2)\}$$
$$H_4 = \langle (1,2,4,3,5) \rangle = \{(), (1,2,4,3,5), (1,4,5,2,3), (1,3,2,5,4), (1,5,3,4,2)\}$$
$$H_5 = \langle (1,2,5,3,4) \rangle = \{(), (1,2,5,3,4), (1,5,4,2,3), (1,3,2,4,5), (1,4,3,5,2)\}$$
$$H_6 = \langle (1,3,4,5,2) \rangle = \{(), (1,3,4,5,2), (1,4,2,3,5), (1,5,2,4,3), (1,5,3,2,4)\}$$

To see the full invariance properties, we need to "projectivize" and use the invariants of binary forms; see section 5.3. Let $y = 1 = \beta_i$. The generalized version of the δ_1's is $\tilde{\delta}_1$, formed by replacing $\alpha_i - \alpha_j$ by $D_{ij} = \det \begin{bmatrix} \gamma_i & \beta_i \\ \gamma_j & \beta_j \end{bmatrix}$ in the formulas defining the δ_i's. In particular,

$$\tilde{\delta}_1 = D_{12}^2 D_{23}^2 D_{34}^2 D_{45}^2 D_{51}^2 - D_{13}^2 D_{35}^2 D_{52}^2 D_{24}^2 D_{41}^2 \tag{33}$$

Since S_5 has six 5-Sylow subgroups let $\delta_1, \ldots, \delta_6$ be the elements associated in this way to the six 5-Sylow's of S_5, i.e., to the six pentagon-opposite pentagon pairs on five given letters. We can write them all explicitly as

$$\begin{aligned}
\tilde{\delta}_2 &= D_{12}^2 D_{23}^2 D_{35}^2 D_{54}^2 D_{41}^2 - D_{13}^2 D_{34}^2 D_{42}^2 D_{25}^2 D_{51}^2 \\
\tilde{\delta}_3 &= D_{12}^2 D_{24}^2 D_{45}^2 D_{53}^2 D_{31}^2 - D_{14}^2 D_{43}^2 D_{32}^2 D_{25}^2 D_{51}^2 \\
\tilde{\delta}_4 &= D_{12}^2 D_{24}^2 D_{43}^2 D_{35}^2 D_{51}^2 - D_{14}^2 D_{45}^2 D_{52}^2 D_{23}^2 D_{31}^2 \\
\tilde{\delta}_5 &= D_{12}^2 D_{25}^2 D_{53}^2 D_{34}^2 D_{41}^2 - D_{15}^2 D_{54}^2 D_{42}^2 D_{23}^2 D_{31}^2 \\
\tilde{\delta}_6 &= D_{13}^2 D_{34}^2 D_{45}^2 D_{52}^2 D_{21}^2 - D_{14}^2 D_{42}^2 D_{23}^2 D_{35}^2 D_{51}^2
\end{aligned} \tag{34}$$

LEMMA 19. *$\delta_i^\sigma = \delta_i$ and $\delta_i^\tau = \delta_i$ for $i = 1, \dots, 6$.*

Clearly, G permutes $\delta_1, \dots, \delta_6$. If G is conjugate to a subgroup of F_5, it fixes one of $\delta_1, \dots, \delta_6$; this fixed δ_i must then lie in $\mathbb{F}$. Let δ_1 as in Eq. (32) and $\delta_2, \dots, \delta_6$ as follows:

(35)

$$\begin{aligned}
\delta_2 =&(\alpha_1-\alpha_2)^2(\alpha_2-\alpha_3)^2(\alpha_3-\alpha_5)^2(\alpha_5-\alpha_4)^2(\alpha_4-\alpha_1)^2-(\alpha_1-\alpha_3)^2(\alpha_3-\alpha_4)^2(\alpha_4-\alpha_2)^2(\alpha_2-\alpha_5)^2(\alpha_5-\alpha_1)^2\\
\delta_3 =&(\alpha_1-\alpha_2)^2(\alpha_2-\alpha_4)^2(\alpha_4-\alpha_5)^2(\alpha_5-\alpha_3)^2(\alpha_3-\alpha_1)^2-(\alpha_1-\alpha_4)^2(\alpha_4-\alpha_3)^2(\alpha_3-\alpha_2)^2(\alpha_2-\alpha_5)^2(\alpha_5-\alpha_1)^2\\
\delta_4 =&(\alpha_1-\alpha_2)^2(\alpha_2-\alpha_4)^2(\alpha_4-\alpha_3)^2(\alpha_3-\alpha_5)^2(\alpha_5-\alpha_1)^2-(\alpha_1-\alpha_4)^2(\alpha_4-\alpha_5)^2(\alpha_5-\alpha_2)^2(\alpha_2-\alpha_3)^2(\alpha_3-\alpha_1)^2\\
\delta_5 =&(\alpha_1-\alpha_2)^2(\alpha_2-\alpha_5)^2(\alpha_5-\alpha_3)^2(\alpha_3-\alpha_4)^2(\alpha_4-\alpha_1)^2-(\alpha_1-\alpha_5)^2(\alpha_5-\alpha_4)^2(\alpha_4-\alpha_2)^2(\alpha_2-\alpha_3)^2(\alpha_3-\alpha_1)^2\\
\delta_6 =&(\alpha_1-\alpha_3)^2(\alpha_3-\alpha_4)^2(\alpha_4-\alpha_5)^2(\alpha_5-\alpha_2)^2(\alpha_2-\alpha_1)^2-(\alpha_1-\alpha_4)^2(\alpha_4-\alpha_2)^2(\alpha_2-\alpha_3)^2(\alpha_3-\alpha_5)^2(\alpha_5-\alpha_1)^2
\end{aligned}$$

Thus, a necessary condition for the (irreducible) polynomial $f(x)$ to be solvable by radicals is that one δ_i lies in $\mathbb{F}$, i.e., that the polynomial

$$g(x) = (x-\delta_1)\cdots(x-\delta_6) \in \mathbb{F}[x] \tag{36}$$

has a root in $\mathbb{F}$. It is also sufficient:

LEMMA 20. *If G fixes one δ_i then G is conjugate to a subgroup of F_5, provided that $\delta_1, \dots, \delta_6$ are all distinct.*

PROOF. To check this it is enough to show that $\delta_1, \dots, \delta_6$ are mutually distinct (under the hypothesis $\Delta_f \neq 0$). hence, we have to show that $\Delta_f \neq 0 \implies \Delta_g \neq 0$. Using computational algebra we find Δ_g and verify that

$$\Delta_g = ((\alpha_1-\alpha_2)(\alpha_3-\alpha_4)(\alpha_4-\alpha_5)(\alpha_3-\alpha_5))^4 \cdot \Delta_f \cdot I_2^2 \cdot I_3 \cdot I_4^2 \cdot I_6^2$$

where I_2, I_3, I_4, and I_6 are given in [8]. Obviously $\Delta_f \neq 0$ implies that $\alpha_i - \alpha_j \neq 0$ for each $i \neq j$. This completes the proof. □

The coefficients of $g(x)$ are symmetric functions in $\alpha_1, \dots, \alpha_5$, hence are polynomial expressions in $c_0, \dots, c_4$. The goal is to find these expressions explicitly. This gives an explicit criterion to check whether $f(x) = 0$ is solvable by radicals.

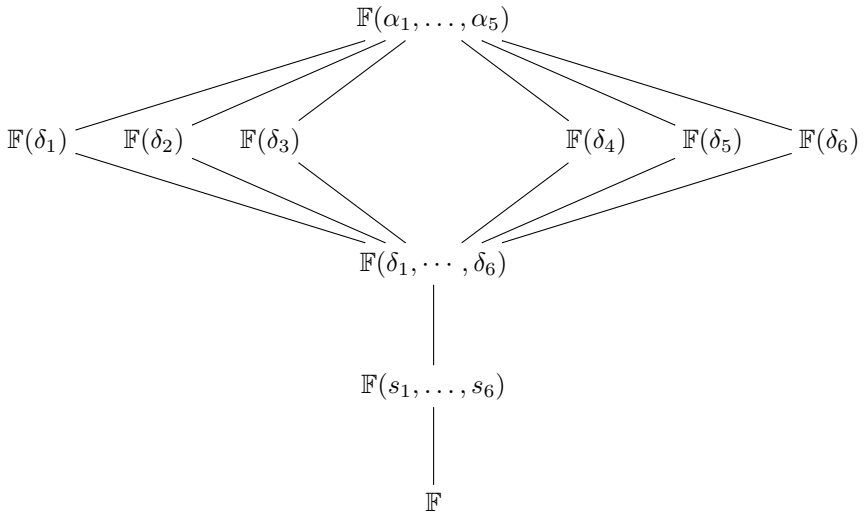
LEMMA 21. *Let $s_r(x_1, \dots, x_6)$, $r = 1, \dots, 6$, be the elementary symmetric polynomials*

$$s_r = \sum_{i_1 < i_2 < \cdots < i_r} x_{i_1} x_{i_2} \dots x_{i_r}. \tag{37}$$

Then $d_r := s_r(\tilde{\delta}_1, \dots, \tilde{\delta}_6)$ is a homogeneous polynomial expression in $b_0, \dots, b_5$ of degree $4r$. These polynomials are invariant under the action of $\mathrm{SL}_2(\mathbb{F})$ on binary quintics: For any $M \in \mathrm{SL}_2(\mathbb{F})$ the quintic f^M has the same associated d_r's.

PROOF. For $\alpha_j := \gamma_j/\beta_j$ we have $\tilde{\delta}_i = (\beta_1 \cdots \beta_5)^4 \delta_i = b_5^4 \delta_i$. Thus $d_r = b_5^{4r}\, s_r(\delta_1, , \dots, \delta_6)$. But the $s_r(\delta_1, , \dots, \delta_6)$ are polynomial expressions in the $c_j = b_j/b_5$, for $j = 0, \dots, 4$. Thus d_r is a rational function in $b_0, \dots, b_5$, where the denominator is a power of b_5. Switching the roles of x and y yields that the denominator is also a power of b_0. Thus it is constant, i.e., d_r is a polynomial in $b_0, \dots, b_5$. If we replace each β_j by $c\beta_j$ for a scalar λ then each $\tilde{\delta}_i$ gets multiplied by λ^4, so d_r gets multiplied by λ^{4r}. Thus d_r is homogeneous of degree $4r$. The rest of the claim is clear. □

(38)



There are four basic invariants of quintics, denoted by J_4, J_8, J_{12}, J_{18}, of degrees 4,8,12 and 18, such that every $\mathrm{SL}(2,\mathbb{F})$-invariant polynomial in $b_0,\ldots,b_5$ is a polynomial in J_4, J_8, J_{12}, J_{18}; see [25].

To define J_4, J_8, J_{12}, we need auxiliary quantities

$$A = \frac{1}{100}\left(20b_4 - 8b_1b_3 + 3b_2^2\right),\ B = \frac{1}{100}\left(100b_5 - 12b_1b_4 + 2b_2b_3\right),\ C = \frac{1}{100}\left(20b_1b_5 - 8b_2b_4 + 3b_3^2\right)$$

and D, E, F, G defined by

$$\begin{vmatrix} 10u + 2b_1v & 2b_1u + b_2v & b_2u + b_3v \\ 2b_1u + b_2v & b_2u + b_3v & b_3u + 2b_4v \\ b_2u + b_3v & b_3u + 2b_4v & 2b_4u + 10b_5v \end{vmatrix} = 10^3(Du^3 + Eu^2v + Fuv^2 + Gv^3)$$

Then J_2, J_8, and J_{12} are given by

$$\begin{aligned} J_4 &= 5^3(B^2 - 4AC) \\ J_8 &= 2^5 \cdot 5^6\left[2A(3EG - F^2) - B(9DG - EF) + 2C(3FD - E^2)\right] \\ J_{12} &= -2^{10} \cdot 5^9 \cdot 3^{-1}\left[4(3EG - F^2)(3FD - E^2) - (9DG - EF)^2\right] \end{aligned} \tag{39}$$

By using special quintics one gets linear equations for the coefficients expressing the d_r's in terms of J_4, J_8, J_{12}. The result is due to Berwick; see [14].

$$\begin{aligned} d_1 &= -10J_4 \\ d_2 &= 35J_4^2 + 10J_8 \\ d_3 &= -60J_4^3 - 30J_4J_8 - 10J_{12} \\ d_4 &= 55J_4^4 + 30J_4^2J_8 + 25J_8^2 + 50J_4J_{12} \\ d_5 &= -26J_4^5 - 10J_4^3J_8 - 44J_4J_8^2 - 59J_4^2J_{12} - 14J_8J_{12} \\ d_6 &= 5J_4^6 + 20J_4^2J_8^2 + 20J_4^3J_{12} + 20J_4J_8J_{12} + 25J_{12}^2 \end{aligned}$$

LEMMA 22. *Let $f(x)$ be a irreducible quintic over $\mathbb{F}$ and $d_1,\ldots,d_6$ defined in terms of the coefficients of $f(x)$ as above. Then $f(x)$ is solvable by radicals if and only if $g(x) = x^6 + d_1x^5 + \cdots d_5x + d_6$ has a root in $\mathbb{F}$.*

Extending the method of invariants becomes harder for higher degree equations. For degree six equations see [3] and [12].

7. Transitivity in S_n

Let S_n be the symmetric group on the set $\Omega = \{1, 2, \ldots, n\}$. A subgroup $G \leq S_n$ is *transitive* if, for any $i, j \in \Omega$, there exists $\sigma \in G$ such that $\sigma(i) = j$. Equivalently, the orbit of any point $i \in \Omega$ under the action of G is Ω.

A subgroup $G \leq S_n$ is *k-transitive* if it acts transitively on the set of ordered k-tuples of distinct elements in Ω. That is, for any two k-tuples $(i_1, \ldots, i_k)$ and $(j_1, \ldots, j_k)$ with $i_a \neq i_b$ and $j_a \neq j_b$ for $a \neq b$, there exists $\sigma \in G$ such that $\sigma(i_a) = j_a$ for all $a = 1, \ldots, k$. The group S_n is n-transitive, and the alternating group A_n is $(n-2)$-transitive for $n \geq 3$.

LEMMA 23 (Order of Transitive Subgroups). *Let $G \leq S_n$ be a transitive subgroup acting on $\Omega = \{1, 2, \ldots, n\}$. Then $|G|$ is divisible by n, and $|G| = n \cdot |\operatorname{Stab}_G(1)|$, where $\operatorname{Stab}_G(1) = \{\sigma \in G \mid \sigma(1) = 1\}$ is the stabilizer of 1.*

PROOF. Since G is transitive, the orbit of 1 under G is Ω, so $|\operatorname{Orbit}_G(1)| = n$. By the Orbit-Stabilizer Theorem, $|G| = |\operatorname{Orbit}_G(1)| \cdot |\operatorname{Stab}_G(1)|$. Thus, $|G| = n \cdot |\operatorname{Stab}_G(1)|$, and n divides $|G|$. □

LEMMA 24 (Classification of Transitive Subgroups). *A transitive subgroup $G \leq S_n$ is either:*

1. Primitive, *if the only G-invariant partitions of Ω are $\{\Omega\}$ and $\{\{1\}, \{2\}, \ldots, \{n\}\}$.*
2. Imprimitive, *if there exists a G-invariant partition of Ω into k blocks of size m, with $k \cdot m = n$, $k > 1$, and $m > 1$.*

PROOF. A partition of Ω is G-invariant if, for every block B in the partition and every $\sigma \in G$, the set $\sigma(B)$ is also a block. If G is transitive, the trivial partitions $\{\Omega\}$ and $\{\{1\}, \ldots, \{n\}\}$ are always G-invariant. If no other G-invariant partitions exist, G is primitive. Otherwise, suppose there exists a non-trivial G-invariant partition $\{B_1, \ldots, B_k\}$, where each block B_i has size m, and $k \cdot m = n$. Since G is transitive, it permutes the blocks transitively, and the action on each block is permutationally isomorphic to a subgroup of S_m. Thus, G is imprimitive, and its action is described by a wreath product structure, such as $H \wr S_k$, where $H \leq S_m$ acts on each block. □

PROPOSITION 9 (Transitive Subgroups of Prime Degree). *Let n be prime, and let $G \leq S_n$ be a transitive subgroup. Then G is one of:*

1. *The cyclic group C_n.*
2. *The dihedral group D_n.*
3. *A Frobenius group $n : k$, where k divides $n - 1$.*
4. *A subgroup of $AGL(1, n)$, the affine general linear group.*
5. *A group containing $PSL(k, q)$, where $n = (q^k - 1)/(q - 1)$.*

PROOF. Since n is prime, any transitive subgroup $G \leq S_n$ has order divisible by n, and G acts on $\Omega = \{1, \ldots, n\}$. Consider the normalizer of a Sylow n-subgroup of G. If G is solvable, Burnside's theorem implies that G is either a cyclic group C_n, a dihedral group D_n, a Frobenius group $n : k$ (where k divides $n - 1$), or a subgroup of $\mathrm{AGL}(1, n)$, which acts on the finite field $\mathbb{F}_n$. If G is not solvable, it must contain a non-abelian simple group. By the classification of finite simple

groups, the only non-solvable transitive groups of prime degree are those containing $\mathrm{PSL}(k,q)$, where $n=(q^k-1)/(q-1)$, acting on the projective space $\mathbb{P}^{k-1}(\mathbb{F}_q)$. For prime n, such groups arise only for specific k and q, such as $\mathrm{PSL}(2,n)$. □

7.1. Computational Enumeration of Transitive Subgroups. The transitive subgroups of S_n can be enumerated using computational group theory tools like GAP, which catalog these subgroups based on their permutation representations. Below, we provide the number of transitive subgroups for select $n \leq 47$ and list all transitive subgroups for $n = 5, 6, 7, 11, 13, 17, 19$.

Table 1 gives the number of transitive subgroups of S_n for select $n \leq 47$, computed using GAP's `TransitiveGroups` function.

n	# Subgroups	n	# Subgroups	n	# Subgroups	n	# Subgroups
5	5	6	16	7	7	8	50
9	34	10	45	11	8	12	301
13	9	14	63	15	104	16	1954
17	10	18	983	19	8	20	1117
21	164	22	59	23	7	24	25000
25	211	26	96	27	2392	28	1854
29	8	30	5712	31	12	33	162
34	115	35	407	36	121279	37	11
38	76	39	306	40	315842	41	10
42	9491	43	10	44	2113	45	10923

TABLE 1. Number of transitive subgroups of S_n for select $n \leq 47$.

Table 2 lists all transitive subgroups of S_n for $n = 5, 6, 7, 11, 13, 17, 19$, using standard group-theoretic notation and GAP identifiers where necessary.

TABLE 2. Transitive subgroups of S_n for $n = 5, 6, 7, 11, 13, 17, 19$.

n	Transitive Subgroups
5	C_5, D_5, $5:4$, A_5, S_5
6	C_6, D_6, $S_3 \wr C_2$, A_4, $C_3 \wr C_2$, $C_2 \wr C_3$, $S_4(6d)$, $S_4(6c)$, $(C_3 \wr C_2):2$, $(C_3 \times C_3):4$, $C_2 \wr S_3$, $\mathrm{PSL}(2,5)$, $\mathrm{PGL}(2,5)$, $S_3 \wr C_2$, A_6, S_6
7	C_7, D_7, $7:3$, $7:6$, $\mathrm{PSL}(3,2)$, A_7, S_7
11	C_{11}, D_{11}, $11:5$, $11:10$, $\mathrm{PSL}(2,11)$, M_{11}, A_{11}, S_{11}
13	C_{13}, D_{13}, $13:3$, $13:4$, $13:6$, $13:12$, $\mathrm{PSL}(3,3)$, A_{13}, S_{13}
17	C_{17}, D_{17}, $17:4$, $17:8$, $17:16$, $\mathrm{PSL}(2,16)$, $\mathrm{P\Sigma L}(2,16)$, $\mathrm{P\Gamma L}(2,16)$, A_{17}, S_{17}
19	C_{19}, D_{19}, $19:3$, $19:6$, $19:9$, $19:18$, A_{19}, S_{19}

8. Resolvents of Polynomials

Resolvents are polynomials constructed from the roots of a given polynomial to determine its Galois group, a transitive subgroup of the symmetric group S_n, as cataloged in Section 7. Following Cohen [7], we define resolvents rigorously, present their theoretical properties with complete proofs, provide computational examples, and conclude with a method to determine the Galois group of irreducible quintics.

Let $f(x) \in \mathbb{F}[x]$ be a monic, separable, irreducible polynomial of degree n over a field $\mathbb{F}$ (typically $\mathbb{F} = \mathbb{Q}$), with roots $\alpha_1, \ldots, \alpha_n$ in a splitting field K over $\mathbb{F}$. The Galois group Gal $(f) = \mathrm{Aut}(K/\mathbb{F})$ acts as a transitive subgroup of S_n by permuting the roots $\{\alpha_1, \ldots, \alpha_n\}$.

Let $G \leq S_n$ be a transitive subgroup containing Gal $(K/\mathbb{F})$, and let $H \leq G$ be a subgroup. In the polynomial ring $\mathbb{F}[\alpha_1, \ldots, \alpha_n]$, where G permutes the root indices, a resolvent polynomial is defined using an element $F \in \mathbb{F}[\alpha_1, \ldots, \alpha_n]$ invariant under H, i.e., $\sigma(F) = F$ for all $\sigma \in H$. The resolvent polynomial is:

$$R_{F,H}(x) = \prod_{\sigma \in T} (x - \sigma(F)),$$

where $T \subseteq G$ is a set of representatives for the left cosets of H in G. Since F is H-invariant, $\sigma(F)$ is independent of the representative: if $\sigma' = \sigma h$, $h \in H$, then $\sigma'(F) = \sigma(h(F)) = \sigma(F)$. The degree of $R_{F,H}(x)$ is $|G : H|$. For $G = S_n$, the degree is $n!/|H|$.

THEOREM 8.1. *The resolvent polynomial $R_{F,H}(x) = \prod_{\sigma \in T}(x - \sigma(F))$ has coefficients in $\mathbb{F}$.*

PROOF. The coefficients of $R_{F,H}(x)$ are elementary symmetric polynomials in $\{\sigma(F) \mid \sigma \in T\}$. For $\gamma \in$ Gal $(K/\mathbb{F}) \subseteq G$:

$$\gamma(\sigma(F)) = (\gamma\sigma)(F).$$

Since $\gamma\sigma \in G$, write $\gamma\sigma = \sigma' h$, $\sigma' \in T$, $h \in H$. Then:

$$(\gamma\sigma)(F) = \sigma'(h(F)) = \sigma'(F).$$

Thus, γ permutes the roots, so the coefficients are invariant under Gal $(K/\mathbb{F})$. By Galois theory, they lie in $\mathbb{F}$, so $R_{F,H}(x) \in \mathbb{F}[x]$. □

THEOREM 8.2. *The resolvent polynomial $R_{F,H}(x)$ factors over $\mathbb{F}$ into irreducible factors, with the number of factors equal to the number of double cosets Gal $(K/\mathbb{F})\backslash G/H$. Each factor's degree is the size of the corresponding double coset.*

PROOF. The roots are $\{\sigma(F) \mid \sigma \in T\}$. Assume distinct $\sigma(F)$ for distinct cosets. For $\gamma \in$ Gal $(K/\mathbb{F})$:

$$\gamma(\sigma(F)) = (\gamma\sigma)(F) = \sigma'(F),$$

where $\gamma\sigma = \sigma' h$. Thus, Gal $(K/\mathbb{F})$ acts on G/H via $\gamma \cdot (\sigma H) = (\gamma\sigma)H$. The irreducible factors correspond to orbits, which are double cosets Gal $(K/\mathbb{F})\backslash G/H$. For a double coset $D =$ Gal $(K/\mathbb{F})\sigma H$, the factor is:

$$\prod_{\tau H \in D} (x - \tau(F)),$$

irreducible over $\mathbb{F}$ as the roots are conjugate. The degree is $|D|$, the number of cosets in D. The minimal polynomial of $\sigma(F)$ has degree equal to the orbit size, so $R_{F,H}(x)$ has as many factors as double cosets, each of degree equal to the double coset size. □

PROPOSITION 10. *If Gal $(K/\mathbb{F}) \subseteq H$, then $R_{F,H}(x)$ has a linear factor over $\mathbb{F}$.*

PROOF. If Gal $(K/\mathbb{F}) \subseteq H$, then for $F = e(F)$, and any $\gamma \in$ Gal $(K/\mathbb{F})$, $\gamma(F) = F$, since H fixes F. Thus, $F \in \mathbb{F}$, and $x - F$ is a linear factor. □

LEMMA 25. *If $R_{F,H}(x)$ has a linear factor over $\mathbb{F}$, then there exists $\sigma \in G$ such that Gal $(K/\mathbb{F}) \subseteq \sigma H \sigma^{-1}$.*

PROOF. If $\sigma(F) \in \mathbb{F}$, then for $\gamma \in \text{Gal } (K/\mathbb{F})$:

$$\gamma(\sigma(F)) = (\gamma\sigma)(F) = \sigma(F).$$

Thus, $\sigma^{-1}\gamma\sigma \in \text{Stab}_G(F) = H$, assuming $\text{Stab}_G(F) = H$. Hence, $\gamma \in \sigma H \sigma^{-1}$, so Gal $(K/\mathbb{F}) \subseteq \sigma H \sigma^{-1}$. □

THEOREM 8.3. *For $H = \text{Stab}_G(B)$, $B \subseteq \{1, \ldots, n\}$ of size k, and $F = \sum_{i \in B} \alpha_i$, if Gal $(K/\mathbb{F}) \subseteq H$, then $R_{F,H}(x)$ has a linear factor. If $R_{F,H}(x)$ is irreducible, then Gal $(K/\mathbb{F}) \not\subseteq \sigma H \sigma^{-1}$ for any $\sigma \in G$.*

PROOF. If Gal $(K/\mathbb{F}) \subseteq H$, then F is fixed by Gal $(K/\mathbb{F})$, so $F \in \mathbb{F}$, and $x - F$ is a linear factor. If $R_{F,H}(x)$ is irreducible, there is one double coset, implying a transitive action on G/H. If Gal $(K/\mathbb{F}) \subseteq \sigma H \sigma^{-1}$, the action fixes σH, contradicting transitivity unless $H = G$. Thus, Gal $(K/\mathbb{F}) \not\subseteq \sigma H \sigma^{-1}$. □

PROPOSITION 11. *Let $H \leq G$ be a maximal subgroup. If $R_{F,H}(x)$ has a linear factor, then either Gal $(K/\mathbb{F}) \subseteq \sigma H \sigma^{-1}$ for some $\sigma \in G$, or Gal $(K/\mathbb{F}) = G$.*

PROOF. By Lemma 25, a linear factor implies Gal $(K/\mathbb{F}) \subseteq \sigma H \sigma^{-1}$. Since H is maximal, if Gal $(K/\mathbb{F}) \not\subseteq \sigma H \sigma^{-1}$, then Gal $(K/\mathbb{F})$ and $\sigma H \sigma^{-1}$ generate a group containing H as a proper subgroup. Maximality implies this group is G, so Gal $(K/\mathbb{F}) = G$. □

THEOREM 8.4. *The discriminant of $R_{F,H}(x)$ is related to the discriminant of $f(x)$ by:*

$$\mathit{Disc}(R_{F,H}) = \mathit{Disc}(f)^{|G:H|} \cdot \prod_{\sigma \neq \tau \in T} (\sigma(F) - \tau(F))^2.$$

PROOF. The discriminant of $R_{F,H}(x) = \prod_{\sigma \in T}(x - \sigma(F))$ is:

$$\text{Disc}(R_{F,H}) = \prod_{\sigma \neq \tau \in T} (\sigma(F) - \tau(F))^2.$$

Express $\sigma(F)$ in terms of the roots α_i. For generic F, the differences $\sigma(F) - \tau(F)$ involve linear combinations of the α_i. The discriminant of $f(x) = \prod_{i=1}^{n}(x - \alpha_i)$ is:

$$\text{Disc}(f) = \prod_{i<j} (\alpha_i - \alpha_j)^2.$$

The product $\prod_{\sigma \neq \tau}(\sigma(F) - \tau(F))$ accounts for differences across cosets, scaled by the structure of G/H. By Galois theory, the squared differences aggregate to a power of $\text{Disc}(f)$, adjusted by the degree $|G : H|$. Thus:

$$\text{Disc}(R_{F,H}) = \text{Disc}(f)^{|G:H|} \cdot \prod_{\sigma \neq \tau} (\sigma(F) - \tau(F))^2,$$

where the additional product depends on the specific F. □

8.1. Choosing resolvents. The choice of F and H is critical for constructing resolvents that probe the structure of Gal $(K/\mathbb{F})$. The element F must be invariant under H, i.e., $\sigma(F) = F$ for all $\sigma \in H$, to ensure $\sigma(F)$ is well-defined for each coset σH. Ideally, the stabilizer $\text{Stab}_G(F) = \{\sigma \in G \mid \sigma(F) = F\}$ equals H, so F distinguishes cosets effectively, producing distinct $\sigma(F)$ values that reflect the double coset structure Gal $(K/\mathbb{F})\backslash G/H$ (Theorem 8.2). The subgroup H is chosen to test specific properties of Gal $(K/\mathbb{F})$, such as transitivity, imprimitivity, or cyclicity, corresponding to the transitive subgroups in Section 7. The following resolvent types are commonly used due to their ability to target these properties while remaining computationally tractable:

- **Linear resolvents**: $F = a_1\alpha_1 + \cdots + a_n\alpha_n$, with distinct $a_i \in \mathbb{F}$ (e.g., $a_i = i$), and $H = \text{Stab}_G(\{1\})$, the point stabilizer. These test the action on individual roots, producing high-degree resolvents (e.g., $n!/|H|$) that probe the full permutation structure, useful for ruling out subgroups with fixed points (e.g., S_{n-1}).
- **Block resolvents**: $F = \sum_{i\in B} \alpha_i$, where $B \subseteq \{1, \ldots, n\}$ is a block, and $H = \text{Stab}_G(B)$, the setwise stabilizer. These test imprimitive actions, where Gal $(K/\mathbb{F})$ preserves a partition (e.g., for dihedral groups like D_5), yielding lower-degree resolvents that are efficient for detecting wreath product structures.
- **Quadratic resolvents**: $F = \alpha_i + \alpha_j$, with $H = \text{Stab}_G(\{i, j\})$. These focus on pairs of roots, testing for 2-transitivity or dihedral subgroups (e.g., D_5), and produce resolvents of moderate degree, balancing computational cost and specificity.
- **Lagrange resolvents**: $F = \sum_{i=1}^{n} \omega^{i-1}\alpha_i$, where $\omega \in \mathbb{F}$ is a primitive n-th root of unity (extending $\mathbb{F}$ if needed), and $H = \langle(1\,2\,\cdots\,n)\rangle \cong C_n$. These target cyclic subgroups (e.g., C_5), historically used to test solvability by radicals, though their high degree requires careful computation.

8.2. Computation of Resolvents. Given G, F, H, and $f(x)$ one can ask if we can compute the corresponding resolvent. Here is the main algorithm how to compute resolvents.

Algorithm 2: Compute Resolvent Polynomial

Input: Polynomial $f(x) \in \mathbb{F}[x]$, degree n, roots $\alpha_1, \ldots, \alpha_n$; group $G \leq S_n$; subgroup $H \leq G$; H-invariant $F \in \mathbb{F}[\alpha_1, \ldots, \alpha_n]$

Output: Resolvent polynomial $R_{F,H}(x)$

```
Compute a set T of left coset representatives of H in G;
for each σ ∈ T do
    Compute σ(F) in terms of α_1, ..., α_n;
    Express σ(F) using elementary symmetric polynomials of f(x);
end
Form R_{F,H}(x) = ∏_{σ∈T} (x − σ(F));
return R_{F,H}(x);
```

However, as expected, this algorithm can become very complex for high degree $f(x)$ and $F(x)$. Below, we describe the symbolic approach how to get the coefficients of the resolvent in terms of the coefficients of $f(x)$.

8.2.1. *Computing resolvents symbolically.* To compute $\mathrm{Res}_G(f, F)(x)$, we express it in the form

$$\mathrm{Res}_G(f, F)(x) = x^k - e_1 x^{k-1} + e_2 x^{k-2} - \cdots + (-1)^k e_k,$$

where k is the number of distinct θ_σ, and e_j are the elementary symmetric polynomials in these values. The steps are as follows.

Step 1: Determine the Degree k The degree k of the resolvent is the size of the orbit of $\theta_e = F(r_1, \ldots, r_n)$ under G. Define the stabilizer subgroup

$$H = \{\sigma \in G \mid F(r_{\sigma(1)}, \ldots, r_{\sigma(n)}) = F(r_1, \ldots, r_n)\}.$$

Then, $k = |G|/|H|$, which counts the number of distinct θ_σ as σ ranges over G. Equivalently, k is the number of distinct evaluations of F over all permutations in G, accounting for symmetries in F. For instance, if F is symmetric under a subgroup of G, H is larger, reducing k.

Step 2: Identify the Roots of the Resolvent The roots of $\mathrm{Res}_G(f, F)(x)$ are the distinct values $\theta_\sigma = F(r_{\sigma(1)}, \ldots, r_{\sigma(n)})$. These are not computed numerically but treated symbolically as expressions in the roots r_i. The elementary symmetric sums of the roots of $f(x)$ are given by Vieta's formulas:

$$s_1 = r_1 + \cdots + r_n = -a_{n-1}, \quad s_2 = \sum_{i<j} r_i r_j = a_{n-2}, \quad \ldots, \quad s_n = r_1 \cdots r_n = (-1)^n a_0.$$

The θ_σ are functions of these roots, permuted by G, and our task is to express the coefficients e_j in terms of the s_i.

Step 3: Compute the Power Sums Define the power sums of the distinct θ_σ as

$$p_m = \sum_\sigma \theta_\sigma^m = \sum_\sigma [F(r_{\sigma(1)}, \ldots, r_{\sigma(n)})]^m,$$

where the sum is over representatives of the k distinct θ_σ. For each $m = 1, 2, \ldots, k$:

- Expand $F(r_{\sigma(1)}, \ldots, r_{\sigma(n)})^m$ as a polynomial in the r_i.
- Sum this expression over all $\sigma \in G$ (or distinct orbit elements).
- Express the result as a polynomial in $s_1, s_2, \ldots, s_n$ using symmetric polynomial identities.

For example, if $F = x_1$, then $\theta_\sigma = r_{\sigma(1)}$, and $p_m = \sum_\sigma r_{\sigma(1)}^m$, which simplifies based on G's action (e.g., for $G = S_n$, $p_m = n(r_1^m + \cdots + r_n^m)$).

Step 4: Relate Power Sums to Elementary Symmetric Sums The coefficients e_j are obtained from the p_m using Newton's identities, which for a set of k

roots are:

$$\begin{aligned} e_1 &= p_1, \\ 2e_2 &= e_1p_1 - p_2, \\ 3e_3 &= e_2p_1 - e_1p_2 + p_3, \\ &\vdots \\ je_j &= \sum_{i=1}^{j-1}(-1)^{i-1}e_{j-i}p_i + (-1)^{j-1}p_j, \quad (j \leq k). \end{aligned}$$

Solve recursively for $e_1, e_2, \ldots, e_k$. Each e_j is a polynomial in $a_0, \ldots, a_{n-1}$ since the p_m are.

Step 5: Construct the Resolvent Polynomial With $e_1, e_2, \ldots, e_k$ computed, the resolvent is

$$\mathrm{Res}_G(f, F)(x) = x^k - e_1x^{k-1} + e_2x^{k-2} - \cdots + (-1)^k e_k.$$

This polynomial has degree k, and its coefficients are fully symbolic in the coefficients of $f(x)$.

We illustrate with an example.

EXAMPLE 3. *Consider $n = 5$, $G = S_5$, $F = x_1 + x_2 + x_3$. Let*

$$f(x) = x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0.$$

Determine the resolvent in terms of coefficients $a_0, \ldots, a_4$.

Let us first determine k. The stabilizer is $H = S_3 \times S_2$, where S_3 permutes $\{1, 2, 3\}$ and S_2 permutes $\{4, 5\}$. Thus, $|H| = 3! \cdot 2! = 12$, and $k = \frac{|S_5|}{|H|} = \frac{120}{12} = 10 = \binom{5}{3}$. The roots are $r_{i_1} + r_{i_2} + r_{i_3}$ for $1 \leq i_1 < i_2 < i_3 \leq 5$, with symmetric sums:

$$s_1 = -a_4, \quad s_2 = a_3, \quad s_3 = -a_2, \quad s_4 = a_1, \quad s_5 = -a_0.$$

We compute power sums $p_m = \sum_{i<j<k}(r_i + r_j + r_k)^m$ *and have:*

$$
\begin{aligned}
p_1 &= 6s_1 = -6a_4,\\
p_2 &= 6(s_1^2 - 2s_2) + 6s_2 = 6a_4^2 - 6a_3,\\
p_3 &= 6(s_1^3 - 3s_1s_2 + 3s_3) + 9(s_1s_2 - 3s_3) + s_3 = -6a_4^3 + 9a_4a_3 + 8a_2,\\
p_4 &= 6a_4^4 - 24a_4^2a_3 + 6a_3^2 + 8a_4a_2 - 24a_1,\\
p_5 &= -6a_4^5 + 30a_4^3a_3 - 10a_4^2a_2 - 20a_4a_3^2 + 30a_4a_1 + 5a_2a_3 - 10a_0,\\
p_6 &= 6a_4^6 - 36a_4^4a_3 + 4a_4^3a_2 + 36a_4^2a_3^2 - 12a_4^2a_1 - 12a_4a_3a_2 + 2a_3^3 + 6a_2^2 + 24a_4a_0 - 12a_1a_3,\\
p_7 &= -6a_4^7 + 42a_4^5a_3 - 14a_4^4a_2 - 56a_4^3a_3^2 + 28a_4^3a_1 + 28a_4^2a_3a_2 - 14a_4a_3^3 + 14a_4a_2^2 - 42a_4^2a_0\\
&\quad - 14a_3^2a_2 + 14a_3a_1 + 14a_2a_0,\\
p_8 &= 6a_4^8 - 48a_4^6a_3 + 12a_4^5a_2 + 72a_4^4a_3^2 - 32a_4^4a_1 - 48a_4^3a_3a_2 + 8a_4^2a_3^3 + 24a_4^3a_0 - 24a_4^2a_3a_1\\
&\quad + 12a_4a_3^2a_2 - 8a_4a_2^2 - 8a_3^2a_1 + 8a_3a_2a_0 - 24a_4a_1a_2 + 24a_1^2,\\
p_9 &= -6a_4^9 + 54a_4^7a_3 - 18a_4^6a_2 - 90a_4^5a_3^2 + 36a_4^5a_1 + 72a_4^4a_3a_2 - 18a_4^3a_3^3 - 54a_4^4a_0 + 18a_4^3a_3a_1\\
&\quad - 18a_4^2a_3^2a_2 + 36a_4^3a_2a_0 + 18a_4^2a_1a_2 + 6a_4a_3^2a_1 - 6a_4a_2a_1 - 18a_3^2a_0 - 6a_3a_2^2 + 18a_3a_1a_0\\
&\quad + 6a_2a_1^2 - 18a_1a_2a_0,\\
p_{10} &= 6a_4^{10} - 60a_4^8a_3 + 20a_4^7a_2 + 108a_4^6a_3^2 - 40a_4^6a_1 - 96a_4^5a_3a_2 + 24a_4^4a_3^3 + 72a_4^5a_0 + 24a_4^4a_3a_1\\
&\quad - 24a_4^3a_3^2a_2 - 40a_4^4a_2a_0 + 8a_4^3a_3^2a_1 - 24a_4^3a_1a_2 + 12a_4^2a_3^3a_2 - 8a_4^2a_3a_1a_2 - 8a_4^2a_2^2a_1\\
&\quad + 24a_4^2a_0a_2 + 4a_4a_3^3a_1 - 4a_4a_3a_2^2 - 12a_4a_3a_0a_1 + 12a_4a_2a_1a_0 + 4a_3^3a_0 + 4a_3^2a_2a_1 - 12a_3^2a_0a_2\\
&\quad - 4a_3a_1^2a_0 + 4a_2^2a_0 - 12a_2a_1a_0^2 + 4a_0^3.
\end{aligned}
$$

Using Newton's identities we get

$$
\begin{aligned}
e_1 &= -6a_4,\ e_2 = 15a_4^2 + 3a_3,\ e_3 = -40a_4^3 + 18a_4a_3 + \frac{8}{3}a_2,\ e_4 = 90a_4^4 - 60a_4^2a_3 - 8a_4a_2 - 3a_3^2 + 6a_1,\\
e_5 &= -198a_4^5 + 165a_4^3a_3 + 15a_4^2a_2 - 45a_4a_3^2 - 18a_4a_1 - 3a_3a_2 + 2a_0,\\
e_6 &= 420a_4^6 - 420a_4^4a_3 - 20a_4^3a_2 + 108a_4^2a_3^2 + 24a_4^2a_1 + 12a_4a_3a_2 + 3a_3^3 - 2a_2^2 - 12a_4a_0 + 2a_1a_3,\\
e_7 &= -858a_4^7 + 1001a_4^5a_3 + 33a_4^4a_2 - 315a_4^3a_3^2 - 54a_4^3a_1 - 54a_4^2a_3a_2 - 9a_4a_3^3\\
&\quad + 6a_4a_2^2 + 18a_4^2a_0 + 3a_3^2a_2 - 3a_3a_1 - a_2a_0,\\
e_8 &= 1716a_4^8 - 2288a_4^6a_3 - 44a_4^5a_2 + 792a_4^4a_3^2 + 88a_4^4a_1 + 176a_4^3a_3a_2 + 24a_4^2a_3^3\\
&\quad - 16a_4^2a_3a_1 - 24a_4^3a_0 - 8a_4^2a_2^2 - 8a_4a_3^2a_2 + 8a_4a_1a_2 - a_3^2a_1 + a_3a_2^2 - a_1^2,\\
e_9 &= -3432a_4^9 + 5148a_4^7a_3 + 66a_4^6a_2 - 2002a_4^5a_3^2 - 176a_4^5a_1 - 528a_4^4a_3a_2 - 66a_4^3a_3^3\\
&\quad + 48a_4^4a_0 + 48a_4^3a_3a_1 + 24a_4^3a_2^2 + 24a_4^2a_3^2a_2 - 24a_4^2a_1a_2 + 2a_4a_3^2a_1 - 2a_4a_2a_1\\
&\quad + 2a_3^2a_0 + 2a_3a_2^2 - 2a_3a_1a_0 - 2a_2a_1^2 + 2a_1a_2a_0,\\
e_{10} &= 6864a_4^{10} - 11440a_4^8a_3 - 88a_4^7a_2 + 5148a_4^6a_3^2 + 352a_4^6a_1 + 1408a_4^5a_3a_2 + 176a_4^4a_3^3\\
&\quad - 112a_4^5a_0 - 112a_4^4a_3a_1 - 80a_4^4a_2a_0 - 48a_4^3a_3^2a_2 + 48a_4^3a_1a_2 - 8a_4^2a_3^3a_2 + 8a_4^2a_3a_1a_2\\
&\quad + 8a_4^2a_2^2a_1 - 8a_4^2a_0a_2 - a_4a_3^3a_1 + a_4a_3a_2^2 + a_3^3a_0 - a_3a_1^2a_0 - a_2^2a_0 + a_1^2a_0 + a_0^2.
\end{aligned}
$$

Then, $Res_{S_5}(f, F)(x) = x^{10} + e_1x^9 + \cdots + e_9x + e_{10}$.

8.2.2. *Computing Resolvents Numerically.* In contrast to the symbolic method outlined in the previous section, Cohen [7] describes a numerical approach to compute the resolvent polynomial $\mathrm{Res}_G(f,F)(x)$ (see Section 6.3 of [7]). This method is particularly useful when exact symbolic computation becomes impractical due to high degree or complexity, and it exploits the fact that, for a polynomial $f(x) \in \mathbb{Z}[x]$ and a suitable F, the resolvent's coefficients are integers. The algorithm approximates the roots of $f(x)$, computes the numerical values of the resolvent's roots, constructs the polynomial, and rounds the coefficients to the nearest integers. Below, we detail this process.

Consider $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 \in \mathbb{Z}[x]$ with roots $r_1, r_2, \ldots, r_n$ in $\mathbb{C}$, a subgroup $G \subseteq S_n$, and a function $F(x_1, \ldots, x_n) \in \mathbb{Z}[x_1, \ldots, x_n]$. The resolvent is

$$\mathrm{Res}_G(f,F)(x) = \prod_{\sigma \in G/H} (x - \theta_\sigma), \quad \theta_\sigma = F(r_{\sigma(1)}, \ldots, r_{\sigma(n)}),$$

where H is the stabilizer of F in G, and $k = |G|/|H|$.

The steps are as follows:

(1) **Approximate the Roots of $f(x)$:** Numerically compute the roots $r_1, r_2, \ldots, r_n$ of $f(x)$ to high precision using a root-finding algorithm (e.g., Newton-Raphson or a polynomial solver like Laguerre's method). For a polynomial over $\mathbb{Z}$, these roots may be real or complex, and precision (e.g., 50-100 decimal places) is chosen to ensure accuracy in subsequent steps. Cohen suggests using a numerical library or software capable of handling complex roots.

(2) **Compute the Resolvent's Roots:** For each coset representative $\sigma \in G/H$, evaluate $\theta_\sigma = F(r_{\sigma(1)}, \ldots, r_{\sigma(n)})$ numerically. Since F is a polynomial with integer coefficients and G permutes the roots, each θ_σ is a complex number. The set $\{\theta_\sigma\}$ contains k distinct values (assuming no unexpected numerical degeneracy), corresponding to the orbit of $F(r_1, \ldots, r_n)$ under G.

(3) **Construct the Polynomial Numerically:** Form the resolvent polynomial

$$\mathrm{Res}_G(f,F)(x) = \prod_{\sigma \in G/H} (x - \theta_\sigma) = x^k - e_1x^{k-1} + e_2x^{k-2} - \cdots + (-1)^k e_k,$$

where e_j are the elementary symmetric sums of the θ_σ. Numerically, compute these coefficients by expanding the product. For small k, this can be done directly; for larger k, use the power sums $p_m = \sum_\sigma \theta_\sigma^m$ and Newton's identities:

$$\begin{aligned} e_1 &= p_1, \\ e_2 &= \frac{p_1e_1 - p_2}{2}, \\ e_3 &= \frac{p_1e_2 - p_2e_1 + p_3}{3}, \\ &\vdots \\ e_k &= \frac{1}{k}\left[\sum_{i=1}^{k-1} (-1)^{i-1} e_{k-i}p_i + (-1)^{k-1}p_k\right]. \end{aligned}$$

Calculate p_m by summing θ_σ^m over all distinct σ.

(4) **Round Coefficients to Integers:** Since $f(x)$ and F have integer coefficients, and G is a permutation group, $\mathrm{Res}_G(f, F)(x) \in \mathbb{Z}[x]$ (assuming $K = \mathbb{Q}$). The numerically computed e_j will be approximate (e.g., 2.9998 or -0.0001), so round each to the nearest integer (e.g., 3 or 0). High precision in root approximation ensures rounding errors are minimal.

(5) **Verification:** Check the resulting polynomial by evaluating it at a few points (e.g., $x = 0, 1$) against the expected integer values, or confirm that its roots (recomputed numerically) match the θ_σ within tolerance.

This method is efficient for large n or complex G, where symbolic computation is infeasible, but requires careful precision management [7].

8.3. Resolvents of quintics. For example, consider $f(x) = x^5 - x - 1$, with $G = S_5$, $H = \mathrm{Stab}_{S_5}(\{1, 2\})$, $F = \alpha_1 + \alpha_2$. The resolvent $R_{F,H}(x)$ has degree $|S_5 : H| = 120/20 = 6$. Compute $F = \alpha_1 + \alpha_2$, express it via symmetric polynomials, and evaluate $\sigma(F)$ for coset representatives. Numerical factorization over $\mathbb{Q}$ (using high precision to avoid errors, as discussed below) yields factors of degrees 2 and 4, indicating multiple double cosets, ruling out $\mathrm{Gal}\,(f) = D_5$. Further tests confirm $\mathrm{Gal}\,(f) = S_5$.

For a Lagrange resolvent, let $H = \langle(1\,2\,\cdots\,5)\rangle$, $F = \sum_{i=1}^{5} \omega^{i-1}\alpha_i$, $\omega = e^{2\pi i/5}$. The resolvent has degree $120/5 = 24$. Compute F, evaluate over cosets, and factor. If irreducible, $\mathrm{Gal}\,(f) \not\subseteq C_5$, as for $x^5 - x - 1$.

Table 3 shows factorization patterns for a block resolvent $F = \alpha_1 + \alpha_2$, $H = \mathrm{Stab}_{S_5}(\{1, 2\})$, for S_5's transitive subgroups:

Group	Factorization Pattern (Degrees)
S_5	(2, 4) or (1, 1, 4)
A_5	(2, 4) or (1, 1, 4)
D_5	(1, 5)
F_5	(6)
C_5	(6)

TABLE 3. Factorization patterns of $R_{F,H}(x)$ for $H = \mathrm{Stab}_{S_5}(\{1, 2\})$ in S_5.

For $n = 5$, test the transitive subgroups C_5, D_5, F_5, A_5, S_5 from Section 7:

- $H = \langle(1\,2\,\cdots\,5)\rangle$, $F = \sum_{i=1}^{5} \omega^{i-1}\alpha_i$: Linear factor implies $\mathrm{Gal}\,(f) = C_5$.
- $H = \mathrm{Stab}_{S_5}(\{1, 2\})$, $F = \alpha_1 + \alpha_2$: Linear factor suggests $\mathrm{Gal}\,(f) = D_5$.
- $H = \mathrm{PSL}(2, 5)$, $F = \sum_{i=1}^{5} a_i\alpha_i$: Linear factor indicates $\mathrm{Gal}\,(f) = F_5$.

THEOREM 8.5. *For an irreducible quintic $f(x) \in \mathbb{Q}[x]$, the Galois group $\mathrm{Gal}\,(f)$ is determined as follows:*

- *Compute the discriminant $\Delta = \mathrm{Disc}(f)$. If Δ is not a square, $\mathrm{Gal}\,(f) = S_5$. If square, proceed.*
- *Test $H = \langle(1\,2\,\cdots\,5)\rangle$, $F = \sum_{i=1}^{5} \omega^{i-1}\alpha_i$. If $R_{F,H}(x)$ has a linear factor, $\mathrm{Gal}\,(f) = C_5$.*
- *Test $H = \mathrm{Stab}_{S_5}(\{1, 2\})$, $F = \alpha_1 + \alpha_2$. If $R_{F,H}(x)$ has a linear factor, $\mathrm{Gal}\,(f) = D_5$.*

- *Test* $H = PSL(2,5)$, $F = \sum_{i=1}^{5} a_i\alpha_i$. *If* $R_{F,H}(x)$ *has a linear factor,* $Gal\ (f) = F_5$.
- *If no linear factors,* $Gal\ (f) = A_5$.

PROOF. The transitive subgroups of S_5 are C_5, D_5, F_5, A_5, S_5. If Δ is not a square, Gal $(f) \not\subseteq A_5$, so Gal $(f) = S_5$. If square, test subgroups using resolvents. By Proposition 10, linear factors detect containment in C_5, D_5, F_5. Proposition 11 ensures that if no linear factors appear for maximal subgroups (D_5, F_5, A_5), then Gal $(f) = A_5$. Theorem 8.3 confirms that irreducibility rules out conjugate subgroups. □

Resolvents systematically identify Galois groups by leveraging factorization patterns and subgroup containment, as shown in Table 3 and Theorem 8.5.

9. Reduction modulo p

The reduction method uses the fact that every polynomial with rational coefficients can be transformed into a monic polynomial with integer coefficients without changing the splitting field. Let $f(x) \in \mathbb{Q}[x]$ be given by

$$f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 \tag{40}$$

Let d be the common denominator of all coefficients $a_0, \cdots, a_{n-1}$. Then $g(x) := d \cdot f(\frac{x}{d})$ is a monic polynomial with integer coefficients. Clearly the splitting field of $f(x)$ is the same as the splitting field of $g(x)$. Thus, without loss of generality we can assume that $f(x) \in \mathbb{Z}[x]$ is a monic polynomial with integer coefficients.

THEOREM 9.1. ***(Dedekind)*** *Let* $f(x) \in \mathbb{Z}[x]$ *be a monic polynomial such that* $\deg f = n$, $Gal_{\mathbb{Q}}(f) = G$, *and* p *a prime such that* $p \nmid \Delta_f$. *If* $f_p := f(x) \mod p$ *factors in* $\mathbb{Z}_p[x]$ *as a product of irreducible factors of degree* $n_1, n_2, n_3, \cdots, n_k$, *then* G *contains a permutation of type* $(n_1)\,(n_2)\cdots(n_k)$

PROOF. van der Warden section 8.10 □

The Dedekind theorem can be used to determine the Galois group in many cases since the *type* of permutation in S_n determines the conjugacy class in S_n. Consider for example polynomials of degree 5. The cycle types for all groups that occur as Galois groups of quintics are given below.

	(2)	$(2)^2$	(3)	(4)	(3)(2)	(5)
S_5	10	15	20	30	20	24
A_5		15	20			24
F_5		5		10		4
D_5		5				4
C_5						4

TABLE 4. Cycle types for Galois groups of quintics

In Tab. 5 we display the table for the type of elements in S_6. As it can be seen from the tables this method works well for degree 5 and 6. Unfortunately it does not always work for degree $d > 6$.

The main question here is how quickly can we find the primes which determine the signature of the group and hopefully the uniquely determine the group.

	()	(2)	(2)(2)	(2)(2)(2)	(3)	(3)(2)	(3)(3)	(4)	(4)(2)	(5)	(6)	$\lvert G\rvert$
S_6	1	15	45	15	40	120	40	90	90	144	120	720
A_6	1	-	45	-	40	-	40	-	90	144	-	360
S_5	1	-	15	10	-	-	20	30	-	24	20	120
$(S_3 \times S_3) \rtimes C_2$	1	6	9	6	4	12	4	-	18	-	12	72
A_5	1	-	15	-	-	-	20	-	-	24	-	60
$C_2 \times S_4$	1	3	9	7	-	-	8	6	6	-	8	48
$(C_3 \times C_3) \rtimes C_4$	1	-	9	-	4	-	4	-	18	-	-	36
$S_3 \times S_3$	1	-	9	6	4	-	4	-	-	-	12	36
S_4	1	-	3	6	-	-	8	6	-	-	-	24
S_4	1	-	9	-	-	-	8	-	6	-	-	24
$C_2 \times A_4$	1	3	3	1	-	-	8	-	-	-	8	24
$C_3 \times S_3$	1	-	-	3	4	-	4	-	-	-	6	18
A_4	1	-	3	-	-	-	8	-	-	-	-	12
D_{12}	1	-	3	4	-	-	2	-	-	-	2	12
S_3	1	-	-	3	-	-	2	-	-	-	-	6
C_6	1	-	-	1	-	-	2	-	-	-	2	6

TABLE 5. Cycle types for Galois groups of sextics

THEOREM 9.2 (Chebotarev Density Theorem). *Let L/K be a Galois extension and C a conjugacy class of $G = Gal\,(L/K)$. Then*

$$\{\mathfrak{p} \mid \mathfrak{p} \text{ is a prime of } K, \mathfrak{p} \nmid \Delta_{L/K}, \sigma_{\mathfrak{p}} \in C\}$$

has density $\#C/\$G$. In particular, there always exist such primes.

10. Databases of irreducible polynomials

10.1. Datasets of irreducible polynomials. In this section we want to create a database of irreducible polynomials $f \in \mathbb{Z}[x]$ of degree $\deg f = n$. Data will be stored in a Python dictionary. A polynomial $f(x) = \sum_{i=0}^{n} a_i x^i$ will be represented by its corresponding binary form $f(x, y) = \sum_{i=0}^{n} a_i x^i y^{n-i}$. Hence our points will be points in the projective space $\mathbb{P}^n_{\mathbb{Q}}$, i.e. points with integer coordinates

$$\mathfrak{p} = [a_n : \cdots : a_0] \in \mathbb{P}^n_{\mathbb{Q}},$$

such that $\gcd(a_0, \ldots, a_n)$. Since $f(x)$ is irreducible over $\mathbb{Q}$ and of degree $\deg f = n$, then $a_n \neq 0$ and $a_0 \neq 0$. Moreover, $\Delta_f \neq 0$.

10.2. Datasets with bounded height. Let us now trying to generate a dataset with a bounded height h as defined in Eq. (5). We will denote the set of such polynomials by $\mathcal{P}^h_n$. In other words

$$\mathcal{P}^h_n := \left\{[a_n : \cdots : a_0] \in \mathbb{P}^n_{\mathbb{Q}} \;\middle|\; a_0 a_n \neq 0, \Delta_f \neq 0, H_{\mathbb{Q}}([a_n : \cdots : a_0]) \leq h\right\}$$

where H_Q is defined as in Eq. (5).

To ensure that the points in the database are not repeated we key the dictionary by the tuples $(a_0, \ldots, a_n)$. A dictionary in Python does not allow key duplicates, which ensures that there are no duplicates in our data. For given h, n the cardinality of $\mathcal{P}^h_n$ is bounded by

$$\#\mathcal{P}^h_n \leq 4h^2(2h+1)^{n-2}$$

The proof is a straightforward counting argument. There are more sophisticated methods to count algebraic points of bounded height on projective spaces; see for

example [11] but we will work only over $\mathbb{Q}$ and our heights will be relatively small which does not allow for much redundant data.

For a degree $d \geq 3$ and height h one can use *Sagemath* and count such points as follows:

```
PP = ProjectiveSpace(d, QQ)
rational_points = PP.rational_points(h)
```

We then *normalize* the data by clearing denominators. Hence, all our data has integer coordinates. Furthermore, we keep only those polynomials which are irreducible over $\mathbb{Q}$. For every point $\mathfrak{p} = [a_n : \cdots : a_0]$ we will compute the following attributes

$$(a_0, \ldots, a_n) : [H(f), [\xi_0, \ldots, \xi_n, \Delta_f], \mathfrak{H}_{\rm k}(\mathfrak{p}), \text{sig}, \text{Gal}\ _{\mathbb{Q}}(f),]$$

where

$H(f)$	Height of $f(x)$defined in *Eq.* (5)
$[\xi_0, \ldots, \xi_n]$	Invariants defined in *section* 5.3
Δ_f	Discriminant of $f(x)$
$\mathfrak{H}_{\rm k}(\mathfrak{p})$	Weighted moduli height as in *Eq.* (16)
sig	Signature
Gal $_{\mathbb{Q}}(f)$	Gap Identity of the Galois group of $f(x)$

Some of the datasets differ for different degrees. For example for quartics, we also compute the invariants T and S as defined in Eq. (8) and the j-invariant. For sextics we compute absolute invariants t_1, t_2, t_3; see [28] for details. We give a slice of the corresponding dictionary for each $d = 3, 4, 5$ which we discuss in the rest of this paper and make all datasets available at [26].

10.3. Cubics. As a simple first exercise we start with irreducible cubics. We create a database of all rational points $[c_0 : c_1 : c_2 : c_3]$ in $\mathbb{P}^3$ with projective height $h \leq 20$ such that

$$f(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3$$

is an irreducible polynomial in $\mathbb{Q}[x]$. Since training a model for determining Gal (f) is trivial in this case we will focus mostly on comparing the naive height with the weighted moduli height and determining how the occurrence of A_3 happens with the increase of h.

A slice of five random elements of our Python dictionary looks like:

Key	Value
(-1, -9, -20, 1)	[20, 98, 3.1463462836, 'A3']
(20, -9, -20, 1)	[20, 1458632, 34.752530588, 'A3']
(8, 12, -20, 1)	[20, 540800, 13.5590472788, 'A3']
(1, 17, -20, 1)	[20, 243602, 22.2162222997, 'A3']
(19, -9, -19, 1)	[19, 1204352, 16.5637384397, 'A3']

where the 'key' has the coefficients of the cubic and the entries in 'values' are respectively: naive height, J_4 invariant, weighted heigh, and the Galois group.

LEMMA 26. *The total number of rational points of heights in (0 , 20] is = 1 299 200. From those there are 1 178 856 irreducible polynomials and only 1328 of them have Galois group C_3. Moreover, the distribution of polynomials with Galois group C_3 with respect to their naive height is given in Fig. 1.*

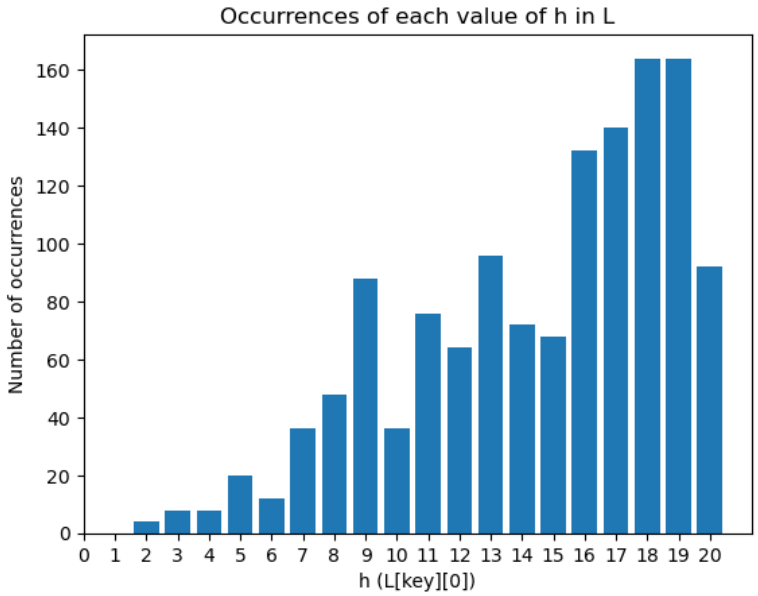


FIGURE 1. This distribution is only for cubics with Galois group C_3.

In [30] we give an estimate on the ratio of the moduli height over the naive height for binary sextics. Such bounds can be given for every degree d polynomial. In out case of cubics the minimum ratio is 0.074 for polynomial $f(x) = 7x^3 - 5x^2 - 16x + 7$ and the maximum ratio is 2.008 for $f(x) = 13x^3 - 19x^2 - 20x + 13$

LEMMA 27. *There are only 40 cubics in the database with height ≤ 5 and Galois group of order 3. The discriminant Δ_f of those forty polynomials has values $\Delta_f = 7^2, 3^4, 13^2, 19^2, 31^2$, and 61^2 as shown in the Tab. 6*

Below is the distribution of points in the database versus the invariant of cubics.

10.4. Quartics. We create a database of all rational points $[c_0 : c_1 : c_2 : c_3 : c_4]$ in $\mathbb{P}^3$ with projective height $h \leq 20$ such that

$$f(x) = c_0 + c_1x + c_2x^2 + c_3x^3 + c_4x^4$$

is an irreducible polynomial in $\mathbb{Q}[x]$. Other than S_4 the other possible Galois groups are C_4, D_4, V_4, and A_4 as explained in Eq. (19). We refer to Eq. (24) for its invariants. However, to avoid denominators we define

$$J_2 = 36 \cdot \xi_0, \qquad J_3 = 216 \cdot \xi_1, \qquad J_6 = \Delta(f, x)$$

TABLE 6. Irreducible degree 3 polynomials of height ≤ 5 and Galois group C_3

#	f	Δ	#	f	Δ	#	f	Δ
1	(1, 3, -4, 1)	7^2	15	(-1, -3, 0, 3)	3^4	29	(1, 2, -5, 1)	19^2
2	(-1, -4, -3, 1)	7^2	16	(1, -3, 0, 3)	3^4	30	(-1, -5, -2, 1)	19^2
3	(1, -1, -2, 1)	7^2	17	(5, 4, -5, 1)	13^2	31	(1, -5, 2, 1)	19^2
4	(1, -2, -1, 1)	7^2	18	(1, 1, -4, 1)	13^2	32	(-1, 2, 5, 1)	19^2
5	(-1, -2, 1, 1)	7^2	19	(5, -3, -2, 1)	13^2	33	(2, -1, -5, 2)	31^2
6	(-1, -1, 2, 1)	7^2	20	(-1, -4, -1, 1)	13^2	32	(2, -5, -1, 2)	31^2
7	(1, -4, 3, 1)	7^2	21	(1, -4, 1, 1)	13^2	35	(-2, -5, 1, 2)	31^2
8	(-1, 3, 4, 1)	7^2	22	(-5, -3, 2, 1)	13^2	36	(-2, -1, 5, 2)	31^2
9	(1, 0, -3, 1)	3^4	23	(-1, 1, 4, 1)	13^2	37	(3, -4, -5, 3)	61^2
10	(3, 0, -3, 1)	3^4	24	(-5, 4, 5, 1)	13^2	38	(3, -5, -4, 3)	61^2
11	(-1, -3, 0, 1)	3^4	25	(-1, -5, -4, 5)	13^2	39	(-3, -5, 4, 3)	61^2
12	(1, -3, 0, 1)	3^4	26	(1, -2, -3, 5)	13^2	40	(-3, -4, 5, 3)	61^2
13	(-3, 0, 3, 1)	3^4	27	(-1, -2, 3, 5)	13^2			
14	(-1, 0, 3, 1)	3^4	28	(1, -5, 4, 5)	13^2			

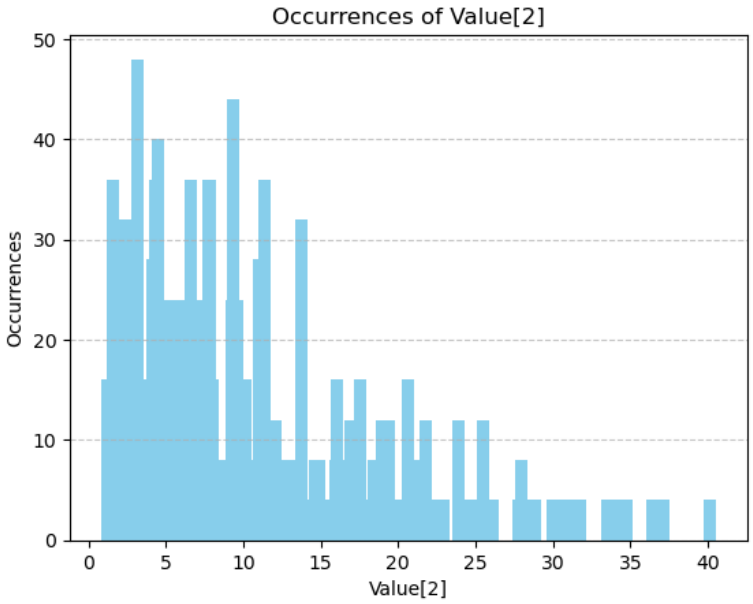


FIGURE 2. The number of occurrences versus the invariants

Hence, for a polynomial $f = [a_0, \ldots, a_4]$ we get

$$J_2 = 12c_0c_4 - 3c_1c_3 + c_2^2$$

$$J_3 = 72c_0c_2c_4 - 27c_0c_3^2 - 27c_1^2c_4 + 9c_1c_2c_3 - 2c_2^3$$

$$\begin{aligned} J_6 =& 256c_0^3c_4^3 - 192c_0^2c_1c_3c_4^2 - 128c_0^2c_2^2c_4^2 + 144c_0^2c_2c_3^2c_4 - 27c_0^2c_3^4 + 144c_0c_1^2c_2c_4^2 - 6c_0c_1^2c_3^2c_4 - 80c_0c_1c_2^2c_3c_4 \\ &+ 18c_0c_1c_2c_3^3 + 16c_0c_2^4c_4 - 4c_0c_2^3c_3^2 - 27c_1^4c_4^2 + 18c_1^3c_2c_3c_4 - 4c_1^3c_3^3 - 4c_1^2c_2^3c_4 + c_1^2c_2^2c_3^2 \end{aligned}$$

One can verify that $J_6 = \frac{1}{27}(4J_2^3 - J_3^2)$. Notice that since J_6 is the discriminant then $J_6 \neq 0$ so we also define the $\mathrm{GL}_2(\mathbb{Q})$-invariant or *j-invariant*

$$j = \frac{J_2^3}{4J_2^3 - J_3^2}$$

A slice of the database for quartics looks as follows:

Key	Value
(1, -2, -2, -2, 1)	[2, [4, -416], 4.5162, 'D(4)', -6400, -1/2700]
(-1, 2, -1, -2, 1)	[2, [1, 110], 3.23853, 'D(4)', -448, -1/12096]

TABLE 7. A slice of the database for quartics

The increase of the number of polynomials with respect to height seems very comparable to degree 3 and 4. We present this graphically in Fig. 3.

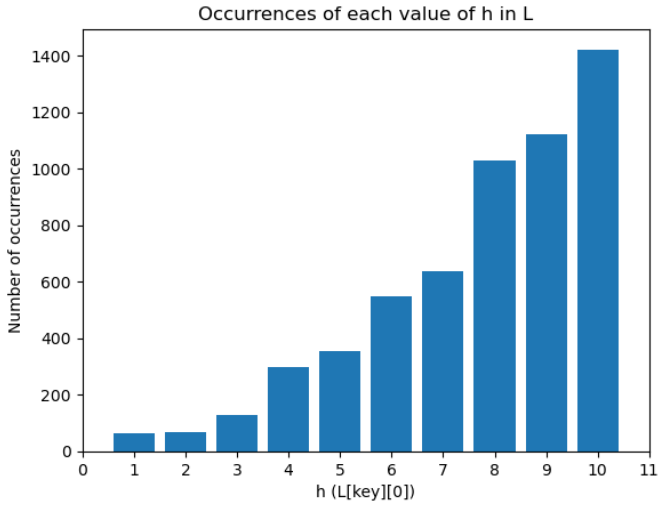


FIGURE 3. This distribution is for quartics with Galois group not isomorphic to S_4.

In [30] we give an estimate on the ratio of the moduli height over the naive height for binary sextics. Such bounds can be given for every degree d polynomial. In the case of quartics the minimum ratio is 0.2236 for the polynomial $f(x) = x^4 - 5x^3 + 10x^2 - 10x + 5$ and the maximum ratio is 3.3959 for $f(x) = x^4 - x^3 - x^2 - x + 1$. The first quartic has Galois group C_4 and the second F_5. We present the ration of the weighted height over the naive height in Fig. 4

There are 5676 irreducible quartics of naive height $h \leq 10$ with Galois group not isomorphic to S_4. From those D_4: 5162 polynomials, A_4: 184 polynomials, V_4: 222 polynomials, and C_4: 108 polynomials. In Fig. 3 we display how the number of such polynomials grows according to the height. The 5676 irreducible quartics are up to $\mathbb{Z}$-equivalence. However, there are only 1231 irreducible quartics up to $\mathbb{Q}$-equivalence, counted by their j-invariant.

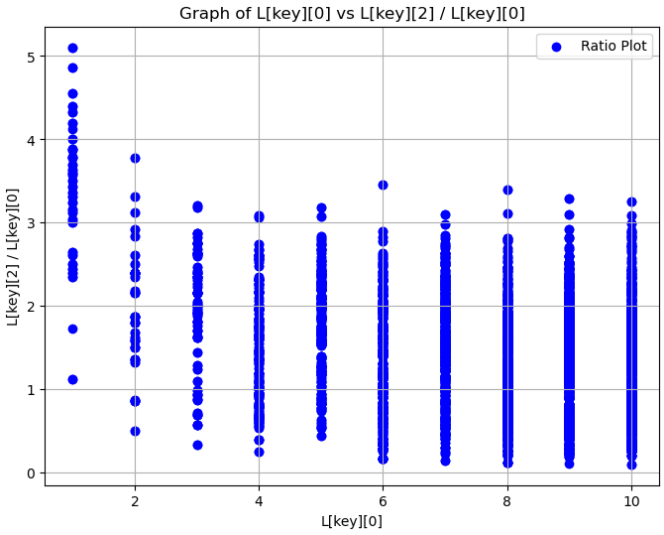


FIGURE 4. The ratio of weighted height with naive height

In [5], being unaware of the weighted height, the authors define the height of a binary quartic as

$$h(f) = \max\{|J_2|^3, |J_3|^2\}$$

Of course this is what we have called the *moduli height* and it is simply the six power $\mathfrak{H}_{\rm k}(f)^6$ of the weighted height. One of the problems considered in [5] is the number of binary quadratic with bounded height. The authors give necessary and sufficient conditions for (J_2, J_3) to be invariants of an integral quartic. We verify such conditions in our database.

The case of quartics is very interesting in its own due to many connections to number theory and elliptic curves and will be the focus of a more detailed investigation in a later stage.

10.5. Quintics. Next we consider the irreducible quintics over $\mathbb{Q}$. Again polynomial will be identified with points $[c_0 : c_1 : c_2 : c_3 : c_4 : c_5]$ in $\mathbb{P}^4$. By Lem. 18 the Galois group of an irreducible quintic is one of the following C_5, D_5, $F_5 = AGL(1,5)$, A_5, S_5. By Eq. (10) the invariants are ξ_0, ξ_1, ξ_2 of order 4, 8, 12 respectively. The expressions of such invariants in **??** suggest we use instead

(41)

$$\begin{aligned} J_4 =& -\frac{625}{2}\cdot\xi_0 = -625c_0^2c_5^2 + 250c_0c_1c_4c_5 - 25c_0c_2c_3c_5 - 40c_0c_2c_4^2 \\ &+ 15c_0c_3^2c_4 - 40c_1^2c_3c_5 - 9c_1^2c_4^2 + 15c_1c_2^2c_5 + 19c_1c_2c_3c_4 - 6c_1c_3^3 - 6c_2^3c_4 + 2c_2^2c_3^2 \\ J_8 =& 1562500\cdot\xi_1 \end{aligned}$$

There are two other invariants J_{12} and J_{18} whch we don't display here and there is a degree 36 homogenous polynomial $F(J_4, J_8, J_{12}, J_{18}) = 0$. This is a homogenous polynomial of degree 36 in terms of coefficients. Hence, a degree two polynomial in J_{18}. According to Dolgachev [10, pg. 152] the discriminant of the quintic is $\Delta = J_4^2 - 128J_8$.

A slice of the dictionary for quintics is given below:

Key	Value
(-2, -1, 0, -2, -2, 1)	[2, [-3264, -8152576, -29726998528], 7.55853, 'F(5) = 5:4']
(1, 0, -1, 2, -2, 1)	[2, [-539, 3599, 116197], 4.81834, 'D(5) = 5:2']
(2, -2, 2, 0, -1, 1)	[2, [-1768, 203456, 379094016], 6.48441, 'A5']

The increase of the number of polynomials with respect to height seems very comparable to degree 3 and 4. We present this graphically in Fig. 5.

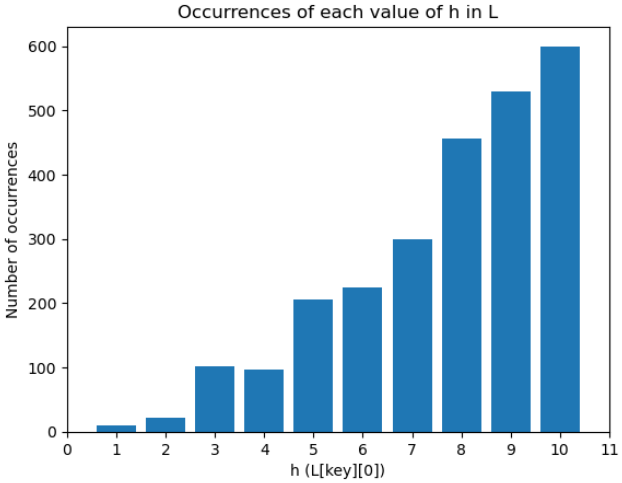


FIGURE 5. This distribution is for quintics with Galois group not isomorphic to S_5.

In [30] we give an estimate on the ratio of the moduli height over the naive height for binary sextics. Such bounds can be given for every degree d polynomial. In the case of quintics the minimum ratio is 0.5353 for the polynomial

$$f(x) = x^5 - 5x^4 + 9x^3 - 9x^2 + 4x - 1$$

and the maximum ratio is 3.7792 for

$$f(x) = x^5 - 2x^4 - 2x^3 - x - 2.$$

The first quintic has Galois group D_5 and the second F_5. We present the ration of the weighted height over the naive height in Fig. 6

LEMMA 28. *From all irreducible quintics in $\mathbb{Z}[x]$ with naive height ≤ 10 there are exactly 20 of them with Galois group C_5, 480 with group F_5, 900 with group D_5, and 1146 with group A_5. Moreover, all polynomials with Galois group C_5 and their invariants are listed in Tab. 8.*

Data in Tab. 8 shows some very interesting trends. First, There are really only 3 quintics with Galois group C_5 up to $\bar{\mathbb{Q}}$-isomorphism since they obviously have the same invariants. This once more stresses the point that the absolute invariants are really the most effective way of dealing with such databases since they considerable decrease the size of the database. Furthermore, by decreasing redundancy the

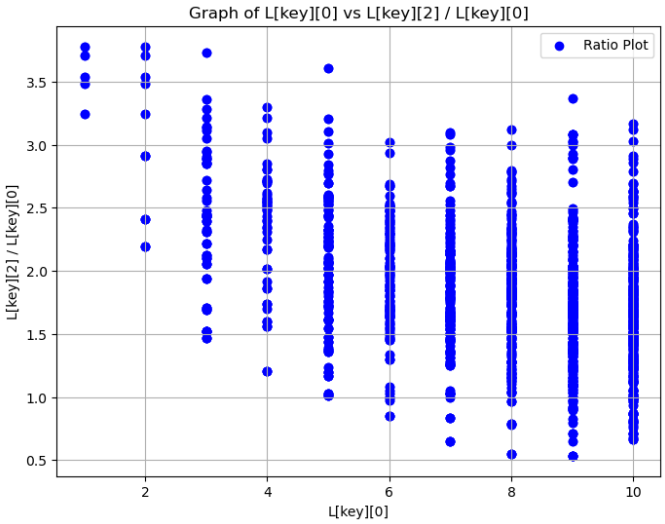


FIGURE 6. The ratio of weighted height with naive height

Key	h	p	wh	
-1, 1, 4, -3, -3, 1	4	[4235, 4026275, -16076916075]	8.06702	C_5
-1, 3, 3, -4, -1, 1	4	[4235, 4026275, -16076916075]	8.06702	C_5
1, 3, -3, -4, 1, 1	4	[4235, 4026275, -16076916075]	8.06702	C_5
1, 1, -4, -3, 3, 1	4	[4235, 4026275, -16076916075]	8.06702	C_5
-1, -2, 5, 2, -4, 1	5	[4235, 4026275, -16076916075]	8.06702	C_5
1, 4, 2, -5, -2, 1	5	[4235, 4026275, -16076916075]	8.06702	C_5
-1, 4, -2, -5, 2, 1	5	[4235, 4026275, -16076916075]	8.06702	C_5
1, -2, -5, 2, 4, 1	5	[4235, 4026275, -16076916075]	8.06702	C_5
1, -6, 10, -1, -6, 1	10	[4235, 4026275, -16076916075]	8.06702	C_5
1, -6, -1, 10, -6, 1	10	[4235, 4026275, -16076916075]	8.06702	C_5
-1, -6, -10, -1, 6, 1	10	[4235, 4026275, -16076916075]	8.06702	C_5
-1, -6, 1, 10, 6, 1	10	[4235, 4026275, -16076916075]	8.06702	C_5
-1, 4, 9, -5, -9, 1	9	[113377, 2971552001, -47471703427379]	18.3498	C_5
-1, 9, 5, -9, -4, 1	9	[113377, 2971552001, -47471703427379]	18.3498	C_5
1, 9, -5, -9, 4, 1	9	[113377, 2971552001, -47471703427379]	18.3498	C_5
1, 4, -9, -5, 9, 1	9	[113377, 2971552001, -47471703427379]	18.3498	C_5
-1, 0, 10, 5, -10, 1	10	[109375, 2392578125, -96893310546875]	18.18568	C_5
-1, 10, -5, -10, 0, 1	10	[109375, 2392578125, -96893310546875]	18.18568	C_5
1, 10, 5, -10, 0, 1	10	[109375, 2392578125, -96893310546875]	18.18568	C_5
1, 0, -10, 5, 10, 1	10	[109375, 2392578125, -96893310546875]	18.18568	C_5

TABLE 8. The only quintics of height ≤ 10 and Galois group isomorphic to C_5

learning process of any AI model becomes more efficient. Some of these issues are further illustrated and discussed in [28].

Second, the polynomials in [27] provide interesting examples of how the height of the binary form can change even for polynomials of such small height. These are very interesting examples in reduction theory; see [29] and more recently [15]

Finally, the above data emphasizes how rare such cases are. There are roughly 20^6 quintic polynomials of height ≤ 10 and from those only three (up to $\bar{\mathbb{Q}}$-isomorphism) have Galois group isomorphic to C_5. Training an AI model to pick such very rare cases might be an impossible task indeed. We will explore that in the next section.

11. Neuro-symbolic networks

A neuro-symbolic network is a type of artificial intelligence system that combines the strengths of neural networks (good at pattern recognition) with symbolic reasoning (based on logic and rules) to create models that can both learn from data and reason through complex situations, essentially mimicking human-like cognitive abilities by understanding and manipulating symbols to make decisions; this approach aims to overcome limitations of either method alone, providing better explainability and adaptability in AI systems. They seem to be the most reasonable choice for our approach since we can use all the theoretical knowledge that we have about polynomials and their Galois groups and somehow incorporate this into some machine learning model. The area of research on deep learning for symbolic mathematics is very active and has had a lot of activity in the ast few years; [1, 2, 9, 18, 21, 23]

11.1. Architecture of Neuro-Symbolic Networks. Let x represent the raw input data. The architecture of a neuro-symbolic network consists of the following key components:

11.1.1. *Input Layer and Preprocessing.* The input x is mapped to a higher-dimensional feature space z_0 through a preprocessing function $f_0 : X \to Z_0$, where X is the input space and Z_0 is the processed feature space. This step typically involves convolutional, recurrent, or embedding layers to transform raw data into structured representations.

11.1.2. *Neural Feature Extraction.* A sequence of neural transformations $f_i : Z_{i-1} \to Z_i$ for $i = 1, \ldots, n$ is applied to extract features. The final output of this stage is a feature representation z_n. These transformations may include convolutional layers for spatial data, recurrent layers for temporal data, or feedforward layers for general patterns:

$$z_n = f_n \circ f_{n-1} \circ \cdots \circ f_1(z_0).$$

11.1.3. *Interface Layer.* The feature representation z_n is mapped to a symbolic representation s through an interface function $\varphi : Z_n \to S$, where S is the symbolic space. Mechanisms such as attention models or learned symbolic encoding are employed. Feedback processes can also map symbolic insights s back into neural spaces Z_n to refine feature extraction:

$$s = \varphi(z_n), \quad z_n' = \psi(s, z_n).$$

11.1.4. *Symbolic Reasoning Layer.* The symbolic representation s undergoes logical or algebraic reasoning. This layer uses symbolic inference mechanisms $R : S \to S'$, such as rule-based systems, constraint solvers, or formal logic:

$$s' = R(s),$$

where S' is the transformed symbolic space.

11.1.5. *Output Integration.* The final output y is derived by integrating the outputs from both neural and symbolic pathways. Let $\rho : S' \times Z_n \to Y$ denote the integration function, mapping the refined symbolic reasoning s' and neural feature representation z_n to the output space Y:

$$y = \rho(s', z_n).$$

11.1.6. *Dynamic Feedback Loops.* Throughout the architecture, feedback loops dynamically adjust both neural and symbolic pathways. Symbolic reasoning s' can guide neural updates, and neural features z_n may suggest new symbolic rules or hypotheses:

$$z_n \to \varphi(s) \to R(s') \to \psi(z_n').$$

This architecture integrates the strengths of neural networks for learning from high-dimensional data and symbolic methods for reasoning, interpretability, and leveraging explicit rules. Neuro-symbolic networks are particularly effective for tasks requiring both data-driven insights and rule-based decision-making. We will try to incorporate some of the above for our particular data with the main focus of determining the Galois group of polynomials. There are many other open questions that one could ask on the data as comparing height of polynomials with the weighted moduli height, classifying equivalence classes of polynomials as described in section 3, investigating Malle's conjecture, and others and for each one of these questions a neuro-symbolic network tailored to the specific question has to be designed.

11.2. Precomputed data for every degree d. For each degree d we precompute two lists:

- " d-grps" which is the list of transitive subgroups of S_d as explained in section 7
- "d-sig" which is the list of the signature for every group in " d-grps"

Such data can be computed using GAP and methods from group theory.

11.3. Signature layer. The first symbolic reasoning layer that we apply to our data is the *signature layer*. This layer for every point $key = (a_0, \dots, a_d)$ creates the polynomial $f(x)$ and computes the factorization $f_p(x)$ for a list of primes p. Normally we use $p = 2, 3, 5, 7$. This signature $\text{sig}(key)$ is compared with the list of possible signatures for the degree d. The field of *groups* for this entry is updated with the list of all groups which admit this signature. If length of $L[key][groups] = 1$ then Gal (f) is uniquely determined and the training is done.

```python
from sympy import symbols, Poly, factor_list
def sig_layer(p):
    x = symbols('x')
    f = sum(a * x**i for i, a in enumerate(p))
    signature = [5]
    primes = [2, 3, 5, 7]
    for prime in primes:
        poly_mod = Poly(f, x, modulus=prime)
        factors = factor_list(poly_mod)[1]  # Get the list of
            factors modulo the prime
        for factor_poly, multiplicity in factors:
            degree = factor_poly.degree()
            if degree > 1 and degree not in signature:  # Avoid
                linear factors and duplicates
                signature.append(degree)
    return signature
```

LISTING 1. Python implementation of the `sig_layer` function.

11.4. Real roots layer. If the polynomial has enough real roots then from **??** the group is A_d or S_d. Computing the real roots is usually easy since it can be done with numerical methods. Hence, for high enough degree d it is usually an efficient method to compute the number of the real roots of $f(x)$.

```python
from sympy import symbols, diff, Poly, sign

def sturm_sequence(P, x):
    P = Poly(P, x)  # Ensure P is treated as a polynomial
    sequence = [P, P.diff(x)]  # Start with P and its
        derivative
    while True:
        remainder = -sequence[-2].rem(sequence[-1])  #
            Polynomial remainder
        if remainder.is_zero:
            break
        sequence.append(remainder)
    return sequence

def count_sign_changes(sequence, value):
    evaluations = []
    for poly in sequence:
        eval_value = poly.eval(value)
        if eval_value == 0:
            evaluations.append(0)  # Consider zero explicitly
        else:
            evaluations.append(sign(eval_value))
    evaluations = [s for i, s in enumerate(evaluations) if i ==
        0 or s != evaluations[i - 1]]
    return len(evaluations) - 1

def real_root_count(P, x, interval=(-1e10, 1e10)):
    a, b = interval
    P = Poly(P.expand(), x)  # Fully expand the polynomial
```

```
    sturm_seq = sturm_sequence(P, x)
    sign_changes_a=count_sign_changes(sturm_seq,a)
    sign_changes_b=count_sign_changes(sturm_seq,b)
    return sign_changes_a - sign_changes_b
```

LISTING 2. Real Root Counting Algorithm

The algorithm for finding the number of real roots of a polynomial using Sturm's theorem involves constructing a Sturm sequence, which starts with the polynomial $f(x)$ and its derivative, followed by successive remainders from polynomial division, with signs reversed. The number of real roots in a given interval is determined by evaluating the sequence at the interval endpoints and counting sign changes in the resulting values. By substituting large finite values ($\pm 10^{10}$) for infinity, the method can approximate the count of real roots over the entire real line. This approach works efficiently for polynomials with integer or rational coefficients.

11.5. Discriminant layer. The discriminant is computed for all polynomials in the precomputed data stage, but it is not factored. This layer is activated only if the entry has as Galois group candidates which are contained or not in the alternating group A_d. Since this layer can slow down considerably the model, we only activate it as a last resort.

11.6. Implementation and efficiency. We implement this approach and test it for quartics and quintics databases that we created for this paper. The case of cubics is quite trivial from the point of view of Galois theory and we ignore it here. While both quartics and quintics are well understood and we don't need any AI model to find out the Galois group, they do provide nice test cases which can tell us how reasonable and efficient such approach is. We study sextics in more detail in [27].

12. Galois Network

We design a network that integrates numerical learning with symbolic reasoning to classify polynomials based on their Galois group properties. The core of this system, which we call the GaloisNetwork, processes polynomial coefficients and leverages mathematical insights to predict the corresponding group labels. This hybrid approach combines the power of deep learning with domain-specific rules, ensuring both accuracy and interpretability.

The input to the network consists of feature vectors derived from polynomial coefficients. We compute these features using mathematical invariants, such as root counts and other Galois group characteristics, creating a robust representation of each polynomial. The features are standardized to improve model performance and are then split into training and validation datasets. Labels representing Galois groups are mapped to numeric values for compatibility with the learning process.

The GaloisNetwork itself is a fully connected feedforward neural network. It begins with an input layer that matches the size of the feature vectors. The network includes three hidden layers, each with 64 neurons and ReLU activation functions, providing the capacity to learn complex patterns in the data. Finally, an output layer produces a probability distribution over all possible Galois group labels using a softmax activation. This architecture allows the network to effectively capture the relationships between features and group classifications.

Training the network involves minimizing a cross-entropy loss function using the Adam optimizer. Over 100 epochs, the network iteratively updates its weights through backpropagation, ensuring that it learns to align its predictions with the true labels. To monitor its progress, we periodically evaluate the model on a validation set, tracking the loss and refining the learning process.

To enhance the model's predictions, we implement a post-processing step that applies domain-specific rules. For example, if the number of real roots of a polynomial exceeds a certain threshold, the prediction is adjusted to align with known Galois group properties. This rule-based layer ensures that the network respects established mathematical principles, making its outputs both reliable and interpretable.

Finally, we evaluate the system using accuracy metrics, confusion matrices, and detailed classification reports. These evaluations demonstrate the effectiveness of combining numerical learning with symbolic reasoning. By integrating these two paradigms, our design not only achieves high accuracy but also maintains alignment with the underlying mathematical structure of the problem, providing a powerful tool for analyzing polynomials through the lens of their Galois groups.

```
coefficients_list = [list(key) for key in L.keys()]
labels = [L[key][3] for key in L.keys()]
features = [compute_galois_features(coeffs) for coeffs in
    coefficients_list]

label_mapping = {label: idx for idx, label in enumerate(set(labels))}
labels_numeric = [label_mapping[label] for label in labels]

scaler = StandardScaler()
features_scaled = scaler.fit_transform(features)
X_train, X_val, y_train, y_val = train_test_split(features_scaled,
    labels_numeric, test_size=0.2, random_state=42)

features_tensor = torch.tensor(X_train, dtype=torch.float32)
features_tensor_validation = torch.tensor(X_val, dtype=torch.float32)
labels_tensor = torch.tensor(y_train, dtype=torch.long)
labels_tensor_validation = torch.tensor(y_val, dtype=torch.long)

class GaloisNetwork(nn.Module):
    def __init__(self, input_size, hidden_size, output_size):
        super(GaloisNetwork, self).__init__()
        self.layers = nn.Sequential(
            nn.Linear(input_size, hidden_size),
            nn.ReLU(),
            nn.Linear(hidden_size, hidden_size),
            nn.ReLU(),
            nn.Linear(hidden_size, hidden_size),
            nn.ReLU(),
            nn.Linear(hidden_size, output_size)
        )
        self.softmax = nn.Softmax(dim=1)

    def forward(self, x):
        return self.softmax(self.layers(x))

input_size = len(features[0])
hidden_size = 64  # Increased for complexity
```

```
output_size = len(label_mapping)

model = GaloisNetwork(input_size, hidden_size, output_size)
criterion = nn.CrossEntropyLoss()
optimizer = optim.Adam(model.parameters(), lr=0.001)

for epoch in range(100):  # Increased epochs for better training
    model.train()
    optimizer.zero_grad()
    outputs = model(features_tensor)
    loss = criterion(outputs, labels_tensor)
    loss.backward()
    optimizer.step()

    if (epoch + 1) % 10 == 0:  # Print every 10 epochs
        model.eval()
        with torch.no_grad():
            val_outputs = model(features_tensor_validation)
            val_loss = criterion(val_outputs, labels_tensor_validation
                )
        print(f"Epoch {epoch+1}, Loss: {loss.item()}, Validation Loss:
             {val_loss.item()}")

# Evaluate model on validation set
model.eval()
with torch.no_grad():
    predictions = model(features_tensor_validation)
    predicted_classes = torch.argmax(predictions, dim=1)
    accuracy = accuracy_score(labels_tensor_validation.cpu().numpy(),
        predicted_classes.cpu().numpy())
    print(f"Validation Accuracy: {accuracy}")

    # Confusion Matrix
    cm = confusion_matrix(labels_tensor_validation.cpu().numpy(),
        predicted_classes.cpu().numpy())
    sns.heatmap(cm, annot=True, fmt='d')
    plt.title('Confusion Matrix')
    plt.ylabel('True label')
    plt.xlabel('Predicted label')
    plt.show()

    # Classification Report
    print(classification_report(labels_tensor_validation.cpu().numpy()
        , predicted_classes.cpu().numpy(), target_names=list(
        label_mapping.keys())))
```

LISTING 3. Python Code for Training a Neural Network

13. Concluding Remarks

This project is a work in progress that uses machine learning to study Galois theory, focusing on polynomials and their Galois groups. By combining traditional algebra with modern computing, we're finding new ways to connect mathematics and data science.

We've shown that supervised learning and neuro-symbolic networks can predict Galois groups and check if polynomials are solvable by radicals. Unsupervised learning has helped us find hidden patterns in polynomial data. We built a detailed

database of irreducible polynomials with known Galois groups, using features like discriminants, root differences, and heights. We also used methods like Julia and Hermite equivalence to organize polynomials into classes and studied how heights help define these classes. Viewing polynomials as points in weighted projective spaces has added a useful geometric perspective.

Moving forward, we plan to expand the database to include polynomials of higher degrees and multiple variables. We could also create new features for machine learning models based on what we learn from the data. More advanced models, like graph neural networks, might better capture how polynomial roots interact. Transfer learning could help apply our findings to more complex problems. We also want to automate classification methods and create tools to visualize results, making this work easier for others to use. Exploring connections to field extensions, algebraic geometry, or even physics could make this approach useful in other areas.

This project shows that machine learning can work with classical math to create new tools for algebraists and reveal new insights. By blending math and computing, we're opening doors to new ways of doing mathematical research.

References

[1] Rashid Barket, Matthew England, and Jürgen Gerhard, *Symbolic integration algorithm selection with machine learning: LSTMs vs tree LSTMs*, Mathematical software—ICMS 2024, [2024] ©2024, pp. 167–175. MR4786719

[2] Rashid Barket, Uzma Shafiq, Matthew England, and Juergen Gerhard, *Transformers to predict the applicability of symbolic integration routines* (2024), available at 2410.23948.

[3] W. E. H. Berwick, *On Soluble Sextic Equations*, Proc. London Math. Soc. (2) **29** (1928), no. 1, 1–28. MR1575303

[4] Manjul Bhargava, Jan-Hendrik Evertse, Kálmán Győry, László Remete, and Ashvin A. Swaminathan, *Hermite equivalence of polynomials*, Acta Arith. **209** (2023), 17–58. MR4665252

[5] Manjul Bhargava and Arul Shankar, *Binary quartic forms having bounded invariants, and the boundedness of the average rank of elliptic curves*, Ann. of Math. (2) **181** (2015), no. 1, 191–242. MR3272925

[6] A. Clebsch and P. Gordan, *Theorie der abelschen funktionen*, Teubner, 1866.

[7] Henri Cohen, *A course in computational algebraic number theory*, Springer-Verlag, Berlin, 1993.

[8] Elira Curri, *On the stability of binary forms and their weighted heights*, Albanian J. Math. **16** (2022), no. 1, 3–23. MR4448533

[9] Tereso del Río and Matthew England, *Lessons on datasets and paradigms in machine learning for symbolic computation: a case study on CAD*, Math. Comput. Sci. **18** (2024), no. 3, Paper No. 17, 27. MR4796805

[10] Igor Dolgachev, *Lectures on invariant theory*, Lond. Math. Soc. Lect. Note Ser., vol. 296, Cambridge: Cambridge University Press, 2003 (English).

[11] Quentin Guignard, *Counting algebraic points of bounded height on projective spaces*, J. Number Theory **170** (2017), 103–141. MR3541701

[12] Thomas R. Hagedorn, *General formulas for solving solvable sextic equations*, J. Algebra **233** (2000), no. 2, 704–757. MR1793923

[13] Marc Hindry and Joseph H. Silverman, *Diophantine geometry*, Graduate Texts in Mathematics, vol. 201, Springer-Verlag, New York, 2000. An introduction. MR1745599 (2001e:11058)

[14] R. Bruce King, *Beyond the quartic equation*, Birkhäuser Boston, Inc., Boston, MA, 1996. MR1401346

[15] Ilias Kotsireas and Tony Shaska, *A machine learning approach of Julia reduction*, RISAT preprints (202412), available at https://www.risat.org/pdf/2024-06.pdf.

[16] Vishwanath Krishnamoorthy, Tanush Shaska, and Helmut Völklein, *Invariants of binary forms*, Progress in Galois theory, 2005, pp. 101–122. MR2148462

[17] Joseph P. S. Kung and Gian-Carlo Rota, *The invariant theory of binary forms*, Bull. Amer. Math. Soc. (N.S.) **10** (1984), no. 1, 27–85. MR722856
[18] Guillaume Lample and François Charton, *Deep learning for symbolic mathematics* (2019), available at 1912.01412.
[19] Shigeru Mukai, *An introduction to invariants and moduli. Transl. from the Japanese by W. M. Oxbury*, Reprint of the 2003 hardback ed., Camb. Stud. Adv. Math., vol. 81, Cambridge: Cambridge University Press, 2012 (English).
[20] P. E. Newstead, *Geometric invariant theory*, Moduli spaces and vector bundles, 2009, pp. 99–127. MR2537067
[21] Kimia Noorbakhsh, Modar Sulaiman, Mahdi Sharifi, Kallol Roy, and Pooyan Jamshidi, *Pretrained language models are symbolic mathematics solvers too!* (2023), available at 2110.03501.
[22] Peter J. Olver, *Classical invariant theory*, London Mathematical Society Student Texts, vol. 44, Cambridge University Press, Cambridge, 1999. MR1694364
[23] Lynn Pickering, Tereso del Río Almajano, Matthew England, and Kelly Cohen, *Explainable AI insights for symbolic computation: a case study on selecting the variable ordering for cylindrical algebraic decomposition*, J. Symbolic Comput. **123** (2024), Paper No. 102276, 24. MR4669630
[24] George Salmon, *Modern higher algebra*, Cambridge University Press, Cambridge, 1876.
[25] I. Schur, *Vorlesungen über Invariantentheorie*, Grundlehren Math. Wiss., vol. 143, Springer, Cham, 1968 (German).
[26] Elira Shaska and Tony Shaska, *Galois theory: A database approach*, RISAT preprints (December 2024), 40.
[27] ———, *Irreducible sextics, invariants, and their galois groups*, RISAT preprints (202412), available at https://www.risat.org/pdf/2024-07.pdf.
[28] ———, *Machine learning for moduli space of genus two curves and an application to isogeny based cryptography* (2024), available at 2403.17250.
[29] T. Shaska, *Reduction of superelliptic Riemann surfaces*, Automorphisms of Riemann surfaces, subgroups of mapping class groups and related topics, 2022, pp. 227–247. MR4375119
[30] T. Shaska and L. Beshaj, *Heights on algebraic curves*, Advances on superelliptic curves and their applications, 2015, pp. 137–175. MR3525576
[31] Bartel Leendert van der Waerden, *Modern algebra*, Springer, Berlin, Heidelberg, 2003. Translated from the German by Fred Blum and John R. Schulenberger.

Department of Computer Science,, College of Enginnering, Oakland University, Rochester, MI, 48309
Email address: elirashaska@oakland.edu

Department of Mathematics and Statistics,, College of Arts and Sciences, Oakland University, Rochester, MI, 48309
Email address: tanush@umich.edu