A MACHINE LEARNING APPROACH TO JULIA REDUCTION

ILIAS KOTSIREAS AND TONY SHASKA

ABSTRACT. We implement a machine learning algorithm based on Julia's original work on reduction of binary forms.

CONTENTS

1. Introduction 1
2. Preliminaries 3
2.1. Reduction theory for binary quadratics 4
3. Julia quadratic and Julia invariant for binary forms 4
3.1. Julia quadratic of binary forms with real coefficients 5
3.2. Julia's quadratic for binary forms with complex coefficients 7
4. Reducing binary forms of higher degree 10
4.1. Binary forms with real coefficients 10
4.2. Binary forms with complex coefficients 10
4.3. Cubics 12
5. The minimal absolute height of binary forms 14
5.1. An algorithm to find the minimum absolute height 14
6. A neural network based on Julia reduction 16
7. Creating a database of binary forms with rational coefficients 17
8. Implementing Julia reduction 18
References 20

1. INTRODUCTION

Reduction of integer binary forms is a classical problem in mathematics. It basically is the idea of picking a coordinate system such that the binary form has "small" coefficients. However, the only case that is fully understood is for quadratics. In 1917, in the first part of his thesis, Gustav Julia suggested a very interested reduction method for an arbitrary degree binary form. It is based on the idea of defining a quadratic (Julia quadratic) $\mathcal{J}_f$ which is covariant under the action of the modular group via coordinate changes. This quadratic is a positive definite quadratic and therefore has only one root in the upper-half complex plane $\mathcal{H}_2$, say α_f. Since $\mathcal{J}(f)$ is an $\mathrm{SL}_2(\mathbb{Z})$-covariant, then bringing α_f to the fundamental domain $\mathcal{F}$ of $\mathrm{SL}_2(\mathbb{Z})$ by a matrix $M \in \mathrm{SL}_2(\mathbb{Z})$, induces an action $f \to f^M$ on binary forms. The form f^M is called **Julia reduction of** f, which is different from *reduction* used in [4, Sec. 4] which means the binary form with the smallest naive height.

In [5] Cremona and Stoll developed a reduction theory in a unified setting for binary forms with real or complex coefficients. A *unique* positive definite Hermitian quadratic $\mathcal{J}_f$ is associated to every binary complex form $f(x, y)$. Since positive definite Hermitian forms parametrize the upper half-space $\mathcal{H}_3$, an extension of the zero map ξ from binary complex forms to $\mathcal{H}_3$ is obtained. The upper half-plane $\mathcal{H}_2$ is contained in $\mathcal{H}_3$ as a vertical cross section (see the following section). When the form $f(x, y)$ has real coefficients, compatibility with complex conjugation forces $\xi(\mathcal{J}_F) \in \mathcal{H}_2$.It is in this sense that working in $\mathcal{H}_3$ unifies the theory of real and complex binary forms. A degree n complex binary form $f(x, y)$ is called *reduced* when its zero map value $\xi(\mathcal{J}_f)$ is in the fundamental domain of the action of the modular group $\mathrm{SL}_2(\mathbb{C})$ on $\mathcal{H}_3$.

For real cubics and quartics, Julia [3] uses geometric constructions to establish the barycentric coordinates $t_1, \dots, t_n$ of $\xi(f)$ in the hyperbolic convex hull of the roots of f. Geometric arguments are also used in [5] for the reduction of binary complex forms. Julia invariant is a multivariable function given in terms of parameters $t_1, \dots, t_s$ and the n-th power of the discriminant of $\mathcal{J}_f$. It has a unique minimum which can be determined via Lagrange multipliers.

In this paper, we design a new neural network where the loss function is exactly Julia invariant, and its minimum is determined via Lagrange multipliers. While the use of Lagrange multipliers in machine learning is not new, the use of such loss function has never been done before as far as we are aware.

While the use of J_f as a loss function comes simply of geometric intuition (as it was Julia's approach), it remains to be investigated if such model can be used in other applications as well.

2. Preliminaries

A **quadratic form over** $\mathbb{R}$ is a function $Q : \mathbb{R}^n \to \mathbb{R}$ that has the form $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ where A is a symmetric $n \times n$ matrix called the **matrix of the quadratic form**. Two quadratic form $F(x, z)$ and $G(x, z)$ are said to be **equivalent over** $\mathbb{R}$ if one can be obtained from the other by linear substitutions. In other words, $G(x, z) = F(ax + bz, cx + dz)$, for some $a, b, c, d \in \mathbb{R}$. Let F, G be quadratic forms and A_F, A_G their corresponding matrices, then $F \sim G$ if and only if A_F is similar to A_G. From now on the terms quadratic form and a symmetric matrix will be used interchangeably.

Let $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ be a quadratic form. The binary quadratic form Q is **positive definite** if $Q(\mathbf{x}) > 0$ for all nonzero vectors $\mathbf{x} \in \mathbb{R}^n$, and Q is **positive semidefinite** if $Q(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$. The binary quadratic form Q is said to be **negative definite** if $Q(\mathbf{x}) < 0$ for all nonzero vectors $\mathbf{x} \in \mathbb{R}^n$, and Q is **negative semidefinite** if $Q(\mathbf{x}) \leq 0$ for all $\mathbf{x} \in \mathbb{R}^n$. Q is **indefinite** if $Q(\mathbf{x})$ is positive for some $\mathbf{x}$'s in $\mathbb{R}^n$, and negative for others. The above definitions of positive definite carry over to matrices and they are found everywhere in the linear algebra literature. A symmetric $n \times n$ matrix A is **positive definite** if the corresponding quadratic form $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is positive definite. Analogous definitions apply for **negative definite** and **indefinite**.

Let $Q(x, z) = ax^2 + bxz + cz^2$ be a binary quadratic in $\mathbb{R}[x, z]$. We will use the following notation to represent the equivalence class of binary quadratics up to a scalar multiple, $Q(x, z) = [a, b, c]$. The **discriminant** of Q is $\Delta = b^2 - 4ac$ and $Q(x, z)$ is positive definite if $a > 0$ and $\Delta < 0$. Denote the set of positive definite binary quadratics with $V^+_{2,\mathbb{R}}$, i.e.

$$V^+_{2,\mathbb{R}} = \left\{ Q(x, z) \in \mathbb{R}[x, z] \;\middle|\; Q(x, z) \textit{ is positive definite} \right\}.$$

Let $\mathrm{SL}_2(\mathbb{R})$ act as usual on the set of positive definite binary quadratic forms

$$\begin{aligned} \mathrm{SL}_2(\mathbb{R}) \times V^+_{2,\mathbb{R}} &\to V^+_{2,\mathbb{R}} \\ \begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{bmatrix} \times \begin{bmatrix} x \\ z \end{bmatrix} &\to Q(\alpha_1 z + \alpha_2 z, \alpha_3 x + \alpha_4 z) \end{aligned}$$

We will denote this new form with $Q^M(x, z) = a'x^2 + b'xz + c'z^2$ where

$$\begin{aligned} a' &= a\alpha_1^2 + b\alpha_1\alpha_3 + c\alpha_3^2 \\ b' &= 2(a\alpha_1\alpha_2 + c\alpha_3\alpha_4) + b(\alpha_1\alpha_4 + \alpha_2\alpha_3) \\ c' &= a\alpha_2^2 + b\alpha_2\alpha_4 + c\alpha_4^2 \end{aligned} \tag{1}$$

and $\Delta' = b'^2 - 4a'c' = (\det M)^2\Delta$. Δ is fixed under the $\mathrm{SL}_2(\mathbb{R})$ action and the leading coefficient of the new form Q^M will be $Q^M(1, 0) = Q(a, c) > 0$. Hence, $V^+_{2,\mathbb{R}}$ is preserved under this action. Consider the map

$$\begin{aligned} \xi : V^+_{2,\mathbb{R}} &\to \mathcal{H}_2 \\ [a, b, c] &\mapsto \xi(Q) = \frac{-b + \sqrt{\Delta}}{2a} \end{aligned} \tag{2}$$

where $\mathrm{Re}(\xi(Q)) = -\frac{b}{2a}$, and $\mathrm{Im}(\xi(Q)) = \frac{\sqrt{|\Delta|}}{2a}$. It is called the **zero map** and is a bijection which gives us a one-to-one correspondence between positive definite quadratic forms and points in $\mathcal{H}_2$.

Let G be a group and X, Y two G-sets. A function $f : X \to Y$ is said to be **G-equivariant** if $f(gx) = gf(x)$, for all $g \in G$ and $x \in X$. In other words the following diagram

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ {\scriptstyle g}\downarrow & & \downarrow{\scriptstyle g} \\ X & \xrightarrow{f} & Y \end{array}$$

commutes. Let $\Gamma := \mathrm{SL}_2(\mathbb{Z})/\{\pm I\}$ be the modular group on $V_{2,\mathbb{R}}^+$ as described above. It also acts (from the right) on $\mathcal{H}_2$ via

$$\begin{aligned} &\mathcal{H}_2 \times \Gamma \to \mathcal{H}_2 \\ &(z, M) \to zM := M^{-1}(z) = \frac{\alpha_4 z - \alpha_2}{\alpha_1 - \alpha_3 z} \end{aligned} \tag{3}$$

Note that the image is also in the upper half-plane, since

$$\mathrm{Im}(M^{-1}(z)) = \det(M^{-1}) \cdot \frac{\mathrm{Im}(z)}{\left\| \alpha_1 - a_3 z \right\|^2}.$$

The zero map $\xi : V_{2,\mathbb{R}}^+ \to \mathcal{H}_2$ is a Γ-equivariant map. In other words, $\xi(Q^M) = M^{-1}\xi(Q)$. In other words,

$$\begin{array}{ccc} V_{2,\mathbb{R}}^+ & \xrightarrow{\xi} & \mathcal{H}_2 \\ {\scriptstyle M}\downarrow & & \downarrow{\scriptstyle M^{-1}} \\ V_{2,\mathbb{R}}^+ & \xrightarrow{\xi} & \mathcal{H}_2 \end{array}$$

2.1. **Reduction theory for binary quadratics.** We denote by $\mathcal{F}$ the fundamental region of the action of Γ on $\mathcal{H}_2$. Define $Q = [a, b, c]$ to be **reduced** if $\xi(Q) \in \mathcal{F}$.

Lemma 1. *The following are true:*

(1) *A positive definite quadratic form $Q \in V_{2,\mathbb{R}}^+$ is reduced iff $|b| \leq a \leq c$.*
(2) *Let Q be a reduced form with fixed discriminant $\Delta = -D$. Then $b \leq \sqrt{D/3}$.*
(3) *The number of reduced forms of a fixed discriminant $\Delta = -D$ is finite.*
(4) *Every positive definite quadratic form Q with fixed discriminant is equivalent to a reduced form of the same discriminant.*

Two reduced binary quadratics are equivalent only in the following two cases $[a, b, a] \sim [a, -b, a]$, and $[a, a, c] \sim [a, -a, c]$. Let $\Delta < 0$ be fixed. Then the class number $h(\Delta)$ is equal to the number of primitive reduced forms of discriminant Δ.

For a binary form $f(x, y) = \sum a_i x^i y^{d-i}$ its **naive height** is defined as $\mathrm{H}(f) = \max\{|a_i|\}$. Let $f(x, z) = ax^2 + bxz + cz^2$ be reduced (i.e. $|b| < a < c$). Then $\mathrm{H}([f]) = c$. Moreover, if f is reduced quadratic then f has minimal height $\mathrm{H}(f)$ in its Γ-orbit.

3. Julia quadratic and Julia invariant for binary forms

In this section we introduce the Julia quadratic of binary forms. The Julia quadratic was introduced in 1917 by Gaston Julia in his PhD thesis; see [3]. It did not get the attention that it deserved. Indeed Julia became known for most of his other work on Julia sets and fractals. In [1] are used ideas of Julia to explore the reduction for cubic binary forms and in [5] is given a generalization of Julia's work for binary forms defined over $\mathbb{C}$.

3.1. **Julia quadratic of binary forms with real coefficients.** We will motivate and define the Julia quadratic of a binary form of degree $n \geq 2$ with real coefficients. We will try to follow as closely as possible the approach and notation used in Julia's original paper [3].

Let $f(x, y) \in \mathbb{R}[x, y]$ be a degree n binary form given as follows:

$$f(x, y) = a_0 x^n + a_1 x^{n-1} y + \cdots + a_n y^n$$

and suppose that $a_0 \neq 0$. Let the real roots of $f(x, y)$ be α_i, for $1 \leq i \leq r$ and the pair of complex roots β_j, $\bar{\beta}_j$ for $1 \leq j \leq s$, where $r + 2s = n$. The form can be factored as

$$f(x, 1) = \prod_{i=1}^{r} (x - \alpha_i) \cdot \prod_{i=1}^{s} (x - \beta_i)(x - \bar{\beta}_i). \tag{4}$$

The ordered pair (r, s) of numbers r and s is called the **signature** of the form f.

We associate to f the two quadratics $T_r(x, 1)$ and $S_s(x, 1)$ of degree r and s respectively given by the formulas

$$T_r(x, 1) = \sum_{i=1}^{r} t_i^2 (x - \alpha_i)^2, \quad \textit{and} \quad S_s(x, 1) = \sum_{j=1}^{s} 2u_j^2 (x - \beta_j)(x - \bar{\beta}_j), \tag{5}$$

where t_i, u_j are to be determined. Then

$$\begin{aligned} T_r(x, 1) &= \left(\sum_{i=1}^{r} t_i^2 \right) x^2 - 2 \left(\sum_{i-1}^{r} t_i^2 \alpha_i \right) x + \left(\sum_{i-1}^{r} t_i^2 \alpha_i^2 \right) \\ S_s(x, 1) &= 2 \left(\sum_{j=1}^{s} u_j^2 \right) x^2 - 4 \left(\sum_{j=1}^{s} u_j^2 \operatorname{Re}(\beta_j) \right) x + 2 \left(\sum_{j=1}^{s} u_j^2 \cdot ||\beta_j||^2 \right). \end{aligned} \tag{6}$$

For a binary form f of signature (r, s) the quadratic Q_f is defined as

$$Q_f(x, 1) = T_r(x, 1) + S_s(x, 1). \tag{7}$$

Let $\beta_i = a_i + b_i \cdot I$, for $i = 1, \ldots, s$. The discriminant of Q_f is a degree 4 homogenous polynomial in $t_1, \ldots t_r$, $u_1, \ldots, u_s$. We would like to pick values for $t_1, \ldots t_r, u_1, \ldots, u_s$ such that this discriminant is square free and minimal. Then we can use the reduction theory of quadratics (with square free, minimal discriminant) to determine the reduced form for Q_f.

For quadratics T and S in Eq. (5) we define

$$\theta_T = \frac{a_0^2 \cdot \Delta_T}{t_1^2 \cdots t_r^2}, \qquad \theta_S = \frac{a_0^2 \cdot \Delta_S}{u_1^4 \cdots u_s^4} \tag{8}$$

Notice that T_r and S_s are given recursively as

$$T_r = T_{r-1} + t_r^2 (x - \alpha_r)^2, \qquad S_s = S_{s-1} + u_s^4 \left(x^2 - 2a_s x + (a_s^2 + b_s^2) \right)$$

The next lemma gives formulas computing the discriminants of T and S.

Lemma 2. *Let T_r and S_s be quadratics given by*

$$T_r(x, 1) = \sum_{i=1}^{r} t_i^2 (x - \alpha_i)^2, \quad \textit{and} \quad S_s(x, 1) = \sum_{j=1}^{s} 2u_j^2 (x - \beta_j)(x - \bar{\beta}_j), \tag{9}$$

where $\beta_i = a_i + I \cdot b_i$, *for* $i = 1, \dots, s$. *Then* $T_r \in V_{2,\mathbb{R}}^+$ *and* $S_s \in V_{2,\mathbb{R}}^+$. *Moreover,*

$$\Delta(T_r) = -4\,(t_1^2 \cdots t_r^2) \sum_{\substack{i,j=1\\ i\neq j\\ n_l\neq i, n_l\neq j}}^{r} \frac{(\alpha_i-\alpha_j)^2}{t_{n_1}^2 \cdots t_{n_l}^2 \cdots t_{n_{r-2}}^2} = -4 \sum_{i<j}^{r} t_i^2 t_j^2\,(\alpha_i-\alpha_j)^2,$$

(10)

$$\Delta(S_s) = -16 \left(\sum_{i<j} u_i^2 u_j^2 \left[(a_i-a_j)^2 + (b_i^2+b_j^2) \right] + \sum_{j=1}^{s} u_j^4 b_j^2 \right).$$

Notice that $\Delta(T_r)$ and $\Delta(S_s)$ can be given recursively as follows

$$\Delta(T_{r+1}) = \Delta(T_r) - 4 \sum_{i=1}^{r} t_i^2 t_{r+1}^2\,(\alpha_i - \alpha_{r+1})^2$$

$$\Delta(S_{s+1}) = \Delta(S_s) - 16 \left(\sum_{i=1}^{s} u_i^2 u_{s+1}^2 \left[(\alpha_1-\alpha_{s+1})^2 + b_i^2 + b_{s+1}^2 \right] + u_{s+1}^4 b_{s+1}^2 \right).$$

Proposition 1. *Let* $f \in V_{n,\mathbb{R}}$ *with signature* (r,s) *and equation as in Eq.* (4). *Then* Q_f *is a positive definite quadratic form with discriminant* $\mathfrak{D}_f$ *given by the formula*

$$\mathfrak{D}_f = \Delta(T_r) + \Delta(S_s) - 8 \sum_{i,j} t_i^2 u_j^2 \left((\alpha_i - a_j)^2 + b_j^2 \right). \tag{11}$$

From the above formula it can be seen that $\mathfrak{D}_f$ is expressed in terms of the root differences. Hence, $\mathfrak{D}_f$ is fixed by all the transpositions of the roots. However, it is not an invariant of the binary form. In order to get an invariant we need to fix it by all symmetries of the roots, hence by an element of order n. Indeed $\mathfrak{D}_f^n$ is an invariant of the binary form f as we will see later. We define the θ_0 of a binary form as follows

$$\theta_0(f) = \frac{a_0^2 \cdot |\mathfrak{D}_f|^{n/2}}{\prod_{i=1}^{r} t_i^2 \prod_{j=1}^{s} u_j^4}. \tag{12}$$

Example 1. *Let* $f \in V_{2,\mathbb{R}}$. *Assume that* f *has signature* $(2,0)$, *say* $f(x,1) = a_0(x-\alpha_1)(x-\alpha_2)$ *and discriminant* $\Delta_f = a_0^2(\alpha_1-\alpha_2)^2$. *Then*

$$Q_f(x,1) = T_2(x,1) = \left(t_1^2+t_2^2\right) x^2 - 2\left(t_1^2\alpha_1 + t_2^2\alpha_2\right) x + \left(t_1^2\alpha_1^2 + t_2^2\alpha_2^2\right)$$

and its discriminant,

$$\mathfrak{D}_f = -4\,(\alpha_1-\alpha_2)^2\, {t_2}^2 {t_1}^2 < 0.$$

Since $t_1^2+t_2^2 > 0$ *and* $\mathfrak{D}_f < 0$, *then* $Q_f \in V_{2,\mathbb{R}}^+$. *Moreover,*

$$\theta_0(f) = \frac{a_0^2 \cdot \sqrt{|\mathfrak{D}_f|^2}}{t_1^2 t_2^2} = 4 \cdot a_0^2 \cdot (\alpha_1-\alpha_2)^2 = 4 \cdot \Delta_f.$$

Assume now that f *has signature (0, 1) (i.e. with no real roots). Then*

$$f(x,1) = a_0(x-\beta)(x-\bar{\beta})$$

for some $\beta = a + bi \in \mathcal{H}_2$ *and discriminant* $\Delta_f = -4a_0^2 b^2$. *Then,*

$$Q_f(x,1) = S_1(x,1) = 2u_1^2\left(x^2 - 2ax + (a^2+b^2)\right)$$

and its discriminant Δ_S *is given by*

$$\mathfrak{D}_f = -16\, u_1^4 b^2 < 0.$$

Thus, since $2u_1^2 > 0$ *and* $\mathfrak{D}_f < 0$ *then* $Q_f \in V_{2,\mathbb{R}}^+$*. Then,*

$$\theta_0(f) = \frac{a_0^2 \cdot \sqrt{|\mathfrak{D}_f|^2}}{u_1^4} = 16a_0^2 b^2 = -4 \cdot \Delta_f.$$

□

Notice that in order for f to be in somewhat "simpler" or "minimal" form we would like the discriminant Δ_f to be minimal. Hence, we would like $\theta_0(f)$ to be minimal. Consider $\theta_0(t_1, \dots, t_r, u_1, \dots, u_s)$ as a multivariable function in the variables $t_1, \dots, t_r, u_1, \dots, u_s$. We would like to pick these variables such that Q_f is a reduced quadratic, hence $\mathfrak{D}_f$ is minimal. This is equivalent to $\theta_0(t_1, \dots, t_r, u_1, \dots, u_s)$ obtaining a minimal value.

Proposition 2. *The function* $\theta_0 : \mathbb{R}^{r+s} \to \mathbb{R}$ *obtains a minimum at a unique point* $(\bar{t}_1, \dots, \bar{t}_r, \bar{u}_1, \dots, \bar{u}_s)$*.*

The proof is an elementary exercise from multivariable calculus. Julia in his thesis [3] proves existence and Stoll, and Cremona prove uniqueness in [5].

Definition 1. *Choosing* $(\bar{t}_1, \dots, \bar{t}_r, \bar{u}_1, \dots, \bar{u}_s)$ *that make* θ_0 *minimal gives a unique positive definite quadratic* $Q_f(x, z)$*. We call this unique quadratic* $Q_f(x, z)$ *for such a choice of* $(\bar{t}_1, \dots, \bar{t}_r, \bar{u}_1, \dots, \bar{u}_s)$ *the* ***Julia quadratic*** *of* $f(x, z)$*, denote it by* $\mathcal{J}_f(x, z)$*, and the quantity* $\theta_f := \theta_0(\bar{t}_1, \dots, \bar{t}_r, \bar{u}_1, \dots, \bar{u}_s)$ *the* ***Julia invariant****.*

The following lemma shows that θ is an invariant of binary forms and $\mathcal{J}$ a covariant of order 2.

Lemma 3. *Consider* $\mathrm{SL}_2(\mathbb{R})$ *acting on* $V_{n,\mathbb{R}}$*. Then* θ *is an* $\mathrm{SL}_2(\mathbb{R})$*- invariant and* $\mathcal{J}$ *is an* $\mathrm{SL}_2(\mathbb{R})$ *covariant of order 2.*

3.2. **Julia's quadratic for binary forms with complex coefficients.** Suppose we are given a binary form $f \in V_{n,\mathbb{C}}$ with $f(x, y) = \sum_{i=0}^n x^{n-i} y^i$ and assume that $a_0 \neq 0$. Then $f(x, y)$ can be factored as

$$f(x, y) = a_0(y_1 x - x_1 y)(y_2 x - x_2 y) \cdots (y_n x - x_n y), \tag{13}$$

for $[x_i, y_i] \in \mathbb{P}^1$, $i = 1, \dots n$. Construct a quadratic form

(14)

$$\begin{aligned} Q(x, y) &= \sum_{i=1}^n t_i^2 \cdot ||y_i x - x_i y||^2 \\ &= \left(\sum_{i=1}^n t_i^2 ||y_i||^2 \right) x\bar{x} - \left(\sum_{i=1}^n t_i^2 y_i \bar{x}_i \right) x\bar{y} - \left(\sum_{i=1}^n t_i^2 x \bar{y}_i \right) \bar{x} y + \left(\sum_{i=1}^n t_i^2 \cdot ||x_i||^2 \right) y\bar{y} \end{aligned}$$

where t_j are non-zero real numbers that have to be determined. Computing the discriminant of the quadratic $Q(x, z)$ and simplifying it we get

$$\mathfrak{D}_f = \sum_{1=i<j=n} t_i^2 t_j^2 \cdot ||y_i x_j - x_i y_j||^2 = \sum_{1=i<j=n} t_i^2 t_j^2 \cdot ||\beta_{ij}||^2. \tag{15}$$

Note that $||\beta_{ij}|| := ||y_i x_j - x_i y_j||$. Since the leading coefficient of Q and $\mathfrak{D}_f$ are both positive then Q is a positive definite quadratic Hermitian form. We define the quantity θ_0 as

$$\theta_0(Q_f) = \frac{||a_0||^2 \cdot \mathfrak{D}_f^{n/2}}{t_1^2 \cdots t_n^2}.$$

Consider θ_0 as a function

$$\begin{aligned} \theta_0 : \mathbb{P}^{n-1} \backslash \{(0, \dots, 0)\} &\to \mathbb{P}^1 \\ (t_1, \dots, t_n) &\mapsto \theta_0(t_1, \dots, t_n). \end{aligned}$$

Since this is a function defined on $\mathbb{P}^{n-1}$ then we take its domain to be

$$D = \left\{ (t_1, \dots, t_n) \in \mathbb{P}^n : \ t_1^2 \cdot t_2^2 \cdots t_n^2 = 1 \right\}.$$

We would like to choose $t_1, \dots, t_n$ such that Q_f is a reduced quadratic, hence a quadratic with minimal discriminant. Then θ_0 obtains a minimum exactly when $\mathfrak{D}_f$ obtains a minimum, under the assumption $t_1^2 \cdots t_n^2 = 1$. Our next task is to determine in what values for $(t_1, \dots, t_n)$ this minimum occurs. For simplicity denote by $h = \mathfrak{D}_f$. To find the critical points in the interior of D we need to solve $\nabla_h = 0$, i.e.

$$2t_i \sum_{\substack{j=1 \\ j \neq i}} t_j^2 \cdot ||y_i x_j - x_i y_j||^2 = 0, \qquad i = 1, \dots n.$$

Note that the only critical point in the interior D° is the tuple $(0, \dots, 0)$, which is not in the domain.

Next, determine the critical points on the boundary of D. Denote by $g = \prod_{i=1}^n t_i^2 = 1$. Using Lagrange multipliers we have to solve the system

$$\begin{cases} \nabla_h = \lambda \nabla_g \\ t_1^2 \cdots t_n^2 = 1 \end{cases}$$

for $\lambda \neq 0$. For convenience denote

$$\boxed{u_i = t_i^2 \quad \textit{and} \quad \alpha_{i,j} = ||\beta_{i,j}||^2 = ||y_i x_j - x_i y_j||^2}$$

and we have

$$\begin{cases} \sum_{\substack{j=1 \\ i \neq j}}^n u_j \cdot \alpha_{i,j} = \lambda \cdot \prod_{i \neq j} u_j, \qquad i = 1, \dots, n \\ \prod_{i=1}^n u_i = 1 \end{cases}$$

or equivalently

$$\begin{cases} u_i \sum_{\substack{j=1 \\ i \neq j}}^n u_j \cdot \alpha_{i,j} = \lambda \\ \prod_{i=1}^n u_i = 1 \end{cases} \tag{16}$$

Let V be the variety defined by the **??**. We have the following.

Theorem 1. *V is a zero dimensional variety over $\mathbb{C}$. Moreover, V has exactly one real point given by*

$$u_i = \frac{2}{n} \cdot \frac{t^2}{(||z - \alpha_i||^2 + t^2)},$$

where t and z satisfy the following system

$$\begin{cases} \displaystyle\sum_{j=1}^n \frac{t^2}{||z - \alpha_j||^2 + t^2} = \frac{n}{2} \\ \displaystyle\sum_{j=1}^n \frac{z - \alpha_j}{||z - \alpha_j||^2 + t^2} = 0 \end{cases} \tag{17}$$

Let $(\bar{u}_1, \dots, \bar{u}_n) \in \mathbb{R}^n$ be the unique real point of V. From now on by θ_f we will denote the function θ_0 evaluated at this unique point. The quadratic $Q(f)$ for the above values $(\bar{u}_1, \dots, \bar{u}_n)$ will be denoted by $\mathcal{J}_f$ and is called **Julia's quadratic**.

Lemma 4. *Let* $\mathrm{SL}_2(\mathbb{C})$ *act on* $V_{n,\mathbb{C}}$*. Then the following are true:*
i) θ_f *is an invariant ii)* $\mathfrak{D}_f^n$ *is an invariant.*

Corollary 1. *Let* $f \in V_{n,\mathbb{C}}$ *and* F_f *its field of moduli. Then,*
i) $\theta_f \in F_f$.
ii) $a_0^4\, \mathfrak{D}_f^n \in F_f(\theta_f^2)$.
iii) $\mathcal{J}_f \in F_f[x, y]$

Proof. It is by definition that $\theta_f \in F_f$ and $\mathcal{J}_f$ has coefficients in $F_f[x, y]$. Part iii) is a consequence of the definition of θ_f. □

Remark 1. *An open question is to express* θ *in terms of generators of the rings of invariants for degree* n *binary forms or absolute invariants of* f *which determine the field of moduli of* f*.*

4. Reducing binary forms of higher degree

In this section we will describe reduction theory of higher degree binary forms. First, we will explain the case of binary forms with real coefficients and then its generalization to binary forms with complex coefficients.

4.1. Binary forms with real coefficients.

To any form $f \in V_{n,\mathbb{R}}$ we associate a positive definite quadratic $\mathcal{J}_f \in V_{2,\mathbb{R}}^+$ as showed above. In **??** we proved that binary quadratic forms in $V_{2,\mathbb{R}}^+$ are in one-to-one correspondence with points in the upper half plane $\mathcal{H}_2$. Hence, we have the following maps

$$\begin{aligned} \zeta : V_{n,\mathbb{R}} &\to V_{2,\mathbb{R}}^+ \to \mathcal{H}_2 \\ f &\mapsto \mathcal{J}_f \mapsto \xi(\mathcal{J}_f). \end{aligned}$$

We call this map the **zero map** and denote it by $\zeta(f) := \xi(\mathcal{J}_f)$.

Proposition 3. *The map $\zeta : V_{n,\mathbb{R}} \to \mathcal{H}_2$ is $\mathrm{SL}_2(\mathbb{R})$-equivariant.*

The proof of the above proposition is easy and it will be proved in the next subsection for the more general case, i.e. binary forms with complex coefficients. A binary form $f \in V_{n,\mathbb{R}}$ is **reduced** if $\zeta(f) \in \mathcal{F}_2$. Next, we will adapt this to binary forms with complex coefficients.

4.2. Binary forms with complex coefficients.

For any form $f \in V_{n,\mathbb{C}}$ the corresponding Julia quadratic is a positive definite Hermitian form. Previously we proved that binary quadratic forms in $\mathrm{Her}^+(\mathbb{C})$ are in a one-to-one correspondence with points in $\mathcal{H}_3$. Hence, we have the maps:

$$\begin{aligned} \zeta : V_{n,\mathbb{C}} &\longrightarrow \mathrm{Her}^+(\mathbb{C}) \longrightarrow \mathcal{H}_3 \\ f &\mapsto \mathcal{J}_f \mapsto \xi(\mathcal{J}_f) \end{aligned}$$

where ξ is as defined in **??**. Note that $\xi(\mathcal{J}_f)$ is the point in $\mathcal{H}_3$ associated to the Hermitian form $\mathcal{J}_f$.

Lemma 5. *The map $j : V_{n,\mathbb{C}} \longrightarrow \mathrm{Her}^+(\mathbb{C})$ is an $\mathrm{SL}_2(\mathbb{C})$-equivariant map, i.e. for every $f \in V_{n,\mathbb{C}}$, $H \in \mathrm{Her}^+(\mathbb{C})$ and $M \in \mathrm{SL}_2(\mathbb{C})$ we have $j(f^M) = j(f)^M$ which is equivalent to saying $H_{f^M} = H_f^M$.*

Proof. We will prove it only for the generators of $\mathrm{SL}_2(\mathbb{C})$, i.e. for $S = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ and $T = \begin{bmatrix} 1 & m \\ 0 & 1 \end{bmatrix}$ where $m \in \mathbb{C}$. First, for $f \in V_{n,\mathbb{C}}$ such that

$$f = a_0(x - \alpha_1 y) \cdots (x - \alpha_n y)$$

and $H \in \mathrm{Her}^+(\mathbb{C})$ we want to prove that $H_{f^S} = H_f^S$. We have

$$f^S = A_0(x - \gamma_1 y) \cdots (x - \gamma_n y)$$

where $A_0 = a_0 \alpha_i^n$ and $\gamma_i = -\frac{1}{\alpha_i}$. The binary quadratic Hermitian form associated to f^S is

$$H_{f^S} = \sum \tau_i^2 ||x - \gamma_i y||^2.$$

On the other side,

$$\begin{aligned} H_f^S &= \sum t_i^2 ||y - \alpha_i(-x)||^2 = \sum t_i^2 \left|\left| \alpha_i \left(x - \frac{y}{-\alpha_i} \right) \right|\right|^2 \\ &= \sum t_i^2 ||\alpha_i||^2 ||x - \gamma_i y||^2. \end{aligned}$$

Notice that for $\tau_i^2 = t_i^2 ||\alpha_i||^2$, we have that $H_f^S = H_{f^S}$. Now let us show $H_{f^T} = H_f^T$. For $f = a_0(x - \alpha_1 y) \cdots (x - \alpha_n y)$ and T as above we have

$$f^T = A_0(x - \gamma_1 y) \cdots (x - \gamma_n y)$$

where $A_0 = a_0$ and $\gamma_i = \alpha_i - m$. The binary quadratic Hermitian form associated to f^T is

$$H_{f^T} = \sum \tau_i^2 ||x - \gamma_i y||^2.$$

On the other side,

$$H_f^T = \sum t_i^2 ||x + my - \alpha_i y||^2 = \sum t_i^2 ||x - (\alpha_i - m)y||^2 = \sum t_i^2 ||x - \gamma_i y||^2.$$

Hence, for $\tau_i^2 = t_i^2$ we have $H_f^T = H_{f^T}$ and we are done. □

Proposition 4. *The map* $\zeta : V_{n,\mathbb{C}} \to \mathcal{H}_3$ *is* $\mathrm{SL}_2(\mathbb{C})$*-equivariant.*

Proof. Let $f \in V_{n,\mathbb{C}}$ and $M \in \mathrm{SL}_2(\mathbb{C})$ be a matrix acting on the given binary form f. We associate to f the Julia quadratic $\mathcal{J}_f$ which is in $\mathrm{Her}^+(\mathbb{C})$. In **??** we proved that the zero map for binary quadratic Hermitian forms is an $\mathrm{SL}_2(\mathbb{C})$-equivariant map. Then we have

$$\begin{aligned} \zeta(f^M) &= \xi(\mathcal{J}_{f^M}) = \xi(\mathcal{J}_f^M) && \text{from Lem. 5} \\ &= M^{-1}\xi(\mathcal{J}_f) && \text{from ??} \\ &= M^{-1}\zeta(f). \end{aligned}$$

Hence, ζ is $\mathrm{SL}_2(\mathbb{C})$-equivariant. This is equivalent to saying that for any $M \in \mathrm{SL}_2(\mathbb{C})$ the following diagram is commutative.

$$\begin{array}{ccccc} V_{n,\mathbb{C}} & \xrightarrow{j} & \mathrm{Her}^+(\mathbb{C}) & \xrightarrow{\xi} & \mathcal{H}_3 \\ \downarrow{\scriptstyle M} & & \downarrow{\scriptstyle M} & & \downarrow{\scriptstyle M^{-1}} \\ V_{n,\mathbb{C}} & \xrightarrow{j} & \mathrm{Her}^+(\mathbb{C}) & \xrightarrow{\xi} & \mathcal{H}_3 \end{array}$$

□

Let K be a field of definition of f. Without loss of generality assume that f has an integral model over $\mathcal{O}_K$. We call $f(x, y)$ to be **reduced** over K if $\zeta(f)$ is in a fixed fundamental domain for the action of Γ_K on $\mathcal{H}_3$, when such a fundamental domain exists.

Definition 2. *Let* $f \in V_{n,\mathbb{C}}$ *be such that it has an integral model over some algebraic number field* K*. We say* $f(x, y)$ *is reduced if* $\zeta(f)$ *is in a fixed fundamental domain for the action of* $\mathrm{SL}_2(\tilde{\mathcal{O}}_K)$ *on* $\mathcal{H}_3$*, when such a domain exists.*

Let f be a given degree n binary form. To find the reduced form in its $\mathrm{SL}_2(\mathcal{O}_K)$-orbit we compute $\zeta(f)$. If $\zeta(f)$ is in the fundamental domain $\mathcal{F}_{\mathcal{O}_K}$ we are done, the given form is the reduced one. Otherwise, compute $M \in \Gamma_{\mathcal{O}_K}$ such that $\zeta(f)^M \in \mathcal{F}_{\mathcal{O}_K}$ and $f^{M^{-1}}$ is the reduced form in its $\mathrm{SL}_2(\mathcal{O}_K)$-orbit.

A natural question to ask is the following; Does the reduced binary form computed this way have minimal height in its $\mathrm{SL}_2(\mathcal{O}_K)$-orbit? We will address this question in the remainder of this section.

Consider f a degree n binary form and K its minimal field of definition. Let $M \in \mathrm{SL}_2(\mathcal{O}_K)$ be a matrix such that f^M is reduced, i.e. $\bar{\xi}(f^M) \in \mathcal{F}_K$ where $\mathcal{F}_K$ is the fundamental domain of $\mathrm{SL}_2(\mathcal{O}_K)$ acting on $\mathcal{H}_3$. First we give a bound on the height of the reduced binary form with respect to the Julia invariant.

Lemma 6. *Let $f \in V_{n,\mathbb{Q}}$ with signature $(n, 0)$. Then*

$$\mathrm{H}(f) \leq c \cdot \theta_f^{n/2},$$

where $c = \left(\frac{1}{3}\right)^{\frac{n^2}{4}} \left(\frac{4}{n-1}\right)^{\frac{n(n-1)}{2}} \frac{1}{a_0^n}$.

Next we see that for binary cubics it is possible to express this bound in terms of the discriminant of the cubic and then we compare this bound with bounds obtained in [2].

Remark 2. *If we consider a binary cubic with signature $(3, 0)$ then from Lem. 6 we have*

$$\mathrm{H}(f) \leq 2^3 \left(\frac{1}{3}\right)^{\frac{9}{4}} \frac{1}{a_0^3} \cdot \theta_f^{3/2}$$

Moreover, $\theta_f = a_0^6\, 3^{\frac{3}{2}}\, |\Delta_f|^{\frac{1}{2}}$, (cf. Section 4.3). We can express the above bound in terms of the discriminant of the binary form f

$$\mathrm{H}(f) \leq 2^3\, a_0^6 \cdot |\Delta_f|^{3/4}.$$

In [2, Thm 2, pg 162] *it is proved that for a binary form f*

$$\mathrm{H}(f) \leq C \cdot |\Delta_f|^{\frac{21}{2}},$$

where C is some constant.

Hence, finding a relation between the $\mathfrak{D}_f$ and Δ_f would give a formula for the Julia invariant θ_f in terms of Δ_f. That would give a bound for $\mathrm{H}(f)$ in terms of Δ_f. This seems to be difficult for $d > 3$.

The main question in reduction of binary forms based on Julia quadratics is the following question: Does the zero-map preserve the naive height? In other words, is it true that

$$\mathrm{H}(\mathcal{J}_{f^M} \leq \mathrm{H}(\mathcal{J}_f) \implies \mathrm{H}(f^M) \leq \mathrm{H}(f)$$

This would guarantee that Julia reduction will always gets the reduced binary forms, because such reduction for quadratic is well defined.

4.3. **Cubics.** Let $f, g \in V_{3,\mathbb{R}}$ be primitive cubic forms

$$f(x,y) = ax^3 + bx^2y + cxy^2 + dy^3$$
$$g(x,y) = a_1x^3 + b_1x^2y + c_1xy^2 + d_1y^3$$

Assume that $\mathrm{H}(f) \leq \mathrm{H}(g)$. How does $\mathrm{H}(\mathcal{J}_f)$ compares with $\mathrm{H}(\mathcal{J}_g)$?

By definition of the naive height we have that

$$\max\{|a|, |b|, |c|, |d|\} \leq \max\{|a_1|, |b_1|, |c_1|, |d_1|\}.$$

Julia quadratics $\mathcal{J}_f$ and $\mathcal{J}_g$ for cubics are given by

$$\mathcal{J}_f(x,y) = (b^2 - 3ac)\,x^2 + (bc - 9ad)\,xy + (c^2 - 3bd)y^2$$
$$\mathcal{J}_g(x,y) = (b_1^2 - 3a_1c_1)\,x^2 + (b_1c_1 - 9a_1d_1)\,xy + (c_1^2 - 3b_1d_1)y^2,$$

see Section 4.3. Since the Julia quadratics are reduced quadratics by reduction theory of binary quadratics and if we assume that $\mathcal{J}_f$ and $\mathcal{J}_g$ are primitive (i.e. $\operatorname{cont}(\mathcal{J}_f) = 1$ and $\operatorname{cont}(\mathcal{J}_g) = 1$) then

$$\mathrm{H}(\mathcal{J}_f) = \left|c^2 - 3bd\right|$$
$$\mathrm{H}(\mathcal{J}_g) = \left|c_1^2 - 3b_1d_1\right|.$$

We want to check whether $\mathrm{H}(\mathcal{J}_f) \leq \mathrm{H}(\mathcal{J}_g)$. Obviously this is not necessary true. Take for example $f(x,y) = x^3 + x^2 + x + 1$ and $g(x,y) = 5x^3 + x^2 + x$. Hence, $\mathrm{H}(f) = 1$ and $\mathrm{H}(g) = 5$, but $\mathrm{H}(\mathcal{J}_f) = |-2| > \mathrm{H}(\mathcal{J}_g) = 1$

However, we are truly interested when $g = f^M$, for some $M \in \mathrm{SL}_2(\mathbb{Z})$, say $M = \begin{bmatrix} \alpha_1, \alpha_2 \\ \alpha_3, \alpha_4 \end{bmatrix}$. Is it really $\mathrm{H}(\mathcal{J}_f) \leq h\left(\mathcal{J}_{f^M}\right)$?

♠♠♠ Tony: [I expect this to be false, which would provide cases when Julia reduction does not give binary forms with minimal height]

This is an optimization problem. It would be interesting to see a geometrical interpretation of it.

♠♠♠ Tony: [We can perform some iterations in python and see what happens. For example take a cubic with big height and iterate with matrices. Probably we want matrices such that the gcd between their entries is 1].

5. The minimal absolute height of binary forms

Let K be a number field and $\mathcal{O}_K$ its ring of integers. We want to develop a reduction theory in the following sense: given a binary form $f(x, y)$ over $\mathcal{O}_K$ we determine its integral model with minimal height $\mathrm{H}(f)$ over $\overline{K}$. Let f and g be two binary forms of degree n and M a matrix in $\mathrm{SL}_2(\mathcal{O}_K)$ such that $g = f^M$. Associate to f and g their Julia quadratics $\mathcal{J}_f$ and $\mathcal{J}_g$. Then $\mathcal{J}_g = \mathcal{J}_f^M$ and $\Delta_{\mathcal{J}_f} = \Delta_{\mathcal{J}_g}$. Hence, the discriminant $\mathfrak{D}_f$ of the Julia quadratic is an invariant of the binary form.

Theorem 2. *Let f be a degree n binary form defined over K and $\mathcal{J}_f$ its Julia quadratic, $\mathfrak{D}_f$ its discriminant, and $L = K(\mathfrak{D}_f)$. Then $[L : K] \leq n$. Let r be the class number of $\mathcal{J}_f$ over L and $M_1, \dots, M_r$ the matrices with entries in $\mathrm{SL}_2(\mathcal{O}_K)$ that send $\mathcal{J}_f$ respectively to $\{J_1, \dots, J_r\}$. The form f^{M_j} for some $j = 1, \dots, r$ has minimal height over $\mathrm{SL}_2(\mathcal{O}_K)$.*

♠♠♠ Tony: [Not worded well] Once we find the "best" binary form amongst all $\mathrm{SL}_2(\mathcal{O}_L)$-orbits we can lower the height of the reduced form if we consider diagonal matrices with entries in $\mathcal{O}_K$. This is done as follows. Let f be a reduced form of degree $n \geq 3$ given by

$$f = a_n x^n + \cdots + a_0 y^n,$$

where $a_0, \dots, a_n \in \mathcal{O}_K$. Consider $M = \mathrm{diag}\ (\alpha, \beta)$ the diagonal matrix with $\alpha, \beta \in \mathcal{O}_K$. Hence, $f^M = (\alpha x, \beta y)$.

Consider $f(\alpha x, y)$. The height $\mathrm{H}(f)$ can be lowered only if all coefficients of $f(\alpha x, y)$ have a common factor. Hence, we must choose α such that $\alpha \,|\, a_0$.

By the same argument, we choose β such that $\beta \,|\, a_n$. Obviously there are only finitely many choices for $M = \mathrm{diag}\ (\alpha, \beta)$. Among all such choices we choose M that gives the smallest height. Obviously, $M \notin \mathrm{SL}_2(\mathcal{O}_K)$ therefore acting with M on the reduced form will lower the height. Hence, we have the following:

Theorem 3. *Let $f = \sum_{i=0}^n a_i x^i y^{n-i}$ be a reduced binary form. Choose $M = \mathrm{diag}\ (\alpha, \beta)$ such that $\alpha \,|\, a_0$ and $\beta \,|\, a_n$ and*

$$\mathrm{H}(f^M) = \min \left\{ \mathrm{H} \left(f^{diag\ (\alpha, \beta)} \right) \right\}$$

Then $\mathrm{H}(f^M) < \mathrm{H}(f)$.

Proof. Let $f = \sum_{i=0}^n a_i x^i y^{n-i}$ be a reduced binary form. Pick α and β such that $\alpha \,|\, a_0$ and $\beta \,|\, a_n$. Then

$$f(\alpha x, \beta y) = \sum_{i=0}^n a_i \alpha^i \beta^{n-i} \, x^i y^{n-i}$$

The content of this new polynomial is $\gcd(a_0, a_1 \alpha \beta^{n-1}, \dots, a_n \alpha^n)$. We choose the form with the smallest height among all primitives of $f(\alpha x, \beta y)$, where α, β are as above. □

5.1. An algorithm to find the minimum absolute height.

We put everything together in the following algorithm, which finds the form with minimal height among all $\mathrm{GL}_2(\mathcal{O}_K)$-orbits is as follows.

Algorithm 1 Computing the binary form with minimal absolute height

Input: A degree n binary form $f(x,y) \in V_{n,\mathcal{O}_K}$
Output: A binary form $F \in V_{n,\mathcal{O}_K}$ which is $\mathrm{GL}_2(\bar{K})$-equivalent to f and has minimal absolute height.

1: Compute the Julia quadratic $\mathcal{J}_f$ associated with the binary form f, as explained in Section 3.2.
2: Compute the zero map $\xi(\mathcal{J}_f) \in \mathbb{H}$ using **??**.
3: Find the matrix A such that $\xi(\mathcal{J}_f)^{A^{-1}} \in \mathbb{F}_{\mathcal{O}_K}$.
4: Assign $f := \textbf{red}\ (f) = f^A$ and $J := J_f^{A^{-1}}$.
5: Compute the discriminant Δ_f of the quadratic form J.
6: Let $L := K(\Delta_f)$ and $h_L(\mathcal{J}) := r$ be the class number of J over L.
7: Determine all quadratics $\{J_1, \dots, J_r\}$ equivalent to J over L, and let $M_1, \dots, M_r \in \mathrm{GL}_2(L)$ be the matrices such that $J = J_i^{M_i}$, for $i = 1, \dots, r$.
8: Compute the set of forms
$$f_1 := f^{M_1}, \dots, f_r := f^{M_r}.$$
9: For each $i = 1, \dots, r$, repeat steps 1-4 to compute $\textbf{red}\ (f_i)$.
10: For each $j = 1, \dots, r$ and $f_j = \sum_{i=0}^{n} a_i x^i y^{n-i}$, do the following: Choose $M = \mathrm{diag}\ (\alpha, \beta)$ such that $\alpha \,|\, a_0$ and $\beta \,|\, a_n$, and pick $g_j := f^{\mathrm{diag}\ (\alpha,\beta)}$ such that
$$\mathrm{H}(f^M) = \min \left\{ \mathrm{H}\left(f^{\mathrm{diag}\ (\alpha,\beta)} \right) \right\}$$
is minimal.
11: Pick the form $F \in V_{n,\mathcal{O}_K}$ with the smallest height among $g_1, \dots, g_r$.
return F

Next we highlight a few remarks about the efficiency of the algorithm.

Remark 3. *For practical purposes computing $\zeta(f)$ numerically is satisfactory since we can find $A \in \Gamma$ such that $\zeta(f)^A \in \mathcal{F}$. Hence, the algorithm can be made rather efficient. The reduced form* **red** *(f) has smaller coefficients and expected minimal height in its Γ-orbit.*

6. A neural network based on Julia reduction

Let k be a field and for any integer $n \geq 1$ denote by $\mathbb{A}_k^n$ (resp. $\mathbb{P}_k^n$) the affine (resp. projective) space over k. When k is an algebraically closed field, we will drop the subscript. A fixed tuple of positive integers $\mathfrak{w} = (q_0, \ldots, q_n)$ is called **set of weights**. The weight of $\alpha \in k$ will be denoted by $\mathbf{wt}(\alpha)$. The set

$$\mathbb{V}_{\mathfrak{w}}^n(k) := \{(x_1, \ldots, x_n) \in k^n \mid \mathbf{wt}(x_i) = q_i,\, i = 1, \ldots, n\}$$

is a graded vector space over k.

A **neuron** is a function $f : \mathbb{V}_{\mathfrak{w}}^n(k) \to k$ such that

$$\alpha_{\mathfrak{w}}(\mathbf{x}) = \sum_{i-0}^{n} w_i x_i + \beta,$$

where $\beta \in k$ is a constant called **bias**. We can generalize neurons to tuples of neurons via

$$\begin{aligned} \phi := \mathbb{V}_{\mathfrak{w}}^n(k) &\to \mathbb{V}_{\mathfrak{w}}^n(k) \\ \mathbf{x} &\to g\left(\alpha_0(\mathbf{x}), \ldots, \alpha_n(\mathbf{x})\right) \end{aligned}$$

for any gives set of weights $\mathfrak{w}_0, \ldots, \mathfrak{w}_n$. Then ϕ is a k-linear function with matrix written as

$$\phi(\mathbf{x}) = W \cdot \mathbf{x} + \beta,$$

for some $\beta \in k^{n+1}$ and W an $n \times n$ matrix with integer entries.

Definition 3. *A function $g : \mathbb{V}_{\mathfrak{w}}^n \to \mathbb{V}_{\mathfrak{w}}^n$ is called an **activation function** while a **network layer** is a function*

$$\begin{aligned} \mathbb{V}_{\mathfrak{w}}^n(k) &\to \mathbb{V}_{\mathfrak{w}}^n(k) \\ \mathbf{x} &\to g\left(W \cdot \mathbf{x} + \beta\right) \end{aligned}$$

*for some g some activation function. A **neural network** is the composition of many layers. The l-th layer*

$$\begin{aligned} \cdots \longrightarrow \mathbb{V}_{\mathfrak{w}}^n(k) &\xrightarrow{\phi_l} \mathbb{V}_{\mathfrak{w}}^n(k) \longrightarrow \cdots \\ \mathbf{x} &\longrightarrow \phi_l(\mathbf{x}) = g_l\left(W^l \mathbf{x} + \beta^l\right), \end{aligned}$$

where g_l, W^l, and β^l are the activation, matrix, and bias corresponding to this layer.

After m layers the output (predicted values) will be denoted by $\hat{\mathbf{y}} = [\hat{y}_1, \ldots, \hat{y_n}]^t$, where

$$\hat{\mathbf{y}} = \phi_m\left(\phi_{m-1}\left(\ldots\left(\phi_1(\mathbf{x})\right)\ldots\right)\right),$$

while the true values by $\mathbf{y} = [y_1, \ldots, y_n]^t$.

For our model the input (in features) will be a vector representing the coefficients of the binary form, say

$$\mathbf{x} = (a_0, \ldots, a_n),$$

where $a_0, \ldots, a_n \in \mathbb{Z}$ and $\gcd(a_0, \ldots, a_n) = 1$.

We will define the **loss function** as the $\theta_0(\mathbf{x})$ function defined in Eq. (12). We will avoid all the discussion of section three and determine the minimum of this function via a machine learning model. The goal of any machine learning model is to minimize the loss function.

7. Creating a database of binary forms with rational coefficients

8. Implementing Julia reduction

```
 # Import required libraries
import numpy as np
import matplotlib.pyplot as plt
from sage.all import *

# Define the polynomial
x = var('x')
f = x^6 + 4*x^4 + 2*x^2 + 1

# Function to compute roots in the upper half-plane
def roots_upper_half_plane(f):
    # Get the roots of the polynomial
    roots = f.roots(multiplicities=False)
    # Filter roots to only include those in the upper half-plane
    upper_half_roots = [r for r in roots if r.imag() > 0]
    return upper_half_roots

# Function to compute the center of gravity of roots
def center_of_gravity(roots):
    if not roots:
        return None
    center = sum(roots) / len(roots)
    return center

# Function to plot roots and center of gravity
def plot_upper_half_roots(roots, center):
    # Separate roots into real and imaginary parts
    real_parts = [r.real() for r in roots]
    imaginary_parts = [r.imag() for r in roots]

    # Create a scatter plot for the roots
    plt.scatter(real_parts, imaginary_parts, color='blue', label='Upper Half Roots')

    # Plot the center of gravity
    if center is not None:
        plt.scatter(center.real(), center.imag(), color='red', label='Center of Gravity', s=100)

    plt.axhline(0, color='black', lw=0.5)
    plt.axvline(0, color='black', lw=0.5)
    plt.xlabel('Real Part')
    plt.ylabel('Imaginary Part')
    plt.title('Roots in the Upper Half Plane')
    plt.legend()
    plt.grid()
    plt.show()

# Function to compute the transformation matrix based on the center
def transformation_matrix(center):
    # Adjust the transformation based on the center coordinates
    if center:
        re = center.real()
        im = center.imag()
        A = Matrix([[1, -re], [0, 1]])  # Example transformation matrix
        return A
    return None

# Function to transform the polynomial using the transformation matrix
```

```
def transform_binary_form(f, A):
    # Transform the polynomial based on the transformation matrix
    x_new = var('x_new')
    return f.subs(x == x_new + A[0, 1] * x_new)


# Main function to find the minimal height form via transformations
def find_minimal_height_form_via_transformation(f):
    # Compute roots in the upper half-plane
    upper_half_roots = roots_upper_half_plane(f)
    # Calculate the center of gravity
    center = center_of_gravity(upper_half_roots)

    # Plot roots and center of gravity
    plot_upper_half_roots(upper_half_roots, center)

    # Compute the transformation matrix
    A = transformation_matrix(center)

    # Transform the polynomial
    f_transformed = transform_binary_form(f, A)

    # Return the polynomial, transformed polynomial, center, and transformation matrix
    return f, f_transformed, A, center
```

References

[1] John E. Cremona, *Reduction of binary cubic and quartic forms*, LMS J. Comput Math **2** (1999), 64–94.

[2] Jan-Hendrik Evertse, *Estimates for reduced binary forms*, J. Reine Angew. Math. **434** (1993), 159–190. MR1195694 (94h:11037)

[3] Gaston Julia, *Étude sur les formes binaires non quadratiques à indéterminées réelles ou complexes.*, Mémories de lAcadémie des Sciences de lInsitut de France **55** (1917), 1–296.

[4] T. Shaska, *Reduction of superelliptic Riemann surfaces*, Automorphisms of Riemann surfaces, subgroups of mapping class groups and related topics, 2022, pp. 227–247. MR4375119

[5] Michael Stoll and John E. Cremona, *On the reduction theory of binary forms*, J. Reine Angew. Math. **565** (2003), 79–99. MR2024647 (2005e:11091)