A NEUROSYMBOLIC FRAMEWORK FOR GEOMETRIC REDUCTION OF BINARY FORMS

ILIAS KOTSIREAS AND TONY SHASKA

ABSTRACT. This paper compares Julia reduction and hyperbolic reduction with the aim of finding equivalent binary forms with minimal coefficients. We demonstrate that hyperbolic reduction generally outperforms Julia reduction, particularly in the cases of sextics and decimics, though neither method guarantees achieving the minimal form. We further propose an additional shift and scaling to approximate the minimal form more closely. Finally, we introduce a machine learning framework to identify optimal transformations that minimize the heights of binary forms. This study provides new insights into the geometry and algebra of binary forms and highlights the potential of AI in advancing symbolic computation and reduction techniques. The findings, supported by extensive computational experiments, lay the groundwork for hybrid approaches that integrate traditional reduction methods with data-driven techniques.

1. INTRODUCTION

Reduction of integer binary forms is a classical problem in mathematics. There are many ways that the term *reduction* is used. Here we will refer to it as the idea of picking a coordinate system such that the binary form has "small" coefficients. This is what is refer in [1] as *Reduction A* versus *Reduction B* which refers to picking the binary form with "smallest" invariants.

From now on, by *reduction* of a binary form $f(x, y)$ defined over a field k, we will refer the process of picking a binary form $g(x, y)$ which is k-equivalent to $f(x, y)$ and has minimal coefficients. The only case that is fully understood and is part of the math folklore is the case of quadratic binary forms.

In his thesis [2] of 1917, G. Julia introduced a reduction theory for binary forms with real coefficients (although explicit and complete answers were provided only in degrees three and four). To every binary form $f(x, y)$ with real coefficients, Julia associated a positive definite quadratic $\mathcal{J}_F$ which is called the *Julia quadratic*. Cremona [3] showed that the coefficients of $\mathcal{J}_F$ are polynomial values of the coefficients of F and this does not happen for higher degree forms. Since positive definite quadratics parametrize $\mathbb{H}_2$, one obtains a well defined map ζ from real binary forms to the upper half-plane. It is called *the zero map* and it is $\mathrm{SL}_2(\mathbb{Z})$-equivariant. If F is a real binary form, then $\zeta(F)$ is a point in the hyperbolic convex hull of the roots of F with non-negative imaginary part. A binary form is called *reduced* if its image via the zero map is in the fundamental domain $\mathcal{F}$ of $\mathrm{SL}_2(\mathbb{Z})$.

In [4] 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 $\mathbb{H}_3$, an extension of the zero map ζ from binary complex forms to $\mathbb{H}_3$ is obtained. The upper half-plane $\mathbb{H}_2$ is contained in $\mathbb{H}_3$ as a vertical cross section (see the following section). When the form $f(x, y)$ has real coefficients, compatibility with complex conjugation forces $\zeta(\mathcal{J}_F) \in \mathbb{H}_2$. It is in this sense that working in $\mathbb{H}_3$ 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 $\zeta(\mathcal{J}_F)$ is in the fundamental domain of the action of the modular group $\mathrm{SL}_2(\mathbb{C})$ on $\mathbb{H}_3$.

In the works cited above, the term *reduced binary form* means reduced in the $\mathrm{SL}_2(\mathbb{Z})$ orbit. It is expected that the reduced forms have smallest size coefficients in such orbit. In [1], the concept of height was defined for forms defined over any ring of integers $\mathcal{O}_K$, for any number field K, and the notion of *minimal absolute height* was introduced and the author suggests an algorithm for determining the minimal absolute height for binary forms.

In [1] the authors introduce an alternative reduction method based purely on a geometric approach. For real cubics and quartics, Julia ([2]) uses geometric constructions to establish the barycentric coordinates $t_1, \ldots, t_n$ of $\zeta(F)$ in the hyperbolic convex hull of the roots of F. Geometric arguments are also used in [4] for the reduction of binary complex forms. In [1] reduction is based solely on a very special geometric point $\zeta_{\mathcal{C}}(F)$ inside the hyperbolic convex hull of the roots of F, namely the *hyperbolic centroid* of these roots. For a finite subset $w_1, \ldots w_n \subset \mathbb{H}_2$, the *hyperbolic centroid* is the unique point x inside their hyperbolic convex hull which minimizes $\sum_{i=1}^{n} \cosh(d_H(\mathbf{x}, w_i)$ (here d_H is the hyperbolic distance). To each real binary form $f(x, y)$ with no real roots, the alternative zero map associates the hyperbolic centroid of its roots. It is shown in [1] that this map is $SL_2(\mathbb{R})$ equivariant and different from Julia's, hence it defines a new reduction algorithm. Although zero maps are different, it seems that the effects of both reductions in decreasing the height are similar. Naturally, one would like to determine how different the zero maps are, or whether one can get examples where the reductions give different results.

The goal of this paper is to explore machine learning techniques, and more specifically neurosymbolic networks, to compare these two types of reduction and further investigate if any of them achieves the minimum height of the binary form. The simplest case would be that of binary sextics, and we will make use of machine learning methods used in [5] for such binary forms. While our methods and algorithms work for any degree, binary sextics and the database of [5,6] provide valuable examples where we can also see how the reduction affects the size of the invariants. We experiment with databases of irreducible quintics in [7] for the case when the binary form has one real root.

Our methods show that, in general, hyperbolic reduction is more effective than Julia reduction. However, it does not always achieve minimal height. In most cases, an **additional shift** is required to reach the minimal height through shifting. Since there is no known method to determine this additional shift, we employ machine learning techniques to further reduce the binary form and obtain its minimal height.

To conclude, the study of binary form reduction is not only a classical topic but also a rich intersection of geometry, algebra, and computational techniques. By applying modern machine learning frameworks, we aim to provide new insights and

algorithms that extend beyond traditional symbolic methods, paving the way for future advancements in this field.

2. PRELIMINARIES

Let k be a field and $k[x, y]$ the ring of polynomials in two variables. A degree n homogenous polynomial $f \in k[x, y]$ can be written as

$$(1) \qquad f(x, y) = \sum_{i=0}^{n} a_i x^i y^{d-i}$$

for $a_0, \ldots, a_n \in k$. Two homogenous polynomials f and g are called equivalent if $f(x, y) = \lambda \cdot g(x, y)$ for some $\lambda \in k^\star$. Equivalence classes of homogenous polynomials are called **binary forms**. The set of degree n binary forms over k will be denoted by $V_{n,k}$. There is a one to one correspondence between $V_{n,k}$ and the projective space $\mathbb{P}_k^n$. Hence, sometimes we will denote the equivalence class of f by $[a_0 : \ldots : a_n]$. The **height** of f (sometimes called the *naive height*) is defined as the height of $[a_0 : \ldots : a_n] \in \mathbb{P}_k^n$ and is denoted by $\mathrm{H}(f)$. It is well-defined. When k has characteristic zero and f is primitive, then $\mathrm{H}(f) = \max\{|a_i|\}$.

A *quadratic form over k* is a function $Q : k^n \to k$ 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 forms $f(x, y)$ and $g(x, y)$ are called *k-equivalent* if one is obtained from the other by linear substitutions. In other words, $f(x, y) = f(ax + by, cx + dy)$, for some $a, b, c, d \in k$. Let f, g be quadratic forms and A_f, A_g their corresponding matrices, then $f \sim g$ if and only if $A_f \sim A_g$.

Let $k = \mathbb{R}$. The binary quadratic form Q is called *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 $\mathbf{x} \in \mathbb{R}^n$.

Let $Q(x, y) = ax^2 + bxy + cy^2$. We will denote the equivalence class of $Q(x, y)$ by $[a, b, c]$. The *discriminant* of Q is $\Delta = b^2 - 4ac$ and $Q(x, y)$ is positive definite if $a > 0$ and $\Delta < 0$. Let

$$V_{2,\mathbb{R}}^+ = \left\{ Q(x, y) \in \mathbb{R}[x, y] \;\middle|\; Q(x, y) \text{ is positive definite} \right\}.$$

Then $\mathrm{SL}_2(\mathbb{R})$ acts on $V_{2,\mathbb{R}}^+$ via

$$\begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} \to Q(\alpha_1 x + \alpha_2 y, \alpha_3 x + \alpha_4 y) =: Q^M(x, y)$$

The discriminant Δ' of Q^M is $\Delta' = (\det M)^2 \cdot \Delta = \Delta$. Hence, 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 $\zeta : V_{2,\mathbb{R}}^+ \to \mathbb{H}_2$

$$(2) \qquad [a, b, c] \mapsto \zeta(Q) = \frac{-b + \sqrt{\Delta}}{2a}$$

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

The group $\Gamma := \mathrm{SL}_2(\mathbb{Z})/\{\pm I\}$ is called the **modular group**. It acts on $V_{2,\mathbb{R}}^+$ as above. It also acts (from the right) on $\mathbb{H}_2$ via $\mathbb{H}_2 \times \Gamma \to \mathbb{H}_2$

$$(3) \qquad\qquad (z, M) \to zM := M^{-1}(z)$$

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

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

This action has a fundamental domain $\mathcal{F}$

$$\mathcal{F} = \left\{ z \in \mathcal{H}_2 \;\middle|\; |z|^2 \geq 1 \text{ and } |Re(z)| \leq 1/2 \right\}$$
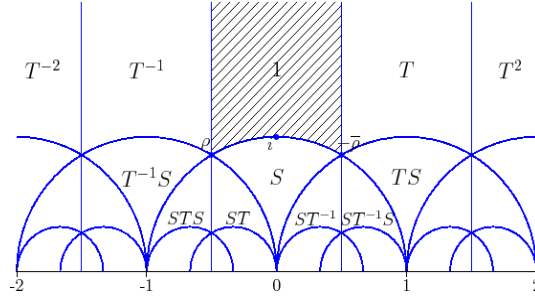
as displayed in Fig. 1.

FIGURE 1. The action of the modular group on $\mathbb{H}_2$.

The zero map is Γ-equivariant, i.e. $\zeta(Q^M) = M^{-1}\zeta(Q)$

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

We define the quadratic $Q = [a, b, c]$ to be a **reduced quadratic** if $\zeta(Q) \in \mathcal{F}$.

Lemma 1. *The following are true:*

(1) $Q \in V_{2,\mathbb{R}}^+$ *is reduced if and only if* $|b| \leq a \leq c$.
(2) *Let Q be a reduced form with $\Delta = -D$. Then* $b \leq \sqrt{D/3}$.
(3) *The number of reduced forms with $\Delta = -D$ is finite.*
(4) *Every $Q \in V_{2,\mathbb{R}}^+$ is equivalent to a reduced one.*

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

2.1. **The hyperbolic plane** $\mathbb{H}_2$. The upper-half plane equipped with the Riemanian metric

$$ds^2 = \frac{dx^2 + dy^2}{y^2}$$

is one of the models of the two dimensional hyperbolic space. The geodesics of the Riemaniann manifold $\mathbb{H}_2$, i.e the hyperbolic equivalents of Euclidean straight lines, are either semicircles $C_{a,b}$ with diameter from $A(a,0)$ to $B(b,0)$ on the real axis, or the vertical rays C_a with origin at $x = a$. In the standard literature, the points $A(a,0), B(b,0)$ are called **the ideal points** of the geodesic $C_{a,b}$, likewise $A(a,0)$ and ∞ are the ideal points of C_a. The ideal points of the geodesic live in the boundary of $\mathbb{H}_2$; see Fig. 2.
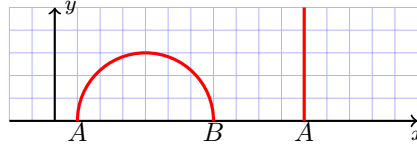

FIGURE 2. Ideal points and the corresponding geodesics

Let $z = x+\mathbf{i}y$, $w = u+\mathbf{i}v$ and z_∞, w_∞ be the ideal points of the geodesic through z, w, where z_∞ is the one closer to z; see Fig. 3.
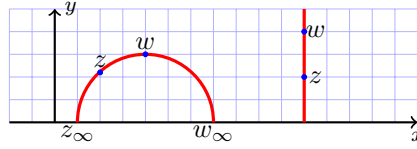

FIGURE 3. Hyperbolic distance between two points z and w

The hyperbolic distance is defined as

$$d_H(z, w) = \ln\left(\frac{|z - w_\infty|}{|w - w_\infty|} \frac{|w - z_\infty|}{|z - z_\infty|} \right).$$

For $x = u$ and $y < v$, the geodesic is the vertical ray C_x. In this case $z_\infty = (x, 0)$, $w_\infty = \infty$ and

$$d_H(z, w) = \ln\left(\frac{v}{y} \right).$$

For $A(a, 0)$ and $z = x + \mathbf{i}y \in \mathbb{H}_2$, define

$$d_H(A, z) := \ln\left(\frac{(x - a)^2 + y^2}{y} \right).$$

Proposition 1. *Let A be one of the ideal points of a geodesic that passes through $z = x + y\mathbf{i}, w = u + v\mathbf{i} \in \mathbb{H}_2$. Then*

$$d_H(z, w) = |d_H(A, z) - d_H(A, w)|.$$

2.2. $\mathbb{H}_2$ as a parameter space for positive definite real quadratic forms.

$\mathbb{H}_2$ parametrizes binary quadratic forms with discriminant $\Delta < 0$ and $a > 0$, while its boundary parametrizes those with discriminant $\Delta = 0$. In [8] it was proved that:

Proposition 2. *Let* $\overline{\mathbb{H}}_2 = \mathbb{H}_2 \cup \partial\mathbb{H}_2 = \mathbb{H}_2 \cup \mathbb{RP}^1$ *and* $\omega_1, \omega_2 \in \overline{\mathbb{H}}_2$. *The quadratics of the form*

$$sQ_{\omega_1} + tQ_{\omega_2}, s > 0, t > 0, s + t = 1$$

parametrize the hyperbolic segment that joins ω_1 *and* ω_2.

This proposition is generalized by induction as follows (see [8]):

Proposition 3. *Let* $\omega_1, \omega_2, ..., \omega_n \in \overline{\mathbb{H}}_2$ *such that for all* i, ω_i *is not in the hyperbolic convex hull of* $\omega_1, \omega_2, ..., \omega_{i-1}$. *Then the convex hull of* $\omega_1, \omega_2, \ldots, \omega_n$ *parametrizes the linear combinations* $\sum_{i=1}^{n} \lambda_i Q_{\omega_i}$ *with* $\lambda_i \geq 0$ *and* $\sum_{i=1}^{n} \lambda_i = 1$.

2.3. The hyperbolic three dimensional space $\mathbb{H}_3$.

As a set, $\mathbb{H}_3 = \mathbb{C} \times \mathbb{R}^+$. Points of $\mathbb{H}_3$ will be written in the form $z + t\mathbf{j}$ where $z \in \mathbb{C}$ and $t > 0$. The equation $t = 0$ represents the floor $\mathbb{C}$ of $\mathbb{H}_3$. The hyperbolic space $\mathbb{H}_3$ is foliated via horospheres

$$H_t := \{z + t\mathbf{j} : \ z \in \mathbb{C}\}$$

which are centered at ∞ and indexed by the height t above $\partial\mathbb{H}_3 = \mathbb{P}^1\mathbb{C}$. The algebra of $\mathbb{H}_3$ is not commutative. Then $\mathbf{j}^2 = -1$, $\mathbf{ij} = -\mathbf{ji}$, $\mathbf{j}z = \bar{z}\mathbf{j}$. The notion of complex modulus extends to $\mathbb{H}_3$: $|z + t\mathbf{j}| = |z|^2 + t^2$. There is a natural isometrical inclusion map $\mathbb{H}_2 \to \mathbb{H}_3$ via $x + \mathbf{i}t \to x + \mathbf{j}t$. Invariant elements of $\mathbb{H}_3$ under the partial conjugation

$$z + \mathbf{j}t \mapsto \bar{z} + \mathbf{j}t$$

are elements of $\mathbb{H}_2$. The hyperbolic metric is

$$ds^2 = \frac{|dz|^2 + dt^2}{t^2}.$$

The geodesics are either semicircles centered on $\mathbb{C}$ and perpendicular to $\mathbb{C}$, or rays $\{z_0 + \mathbf{j}t\}$ perpendicular to $\mathbb{C}$.
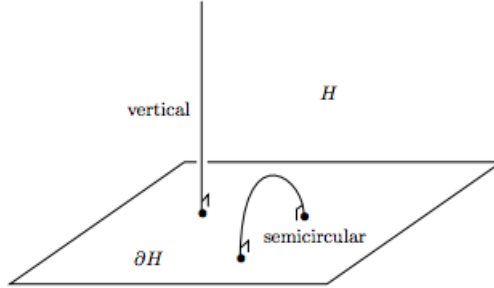

FIGURE 4. Geodesics in $\mathbb{H}_3$

For $\omega = z + t\mathbf{j} \in \mathbb{H}_3$ and $w + 0\mathbf{j} \in \mathbb{C}$ in the floor, define

$$d_H(\omega, w) := \frac{|z - w|^2 + t^2}{y}.$$

Proposition 4. *If one of the ideal points of the geodesic through ω_1, ω_2 is at w, then*

$$d_H(\omega_1, \omega_2) = |d_H(\omega_1, w) - d_H(\omega_2, w)|.$$

There is a right action of $\mathrm{SL}_2(\mathbb{C})$ on $\mathbb{H}_3$. If $M \in \mathrm{SL}_2(\mathbb{C})$ and $M^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, its action is described as follows

$$(z + \mathbf{j}t) \cdot M = M^{-1}(z + \mathbf{j}t) = [a(z + \mathbf{j}t) + b][(c(z + \mathbf{j}t) + d]^{-1},$$

where the inverse indicates the right inverse in the non commutative structure of $\mathbb{H}_3$. Note that for $t = 0$ we get the standard $\mathrm{SL}_2(\mathbb{C})$-action on the boundary $\mathbb{CP}^1$ of $\mathbb{H}_3$.

Lemma 2. *The action of $\mathrm{SL}_2(\mathbb{C})$ on $\mathbb{H}_3$ can be written in the form*

$$(z + \mathbf{j}t) \cdot M = \frac{(az + b)\overline{(cz + d)} + a\bar{c}t^2 + \mathbf{j}t}{|cz + d|^2 + |c|^2 t^2}.$$

2.3.1. *$\mathbb{H}_3$ as a parameter space for positive definite Hermitian quadratic forms.* Let

$$H(x, y) = a|x|^2 - bx\bar{y} - \bar{b}\bar{x}y + c|y|^2, a, c \in \mathbb{R}$$

be a Hermitian quadratic form with homogeneous variables $[x, y] \in \mathbb{P}^1\mathbb{C}$. Notice that the values of $H(x, y)$ are always real. Let $\Delta = ac - |b|^2$ be its discriminant. Then

$$H(x, y) = a[x - (\bar{b}/a)y]^2 + (\Delta/a)y^2,$$

hence $H(x, y) > 0$ for all (x, y) when $\Delta > 0, a > 0$. Such a form is called **positive definite.** Denote the set of all positive definite Hermitian forms by $V_{2,\mathbb{C}}^+$. There is an $\mathrm{SL}_2(\mathbb{C})$ action on $V_{2,\mathbb{C}}^+$ similar to the real case. The natural $\mathrm{SL}_2(\mathbb{R})$ equivariant inclusion $\psi : V_{2,\mathbb{R}}^+ \to V_{2,\mathbb{C}}^+$ via

$$\psi(ax^2 - 2bxy + cy^2) = a|x|^2 - bx\bar{y} - \bar{b}\bar{x}y + c|y|^2,$$

gives rise to an extension of the zero map.

Definition 1. *The zero map $\zeta : V_{2,\mathbb{C}}^+ \to \mathbb{H}_3$ is defined via*

$$(4) \qquad \zeta(a|x|^2 - bx\bar{y} - \bar{b}\bar{x}y + c|y|^2) = \frac{\bar{b}}{a} + \mathbf{j}\frac{\sqrt{\Delta}}{a}$$

Proposition 5. *The map ζ is $\mathrm{SL}_2(\mathbb{C})$ equivariant.*

The hyperbolic space $\mathbb{H}_3$ parametrizes (up to a constant factor) positive definite $(\Delta > 0, a > 0)$ Hermitian forms via the inverse map

$$\zeta^{-1}(\omega) = \zeta^{-1}(z + \mathbf{j}t) = |x|^2 - \bar{z}\bar{x}y - zx\bar{y} + (|z|^2 + t^2)|y|^2 = H_\omega,$$

while the boundary $\mathbb{CP}^1 = \mathbb{C} \cup \infty$ of $\mathbb{H}_3$ parametrizes the decomposable $(\Delta = 0)$ Hermitian forms

$$H_\beta = (x - \bar{\beta}y)(\bar{x} - \beta\bar{y}) \text{ for } \beta \in \mathbb{C}, \text{ and } H_\infty = |y|^2,$$

Just as in the case of $\mathbb{H}_2$, we have the following proposition:

Proposition 6. *Let $\overline{\mathbb{H}}_3 = \mathbb{H}_3 \cup \partial\mathbb{H}_3 = \mathbb{H}_3 \cup \mathbb{CP}^1$. The hyperbolic convex hull of $\omega_1, \omega_2, ..., \omega_n \in \overline{\mathbb{H}}_3$ parametrizes Hermitian forms $\sum_{i=1}^n \lambda_i H_{\omega_i}$ with $\lambda_i > 0$ for $i = 1, 2, ..., n$ and $\sum_{i=1}^n \lambda_i = 1$.*

The equivariant connection between the geometry of hyperbolic spaces and the algebra of positive definite forms, which extends to the boundary as well, can be expressed in the following equivariant commutative diagram:

$$
\begin{array}{ccc}
V_{2,\mathbb{R}}^+ & \xrightarrow{\ \zeta\ } & \mathbb{H}_2 \\
\downarrow & & \downarrow \\
V_{2,\mathbb{C}}^+ & \xrightarrow{\ \zeta\ } & \mathbb{H}_3
\end{array}
$$

Next, we will see how to use the equivariance of the zero map to construct a reduction method.

3. REDUCTION OF BINARY FORMS

3.1. **Julia reduction.** Let $f(x, y) \in \mathbb{R}[x, y]$ be a degree n binary form given as in Eq. (1). The form is factored as

$$(5) \qquad f(x, 1) = \prod_{i=1}^{r}(x - \alpha_i) \cdot \prod_{i=1}^{s}(x - \beta_i)(x - \bar{\beta}_i).$$

where $\alpha_1, \ldots, a_r \in \mathbb{R}$ and $\beta_1, \ldots, \beta_s$ are in the upper half complex plane, denoted by $\mathbb{H}_2$. The ordered pair (r, s) is called the *signature* of f. We associate to f the quadratic Q_f

$$(6) \qquad Q_f(x, y) = \sum_{i=1}^{r} t_i^2 (x - \alpha_i y)^2 + \sum_{j=1}^{s} 2u_j^2 (x - \beta_j y)(x - \bar{\beta}_j y),$$

where t_i, u_j are to be determined. Let $\beta_j = a_j + b_j i$, for $j = 1, \ldots, s$. 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.

Q_f is a positive definite quadratic form with discriminant $\mathfrak{D}_f$; which is expressed in terms of the root differences; see [1]. Hence, $\mathfrak{D}_f$ is fixed by all the transpositions of the roots. Indeed $\mathfrak{D}_f^n$ is an invariant of the binary form f. We define the θ_0 of f as

$$(7) \qquad \theta_0(f) = \frac{a_0^2 \cdot |\mathfrak{D}_f|^{n/2}}{\prod_{i=1}^{r} t_i^2 \prod_{j=1}^{s} u_j^4}.$$

Consider $\theta_0(t_1, \ldots, t_r, u_1, \ldots, u_s)$ as a multivariable function in the variables $t_1, \ldots, t_r, u_1, \ldots, u_s$. We would like to pick these variables such $\mathfrak{D}_f$ is minimal. This is equivalent to $\theta_0(t_1, \ldots, t_r, u_1, \ldots, u_s)$ obtaining a minimal value.

Proposition 7. *The function $\theta_0 : \mathbb{R}^{r+s} \to \mathbb{R}$ obtains a minimum at a unique point $(\bar{t}_1, \ldots, \bar{t}_r, \bar{u}_1, \ldots, \bar{u}_s)$.*

Choosing $(\bar{t}_1, \ldots, \bar{t}_r, \bar{u}_1, \ldots, \bar{u}_s)$ that make θ_0 minimal gives a unique positive definite quadratic $Q_f(x, y)$. We call this unique quadratic $Q_f(x, y)$ for such a choice of $(\bar{t}_1, \ldots, \bar{t}_r, \bar{u}_1, \ldots, \bar{u}_s)$ the **Julia quadratic** of $f(x, y)$ and denote it by $\mathcal{J}_f(x, y)$. The quantity

$$\theta_f := \theta_0(\bar{t}_1, \ldots, \bar{t}_r, \bar{u}_1, \ldots, \bar{u}_s)$$

is called the **Julia invariant**.

Lemma 3. *Let $f \in V_{n,\mathbb{R}}$. Then θ_f is an $\mathrm{SL}_2(\mathbb{R})$- invariant and $\mathcal{J}_f$ is an $\mathrm{SL}_2(\mathbb{R})$ covariant of order 2.*

Hence we have $\mathbb{V}_{n,R} \to V_{2,\mathbb{R}}^+ \to \mathbb{H}_2$ via

$$f(x,y) \to \mathcal{J}_f(x,y) \to \zeta(\mathcal{J}_f) =: \zeta(f)$$

The map

$$\zeta : V_{n,\mathbb{R}} \to \mathbb{H}_2$$

is called **Julia zero map** and it is $\mathrm{SL}_2(\mathbb{R})$-equivariant; see [2, 4]. The zero map extends to

$$\xi : V_{n,\mathbb{C}} \to \mathbb{H}_3$$

via $\xi(f) = \xi(\mathcal{J}_f) \in \mathbb{H}_3$, a point in the hyperbolic convex hull of the roots of f. The form f is called **Julia reduced** if $\xi(f)$ is in the fundamental domain $\mathcal{F}$ of $\mathrm{SL}_2(\mathbb{C})$.

If Julia quadratic preserves the height, then Julia reduction would give a form with minimal height. However, this is not true as shown in [1] for cubics and it will shown again in the coming section.

3.2. **Hyperbolic reduction.** Hyperbolic reduction was introduced in [1] when authors showed that using the hyperbolic centroid for the zero map instead of the center of mass gives different results from Julia reduction. Below, we briefly describe this reduction in detail, since it is less known than Julia reduction and provide explicit formulas how to compute the hyperbolic centroid for a set of points in the upper complex plane. For further details see [1].

Definition 2. *The hyperbolic centroid, or simply centroid,*

$$\mathcal{C}_H(\alpha_1, \alpha_2, ..., \alpha_n)$$

of the collection $\{\alpha_j \in \mathbb{H}_2 \mid j = 1, 2, ..., n\}$ is the unique point $t + \mathbf{i}u \in \mathbb{H}_2$ that minimizes

$$\sum_{j=1}^n \frac{(t - x_j)^2 + (u - y_j)^2}{u y_j}.$$

Proposition 8. *The centroid $\mathcal{C}_H = t + \mathbf{i}u \in \mathbb{H}_2$ of $\alpha_1, \alpha_2, ..., \alpha_n$ satisfies*

$$t = \sum_{i=1}^n \left(\frac{y_1 y_2 \cdots y_{i-1} y_{i+1} \cdots y_n}{\mathfrak{s}_{n-1}(y_1, y_2, ..., y_n)} \right) x_i$$

(8)
$$|\mathcal{C}_H|^2 = \sum_{i=1}^n \left(\frac{y_1 y_2 \cdots y_{i-1} y_{i+1} \cdots y_n}{\mathfrak{s}_{n-1}(y_1, y_2, ..., y_n)} \right) |\alpha_i|^2$$

$$Q_{\mathcal{C}_H}(x,y) = \sum_{i=1}^n \left(\frac{y_1 y_2 \cdots y_{i-1} y_{i+1} \cdots y_n}{\mathfrak{s}_{n-1}(y_1, y_2, ..., y_n)} \right) Q_{\alpha_i}(x,y).$$

It follows that as a point in the hyperbolic convex hull of $\alpha_1, \alpha_2, ..., \alpha_n$, the centroid is represented by the linear combination positive definite quadratic

$$Q_{\mathcal{C}_H}(x,y) = \sum_{i=1}^n \left(\frac{y_1 y_2 \cdots y_{i-1} y_{i+1} \cdots y_n}{\sigma_{n-1}(y_1, y_2, ..., y_n)} \right) Q_{\alpha_i}(x,y).$$

All equations in Eq. (8) are described in terms of the function defined by

$$\psi : \mathbb{R}^n \times \mathbb{R}^n_{>0} \mapsto \mathbb{R}$$

$$\psi\left((x_1, \ldots, x_n), (y_1, \ldots, y_n)\right) = \sum_{i=1}^{n} \left(\frac{y_1 y_2 \cdots y_{i-1} y_{i+1} \cdots y_n}{\mathfrak{s}_{n-1}(y_1, y_2, \ldots, y_n)} \right) x_i.$$

The function ψ has symmetries and is a convex linear combination of x_i's with weights that depend only on $y_1, \ldots, y_n$. It is probably a well-known and standard function in areas where symmetries and group actions are relevant.

Let $V^+_{2n, \mathbb{R}}(0, n)$ denote binary forms of degree $2n$ with real coefficients and no real roots. Every $f(x, y) \in V^+_{2n, \mathbb{R}}(0, n)$ can be factored

$$f(x, y) = \prod_{j=1}^{n} Q_{\alpha_j}(x, y)$$

where $\alpha_j = x_j + \mathbf{i} y_j$, and

$$Q_{\alpha_j}(x, y) = (X - \alpha_j Z)(X - \overline{\alpha_j} Z).$$

The **centroid zero map**

$$\xi_{\mathcal{C}} : V^+_{2n, \mathbb{R}}(0, n) \to \mathbb{H}_2$$

is defined via

$$\xi_{\mathcal{C}}(F) := \mathcal{C}_H = \mathcal{C}_H(\alpha_1, \alpha_2, \ldots, \alpha_n).$$

The form

$$\mathcal{J}_f^{\mathcal{C}} := (x - \mathcal{C}_H\, y)(x - \overline{\mathcal{C}_H}\, y)$$

$$= \sum_{j=1}^{n} \left(\frac{y_1 y_2 \cdots y_{j-1} y_{j+1} \cdots y_n}{\mathfrak{s}_{n-1}(y_1, y_2, \ldots, y_n)} \right) Q_{\alpha_j}(x, y)$$

is called the **centroid quadratic** of F.

The reduction theory based on the centroid proceeds as in the case of Julia reduction. Let $f(x, y)$ be a real binary form with no real roots. If $\xi_{\mathcal{C}}(f) \in \mathcal{F}$ then f is reduced. Otherwise, let $M \in \mathrm{SL}_2(\mathbb{R})$ such that $M^{-1}\xi_{\mathcal{C}}(f) \in \mathcal{F}$. The form f reduces to $f^M(x, y)$.

In [8] it was given a formula for computing the hyperbolic centroid:

Proposition 9. *Let $F(X, Z)$ be a totally complex form factored over $\mathbb{R}$ as below*

$$F(X, Z) = \prod_{i=1}^{n} (X^2 + a_i X Z + b_i Z^2)$$

Denote by $d_i = \sqrt{4b_i - a_i^2}$, for $i = 1, \ldots, n$ the discriminants for each factor of $f(x, y)$. Let

$$\mathfrak{s}_{n-1} = \sum_{i=1}^{r} d_1 \cdots d_{i-1} \hat{d}_i d_{i+1} \cdots d_r,$$

where $\hat{x}$ denote a missing x, and

$$\mathbf{a} = (a_1, \ldots, a_n), \quad \mathbf{b} = (b_1, \ldots, b_n), \quad \mathbf{d} = (d_1, \ldots, d_n).$$

The centroid quadratic of $f(x, y)$ is given by

$$\mathcal{J}_F^{\mathcal{C}} = \sum_{i=1}^{n} \left(\frac{d_1 d_2 \cdots d_{i-1} d_{i+1} \cdots d_n}{\mathfrak{s}_{n-1}} \right) (X^2 + a_i X Z + b_i Z^2).$$

The centroid zero map $\xi_{\mathcal{C}}(F) = t + \mathbf{i}u \in \mathbb{H}_2$ is given by

$$t = -\frac{1}{2} \sum_{i=1}^{n} \frac{d_1 \cdots d_{i-1} d_{i+1} \cdots d_n}{\mathfrak{s}_{n-1}} a_i = \psi(\mathbf{d}, \mathbf{a}),$$

$$u^2 = \frac{1}{4\mathfrak{s}_{n-1}^2} \prod_{i=1}^{n} d_i \left(\mathfrak{s}_{n-1} \sum_{i=1}^{n} d_i + \sum_{i}^{n} d_1 \cdots \hat{d}_i \cdots \hat{d}_j \cdots d_n (a_i - a_j)^2 \right)$$

$$|\xi_{\mathcal{C}}(F)|^2 = \sum_{i=1}^{n} \frac{d_1 \cdots d_{i-1} d_{i+1} \cdots d_n}{\mathfrak{s}_{n-1}} b_i = \psi(\mathbf{d}, \mathbf{b}).$$

The reduction is defined over $\mathbb{Q}(d_1, d_2, ..., d_n)$

Even for hyperbolic reduction, similarly to the Julia reduction, the main question is the same: does the reduced binary form have minimal height? An affirmative answer to this question is very unlikely (as it was the case for Julia reduction). Moreover, we would like to know which one performs better in general. This will be investigated next.

4. A DATABASE OF BINARY FORMS

Next we want to construct a database of binary forms so we can possibly discover properties of our reduction methods and design the best possible model for reduction. In building a database of binary forms we can follow two main methods.

First, we can use databases of binary forms from [5]. Such degree $n > 2$ binary forms are points in the projective space $\mathbb{P}^n$. However, because of the way such databases were constructed most of those binary forms have minimal hight and would be useless to us for illustrating reduction methods. In order to have this data $\{f\}_S$, for some index set S, useful for training, we can randomly act on each binary form with random matrices $M \in \mathrm{SL}_2(\mathbb{Z})$. The new data $\{f^M\}_S$, will not, in general, have binary forms with minimum heights. However, we can design a machine learning model based on $\{f^M\}_S$, and do the training of this model based on $\{f\}_S$.

Second , we start with roots in the hyperbolic plane $\mathbb{H}_2$. We create a database of n-gons with vertices $\alpha_1, \ldots, \alpha_r \in \mathbb{H}_2$. For simplicity of the argument here we assume we have no real roots, even though the method can be easily extended in this case. Thus our binary form will be

$$f(x, y) = \prod_{j=1}^{r} \left(x^2 - 2\operatorname{Re}(a_j)xy + ||\alpha_j||^2 y^2 \right).$$

Binary forms of this type are called *totally complex forms*. In order to have $f(x, y)$ with integer coefficients we can further assume that $\alpha_1, \ldots, \alpha_r \in \mathbb{H}_2$ are Gaussian integers.

To control the location of the polygons we can assume that the roots $\alpha_1, \ldots, \alpha_n$ are always picked between radii r_1 and r_2. This assures that we don't take ones close to the fundamental domain (so the affects of the reduction are more visible) or we don't have floating issues in the case of very large coordinates. The main question here becomes how to pick r_1 and r_2 so we can get a database of preferable size.

The number of Gaussian integers in this region is roughly $\pi(r_2^2 - r_1^2)$ is related to the famous *Gauss circle problem*. Hence, we can always have some estimate of how many points we will have in the region and therefore the number of n-gons, which

is much bigger then the number of points in the region. As you will see below, there are $37\,090\,735$ triangles for $r_1 - 1$ and $r_2 = 20$ and $8\,936\,928$ pentagons for $r_1 = 1$ and $r_2 = 7$.

These two very different approaches of creating a database of binary forms are mostly forced upon us by the strategies of building a training model.

The algebraic approach would be to ignore the geometry (roots of binary forms) and express the Julia invariant in terms of the coefficients of the form $a_0, \ldots, a_n$. Since the Julia invariant is an invariant of the form then it must be expressed in terms of such coefficients. We can design a neural network such that the loss function is precisely this invariant. This would be very effective because the minimum of the loss function would determine precisely the value of the zero map and therefore the transformation needed to get the Julia reduction of the form. There is one major problem with this approach. As Beshaj showed in her thesis [9] computation of the Julia invariant symbolically is extremely difficult even for small degree forms such as quartics, quintics, and sextics.

However, geometrically this can be done rather easily for each n-gon as we illustrate next. We can numerically compute the roots of $f(x)$ in the hyperbolic plave $\mathbb{H}_2$ including the real roots. Using the method described in Section 3.2 we can find the hyperbolic centroid of such roots. Even though this is computed numerically, we can always estimate a matrix $M \in \mathrm{SL}_2(\mathbb{Z})$ such that the hyperbolic centroid is in the fundametal region $\mathcal{F}$.

4.1. **Triangles and binary sextics.** We constructed a database of triangles for $r_1 = 1$ and $r_2 = 20$. There are 37 090 735 such triangles in $\mathbb{H}_2$. The data is organized in a dictionary as:

$$(\alpha_1, \alpha_2, \alpha_3) : [[c_0, \ldots, c_6], [x_1, y_1], [x_2, y_2]],$$

where $(\alpha_1, \alpha_2, \alpha_3)$ is the key, $[c_0, \ldots, c_6]$ are the coefficients of $f(x, y)$, $[x_1, y_1]$ is the center of mass, and $[x_2, y_2]$ the hyperbolic centroid.

Among all such triangles we are interested on the ones where the distance between the center of mass and the hyperbolic centroid are the biggest. Out of $37\,090\,735$ such triangles, the one where this distance is maximum is for the triangle with vertices

$$\alpha_1 = 1 + 19i, \alpha_2 = 2 + 19i, \alpha_3 = 19 + i.$$

The corresponding sextic is

$$f(x, y) = (x^2 - 38x + 362)(x^2 - 4x + 365)(x^2 - 2x + 362) = x^6 - 44x^5 y$$
$$+ 1325x^4 y^2 - 32280x^3 y^3 + 480964x^2 y^4 - 5809376xy^5 + 47831060y^6$$

with height $\mathrm{H}(f) = 47\,831\,060$.

The center of mass has coordinates $\left(\frac{22}{3}, 13\right)$ and the hyperbolic center $\left(\frac{52}{3}, \frac{271}{100}\right)$. To shift the center of mass to the fundamental domain we shift by seven units to the left ($z \to z - 7$) which correspond to the matrix $M = \begin{bmatrix} 1 & -7 \\ 0 & 1 \end{bmatrix}$. The Julia reduced form of $f(x, y)$ is $f^{M^{-1}} = f(x + 7, y)$, which is

$$\mathrm{Jul}\, f = (x^2 - 24x + 145)(x^2 + 10x + 386)(x^2 + 12x + 397)$$

with height $\mathrm{H}(\mathrm{Jul}\, f) = 22\,220\,090$, a significant improvement from the original height. The hyperbolic reduction would correspond to the shifting $f(x + 17, y)$ and
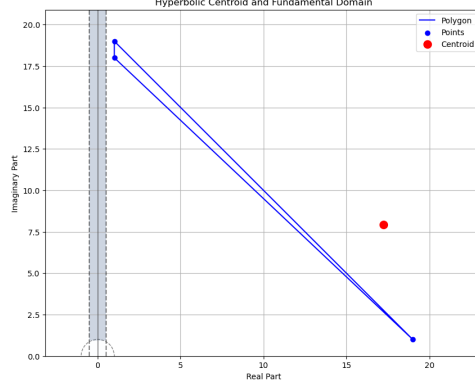
FIGURE 5. The triangle with the maximum distance between the center of mass and hyperbolic centroid among $37\,090\,735$ triangles.

give

$$\operatorname{Hyp} f = (x^2 - 4x + 5)(x^2 + 30x + 586)(x^2 + 32x + 617)$$

with a height $\operatorname{H}(\operatorname{Hyp} f) = 1\,807\,810$, a significant improvement from the Julia reduction. This is the first example that we know where the hyperbolic reduction gives a much smaller height than the Julia reduction.

However, something amazing happens here. The height continues to get smaller if we shift to the left and achieve it minimum for $f(x + 19, y)$, where the form becomes

$$f(x + 19, y) = (x^2 + 1)(x^2 + 34x + 650)(x^2 + 36x + 685)$$

which has height $\operatorname{H}(f(x + 19), y) = 447\,809$. Is this the minimal height in the Γ-orbit? Or we could ask even more, is this the minimal absolute height (i.e. the smallest height among all Γ-orbits)? Notice that no transformation via diagonal matrices would lower the height here; see [1]. Hence, very likely this is the minimal height.

A similar example where the Julia reduction was computed algebraically was given in [10, Example 1]. For the same example in [8] the hyperbolic center was computed and shown that was different from the center of mass. However, they were too close to each other that the reduction both ways held the same result. That was the reason that we looked though our large database for the example were the distance between the two centroids was maximum.

In [10] was also shown that for binary sextics with extra involutions the center of mass was always in the i-axis. That is because such sextics have roots symmetric to this axis. To avoid such cases of reducible forms we picked our triangles to by always with positive real part.

As far as we are aware, this is the first example where the two reductions are shown to give different results. This example shows that neither Julia reduction, nor hyperbolic reduction achieve minimal height. Moreover, it seems to suggest that the hyperbolic reduction is a more natural approach since it preserves better the geometry of the hyperbolic plane.

We compared both reductions from all the data for triangles between circles $r = 1$ and $r = 5$ and found out that from all 518 665 binary sextics we have:

$$\begin{aligned}
\text{Hyperbolic reduction:} &\quad 270\,997 \\
\text{Julia reduction:} &\quad 75\,993 \\
\text{Same result:} &\quad 171\,675
\end{aligned}$$

Hence, in this case the hyperbolic reduction performs considerably better than the Julia reduction. This suggests that some mixture of the two methods might be more suitable. Next, we will see how each reduction performs in the case of binary decimics.

4.2. **Pentagons and binary decimics.** We follow the same approach with the same assumptions as for the case of triangles. Hence, we want to build a database of pentagons with vertices in the hyperbolic plane and with $\text{Re}(z) > 0$. Since there will be more possible combinations in this case, we only for radius up to $r \le 7$. For each one of such pentagons we have a totally complex binary decimic.

In Fig. 6 we graph the pentagon where such distance is the maximum between all 8 936 928 pentagons for $r_1 = 1$ and $r_2 = 7$. It belongs to roots
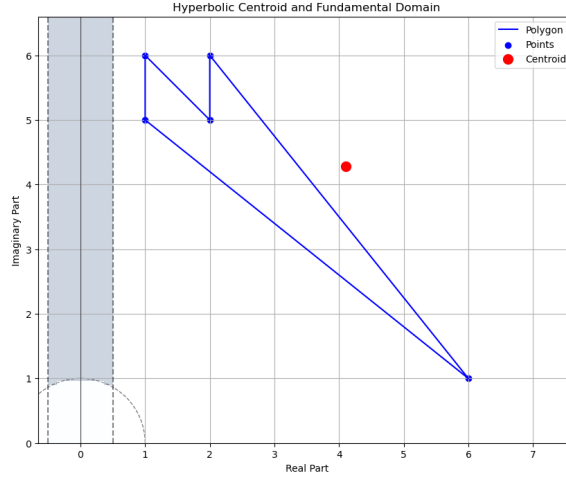

FIGURE 6. The pentagon with the maximum distance between the center of mass and hyperbolic centroid among 8 936 928 pentagons.

$$1 + 5i,\ 1 + 6i,\ 2 + 6i,\ 3 + 3i,\ 6 + i.$$

The corresponding degree ten binary form has height

$$\text{H}(f) = 25\,627\,680$$

The center of mass has coordinates $(2.6, 4.8)$ and the hyperbolic centroid is $(4, 24, 2.94)$. The shift corresponding to the Julia reduction (resp. hyperbolic reduction) is $f(x + 3, y)$ (resp. $f(x + 4, y)$) and has height $\text{H}(Jul(f)) = 3\,862\,800$ (resp. $\text{H}(\text{Hyp}\,f) = 3\,060\,000$). Hence, again the hyperbolic reduction gets a better height, but not the minimal height, which is obtained for $f(x+5, y)$ and it is $h = 2\,494, 440$. The minimal polynomial is

$$f(x) = (x^2 - 2x + 2)(x^2 + 4x + 13)(x^2 + 6x + 45)(x^2 + 8x + 52)$$

Hence, it seems as what is happening is very similar as in the case of the triangles in the sense that it is closer to the far right vertex of the n-gon, but it does not exactly at the vertex as in the case of triangles.

We compared both reductions from all the data for triangles between circles $r = 1$ and $r = 4$ and found out that from all 11 628 binary decimics we have:

$$\begin{array}{ll} \text{Hyperbolic reduction:} & 2\,367 \\ \text{Julia reduction:} & 797 \\ \text{Same result:} & 8\,464 \end{array}$$

Hence, again the hyperbolic reduction performs considerably better than the Julia reduction. For the strip $r_1 = 1$ and $r_2 = 5$ we get 278 256 sextics from which

$$\begin{array}{ll} \text{Hyperbolic reduction:} & 81\,034 \\ \text{Julia reduction:} & 33\,213 \\ \text{Same result:} & 164\,009 \end{array}$$

4.3. **Finding the minimal form.** From the above work is clear that none of the two methods will determine the minimal form in every case. Moreover, even one method is better than the other, it does not mean that is has reached the minimal form. There are two types of transformations that could be used to decrease the height of binary forms: *shifts* and *rescaling*. Shifts, $f(x + d)$ which have been discussed above will send a monic polynomial to a monic polynomial, therefore $f(x+d)$ is also primitive. However, transformations $x \to \lambda x$ for the appropriate nonzero λ can produce a binary form $f(\lambda x)$ which is not primitive. This new binary form can have smaller height less that the original forms, since when we compute the height we must divide by the content of $f(\lambda x)$.

Shifts: After shifting the form using the "best" reduction from Julia reduction or hyperbolic reduction above, we might need another *additional shift* to reach the minimal height. For our experiments above for an additional shift $d \le 3$ we always reach the minimal form, but most likely this is due to the size of our data. It is an open problem to bound the size of this additional shift. In the next section we will design a layer which determines this additional shift. It is based on the fact that while d is increasing in absolute value then the height of $f(d \pm i)$ decreases until it reaches the minimum value and then it starts increasing again.

Scaling: We can lower the height of the binary form by transformations of the form $x \to \lambda x$. This used the fact that the height of the binary form $f = [a_0 : \ldots : a_n]$ is the maximum of the absolute values $|a_i|$ when the form is primitive. Hence, if we pick λ such that we maximize the greatest common divisor of the coefficients. This was considered in [10] and also in [1] in terms of diagonal matrices. We will take a slightly different approach here.

Let $f = [a_0 : a_1 : \ldots : a_n]$. Let $\mathfrak{w} = (0, 1, 2, \ldots, n)$ be a set of heights. Consider $\mathfrak{p}_f = [a_1 : \ldots : a_n] \in \mathbb{P}_\mathfrak{w}$ as a point in the weighted projective space $\mathbb{P}_\mathfrak{w}$. We will denote by $\mathfrak{h}_\mathfrak{w}(\mathfrak{p}_f)$ the weighted height of $\mathfrak{p}_f$ and by $\mathrm{wgcd}_\mathfrak{w}(a_1, \ldots, a_n)$ the weighted greatest common divisor with respect to the weights $\mathfrak{w}$; see [11] for details. Let $m = \mathrm{lcm}(1, \ldots, n)$. Then, it was proved in [11] that

$$\mathfrak{h}(a_1, \ldots, a_n) < \mathrm{H}\left((a_1, \ldots, a_n)\right)^{\frac{1}{m}}$$

Lemma 4. *Let f and w as above and p and q integers such that*

$$p = \gcd(a_0, q). \quad where \; q = wgcd_{\mathfrak{w}}(a_1, \ldots, a_n)$$

Then $f(px, y)$ has the minimum height that can be achieved by scaling

Proof. Suppose that there is $g(x, y)$ which is obtained from $f(x, y)$ by scaling and has smaller height. Then, there exists a non-zero λ such that $f(\lambda x, y) = g(x, y)$. Hence, coefficients change as

$$f = [a_0 : \ldots : 1] \to [a_0 : \lambda a_1 : \ldots : \lambda^{n-1} a_{n-1} : \lambda^n] = g$$

That means that $\lambda | q$ and since we are assuming that the height of g is smaller than f, that implies that $\lambda | a_0$. Hence, $\lambda | p$. That completes the proof. □

Lemma above provides an algorithm how to get the form with minimum height and will b implemented as the *scaling layer* in the next section.

5. A Machine Learning Approach to Reducing Binary Forms

Determining the transformation that reduces a binary form to its minimal height remains an open and challenging problem. Historically, Julia reduction was considered the most effective method of reduction for binary forms. It generalized the reduction of quadratics, which successfully minimizes the discriminant and the height. This motivated attempts to generalize reduction to higher-degree forms. However, in contrast to quadratics, higher-degree forms involve multiple invariants, making the minimization problem more complex. Minimizing these invariants, referred to as *Reduction A* in [1], can be achieved using weighted greatest common divisors and weighted heights, as discussed in [12,13]. However, minimizing the invariants does not necessary means minimizing the coefficients, which is a complex arithmetic problem.

Despite the progress made by these approaches, neither Julia reduction nor hyperbolic reduction guarantees achieving the minimal form for binary forms of arbitrary degree. To address this limitation, we propose a novel machine learning framework designed to predict transformations that effectively minimize the height of binary forms. Our approach combines neural networks with symbolic layers to improve the model's accuracy and interpretability.

5.1. **Architecture of the Model.** The input to the model is a degree $n \geq 2$ binary form, represented as a projective point:

$$[c_0 : \ldots : c_n] \in \mathbb{P}^n.$$

The model is composed of the following layers:

Roots layer: In this layer, we compute the roots of the binary form in the upper half-plane $\mathbb{H}_2$ numerically. This provides the essential geometric data for subsequent computations. The Python code for this computation is provided below.

```
1  def roots_upper_half_plane(f, precision=10):
2      f_poly = sp.Poly(f, x)
3      roots = f_poly.all_roots()
4      upper_half_roots = []
5      for r in roots:
6          r_num = sp.N(r, precision)
7          if hasattr(r_num, 'imag') and r_num.imag > 0:
```

```
8              upper_half_roots.append(r_num)
9          return upper_half_roots
```

Hyperbolic layer: This layer computes the hyperbolic centroid of the roots in $\mathbb{H}_2$ using the formula from Prop. 9. The centroid serves as a geometric invariant that guides the reduction process.

```
1  def hyperbolic_centroid(points):
2      x_coords = [x for x, _ in points]
3      y_coords = [y for _, y in points]
4      n = len(points)
5      s_n_minus_1 = sum(prod(y_coords[:i] + y_coords[i + 1:]) for i in
            range(n))
6      t = sum(prod(y_coords[:i] + y_coords[i + 1:]) * x_coords[i] for i
            in range(n)) / s_n_minus_1
7      norm_squared = sum(
8          prod(y_coords[:i] + y_coords[i + 1:]) * (x_coords[i] ** 2 +
                y_coords[i] ** 2)
9          for i in range(n)
10      ) / s_n_minus_1
11      u = (norm_squared - t**2) ** 0.5
12      return round(t, 2), round(u, 2)
```

Direction layer: While the hyperbolic centroid provides a useful geometric indicator, it does not guarantee the minimal form. The direction layer determines the optimal shift direction in $\mathbb{H}_2$ to further reduce the height of the binary form. This step refines the reduction process by identifying the transformation that leads to the most significant height reduction.

```
1  def direction_layer(L):
2      for key, value in L.items():
3          coefficients = [int(c) for c in key.strip('[]').split(',')]
4          original_height = height(coefficients)
5          transformed_plus = [
6              sum(coefficients[j] * comb(j, i) for j in range(i, len(
                    coefficients)))
7              for i in range(len(coefficients))
8          ]
9          height_plus = height(transformed_plus)
10          transformed_minus = [
11              sum(coefficients[j] * (-1)**(j - i) * comb(j, i) for j in
                    range(i, len(coefficients)))
12              for i in range(len(coefficients))
13          ]
14          height_minus = height(transformed_minus)
15          if height_plus < original_height:
16              value.append('+')
17          if height_minus < original_height:
18              value.append('-')
19      return L
```

Scaling layer: The scaling layer handles reductions up to $\mathrm{GL}_2(\mathbb{Q})$-equivalence by applying a scaling transformation of the form $x \to \lambda x$ for some $\lambda \in \mathbb{Q}^\star$. This step

ensures that the resulting binary form achieves a minimal height with respect to $GL_2(\mathbb{Q})$-equivalence. The theoretical basis for this layer is provided by the scaling lemma (see Lem. 4).

```python
def scaling_layer(key):
    from math import gcd
    def divisors(x):
        x = abs(x)
        return [i for i in range(1, x + 1) if x % i == 0]
    key = primitive(key)
    coefficients = key
    n = len(coefficients) - 1
    c_0 = coefficients[0]
    possible_d = divisors(c_0)
    for d in possible_d[1:]:
        if c_0 % (d ** n) != 0:
            continue
        valid = True
        for i, c_i in enumerate(coefficients):
            if c_i % (d ** (n - i)) != 0:
                valid = False
                break
        if valid:
            g = [c * (d ** i) for i, c in enumerate(coefficients)]
            return primitive(g)
    return key
```

Having introduced the layers of the machine learning model, we now turn to the details of its implementation and the challenges encountered during training.

5.2. **Implementation Details.** Our implementation is designed to handle binary forms of various degrees, including degrees 5, 6, and 10, with detailed databases described in Section 4. All datasets and code will be made publicly available.

Initial attempts to use unsupervised machine learning models achieved low accuracy rates of 10–20%. However, the inclusion of symbolic layers significantly improved performance, demonstrating the value of combining neural networks with symbolic computation.

A major challenge in training the model was the lack of reliable, large-scale datasets for higher-degree binary forms that include their corresponding minimal forms. While it is straightforward to generate large datasets of binary forms, these datasets often lack the necessary ground truth for minimal reductions. To overcome this, we employed alternative methods to construct training data, combining algorithmic reduction techniques with symbolic computations to approximate minimal forms.

6. Conclusions

Binary forms have been the focus of classical mathematics and continue to be the focal point of current research; see [1, 11, 12, 14–16]. This study provides a comparative analysis of Julia reduction and hyperbolic reduction for finding equivalent binary forms with minimal coefficients. Our results demonstrate that hyperbolic reduction generally achieves better outcomes than Julia reduction, particularly for sextics and decimics. However, neither method guarantees achieving the minimal

form, highlighting the need for additional transformations. To address this, we introduced an additional shift and scaling approach that further reduces the form, offering an improved but not absolute solution.

A significant contribution of this work is the proposal of a machine learning framework to determine optimal transformations. This approach bridges traditional mathematical methods with data-driven techniques, offering a novel perspective on the problem. The success of this framework suggests that machine learning can be a valuable tool in exploring the complex landscape of binary forms, particularly in identifying patterns and relationships that are difficult to capture through classical methods alone.

Despite these advancements, certain limitations remain. Both Julia and hyperbolic reduction methods are heuristic in nature and do not guarantee a minimal form, and the proposed machine learning framework requires further development to generalize across a wider range of forms. Additionally, the reliance on computational experiments necessitates high computational resources, which may limit the scalability of the methods.

Looking forward, there are several promising directions for future research. First, enhancing the machine learning framework with larger and more diverse training datasets could improve its robustness and accuracy. Second, exploring connections between reduction theory and other areas of computational mathematics, such as lattice reduction or invariant theory, may yield new insights. Finally, developing theoretical guarantees for achieving minimal forms under specific conditions remains an open and intriguing question.

This work lays a foundation for integrating classical reduction techniques with modern computational tools, offering both practical solutions and a deeper understanding of binary forms. By combining traditional methods with machine learning, we take a step toward more effective and generalizable approaches to symbolic computation and reduction.

References

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

[2] 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.

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

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

[5] Elira Shaska and Tony Shaska, *Machine learning for moduli space of genus two curves and an application to isogeny based cryptography* (2024), available at 2403.17250.

[6] L. Beshaj, R. Hidalgo, S. Kruk, A. Malmendier, S. Quispe, and T. Shaska, *Rational points in the moduli space of genus two*, Higher genus curves in mathematical physics and arithmetic geometry, 2018, pp. 83–115. MR3782461

[7] Elira Shaska and Tony Shaska, *Polynomials, galois groups, and deep learning*, RISAT preprints (202412), available at https://www.risat.org/pdf/2024-05.pdf.

[8] A. Elezi and T. Shaska, *Reduction of binary forms via the hyperbolic centroid*, Lobachevskii J. Math. **42** (2021), no. 1, 84–95. MR4230815

[9] Lubjana Beshaj, *Integral binary forms with minimal height*, Ph.D. Thesis, 2016. Thesis (Ph.D.)–Oakland University. MR3579531

[10] _____, *Minimal integral Weierstrass equations for genus 2 curves*, Higher genus curves in mathematical physics and arithmetic geometry, [2018] ©2018, pp. 63–82. MR3782460

[11] L. Beshaj, J. Gutierrez, and T. Shaska, *Weighted greatest common divisors and weighted heights*, J. Number Theory **213** (2020), 319–346. MR4091944

[12] Sajad Salami and Tony Shaska, *Local and global heights on weighted projective varieties*, Houston J. Math. **49** (2023), no. 3, 603–636 (English).

[13] _____, *Vojta's conjecture on weighted projective varieties* (2024), available at 2309.10300.

[14] Elira Shaska and Tony Shaska, *Irreducible sextics, invariants, and their galois groups*, RISAT preprints (202412), available at https://www.risat.org/pdf/2024-07.pdf.

[15] T. Shaska, *Artificial neural networks on graded vector spaces* (2024), available at 2407.19031.

[16] 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

Department of Computer Science, Wilfrid Laurier University, Waterloo, Ontario N2L 3C5, CANADA

Email address: ikotsireas@wlu.ca

Department of Mathematics and Statistics, Oakland University, Rochester, MI, 48309, USA

Email address: tanush@umich.edu