2002

An Effective Version of Belyi's Theorem

Lily S. Khadjavi
Loyola Marymount University, lily.khadjavi@lmu.edu

Recommended Citation

Khadjavi, Lily S. “An Effective Version of Belyi’s Theorem.” Journal of Number Theory, vol. 96, no. 1, Jan. 2002, pp. 22–47.

Journal of Number Theory **96**, 22–47 (2002)
doi:10.1006/jnth.2001.2748

An Effective Version of Belyi's Theorem

Lily S. Khadjavi

Department of Mathematics, Loyola Marymount University, Los Angeles, California 90045
E-mail: lkhadjav@lmu.edu

Communicated by H. Darmon

Received October 25, 1999; revised June 15, 2001

We compute bounds on covering maps that arise in Belyi's Theorem. In particular, we construct a library of height properties and then apply it to algorithms that produce Belyi maps. Such maps are used to give coverings from algebraic curves to the projective line ramified over at most three points. The computations here give upper bounds on the degree and coefficients of polynomials and rational functions over the rationals that send a given set of algebraic numbers to the set $\{0, 1, \infty\}$ with the additional property that the only critical values are also contained in $\{0, 1, \infty\}$.

1. INTRODUCTION

In "ABC Implies Mordell" [4], Elkies proves an effective version of Mordell's conjecture assuming an effective ABC conjecture. In fact, the Mordell conjecture has already been proven, in several different ways, while the ABC conjecture remains open. However, none of these proofs is effective; that is, none produces an actual upper bound on the size of the rational points on a curve or a procedure to provably find all the rational points. Elkies makes use of a result of Belyi, which for an algebraic curve defined over $\bar{\mathbf{Q}}$ gives the existence of a covering map from the curve to the projective line ramified over at most three points [2]. Belyi's result is used, for example, in work on the inverse Galois problem. In fact, the converse holds as well and was known prior to Belyi. See [11, 13] for a more recent presentation of a proof.

In particular, Belyi provides an algorithm that given a finite subset S of $\bar{\mathbf{Q}}$ produces a non-constant function $R(x) \in \mathbf{Q}(x)$ such that both the image of S under R and the critical values of R are contained in $\{0, 1, \infty\}$. In other words, R gives a covering of $\mathbf{P}_1(\mathbf{C})$ to itself ramified only over $\{0, 1, \infty\}$ with $R(S) \subset \{0, 1, \infty\}$. To obtain such a covering map from the curve to the projective line, one can compose an arbitrary non-constant (hence surjective) rational function with a function R appropriately generated by the algorithm. Notable for us is that Belyi's proof not only demonstrates the existence of such a map but gives an explicit construction. It is this fact, that

0022-314X/02 $35.00

Belyi's theorem is constructive, which Elkies exploits. Given the apparent effectiveness of Belyi's theorem, Elkies can thus build his Mordell bound on the degree and height of the coefficients of such a covering map.

In this paper, we compute an actual upper bound on the degree and height of a polynomial over $\mathbf{Q}$ that is a Belyi function for a finite set $S \subset \bar{\mathbf{Q}}$. Our bound is a function of both the size of the set S and the maximal height of an element in S, and is proven using a modification of Belyi's original algorithm. For other presentations of this algorithm, one can see Serre's book on the Mordell–Weil Theorem [12], among others. Next, we consider the case of a rational function in $\mathbf{Q}(x)$ that is a Belyi function for S. Applying similar techniques to an algorithm that generates rational functions, we prove better upper bounds. Geometrically, the polynomial case corresponds to requiring that the map be totally ramified over infinity, and the rational function case relaxes this condition. These bounds are a first step toward realizing the ingredients in Elkies' expression.

Moreover, the bounds can be interpreted in other contexts, for example, bounding the number of edges of a Grothendieck *dessin* (see [5, 11]) given by the map. Lower bounds are also of interest, and so we note a lower bound on the degree of such a polynomial, an easy consequence of the Riemann–Hurwitz Formula. See [9] for more on lower bounds. Liţcanu also proves an upper bound on the degree of a rational Belyi function for the case $S = \{0, 1, a, \infty\}$ with $a \in \bar{\mathbf{Q}}$, independently using methods similar to the ones here.

The main results are summarized in the following series of theorems; we will in fact prove sharper versions of these.

Take S to be a finite, non-empty set in $\bar{\mathbf{Q}}$, closed under the action of the Galois group $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q})$, of cardinality s and of height H_S. See Section 2 for the definition of the height of a set, H_S, which bounds the size of each element of S, and the height of a function, $H(f)$, which bounds the coefficients.

THEOREM 1.1. *Given the set S as above, there exists a non-constant function $R(x) \in \mathbf{Q}(x)$ with $R(S) \subset \{0, 1, \infty\}$, ramified over at most $\{0, 1, \infty\}$, such that*:

- *if $s < 3$, then* $\deg(R) \leqslant 2$,
- *if $s \geqslant 3$ and $S \subset \mathbf{Q}$, then*

$$\deg(R) < 2^{s^2} H_S^{3s^2}$$

- *and otherwise if $s \geqslant 3$ and $S \not\subset \mathbf{Q}$, then*

$$\deg(R) < (4sH_S)^{9s^3 2^{s-2} s!}.$$

THEOREM 1.2. *There exists $R(x)$ as in the previous theorem such that*:

- *if* $s = 1$, *then* $H(R) \leqslant H_S$,
- *if* $s = 2$ *and* $S \subset \mathbf{Q}$, *then* $H(R) \leqslant 2H_S^2$,
- *if* $s = 2$ *and* $S \not\subset \mathbf{Q}$, *then* $H(R) \leqslant 2^2 H_S^4$,
- *if* $s \geqslant 3$ *and* $S \subset \mathbf{Q}$, *then*

$$H(R) < (2H_S)^{2^{s^2} H_S^{3s^2}}$$

- *and otherwise if* $s \geqslant 3$ *and* $S \not\subset \mathbf{Q}$, *then*

$$H(R) < (2sH_S)^{(2sH_S)^{3s!2^s s^3}}.$$

THEOREM 1.3. *Let* $B(x) = x^x$ *and let* $B_i(x) = B \circ B \circ \cdots \circ B$ *be the composition of i factors B. Let* $B_0(x) = x$. *There exists a non-constant polynomial* $P(x) \in \mathbf{Q}[x]$ *with* $P(S \cup \{zeroes\ of\ P'\}) \subset \{0, 1\}$ *such that*:

- *if* $s < 3$, *then* $\deg(P) \leqslant 2$,
- *if* $s \geqslant 3$, *and* $S \subset \mathbf{Q}$, *then*

$$\deg(P) < (B_{s-3}(2^{s-2} H_S^3))^2$$

- *and otherwise if* $s \geqslant 3$ *and* $S \not\subset \mathbf{Q}$, *then*

$$\deg(P) < (B_{s-3}((16sH_S)^{9s2^{s-3}s!}))^2.$$

THEOREM 1.4. *Let* $G(x) = x^{2x}$ *and let* $G_i(x) = G \circ G \circ \cdots \circ G$ *be the composition of i factors G with* $G_0(x) = x$. *Set* $M < (2^3 sH_S)^{s2^s s!}$. *Then there exists* $P(x)$ *as in the previous theorem, such that*:

- *if* $s = 1$, *then* $H(P) \leqslant H_S$,
- *if* $s = 2$ *and* $S \subset \mathbf{Q}$, *then* $H(P) \leqslant 2H_S^2$,
- *if* $s = 2$ *and* $S \not\subset \mathbf{Q}$, *then* $H(P) \leqslant 2^2 H_S^4$, *and*
- *otherwise if* $s \geqslant 3$, *then* $H(P) < G_{s-3}(M)^{3G_{s-3}(M)^2}$.

Although we will show slightly sharper bounds in Sections 3 and 4, these give an indication of the behavior relative to S of the degree and height obtained following Belyi's algorithm. For example, if S contains many rational as opposed to algebraic elements, one achieves better bounds.

The paper is arranged as follows. In Section 2, we develop the height machinery that will serve us in all the computations of the following sections. Section 3 first presents the algorithm of Belyi's original paper [2], slightly modified (with little effect on the bounds) to produce polynomials. Next, the height properties are applied to analyzing the algorithm, and given a starting set of ramification points, we prove upper bounds for the degree and the coefficients of the polynomial. The bulk of the section is devoted to a rather detailed computation of the effect of the algorithm on the heights of the ramification points in the first stage, as these results return in the cases of both the coefficient and the rational function bounds.

The computations in Section 4 parallel these and are presented in less detail. First, we give an algorithm for producing Belyi maps that are rational functions. No longer requiring that our map be a polynomial, we profit greatly with improved upper bounds, again proving results first for the degree and then the coefficients of the map. Finally, in Section 5, we make some basic notes about lower bounds; we also make some remarks and give some simple examples related to upper bounds and elliptic curves.

2. HEIGHTS

2.1. Definition and Conventions. Let $\bar{\mathbf{Q}}$ denote an algebraic closure of the rationals, $\mathbf{Q}$, and α denote an algebraic number in $\bar{\mathbf{Q}}$. For any number field K, we can choose a set of normalized valuations of K in the following way. Each valuation is either an extension of the ordinary absolute value on $\mathbf{Q}$ or else an extension of a p-adic valuation. To choose a normalization, fix an absolute value $||\cdot||_v$, induced by an embedding σ of K into $\mathbf{C}$, to be $|\sigma(\cdot)|^{[K_v\,:\,\mathbf{Q}_v]}$, where K_v is the completion of K with respect to v and $|\cdot|$ is the usual absolute value on $\mathbf{C}$, i.e., $|2| = 2$. Then $[K_v : \mathbf{Q}_v]$ is either 1 or 2 as K_v is $\mathbf{R}$ or $\mathbf{C}$, and we call v real or complex accordingly. We require that the product formula holds: for all $\alpha \in K^*$,

$$\prod_v ||\alpha||_v = 1.$$

It then follows that for each non-archimedean v, induced by a prime ideal P in the ring of integers of K, we have

$$||\alpha||_v = \mathbf{N}(P)^{-\mathrm{ord}_P(\alpha)},$$

where $\mathbf{N}(P)$ is the norm of P, i.e., the size of the residue class field.

We define the height, $H(\alpha)$, of α to be

$$H(\alpha) = \left(\prod_v \sup(1, ||\alpha||_v)\right)^{1/[K:\mathbf{Q}]}$$

with the product running over the set of normalized valuations v of K.

It is a basic property of heights that $H(\alpha)$ is independent of the choice of K. (See [8, 12] for more on such heights and valuations.)

Similarly, we define the height of a polynomial, $H(f)$. Let $f(x) = \sum_{i=0}^{n} a_i x^i$ where the a_i are algebraic numbers, and choose a number field K containing the a_i. For example, one may take $K = \mathbf{Q}(a_0, \ldots, a_n)$. Then we define

$$H(f) = \left(\prod_v \sup(1, \|a_0\|_v, \ldots, \|a_n\|_v) \right)^{1/[K:\mathbf{Q}]},$$

again with the product running over the set of normalized valuations of K. As above, the definition is independent of the choice of K.

It is common to define a similar height without the 1 in the supremum, which one might refer to as a projective height, as in that case one would have $H(f) = H(\lambda f)$ for any scalar λ. However, because we require effective bounds on f, here we use an "affine" height, which captures any such scaling (see Remark 2.1).

Finally, consider the case of a rational function, h. Let K be the smallest number field such that $h(x) \in K(x)$. We shall define the height $H(h)$ of $h(x)$ by

$$H(h) = \min_{f,g} \{\max \{H(f), H(g)\}\},$$

where the minimum runs over polynomials f and g with $f(x), g(x) \in K[x]$, relatively prime in $K[x]$, such that $h(x) = f(x)/g(x)$.

Remark 2.1. A disadvantage of this definition is that given some $h(x) \in K(x)$, it may not be easy to immediately compute its height, $H(h)$. However, the main goal is to have a height function which allows for effective bounds, which this one does. In particular, for heights of algebraic numbers, there is the useful result known as Northcott's Theorem (for a proof see [12]): Given a fixed constant c and a number field K, there are only finitely many algebraic numbers in K of height less than c. Similarly, one can show that for a fixed constant d, there are only finitely many polynomials in $K[x]$ and rational functions in $K(x)$ with degree less than d and height less than c. Thus bounding the height and degree of a function over a number field is effective in the sense that only finitely many functions will satisfy that bound.

2.2. Lemmas on Heights. We will use the following "library" of bounds. Let α and β denote algebraic numbers and f and g denote non-constant polynomials of degree n and m, respectively, with algebraic coefficients. The bounds follow from standard properties of valuations and heights, although we note that [8, 12] use a projective rather than affine height. Nevertheless, the arguments are straightforward; detailed proofs are in [7].

Property 2.1. *$H(f(\alpha)) \leqslant (n+1)H(\alpha)^n H(f)$.*

Property 2.2. *$H(f') \leqslant nH(f)$, where f' is the derivative of f.*

Property 2.3. *If $\beta_1, \dots, \beta_n$ are the zeroes of f, then*

$$\prod_{i=i}^{n} H(\beta_i) \leqslant 2^n H(f).$$

Property 2.4. *Let f_i be a polynomial of degree n_i. Then*

$$H\left(\prod_{i=1}^{k} f_i\right) \leqslant \prod_{i=1}^{k} (n_i+1) \prod_{i=1}^{k} H(f_i).$$

Property 2.5. *If f is the minimal polynomial of β, then*

$$H(f) \leqslant 2^n H(\beta)^n.$$

Remark 2.2. Although for projective heights this sort of relation is proven in a manner similar to Property 2.3 (see [8]), here it follows immediately as a special case of Property 2.4. Take $f_i = x - \beta_i$ where the β_i range over the Galois conjugates of β, and note that the heights of Galois conjugates are equal.

Property 2.6. *If f and g are monic of degree n and m, respectively, then*

$$H(f \circ g) \leqslant 2^{2n}(m+1)^n H(g)^n H(f).$$

The remaining height properties involve polynomials of specific form.

Property 2.7. *If $\alpha_1, \dots, \alpha_{n-1}$ are the zeroes of $f'(x)$, then*

$$H\left(\prod_{i=1}^{n-1} (x - f(\alpha_i))\right) \leqslant 2^{n^2-1}(n+1)^{n-1} n^n H(f)^{2n-1}.$$

Property 2.8. *If $h(x) = \frac{x-\alpha_1}{\alpha_2-\alpha_1}$, where α_1 and α_2 are algebraic numbers, then*

$$H(h) \leqslant 2H(\alpha)^2$$

with $H(\alpha) = \max\{H(\alpha_1), H(\alpha_2)\}$; and

$$H(h \circ f) \leqslant 2H(f)H(h).$$

Property 2.9. *If $j(x) = \beta x$, then $H(j \circ f) \leqslant H(j)H(f)$.*

PROPERTY 2.10. *If* $k(x) = x^a(1-x)^b$, *then*

$$H(k \circ f) \leqslant 2^b(n+1)^{a+b} H(f)^{a+b}$$

and

$$H(k) \leqslant 2^{\deg(k)}.$$

3. POLYNOMIAL CASE

3.1. Notation and Algorithm. In this section, we fix the notation and describe the algorithm that we will work with for the polynomial bound. Within the algorithm, the functions are detailed in a way that is useful later for computing the bounds.

Let S be a finite, non-empty set of algebraic numbers closed under Galois action, i.e., given any set of algebraic numbers, construct S by adjoining all elements which are Galois conjugates of elements in the original set. Let r be the number of elements of S in $\mathbf{Q}$ and let t be the number of distinct conjugacy classes of elements of S in $\bar{\mathbf{Q}} \backslash \mathbf{Q}$.

Set $D = \sum_{i=1}^{t} d_i$, where d_i is the size of each such conjugacy class, so we have

$$\#S = D + r.$$

Let $H(\alpha)$ denote the height of α and $H(f)$ denote the height of the polynomial f, both defined above, and set

$$H_S = \max_{\alpha \in S} \{H(\alpha)\}.$$

Belyi's algorithm to construct a polynomial $P(x) \in \mathbf{Q}[x]$ with the property that $P(S) \subset \{0,1\}$ and that $P(\{\text{zeroes of } P'\}) \subset \{0,1\}$ is naturally divided into two stages. (With a minor adaptation of the original algorithm in [2], we produce polynomials, rather than rational functions. As noted in the introduction, geometrically this condition corresponds to restricting to maps which are totally ramified over infinity.) In the first stage, the set of ramification points are mapped from $\bar{\mathbf{Q}}$ to $\mathbf{Q}$, without necessarily reducing the cardinality of the set, and in the second stage the set of ramification points is reduced to $\{0,1\}$.

Stage I: Passing from $\bar{\mathbf{Q}}$ *to* $\mathbf{Q}$: Let $f_0(x)$ be the minimal polynomial for the set of all irrational elements of S, so $f_0(x)$ is of degree D. We proceed inductively. Let $f_i(x)$ be the minimal polynomial of the set $\{f_{i-1}(a) : f'_{i-1}(a) = 0\}$. Observe that $f_i(x)$ is of degree strictly less than

$f_{i-1}(x)$, so we have a set $\{f_k(a) : f'_{k-1}(a) = 0\} \subset \mathbf{Q}$ after k is at most $D - 2$. Let

$$F_S(x) = f_k \circ f_{k-1} \circ \cdots \circ f_0.$$

Stage II: *Passing from* $\mathbf{Q}$ *to* $\{0, 1\}$. We now work with the set $T \subset \mathbf{Q}$ defined by

$$T = \{F_S(a) : F'_S(a) = 0\} \cup \{F_S(S)\}.$$

Let $T = \{\beta_1, \beta_2, \ldots, \beta_n\}$, where $n \leqslant \# S$, ordered such that $\beta_i < \beta_{i+1}$. Scale T to the unit interval $[0, 1]$ with the map

$$h_0(x) = \frac{x - \beta_1}{\beta_s - \beta_1}.$$

(If $n \leqslant 2$, we are done.) Inductively, we will map $\{0, 1, \beta\}$ to some $\{0, \beta'\}$ and rescale $n - 2$ times, as follows.

One can represent any rational in $h_0(T)$ as a quotient $\frac{a}{a+b}$ with integral a and b; in particular, choose the rational *of maximal height* in the set. This choice is not necessary for Belyi's algorithm but is useful for computing our bound. Then let

$$g_1(x) = x^a(1 - x)^b.$$

Note that because T is scaled to $[0, 1]$, we have that g_1 is a polynomial. Note also that the critical values of $g_1(x)$ are exactly 0 and $g_1(\frac{a}{a+b})$, and one can check that this last value is of largest height in $g_1(h_0(T))$, that is, $g_1(\frac{a}{a+b}) = \max g_1(h_0(T))$. Hence we rescale our new set to the unit interval with

$$h_1(x) = \frac{x}{g_1(\frac{a}{a+b})}.$$

Next choose the rational of maximal height in the rescaled set, construct a corresponding polynomial g_2 of the same form as g_1, and rescale with the corresponding h_2. Repeating this construction, for T of size n, the process terminates after $n - 2$ steps. In general, it is always true that $h_i(x)$ will have a factor of $g_i(\frac{c}{c+d})$, where $\frac{c}{c+d}$ is the rational of maximal height which was chosen to construct g_i.

Let

$$F_T(x) = h_{n-2} \circ g_{n-2} \circ h_{n-3} \circ \cdots \circ g_1 \circ h_0(x).$$

One can then check that $P(x) = F_T \circ F_S$ has the desired ramification properties.

3.2. Bounding the Degree. We first compute a bound on the degree of $P(x)$. It is clear after Stage I that F_S, which is the composition of f_i of descending degrees, is of degree at most $D!$. The problem, then, is to bound the degree of F_T. As the h_i are linear, it is the g_i that remain to be bounded. Recall that the g_i are of the form $x^a(1-x)^b$, i.e., of degree $a+b$, where the size of $a+b$ depends on a rational that has been scaled between 0 and 1. The central idea of the proof rests on the following simple observation: the degree of g_i is the height of this rational. (Note that $H(\frac{p}{q}) = \sup\{|p|, |q|\}$ for $p, q \in \mathbf{Z}$.)

Thus we need, at each step of Stage II, to bound the heights of the points besides 0 and 1, starting with a bound on the heights of the elements of T. Let

$$H_T = \max_{\beta \in T} \{H(\beta)\}.$$

PROPOSITION 3.1. *If S is as above, of height H_S and of size $s = D + r$, where D is the number of irrational elements of S, and with T as above of height H_T, then if $D \neq 0$, one has*

$$H_T \leqslant 48(2^2 D H_S)^{3D2^{D-3}D!}.$$

Otherwise if $D = 0$, then $H_T = H_S$.

Proof. For the case $D = 0$, one would simply take $F_S(x) = x$ and continue to Stage II of the algorithm with $T = S$. In general, H_T is the maximum of $H(F_S(S))$ and $H(F_S(\text{zeroes of } F_S'))$. Thus, we must bound first $H(F_S)$ and then both $H(F_S')$ and the heights of the zeroes of F_S', which we do in the following series of lemmas. In the first two, we bound the composition factors f_i that make up F_S. In Lemmas 3.1 and 3.4, we compute a bound on F_S and then the image of S, $F_S(S)$. Finally, in Lemma 3.5, we consider the critical values of F_S, hence bounding H_T and completing the proof of the proposition. ∎

Note that if $D \neq 0$, we have in fact that $D \geqslant 2$, since S is closed under Galois action. We may set $K = \mathbf{Q}(S)$.

LEMMA 3.1. *For f_i as above, of degree d_i,*

$$H(f_0) \leqslant D^t 2^D H_S^D$$

and

$$H(f_i) \leqslant 2^{d_i^2} d_i^{2d_i} H(f_{i-1})^{2d_i}.$$

Proof. We first bound $H(f_0)$. Let $\alpha_1, \ldots, \alpha_t$ be a list of conjugacy class representatives of S. Write

$$f_0 = \prod_{i=1}^{t} m_{\alpha_i},$$

where m_{α_i} is the minimal polynomial for each conjugacy class. From Property 2.5, we know

$$H(m_{\alpha_i}) \leqslant 2^{d_i} H(\alpha_i)^{d_i}.$$

Then by Property 2.4, we have

$$H(f_0) \leqslant \prod_{i=1}^{t} (d_i + 1) \prod_{i=1}^{t} (2^{d_i} H(\alpha_i)^{d_i}).$$

Since $H_S \geqslant H(\alpha_i)$, if $t > 1$ (so $D \geqslant d_i + 1$), we obtain

$$H(f_0) \leqslant D^t 2^D H_S^D.$$

If $t = 1$, then $f_0 = m_\alpha$ and $H(f_0) \leqslant 2^D H_S^D$ by Property 2.5, so the inequality still holds.

Next, we construct f_i. Recall that f_i is the minimal polynomial of the image of the zeroes of f'_{i-1} under f_{i-1}, that is,

$$f_i = \prod_{k=1}^{d-1} (x - f_{i-1}(\beta_k)),$$

where d is the degree of f_{i-1} and where the β_k run over the zeroes of f'_{i-1}. (Note that $f_i \in \mathbf{Q}[x]$ because all conjugates of $f_{i-1}(\beta_k)$ are in the product.) Thus applying Property 2.7 gives

$$H(f_i) \leqslant 2^{d^2-1} (d+1)^{d-1} d^d H(f_{i-1})^{2d-1}.$$

For simplicity in later computations, round this up to:

$$H(f_i) \leqslant 2^{d^2} d^{2d} H(f_{i-1})^{2d}. \qquad \blacksquare$$

LEMMA 3.2.

$$H(f_i) \leqslant H(f_0)^{2^i D!/(D-i)!} 2^{\sum_{k=1}^{i} (D-k+1)^2 2^{i-k} (D-k)!/(D-i)!} \times \prod_{k=1}^{i} (D-k+1)^{2^{i-k+1}(D-k+1)!/(D-i)!}$$

and

$$H(F_S) \leqslant 2^{2((D-1)!+(D-2)!+\cdots+2!)} \prod_{i=1}^{D-2} \left(\frac{D!}{(D-i)!}+1\right)^{(D-i)!} \prod_{i=0}^{D-2} H(f_i)^{(D-i-1)!}.$$

Proof. Bounding $H(f_i)$ follows inductively from the previous inequality on f_i, with f_0 of degree D; combining like terms yields the desired result.

Next, for any monic polynomials f_i of degree d_i, we can bound the polynomial $F_k = f_k \circ f_{k-1} \circ \cdots \circ f_0$, with repeated applications of Property 2.6. This yields

$$H(F_k) \leqslant 2^{2\sum_{i=1}^{k} (d_i d_{i+1} \cdot \ldots \cdot d_k)} \prod_{i=1}^{k} (d_{i-1} d_{i-2} \cdot \ldots \cdot d_1 d_0 + 1)^{d_i d_{i+1} \cdot \ldots \cdot d_k}$$
$$\times \prod_{i=0}^{k} H(f_i)^{d_0 d_1 \cdot \ldots \cdot d_k / d_0 d_1 \cdot \ldots \cdot d_i}.$$

Given $\deg(f_i) = D - i$, we obtain the desired inequality for F_{D-2}, which bounds F_S. (Note that if $D = 2$, then $F_S = f_0$.) ∎

LEMMA 3.3.

$$H(F_S) < 2^{D2^{D-1}D!} (DH_S)^{D2^{D-2}D!}.$$

Proof. We apply the two inequalities of the previous lemma and simplify the resulting expression. For example,

$$\begin{aligned} &2^{2((D-1)!+(D-2)!+\cdots+2!)} \\ &\quad = 2^{2(D-1)!(1+1/(D-1)+\cdots+1/[(D-1)(D-2)\ldots 3])} \\ &\quad \leqslant 2^{2^2(D-1)!}. \end{aligned}$$

To bound the product

$$\begin{aligned} &\prod_{i=1}^{D-2} \left(\frac{D!}{(D-i)!}+1\right)^{(D-i)!} \\ &\quad = (D+1)^{(D-1)!} (D(D-1)+1)^{(D-2)!} \cdot \ldots \cdot (D(D-1)(D-2) \cdot \ldots \cdot 3 + 1)^{2!}, \end{aligned}$$

note that each factor except the first, $(D+1)$, is less than D^i. Then taking logs to the base D, we get

$$(D-1)!\left(\log(D+1) + \frac{2}{D-1} + \frac{3}{(D-1)(D-2)} + \cdots + \frac{D-2}{(D-1)(D-2)\ldots 3}\right).$$

For $D \geqslant 3$, this is

$$\leqslant (D-1)!\left(\frac{3}{2} + (D-3)\right) \leqslant D!$$

and hence the product is bounded by $D^{D!}$. If $D = 2$, then $F_S = f_0$, and the lemma holds trivially.

However, the main contribution to $H(F_S)$ comes from the product of the terms of the form $H(f_i)^{(D-1-i)!}$. Using Lemma 3.2 to write these in terms of $H(f_0)$ and simplifying, pulling out the dominant term as in the previous examples but with a bit more algebra, we find that

$$\prod_{i=0}^{D-2} H(f_i)^{(D-i-1)!} < 2^{D!D2^{D-2}} D^{3\cdot 2^{D-3}D!} H(f_0)^{2^{D-2}D!}.$$

In fact, the roundoff is great enough to easily absorb the two previous contributions of $2^{2^2(D-1)!}$ and $D^{D!}$. We substitute for $H(f_0)$ using Lemma 3.1, which says $H(f_0) \leqslant D^t 2^D H_S^D$. Since $D \geqslant 2$, we have that t, the number of conjugacy classes in $S \backslash \mathbf{Q}$, is at most $D/2$. Combining all these, we obtain

$$H(F_S) < 2^{D2^{D-1}D!}(DH_S)^{D2^{D-2}D!}. \quad \blacksquare$$

From here on we omit most of the detail of algebraic computations.

Lemma 3.4.

$$H(F_S(S)) \leqslant (D!+1)2^{D2^{D-1}D!} D^{D(2^{D-2})D!} H_S^{D(2^{D-2})D!+D!}.$$

Proof. We apply Property 2.1, noting F_S has degree $D!$, so

$$H(F_S(S)) \leqslant (D!+1)(H_S)^{D!} H(F_S).$$

The previous lemma bounds $H(F_S)$, completing the proof. (In fact the factor $D!+1$ could be absorbed in earlier roundoff, for example in the exponent of 2.) $\blacksquare$

Lemma 3.4 gives us a potential bound for H_T, where $T = \{\text{critical values of } F_S\} \cup \{F_S(S)\}$. Thus for Proposition 3.1, it remains only to bound the height of the critical values.

Lemma 3.5. *If U is the set of zeroes of F_S', then*

$$H(F_S(U)) \leqslant 48(4DH_S)^{3D2^{D-3}D!}.$$

Proof. Let U_i denote the zeroes of f_k'. By a simple application of the chain rule, one finds that

$$F_S(U) = f_{D-2}(U_{D-2}) \cup f_{D-2} \circ f_{D-1}(U_{D-1}) \cup \cdots \cup f_{D-2} \circ f_{D-1} \circ \cdots \circ f_0(U_0).$$

Note, however, by construction of the f_i, that

$$f_i(f_{i-1}(U_{i-1})) = 0.$$

Thus, one has

$$F_S(U) = f_{D-2}(U_{D-2}) \cup \{0\} \cup f_{D-2}(0) \cup f_{D-2} \circ f_{D-1}(0) \cup \cdots \cup f_{D-2} \circ f_{D-1} \circ \cdots \circ f_2(0).$$

Because the bound on $H(F_S)$ is greater than or equal to $H(f_{D-2} \circ \cdots \circ f_k)$, with $0 \leqslant k \leqslant D-2$, following the same bounding techniques we would find all of these to be of smaller height than $F_S(S)$ except for possibly $f_{D-2}(U_{D-2})$. Hence for our purposes it suffices to bound $H(f_{D-2}(U_{D-2}))$ and check that this is greater than or equal to $H(F_S(S))$, as given in Lemma 3.4.

First, we bound U_{D-2} in terms of f_{D-2}. Recall that $\deg(f_{D-2}) = 2$, so by Property 2.2, $H(f_{D-2}') \leqslant 2H(f_{D-2})$ and by Property 2.3, $H(U_{D-2}) \leqslant 2H(f_{D-2}')$. Combining these with Property 2.1, we have that

$$H(f_{D-2}(U_{D-2})) \leqslant 3 \cdot 2^4 \cdot H(f_{D-2})^3.$$

Lemma 3.2 gives a bound for $H(f_i)$, and using Lemma 3.1 for $H(f_0)$, we find

$$H(f_{D-2}) < 2^{D2^{D-2}D!}(DH_S)^{D2^{D-3}D!}. \qquad \blacksquare$$

Note that the bound from Lemma 3.4 is smaller than the one we get here, so Lemma 3.5 also gives the upper bound on H_T, proving Proposition 3.1.

Now we are ready to bound the degrees of the maps in Stage II of the algorithm.

Proposition 3.2. *Let $G(x) = x^{2x}$ and let $G_i(x) = G \circ G \circ \cdots \circ G$ be the composition of i factors G with $G_0(x) = x$. For S, H_S, H_T as above, with $F_T = h_{n-2} \circ g_{n-2} \circ h_{n-3} \circ \cdots \circ g_1 \circ h_0$, if the cardinality of S is $s \geqslant 3$, then with $n \leqslant s$, we have*

$$\deg(F_T) \leqslant \prod_{i=0}^{s-3} G_i(2H_T^3).$$

Proof. Recall that the h_i are linear and the degree of each map g_i depends on the height of the point generating it. Let us observe what g_i and h_i do to the height—which is bounded by the maximal denominator—at each step with $i \geqslant 1$, as this will determine the degree of g_i. At this stage, we assume each element is in $[0, 1]$. Given $\alpha = \frac{c}{c+d}$, $g_i(x) = x^a(1-x)^b$, with $a + b \geqslant c + d$ (by our choice of maximal height each time to generate g_i), we have

$$g_i(\alpha) = \frac{c^a d^b}{(c+d)^{a+b}},$$

noting, in particular, the denominator

$$(c+d)^{a+b} \leqslant (a+b)^{a+b}.$$

Thus applying g_i takes the maximal height H of the set to at most H^H. As $h_i(x) = \frac{x}{\alpha}$, one can easily check that applying h_i at most squares the height of the set. This implies that

$$\deg(g_i) \leqslant (\deg(g_{i-1})^{\deg(g_{i-1})})^2$$

for $i > 1$.

Define

$$G(x) = (x^x)^2$$

and

$$G_i(x) = G \circ G \circ \cdots \circ G,$$

the composition of i factors G. Let $G_0(x) = x$. Then $\deg(g_i)$ is bounded by $G_{i-1}(\deg g_1)$. We can thus bound each g_i, of which there are at most $s - 2$, inductively.

So, it remains only to bound $\deg(g_1)$. Given H_T, we first scale by

$$h_0(x) = \frac{x - \beta_1}{\beta_n - \beta_1}.$$

By Property 2.8, the height of h_0 is $\leqslant 2H_T^2$. Hence by Property 2.1,

$$H(h_0(T)) \leqslant 2H_T^3.$$

Thus the maximal denominator in the scaled set is of height at most $2H_T^3$, i.e.,

$$\deg(g_1) \leqslant 2H_T^3. \qquad \blacksquare$$

Given Propositions 3.1 and 3.2, we can now bound the degree of P in terms of the set S as the product of $\deg(F_S)$ and $\deg(F_T)$.

THEOREM 3.1. *Let* $G(x) = x^{2x}$ *and let* $G_i(x) = G \circ G \circ \cdots \circ G$ *be the composition of* i *factors* G *with* $G_0(x) = x$. *Given a finite, non-empty set* $S \subset \bar{\mathbf{Q}}$, *closed under Galois action, of cardinality* s, *with* D *irrational elements, and of height* H_S, *there exists a non-constant polynomial* $P(x) \in \mathbf{Q}[x]$ *with* $P(S \cup \{\textit{zeroes of } P'\}) \subset \{0, 1\}$ *such that*:

- *if* $s < 3$, *then* $\deg(P) \leqslant 2$,
- *if* $s \geqslant 3$ *and* $S \subset \mathbf{Q}$, *then*

$$\deg(P) < \prod_{i=0}^{s-3} G_i(2H_S^3)$$

- *and otherwise if* $s \geqslant 3$ *and* $S \not\subset \mathbf{Q}$, *then*

$$\deg(P) < \prod_{i=0}^{s-3} G_i((2^4 DH_S)^{3^2 D 2^{D-3} D!}).$$

Proof. The existence of such a polynomial P follows from Belyi's theorem. First, assume $s \geqslant 3$. From Proposition 3.2, we have

$$\begin{aligned}\deg(P) &\leqslant D! \deg(F_T)\\ &\leqslant D! \prod_{i=0}^{s-3} G_i(2H_T^3),\end{aligned}$$

where $G(x) = x^{2x}$ and $G_i(x) = G \circ G \circ \cdots \circ G$ the composition of i factors G. If $D < 2$, then $H_T = H_S$. Otherwise, from Proposition 3.1,

$$H_T \leqslant 48(2^2 DH_S)^{3D2^{D-3}D!}.$$

It follows that

$$\deg(P) < \prod_{i=0}^{s-3} G_i(2^{3^2 D 2^{D-1} D!}(DH_S)^{3^2 D 2^{D-3} D!}),$$

where the desired inequality follows from observing that $D!$ and the constant factors can be absorbed into an overestimate for H_T. If D is much less than s, this expression in terms of D gives a better bound than the theorem of the Introduction.

The case $s \leqslant 2$ is much shorter: if $S \subset \mathbf{Q}$, one linear rescaling step suffices to send S to $\{0,1\}$, so $\deg(P) = 1$. Otherwise we have $S \subset \bar{\mathbf{Q}} \backslash \mathbf{Q}$, and a quadratic polynomial will map S to $\mathbf{Q}$ and then a linear rescaling to $\{0,1\}$, so $\deg(P) = 2$. ∎

As we are primarily interested in the dominating terms, rather than a detailed bound, for the theorem we could bound P even more crudely: let $B(x) = x^x$, with $B_i(x)$ the composition as before, so for $S \subset \mathbf{Q}$, we can write

$$\deg(P) < \prod_{i=0}^{s-3} B_i(2^{s-2} H_S^3),$$

noting $D \leqslant s$. Since the final B_i is so far greater than product of all the previous ones, we can even write

$$\deg(P) < (B_{s-3}(2^{s-2} H_S^3))^2.$$

Doing the same for the bound in the case $S \not\subset \mathbf{Q}$, we have

$$\deg(P) < (B_{s-3}(2^{s-2} \cdot 2H_T^3))^2.$$

Hence, after some algebraic manipulation,

$$\deg(P) < (B_{s-3}(2^{3^2 s 2^{s-1} s!} (DH_S)^{3^2 s 2^{s-3} s!}))^2,$$

which proves the theorem of the Introduction.

Essentially, one can think of the input to the function B_i as the buildup of the height and degree from Stage I of the algorithm and the main growth from the B_i itself as the result of composing polynomials of the form $x^a(1-x)^b$ in Stage II.

3.3. Bounding the Coefficients. The arguments for bounding the coefficients of P are similar to those for the degree, so we present them in less detail. We have already bounded the height of F_S (Lemma 3.3), so it is easy to bound $H(h_0(F_S))$, the height of the polynomial after the first rescaling in Stage II. Thus, it remains to bound the heights of the rest of the polynomials used in Stage II. Because these polynomials are all linear or of the special form $x^a(1-x)^b$, we then apply Properties 2.9 and 2.10 to bound their compositions.

Before examining the most general case, we note that for S of cardinality $s < 3$, as in the case of the degree, $H(P)$ is easy to bound. If $s = 1$, then P is linear and $H(P) \leqslant H_S$ by Property 2.9. If $s = 2$ and $S \subset \mathbf{Q}$, then $H(P) \leqslant 2H_S^2$, by Property 2.8. Finally, if $s = 2$ and $S \not\subset \mathbf{Q}$, a straightforward computation composing the minimal polynomial for S with an appropriate linear rescaling gives a polynomial P with $H(P) \leqslant 2^2 H_S^4$.

Thus we consider $s \geqslant 3$. Assume in the following lemmas that D is nonzero, so $D \geqslant 2$. (Otherwise take F_S to be the identity, H_S in place of H_T, and without loss of generality, s in place of D.)

Lemma 3.6. *For h_0 as in the algorithm, we have*

$$H(h_0(F_S)) < 2^{3D2^D D!}(DH_S)^{3D2^{D-1}D!}.$$

Proof. By Property 2.8, we have $H(h_0(F_S)) \leqslant 2H(F_S)H(h_0)$ and $H(h_0) \leqslant 2H_T^2$. We have proved that H_T is bounded by $48(2^2 DH_S)^{3D2^{D-3}D!}$ in Proposition 3.1, and $H(F_S)$ is bounded by $2^{D2^{D-1}D!}(DH_S)^{D2^{D-2}D!}$ in Lemma 3.3, so combining these and simplifying we have

$$H(h_0(F_S)) < 2^{3D2^D D!}(DH_S)^{D2^D D!}. \qquad \blacksquare$$

Lemma 3.7. *For h_i and g_i as in the algorithm, $i \geqslant 1$, we have*

$$H(h_i) \leqslant \deg(g_i)^{\deg(g_i)}$$

and

$$H(g_i) \leqslant 2^{\deg(g_i)}.$$

Proof. Recall that $h_i(x) = \frac{x}{g_i(\frac{a}{a+b})}$, with $a + b \geqslant a$ and $\deg(g_i) = a + b$. Hence this has height bounded by $(a+b)^{a+b}$. The second inequality is Property 2.10. $\blacksquare$

But now all the machinery is in place to bound $H(P)$ inductively. For example, by Property 2.10,

$$H(g_1(h_0(F_S))) \leqslant (2(D! + 1)H(h_0(F_S)))^{\deg(g_1)}.$$

Then by Property 2.9,

$$\begin{aligned} H(h_1(g_1(h_0(F)))) &\leqslant H(h_1)(2(D! + 1)H(h_0(F)))^{\deg(g_1)} \\ &\leqslant \deg(g_1)^{\deg(g_1)}(2(D! + 1)H(h_0(F)))^{\deg(g_1)}. \end{aligned}$$

In general, for any polynomial f,

$$H(h_i \circ g_i \circ f) \leqslant (2c_i(\deg(f) + 1)H(f))^{c_i},$$

where $c_i = \deg(g_i)$. Continuing in this fashion, alternately applying Properties 2.9 and 2.10, yields the following formula:

$$\begin{aligned} & H(h_k \circ g_k \circ h_{k-1} \circ g_{k-1} \circ \cdots \circ h_0 \circ F_S) \\ & < 2^{\sum_{i=1}^{k} (c_i c_{i+1} \ldots c_k)} H(h_0(F_S))^{c_1 c_2 \ldots c_k} \prod_{i=1}^{k} c_i^{c_i c_{i+1} \ldots c_k} \\ & \times \prod_{i=1}^{k} (c_{i-1} c_{i-2} \cdot \ldots \cdot c_1 c_0 + 1)^{(c_i c_{i+1} \ldots c_k)} \end{aligned}$$

with $c_0 = \deg(F_S) = D!$.

Now note that $H(P) = H(F_T \circ F_S) = H(h_{s-2} \circ g_{s-2} \circ h_{s-3} \circ g_{s-3} \circ \cdots \circ h_0 \circ F_S)$, with s the size of S, so for $k = s - 2$, the formula gives the bound on the height of the polynomial. Recall the behavior of the degrees of the g_i, given by $c_i \leqslant c_{i-1}^{2c_{i-1}}$ for $i \geqslant 1$. Thus we have shown the following theorem.

Theorem 3.2. *Let $G(x) = x^{2x}$ and let $G_i(x) = G \circ G \circ \cdots \circ G$ be the composition of i factors G with $G_0(x) = x$. Given a finite, non-empty set $S \subset \bar{\mathbf{Q}}$, closed under Galois action, of cardinality s, with D irrational elements, and of height H_S, there exists a non-constant polynomial $P(x) \in \mathbf{Q}[x]$ with $P(S \cup \{zeroes\ of\ P'\}) \subset \{0, 1\}$ such that:*

- *if $s = 1$, then $H(P) \leqslant H_S$,*
- *if $s = 2$, and $S \subset \mathbf{Q}$, then $H(P) \leqslant 2H_S^2$,*
- *if $s = 2$ and $S \not\subset \mathbf{Q}$, then $H(P) \leqslant 2^2 H_S^4$,*
- *and otherwise if $s \geqslant 3$, then with $c_0 = D!$, $c_1 = 2H_T^3$, and $c_i = G_{i-1}(c_1)$, and finally $M < (2^3 D H_S)^{D 2^D D!}$, we have*

$$\begin{aligned} H(P) & < 2^{\sum_{i=1}^{s-2} (c_i c_{i+1} \ldots c_{s-2})} (M)^{c_1 c_2 \ldots c_{s-2}} \prod_{i=1}^{s-2} c_i^{c_i c_{i+1} \ldots c_{s-2}} \\ & \times \prod_{i=1}^{s-2} (c_{i-1} c_{i-2} \ldots c_1 c_0 + 1)^{c_i c_{i+1} \ldots c_{s-2}} \end{aligned}$$

This expression is extremely cumbersome, though. For such a rough bound, it suffices to note that every term with a contribution of c_{s-2} far outweights any other. We omit the details, but one can obtain

$$H(P) < c_{s-2}^{3c_{s-2}^2}.$$

4. RATIONAL FUNCTION CASE

4.1. Notation and Algorithm. As in the polynomial case, we first present the algorithm for producing rational functions which we will use to compute the bound. There are a number of algorithms one might construct oneself or choose from; for example, as mentioned earlier, Belyi's original algorithm produced rational functions. However, we are interested in achieving bounds much lower than those in the polynomial case and thus in particular, in avoiding the many compositions of Stage II of the previous algorithm. See a more recent preprint of Belyi's, [1], for one possible improved method.

The algorithm stated here uses essentially the same method as [1] but has a very economical presentation. We thank Frits Beukers for drawing our attention to this idea of Jean Marc Couveignes'. With the same notation given earlier, that is, S a finite, non-empty set of algebraic numbers closed under Galois action with size s and height H_S, the algorithm produces $R(x) \in \mathbf{Q}(x)$ where $R(S) \subset \{0, 1, \infty\}$ and $R(x)$ is ramified over at most those three points.

Stage I. Follow Stage I of the previous algorithm to construct $F_S(x) \in \mathbf{Q}[x]$. Then $T \subset \mathbf{Q}$ is given by

$$T = \{F_S(a) : F_S'(a) = 0\} \cup \{F_S(S)\}.$$

Stage II. Let $T = \{0, \beta_1, \ldots, \beta_n\}$, $n<s$. (Note that while the β_i are not necessarily ordered, we do distinguish zero in the set.) Let

$$G_T(x) = \prod_{i=1}^{n} \left(1 - \frac{x}{\beta_i}\right)^{r_i},$$

where the r_i are non-zero integral solutions to the system of equations

$$\sum_{i=1}^{n} \frac{r_i}{\beta_i^k} = 0 \qquad \text{for } k = 1, \ldots, n-1.$$

Set $R(x) = G_T \circ F_S$, finishing the algorithm.

For completeness, we check that with this construction G_T satisfies the desired ramification properties and make a few comments. The system of equations for the r_i come from the following constraint: let $G_T = 1 + cx^n +$ (higher-order terms). Consider the logarithmic derivative, given by

$$\frac{G_T'}{G_T} = \sum_{i=1}^{n} \frac{-r_i/\beta_i}{1 - x/\beta_i}.$$

This must be of the form $cx^{n-1}+$ (higher-order terms), which yields the equations for the r_i. But then as zero is the root of $G'_T/G_T = 0$ with multiplicity $n-1$, no new ramification has been introduced by G_T.

Note also that each β_i is sent to 0 or ∞ as r_i is positive or negative, and 0 is sent to 1. Thus it remains only to show that no r_i can be zero, but this follows from the fact that the system of equations generate a Vandermonde matrix. More specifically, solving explicitly for the r_i, say by Cramer's rule, involves determinants of Vandermonde matrices, which are given by the product of differences of the matrix entries and hence are never zero here. This remark is of computational interest in generating Belyi functions not only because it shows there is a solution where every r_i is non-zero, which is necessary to control the ramification, but also because ordinarily using determinants is a computationally expensive operation to implement yet in the special case of Vandermonde matrices can be practical.

4.2. Bounding the Degree. To bound the degree of $R(x)$, where $R(x) = G_T \circ F_S$, since we already have that F_S is of degree at most $D!$, we must simply bound the degree of G_T. In fact, most of the work for this has been done by computing a bound for the height H_T of the set T, which is given in Proposition 3.1, Section 3.2. We also require the following result regarding solutions to the system of equations that generate the exponents in G_T.

PROPOSITION 4.1. *If $T = \{\beta_1, \ldots, \beta_n\} \in \mathbf{Q}\backslash\{0\}$ is of height H_T, then there exists a non-trivial solution $(r_1, \ldots, r_n) \in \mathbf{Z}^n$ for the system of equations*

$$\sum_{i=1}^{n} \frac{r_i}{\beta_i} = 0,$$
$$\sum_{i=1}^{n} \frac{r_i}{\beta_i^2} = 0,$$
$$\vdots$$
$$\sum_{i=1}^{n} \frac{r_i}{\beta_i^{n-1}} = 0$$

with

$$H(r_i) \leqslant 2^{(n-2)(n-1)} H_T^{(3n-4)(n-1)}.$$

Proof. The existence of such a non-zero solution follows from linear algebra, and we can solve directly for the r_i's, using Cramer's rule to first obtain a solution in $\mathbf{Q}^n$. Setting $r_n = -1$, we have

$$r_i = \frac{\det M_i}{\det A},$$

where A is the $(n-1) \times (n-1)$ matrix with entries $(a_{jk}) = \frac{1}{\beta_k^j}$ and M_i is the matrix A with the ith column replaced by entries of $(a_{ji}) = \frac{1}{\beta_n^j}$.

As A and M_i are Vandermonde matrices (after pulling out constants), the expression for r_i is readily simplified to

$$r_i = \frac{\beta_i}{\beta_n} \prod_{k=1}^{n-1} \left(\frac{1}{\beta_n} - \frac{1}{\beta_k} \right) \Big/ \prod_{\substack{k=1 \\ k \neq i}}^{n} \left(\frac{1}{\beta_i} - \frac{1}{\beta_k} \right).$$

Applying properties of heights and valuations, we obtain

$$H(r_i) \leqslant 2^{(n-2)} H_T^{(3n-4)}$$

for $1 \leqslant i \leqslant n-1$ and

$$H(r_n) = 1.$$

Therefore, we can scale up to obtain integral solutions with

$$H(r_i) \leqslant (2^{(n-2)} H_T^{(3n-4)})^{(n-1)},$$

which gives the proposition. ▮

Remark 4.1. Had the system of equations not been of such a special form, one could still easily bound solutions with Siegel's Lemma. Given m equations, n unknowns, and integral coefficients bounded in absolute value by C, there is a non-trivial solution set bounded in every entry by $1 + (nC)^{n/(m-n)}$. In fact, one can use a sharper form of Siegel's Lemma, such as Bombieri and Vaaler proved in [3]. However, their result benefits from reducing by common divisors of minors of a certain determinant, which we cannot necessarily do here. Also, it should be noted that Siegel's Lemma is stronger when there is a greater difference between the number of equations and number of unknowns; here the difference is only one.

THEOREM 4.1. *Given a finite, non-empty set* $S \subset \bar{\mathbf{Q}}$, *closed under Galois action, of cardinality* s, *with* D *irrational elements, and of height* H_S, *there exists a non-constant function* $R(x) \in \mathbf{Q}(x)$ *with* $R(S) \subset \{0, 1, \infty\}$, *ramified over at most* $\{0, 1, \infty\}$, *such that*

- *if* $s < 3$, *then* $\deg(R) \leqslant 2$,

- *if* $s \geqslant 3$ *and* $S \subset \mathbf{Q}$, *then*

$$\deg(R) < (s-1)2^{(s-2)(s-1)}H_S^{(3s-4)(s-1)},$$

- *and otherwise if* $s \geqslant 3$ *and* $S \not\subset \mathbf{Q}$, *then*

$$\deg(R) < (4DH_S)^{3^2 D 2^{D-2} D!(s-2)^2}.$$

Proof. The existence of $R(x)$ follows from the algorithm of Section 4.1 where the discussion shows that R has the required ramification properties. For the case when $s < 3$, the argument is the same as that for the polynomial case. Otherwise, to bound the degree of $R = G_T \circ F_S$, we use that $\deg(G_T) \leqslant (s-1)\max_i |r_i|$, since there are at most $(s-1)$ exponents r_i contributing to the degree; also, we know $\deg(F_S) \leqslant D!$.

If $S \subset \mathbf{Q}$, then $D = 0$ and $H_T = H_S$, so the previous proposition gives the bound

$$\deg(R) < (s-1)2^{(s-2)(s-1)}H_S^{(3s-4)(s-1)},$$

which is rounded up further in the Introduction. Finally, if $S \not\subset \mathbf{Q}$, again we apply the previous proposition along with the bound for H_T proved in Section 3.2, Proposition 3.1. This yields

$$\deg(R) < D!(s-1)2^{(s-2)(s-3)}(48(2^2 DH_S)^{3D2^{D-2}D!})^{(3s-7)(s-2)}.$$

If D is much less than s, note that this expression gives a better bound than that stated in the theorem of the Introduction. Since $s \geqslant D$, we can write this bound in terms of s. First rounding up for simplicity, we get

$$\begin{aligned}\deg(R) &< (4DH_S)^{3^2 D 2^{D-2} D!(s-2)^2}\\ &\leqslant (4sH_S)^{3^2 2^{s-2} s! s^3}. \qquad \blacksquare\end{aligned}$$

4.3. Bounding the Coefficients. Until now we have not used the height of a rational function; recall that this is bounded by the height of its numerator and denominator. Since $R(x) = G_T \circ F_S$, given information about the height of G_T and F_S, one might use composition properties modified for rational functions to bound the height. However in this case, because G_T is of such a specific form, that is, the product of binomials, we will profit more from simply using Property 2.4, which bounds products. We prove the following result.

THEOREM 4.2. *Given a finite, non-empty set* $S \subset \bar{\mathbf{Q}}$, *closed under Galois action, of cardinality* s, *with* D *irrational elements, and of height* H_S, *there* *exists a non-constant function* $R(x) \in \mathbf{Q}(x)$ *with* $R(S) \subset \{0, 1, \infty\}$, *ramified over at most* $\{0, 1, \infty\}$, *such that*:

- *if* $s = 1$, *then* $H(R) \leqslant H_S$,
- *if* $s = 2$ *and* $S \subset \mathbf{Q}$, *then* $H(R) \leqslant 2H_S^2$,
- *if* $s = 2$ *and* $S \not\subset \mathbf{Q}$, *then* $H(R) \leqslant 2^2 H_S^4$,
- *if* $s \geqslant 3$ *and* $S \subset \mathbf{Q}$, *then*
$$H(R) < (4H_S)^{(s-1)2^{(s-3)(s-2)} H_S^{(3s-7)(s-2)}}$$
- *and otherwise if* $s \geqslant 3$ *and* $S \not\subset \mathbf{Q}$, *then*
$$H(R) < (2DH_S)^{(2DH_S)^{3D2^D D! s^2}}.$$

Proof. The first cases, for $s \leqslant 2$, are as in the polynomial case. For $s \geqslant 3$, following the algorithm given in Section 4.1, we have that

$$R(x) = \prod_{i=1}^{s-1} \left(1 - \frac{F_S(x)}{\beta_i}\right)^{r_i}$$

is a function satisfying the desired ramification properties with F_S, β_i, and r_i as given in the algorithm. Let $R^+(x)$ and $R^-(x)$ denote the numerator and denominator in $\mathbf{Q}[x]$ of $R(x)$. To bound $H(R)$, it suffices to put an upper bound on $R^+(x)$ (or equivalently on $R^-(x)$). From Property 2.4, we get

$$H(R^+(x)) \leqslant \prod (D! + 1) \prod H\left(1 - \frac{F_S(x)}{\beta_i}\right),$$

where both products are taken over the sum of the positive r_i, with each β_i appearing r_i times. Now for a single factor, by Property 2.8 we have

$$H\left(1 - \frac{F_S(x)}{\beta_i}\right) \leqslant 2H(\beta_i)H(F_S),$$

which is

$$\leqslant 2H_T H(F_S)$$

for all i. Let M denote the maximum of the r_i, so the sum of the r_i is at most nM, where n is the number of non-zero elements of T, as in the algorithm. Thus we have

$$H(R^+(x)) \leqslant ((D! + 1)2H_T H(F_S))^{nM}$$

with $n \leqslant s - 1$, as not all the r_i are of the same sign.

For the case when $S \subset \mathbf{Q}$, we have $D = 0$ and $F_S(x) = x$, so $H_T = H_S$, $H(F_S) = 1$, $n = s - 1$, and $M = H(r_i)$ from Proposition 4.1. This gives

$$H(R^+(x)) \leqslant (4H_S)^{(s-1)2^{(s-3)(s-2)} H_S^{(3s-7)(s-2)}}$$

and thus, we can round up to

$$H(R) < (2H_S)^{2^{s^2} H_S^{3s^2}}.$$

Otherwise, it remains simply to use the prior bound work, where M is again given by the height $H(r_i)$ in Proposition 4.1, H_T is bounded in Proposition 3.1, and $H(F_S)$ is bounded in Lemma 3.3. Also, we have $n = s - 2$, since there are at most $s - 1$ exponents r_i of which at least one is of sign different than the others. After some algebraic manipulation, one can obtain

$$H(R) < (2DH_S)^{(2DH_S)^{3D2^D D! s^2}}.$$

The details are left to the reader. Leaving D in the expression allows one to profit in the case where D is much less than s; noting $s \geqslant D$ gives the theorem of the Introduction. ∎

5. REMARKS

5.1. A Lower Bound on the Polynomial Degree. It is natural to ask for a lower bound of a Belyi map, which, at least for the degree, is an easy consequence of the Riemann–Hurwitz formula. For the polynomial case, consider P mapping from $\mathbf{P}_1(\mathbf{C})$ to $\mathbf{P}_1(\mathbf{C})$, totally ramified over ∞. Then by Riemann–Hurwitz, one finds that the inverse image of $\{0, 1\}$, our only other ramification points, is of size at most $\deg(P) + 1$. In other words, one has

$$\deg(P) \geqslant \#S - 1.$$

In fact this polynomial degree bound is sharp, for example, in the simple case $S = \{1, 0, -1\}$ and $P(x) = x^2$ (or more generally, for ζ_n an nth root of unity, $S = \{0, \zeta_n, \zeta_n^2, \ldots, \zeta_n^{n-1}, 1\}$ and $P(x) = x^n$).

5.2. Regarding Upper Bounds and Curves. The upper bounds given in the theorems are by no means sharp. From the outset, the height lemmas are chosen to be easy to work with and compose rather than to give as sharp as possible a result. However, the bounds do give an indication of the general growth of the Belyi map relative to the size of S following Belyi's algorithm. In practice, one would like to start with a curve, C, rather than a set of ramification points in $\mathbf{P}_1$.

As an example, consider the case of an elliptic curve, E. Given in Legendre normal form, we have (for $\lambda \in \mathbf{C}\backslash\{0, 1\}$)

$$E : y^2 = x(x - 1)(x - \lambda).$$

Every elliptic curve E defined over a number field K is isomorphic (over $\bar{K}$) to such a form. Projection in the x-coordinate gives a degree two map ramified over $\{0, 1, \lambda, \infty\}$. Therefore, one can apply the theorems of Section 3 or 4 to obtain a bound on the degree of a Belyi map as a function of $s = \deg(\lambda) + 2$ and $H_S = H(\lambda)$.

Given a specific curve, or restricting to some set of curves, typically ad hoc methods will give lower Belyi map bounds than the algorithms do. In [6], Jones and Singerman give the example of the nth Fermat Curve given by $x^n + y^n = 1$ of genus $(n - 1)(n - 2)/2$. The projection map, say in the x-coordinate, gives a map of degree n ramified over the nth roots of unity, which we can then compose with the x^n map noted in the previous section. One can check that this gives a map of degree n^2 ramified over at most $\{0, 1, \infty\}$. For the case of elliptic curves, one might do better starting with a minimal Weierstrass form.

5.3. Belyi Maps for Non-zero Characteristic. Although Belyi's theorem is for curves over number fields, it is of interest to ask what happens over characteristic $p \neq 0$. Let $\bar{\mathbf{F}}_p$ denote the algebraic closure of the finite field $\mathbf{F}_p$. Working in characteristic p, with maps from $\mathbf{P}_1(\bar{\mathbf{F}}_p)$ to itself, it is in fact easy to kill any ramification while using pth powers so that no new ramification is introduced (forcing that the derivative has coefficients of p). See [10, p. 335], for one possible proof.

Saïdi notes that for a curve C, if $p > 2$ then by a result of Fulton's there is a map from C to $\mathbf{P}_1(\bar{\mathbf{F}}_p)$ that is tamely ramified over a finite set S, in particular, of ramification degree 2 over each point. (We assume the curve is algebraic over $\bar{\mathbf{F}}_p$.) Thus one can construct a Belyi-like map, unramified outside $\{0, 1, \infty\}$ from C to $\mathbf{P}_1(\bar{\mathbf{F}}_p)$.

ACKNOWLEDGMENTS

The referee's encouraging comments and suggestions were most appreciated. I also thank the U.C. Berkeley math department, where the work was done, and M.S.R.I., for its hospitality. In particular, it is a pleasure to thank Hendrik Lenstra for introducing me to Belyi's theorem and raising the question of effective bounds.

REFERENCES

1. G. V. Belyi, Another proof of the three points theorem, Max Planck Institut preprint, April 1997.

2. G. V. Belyi, On the Galois extensions of the maximal cyclotomic field (trans. by N. Koblitz), *Math. USSR-Izv.* **14**(2) (1980), 247–253.
3. E. Bombieri and J. Vaaler, On Siegel's lemma, *Invent. Math.* **73** (1983), 11–32.
4. N. D. Elkies, ABC implies Mordell, *Internat. Math. Res. Notices* **7** (1991) [Appendix to Duke J. Math. **64**(3) (1991)].
5. A. Grothendieck, *Esquisse d'un Programme*, pp. 5–48 (English translation, pp. 243–284) in [11].
6. G. Jones and D. Singerman, Belyi functions, hypermaps, and Galois groups, *Bull. London Math. Soc.* **28** (1996), 561–590.
7. L. Khadjavi, "An Effective Version of Belyi's Theorem," Dissertation, Univ. of California, Berkeley, May 1999.
8. S. Lang, "Fundamentals of Diophantine Geometry," Springer-Verlag, New York, 1983.
9. R. Liţcanu, "Etude du Degré des Morphismes de Belyi," Thése, Univ. Paris, XI Orsay, September 1999.
10. M. Saïdi, Revêtements modérés et groupe fondamental de graphe de groupes, *Compositio Math.* **107**(3) (1997), 319–338.
11. L. Schneps and P. Lochak (Ed.), "Geometric Galois Actions 1. Around Grothendieck's Esquisse d'un Programme," London, Mathematical Society Lecture Note Series, Vol. 242, Cambridge Univ. Press, Cambridge, UK, 1997.
12. J.-P. Serre, "Lectures on the Mordell Weil Theorem" (trans. by M. Brown), Aspects of Math E. Vol. 15, Vieweg und Sohn, Braunschweig, 1989.
13. J. Wolfart, *The 'obvious' part of Belyi's theorem and Riemann surfaces with many automorphisms*, pp. 97–112 in [11].