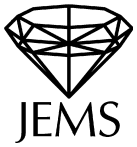
J. Eur. Math. Soc. 14, 175–208 

DOI 10.4171/JEMS/299



Clemens Fuchs · Umberto Zannier

# Composite rational functions expressible with few terms

Received April 19, 2010

**Abstract.** We consider a rational function $f$ which is 'lacunary' in the sense that it can be expressed as the ratio of two polynomials (not necessarily coprime) having each at most a given number $\ell$ of terms. Then we look at the possible decompositions $f(x) = g(h(x))$, where $g, h$ are rational functions of degree larger than 1. We prove that, apart from certain exceptional cases which we completely describe, the degree of $g$ is bounded only in terms of $\ell$ (and we provide explicit bounds). This supports and quantifies the intuitive expectation that rational operations of large degree tend to destroy lacunarity.

As an application in the context of algebraic dynamics, we show that the minimum number of terms necessary to express an iterate $h^{\circ n}$ of a rational function $h$ tends to infinity with $n$, provided $h(x)$ is not of an explicitly described special shape. The conclusions extend some previous results for the case when $f$ is a Laurent polynomial; the proofs present several features which have not appeared at all in the special cases treated so far.

## 1. Introduction and results

In this paper we are concerned with 'lacunary' rational functions; by this we mean expressions $f(x) = P(x)/Q(x)$, where $P, Q$ are polynomials (not necessarily coprime) having altogether at most a given number $\ell$ of terms. We think of the number of terms as being bounded whereas we allow the degrees of the terms and their coefficients to be arbitrary.

More specifically, the paper studies the *decomposability* of such a lacunary function, namely the equation $f(x) = g(h(x))$, where $g, h$ are rational functions of degree $> 1$.

We remark that lacunary polynomials (or rational functions) appear in many mathematical investigations: the binomials have an ancient history; reducibility of trinomials and $\ell$-nomials was studied by Selmer, ..., Schinzel (see e.g. [14, Chs. 5, 6]). Lacunary polynomials are meaningful also because they appear as expressing restrictions of any prescribed regular function on $\mathbb{G}_{\mathrm{m}}^n$ to varying 1-dimensional algebraic subgroups.

Generally speaking, lacunary polynomials also appear in computational issues since we may 'write down' a formula for a polynomial with a given number of terms, thinking

C. Fuchs: Department of Mathematics, ETH Zurich, Rämistrasse 101, 8092 Zürich, Switzerland; e-mail: clemens.fuchs@math.ethz.ch

U. Zannier: Scuola Normale Superiore, Piazza dei Cavalieri, 7, 56126 Pisa, Italy; e-mail: u.zannier@sns.it

of the degrees and coefficients as indeterminates. (For instance lacunary polynomials appear as one-way functions in cryptography.) We may further quote work in real geometry (see e.g. the survey [9]) showing that sometimes it is possible to have control on objects defined by equations with boundedly many terms, even if they have large degree.

And also 'decomposability' of a rational function $f$ is relevant in arithmetical and algebraic questions; it is for instance deeply related to the Galois group of the rational cover $f : \mathbb{P}_1 \to \mathbb{P}_1$, and has impact in Diophantine equations with separated variables (see e.g. [1, 3]).

The paper [19] studied the decomposability of lacunary polynomials; note that since the degree of a polynomial with a fixed number of terms may be arbitrarily large, it does not appear an obvious issue to check whether it is decomposable. It has been proved therein that *if $f$ is a 'lacunary' polynomial, say with at most $\ell$ non-constant terms, and if $f(x) = g(h(x))$ for polynomials $g, h$ of degree larger than 1, then either $h(x)$ has the special shape $ax^n + b$ or* $\deg g$ *is bounded in terms of $\ell$ only.* In other words, roughly speaking, *if $g$ has 'large' degree then $g(h(x))$ has 'many' terms* (or else $h$ is very special). So this kind of result supports and quantifies the intuitive expectation that polynomial operations of large degree tend to destroy lacunarity.

Moreover, this bound led in [20] to a proof of a conjecture of Schinzel, and even to describe 'algorithmically' all the possible decompositions of a polynomial with a given number of terms.

In [21] a similar bound was extended to Laurent polynomials (with a new unavoidable exceptional shape) also with the purpose of application to certain questions studied by Watt and Zieve on symbolic polynomials [17]. Actually, these last authors would also require, for further applications, *some extension of such bounds to general rational functions*, and on the other hand this could also admit applications to a generalization of the results of [20].

It is the purpose of this paper to carry out this program. We remark at once that this extension does not appear to follow by a mere straightforward modification of the methods of [19, 21], as several new crucial obstacles appear. (For the treatment of one case we have even found it necessary to rely on a paper using the classification of finite simple groups.)

As a simple immediate application in the context of one-variable dynamics, we shall also prove a corollary bounding below the number of terms of an $n$-fold iterate $h^{\circ n}$ of a rational function $h$. (This was done in [19] for Laurent polynomials.)

**Notation.**

- We let $k$ be a field of characteristic zero. For the present purposes it causes no loss of generality to suppose, as we shall do, that $k$ is algebraically closed. As usual we set $k^* = k\backslash\{0\}$.
- The degree of $f \in k(x)$ is defined by $\deg f = [k(x) : k(f(x))]$. In case $f(x) = p(x)/q(x)$, with $p, q \in k[x]$ coprime, $\deg f$ is just the maximum of the degrees of $p$ and $q$ respectively. (We shall also use degrees in function fields other than $k(x)$; see Section 2.)

- We denote by $\mathrm{PGL}_2(k) = \mathrm{Aut}(\mathbb{P}_1(k))$ the group of linear fractional transformations, i.e. rational functions of the form $(ax+b)/(cx+d)$ with $a, b, c, d \in k$, $ad - bc \neq 0$, and by $\infty = (0:1)$ the unique point at infinity of $\mathbb{P}_1(k)$.
- We say that a rational function $f \in k(x)$ is *decomposable* if there is a non-trivial decomposition of $f$ as a rational function in $k(x)$, i.e. an equation $f(x) = g(h(x))$ with $g, h \in k(x)$, $\deg g, \deg h > 1$. Observe that we can modify such decompositions by replacing $g, h$ resp. by $g \circ \lambda^{-1}, \lambda \circ h$ with any $\lambda \in \mathrm{PGL}_2(k)$. We remark that if $f$ is a polynomial or a Laurent polynomial, then we can easily normalize $g, h$ correspondingly; see Remark 1.3 below.
  Accordingly, we say that $f$ is *indecomposable* if the equality $f(x) = g(h(x))$, $g, h \in k(x)$, implies $\deg g = 1$ or $\deg h = 1$.
- We denote by $T_n$ the Chebyshev polynomial of degree $n$; it is uniquely determined by the equation $T_n(x + x^{-1}) = x^n + x^{-n}$.

We are now ready to formulate the main result.

**Main Theorem.** *Let $\ell$ be a positive integer and let $f \in k(x)$ be expressible as $f(x) = P(x)/Q(x)$, where $P, Q \in k[x]$ have altogether at most $\ell$ terms. Suppose that $f(x) = g(h(x))$, where $g, h \in k(x)$ and where $h(x)$ is not of the shape $\lambda(ax^n + bx^{-n})$ for any $a, b \in k$, $n \in \mathbb{N}$ and $\lambda \in \mathrm{PGL}_2(k)$. Then*

$$\deg g \leq 2016 \cdot 5^{\ell}.$$

Note that in this statement we are not assuming that $P$, $Q$ are coprime (which would lead to an easier discussion). We shall refer to the shape $\lambda(ax^n + bx^{-n})$ as the *forbidden shape* (for the Main Theorem).

Here is the simple application to iterates alluded to above, where we denote by $h^{\circ n}$ the $n$th iterate $h(h(\ldots h(x))\ldots)$:

**Corollary.** *Let $q \in k(x)$ be non-constant and let $h \in k(x)$ be of degree $d \geq 3$. Suppose that $h(x)$ is not conjugate (with respect to the group action given by $\mathrm{PGL}_2(k)$ on $k(x)$) to $\pm x^d$ or to $\pm T_d(x)$. Then, for any integer $n \geq 3$, we cannot express $h^{\circ n}(q(x))$ as a ratio of polynomials having altogether less than $\frac{1}{\log 5}((n-2)\log d - \log 2016)$ terms.*

Note that in the cases $h(x) = x^d$ and $h(x) = T_d(x)$ we cannot obtain any bound tending to infinity with $d$, in view of the equations $h^{\circ n}(x) = x^{d^n}$ in the first case and $h^{\circ n}(x + x^{-1}) = x^{d^n} + x^{-d^n}$ in the second case. We have restricted to $\deg h \geq 3$ for simplicity. If $\deg h = 2$ we may obtain a similar statement with a slightly more boring proof, or by applying this one with $h(h(x))$ in place of $h(x)$. Inspection shows that the proof yields further precision on the shape of $q(x)$ if the conclusion does not hold.

The Main Theorem takes into account the most general case, but it subdivides into a number of subcases according to the shape of $f$ and $h$, where the bound takes different forms and the proofs can be more or less involved. Therefore we add to this statement the corresponding statements for the subcases; in turn, the Main Theorem follows easily from these statements.

**Theorem 1.1** (Polynomial case). *Let $\ell$ be a positive integer and let $f \in k[x]$ be a polynomial with $\ell$ non-constant terms. Suppose that $f(x) = g(h(x))$, where $g, h \in k[x]$ and where $h(x)$ is not of the shape $ax^n + b$ for $a, b \in k$, $n \in \mathbb{N}$. Then*

$$\deg f + \ell - 1 \leq 2\ell(\ell-1)\deg h \quad \textit{and} \quad \deg g \leq 2\ell(\ell-1).$$

This is [19, Theorem 1]. We remark (as is pointed out in [19]) that the exclusion of polynomials $h$ of the shape $ax^n + b$ is really necessary, as is shown by simple examples like $g(x) = g^*(x - b)$, where $g^* \in k[x]$ has $\ell$ non-constant terms.

**Theorem 1.2** (Laurent case). *Let $\ell$ be a positive integer and let $f \in k[x, x^{-1}]\backslash k[x]$ be a Laurent polynomial with $\ell$ non-constant terms. Suppose that $f(x) = g(h(x))$, where $g \in k[x]$, $h \in k[x, x^{-1}]$ and where $h(x)$ is not of the shape $ax^n + b + cx^{-n}$ for $a, b, c \in k$, $n \in \mathbb{N}$. Then*

$$\deg f \leq 2(2\ell-1)(\ell-1)(\deg h - 2) \quad \textit{and} \quad \deg g \leq 2(2\ell-1)(\ell-1).$$

This follows at once from [21]. We again remark (as in [21]) that the exclusion of Laurent polynomials $h$ of the shape as in the theorem cannot be avoided, as follows e.g. from the example $T_n(x + x^{-1}) = x^n + x^{-n}$. Thus these exceptions also have to be taken into account in the present Main Theorem.

**Remark 1.3.** Suppose that $f \in k[x]$ is a polynomial which is decomposable as a rational function: $f = g \circ h$, $g, h \in k(x)$. Then observe that, by changing $h$ into $\lambda \circ h$ for a $\lambda \in \mathrm{PGL}_2(k)$, and changing $g$ into $g \circ \lambda^{-1}$, which does not change $\deg g, \deg h$, we may assume that $g, h$ are polynomials: in fact, the preimage $f^{-1}(\infty)$ consists of just $\infty$ (since $f$ is a polynomial) and, by choosing $\lambda$ suitably, the same becomes true for $g$ and $h$. Thus $f$ is decomposable also as a polynomial.

Similarly for the case when $f$ is a Laurent polynomial, but neither $f \in k[x]$ nor $f \in k[x^{-1}]$: First, the preimage $f^{-1}(\infty)$ is $\{0, \infty\}$. If $f = g \circ h$, then $g^{-1}(\infty)$ is either $\{P\}$ or $\{P, Q\}$, $P, Q \in \mathbb{P}_1(k)$, $P \neq Q$ and we have $h^{-1}(P) = \{0, \infty\}$ or $h^{-1}(\{P, Q\}) = \{0, \infty\}$, respectively. In the first case we may choose $\lambda$ so that $P = \infty$ and thus $g$ is a polynomial and $h$ a Laurent polynomial. In the second case, if $h^{-1}(P) = \{0\}, h^{-1}(Q) = \{\infty\}$, we choose $\lambda$ so as to transform $Q$ into $\infty$ and $P$ into 0 (this is possible since $\mathrm{PGL}_2(k)$ is doubly transitive). Then $g$ becomes a Laurent polynomial and $h(x) = cx^n$ becomes of the forbidden shape.

(In particular, this shows that when $h$ is a (Laurent) polynomial, the exceptional shapes of the Main Theorem may be reduced to the shapes stated in Theorems 1.1 and 1.2.)

The normalization described in this remark will be used repeatedly in what follows.

These two theorems present the current state of the problem. In the next statements we give the new results, which together with the stated theorems will yield the Main Theorem.

With the phrasing '(non-)Laurent case' we refer to the shape of $h$; however, even when $h$ is a (Laurent) polynomial, the situation differs from the previous two theorems in that we compose $h$ with a rational function $g$ (rather than a polynomial).

Of course it may happen that $h \notin k[x, x^{-1}]$ but $\lambda \circ h \in k[x, x^{-1}]$ for a suitable $\lambda \in \mathrm{PGL}_2(k)$, so that after replacing $g, h$ resp. by $g \circ \lambda^{-1}, \lambda \circ h$ we are in the Laurent case. To express rapidly this fact, we have found it convenient to introduce the following simple definition:

**Definition NL.** *$h$ satisfies* NL $:\Leftrightarrow$ For any $\lambda \in \mathrm{PGL}_2(k)$: $\lambda \circ h \notin k[x, x^{-1}]$.

Namely, $h$ satisfies NL (Non-Laurent) if $h$ cannot be turned into a Laurent polynomial by applying a linear fractional transformation on the left.

This definition allows us to reduce more simply the remaining cases to the following theorems:

**Theorem 1.4** (Laurent case for rational functions). *Let $\ell$ be a positive integer and let $f \in k(x)$ be expressible as $f(x) = P(x)/Q(x)$, where $P, Q \in k[x]$ have altogether at most $\ell$ terms. Suppose that $f(x) = g(h(x))$, where $g \in k(x)$, $h \in k[x, x^{-1}]$, and where $h(x) \neq ax^n + b + cx^{-n}$ for any $a, b, c \in k, n \in \mathbb{N}$. Then we have the following two cases:*

(i) *If $P, Q$ are coprime in $k[x]$, then*

$$\deg f \leq 2(2\ell - 1)(\ell - 1) \deg h, \quad \deg g \leq 2(2\ell - 1)(\ell - 1).$$

(ii) *In general,*

$$\deg f \leq (2016 \cdot 5^{\ell})(\deg h - 1), \quad \deg g \leq 2016 \cdot 5^{\ell}.$$

Note that the inequalities on the right for $\deg g$ follow from the respective ones on the left for $\deg f$, in view of $\deg f = \deg g \deg h$.

We remark that our proof of the last case (ii) in this statement is the most involved step in the whole paper and requires a result of Müller [11, 12] relying on the classification of finite simple groups. This is not needed in the case (i), when $f$ is in reduced form, since we may apply [21, Theorem 2*] to the numerator and the denominator to get the conclusion; for this case we do not even need the auxiliary results of Section 2 or the normalizations from Section 3; we give the proof, which is independent of the rest, in Subsection 5.1.

We also point out that Theorem 1.4(ii) gives a partly new proof for Theorems 1.1, 1.2 (with a worse bound for $\deg f$).

**Theorem 1.5** (Non-Laurent case for rational functions). *Let $\ell$ be a positive integer and let $f \in k(x)$ be expressible as $f(x) = P(x)/Q(x)$, where $P, Q \in k[x]$ have altogether at most $\ell$ terms. Suppose that $f(x) = g(h(x))$, where $g, h \in k(x)$, $h$ satisfies* NL *and for every equation $h(x) = p(q(x))$ either $q$ satisfies* NL *or $q(x) = \lambda(ax^n + bx^{-n})$ for a $\lambda \in \mathrm{PGL}_2(k)$ and some $a, b \in k, n \in \mathbb{N}$. Then*

$$\deg f \leq (267 \cdot 5^{\ell})(\deg h - 1), \quad \deg g \leq 267 \cdot 5^{\ell}.$$

Before we continue with some remarks on the stated results we now give at once the deduction of the Main Theorem from the other theorems; the proof is almost immediate:

*Deduction of the Main Theorem from Theorems 1.1, 1.2, 1.4 and 1.5.* If $f$ is a polynomial or a Laurent polynomial that is decomposable as $f(x) = g(h(x))$ with $h(x)$ not of the forbidden shape (of the Main Theorem), then by Remark 1.3 we may change the decomposition (from $f = g \circ h$ to $f = (g \circ \lambda^{-1}) \circ (\lambda \circ h)$ with $\lambda \in \mathrm{PGL}_2(k)$) so that for the new composition factors we have either $g, h \in k[x]$ or $g \in k[x], h \in k[x, x^{-1}]$; then the result follows by Theorem 1.1 or Theorem 1.2 respectively (observe that by the assumptions of the Main Theorem the modified $h$ certainly cannot be of the forbidden shape of the respective theorem). (We mention that as noted right before Theorem 1.5, this opening paragraph could be skipped and subsumed in the arguments below.)

Now let $f \in k(x)$ and $f(x) = g(h(x))$ with $h(x)$ not of the forbidden shape (of the Main Theorem). If $h$ satisfies NL and for every equation $h(x) = p(q(x))$ either $q$ satisfies NL or $q(x) = \lambda(ax^n + bx^{-n})$ for a $\lambda \in \mathrm{PGL}_2(k)$ and some $a, b \in k, n \in \mathbb{N}$, then the conclusion follows from Theorem 1.5.

Otherwise we have two cases:

In the first case, $h$ does not satisfy NL, so there is a homography $\lambda \in \mathrm{PGL}_2(k)$ with $\lambda \circ h \in k[x, x^{-1}]$; now, on replacing $g, h$ resp. by $g \circ \lambda^{-1}, \lambda \circ h$, we may assume that $h \in k[x, x^{-1}]$, and the conclusion follows from Theorem 1.4.

In the second case there exists an equation $h(x) = p(q(x))$ with $p, q \in k(x)$ such that $\lambda^* \circ q \in k[x, x^{-1}]$ for a suitable $\lambda^* \in \mathrm{PGL}_2(k)$ and $q(x)$ is not of the stated shape $\lambda(ax^n + bx^{-n})$. In this case, by changing the equation $h = p \circ q$ into $h = (p \circ (\lambda^*)^{-1}) \circ (\lambda^* \circ q)$, we may again assume that $q \in k[x, x^{-1}]$, and *a fortiori* we have $q(x) \neq ax^n + b + cx^{-n}$ for any $a, b, c \in k, n \in \mathbb{N}$. Also, we now have the decomposition $f(x) = (g \circ p)(q(x))$, and we may apply to it Theorem 1.4 with $g \circ p$ in place of $g$ and $q$ in place of $h$. Since $\deg(g \circ p) \geq \deg g$, the result again follows.

Since there are no further cases, the deduction of the Main Theorem is complete. □

**Remark 1.6.** (i) We have not made any special effort to derive good numerical constants in the theorems; however, the exponential dependence of the bound on $\ell$ cannot be improved using our way of reasoning. (This step occurs in the subdivision—introduced below—of the set of exponents appearing in $P$, $Q$ into 'large' and 'small' ones; such a subdivision was not necessary and not taken into account in the polynomial and Laurent case treated in previous papers, which explains the better dependence on $\ell$ obtained therein.) Also, we do not have lower bounds better than linear in $\ell$ and we do not have definite ideas on what could be a best-possible shape for them.

(ii) Note that in the most general case we do not assume that the 'lacunary' expression $f(x) = P(x)/Q(x)$ for $f$ is in reduced form.

(iii) The issue arises if more is true, namely a similar bound on the mere assumption that the numerator is lacunary. This in turn naturally leads to the following strictly related question:

**Question.** *Suppose that $f$ is a polynomial with at most $\ell$ terms, and that it is divisible by $g(h(x))$ for $g, h \in k[x]$, $\deg g, \deg h \geq 2$, $h(x)$ not of the form $ax^n + b$, $a, b \in k$, $n \in \mathbb{N}$. Is is true that $\deg g$ is necessarily bounded in terms of $\ell$ only?*

We feel that this question might have a positive answer in general, but certainly a proof would require new methods. Some evidence for a positive answer is provided

by the case when $f$ has at most two terms; we treat this special case in the appendix to this paper.

The **strategy** of our method is very roughly as follows: We choose a conjugate $y \neq x$ of $x$ over the field $k(h)$, so we have $h(x) = h(y)$. The equation $f(x) = g(h(x)) \in k(h)$ then implies $f(x) = f(y)$. Thus we end up with $P(x)Q(y) - P(y)Q(x) = 0$. We view this last equation as an $S$-unit equation with 'few' terms (here we use the lacunarity), and apply a theorem of Brownawell–Masser (cf. Lemma 2.2) to bound the maximal degree involved. However, this theorem can be applied directly only to equations in which no proper subsum vanishes. Hence we are led to study the vanishing subsums of the numerator of $f(x) - f(y)$. If $f$ is a (Laurent) polynomial, this is easier because each vanishing subsum may then be written as $p(x) - q(y) = 0$ and has the same 'separated variables' structure as the original equation. (This analysis was the one carried out in the previous papers on this topic.)

On the other hand, if $f$ is not a Laurent polynomial, the vanishing subsums may have *a priori* a different structure. To study them we distinguish between 'large' and 'small' exponents. In order to take advantage of the upper bound alluded to above, we are led to investigate lower bounds for the degree $\deg_K(z) = [K : k(z)]$ in $K = k(x, y)$ of expressions of the form $z = x^m y^n$ with $m, n \in \mathbb{Z}$; we need a bound of the shape $\deg_K(z) \geq \varepsilon \max\{|m|, |n|\} \deg_K(x)$ for some specific number $\varepsilon > 0$, where the dependence on $\deg_K(x)$ is essential in comparison with the upper bound that follows from the Brownawell–Masser inequality. (This problem reflects another major difference with the case of Laurent polynomials.)

In order to overcome this issue we use the partition of the exponents mentioned before and moreover we require detailed information on the irreducible factors of $h(Y) - h(x)$ and their associated Puiseux series; in turn, this involves a deep result by Müller in one of the cases to be considered. We do not enter now in any detail concerning this point; we shall comment further on this at the end of Section 3.

The paper is organized as follows: In the next section we will collect some auxiliary results which are all well known and which are needed in our proof. We formulate these results in a way suitable for our applications. In Section 3 we start with some first reductions and normalizations and we study lower bounds for the degree of $x^m y^n$. In Section 4 we give the proof of a 'big' part of our theorem; in other words we prove Theorem 1.5. In this case a lemma by Hajós (cf. Lemma 2.1) will also be of importance. In Section 5 we prove the result in the Laurent case for rational functions (that is, Theorem 1.4), and then in the last section we give the proof of the corollary. Finally, in the appendix we discuss the aforementioned result concerning the more general question.

## 2. Auxiliary results

In this section we collect some auxiliary results that we will use in the proof of our theorems and that already appear in the literature; in general we shall give no proofs.

**Lemma 2.1** (Hajós lemma). *Let $f \in k[x]$ have $\ell$ non-constant terms and let $\beta \in k^*$. Then* $\operatorname{ord}_\beta(f(x) - f(\beta)) \leq \ell$.

The proof of this lemma is in fact very easy by calculating the first $\ell + 1$ derivatives of $f(x) - f(\beta)$ and then reducing the resulting linear system of equations in the $\ell + 1$ unknown coefficients to a system of Vandermonde type.

Before we state our next lemma we briefly review the theory of valuations/places on function fields (for which we refer e.g. to [10]).

For each $\theta \in k$ we have a place of $k(x)$ whose valuation $v_\theta$ expresses the order of vanishing at $\theta$, namely is given by $v_\theta(f) = n$ if $f(x) = (x-\theta)^n g(x)$ with $g(\theta) \neq 0$. Moreover, there is a place $\infty$ whose valuation $v_\infty$ is given by $v_\infty(f) = \deg Q - \deg P$ for $f(x) = P(x)/Q(x)$ with $P, Q \in k[x]$. As a matter of terminology, if $f$ has a pole at $v$, i.e. $v(f) < 0$, then the order $\mathrm{ord}(f)$ of $v$ *as a pole* of $f$ is given by $-v(f) > 0$.

In this way all (normalized) valuations of $k(x)/k$ are obtained.

Now, let $K$ be a finite extension of $k(x)$. Each of the places of $k(x)$ can be extended in at most $[K : k(x)]$ ways to a place on $K$ and in this way all places $v$ on $K$ are obtained (normalized so that $v(K\backslash\{0\}) = \mathbb{Z}$). Furthermore, the places of $K$ are in a one-to-one correspondence with the points over $k$ of a(ny) nonsingular complete curve over $k$ with function field $K$; in particular, for the rational function field $k(x)$ the places are in one-to-one correspondence with the points of $\mathbb{P}_1(k)$.

For an element $f \in K$ we define the *degree* (sometimes also called the *height*) of $f$ with respect to $K$ by

$$\deg_K(f) = \sum_v \max\{0, v(f)\} = -\sum_v \min\{0, v(f)\},$$

where the sum is taken over all places of $K$; thus it is just the number of zeros respectively poles of $f$ counted according to multiplicity. Equivalently, $\deg_K(f) = [K : k(f)]$. A useful property for us is $\deg_K(f) - \deg_K(g) \leq \deg_K(fg) \leq \deg_K(f) + \deg_K(g)$ for $fg \neq 0$.

As in the quoted previous papers on this topic, we shall need the following crucial result on $S$-units in function fields.

**Lemma 2.2** (Brownawell–Masser). *Let $K/k$ be a function field (in one variable), of genus $\mathfrak{g}$, and let $z_1, \ldots, z_s \in K$ be not all constant and such that $1 + z_1 + \cdots + z_s = 0$. Suppose also that no proper subsum of the left side vanishes. Then*

$$\max\{\deg_K(z_i)\} \leq \binom{s}{2}(2\mathfrak{g} - 2 + |S|),$$

*where $S$ is any set of places of $K$ containing all zeros and poles of the $z_i$.*

We have taken this formulation from [19, 21], but we mention again that it is an immediate consequence of [4] (or of [18]).

Our next lemma involves permutation groups, so we recall some standard definitions before giving the statement. Let $\mathcal{S}_n$ and $\mathcal{A}_n$ denote the symmetric group and the alternating group of $n$ elements respectively. Assume that a group $G$ is acting on a set $\Omega$ and let $\alpha \in \Omega$. The *orbit* of $\alpha$ under $G$ is defined by $G\alpha = \{g\alpha : g \in G\}$ and the point stabilizer of $\alpha$ in $G$ by $G_\alpha = \{g \in G : g\alpha = \alpha\}$. A group is said to act *transitively* on $\Omega$ if it has

only one orbit, i.e. $G\alpha = \Omega$ for all $\alpha \in \Omega$. Let $m$ be an integer with $1 \le m \le |\Omega|$ and denote by $\Omega^m$ the $m$th cartesian power of $\Omega$. Then $G$ acts on $\Omega^m$ componentwise and the subset $\Omega^{(m)}$ of unordered $m$-tuples of distinct points is $G$-invariant for every choice of $G$ and $m$. We say that $G$ is *$m$-transitive* if $G$ is transitive on $\Omega^{(m)}$. Finally, let $G$ be a group acting transitively on a set $\Omega$ with at least two points. We call $G$ *primitive* if each point stabilizer $G_\alpha$ is a maximal subgroup of $G$. For these definitions we refer e.g. to [5].

**Lemma 2.3** (Müller). *Let $h$ be an indecomposable Laurent polynomial, written as $h(x) = p(x)/x^e$, $p \in k[x], e \in \mathbb{N}$, and suppose that $d = \deg h = \deg p > e > 0$. Then the Galois group $G$ of the equation $h(X) - h(x)$ over $k(h(x))$, as a permutation group on the $d$ roots, is primitive and contains an element with exactly two cycles of lengths say $n$ and $d - n$; it satisfies at least one of the following:*

(i) $\mathcal{A}_d \subseteq G$,
(ii) $d = m^2$ *with* $1 < m \in \mathbb{N}, n = ma$ *with* $\gcd(m, a) = 1$, $G = (\mathcal{S}_m \times \mathcal{S}_m) \rtimes \mathbb{Z}/(2)$ *and the stabilizer of a point is conjugate to* $(\mathcal{S}_{m-1} \times \mathcal{S}_{m-1}) \rtimes \mathbb{Z}/(2)$,
(iii) $d \le 64$.

This statement follows immediately from the paper [11] (which in turn is essentially contained in [12, 13]); more precisely it follows from Theorem 4.8, the additional information in case (ii) comes from Theorem 3.3 that contains the main classification result for primitive permutation groups with cyclic two-orbit. The upper bound appearing in case (iii) could be replaced by an explicit list of sporadic groups, but since this has only a small impact on the numerical estimates in our theorems, we prefer not to deal with such exceptional cases of small degree.

We mention that $G$ being 2-transitive amounts to $(h(X) - h(Y))/(X - Y)$ being absolutely irreducible over $k$. This property (not always satisfied) is important in our application to the Laurent case for rational functions. (For polynomials the corresponding classification is due to Fried [7]; the paper of Müller [11] also records precise information on the Galois group in this case, namely for an indecomposable $h \in k[x]$ with $d = \deg h \ge 32$ the Galois group $G$ of $h(X) - h(x)$ over $k(h(x))$ is either $\mathbb{Z}/(p)$ or the dihedral group of $p$ elements for some prime $p = d$ or $\mathcal{A}_d$ or $\mathcal{S}_d$; see [11, Theorem 4.9, p. 63].)

We end this section by collecting some information on the Puiseux series for the equation $h(Y) = h(x)$ for an $h \in k(x)$.

**Lemma 2.4.** *Let $h(x) = x^e p(x)/q(x)$ with $p, q \in k[x]$ coprime, $p(0)q(0) \ne 0$, $e \in \mathbb{Z}\backslash\{0\}$. Define $s = e + \deg p - \deg q$. Assume that $h(\infty) = \infty$, i.e. $s > 0$. Let $y$ be a solution of the equation $h(Y) - h(x) = 0$ in an algebraic closure of $k(x)$.*

*For $e < 0$, the dominant terms of the possible Puiseux series for $y$ are as follows: At $x = 0$ there are precisely*

1. *$|e|$ Puiseux series with $y = \theta x + \cdots$, $\theta^{|e|} = 1$,*
2. *$s$ Puiseux series with $y = \theta x^{-|e|/s} + \cdots$, $\theta^s = 1$,*
3. *$\deg q$ Puiseux series with $y = \theta + \cdots$, where $\theta \in k^*$ satisfies $q(\theta) = 0$,*

*and at $x = \infty$,*

1. *$s$ Puiseux series with $y = \theta x + \cdots$, $\theta^s = 1$,*
2. *$|e|$ Puiseux series with $y = \theta x^{-s/|e|} + \cdots$, $\theta^{|e|} = 1$,*
3. *$\deg q$ Puiseux series with $y = \theta + \cdots$, where $\theta \in k^*$ satisfies $q(\theta) = 0$.*

*For $e > 0$, the first terms of the Puiseux series are as follows: At $x = 0$ there are precisely*

1. *$e$ Puiseux series with $y = \theta x + \cdots$, $\theta^e = 1$,*
2. *$\deg p$ Puiseux series with $y = \theta + \cdots$, where $\theta \in k^*$ satisfies $p(\theta) = 0$,*

*and at $x = \infty$,*

1. *$s$ Puiseux series with $y = \theta x + \cdots$, $\theta^s = 1$,*
2. *$\deg q$ Puiseux series with $y = \theta + \cdots$, where $\theta \in k^*$ satisfies $q(\theta) = 0$.*

We remark that the dots here indicate that the terms which follow have a higher order of zero than the first one (at the relevant point).

*Proof.* We start with the case $e < 0$ and $x = 0$. Observe that $\deg h = \deg p = s + |e| + \deg q > 0$. From the definition of $y$ we have

$$h(x) = \frac{p(x)}{x^{|e|}q(x)} = \frac{p(y)}{y^{|e|}q(y)} = h(y). \tag{1}$$

At $x = 0$ we have $h = \infty$ and therefore either $y = 0$ or $y = \infty$ or $y = \theta \in k^*$ with $q(\theta) = 0$. In the first case, equation (1) implies that the Puiseux expansion at $x = 0$ of $y$ satisfies $y^{|e|} + \cdots = x^{|e|} + \cdots$ and therefore $y = \theta x + \cdots$ for $\theta$ an $|e|$th root of unity. Clearly, we have $|e|$ series of this type. In the second case, (1) implies the relation

$$x^{|e|} + \cdots = \frac{1}{y^s} + \cdots$$

and therefore $y = \theta x^{-|e|/s} + \cdots$ for $\theta$ an $s$th root of unity. Altogether we have $s$ series of that type. Finally in the third case, (1) means that we have $y = \theta + \cdots$, where $\theta \in k^*$ is a root of $q$. Altogether there are $\deg q$ series of this type. Thus we have the result in the case $x = 0$.

The Puiseux factorization of $h(Y) - h(x)$ at $x = \infty$ and the case $e > 0$ can be obtained following the same line of arguments. □

We remark that in all cases when the series looks like $y = \theta + \cdots$, the root $\theta \in k^*$ of $q$ respectively $p$ appears as many times as its multiplicity; we shall not need this.

## 3. Normalizations and preliminaries on degrees

Let $\ell$ be a positive integer. The proofs of Theorems 1.4(ii) and 1.5 will be done by induction on $\ell$.

So let $f \in k(x)$ with $f(x) = P(x)/Q(x) = g(h(x)) \in k(x)$, where $g, h \in k(x)$, $\deg g, \deg h > 1$, and where $P, Q \in k[x]$ are not necessarily coprime; however, we assume they are not both divisible by $x$, which we may without affecting their number

of terms. We assume that the total number of terms appearing in $P, Q$ is $\ell$. By the last normalization it follows that there are exactly $\ell - 1$ non-constant terms; in particular, this implies that $\ell \geq 2$.

We shall assume throughout that $h(x)$ is not of the forbidden shape $\lambda(ax^n + bx^{-n})$ for any $a, b \in k$, $n \in \mathbb{N}$ and $\lambda \in \mathrm{PGL}_2(k)$. We remark that we may consider the statements also for $\deg h = 1$ in which case they are empty (and so trivially true) since then $h(x)$ has forbidden shape.

As a second normalization we show that we may reduce to the case when

$$h(\infty) = \infty \quad \text{and} \quad g(\infty) = \infty.$$

Let us first check that we can assume $h(\infty) = \infty$. In fact, suppose first that we are dealing with Theorem 1.4(ii), i.e. that $h$ is a Laurent polynomial. Then either $h(\infty) = \infty$ or $h \in k[x^{-1}]$. In this last case we just replace $x$ by $x^{-1}$; observe that we continue to be in the case of Theorem 1.4(ii) (i.e. the number of terms of $f(x^{-1}) = P(x^{-1})/Q(x^{-1}) = x^{\deg Q - \deg P}(x^{\deg P} P(x^{-1})/(x^{\deg Q} Q(x^{-1}))$ is the same as that for $f$ and the new $h$ is also a Laurent polynomial). Suppose now that we are dealing with an $h$ satisfying the assumptions of Theorem 1.5. In this case we replace $g, h$ resp. by $g \circ \lambda^{-1}, \lambda \circ h$ for a suitable $\lambda \in \mathrm{PGL}_2(k)$, i.e. such that $h(\infty) = \lambda^{-1}(\infty)$. Again, we remain in the case of Theorem 1.5. Note that the modified $h$ continues not to be of the forbidden shape.

Now, acting on the left of $g$ (and thus also of $f$) by a suitable element in $\mathrm{PGL}_2(k)$, we may further achieve that $g(\infty) = \infty$. This normalization is harmless; it modifies neither the degree of $g$, nor the function $h$, nor the number of terms in a suitable expression for $f$: in fact this replaces $P, Q$ by two independent linear forms in them.

In conclusion, from now on we assume that $h(\infty) = g(\infty) = \infty$.

Observe also that in proving the theorems we may assume, by a further induction on $\deg h$, that $h \notin k(x^n)$ for any $n > 1$. The theorems hold for $\deg h = 1$ since then the statements are empty.

For the induction step assume that we have $h(x) = h^*(x^m)$ with $h^* \in k(x)$ and $m \in \mathbb{N}, m > 1$. Then $f(x) = g(h(x)) = g(h^*(x^m)) = f^*(x^m)$ for $f^* = g \circ h^* \in k(x)$. By grouping the terms of $P, Q$ with respect to the residue class modulo $m$ of their degrees, we can write $P(x) = P_0(x^m) + P_1(x^m)x + \cdots + P_{m-1}(x^m)x^{m-1}$, $Q(x) = Q_0(x^m) + Q_1(x^m)x + \cdots + Q_{m-1}(x^m)x^{m-1}$ with $P_0, \ldots, P_{m-1}, Q_0, \ldots, Q_{m-1} \in k[x]$. Since $P(x) = f^*(x^m)Q(x)$ it follows that $P_i(x^m) = f^*(x^m)Q_i(x^m)$ for $i = 0, 1, \ldots, m-1$. Pick a $j \in \{0, 1, \ldots, m-1\}$ with $Q_j \neq 0$. Then $P_j(x) = f^*(x)Q_j(x)$ and thus

$$g(h^*(x)) = f^*(x) = \frac{P_j(x)}{Q_j(x)},$$

where the terms in $P_j, Q_j$ on the right form a subset of the terms of $P, Q$.

Observe that $h \in k[x, x^{-1}]$ if and only if $h^* \in k[x, x^{-1}]$. Also, if $h$ satisfies NL and for every equation $h(x) = p(q(x))$ either $q$ satisfies NL or $q(x) = \lambda(ax^n + bx^{-n})$ for a $\lambda \in \mathrm{PGL}_2(k)$, $n \in \mathbb{N}$ and $a, b \in k$, then the same is true for $h^*$. In conclusion, if we are in the cases of Theorems 1.4(ii), 1.5 for $h$, then we are in the respective cases for $h^*$.

Thus each of the statements will follow by induction once we have treated the case of $h$ not a rational function in $x^n$, for any $n > 1$, since $g$ has not changed and moreover $h^*(x)$ cannot be of the forbidden shape for otherwise $h(x)$ would also be.

Next we show that it is sufficient to prove the theorems for $h$ having one of the following two additional properties (which exclude each other):

(H1) $h \notin k(x^n)$ for any $n > 1$ and $h$ is indecomposable.
(H2) $h \notin k(x^n)$ for any $n > 1$, $h$ is decomposable as $h(x) = \tilde{h}(x + \eta x^{-1})$ where $\eta \in k^*$, and moreover every decomposition $h(x) = p(q(x))$ with $\deg p, \deg q > 1$ has $q(x) = \lambda(ax + bx^{-1})$ for some $a, b \in k^*$ and a $\lambda \in \mathrm{PGL}_2(k)$.

As to (H2), note that $h(x) = h(\eta x^{-1})$. If we also have $h(x) = h(cx^{-1})$ for some $c \in k^*$, then $h(x) = h((c/\eta)x)$, whence $c/\eta$ is a root of unity, say of exact order $n$, so $h(x) = h^*(x^n)$ for some $h^* \in k(x)$. Since this can happen only for $n = 1$, we see that $c = \eta$, so in particular $\eta$ is uniquely determined.

Note also that since $h$ is not of the forbidden shape, under (H2) we necessarily have $\deg \tilde{h} > 1$.

We prove this reduction to either (H1) or (H2) by induction on $\deg h$. We already noticed that we may suppose $h \notin k(x^n), n > 1$. Assume now that the theorems are proved in degrees lower than $\deg h$, the case $\deg h = 1$ being empty since $h$ is then of the forbidden shape.

First assume that $h$ is indecomposable. Then it satisfies (H1) and we are done if we know the result under this assumption.

Let us now assume that $h$ is decomposable. We separately consider the cases of Theorem 1.4(ii) and Theorem 1.5.

In the case of Theorem 1.4(ii) we know that $h \in k[x, x^{-1}]$ and that $h(x)$ is not of the shape $ax^n + b + cx^{-n}, a, b, c \in k, n \in \mathbb{N}$. Suppose that we have a decomposition $h = p \circ q$. By Remark 1.3 we may further assume (on replacing $p, q$ resp. by $p \circ \lambda^{-1}, \lambda \circ q$ for a suitable $\lambda \in \mathrm{PGL}_2(k)$) that $q$ is a Laurent polynomial.

If in some such decomposition, $q(x)$ is not of the forbidden shape, we may replace $g, h$ resp. by $g \circ p, q$, and apply induction on $\deg h$. Otherwise (i.e. if we cannot find such a decomposition) in every decomposition $h(x) = p(q(x))$ with $q \in k[x, x^{-1}]$, the inner factor $q(x)$ is necessarily of the forbidden shape (with $n = 1$). This says that either (H1) or (H2) holds.

Now we come to Theorem 1.5. Clearly, $h$ is not of the forbidden shape since it satisfies NL. Let $h(x) = p(q(x))$ be a decomposition of $h$; then by the assumption for Theorem 1.5 either $q$ satisfies NL or (again on replacing $p, q$ resp. by $p \circ \lambda^{-1}, \lambda \circ q$ for a suitable $\lambda \in \mathrm{PGL}_2(k)$) we have $q(x) = ax + bx^{-1}, a, b \in k$. (Observe in fact that $q$ does not lie in $k(x^n)$ for any $n > 1$ because $h$ does not.) If $q$ satisfies NL, then every decomposition of $q$ has an inner composition factor satisfying NL or being forbidden, because otherwise $h$ would not have this property; so we can argue by induction as before and replace $h$ by $q$. Sooner or later we arrive at a stage where either $h$ is indecomposable or for every such decomposition of $h$, we have $q(x)$ of the forbidden shape $\lambda(ax + bx^{-1}), a, b \in k, \lambda \in \mathrm{PGL}_2(k)$. This means that (H1) or (H2) is satisfied, as desired.

Altogether it follows that, in proving either Theorem 1.4(ii) or Theorem 1.5, we may assume that (H1) or (H2) holds.

Hence, let us assume from now on that $h$ is not of the forbidden shape but that it satisfies either (H1) or (H2).

Furthermore, from now on, if we use the symbols $\tilde{h}$ resp. $\eta$, we always refer to the special decomposition for $h$ appearing in (H2); in particular, $\eta$ denotes the (unique) constant appearing therein.

To go on, we again distinguish two cases, in which we slightly modify the notation of Lemma 2.4 by changing the sign of $e$ so that it is always positive:

- $h(0) = \infty$; now we may assume $\deg h > 2$ since otherwise $h(x)$ would be of the forbidden shape. Hence, we may write
$$h(x) = \frac{\tilde{p}(x)}{x^e \tilde{q}(x)}$$
with $e > 0$, $\tilde{p}, \tilde{q} \in k[x]$ coprime satisfying $\tilde{p}(0)\tilde{q}(0) \neq 0$, $d = \deg \tilde{p} = \deg h > 2$, and finally $s = d - e - \deg \tilde{q} > 0$.
- $h(0) \neq \infty$; by replacing $g, h$ resp. by $g \circ \lambda^{-1}, \lambda \circ h$ with $\lambda \in \mathrm{PGL}_2(k)$ defined by $\lambda(x) = x - h(0)$, we may assume $h(0) = 0$ and thus we may write
$$h(x) = \frac{x^e \tilde{p}(x)}{\tilde{q}(x)}$$
with $e > 0$, $\tilde{p}, \tilde{q} \in k[x]$ coprime satisfying $\tilde{p}(0)\tilde{q}(0) \neq 0$, $d = e + \deg \tilde{p} = \deg h \geq 2$, and finally $s = \deg \tilde{q} - d > 0$.

(We point out that throughout the proof the symbols $\tilde{p}, \tilde{q}, e, s$ and $d$ will only be used to refer to the quantities related to $h$ that are introduced above.)

In the first case, namely for $e < 0$ or equivalently $h(0) = \infty$, we have the following important lemma that plays a central role in the proofs.

Before stating it, we anticipate a simple remark that will be useful throughout: If $K$ is a function field over $k$ containing both $x, y$, with $h(x) = h(y)$, then
$$\deg_K(x) = \deg_K(y).$$
In fact, this follows from the general equality $\deg_K(h(u)) = (\deg h)\deg_K(u)$ for any $u \in K \setminus k$.

**Lemma 3.1.** *Assume that $h(0) = \infty$ and that either* (H1) *or* (H2) *holds. Then there exists a conjugate $y$ of $x$ over $k(h)$, different from $x$ and also from $\eta x^{-1}$ in case* (H2)*, such that the following holds, on putting $K = k(x, y)$:*

(i) *if $u \in k(x)$ is such that $u(x) = u(y)$, then $u \in k(h)$,*
(ii) *for all $m, n \in \mathbb{N}$ we have*
$$\deg_K(x^m y^n) \geq 0.12 \max\{m, n\} \deg_K(x).$$

*Moreover, if $\min\{e, s\} \leq 0.75 \max\{e, s\}$, then* (ii) *holds for all $m, n \in \mathbb{Z}$.*

We remark that the precise numerical values here are relatively immaterial; for instance, any larger number in place of 0.75 would lead to a similar statement with 0.12 replaced by some corresponding positive number. Ultimately, this would only result in different numerical constants in the Main Theorem. (We only observe that the larger constant we get in the lower bound (ii), the better numerical constant will appear in the theorems.)

*Proof.* Let $y \neq x$ be a conjugate of $x$ over $k(h)$, i.e. we have $h(x) = h(y)$. We define $L = \{u \in k(x) : u(x) = u(y)\}$. Note that $L$ is a subfield of $k(x)$ and that $k(h) \subseteq L \subseteq k(x)$. Since $y$ is different from $x$, we have $L \neq k(x)$. By the Lüroth Theorem (cf. [15]) it follows that $L = k(z)$ for a $z \in k(x)$; thus we get $h(x) = h^*(z)$ with $h^* \in k(x)$. Since $L \neq k(x)$, we have $\deg z > 1$. Now let us distinguish between (H1) and (H2).

We first consider (H1): In this case $h$ is indecomposable, so since $\deg z > 1$, we have $\deg h^* = 1$, i.e. $L = k(h)$. It clearly follows that requirement (i) is satisfied for any such $y$.

Now we turn to (H2): We have $h(x) = \tilde{h}(q(x))$ where $q(x) = x + \eta x^{-1}$ and $\eta \neq 0$. Then $h(\eta x^{-1}) = h(x)$ (because $q(\eta x^{-1}) = q(x)$). Hence $\eta x^{-1}$ is a conjugate of $x$ but we have excluded it too. (In fact recall that working under (H2), we assume $y \neq x$ and $y \neq \eta x^{-1}$.)

Observe that since $h(x)$ is not of the forbidden shape, we have $\deg \tilde{h} > 1$, so in particular $\deg h > 2$.

Going back to the above decomposition, we assume first that $\deg h^* > 1$. Then $h = h^* \circ z$ with $\deg z > 1$ and $\deg h^* > 1$. By (H2) this implies $z(x) = \lambda(ax + bx^{-1})$ for some $a, b \in k^*$ and $\lambda \in \mathrm{PGL}_2(k)$. We have already seen (in the discussion below (H2)) that this implies $b/a = \eta$. It follows that $q(x) = c\lambda^{-1}(z)$ for $c = 1/a \in k$, and since $L = k(z)$, we also have $q \in L$, which means $q(x) = q(y)$. But this is a quadratic equation in $y$ which has the two roots $y = x$ and $y = \eta x^{-1}$, and we have excluded them from the beginning. Therefore this case cannot occur.

Thus we have $\deg h^* = 1$. However, this means that $L = k(h)$ and so (i) is satisfied in this case too.

It follows that, apart from the choice $y = x$ and possibly $y = \eta x^{-1}$ (which may only occur in case (H2)), any other conjugate $y$ of $x$ over $k(h)$ will automatically satisfy (i). (Note that such a conjugate certainly exists, because we have $\deg h > 2$ since otherwise $h(0) = h(\infty) = \infty$ implies that $h(x)$ is of the forbidden shape.)

We now turn to (ii) and to the additional statement.

For (ii) let $z = x^m y^n$ with $m, n \in \mathbb{N}$. We shall show that there exists a suitable choice for the conjugate $y$ ($y \neq x$ and $y \neq \eta x^{-1}$ in case (H2)) so as to satisfy (ii) for all such $m, n$; this choice will be made along the proof because we think this leads to better clarity.

However, let us immediately note that if $|m - n| \geq 0.12 \max\{m, n\}$, then our inequality is true for any choice of the conjugate. In fact, taking into account that $\deg_K(x) = \deg_K(y)$, from the properties of the degree we infer that in this case we have

$$\deg_K(z) \geq (\max\{m, n\} - \min\{m, n\}) \deg_K(x) \geq 0.12 \max\{m, n\} \deg_K(x),$$

proving (ii).

Therefore from now on in the proof of (ii), we shall choose the sought-for conjugate assuming that $m \geq 0.88n > 0$.

For a given choice of the conjugate $y$, we also put $\delta = \delta_y := [k(x, y) : k(x)]$, so $\deg_K(x) = \deg_K(y) = \delta$.

Finally, we put $\varepsilon = 0.14$. We will show that $\deg_K(z) \geq \varepsilon m\delta$; this then implies that $\deg_K(z) \geq 0.88\varepsilon \max\{m, n\}\delta \geq 0.12 \max\{m, n\}\delta$ and thus the desired inequality.

We count the zeros of $z$ through the zeros of $x$. Assume that $v$ is a zero of $x$, i.e. $v(x) > 0$. We have now two possibilities for $y$:

If $v(y) \geq 0$, then $v(z) \geq mv(x) \geq \varepsilon mv(x) > 0$.

If $v(y) < 0$, then by Lemma 2.4 (or just by using the equation $h(x) = h(y)$) we have $-sv(y) = ev(x)$ and therefore

$$v(z) = \left(m - n\frac{e}{s}\right)v(x) = m\left(\frac{s}{e} - \frac{n}{m}\right)\frac{e}{s}v(x).$$

If

$$\frac{s}{e} - \frac{n}{m} \geq \varepsilon\frac{s}{e}, \tag{2}$$

then we get

$$\deg_K(z) = \sum_v \max\{0, v(z)\} \geq \sum_{v(x)>0} \max\{0, v(z)\} \geq \varepsilon m \sum_{v(x)>0} v(x) = \varepsilon m\delta,$$

because $\delta = \deg_K(x) = \sum_{v(x)>0} v(x)$.

The same inequality also follows if we assume that, in the irreducible factor of $h(Y) - h(x)$ defining $y$, there are at least $\varepsilon\delta$ Puiseux series at $x = 0$ of type 1 or 3, in the notation of Lemma 2.4; in fact, in that case we have

$$\deg_K(z) \geq m \sum v(x) \geq \varepsilon m\delta,$$

where the sum is taken over all zeros of $x$, and corresponding to a Puiseux series at $x = 0$ of type 1 or 3 in Lemma 2.4.

So in proving our conclusion let us assume that (2) is not true and that the number of Puiseux series at $x = 0$ of type 1 and 3 is less than $\varepsilon\delta$.

Now we count the poles of $z$ through the poles of $x$.

If $v(x) < 0$, then either $v(y) \leq 0$ and therefore $v(z) \leq mv(x) \leq \varepsilon mv(x) < 0$, or $v(y) > 0$. In the last case, again by Lemma 2.4, we have $ev(y) = -sv(x)$, which implies

$$v(z) = m\left(\frac{e}{s} - \frac{n}{m}\right)\frac{s}{e}v(x).$$

As before we get $\deg_K(z) \geq \varepsilon m\delta$ unless

$$\frac{e}{s} - \frac{n}{m} < \varepsilon\frac{e}{s}$$

and unless the number of Puiseux series at $x = \infty$ of type 1 and 3 is less than $\varepsilon\delta$.

In the remaining cases we have by the last displayed inequality and by (2) that

$$(1-\varepsilon)\frac{e}{s} < \frac{n}{m} \quad \text{and} \quad (1-\varepsilon)\frac{s}{e} < \frac{n}{m}.$$

But since $m \geq 0.88n > 0$ we get

$$\frac{\max\{e,s\}}{\min\{e,s\}} < \frac{n}{(1-\varepsilon)m} \leq \frac{1}{0.88(1-\varepsilon)} \leq \frac{1}{0.75}.$$

It follows that

$$\min\{e,s\} > 0.75\max\{e,s\}. \tag{3}$$

Moreover, we also see that the number of Puiseux series of type 1 and 3 at $x=0$ and at $x=\infty$ is less than $\varepsilon\delta$ in every irreducible factor defining a $y \neq x, \eta x^{-1}$ over $k(h)$.

The number of Puiseux series at $x=0$ of type 1 and 3 in the Puiseux factorization of the numerator of $(h(Y)-h(x))/(Y-x)$ is $e+\deg\tilde{q}-1 = d-s-1$, which follows from Lemma 2.4. (The subtraction of 1 comes from the factor $Y-x$ which clearly contains exactly one series of type 1 or 3 at $x=0$, whereas the factor $xY-\eta$, in case (H2), has no series of this type at $x=0$ at all.)

Since we are in the case where we assume that in each irreducible factor, of degree $\delta$, of the numerator of $(h(Y)-h(x))/(Y-x)$ there are at most $\varepsilon\delta$ such series, by summing up the contributions from each factor we conclude that $d-s-1 < \varepsilon(d-1)$ and $(1-\varepsilon)d+\varepsilon-1 < s$. From (3) we get $0.75s < e$. Hence, we obtain $d-\deg\tilde{q} = e+s > 1.75s > 1.75(1-\varepsilon)d-1.75(1-\varepsilon)$ and therefore $0 \leq \deg\tilde{q} < (-0.75+1.75\varepsilon)d+1.75(1-\varepsilon) \leq 3(-0.75+1.75\varepsilon)+1.75(1-\varepsilon) = -0.01 < 0$. (In fact we have used that by our choice $\varepsilon = 0.14$ we have $-0.75+1.75\varepsilon < 0$ and that $d \geq 3$.)

This contradiction implies that in at least one irreducible factor the number of Puiseux series at $x=0$ of type 1 and 3 is $\geq \varepsilon\delta$ and by choosing $y$ as a root corresponding to such a factor we deduce the desired inequality. (We could have also argued with the number of Puiseux series at $x=\infty$ of type 1 and 3, using the fact that $s > 0.75e$.)

Finally, we prove the remaining part of the statement. We have already seen that, if $\min\{e,s\} \leq 0.75\max\{e,s\}$, then we get the desired lower bound for $\deg_K(z)$ for all positive exponents independently of the choice of the irreducible factor from which we take $y$ (excluding, as always, $y=x$ and $y=\eta x^{-1}$ in case (H2)). Observe also that the case of both exponents negative follows trivially from the positive case since $\deg_K(z) = \deg_K(1/z)$.

Now we consider exponents of opposite sign. Let $z = x^m y^n$ with $m,n \in \mathbb{Z}$ and $mn < 0$. As above, if $\big||m|-|n|\big| \geq 0.12\max\{|m|,|n|\}$, then the desired inequality holds independently of the choice of the conjugate.

Therefore, and since $\deg_K(z) = \deg_K(1/z)$, we may just consider the cases $z = x^m y^{-n}$ with $m \geq 0.88n > 0$. We put $\varepsilon = 0.14$ as before. We will again show that $\deg_K(z) \geq \varepsilon m\delta$, from which it then follows $\deg_K(z) \geq 0.88\varepsilon\max\{m,n\}\delta \geq 0.12\max\{m,n\}\delta$, proving the desired inequality.

As above, we count the zeros of $z$ through the zeros of $x$. Let $v$ be a zero of $x$. If $v(y) \leq 0$, then $v(z) \geq mv(x)$. Therefore, if there are $\geq \varepsilon\delta$ Puiseux series at $x=0$ of type 2 or 3 in at least one irreducible factor, then we can choose it to define $y$, and the claimed

inequality follows. Similarly, arguing with the poles of $x$, it suffices that at least one of the factors contains $\geq \varepsilon\delta$ Puiseux series at $x = \infty$ of type 2 or 3 since then we again get the claimed inequality for the corresponding $y$.

We now show that the assumption $\min\{e, s\} \leq 0.75 \max\{e, s\}$ implies the existence of such a factor.

Assume that $0.75s \geq e$ and that at $x = 0$ all factors different from $Y - x$, $xY - \eta$ have at most $\varepsilon\delta$ Puiseux series at $x = 0$ of type 2 or 3. The arguments under (H1) and (H2) are slightly different; we first consider (H2): Then $d - e - 1 = s + \deg \tilde{q} - 1 < \varepsilon(d - 1)$ (now subtraction of 1 comes from the factor $xY - \eta$ which clearly contains exactly one series of type 2 or 3 at $x = 0$, whereas the factor $Y - x$ has no series of this type at $x = 0$ at all). It follows that $d - \deg \tilde{q} = e + s \leq 1.75s < 1.75\varepsilon(d - 1) - 1.75 \deg \tilde{q} + 1.75$ and therefore $0 \leq 0.75 \deg \tilde{q} < (1.75\varepsilon - 1)d + 1.75(1 - \varepsilon) \leq 3(1.75\varepsilon - 1) + 1.75(1 - \varepsilon) = 3.5\varepsilon - 1.25 = -0.76 < 0$, a contradiction (by our choice $\varepsilon = 0.14$). In case (H1) the assumption that at $x = 0$ all factors different from $Y - x$ (now there is no factor $xY - \eta$) have at most $\varepsilon\delta$ Puiseux series of type 2 or 3 implies $d - e = s + \deg \tilde{q} < \varepsilon d$, and then the contradiction follows from $0 \leq 0.75 \deg \tilde{q} < (1.75\varepsilon - 1)d < 0$ by our choice of $\varepsilon$.

If $0.75e \geq s$ one can argue similarly with the Puiseux series at $x = \infty$ by using the fact that $\varepsilon = 0.14$. Thus the proof of the statement is complete. □

In the second case, namely $e > 0$ or equivalently $h(0) = 0$, the situation is simpler and we need less. Here we have the following

**Lemma 3.2.** *Assume that* $h(0) = 0$ *and that either* (H1) *or* (H2) *holds. Then there exists a conjugate* $y$ *of* $x$ *over* $k(h)$*, different from* $x$ *and also from* $\eta x^{-1}$ *in case* (H2)*, such that the following holds:*

(i) *if* $u(x) = u(y)$, $u \in k(x)$, *then* $u \in k(h)$,
(ii) *for all* $m, n \in \mathbb{N}$ *we have*

$$\deg_K(x^m y^n) \geq \max\{m, n\} \deg_K(x),$$

*where* $K = k(x, y)$.

*Proof.* For (i) we can argue as in the previous lemma; just observe that (H2) is clearly excluded since we have $h(\infty) = \infty$ and $h(0) = 0$, hence $h(x)$ cannot be of the shape $\tilde{h}(x + \eta x^{-1})$ with $\eta \neq 0$. It follows that for every conjugate $y$ of $x$ over $k(h)$ that is different from $x$, condition (i) will be satisfied. Let $y$ be such a conjugate and define $\delta = [k(x, y) : k(x)]$.

Now we prove (ii). Let $z = x^m y^n$, $m, n \in \mathbb{N}$. We may assume that $m \geq n \geq 0$ since otherwise we just have to exchange the roles of $x$ and $y$ (observe that here we have already chosen the conjugate $y$) in the arguments below. As in the proof of Lemma 3.1 we count the zeros of $z$ by going through the zeros of $x$. Assume that $v(x) > 0$. Clearly, $v(y) < 0$ is impossible and so we have $v(y) \geq 0$. In this case we get $v(z) \geq mv(x)$. It immediately follows that

$$\deg_K(z) \geq m\delta$$

independently of the factor from which we choose $y$. This already proves the statement. □

Note that in both cases for $h$ (that is, for both possible values for $h(0)$) we find, for every conjugate $y$ of $x$ over $k(h)$, that $\deg_K(x) = \deg_K(y) \le d - 1$ (as follows immediately from $h(x) = h(y)$).

**Further strategy.** Let us emphasize that the lower bounds obtained in the last two lemmas work only on additional conditions, namely concerning the ratio $s/e$ or the signs of $m, n$. In the non-Laurent case and for $h(0) = 0$ there are now different devices which prevent us from the need of more, but in the remaining situations, i.e. in the Laurent case with $h(0) = \infty$, we need a bound which does not depend on these issues: Therefore, to overcome this difficulty, we start there by choosing a factor (or equivalently a conjugate $y$) so that Lemma 3.1(i) holds. Further, if $s/e$ is not near to 1 we can choose it so that the bound holds for all $m, n$, whereas if $s/e$ is near to 1, we can only say that the bound holds for non-negative $m, n$ for an appropriate choice of the factor. Hence, no difficulty arises if $s/e$ is not near to 1. In the case when $s/e$ is near to 1, we could first choose the factor so that the bound holds for non-negative $m, n$, but then the factor is fixed and we cannot change it anymore. This means that if we want the estimate for all $m, n \in \mathbb{Z}$ we need *a priori* more information on the possible factors, and this is the place where the result of Müller [11, 12] will play an essential role (see Subsection 5.2 for the precise statement and the detailed proof of this).

After these remarks we are now ready to prove Theorems 1.4(ii) and 1.5, which we will do in the following two sections. Before we start we summarize, for the reader's convenience, some points of the proof.

We are arguing by induction on $\ell$ and, for given $\ell$, also by a second induction on $\deg h$.

Further, we recall the normalizations obtained so far: We may assume that $P$, $Q$ are not both divisible by $x$, that $g(\infty) = h(\infty) = \infty$, that $h(0) \in \{0, \infty\}$ and that $h$ satisfies (H1) or (H2). (This last reduction has been shown in Section 3, before Lemma 3.1.)

## 4. The non-Laurent case for rational functions

In this section we prove Theorem 1.5. Let $f(x) = g(h(x))$ with the notation and normalizations just recalled.

Note that the assumption that $h$ satisfies NL implies that $\deg \tilde{q} \ge 1$ and $\deg \tilde{p} \ge 1$, where we refer to the equations $h(x) = x^{\pm e}\tilde{p}(x)/\tilde{q}(x)$ displayed just before Lemma 3.1. (E.g., if $\tilde{p}$ were constant, $1/h$ would be a Laurent polynomial.)

As before, we shall consider two cases depending on whether $h(0) = \infty$ or $h(0) = 0$.

### 4.1. *Proof of Theorem 1.5 in the case* $h(0) = \infty$

First we shall use the lacunarity of $f$ to partition the set $\Sigma$ of exponents appearing in $P$, $Q$ into two groups with a controlled gap. We write

$$\Sigma = \{0 = m_1 < m_2 < \cdots < m_\ell\}$$

for the set of exponents appearing in $P$, $Q$ (ordered by size); in particular, we have $m_\ell \geq \deg f$ (with equality if $P$, $Q$ are coprime). Observe that $|\Sigma| = \ell \geq 2$. We partition $(0, m_\ell]$ into infinitely many intervals $(m_\ell/3, m_\ell]$, $(m_\ell/9, m_\ell/3]$, .... It follows that at least one interval, say $(m_\ell/3^l, m_\ell/3^{l-1}]$ with $2 \leq l \leq \ell$, is disjoint from $\Sigma$. We set

$$\Sigma_L = \{m \in \Sigma : m > m_\ell/3^{l-1}\}, \qquad \Sigma_S = \{m \in \Sigma : m \leq m_\ell/3^l\}.$$

So $\Sigma = \Sigma_L \cup \Sigma_S$ and any difference $m' - m$ with $m' \in \Sigma_L$, $m \in \Sigma_S$ satisfies

$$m' - m > (2/3^l)m_\ell.$$

Moreover, the difference between any two elements in $\Sigma_S$ is at most $m_\ell/3^l$.

Let $y$ be the conjugate of $x$ over $k(h)$ from Lemma 3.1. From the equation $f(x) = g(h) \in k(h)$ we deduce $f(x) = f(y)$. Writing $f(x) = P(x)/Q(x)$ we get the equation

$$P(x)Q(y) - P(y)Q(x) = 0.$$

This gives a relation involving terms of the form $x^m y^n$ with $m, n \in \Sigma$ and with at most $\ell^2 - 1$ such summands. (Recall in fact that $f(\infty) = \infty$, so $m_\ell$ appears only in $P$ and this implies that $x^{m_\ell} y^{m_\ell}$ does not occur in the relation above.)

We say that a term of the said form is of type LL if $(m, n) \in \Sigma_L \times \Sigma_L$; similarly terms of type SL, LS and SS are defined.

We partition the sum on the left of this equation into minimal vanishing subsums, where no further proper subsum vanishes. (This partition may be done in several ways; we can choose freely one of them.) Observe that in order to apply Lemma 2.2 to one of these minimal vanishing subsums we first have to normalize it, dividing by one of the terms, so that the constant term 1 appears in the sum.

Let us assume that there is such a minimal vanishing subsum containing a term of type LL and another one of type SS. The ratio $z$ of these terms is of shape $x^m y^n$ where $m, n \in \Sigma_L - \Sigma_S$. So $m$ and $n$ are $> (2/3^l)m_\ell$. An application of Lemma 3.1 leads to

$$\deg_K(z) \geq \frac{2\varepsilon}{3^l} m_\ell \delta \geq \frac{2\varepsilon}{3^\ell} m_\ell \delta,$$

where $K = k(x, y)$, $\delta = \deg_K(x) = \deg_K(y) = [K : k(x)]$ and $\varepsilon = 0.12$. Then, by Lemma 2.2, we get

$$\frac{2\varepsilon}{3^\ell} m_\ell \delta \leq \binom{\ell^2 - 2}{2} (2\mathfrak{g} - 2 + |S|),$$

where $\mathfrak{g}$ is the genus of $k(x, y)$. We have

$$\mathfrak{g} \leq (\delta - 1)^2$$

(where we use the fact that $\delta = \deg_K(x) = \deg_K(y)$, and therefore the bound follows from Castelnuovo's inequality; cf. [15, III.10.3 Theorem]).

Further, the set $S$ consisting of the zeros and poles of $x$ and $y$ in $K$ satisfies $|S| \leq 2(\deg_K(x) + \deg_K(y)) \leq 4\delta$ and so $2\mathfrak{g} - 2 + |S| \leq 2(\delta-1)^2 - 2 + 4\delta = 2\delta^2$. Therefore we get

$$m_\ell \leq \frac{3^\ell}{2\varepsilon}\binom{\ell^2-2}{2}2\delta \leq \frac{3^\ell}{\varepsilon}\binom{\ell^2-2}{2}(d-1) \leq (267\cdot 5^\ell)(d-1),$$

where we have used the inequality

$$\binom{\ell^2-2}{2} = \frac{(\ell^2-2)(\ell^2-3)}{2} \leq 32\cdot\left(\frac{5}{3}\right)^\ell. \tag{4}$$

(Here the constant 32 is obtained by bringing the exponential term to the left and calculating the maximal values of the resulting function in $\ell$ by differentiation; this gives a polynomial equation of degree 4 that has its positive roots at $1.5717\ldots$ and $8.1381\ldots$.)

If there is a minimal vanishing subsum that involves both terms of type LS and of type SS, then the ratio $z$ is of the shape $z = x^m y^n$ with $m \geq (2/3^l)m_\ell$ and $|n| \leq m_\ell/3^l$. Hence,

$$\deg_K(z) \geq |m|\deg_K(x) - |n|\deg_K(y) = (|m| - |n|)\delta \geq \frac{1}{3^\ell}m_\ell\delta,$$

leading to an even better bound than before. We can argue similarly for subsums involving terms of type SL and SS.

We may therefore assume that each minimal vanishing subsum involves either just terms of the type SS or the rest (i.e. $(\Sigma_L\times\Sigma_S)\cup(\Sigma_S\times\Sigma_L)\cup(\Sigma_L\times\Sigma_L)$). Let us then write

$$P(x) = a(x) + A(x), \quad Q(x) = b(x) + B(x),$$

where the capitals involve precisely those exponents which lie in $\Sigma_L$. The full relation is

$$\begin{aligned} a(x)b(y) - a(y)b(x) + A(x)B(y) - A(y)B(x) \\ + A(x)b(y) - A(y)b(x) + a(x)B(y) - a(y)B(x) = 0, \end{aligned}$$

and the above shows that we may assume

$$a(x)b(y) = a(y)b(x).$$

Now we show that we can also reduce to the assumption $a(x)b(x)\neq 0$. First, $a(x)\neq 0$ since $a(x)$ involves a constant term; this follows from $P(0)\neq 0$, which in turn is implied by $f(0) = g(h(0)) = g(\infty) = \infty$. (Recall also that, although $P$, $Q$ are not supposed to be coprime, we are working under the harmless normalization that they do not both vanish at 0.)

For the reduction to the crucial fact that $b(x)\neq 0$ we will use that we are in the non-Laurent case (in fact we only need $\deg\tilde{q}\geq 1$, which is implied by NL). We start by writing

$$f(x) = \frac{P(x)}{Q(x)} = \frac{P(x)}{x^E\tilde{Q}(x)}$$

with $E > 0$, $P(0)\tilde{Q}(0) \neq 0$. (Observe again that $f(0) = g(h(0)) = g(\infty) = \infty$.) Moreover, $P(x)$ and $x^E\tilde{Q}(x)$ have $\ell - 1$ non-constant terms altogether. We write

$$g(x) = ax^{\operatorname{ord}_\infty(g)} + \cdots$$

up to terms of smaller order at $x = \infty$, where $a \in k^*$. Recall that $h$ satisfies NL, so $\tilde{q}$ is not constant and we may pick a root $\theta \in k^*$ of $\tilde{q}$, say of multiplicity $\mu \geq 1$. Then at $x = \theta$ we have

$$g(h(x)) = \frac{b}{(x-\theta)^{\mu \operatorname{ord}_\infty(g)}} + \cdots$$

for some $b \in k^*$. This implies that $\tilde{Q}(x)$ has $\theta$ as a root with multiplicity $\geq \mu \operatorname{ord}_\infty(g)$. So with Lemma 2.1 we get $\operatorname{ord}_\infty(g) \leq \mu \operatorname{ord}_\infty(g) \leq \operatorname{ord}_\theta(\tilde{Q}) \leq \ell - 1$. It follows that

$$E = \operatorname{ord}_0(f) = e \operatorname{ord}_\infty(g) \leq (\ell-1)e \leq (\ell-1)(d-1) \tag{5}$$

and thus $Q(x)$ involves the term $x^E$ with $E \leq (\ell-1)(d-1)$.

If $E \in \Sigma_L$, then

$$m_\ell \leq (\ell-1)3^{\ell-1}(d-1) \leq 5^\ell(d-1)$$

and we are done (here we can e.g. use the fact that the real function $\sqrt[x]{x}$ takes its maximum at $x = \exp(1) = 2.7182\ldots$).

Hence, we may assume that $E \in \Sigma_S$, implying $b(x) \neq 0$. Now this gives

$$\frac{a(x)}{b(x)} = \frac{a(y)}{b(y)}$$

and by the property of $y$ expressed in Lemma 3.1(i) we deduce

$$\frac{a(x)}{b(x)} = \varphi(h)$$

where $\varphi \in k(x)$. Let $x^M$ be the largest power of $x$ dividing both $A(x)$ and $B(x)$. Write $A(x) = x^M\tilde{A}(x)$, $B(x) = x^M\tilde{B}(x)$. Since $M \in \Sigma_L$, we note

$$M > m_\ell/3^{l-1}. \tag{6}$$

Furthermore, we have

$$f(x) = \frac{a(x) + x^M\tilde{A}(x)}{b(x) + x^M\tilde{B}(x)} = g(h).$$

Define $\psi(x) := g(x) - \varphi(x)$, so

$$\psi(h) = g(h) - \varphi(h) = f(x) - \frac{a(x)}{b(x)} = x^M\frac{\tilde{A}(x)b(x) - a(x)\tilde{B}(x)}{b(x)Q(x)}.$$

Now, by a Hajós argument similar to the one before, we first have

$$v_0(\psi(h)) \geq M - v_0(b) - v_0(Q) \geq M - 2E \geq M - 2(\ell-1)(d-1),$$

where we have used (5) again. We may assume that $M - 2(\ell-1)(d-1) > 0$ since otherwise

$$m_\ell < 2(\ell-1)3^{\ell-1}(d-1) \le (2\cdot 5^\ell)(d-1)$$

and we are done. It follows that $v_0(\psi(h)) = \mathrm{ord}_0(h)v_\infty(\psi) = ev_\infty(\psi) > 0$, and therefore $v_\infty(\psi) > 0$ since we have $e > 0$; in particular, we get $\psi(\infty) = 0$ and

$$e\, v_\infty(\psi) = v_0(\psi(h)) \ge M - 2(\ell-1)(d-1). \tag{7}$$

Now pick as before a $\theta \in k^*$ with $\tilde{q}(\theta) = 0$, so $h(\theta) = \infty$ and

$$v_\theta(\psi(h)) = v_\theta(\tilde{q})v_\infty(\psi) > 0.$$

Therefore $\tilde{A}(x)b(x) - a(x)\tilde{B}(x)$ vanishes at $\theta$ of order at least $v_\infty(\psi) > 0$.

Now, if $\tilde{A}(x)b(x) \ne a(x)\tilde{B}(x)$, the difference has $\le \ell^2$ terms (in fact every term has a degree of the form $m + m' - M$ with $m, m' \in \Sigma$). So by Lemma 2.1 again

$$v_\infty(\psi) \le v_\theta(\tilde{A}b - a\tilde{B}) \le \ell^2,$$

whence by (7) it follows that $M \le \ell^2 e + 2(\ell-1)(d-1) \le \ell(\ell+2)(d-1)$ and by using (6) we get

$$m_\ell < \ell(\ell+2)3^{\ell-1}(d-1) \le (2\cdot 5^\ell)(d-1)$$

and we are done. (The second inequality is obtained as in (4).)

On the other hand, suppose $\tilde{A}(x)b(x) = a(x)\tilde{B}(x)$. Since $\tilde{A}(x) \ne 0$ (otherwise $\Sigma_L$ would be empty), we get $\tilde{B}(x) \ne 0$ and so

$$\frac{\tilde{A}(x)}{\tilde{B}(x)} = \frac{a(x)}{b(x)} = \varphi(h) = f(x)$$

and we have expressed $f$ with a proper subset of $\Sigma$. In this case we can argue by induction on $\ell$, the case $\ell = 2$ being trivial (e.g. we can repeat all arguments of this proof and observe that $b(x) \ne 0$ is impossible since then $\Sigma$ contains at least the three elements $0, E, m_\ell$). Therefore $\deg\varphi \le 267\cdot 5^{\ell-1}$ and thus $\deg f = \deg\varphi \cdot d \le (267\cdot 5^\ell)(d-1)$.

Summing up we have proved that in all cases

$$\deg f \le m_\ell \le (267\cdot 5^\ell)(d-1),$$

which is the desired inequality. □

*4.2. Proof of Theorem 1.5 in the case $h(0) = 0$*

Recall that both $\tilde{p}, \tilde{q}$ are non-constant; in fact here we will only use the assumption $\deg \tilde{p} \geq 1$. We start by observing that we may also assume, on subtracting a constant from both $g$, $f$, that

$$g(0) \in \{0, \infty\}.$$

This is like saying that $v_0(g) \neq 0$ and it implies $f(0) = g(h(0)) = g(0) \in \{0, \infty\}$.

As before we partition the set $\Sigma$ consisting of the terms contained in the numerator and denominator of $f$ into the two disjoint sets $\Sigma_S$ and $\Sigma_L$. We investigate the equation $f(x) - f(y) = 0$, where $y$ is the conjugate obtained in Lemma 3.2, and partition the terms in $P(x)Q(y) - P(y)Q(x)$ into minimal sets with vanishing sum.

For each monomial $z = x^m y^n$ that appears as a ratio of a monomial of type LL, or LS, or SL and a monomial of type SS in a minimal vanishing subsum, we get

$$\deg_K(z) \geq \frac{1}{3^l} m_\ell \delta \geq \frac{1}{3^\ell} m_\ell \delta,$$

where $\delta = \deg_K(x)$. Thus, by Lemma 2.2,

$$\frac{1}{3^\ell} m_\ell \delta \leq \binom{\ell^2 - 2}{2} (2\mathfrak{g} - 2 + |S|).$$

We have $|S| \leq 4\delta$ and thus $2\mathfrak{g} - 2 + |S| \leq 2(\delta - 1)^2 - 2 + 4\delta = 2\delta^2$. Therefore we get

$$m_\ell \leq 3^\ell \binom{\ell^2 - 2}{2} 2\delta \leq (64 \cdot 5^\ell)(d - 1)$$

by using the inequality (4).

Otherwise we have again $a(x)b(y) - a(y)b(x) = 0$. Since $g(0) \in \{0, \infty\}$, we get

$$g(x) = cx^{v_0(g)} + \cdots,$$

where $c \in k^*$. Now as before we write

$$f(x) = \frac{P(x)}{Q(x)} = x^E \frac{\tilde{P}(x)}{\tilde{Q}(x)}$$

with $E \in \mathbb{Z}\backslash\{0\}$, $\tilde{P}(0)\tilde{Q}(0) \neq 0$. (Observe that $f(0) \in \{0, \infty\}$.) Moreover, $x^{\max\{0,E\}}\tilde{P}(x)$ and $x^{-\min\{0,E\}}\tilde{Q}(x)$ have $\ell - 1$ non-constant terms altogether. Since $\tilde{p}$ is not constant, we let $\theta \in k^*$ be a root of $\tilde{p}$ of multiplicity $\mu \geq 1$. We have $v_\theta(g(h)) = \mu v_0(g)$ and thus either $\tilde{P}(x)$ or $\tilde{Q}(x)$ has $\theta$ as a root with multiplicity $\geq \mu|v_0(g)|$. By Lemma 2.1 we get $\mu|v_0(g)| \leq \ell - 1$. Finally, we obtain

$$|E| = e|v_0(g)| \leq e(\ell - 1) \leq (\ell - 1)(d - 1).$$

We may assume that $|E| \in \Sigma_S$ (otherwise we are done), whence both $a(x)$ and $b(x)$ are non-zero since now one of the two contains the constant term and the other one $x^{|E|}$, and thus we infer that

$$\frac{a(x)}{b(x)} = \varphi(h)$$

and

$$\psi(h) = f(x) - \varphi(h) = x^M \frac{\tilde{A}(x)b(x) - a(x)\tilde{B}(x)}{b(x)Q(x)},$$

where $x^M$ is the largest power of $x$ dividing both $A(x)$ and $B(x)$. In particular, $M \in \Sigma_L$, i.e. $M > m_\ell/3^{l-1}$ (that is, (6) in the previous case). Now the factor $x^{|E|}$ may appear in the numerator or the denominator of the second factor. In any case

$$v_0(\psi(h)) \geq M - 2|E| \geq M - 2(\ell-1)(d-1).$$

We may again assume that $M - 2(\ell-1)(d-1) > 0$ (otherwise we get the asserted inequality) and thus $v_0(\psi) > 0$; in particular, $\psi(0) = 0$. Hence, we get

$$ev_0(\psi) = v_0(\psi(h)) \geq M - 2(\ell-1)(d-1).$$

Again if $\tilde{p}(\theta) = 0$ for $\theta \in k^*$, then $h(\theta) = 0$ and this gives

$$v_\theta(\psi(h)) = \mathrm{ord}_\theta(\tilde{p})v_0(\psi) \geq v_0(\psi) \geq \frac{1}{d-1}(M - 2(\ell-1)(d-1)).$$

If $\tilde{A}(x)b(x) - a(x)\tilde{B}(x) \neq 0$, then by Lemma 2.1 the left hand side of the above inequality is $\leq \mathrm{ord}_\theta(\tilde{A}b - a\tilde{B}) \leq \ell^2$ and we get the desired bound as before. Otherwise, the result follows by induction on $\ell$.

Putting all the upper bounds for $m_\ell$ together we get

$$\deg f \leq m_\ell \leq (267 \cdot 5^\ell)(d-1),$$

which is what we want. □

## 5. The Laurent case for rational functions

In this section we prove Theorem 1.4. We start with the case when $f(x) = P(x)/Q(x)$ is in reduced form, i.e. with $P, Q \in k[x]$ coprime; here we do not need any of the normalizations from Section 3.

### *5.1. Proof of Theorem 1.4 for reduced $f$*

We write $g(x) = g_1(x)/g_2(x)$ with $g_1, g_2 \in k[x]$ coprime and thus we get

$$\frac{P(x)}{Q(x)} = \frac{g_1(h(x))}{g_2(h(x))}.$$

Since the quotient on the left side is reduced and $g_1, g_2$ are coprime, it follows that $P(x)P_1(x) = g_1(h(x))$, $Q(x)Q_1(x) = g_2(h(x))$ with $P_1, Q_1$ units in $k[x, x^{-1}]$, i.e. of the form $x^{\pm n}$ with $n \in \mathbb{N}$. Now since $PP_1$ has the same number of terms as $P$, and $QQ_1$ has the same number of terms as $Q$, the result follows at once from [21, Theorem 2*]. □

We are left with the general case in which $P, Q$ may not be coprime.

Now we argue by induction on $\ell$ and we remind the reader that, by a further induction on $\deg h$, we have reduced to the case when either (H1) or (H2) holds. (We recall that this reduction has been shown in Section 3, before Lemma 3.1.)

Let $f(x) = g(h(x))$ with $h \in k[x, x^{-1}]$ and with $h(x)$ not of the shape $ax+b+cx^{-1}$, $a, b, c \in k$, where we again use the notation and normalizations described in Section 3. Therefore, we may now write

$$h(x) = x^{\pm e}\tilde{p}(x), \quad \tilde{p}(0) \neq 0,$$

and we again have to consider the cases $h(0) = \infty$ or $h(0) = 0$ (i.e. minus or plus sign respectively). Observe that $\deg \tilde{p} \geq 1$ since otherwise $h$ would be of the forbidden shape.

### 5.2. *Proof of Theorem 1.4 in the case* $h(0) = \infty$

We start by proving that there is a conjugate $y$ of $x$ over $k(h)$ satisfying (i) of Lemma 3.1 and

$$\deg_K(z) \geq \frac{1}{63}\max\{|m|, |n|\}\deg_K(x) \tag{8}$$

for all $z = x^m y^n$ with $m, n \in \mathbb{Z}$. We set $\delta = \deg_K(x) = \deg_K(y)$ as usual.

We first show this property whenever $y \neq x, \eta x^{-1}$ and $\deg_K(y) \leq 63$. Note that this last inequality holds, in particular, for all $h$ with $\deg h \leq 64$.

Observe that for every conjugate $y$ of $x$ over $k(h)$ different from both $x$ and $\eta x^{-1}$, condition (i) is fulfilled; this follows as in the proof of Lemma 3.1. Since $\deg h > 2$ such a conjugate certainly exists and we go on to show that it necessarily satisfies also (8) for the relevant values of $m, n$.

To show (8), note that the divisors of $x, y$ (in the function field $K = k(x, y)$) are not proportional (otherwise $xy^{\pm 1} \in k^*$, but in the opening arguments for Lemma 3.1 we have already seen that only $Y - x, xY - \eta$ would be admissible, and we have excluded these possibilities).

Assume first that $m \geq n \geq 0$. If there is a valuation $v$ with $v(x) > 0$ and $v(y) \geq 0$, then $v(z) = mv(x) + nv(y) \geq m \geq m \deg_K(x)/63$ and we are done.

Otherwise for every $v$ with $v(x) > 0$ we have $v(y) < 0$; it follows that there is one such $v$ with $v(x) \neq |v(y)|$ (here we use the assumption that the divisors are not proportional), and since $\deg_K(x) = \deg_K(y)$, there exists a $v$ with $v(x) > |v(y)| \geq 1$. Thus $v(z) = mv(x) - n|v(y)| \geq m + (m-n)|v(y)| \geq m$. Now assume that $m \geq -n > 0$. If there is a $v$ with $v(x) > 0$, $v(y) \leq 0$, then $v(z) \geq m$; otherwise we have $v(y) > 0$ for all $v$ with $v(x) > 0$ and as above there is such a $v$ with $v(x) > v(y)$, which then implies $v(z) = m + (m+n)v(y) \geq m$. It follows that

$$\deg_K(z) \geq \max\{|m|, |n|\} \geq \frac{1}{63}\max\{|m|, |n|\}\delta.$$

The key point in this argument is that $\deg h$ is absolutely bounded, and hence the same holds for $\delta \leq \deg h - 1$. So we do not have to worry about the dependence of the lower bound on $\delta$. In turn, this implies that we have a lower bound for $\deg_K(z)$ without using Lemma 3.1.

From now on we therefore assume that $\delta \geq 64$.

Now if $0.75 \max\{e, s\} \geq \min\{e, s\}$, the inequality (8) follows immediately from the additional statement in Lemma 3.1. We therefore assume that $0.75 \max\{e, s\} < \min\{e, s\}$ or equivalently $0.75e < s < e/0.75$.

If $h$ satisfies (H1), then $h$ is indecomposable and we apply Müller's theorem (cf. Lemma 2.3). Clearly, the sporadic cases are covered (by the condition $\delta \geq 64$). Thus let us now only look at infinite families and assume that $\delta \geq 64$.

If we are in case (i) of Lemma 2.3, i.e. if $G \supseteq \mathcal{A}_d$, then $(h(Y) - h(x))/(Y - x)$ is absolutely irreducible and thus we take it as the defining polynomial for $y$. We have $\delta = d - 1$ and clearly $y \neq x, \eta x^{-1}$, which implies (i) of Lemma 3.1. Here it follows, by using $s = d - e > d - (s/0.75)$ and $e = d - s > d - (e/0.75)$ (observe that $\deg \tilde{q} = 0$ and therefore there are no Puiseux series of type 3 in Lemma 2.4), that there are at least

$$s > \frac{d}{2.34} \geq \frac{\delta}{62}$$

Puiseux series at $x = 0$ of type 2 and at least

$$e - 1 > \frac{d}{2.34} - 1 \geq \frac{\delta}{62}$$

Puiseux series at $x = 0$ of type 1 in Lemma 2.4 in the factor defining $y$. Then the lower bound for $\deg_K(z)$ follows by the same arguments as used in the proof of Lemma 3.1: Let $z = x^m y^n$ with $m, n \in \mathbb{Z}$. If $\big||m| - |n|\big| \geq (1/63) \max\{|m|, |n|\}$, then we are done by properties of the degree, and if this is not the case, then it is enough to consider $z = x^m y^n$ with $m \in \mathbb{N}, n \in \mathbb{Z}$ and $m \geq (62/63)|n| > 0$. Since we have already proved that there are $\geq (1/62)\delta$ Puiseux series at $x = 0$ of type 1 and 2 in Lemma 2.4, it follows that $\deg_K(z) \geq (1/62)m\delta \geq (1/63) \max\{m, |n|\}\delta$. (This is obtained, as before, by counting the zeros of $z$ by going through the zeros of $x$ and using the fact that in both cases $n \geq 0$ resp. $n < 0$ there are $\geq (1/62)\delta$ places that contribute at least $m$ to the lower bound.) So we are done again.

Otherwise we are in case (ii) of Lemma 2.3. It follows that $d = m^2$ with $m \geq 8$ and $G = (\mathcal{S}_m \times \mathcal{S}_m) \rtimes \mathbb{Z}/(2)$. The orders of the distinct orbits induced by the action of $G$ (that are the degrees in $Y$ of the irreducible factors of $h(Y) - h(x)$) are $(m-1)^2$, $2m-2$ and 1. Let $y$ be defined by the factor corresponding to $(m-1)^2$. Clearly, (i) of Lemma 3.1 is satisfied. Since we are assuming that $0.75e < s < e/0.75$, which implies $s > m^2/2.34$ and $e > m^2/2.34$, similarly to the above we find that there are at least

$$e - 2m + 1 > \frac{m^2}{2.34} - 2m + 1 \geq \frac{1}{62}(m-1)^2$$

Puiseux series at $x = 0$ of type 1 and at least

$$s - 2m + 1 > \frac{1}{62}(m-1)^2$$

Puiseux series at $x = 0$ of type 2 in the defining equation of $y$ (with $\delta = (m-1)^2$). This again implies the asserted lower bound for $\deg_K(z)$.

Finally assume that $h$ satisfies (H2), so $h(x) = \tilde{h}(x+\eta x^{-1})$, $\eta \in k^*$. After rescaling $x$ we may assume that $\eta = 1$, i.e. $h(x) = \tilde{h}(x + x^{-1})$. Observe that $d = 2 \deg \tilde{h}$; we set $n := \deg \tilde{h} = d/2 \geq 32$.

Since $h$ is a Laurent polynomial, it follows that $\tilde{h} \in k[x]$ (otherwise $h$ would have a pole outside $\{0, \infty\}$) and since every decomposition $h = p \circ q$ has $q(x) = \lambda(ax + bx^{-1})$ for $a, b \in k$, $\lambda \in \mathrm{PGL}_2(k)$, it also follows that $\tilde{h}$ is indecomposable.

So we have (e.g. by [14, Theorem 10, p. 52]) three cases: Either $\tilde{h}$ is related to a cyclic or a Chebyshev polynomial (i.e. $h(x) = c_1(x + x^{-1} + c_2)^n + c_3$ or $h(x) = c_1 T_n(c_2(x+x^{-1})+c_3)+c_4$, $c_1, c_2, c_3, c_4 \in k$), or $H(U, V) := (\tilde{h}(U) - \tilde{h}(V))/(U - V)$ is absolutely irreducible.

In the first two cases $H(U, V)$ splits into factors of degree 1 or, by [14, Lemma 1, p. 52], into factors of degree 2 and therefore $\delta \leq 4$. Thus we can forget about these cases since we are assuming here that $\delta \geq 64$.

Hence, in the following we assume that we fall into the third case, i.e. $H$ is absolutely irreducible. We now apply a result analogous to Müller's quoted above, but for the simpler case of polynomials (instead of Laurent polynomials); it appears in Müller's paper [11, Theorem 4.9, p. 63] (but is proved elsewhere). We may already forget about the sporadic cases and thus consider only the infinite families (given at the beginning of the cited statement). The first cases (a), (b) correspond to the cyclic and Chebyshev case (in fact now the Galois group structure says that $H$ is not absolutely irreducible). Hence these cases have already been taken into account.

Therefore we concentrate on case (c), namely we assume that the Galois group $G$ of the equation $\tilde{h}(U) - \tilde{h}$ over $k(\tilde{h})$ is either $\mathcal{A}_n$ or $\mathcal{S}_n$ (remember that now $n = \deg \tilde{h}$); recall moreover that $n \geq 32$. We denote by $L$ the splitting field of this equation over $k(\tilde{h})$. We let $u, v \in L$ be distinct solutions to this equation, so $u \neq v$ and $\tilde{h} = \tilde{h}(u) = \tilde{h}(v)$.

We have the field inclusions $k(\tilde{h}) \subseteq k(u) \subseteq k(u, v) \subseteq L$. The Galois group of $L/k(u)$ corresponds to the stabilizer in $G$ of 1, say, whereas the Galois group of $L/k(u, v)$ corresponds to the stabilizer of both 1 and 2. These two subgroups in turn correspond to natural inclusions, either $\mathcal{A}_{n-2} \subseteq \mathcal{A}_{n-1}$ or $\mathcal{S}_{n-2} \subseteq \mathcal{S}_{n-1}$ (depending on whether $G$ is $\mathcal{A}_n$ or $\mathcal{S}_n$).

Suppose now that there is a field $F$, quadratic over $k(u)$ and with $k(u) \subseteq F \subseteq k(u, v)$. This would correspond to a group $\Gamma$ of index 2 either in $\mathcal{A}_{n-1}$ or in $\mathcal{S}_{n-1}$, and with either $\mathcal{A}_{n-2} \subseteq \Gamma \subseteq \mathcal{A}_{n-1}$ or $\mathcal{S}_{n-2} \subseteq \Gamma \subseteq \mathcal{S}_{n-1}$ in the respective situations for $G$. Say that $G = \mathcal{A}_n$: since $n \geq 32$, $\mathcal{A}_{n-1}$ is simple, so we cannot have $[\mathcal{A}_{n-1} : \Gamma] = 2$. Similarly if $G = \mathcal{S}_n$: we would have $\Gamma = \mathcal{A}_{n-1}$, contrary to $\mathcal{S}_{n-2} \subseteq \Gamma$. We conclude that such a quadratic extension does not exist.

Let us now define $x$ as a solution of $X + X^{-1} = u$, in an algebraic closure of $k(u)$ containing $L$; so $x + x^{-1} = u$. The other solution is $x^{-1}$, so $k(x)$ is independent of the solution, and $k(x)/k(u)$ is quadratic. Hence, $k(x)$ is not included in $k(u, v)$, and therefore $x$ has degree 2 over $k(u, v)$.

Similarly, we may define $y$ as a solution of $Y + Y^{-1} = v$ in the same algebraic closure, and then $y$ also has degree 2 over $k(u, v)$.

Note that $\tilde{h}(u) = h(x)$, $\tilde{h}(v) = h(y)$, and $y$ is a conjugate of $x$ over $k(h)$.

Now, $k(x, y)$ has either degree 4 or degree 2 over $k(u, v)$. We get the tower of fields $k(u) \subseteq k(x) \subseteq k(x, y) = K$ given by the defining equations $x + x^{-1} = u$ and $h(x) - h(y) = 0$ with the respective degrees $[k(x) : k(u)] = 2$ and $[K : k(x)] = \delta = \deg_K(x)$. Also we have $k(u) \subseteq k(u, v) \subseteq K$; the first extension is given by $H(u, v) = 0$ and has therefore degree $[k(u, v) : k(u)]$ equal to $n - 1 = d/2 - 1$, and the second is given by $x + x^{-1} = u$, $y + y^{-1} = v$.

If $[K : k(u, v)] = 4$, then a comparison of the degree $[K : k(u)]$ obtained from the two towers gives $\delta = d - 2$. Hence the numerator of

$$\frac{h(Y) - h(X)}{(Y - X)(YX - 1)}$$

is absolutely irreducible and is therefore the only choice as defining polynomial for $y$. As before we can use $0.75e < s < e/0.75$, and the fact that there are $s - 1$ Puiseux series of type 2 and $e - 1$ Puiseux series of type 1 at $x = 0$ in this irreducible factor of $h(X) - h(Y)$ to get the desired lower bound for $\deg_K(z)$.

We are left with the case when $[K : k(u, v)] = 2$, and we proceed to prove that this case cannot occur at all. Note that $k(x) = k(u, \sqrt{u^2 - 4})$ and $k(y) = k(v, \sqrt{v^2 - 4})$. Hence, $K = k(x, y) = k(u, v)(\sqrt{u^2 - 4}, \sqrt{v^2 - 4})$. Since none of the fields $k(x), k(y)$ is included in $k(u, v)$, we infer that none of $u^2 - 4$, $v^2 - 4$ is a square in $k(u, v)$ but the ratio $(v^2 - 4)/(u^2 - 4)$ is a square in $k(u, v)$.

We distinguish between two further cases.

- $\tilde{h}(2) \neq \tilde{h}(-2)$: Since $(v^2 - 4)/(u^2 - 4)$ is a square in $k(u, v)$, the function $u - 2$ must have even order at all the places of $k(u, v)$ lying above the place $u = 2$ of $k(u)$, except possibly for the places with $v = 2$ ($v = -2$ does not lie above $u = 2$ because $\tilde{h}(2) \neq \tilde{h}(-2)$). Since $H(U, V)$ is absolutely irreducible, the places of $k(u, v)$ above $u = 2$ correspond to the Puiseux series of $\tilde{h}(Z) - \tilde{h}(u)$, as series $Z = Z(u)$ centered at $u = 2$, where we disregard the 'trivial' series $Z = u = 2 + (u - 2)$.
  It is very easy to determine some features of these series: Let $z = \xi$ be a root of multiplicity $\mu_\xi$ of the equation $\tilde{h}(z) - \tilde{h}(2) = 0$; then the ramification index at $v = \xi$ above $u = 2$ is $\mu_\xi / \gcd(\mu_\xi, \mu_2)$ (see e.g. [8]). Since $(v^2 - 4)/(u^2 - 4)$ has even order at all places, this index has to be even for $\xi \neq 2$ (note that in the present case the value $\xi = -2$ does not appear). Hence in particular $\mu_\xi$ is even for $\xi \neq 2$. Note that $\mu_2$ cannot be even, for otherwise $\tilde{h}(Z) - \tilde{h}(2)$ would be a square, and $H(U, V)$ would be reducible, a contradiction. Similarly at $u = -2$. We deduce that

  $$\tilde{h}(Z) - \tilde{h}(\pm 2) = (Z \mp 2) Q_\pm^2(Z)$$

  for a suitable $Q_\pm \in k[Z]$. By [14, Lemma 4, p. 27] (applied with $q_1 = \xi_1 = 2$, $q_2 = \xi_2 = -2$), we have $\tilde{h}(Z) = \pm T_n(Z)$ where $T_n$ is the $n$th Chebyshev polynomial. But then again $H(U, V)$ would be reducible, which is not the case.
- $\tilde{h}(2) = \tilde{h}(-2)$: Now the same argument as above shows again that $\mu_\xi$ is an even integer for $\xi^2 \neq 4$. Also, putting $\mu_2 = \alpha_+ \rho$, $\mu_{-2} = \alpha_- \rho$, where $\rho = \gcd(\mu_2, \mu_{-2})$, it is easy to see, by looking at Puiseux series, that the ramification index of any place in $k(u, v)/k(u)$ with $v = -2$ above $u = 2$ is $\alpha_-$, and the order of $(v^2 - 4)/(u^2 - 4)$ at

such a place is $\alpha_-((\mu_2/\mu_{-2})-1)=\alpha_+-\alpha_-$. Hence the present assumption implies that $\alpha_-$ and $\alpha_+$ are odd and $\mu_2-\mu_{-2}$ is even.
Hence $\tilde{h}(u)-\tilde{h}(2)=(u^2-4)^m R^2(u)$ for a polynomial $R\in k[u]$ and an integer $m\in\{0,1\}$, whence $h(x)-h(1)=(x-x^{-1})^{2m}R^2(x+x^{-1})$ is a square $Q^2(x)$ in $k(x)$. This however contradicts condition (H2), because we would have the decomposition $h(x)=p(q(x))$ with $p(x)=x^2+h(1)$, $q(x)=Q(x)$. (Note that $\deg Q=\deg h/2\geq 32$.)

In conclusion, in the cases under consideration the degree $[K:k(u,v)]$ cannot be 2 but must be 4, and we are done as remarked above.

After all of this work we have merely achieved (8); however with this tool we are ready to give the proof of the statement in question. We study the equation $f(x)-f(y)=0$ and take the same partition coming from $\Sigma=\Sigma_L\cup\Sigma_S$ as before. Suppose that a vanishing minimal subsum of $f(x)-f(y)$ contains terms of type LL and LS, say $x^m y^n$ and $x^{m'}y^{n'}$. Then the ratio is $z=x^{m-m'}y^{n-n'}$, where $|n-n'|>(2/3^l)m_\ell$, and hence

$$\deg_K(z)\geq\frac{2}{63\cdot 3^\ell}m_\ell\delta,$$

where we have used (8). So by Lemma 2.2 and by using $2\mathfrak{g}-2+|S|\leq 2\delta^2$ and the inequality (4) we get

$$\deg f\leq m_\ell\leq\frac{63\cdot 3^\ell}{2}\binom{\ell^2-2}{2}2\delta\leq(2016\cdot 5^\ell)(d-1),$$

and similarly if two of the four sets contain terms from some minimal subsum. Hence, we may assume that each of the four sets is a union of vanishing subsums, so

$$\begin{aligned}a(x)b(y)&=a(y)b(x),\\ A(x)B(y)&=A(y)B(x),\\ A(x)b(y)&=B(x)a(y),\\ a(x)B(y)&=b(x)A(y).\end{aligned}$$

Suppose that some among $a(x)$, $A(x)$, $b(x)$, $B(x)$ vanish. We know $A(x)\neq 0$ (it contains the term $x^{m_\ell}$ with the maximal degree) and also $a(x)\neq 0$ (it contains the constant term). If $b(x)=0$, we have $B(x)a(y)=A(x)b(y)=0$, so $B(x)=0$, which is impossible. If $B(x)=0$, then we similarly conclude $b(x)=0$, which is again impossible. So $a(x)A(x)b(x)B(x)\neq 0$. We find

$$\frac{a(x)}{b(x)}=\frac{a(y)}{b(y)}=\frac{A(x)}{B(x)}=\frac{A(y)}{B(y)}=\varphi(h)$$

for some $\varphi\in k(x)$. Thus $a(x)=\varphi(h)b(x)$, $A(x)=\varphi(h)B(x)$ and therefore $a(x)+A(x)=\varphi(h)(b(x)+B(x))$, implying

$$f(x)=\frac{a(x)+A(x)}{b(x)+B(x)}=\varphi(h)=\frac{a(x)}{b(x)}.$$

By induction on $\ell$ we get $\deg\varphi \le 2016\cdot 5^{\ell-1}$ and so $\deg f = \deg\varphi\cdot d \le (2016\cdot 5^{\ell})(d-1)$. (The case $\ell = 2$ again follows, e.g. by following the arguments above and observing that the last case is impossible since $\Sigma$ then contains at least three elements.) Thus the statement is proved. □

*5.3. Proof of Theorem 1.4 in the case $h(0) = 0$*

This last part of the proof is exactly like the one for the non-Laurent case in Subsection 4.2 above: We just have to observe that $h(x) = x^e\tilde{p}(x)$, where $\tilde{p}$ is a non-constant polynomial, not vanishing at 0. The arguments therein do not require any modification. □

## 6. Proof of the Corollary

In this section we give the proof of the Corollary. We argue similarly to [6, Lemma 2] or [19].

Suppose that one among $q(x), h(q(x)), h(h(q(x)))$ is not of the forbidden shape (of the Main Theorem). Then we may apply the theorem with $h^{\circ(n-i)}$ in place of $g$ (with suitable $i \le 2$) and $h^{\circ i}\circ q$ in place of $h$. If $h^{\circ n}(q(x)) = P(x)/Q(x)$, where $P, Q$ are polynomials having altogether at most $\ell$ terms, then the conclusion of the Main Theorem gives $2016\cdot 5^{\ell} \ge \deg g = (\deg h)^{n-i} = d^{n-i} \ge d^{n-2}$, whence

$$\ell \ge \frac{1}{\log 5}((n-2)\log d - \log 2016),$$

proving the conclusion.

So let us assume that each among $q(x), h(q(x)), h(h(q(x)))$ is of the forbidden shape, i.e. $\lambda(ax^n + bx^{-n})$, $a, b \in k$, $n \in \mathbb{N}$, $\lambda \in \mathrm{PGL}_2(k)$.

Take this expression for $q(x)$ and assume first that $ab \ne 0$. On rescaling $x$, setting $x$ in place of a suitable power of it, and changing $h$ to $\lambda^{-1}\circ h\circ\lambda$, we may write $q(x) = x + x^{-1}$. (Note that these substitutions do not affect the conclusions.)

Then we have $h(x + x^{-1}) = \lambda_1(a_1x^r + b_1x^{-r})$ for $a_1, b_1 \in k$, $\lambda_1 \in \mathrm{PGL}_2(k)$ and some $r > 0$. Since the left side is invariant under $x \mapsto x^{-1}$, we must have $a_1 = b_1 \ne 0$, and then $r = d = \deg h$. Recall now that $T_d(x + x^{-1}) = x^d + x^{-d}$ for the Chebyshev polynomial $T_d$ of degree $d$, so $h(x) = \lambda_1(a_1T_d(x)) = (\lambda_1^*\circ T_d)(x)$ for a $\lambda_1^* \in \mathrm{PGL}_2(k)$.

Further, we have $h(h(q(x))) = \lambda_2(a_2x^m + b_2x^{-m})$ for $a_2, b_2 \in k$, $\lambda_2 \in \mathrm{PGL}_2(k)$, $m \in \mathbb{N}$. Again, $a_2 = b_2$, and $d$ is a divisor of $m$: $m = ld$ for a positive integer $l$. The above equation yields $h(\lambda_1^*(T_d)) = \lambda_2^*(T_m) = \lambda_2^*(T_{ld}) = \lambda_2^*(T_l(T_d)))$ for suitable $\lambda_2^* \in \mathrm{PGL}_2(k)$, where we have used the composition property $T_{rs} = T_r\circ T_s$ for positive integers $r, s$. Then $h\circ\lambda_1^* = \lambda_2^*\circ T_l$, whence $l = d$ and, on taking $\lambda_3 = (\lambda_1^*)^{-1}\circ\lambda_2^* \in \mathrm{PGL}_2(k)$, we get

$$T_d\circ\lambda_1^* = \lambda_3\circ T_d.$$

Now, $T_d$ is a polynomial of degree $d$. Since $d \ge 3$, $\infty$ is the unique totally ramified point of the map $x \mapsto T_d(x)$ and it follows that $\lambda_1^*, \lambda_3$ must fix $\infty$. We therefore invoke

[14, Lemma 5, p. 28] to conclude that $\lambda_1^*(x) = \pm x$, $\lambda_3(x) = (\pm 1)^d x$. This implies $\lambda_2^*(x) = (\pm 1)^{d+1} x$ and also $h(x) = (\pm 1)^d T_d(x)$.

Consider now the case when $ab = 0$ in the expression for $q$. Up to conjugation by an element in $\mathrm{PGL}_2(k)$ and on setting $x$ in place of a suitable power of it we may now assume that $q(x) = x$. Write then as above $h(x) = \lambda_1(a_1 x^d + b_1 x^{-d})$ with (new) $a_1, b_1 \in k$ and $\lambda_1 \in \mathrm{PGL}_2(k)$. If $a_1 b_1 \neq 0$, the arguments are as before. Then assume $a_1 b_1 = 0$, so $h(x) = \lambda_1^*(x^d)$ for a $\lambda_1^* \in \mathrm{PGL}_2(k)$. Similarly to the case treated above, we also obtain $\lambda_1^*(x)^d = \lambda_3(x^d)$, and the conclusion easily follows. □

## 7. Appendix: Composite factors of binomials

We prove here that the question posed at the end of the introduction (see Remark 1.6(iii)) has an affirmative answer in the special case of binomials. To recall this, let us assume that a binomial $x^{m'}(x^m + a)$ with $0 \leq m' < m$, $m, m' \in \mathbb{N}, a \in k$, is divisible by the composite $g(h(x))$ of two polynomials $g, h \in k[x]$, where as usual $\deg g, \deg h \geq 2$ and $h(x)$ is not of the shape $bx^n + c$, $b, c \in k, n \in \mathbb{N}$. We thus have the equation

$$x^{m'}(x^m + a) = q(x)g(h(x)), \quad q(x) \in k[x]. \tag{9}$$

We prove

**Theorem 7.1.** *Equation* (9) *implies* $\deg g \leq 24$.

*Proof.* If $a = 0$, then $g(h(x))$ is a power of $x$. Hence, for every root $\xi$ of $g$ we have $h(x) - \xi = bx^d$, where $b$ and $d$ are the leading coefficient and degree of $h$ respectively. Hence, $h$ is of the exceptional shape. Let us then assume that $a \neq 0$. By rescaling we may further assume $a = -1$, so the equation becomes

$$x^{m'}(x^m - 1) = q(x)g(h(x)).$$

We note at once that we may reduce to the case $h \notin k[x^s]$ for any $s > 1$. To prove this reduction we argue by induction on $\deg h$, similarly to what we did in connection with the previous results. Assuming $h(x) = h^*(x^s)$ for a polynomial $h^* \in k[x]$, we write $q(x) = q_0(x^s) + xq_1(x^s) + \cdots + x^{s-1}q_{s-1}(x^s)$ for polynomials $q_0, q_1, \ldots, q_{s-1} \in k[x]$. Note that $x^{m'}(x^m - 1) = \sum_{i=0}^{s-1} x^i q_i(x^s) g(h^*(x^s))$, where all the terms in the summand indexed by $i$ on the right have degree $\equiv i \pmod s$.

If $m \not\equiv 0 \pmod s$, then we conclude that $x^{m'}$ is of the shape $x^i q_i(x^s) g(h(x))$, so we fall in a case just excluded in the opening argument. Therefore we have $m \equiv 0 \pmod s$, and for some $i$, $x^{m'}(x^m - 1) = x^i q_i(x^s) g(h(x))$. Hence, $x^{(m'-i)/s}(x^{m/s} - 1) = q_i(x)g(h^*(x))$; this is an equation of the same type as before, with the same $g$ but lowered degrees. So eventually we reach a situation when $h \notin k(x^s)$ for any $s > 1$, which we shall assume from now on.

For each root $\xi$ of $g$, all the roots of $h(x) - \xi$ are either 0 or roots of unity. Here 0 may appear at most for one $\xi$, and the other roots are simple. Also, for all $\xi$ the set of non-zero roots, denoted by $S_\xi$, is non-empty. We have the equation

$$h(\theta) - h(\zeta) = 0$$

for all $\theta, \zeta \in S_\xi$.

Now, by a theorem of Beukers and Smyth [2] (see also [16] for the main argument of their proof), the number of pairs $(\theta, \zeta)$ of roots of unity which lie on an irreducible curve $f(\theta, \zeta) = 0$, when $f$ is not of the special shape $bx^n y^{n'} + c$ or $bx^n + cy^{n'}$, is bounded by $11(\deg f)^2$.

Let us detect the special factors dividing $h(x) - h(y)$. One factor is $x - y$. If another factor is of the above shape, we have $h(x^n) = h(cx^{n'})$ for some constant $c \in k^*$ and coprime exponents $n, n' \in \mathbb{Z}$. Necessarily $n = n'$, so we may assume $h(x) = h(cx)$. Then $c$ is a root of unity and $h(x) = h^*(x^e)$ for $h^* \in k[x]$ and an $e > 1$, a case which has been excluded above.

So there are no other such special factors, whence

$$|\{(\theta, \zeta) : \theta \neq \zeta,\ h(\theta) = h(\zeta),\ \theta, \zeta \text{ roots of unity}\}| \leq 11(\deg h)^2.$$

Hence,

$$\sum_{g(\xi)=0} |S_\xi|(|S_\xi| - 1) \leq 11(\deg h)^2.$$

On the other hand, since $|S_\xi| = \deg h$ for all but possibly one $\xi$, in which case $|S_\xi| \geq 1$, we have

$$\sum_{g(\xi)=0} |S_\xi| \geq \deg g \deg h - \deg h + 1 = \deg h(\deg g - 1) + 1.$$

Using the Cauchy–Schwarz inequality we get

$$\sigma^2 := \Big( \sum_{g(\xi)=0} |S_\xi| \Big)^2 \leq \deg g \sum_{g(\xi)=0} |S_\xi|^2,$$

and thus

$$11(\deg h)^2 \geq \sum_{g(\xi)=0} |S_\xi|(|S_\xi| - 1) \geq \frac{\sigma^2}{\deg g} - \sigma = \sigma\left(\frac{\sigma}{\deg g} - 1\right).$$

Furthermore, this implies

$$11(\deg h)^2 \geq \deg h(\deg g - 1)\left(\frac{\deg h(\deg g - 1)}{\deg g} + \frac{1}{\deg g} - 1\right).$$

Dividing by $(\deg h)^2$ and using $\deg g, \deg h \geq 2$, it follows that

$$11 \geq (\deg g - 1)\left(\frac{\deg g - 1}{\deg g} - \frac{1}{\deg h} + \frac{1}{\deg h \deg g}\right) \geq \frac{(\deg g - 1)^2}{2 \deg g} \geq \frac{\deg g}{2} - 1,$$

and finally $\deg g \leq 24$. □

**Remark 7.2.** We remark that there are certainly non-trivial (that is, with $\deg g > 1$) solutions to (9). For instance, observe that

$$x^{n(n-1)+1} - x = (x-1)^2 h(x)(h(x)-1) = (x-1)^2 g(h(x)),$$

with

$$h(x) = \frac{x^n - 1}{x - 1}, \quad g(x) = x(x-1).$$

This provides an example with $\deg g = 2$.

Most probably, however, the estimate $\deg g \leq 24$ may be sharpened, possibly to $\deg g \leq 2$, which would then be best possible. It is also possible that there are no solutions in which $\deg g \geq 2$ and $m' = 0$.

These questions amount to certain systems of equations to be solved in roots of unity. We plan to return to them in a future paper.

On the other hand, the general question of divisibility of an $\ell$-nomial by a composite factor $g(h(x))$ appears to be difficult and not in the range of the methods of the present paper.

*Acknowledgments.* This paper was partly written during a stay of the second author at ETH in October 2009; he is grateful for financial support and hospitality during the stay. Furthermore, both authors would like to thank M. Fried and P. Müller for very helpful discussions and references.

## References

[1] Avanzi, R., Zannier, U.: The equation $f(X) = f(Y)$ in rational functions $X = X(t)$, $Y = Y(t)$. Compos. Math. **139**, 263–295 (2003) Zbl 1050.14020 MR 2041613

[2] Beukers, F., Smyth, C. J.: Cyclotomic points on curves. In: Number Theory for the Millennium, I (Urbana, IL, 2000), A K Peters, Natick, MA, 67–85 (2002) Zbl 1029.11009 MR 1956219

[3] Bilu, Y., Tichy, R.: The Diophantine equation $f(x) = g(y)$. Acta Arith. **95**, 261–288 (2000) Zbl 0958.11049 MR 1793164

[4] Brownawell, D., Masser, D.: Vanishing sums in function fields. Math. Proc. Cambridge Philos. Soc. **100**, 427–434 (1986) Zbl 0612.10010 MR 0857720

[5] Dixon, J. D., Mortimer, B.: Permutation Groups. Grad. Texts in Math. 163, Springer, New York (1996) Zbl 0951.20001 MR 1409812

[6] Dvornicich, R., Zannier, U.: Cyclotomic Diophantine problems (Hilbert irreducibility and invariant sets for polynomial maps). Duke Math. J. **139**, 527–554 (2007) Zbl 1127.11040 MR 2350852

[7] Fried, M.: On a conjecture of Schur. Michigan Math. J. **17**, 41–55 (1970) Zbl 0169.37702 MR 0257033

[8] Fried, M.: Arithmetical properties of function fields. II. The generalized Schur problem. Acta Arith. **25**, 225–258 (1974) Zbl 0229.12020 MR 0444618

[9] Gabrielov, A., Vorobjov, N.: Complexity of computations with Pfaffian and Noetherian functions. In: Normal Forms, Bifurcations and Finiteness Problems in Differential Equations, NATO Sci. Ser. II Math. Phys. Chem. 137, Kluwer, Dordrecht, 211–250 (2004) MR 2083248

[10] Lang, S.: Introduction to Algebraic and Abelian Functions. 2nd ed., Grad. Texts in Math. 89, Springer, New York (1982) Zbl 0513.14024 MR 0681120

[11] Müller, P.: Permutation groups with a cyclic two-orbits subgroup and monodromy groups of Siegel functions. arXiv:math/0110060

[12] Müller, P.: Permutation groups with a cyclic two-orbits subgroup and monodromy groups of Laurent polynomials. Ann. Scuola Norm. Sup. Pisa **12** (2013), to appear

[13] Müller, P.: Cofinite integral Hilbert sets, Habilitationsschrift, Heidelberg (1999)

[14] Schinzel, A.: Polynomials with Special Regard to Reducibility. Encyclopedia Math. Appl. 77, Cambridge Univ. Press, Cambridge (2000) Zbl 0956.12001 MR 1770638

[15] Stichtenoth, H.: Algebraic Function Fields and Codes. Universitext, Springer, Berlin (1993) Zbl 0816.14011 MR 1251961

[16] Sturmfels, B.: Polynomial equations and convex polytopes. Amer. Math. Monthly **105**, 907–922 (1998) Zbl 0988.52021 MR 1656923

[17] Watt, S., Zieve, M.: Functional composition of symbolic Laurent polynomials. In preparation

[18] Zannier, U.: Some remarks on the S-unit equation in function fields. Acta Arith. **64**, 87–98 (1993) Zbl 0786.11019 MR 1220487

[19] Zannier, U.: On the number of terms of a composite polynomial. Acta Arith. **127**, 157–167 (2007) Zbl 1161.11003 MR 2289981

[20] Zannier, U.: On composite lacunary polynomials and the proof of a conjecture of Schinzel. Invent. Math. **174**, 127–138 (2008) Zbl 1177.12004 MR 2430978

[21] Zannier, U.: Addendum to the paper: On the number of terms of a composite polynomial. Acta Arith. **140**, 93–99 (2009) Zbl 1161.11003 MR 2557855